



Rint  
Réseau  
international  
de néologie  
et de terminologie

14

Revue semestrielle  
coéditée par l'Agence de  
coopération culturelle et  
technique et la  
Communauté française de  
Belgique.

N° 14 - décembre 1995.

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



# Terminologies nouvelles

## Le Rint

Présentation,  
*par Louis-Jean Rousseau.*  
Page 2

Le Rint, bilan et perspectives,  
*par Louis-Jean Rousseau.*  
Page 3

## Afrique

La terminologie dans le contexte de  
l'aménagement linguistique de la  
République centrafricaine,  
*par Michel Koyt.*  
Page 6

## Canada

L'utilisation de la bitextualité pour la  
validation des traductions :  
le système *Transcheck*,  
*par Elliott Macklovitch.*  
Page 10

*TransSearch* : un concordancier  
bilingue pour la traduction et la  
terminologie,  
*par Pierre Isabelle.*  
Page 21

## Communauté française de Belgique

La base de données liégeoise du  
*Robert & Collins* : un outil pour le  
lexicographe, le linguiste ou le  
traducteur,  
*par Thierry Fontenelle.*  
Page 24

## France

*Trader*, aux frontières du néologisme,  
*par Chantal Girardin.*  
Page 31

Analyse de textes spécialisés pour le  
recueil de données terminologiques,  
*par Anne Condamines.*  
Page 35

Pour une analyse linguistique des  
termes dans la phrase,  
*par Jacques Boissy.*  
Page 43

*Balnéo* : un projet de réseau  
informatique pour la veille  
néologique,  
*par André Le Meur et Loïc Depecker.*  
Page 48

## Québec

Une autre façon de mesurer  
l'utilisation du français dans les  
entreprises,  
*par Carl A. Teixeira.*  
Page 54

## Suisse

La terminologie dans les pays  
germanophones : vent nouveau sur la  
collaboration,  
*par Rolf Moos.*  
Page 63

## Union latine

Vers la planification linguistique du  
portugais : Portugal, Brésil, Afrique.

Quelques réflexions,  
*par Enilde Faulstich.*  
Page 66

Développement harmonisé et  
distribution des ressources  
terminologiques : le Réseau panlatin  
de terminologie - Realiter,  
*par Giovanni Adamo.*  
Page 77

Principes méthodologiques du travail  
terminologique au sein du Réseau de  
terminologie - Realiter,  
*par Louis-Jean Rousseau.*  
Page 82

## Terminotique

Gestion assistée des liens entre  
concepts,  
*par Caroline de Schaetzen.*  
Page 85

## Bibliographie de la néologie

Rubrique préparée par  
*John Humbley.*  
Page 92

## Inventaire

Inventaire des travaux en cours et des  
projets de terminologie.  
Page 94

## En bref

Publications.  
Page 99

Congrès, colloques, séminaires.  
Page 102

# Sommaire

## Présentation

**C**e nouveau numéro de *Terminologies nouvelles* est, par la diversité de son contenu, à l'image du Rint dont les préoccupations embrassent toutes les composantes de la terminologie et de l'aménagement de la terminologie, comme le démontre d'ailleurs le bilan qui ouvre ce numéro.

La terminologie, en tant que champ de connaissance et d'expérimentation, s'élargit : en témoignent les articles sur l'aménagement terminologique en Afrique, sur la mesure de l'utilisation de la terminologie française et sur le recours aux outils informatisés pour le traitement des technolectes.

Par ailleurs, les textes présentés ici sous la bannière du module suisse et sous celle de l'Union latine illustrent bien l'ouverture du Rint sur d'autres pratiques qui ne sont pas étrangères à ses propres travaux. Ainsi, les travaux sur l'établissement des principes méthodologiques du travail terminologique du Réseau panlatin de terminologie (Realiter) sont dans la droite ligne des réflexions entreprises par le Rint depuis la fin des années 80.

À compter du numéro 15, des changements seront apportés à la présentation de *Terminologies nouvelles*. Ainsi, les articles ne seront plus regroupés par modules, de façon à

rendre plus évidente l'ouverture de la revue à toutes les contributions, quelle qu'en soit l'origine géographique ou institutionnelle. De plus, tous les numéros de *Terminologies nouvelles* seront thématiques. Le numéro 15 portera sur les banques de terminologie et le numéro 16 sera consacré à la dialectologie terminologique. Enfin, une nouvelle chronique traitant des questions de normalisation sera créée.

Nous souhaitons par ces innovations élargir le bassin des auteurs de la revue et intéresser un public plus étendu par la couverture d'un éventail plus grand de sujets.

*Le secrétaire général du Rint,  
Louis-Jean Rousseau.*

# Le Rint

# Le Rint, bilan et perspectives

## 1 Un bilan riche et varié

**A**u cours des toutes dernières années, le Rint a, par ses travaux, consolidé ses acquis et préparé la mise en œuvre d'un plan d'intervention majeur en matière de développement terminologique. En effet, depuis 1989, le Rint s'est donné, grâce à une série de rencontres scientifiques et à ses travaux de recherche, des assises théoriques et méthodologiques solides faisant la synthèse des différentes expertises présentes au sein du Réseau. Les résultats de ces travaux ont été pour la plupart publiés dans la revue *Terminologies nouvelles*, dont le numéro 12, consacré à l'implantation des terminologies, a été récemment publié. Le Rint est en effet un rassemblement d'organismes chargés de définir et de mettre en œuvre des politiques linguistiques, des centres de recherche en terminologie et des universités. Toutes les facettes de la terminologie y sont représentées, qu'il s'agisse de la recherche théorique et appliquée, de la production et de la diffusion de la terminologie, de la traduction spécialisée et de l'aménagement linguistique dans son ensemble.

Pendant cette période, le Rint s'est également préoccupé du développement d'outils informatisés pour le travail terminologique tels le système *Nomino* (anciennement *Termino*) pour le dépouillement terminologique assisté par ordinateur,

et le système *Balnéo* pour l'échange par Internet des données terminologiques.

De plus, le Rint a déjà à son crédit l'élaboration et la publication de vocabulaires spécialisés tant en français que dans les langues partenaires. Ainsi, à titre d'exemples, le Rint travaille actuellement à la réalisation des ouvrages suivants : *Dictionnaire de l'eau* (anglais-français), *Vocabulaire du tourisme* (malgache-français), *Lexique informatique* (arabe-français-anglais), *Vocabulaire de la mécatronique* (anglais-français), *Vocabulaire des élections* (français-yoruba-fon-kanuri-tamajàk-pulaar-bambara-wolof). Ces travaux terminologiques ont permis la validation des principes méthodologiques adoptés par le Rint en 1990 et l'expérimentation de la coopération en matière de terminologie multilingue.

Le Rint compte maintenant 20 pays ou États membres répartis en 12 modules. De plus, la participation de l'Union latine aux travaux du Rint permet l'établissement de liens fonctionnels avec des organismes à vocation linguistique et terminologique dans un grand nombre de pays de langue latine. Le soutien de l'Agence de coopération culturelle et technique a permis au Rint de devenir un acteur important en matière d'aménagement linguistique. Le Rint s'apprête maintenant à relever des défis de taille en matière de développement terminologique en vue de résultats concrets répondant aux attentes des pays membres.

## 2 Perspectives pour 1996-1997

Pour 1996-1997 et sans doute pour les années subséquentes, le Rint s'est donné les priorités suivantes.

### 2.1 Développement terminologique et banques de terminologie

Ce programme correspond aux principaux objectifs et même à la raison d'être du Rint. Ce programme mobilisateur, qui donnera un contenu terminologique aux autoroutes de l'information, sera déployé en quatre volets : inventaire des ressources terminologiques; veille terminologique et néologique; production et publication de dictionnaires terminologiques; banques de terminologie.

#### 2.1.1 Inventaire des ressources terminologiques

Depuis sa création, le Rint dresse un inventaire permanent des travaux terminologiques publiés et des travaux en cours. En 1995, nous avons ajouté à cet inventaire une enquête sur les banques de terminologie dont le français est une langue de travail. Ce travail se poursuivra dans les années à venir afin de pouvoir diffuser une information complète et à jour sur les ressources disponibles. Cet inventaire est un outil de diffusion et un outil de gestion dans la mesure où il permet de mieux coordonner les travaux en évitant de faire double emploi et en permettant d'établir des collaborations à l'échelle internationale.

#### 2.1.2 Veille terminologique et néologique

Ce volet consiste à organiser la collecte et l'échange de matériaux terminologiques, surtout en ce qui a

trait aux termes nouveaux, en français ou dans les langues partenaires, de façon à accélérer la mise à jour des banques de terminologie et des dictionnaires spécialisés. Des économies importantes peuvent être réalisées dans l'élaboration des terminologies par la mise en commun des matériaux terminologiques dans la francophonie. Il est prévu de mettre en œuvre dans ce projet des produits des industries de la langue développés spécialement pour les travaux du Rint tels le système *Adepte-Nomino* (dépouillement terminologique automatisé des banques de textes), les banques de terminologie comme corpus d'exclusion, le système *Balnéo*, mis au point par le Rint pour assurer l'échange des données terminologiques par Internet (voir l'article consacré à ce logiciel). Les travaux porteront sur les domaines correspondant aux besoins prioritaires des pays participants. Ce système d'échange sera ouvert, selon certaines balises à un grand nombre de participants, tant dans la francophonie que dans les organisations internationales. En 1996, le Rint doit terminer la mise au point et l'expérimentation de *Balnéo*, et assurer son implantation, y compris dans les pays du Sud, en fonction de l'accessibilité au Réseau Internet.

#### 2.1.3 Production et diffusion de vocabulaires

La production de dictionnaires terminologiques, qui répond à un besoin de diffusion du vocabulaire scientifique et technique, est une suite logique des travaux d'inventaire et de veille évoqués ci-dessus. Ces produits peuvent être diffusés sur support électronique et sous forme d'imprimés, selon les besoins et les situations de diffusion. Le Rint a déjà publié un certain nombre de dictionnaires et d'autres sont présentement en chantier. Au cours des deux dernières années, plusieurs projets ont été entrepris par les pays du Sud. Ces

projets multilingues sont la plupart du temps le fruit d'un travail de coopération entre plusieurs pays en réponse à des besoins communs.

#### 2.1.4 Banques de terminologie

Plusieurs banques de terminologie sont représentées au sein du Rint et de plus tous les modules sont des utilisateurs des banques de terminologie. Le Rint porte un intérêt croissant à la question des banques. L'inventaire des banques de terminologie en cours d'élaboration démontre qu'à côté des grandes banques de terminologie qui traitent la plupart des domaines, il existe un foisonnement de petites banques spécialisées. Cette situation qui va croissant rend impossible l'idée de la fusion des banques en une seule. Le Rint est d'avis que ces banques sont complémentaires et que les usagers souhaitent avoir accès à toutes les banques susceptibles de répondre à leurs besoins. Les deux questions d'actualité sont celles de l'accès aux banques, notamment par le biais des autoroutes de l'information, et de la mise à jour de ces banques. Le plan de développement terminologique du Rint est susceptible d'apporter une réponse à la mise à jour des banques.

Pour faire le point sur la question des banques, le Rint a tenu à Québec en janvier 1996 une table ronde internationale (les actes de cette table ronde seront publiés dans le numéro 15 de *Terminologies nouvelles*) au cours de laquelle ont été débattues les questions des droits d'auteur, de la gestion et du développement des banques, de leur diffusion, des développements technologiques et des perspectives de collaboration entre les banques. Le Rint souhaitait par cette initiative créer un forum permanent des banques de terminologie, de la francophonie et des autres aires linguistiques, de façon à rendre concrète cette volonté de coopération et à rassembler toutes les expertises

existant en cette matière. Des rencontres biennales entre exploitants et usagers des banques de terminologie seront ainsi envisagées.

Par ailleurs, le Rint a entrepris une étude sur la perspective de la création de banques de terminologie en Afrique, afin de soutenir les nombreux travaux d'aménagement terminologique menés dans plusieurs pays.

## 2.2 Formation

La mise en place de projets de développement terminologique suscite des besoins de formation continue. Une enquête menée au sein du Rint a mis en relief la nature de ces besoins. Il s'agit principalement de formation pratique s'adressant à des spécialistes engagés dans des travaux d'aménagement linguistique et donc ayant déjà une formation de base. Les thématiques souhaitées sont les suivantes : gestion des programmes d'aménagement terminologique, perfectionnement en matière de travaux terminologiques, notamment formation orientée sur les langues partenaires, utilisation des outils terminotiques et informatiques. En 1995, le Rint et l'École internationale de Bordeaux ont tenu un stage sur l'informatisation des travaux terminologiques en langues africaines. Le Comité d'orientation souhaite inscrire chaque année au programme du Rint une activité de formation. Cette activité pourrait s'organiser en collaboration avec l'ACCT et toute autre organisation intéressée. Au cours du prochain biennium, il conviendrait de tenir un stage sur les questions méthodologiques, notamment sur la veille terminologique et l'utilisation du système *Balnéo* et un autre stage sur l'informatisation des travaux terminologiques portant sur les langues partenaires.

## 2.3 Terminotique

La mise sur pied de programmes de développement terminologique rend nécessaire, pour des raisons d'efficacité, le recours à des outils informatiques adaptés à la production, l'échange et la diffusion des données terminologiques. Le Rint a investi dans le développement d'outils adaptés et, compte tenu de l'importance des outils informatisés et de la complexité de ces sujets, il a été jugé nécessaire de créer au sein du Réseau un groupe de travail terminotique. Ce groupe de travail a pour mandat de favoriser la conception, le développement, l'évaluation et l'utilisation au sein du Rint de produits et d'outils informatiques adaptés aux besoins du Réseau. Le groupe de travail mènera notamment un inventaire permanent des produits terminotiques intéressant le Rint.

## 2.4 Normalisation

L'importance grandissante de la normalisation et plus particulièrement de la normalisation terminologique et de la normalisation dans les technologies de l'information risque d'avoir une incidence importante sur le statut du français. Il devient, dans cette perspective, absolument essentiel que le Rint s'intéresse aux aspects terminologiques de cette question. Le Rint souhaite établir des liens avec les organisations internationales de normalisation. Des actions en ce sens s'inscriront dans la programmation du Rint au cours des années prochaines.

*Le secrétaire général du Rint,  
Louis-Jean Rousseau.*

# La terminologie dans le contexte de l'aménagement linguistique de la République centrafricaine

**L'**accession à la souveraineté nationale et internationale de la République centrafricaine a été assortie d'un mouvement de revendication culturelle qui a abouti dès 1963 à la reconnaissance officielle du sängö et à son adoption comme langue nationale de la jeune république. Depuis lors, l'État et la société civile s'emploient sans relâche à la valorisation de cette langue qui a fini par être un instrument pour le développement économique et social.

Cette valorisation passe nécessairement par deux voies, dont la première consiste en l'aménagement du statut de la langue. Celle-ci s'est réalisée à pas lents, mesurés, voire prudents. La deuxième vise le corpus: élaboration des grammaires, aménagement terminologique, normalisation, implantation; autrement, toutes choses qui nécessitent des ressources humaines, un cadre de travail, une planification, un suivi, une évaluation.

Dans le cas centrafricain, il s'agit d'un changement linguistique naturel, mû par la volonté de la base dans un contexte où tout était à créer. Mais une volonté populaire tenace devait aider au développement d'une activité lexicographique et enfin terminologique, vers un changement linguistique dit planifié.

## 1 Les premiers jalons de l'aménagement linguistique

### 1.1 Les travaux de première génération.

La volonté de promouvoir et de valoriser le sängö a conduit les gouvernements centrafricains qui se sont succédé à adopter divers textes à caractère législatif et réglementaire. Le décret portant création d'une commission nationale pour l'étude du sängö, pris en 1965, fut un événement marquant. Cette commission avait eu pour mission de proposer une orthographe du sängö et d'élaborer un dictionnaire. Ainsi, l'État donnait officiellement le ton pour la recherche lexicographique. Quelques années plus tard, en 1974, un autre texte devait porter création de l'Institut pédagogique national, dont une des missions avait été de préparer l'introduction du sängö dans l'enseignement. Dès lors, furent engagés des travaux de recherche terminologique en vue de l'élaboration de lexiques sängö des diverses disciplines scolaires, ainsi que pour l'alphabétisation des adultes.

Un champ plus vaste de recherche terminologique fut ouvert en 1976 dans l'optique de l'enrichissement et la modernisation du stock lexical du sängö afin de le rendre apte à exprimer les réalités contemporaines. Suite à ces travaux,

# Afrique

deux lexiques spécialisés avaient été publiés en 1983 et 1984. Il faut cependant reconnaître qu'il s'agissait là de traductions de termes à partir d'une liste d'items préétablie. L'objectif avait été d'attribuer à chaque item un équivalent dans la langue d'arrivée. Ces travaux dits de première génération avaient été réalisés par du personnel non qualifié, rapidement initié aux techniques et procédures de créations lexicales, travaillant essentiellement sur des fiches terminologiques simplifiées, limitées à l'entrée, l'équivalent et la catégorie grammaticale.

L'apport du secteur privé et particulièrement des Églises fut remarquable. En effet, animées par le souci de mettre la sainte parole à la portée de leurs ouailles, les Églises ont procédé à la traduction des Saintes Écritures dans nombre de langues locales, ainsi qu'en sängö. Il en est résulté une production terminologique du domaine liturgique, découlant d'une pratique de recherche s'apparentant à maints égards à la terminologie traductionnelle.

Les contributions des chercheurs indépendants furent significatives. Il est aisé de citer le dictionnaire sängö-français de Luc Boucquiaux (Boucquiaux 1978) ainsi que les nombreux essais de créations terminologiques dus à Marcel Diki-Kidiri. Les efforts de création lexicale des journalistes centrafricains, et particulièrement de Jean Willibiro Passy, ont constitué des apports non négligeables qui ont marqué le début de l'activité terminologique en Centrafrique.

## 1.2. Les travaux de deuxième génération

La signature d'un contrat de coopération linguistique entre l'État centrafricain et l'Agence de coopération culturelle et technique

(ACCT) apporta une impulsion nouvelle à la recherche linguistique et terminologique centrafricaine. Le gouvernement désigna cette fois-ci l'Institut de linguistique appliquée de l'Université de Bangui comme l'agent d'exécution du contrat. Cet institut reçut par ce fait même une nouvelle mission qui consistait à réaliser les programmes de recherche en vue de la promotion du sängö et de son adaptation à l'expression des réalités du monde moderne. Ceux-ci couvraient plusieurs domaines, à savoir la grammaire, la sociolinguistique, la lexicographie et la terminologie. Ces derniers visent l'élaboration d'un dictionnaire monolingue, d'un dictionnaire orthographique ainsi que de nombreux lexiques spécialisés.

Pour s'acquitter de sa mission, l'institut mit au point une orthographe officielle du sängö, en collaboration avec le centre Lacito<sup>(1)</sup> du CNRS. Cette orthographe fut adoptée par décret présidentiel. L'institut expérimenta une méthodologie de recherche calquée sur le modèle nord-américain, particulièrement celui de l'Office de la langue française. Elle s'en écarte cependant par le fait qu'il privilégie les secteurs d'activités plutôt que des domaines. Exploitant judicieusement les ressources néologiques de la langue, les terminologues attachés à ces programmes ont produit une masse impressionnante de termes techniques. L'apparition de l'informatique dans les travaux a permis d'opérer le passage de la fiche sur support papier à la fiche électronique, rendant plus aisé le travail de terminologue. Aussi, une importante base de données textuelles fut-elle réunie à partir des textes en tous genres écrits en sängö.

Une émission radio-diffusée sur le bon usage en sängö, fort appréciée par la population, a permis tant soit peu de faire reculer certains préjugés tenaces et de rompre certaines

barrières de résistance à la promotion de l'usage du sängö écrit dans la vie publique nationale et à son introduction dans les systèmes éducatifs.

## 2 Le tournant décisif

### 2.1 La promotion du statut du sängö

La promotion du sängö au statut de langue officielle conjointement au français, intervenue en 1991, marquera un tournant décisif dans la politique linguistique du pays. Une politique manifeste d'aménagement linguistique systématique s'est fait jour. Ainsi, en 1992, après de longs débats, fut créé le Conseil national d'aménagement linguistique (CNAL), dont la mission est de définir, proposer et conduire la politique linguistique centrafricaine et de veiller à ce que le sängö devienne une langue de communication écrite et orale dans la vie publique nationale.

Le conseil adopta une nouvelle méthode de recherche terminologique et phraséologique s'appuyant sur le dépouillement des ethnotextes. Cette méthode constitue le cadre de référence pour l'élaboration des vocabulaires et lexiques spécialisés bilingues. De même, le Conseil créa la banque de données terminologiques centrafricaines (Goleba). Cette banque de données informatisées est alimentée par une vingtaine de terminologues permanents travaillant sur les vocabulaires des différents secteurs d'activités. Le personnel vacataire est constitué pour la plupart d'étudiants finalistes.

(1) Langues et civilisations à traditions orales.

Le secteur privé emboîta le pas au Conseil national. Des organisations non gouvernementales et institutions privées ont participé à l'élaboration des vocabulaires techniques, à la traduction et à la promotion littéraire sängö.

## 2.2 Les obstacles à franchir

Mais à ce jour, la production terminologique reste encore très loin de satisfaire les besoins terminologiques de la nation. De plus, cette production ne connaît qu'une diffusion limitée; le contexte ne permettant pas l'utilisation du vocabulaire technique en question, vu que toute l'activité administrative et industrielle fonctionne en français. Enfin, toutes les situations formelles utilisent le français même à l'oral. Le recours au sängö n'intervient que lorsque le besoin se fait sentir de se faire comprendre par le peuple.

## 3 Vers un aménagement linguistique planifié

### 3.1 Le contexte

Les enquêtes sociolinguistiques menées ces dernières années ont fait prévaloir des attitudes positives envers la promotion du sängö. Cependant, elles semblent révéler que l'action d'aménagement reste timide et n'avance qu'à petits pas. En effet, il est ressorti des enquêtes que le sängö et le français se sont partagé les situations : le français apparaît autant à l'écrit qu'à l'oral, tandis que le sängö est confiné au seul domaine de l'oral. Le défi à relever est de faire passer l'usage du sängö au domaine de l'écrit. On comprend aisément que cette langue doit voir son stock lexical enrichi et modernisé afin de lui permettre d'assumer le rôle qui lui est dévolu désormais.

### 3.2 Les options politiques

L'avènement de la démocratie en République centrafricaine a entraîné le besoin de l'instauration d'un dialogue entre le peuple et les élus, entre les autorités de l'État et le peuple et vice versa. Cette perspective nouvelle ouvre la voie à un changement linguistique planifié plus rigoureux devant permettre l'utilisation du sängö écrit dans la vie publique, conformément à la politique de bilinguisme convivial instituée par la Constitution de décembre 1994 qui proclame : «Les deux langues officielles de l'État sont le sängö et le français».

Partant de ce principe de base, cette même Constitution affirme en son article 57 : «les règles relatives au plan d'aménagement et d'implantation progressive et généralisée du sängö ressortissent au domaine de la loi».

La loi d'orientation sur l'éducation et la formation, quant à elle, énonce : «Le sängö et le français sont les deux langues d'enseignement (...) Une politique rationnelle d'utilisation des deux langues officielles dans les services de l'État devra être promue pour soutenir l'usage de ces deux langues dans l'enseignement (...) Un décret précisera avant 1996 le plan d'introduction du sängö dans l'enseignement ainsi que les modalités d'utilisation équilibrée et harmonieuse des deux langues officielles dans l'enseignement».

Compte tenu de ces orientations politiques, le Conseil national s'est fixé quatre grandes lignes de recherche terminologique.

Le premier domaine vise des travaux terminologiques des secteurs scientifiques, techniques et technologiques devant permettre la préparation des ouvrages scolaires en vue de l'enseignement en sängö.

Le second domaine couvre les terminologies des activités des

secteurs économiques et financiers en vue de répondre au besoin du monde de la production (secteur agro-pastoral, projet de développement, banque, commerce, assurance...).

Le troisième champ couvre les terminologies des secteurs d'activités administratives, la législation du travail, les conventions collectives, les activités judiciaires, les voiries municipales.

Le quatrième champ couvre les terminologies des secteurs de pointe, particulièrement le domaine de l'informatique et ses applications (bureautique).

Un partenariat avec le secteur privé et l'Université s'avère nécessaire pour relever le défi de l'aménagement terminologique.

### 3.3 La normalisation et l'implantation

Au stade actuel, la définition d'une norme du sängö est à l'ordre du jour. La codification de la norme orthographique sera faite avec la publication prochaine du dictionnaire orthographique sängö (Dior) dont l'élaboration est achevée depuis peu. La codification lexicale, quant à elle, est chose faite, grâce à l'achèvement du dictionnaire sängö dont la publication par le Conseil est prévue pour l'année 1996.

Les comités des langues sängö et de langue nationale et média créés par arrêté en 1984 assurent leur mission de création terminologique pour le premier et de normalisation pour le second.

L'élaboration et l'adoption du plan d'aménagement et d'implantation progressive du sängö dans la vie publique nationale donnera l'occasion de fixer les organes chargés de son exécution et de sa mise en œuvre. La tendance qui se dessine est celle de la mise en place d'une commission nationale de terminologie au niveau du Conseil,

composée de personnalités dont les compétences sont reconnues dans les domaines et des représentants de la société civile et des secteurs économiques. Mais comme tout changement linguistique est une œuvre de longue haleine, le Conseil a arrêté une stratégie qui lui permet d'opérer un changement durable. En effet, le Conseil articule son action principalement autour de l'introduction du sängö dans les systèmes éducatifs. Cette stratégie qui requiert du temps pour porter ses fruits présente l'avantage comparatif de former des citoyens nouveaux, capables de travailler dans les deux langues officielles du pays.

L'œuvre d'implantation qui s'appuie sur le système éducatif gagnera également à être accompagnée d'une action d'affichage public en sängö, de rédaction de formulaires bilingues, d'élaboration d'actes de naissance bilingues, de cartes d'invitation bilingues, etc. afin de créer un environnement social bilingue. Les médias nationaux, les confessions religieuses et la musique profane ont donné le ton depuis bien longtemps déjà.

*Michel Marie Koyt,  
Conseil national d'aménagement  
linguistique,  
Bangui,  
République centrafricaine.*

# L'utilisation de la bitextualité pour la validation des traductions : le système *TransCheck* <sup>(1)</sup>

Cet article décrit la première version prototype de *TransCheck*, système qui permet de détecter automatiquement certains types d'erreurs de traduction et qui est basé sur la notion de bitexte, ou corpus composé de textes sources, alignés avec leur traduction.

Nous y présentons également l'analyse des résultats de traductions publiées, la description de certains des problèmes qui échappent actuellement au système, ainsi que les grandes lignes des projets de développement du premier prototype de *TransCheck*.

Termes-clés : traduction; détection d'erreur; alignement de corpus; bitexte; français; anglais.

(1) Article déjà paru en 1995 dans *L'Actualité terminologique*, vol. 28, n°3, publication du Bureau de la traduction de Travaux publics et services gouvernementaux Canada, et reproduit avec la permission des éditeurs.

(2) À leur décharge, il faut dire que ces traducteurs travaillent dans un climat de grande tension, car ils doivent traduire les débats parlementaires du jour même pour qu'ils puissent être publiés le lendemain matin. Dans ces conditions, et compte tenu de la quantité de textes que nous avons examinés, ce qui est réellement étonnant, c'est la rareté des erreurs relevées dans le *Journal des débats*.

(3) À part les exemples (xv) et (xvi). Voir les notes 10 et 14 plus loin.

(4) Pour plus de détails sur l'algorithme d'alignement utilisé actuellement au Citi, et qui présente un taux de succès de plus de 98%, voir Simard *et al.* (1992).

## 1 Détecteur d'erreurs

Comme tous les gens qui rédigent beaucoup, les traducteurs passent généralement leurs traductions au correcteur orthographique ou stylistique avant de les livrer, afin de déceler les erreurs simples qui auraient pu leur échapper. Étant unilingues, ces outils rédactionnels sont cependant tout à fait incapables de détecter les erreurs de traduction même les plus flagrantes, car ces erreurs sont bilingues de nature et dépendent des relations qui existent entre deux textes de langues différentes. Cet article décrit un nouveau genre d'outil rédactionnel expressément conçu pour détecter les erreurs comme celles qui sont présentées au tableau n°1, que l'on trouvera à la fin de l'article.

Tous les exemples énumérés au tableau n°1 sont authentiques : ils sont tirés du *Journal des débats* de la Chambre des communes du Canada, de sorte que les erreurs qu'ils contiennent ont en réalité été commises par des traducteurs qui sont parmi les plus compétents au Canada <sup>(2)</sup>.

Les erreurs énumérées au tableau n°1 ont été détectées automatiquement par le système de vérification de traduction applé

*TransCheck* <sup>(3)</sup>, qui fait partie d'une nouvelle famille d'aides à la traduction mises au point au Centre d'innovation en technologies de l'information (Citi) (Isabelle *et al.* 1993). Ces outils sont tous basés sur des algorithmes d'alignement bitextuel comme ceux qui ont été proposés initialement par Gale et Church (1991), et par Brown *et al.* (1991). Que font ces algorithmes ? En deux mots, ils calculent avec précision les liens entre les segments correspondants d'un texte source et de sa traduction, sans égard à la longueur des textes <sup>(4)</sup>. Suivant la terminologie utilisée par Harris (1988), un corpus de textes sources/cibles, alignés de cette façon, est souvent appelé un bitexte. Dans les travaux que nous venons de citer, les segments alignés d'un bitexte sont généralement des phrases, mais il est également possible, en principe, d'établir des liens entre des unités plus longues ou plus courtes. Bien sûr, une phrase du texte source peut être explicitement reliée à deux ou plusieurs phrases du texte cible, et vice-versa.

Dans plusieurs publications récentes du Citi – notamment Isabelle *et al.* (1993), Isabelle (1992) et Macklovitch (1992) –, on a avancé que le concept de bitextualité permet le développement d'une toute nouvelle génération d'aides à la traduction, dont la plus évidente est peut-être le concordancier bilingue,

# Canada

comme le système *TransSearch* mis au point au Citi. Un programme de concordance standard est une sorte de système d'interrogation de base de données qui recherche, extrait et affiche dans leur contexte toutes les occurrences d'un terme ou d'une expression contenues dans une base de données textuelles. Ce qu'un concordancier bilingue comme *TransSearch* fait, en outre, c'est afficher, en regard de chacune de ces occurrences, la ou les phrases qui en constituent la traduction, selon les calculs préalables des programmes d'alignement. Ce simple (du point de vue conceptuel) couplage de segments en langue de départ et en langue d'arrivée présente cependant d'énormes avantages : il permet aux traducteurs d'accéder facilement, au moyen de l'élégante interface graphique de *TransSearch*, à toutes les richesses enfouies dans leur production antérieure. Et, comme l'a justement souligné Pierre Isabelle (1992), «la masse des traductions produites chaque année contient infiniment plus de solutions à plus de problèmes que tous les outils de référence existants et imaginables».

Les requêtes que l'utilisateur peut soumettre à *TransSearch* peuvent porter sur des mots simples ou des expressions complexes, et elles peuvent être unilingues ou bilingues. Par exemple, l'utilisateur peut demander au système d'afficher toutes les phrases de sa base de données qui renferment le terme *gridlock*, ou toutes les paires alignées dans lesquelles la phrase anglaise contient le terme *librairie*. Dans ce cas, les occurrences repérées par le système constitueraient des erreurs de traduction, car *library* et *librairie* sont de faux amis : bien qu'ils soient étymologiquement apparentés et encore semblables morphologiquement, ces deux termes ne peuvent jamais être utilisés comme des équivalents traductionnels, car ils ont maintenant des sens

complètement différents (*librairie* correspond au terme anglais *bookstore*).

Supposons qu'on compile une liste exhaustive de ces faux amis. Supposons en outre qu'on puisse aligner automatiquement une ébauche de traduction avec son texte source et qu'on dispose aussi d'une procédure automatique permettant d'appliquer cette liste de faux amis au bitexte ainsi obtenu, en tant que requêtes en lot soumises à *TransSearch*. Supposons enfin qu'on dispose d'une interface nous permettant d'examiner et d'éditer les résultats de ces requêtes. Ce qu'on obtiendrait alors, en fait, c'est un simple vérificateur de traduction. Simple, en ce sens qu'il pourrait vérifier un premier jet de traduction pour y détecter un type évident d'erreurs de traduction, c'est-à-dire les faux amis. Et en fait, c'est exactement de cette façon que *TransCheck* a d'abord été conçu et développé au Citi – comme une extension de notre concordancier bilingue.

## 2 Erreurs de traduction

Les faux amis complets comme *library/librairie* constituent un bon point de départ pour le développement d'un vérificateur de traduction car, par définition, les mots qui constituent ces paires ne peuvent jamais être des traductions réciproques. Mais ce n'est pas là le seul critère d'inclusion dans un vérificateur de traduction; si ce l'était, le lexique bilingue du programme (ou anti-lexique, comme il serait plus juste d'appeler une base de données de couples illicites) contiendrait un nombre énorme d'entrées. Il est évident que le nombre de mots qui ne peuvent jamais être des équivalents traductionnels réciproques est infiniment plus grand que le nombre de paires de mots qui peuvent l'être.

Pour n'en donner qu'un exemple farfelu, nous sommes tout à fait certains que le mot anglais *very* ne peut jamais être traduit en français par le mot *courgette* et, pourtant, nous ne voudrions pas inclure cette paire dans l'anti-lexique de *TransCheck*. Ce qui distingue la paire *library/librairie* de la paire *very/courgette*, c'est qu'il est beaucoup plus probable que la première apparaisse dans une traduction de l'anglais au français, précisément parce que la similitude morphologique des deux mots peut aisément amener le traducteur à oublier leur dissimilitude sémantique. D'une façon plus générale, les erreurs que nous voulons que notre programme détecte sont précisément les erreurs courantes les plus susceptibles de se glisser dans des traductions ou qui, selon les réviseurs, sont effectivement souvent commises par les traducteurs.

Pour nous aider à compiler un inventaire de ce genre d'erreurs pour le premier prototype de *TransCheck*, nous avons consulté plusieurs ouvrages de référence bien connus traitant des difficultés de traduction de l'anglais au français (Copron 1982, Dagenais 1984, De Villers 1988, Rey 1984, Van Roey *et al.* 1988). Nous avons extrait les descriptions de plus de 2 800 difficultés de traduction, que nous avons transcrites et stockées dans de simples enregistrements de base de données comprenant les champs suivants :

- Français incorrect;
- Anglais (correspondant);
- Français correct;
- Type d'erreur;
- Source;
- Commentaires.

Ces descriptions statiques ont ensuite été converties par un autre programme en requêtes *TransSearch*, sous la forme générale suivante :

(i) {e(mot+/cat)}/=f(mot+/cat)  
où «e(mot)» désigne un ou plusieurs mots anglais, «f(mot)» désigne un ou

plusieurs mots français, «cat» désigne une catégorie morphosyntaxique et le symbole facultatif «+» correspond à la forme de base du mot, plus toutes ses variantes flexionnelles. Les accolades ({} ) indiquent le caractère facultatif du segment en langue source <sup>(5)</sup>.

Lorsque les e(mot) et f(mot) sont des éléments uniques et légitimes de leur vocabulaire respectif et qu'ils sont étymologiquement apparentés, la requête correspond à une recherche de faux amis. Mais, bien sûr, il n'est pas toujours nécessaire que ces mots soient étymologiquement apparentés; dans les ouvrages de référence que nous avons consultés, nous avons trouvé de nombreux exemples de traductions illicites entre des paires qui n'étaient pas morphologiquement similaires ou dérivées d'une racine commune. Pour reprendre la terminologie utilisée par une de ces autorités, ces paires sont appelées des *traductions impropres* et elles ont fait l'objet, dans notre base de données, de la même description formelle que

(5) En réalité, l'opérateur «=/> n'apparaît pas dans les requêtes soumises à *TransSearch*. Nous l'avons inclus ici pour souligner le fait que les couples anglais-français ne sont pas des traductions réciproques possibles. Plusieurs des champs de base de données mentionnés plus tôt n'apparaissent pas non plus dans les requêtes formelles, mais sont fournis à titre de renseignements supplémentaires durant la séance d'édition avec *TransCheck*.

(6) Cela ne veut pas dire que la détection des emprunts soit toujours aussi simple. Parmi les complications auxquelles nous nous sommes heurtés, signalons la difficulté de générer les variantes morphologiques des formes non françaises comme *backer*, ainsi que le problème posé par les formes empruntées qui se trouvent coïncider avec un mot parfaitement légitime en langue d'arrivée, comme le mot *pin* qui est utilisé incorrectement en français pour désigner une broche, par opposition au mot français *pin*, qui désigne un type de conifère.

les faux amis. Le mot *cabaret*, par exemple, est tout à fait correct en français, mais ce n'est pas une traduction acceptable du mot anglais *tray*; le mot *cabaret* désigne un type de bar et la traduction française correcte de *tray* est *plateau*. Ainsi, si notre vérificateur de traduction repérait deux segments alignés dont l'anglais contiendrait le mot *tray* ou *trays* et le français contiendrait le mot *cabaret*, il aurait selon toute vraisemblance détecté une erreur de traduction, en supposant, bien sûr, que le mot *cabaret* ne soit pas la traduction d'un autre mot de la phrase anglaise. (Nous reviendrons plus loin sur cette question.) On trouvera d'autres exemples de traductions impropres repérées dans le *Journal des débats* aux entrées (v-vii) du tableau n°1, à la fin de l'article.

En dépouillant les ouvrages de référence traitant des difficultés de traduction (*op.cit.*), nous avons relevé d'autres types de difficultés de traduction que nous avons décidé d'inclure dans le prototype de *TransCheck*. Lorsque deux langues sont en contact étroit, comme c'est le cas de l'anglais et du français au Canada, les emprunts illicites constituent une source très fréquente d'erreurs de traduction. Pour les fins de nos travaux, un emprunt est simplement défini comme un mot anglais apparaissant dans un texte français, et vice-versa. En réalité, les choses ne sont pas si simples : toutes les langues font constamment des emprunts et, dans certains cas, les autorités en la matière ont des opinions divergentes quant au statut de naturalisation des mots étrangers. Dans la grande majorité des cas, cependant, il n'y a pas de controverse. Dans le cas de mots tels que *lunch*, *cool* ou le verbe *backer*, le vérificateur de traduction n'a qu'à détecter leur présence dans un texte français pour signaler un emprunt. Il n'est pas essentiel ici de déterminer leur catégorie lexicale, ni même de vérifier

la présence de la forme correspondante dans le segment anglais aligné <sup>(6)</sup>. En fait, on peut même se demander s'il est nécessaire d'inclure, dans la base de données de *TransCheck*, des entrées décrivant de tels emprunts illicites; étant donné que ces emprunts ne font pas partie du lexique français, un correcteur orthographique unilingue devrait pouvoir les détecter en tant que mots inconnus. Cependant, ce que les entrées de la base de données de *TransCheck* peuvent faire de plus qu'un correcteur orthographique, c'est indiquer à l'utilisateur la forme correcte qu'il devrait utiliser en français au lieu de l'emprunt illicite. Découvrir, par exemple, que *briefing* n'est pas un mot correct en français est une chose, en connaître l'équivalent français correct en est une autre. (On trouvera, au tableau n°1 (xi-xiv), d'autres exemples d'emprunts illicites.)

Les calques constituent un autre type courant d'interférence linguistique. Un calque est la traduction littérale, c'est-à-dire mot à mot, d'une expression complexe en langue source, et dont le résultat est inacceptable en langue cible. Dans certains cas, l'expression calquée peut être parfaitement correcte sur le plan syntaxique, et transparente sur le plan sémantique, mais elle ne correspond tout simplement pas au terme accepté ou normalisé en langue cible. Ainsi, *certificat de naissance* est composé de mots français corrects et combinés conformément aux règles de formation des syntagmes nominaux français; le seul problème, c'est que le français utilise un autre terme pour désigner le *birth certificate* anglais, soit *extrait d'acte de naissance*. Dans d'autres cas, l'expression calquée contrevient aux règles de la grammaire française; c'est le cas de l'expression *à la journée longue*. En français, l'adjectif temporel *long* ne suit pas habituellement le nom qu'il modifie et la façon correcte de

traduire l'expression anglaise *all day long* est à *longueur de journée*. Pour ces deux types de calques, comme dans le cas des emprunts mentionnés ci-dessus, *TransCheck* n'a pas nécessairement besoin de vérifier la présence de l'expression anglaise correspondante dans le segment aligné pour pouvoir détecter une erreur. En fait, en règle générale, lorsqu'un mot ou un syntagme douteux peut être clairement identifié sans recourir au segment aligné de l'autre texte, les requêtes unilingues sont probablement préférables, car on court toujours le risque, dans une relation aussi libre que la traduction, que l'expression spécifiée ne se trouve pas dans le texte source et, par conséquent, que le système omette une erreur faute d'avoir pu établir la correspondance. (On trouvera, au tableau n° 1 (viii-x), d'autres exemples de calques.)

### 3 Résultats préliminaires

Afin de tester le premier prototype de *TransCheck*, nous l'avons appliqué à cinq échantillons de 100 000 mots de texte traduit : trois de ces échantillons étaient tirés du *Journal des débats* de la Chambre des communes et les deux autres, de manuels opérationnels utilisés par le ministère de la Défense nationale (MDN) <sup>(7)</sup>. Évidemment, *TransCheck* est conçu pour vérifier des ébauches de traduction et, dans tous ces cas, les échantillons soumis à la validation étaient des traductions peaufinées et publiées. C'est donc dire qu'on ne peut pas accorder beaucoup d'importance au nombre absolu d'erreurs détectées, pas plus qu'on

(7) L'un de ces manuels traitait de l'utilisation des bataillons d'infanterie, tandis que l'autre portait sur la formation des tireurs d'élite.

envisagerait d'évaluer l'efficacité d'un correcteur orthographique en fonction des résultats obtenus en l'appliquant à des textes unilingues publiés. De plus, comme il s'agissait d'un premier prototype que nous étions en train de développer, nous désirions l'utiliser pour évaluer un certain nombre d'hypothèses, par exemple, la possibilité de relever toutes les expressions multilexicales (syntagmes) en n'utilisant que des requêtes unilingues. L'adoption de telles hypothèses introduit aussi un élément de distorsion dans les données relatives à la performance de *TransCheck*.

Malgré tout, il était important pour nous d'avoir un aperçu du genre d'erreurs que *TransCheck* était capable de détecter, ainsi que du niveau de bruit <sup>(8)</sup> qu'il produisait, ne serait-ce que pour être en mesure d'évaluer la viabilité générale d'un vérificateur de traduction qui utilise un programme d'alignement de phrases pour détecter des correspondances lexicales illicites. Le tableau n° 2, à la fin de l'article, résume les résultats obtenus par *TransCheck* sur les cinq corpus d'essai.

Ce qu'il faut d'abord remarquer à propos de ces résultats, ce sont les taux de précision impressionnants obtenus par *TransCheck* sur les corpus tirés du *Journal des débats* et, dans une moindre mesure, sur le premier corpus tiré des manuels du MDN. Ce que ces résultats nous apprennent, en d'autres termes, c'est que la grande majorité des erreurs potentielles repérées dans ces corpus s'avèrent des erreurs véritables, en ce sens qu'elles correspondent exactement aux difficultés linguistiques décrites par les autorités que nous avons consultées. Nous exposerons plus loin les raisons des résultats inférieurs obtenus par le système avec les manuels du MDN. Ce qu'il faut souligner ici, cependant, c'est que, parmi tous les cas de bruit relevés dans les cinq corpus, il n'y en a qu'un ou deux qui sont attribuables à la

grossièreté des alignements réalisés par le système, et aucun qui n'est attribuable à un mauvais alignement des phrases source et cible. Cela confirme les premiers résultats décrits dans Isabelle *et al.* (1993) et semble justifier l'affirmation qui y est faite qu'un vérificateur de traduction basé sur un programme d'alignement de phrases et un programme de catégorisation grammaticale pourrait bien constituer une plate-forme suffisante pour le développement d'applications pratiques.

Plusieurs autres aspects des résultats présentés au tableau n°2 méritent qu'on s'y attarde, notamment le nombre considérablement moindre d'erreurs repérées dans le corpus *Journal des débats-2*, comparativement au corpus *Journal des débats-1*. Cette baisse importante découle du fait que nous avons omis de spécifier l'anglais comme langue source dans le premier échantillon du *Journal des débats*, de telle sorte qu'un grand nombre des erreurs détectées dans ce corpus apparaissaient en réalité dans les discours des parlementaires francophones. Comme nous nous intéressions davantage aux erreurs des traducteurs français qu'à celles des députés francophones, nous avons réglé ce paramètre pour que l'anglais soit exigé comme langue source dans les quatre autres échantillons d'essai <sup>(9)</sup>. Le nombre total d'erreurs

(8) Le mot *bruit* est employé ici dans un sens analogue à celui qui est le sien en recherche documentaire pour désigner des résultats qui ne sont pas pertinents, dans ce cas-ci, parce qu'ils ne correspondent pas à des erreurs de traduction véritables.

(9) Bien qu'un vérificateur de traduction puisse, en principe, être bidirectionnel, *TransCheck* présente une nette orientation anglais-français, principalement attribuable au fait que les ouvrages de référence que nous avons consultés se concentrent presque exclusivement sur les problèmes d'usage du français.

détectées dans ces échantillons serait par conséquent beaucoup plus élevé si nous n'avions pas modifié le paramètre langue source.

Une autre constante qui se dégage de l'analyse des échantillons d'essai est la suivante : la plupart des erreurs détectées dans un texte donné sont attribuables à un nombre relativement restreint de difficultés linguistiques spécifiées dans la base de données de *TransCheck*. En d'autres termes, la plupart des requêtes soumises par le système ne trouvent aucune correspondance dans le texte traduit. Nous avons mentionné plus tôt que la base de données actuelle de *TransCheck* contient environ 2 800 entrées; dans chacun des quatre derniers échantillons d'essai, à peine 25 de ces entrées ont permis de détecter des erreurs (bien que les requêtes n'aient pas été nécessairement les mêmes pour tous les échantillons et qu'une même requête ait pu évidemment repérer plusieurs erreurs dans le même texte). Nous ne nous attendons pas à ce que cela change beaucoup, même en appliquant *TransCheck* à des ébauches de traduction plus grossières. Ce phénomène est attribuable à la nature de la base de données de *TransCheck*, qui vise à décrire le plus grand nombre possible d'erreurs de traduction attestées et les plus probables, quel que soit le domaine ou le type de texte. Il reste qu'une traduction donnée ne contiendra normalement qu'une fraction de ces erreurs; donc, il est tout à fait naturel que la plupart des requêtes traitées par *TransCheck* ne donnent aucun résultat, car même les premiers jets les plus grossiers contiennent beaucoup plus de segments rendus correctement que de segments contenant des erreurs.

Du relativement petit nombre d'entrées qui produisent effectivement des correspondances, il y en a quelques-unes qui tendent à réapparaître, parfois très

fréquemment, dans chacun des échantillons d'essai. Dans le *Journal des débats*, par exemple, le terme *caucus* est repéré à de nombreuses reprises, ce qui est également le cas du couple *deception/déception* dans les corpus tirés des manuels du MDN. Ce ne sont donc pas des fautes d'inattention; leur répétition indique plutôt une sérieuse divergence d'opinion entre au moins l'une des autorités que nous avons consultées et les traducteurs qui produisent ces textes. D'ailleurs, en consultant d'autres références (comme *Termium*<sup>®</sup>), nous pouvons constater que ces deux termes sont acceptés et considérés comme corrects. Ce genre de divergence entre autorités suscite un problème épineux pour le développement d'un vérificateur automatique de traduction, et l'attitude que nous avons adoptée peut se résumer comme suit : ce n'est pas à nous de trancher. Nous n'avons pas la prétention d'être des experts, et encore moins des prescripteurs, en matière d'usage correct du français, et nous ne prenons pas position vis-à-vis des erreurs répertoriées par les autorités que nous avons consultées. Pour chaque erreur qu'il détecte, le système fournit un renvoi à l'ouvrage dans lequel cette difficulté est décrite. Si les utilisateurs éventuels ne sont pas d'accord avec l'opinion de cette autorité, il faudra qu'ils puissent neutraliser l'entrée correspondante dans *TransCheck*, tout comme les gens qui utilisent des correcteurs stylistiques peuvent désactiver certaines règles ou certains ensembles complets de règles. Plus généralement, *TransCheck* est un outil qui contribuera à l'application des normes de traduction, normes au pluriel, car il n'existe sûrement pas de norme de traduction unique qui s'applique à tous les domaines et à tous les types de textes. Voilà pourquoi la constitution d'une base de données élémentaire sera certainement plus difficile pour les

concepteurs d'un vérificateur de traduction qu'elle ne le sera, disons, pour les concepteurs d'un correcteur orthographique. Mais quelle que soit la norme (ou les normes) que les utilisateurs décideront d'adopter pour leur usage personnel ou pour celui de leurs divers clients, nos résultats préliminaires portent à conclure qu'un outil comme *TransCheck* les aidera néanmoins à détecter un nombre appréciable d'erreurs dans leurs textes.

Les niveaux de bruit plus élevés qu'on peut constater, au tableau n°2, dans les résultats obtenus par *TransCheck* sur les deux échantillons tirés des manuels du MDN, sont principalement attribuables à un petit nombre d'erreurs de description dans la base de données du système, dont certaines renvoient à des termes qui reviennent parfois très fréquemment dans les manuels militaires. Notre analyse de ces résultats révèle qu'il s'agit, dans la plupart des cas, de règles de traduction dépendant du contexte, qui ont été incorrectement caractérisées, soit par nous, soit par une de nos autorités, comme des interdictions traductionnelles absolues<sup>(10)</sup>. Nous avons développé *TransCheck* en nous servant du *Journal des débats* comme corpus de référence et il se peut que nous ayons involontairement adapté, jusqu'à un certain point, les entrées lexicales du système au langage parlementaire. Il n'est donc pas tout à fait surprenant

(10) Selon une de nos autorités, par exemple, le nom français *raid* n'est pas une traduction correcte de l'anglais *raid*, quand il s'agit d'une intervention policière, bien qu'il le soit, en l'occurrence, quand il s'agit d'une intervention militaire. Quant à nous, nous avons entré, à tort, le verbe *grader* comme faux ami complet, ignorant son emploi correct au sens de «calibrer», soit la façon dont ce verbe est employé dans le manuel de formation du MDN à l'intention des tireurs d'élite.

que, lorsque le système est appliqué à des textes relevant d'un domaine très différent, certains des ajustements qui entrent inévitablement dans l'élaboration d'un dictionnaire automatisé ne donnent pas les résultats attendus. L'autre principale source de bruit dans les résultats obtenus des échantillons du MDN est attribuable à certaines lacunes des modules de découpage et d'analyse lexicale de *TransCheck* : par exemple, l'abréviation *s/off* (sous-officier) est incorrectement découpée en deux unités et l'unité *off* est identifiée comme un emprunt. Nous devrions être en mesure de corriger ces problèmes sans trop de difficulté.

Voici maintenant quelques mots sur la forme actuelle des sorties de *TransCheck*. Pour ce premier prototype de démonstration, nous avons utilisé comme interface un éditeur *Emacs* standard (voir les deux recopies d'écran présentées à la fin du texte). Bien qu'elle ne soit pas aussi élégante que l'interface graphique utilisée pour *TransSearch*, l'interface *Emacs* offre beaucoup de souplesse aux programmeurs et aux linguistes qui travaillent au développement de *TransCheck*. La fenêtre supérieure de chaque écran affiche le bitexte aligné côte à côte, le segment visé par la requête étant affiché en gras et l'erreur potentielle détectée par le système étant soulignée. La fenêtre inférieure affiche l'entrée correspondante dans la base de données, la requête compilée et le résultat de l'analyse grammaticale des segments source et cible, qui se résume actuellement à une analyse de la catégorie grammaticale. Ainsi, dans la première recopie d'écran, tel type d'erreur détectée est un calque : selon le *Multidictionnaire des difficultés de la langue française*, l'expression *en devoir* n'est pas une traduction acceptable de l'anglais *on duty*; le syntagme prépositionnel qui devrait être utilisé est *de service* ou *en service*. Dans la deuxième recopie d'écran, l'entrée de

la base de données nous apprend que le verbe *nominer* est un emprunt illicite en français; la forme correcte est *mis en nomination*.

## 4 Extension du prototype actuel

Les données présentées au tableau n°2 sont révélatrices de la qualité des résultats produits par *TransCheck*, pour ce qui est du rapport bruit/erreurs véritables détectées. Malheureusement, nous ne disposons pas actuellement de données sur le «silence» du système, c'est-à-dire sur la proportion d'erreurs véritables qui échappent à *TransCheck*. Pour obtenir de telles données, il nous faudrait appliquer le système à une traduction-étalon, dont toutes les erreurs auraient été préalablement décelées par un expert; cet exercice nous permettrait de calculer le nombre d'erreurs détectées et le nombre d'erreurs omises par le système. Néanmoins, d'après les mois de travaux préliminaires consacrés au développement du premier prototype et d'après l'analyse des premiers résultats décrits ci-dessus, nous savons que la version actuelle du système détecte à tort certains segments qui ne contiennent pas d'erreurs et, surtout, qu'il existe de nombreux types de difficultés de traduction qui ne peuvent être saisis par son modèle descriptif rudimentaire.

Comme nous l'avons déjà indiqué, nous savons que l'alignement de deux textes au niveau de la phrase n'est pas toujours adéquat; à l'occasion, les phrases alignées contiennent des paires de mots ou d'expressions interdites, mais qui ne sont pas des traductions réciproques<sup>(11)</sup>. La solution à ce problème est évidente : nous devons obtenir des alignements bitextuels plus subtils, qui s'établissent non plus au niveau de la phrase, mais au

niveau du syntagme et du mot. Des progrès ont été réalisés dans ce sens, au Citi et ailleurs (voir, en particulier, Dagan *et al.* 1993).

Ensuite, il faut considérer les nombreuses difficultés de traduction dont la description exige plus d'information contextuelle que la catégorisation grammaticale élémentaire actuellement réalisée par *TransCheck*. Prenons, par exemple, le verbe français *débuter* : on peut l'utiliser pour traduire le verbe anglais *to start*, mais seulement lorsqu'il est intransitif. Pour que *TransCheck* arrive à ne détecter que les occurrences erronées de *débuter*, il faudrait enrichir le système d'un analyseur syntaxique capable de faire la distinction entre verbes transitifs et intransitifs. Et bien sûr, le même raisonnement s'applique aux contraintes sémantiques visant les types d'arguments. Par exemple, selon une des autorités que nous avons consultées, le verbe français *touer* ne peut accepter que des compléments directs appartenant à la classe des embarcations, contrairement à son sosie anglais *to tow*, qui peut avoir pour complément n'importe quel type de véhicule. Là encore, pour que *TransCheck* arrive à distinguer les occurrences correctes des occurrences incorrectes du verbe français, il lui faudra évidemment incorporer un analyseur sémantique.

En fait, si Martin Kay a raison d'affirmer qu'«il n'y a rien qu'une personne puisse savoir, ressentir ou rêver qui ne soit crucial à la bonne traduction d'un texte»<sup>(12)</sup>, alors il n'y

(11) On trouvera un exemple de ce genre de bruit au tableau n°1 (xv). Dans ce cas, la paire illicite est constituée des noms *grant/octroi*, car le français *octroi* désigne uniquement l'action d'accorder une subvention, et non la subvention elle-même. Le problème est que la phrase française renferme aussi le nom *subvention*, qui est la traduction correcte de *grant*.

(12) Tiré (et traduit) de l'avant-propos de Hutchins et Somers (1992).

a, en principe, aucune limite aux connaissances linguistiques et extralinguistiques qu'un vérificateur de traduction devrait incorporer afin de pouvoir détecter toutes les erreurs de traduction possibles. Réviser une traduction n'exige pas moins de connaissances qu'il n'en faut pour produire un premier jet, et un vérificateur de traduction qui vise à émuler parfaitement un réviseur humain devra posséder un niveau de compréhension qui va bien au-delà du texte littéral pour englober toute l'intelligence qu'un traducteur compétent apporte à sa tâche. Il est bien évident qu'un tel système n'est pas près de faire irruption sur le marché, et il ne fait aucun doute que son apparition devra attendre l'avènement de la traduction entièrement automatique de grande qualité. Il y a, cependant, une différence majeure entre une approximation partielle du vérificateur de traduction idéal, comme le modeste prototype décrit dans cet article, et une approximation partielle d'un système de traduction automatique idéal. Un vérificateur de traduction moins qu'exhaustif, tout comme un correcteur orthographique incomplet, peut quand même être utile : dans un cas comme dans l'autre, les erreurs que le système arrive à détecter automatiquement permettent à l'utilisateur d'améliorer la qualité du texte final. Il n'est pas du tout évident que les systèmes de traduction automatique actuels, qui fonctionnent nécessairement à partir d'une compréhension incomplète du texte à traduire, puissent rendre un service comparable au traducteur humain. Il faut souvent plus de temps et d'effort pour corriger et remanier une approximation partielle de traduction produite par un système de traduction automatique qu'il n'en faut à un traducteur humain pour produire une version correcte à partir de zéro. Voir Pierre Isabelle (1993), où cet argument en faveur des outils

de traduction, par opposition à la traduction automatique classique, est présenté de façon plus détaillée.

Par rapport aux exemples relativement simples de complémentation syntaxique et sémantique cités à la page précédente, on pourrait légitimement se demander si la détection de ces difficultés nécessite le recours à un vérificateur de traduction bilingue, ou si un vérificateur grammatical unilingue français ne ferait pas tout aussi bien l'affaire. Après tout, quand les experts, par exemple, qualifient d'«anglicisme» l'emploi transitif du verbe *contribuer*, en réalité, ils nous informent de la genèse du problème, c'est-à-dire du fait que les francophones ont emprunté la façon dont le verbe se construit en anglais. Cependant, quelle que soit son origine, cette difficulté demeure essentiellement un problème d'usage du français dont la détection automatique ne nécessite pas absolument le recours à la source d'une traduction. On pourrait facilement imaginer exactement le même problème survenant dans un document unilingue français n'ayant fait l'objet d'aucune traduction. En fait, on pourrait très bien considérer tous les problèmes traités par *TransCheck* dans l'optique d'un vérificateur de langue cible unilingue, même les faux amis comme *library/librairie*. De ce point de vue, le mot français *librairie* a deux sens, un sens correct et un incorrect. Quand ce mot apparaît dans une traduction, nous pouvons utiliser le texte source pour faire la distinction entre les deux sens : si *librairie* est utilisé pour traduire le mot anglais *library*, il est utilisé incorrectement. Mais, au fond, le recours au texte source n'est qu'une façon d'élucider le sens du mot, et on peut en imaginer d'autres.

En pratique, cependant, la question soulevée ci-dessus est quelque peu hypothétique <sup>(13)</sup>. Si nous disposons d'une grammaire

générative complète et calculable d'une langue, nous pourrions peut-être l'utiliser pour valider automatiquement la correction de tout texte rédigé dans cette langue, en fonction des formes correctes énumérées par cette grammaire et en fonction du complément de cet ensemble constitué par les formes incorrectes <sup>(14)</sup>. Il va sans dire que cette grammaire n'existe pas et que tant qu'il en sera ainsi, la meilleure chose à faire consiste peut-être à décrire explicitement certaines des erreurs qui apparaissent fréquemment dans cette langue. C'est exactement ce que le système *TransCheck* tente de faire, particulièrement pour les erreurs qui sont susceptibles d'être commises par suite du contact entre deux langues; et pour chaque erreur de ce genre, il propose une correction possible.

Les extensions du modèle linguistique de *TransCheck*, auxquelles nous avons fait allusion plus haut, exigeront des efforts de recherche considérables, surtout si nous voulons que les résultats de l'analyse syntaxique et sémantique atteignent le niveau d'exactitude et de robustesse offert par le programme de catégorisation grammaticale que le système emploie actuellement. À court terme, nous prévoyons de procéder à des perfectionnements moins ambitieux, comprenant la vérification des expressions numériques entre

(13) En premier lieu, à notre connaissance, les correcteurs de grammaire française qu'on trouve actuellement sur le marché ne vérifient guère que l'accord grammatical, par exemple entre le verbe et le sujet, et même à ce niveau, leur performance est loin d'être brillante.

(14) Précisons que même cette grammaire ne pourrait pas nous aider à repérer les phrases de cette langue qui sont correctes, mais ne communiquent pas correctement le sens voulu par l'auteur.

textes source et cible <sup>(15)</sup>, la détection des omissions d'unités importantes dans le texte traduit (c'est-à-dire des omissions de phrases et de paragraphes) et, peut-être, la vérification de la cohérence terminologique. Ces nouvelles fonctions seront décrites de façon plus détaillée dans les prochains articles faisant le point sur *TransCheck*.

## 5 Conclusion

Nous avons affirmé, au début de cet article, que l'avènement de la bitextualité permettait d'envisager un nouveau type d'outil rédactionnel pour traducteurs, un outil capable de détecter les erreurs de traduction qui échappent, de par leur nature, aux correcteurs orthographiques ou de grammaire unilingues. Au cours de l'exposé, nous avons quelque peu nuancé cette affirmation. Nous avons vu plusieurs types d'erreurs de traduction, notamment les emprunts illicites et les calques grammaticaux, qui pourraient, en principe, être détectées mais peut-être pas facilement corrigées par un outil rédactionnel unilingue. Pour d'autres types d'erreurs de traduction, en particulier celles qui sont caractérisées par le fait que le texte cible demeure grammaticalement correct mais ne véhicule pas le même sens que son pendant en langue source, un outil de validation bitextuelle comme *TransCheck* semble constituer la meilleure solution possible à l'heure actuelle. Nous avons avancé que même une approximation partielle d'un tel vérificateur de traduction pourrait être utile au traducteur ou au réviseur humain, et nous avons esquissé nos plans à court et à long

(15) On trouvera, au tableau n°1 (xvi), un exemple de ce genre d'erreur, relevé dans le *Journal des débats*, que nous espérons pouvoir détecter ultérieurement avec *TransCheck*.

terme pour augmenter les capacités du premier prototype de *TransCheck*.

*Elliott Macklovitch,*  
*Centre d'innovation en technologies*  
*de l'information,*  
*Laval (Québec),*  
*Canada.*

## Remerciements

Le système *TransCheck* décrit dans cet article est le fruit des efforts concertés de plusieurs chercheurs du Citi, dont j'aimerais souligner ici l'apport. Ce sont : Pierre Isabelle, directeur du groupe de traduction assistée par ordinateur du Citi, Monique Chevalier, George Foster, Marie-Louise Hanna, François Perrault, Xiaobo Ren, Michel Simard et Caroline Viel, que je tiens à remercier de l'excellent travail qu'ils ont accompli. J'assume évidemment l'entière responsabilité des erreurs qui auraient pu se glisser dans ce rapport.

## Bibliographie

Brown (P.), Lai (J.) et Mercer (R.), 1991 : «Aligning sentences in parallel corpora», dans *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, Berkeley, Californie, p. 169-176.

Colpron (Gilles), 1982 : *Dictionnaire des anglicismes*, Laval (Québec), Éditions Beauchemin.

Dagan (Ido) Church (Kenneth) et Gale (William), 1993 : «Robust Bilingual Word Alignment for Machine-Aided Translation», dans *Proceedings of the Workshop on Very Large Corpora*, Columbus, Ohio.

Dagenais (Gérard), 1984 : *Dictionnaire des difficultés de la langue française au Canada*, Boucherville (Québec), Les Éditions françaises.

De Villers (Marie-Éva), 1988 : *Multidictionnaire des difficultés de la*

*langue française*, Montréal, Éditions Québec/Amérique.

Gale (William) et Church (Kenneth), 1991 : «A Program for Aligning Sentences in Bilingual Corpora» dans *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, Berkeley, Californie, p. 177-184.

Harris (Brian), 1988 : «A New Concept in Translation Theory», dans *Language Monthly*, n°54.

Hutchins (John) et Somers (Harold), 1992 : *An Introduction to Machine Translation*, Academic Press.

Isabelle (Pierre), 1992 : «Bi-text: Toward a new Generation of Support Tools for Translation and Terminology», dans *Citi Technical Report*, 29. Publié en français sous le titre «La bitextualité : vers une nouvelle génération d'aides à la traduction et à la terminologie» dans *Meta*, vol. 37, n° 4, p. 721-737.

Isabelle (Pierre), 1993 : «Machine-Aided Human Translation and the Paradigm Shift», dans *Proceedings of the Fourth Machine Translation Summit*, Kobe.

Isabelle (Pierre), Dymetman (Marc), Foster (George), Jutras (Jean-Marc), Macklovitch (Elliott), Perrault (François), Ren (Xiaobo) et Simard (Michel), 1993 : «Translation Analysis and Translation Automation», dans *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation*, Kyoto.

Macklovitch (Elliott), 1992 : «Corpus-Based Tools for Translators», dans *Proceedings of the 33rd Annual Conference of the American Translators Association*, San Diego, Californie, p. 317-328.

Rey (Jean), 1984 : *Dictionnaire sélectif et commenté des difficultés de la version anglaise*, Paris, Éditions Ophrys.

Simard (Michel), Foster (George) et Isabelle (Pierre), 1992 : «Using Cognates to Align Sentences in Parallel Corpora», in *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation*, Montréal, p. 67-81.

Van Roey (Jacques) Granger (Sylviane) et Swallow (Helen), 1988 : *Dictionnaire des faux amis français-anglais*, Paris, Duculot.

TABLEAU 1

SOURCE	TRADUCTION
(i) ... the overall long term survival of any type and <b>stage</b> of childhood cancer ...	... le taux de survie à long terme, pour tous types et à tous les <b>stages</b> de cancer ... (19/05/92)
(ii) ... to ensure progress toward a world free of injustice, discrimination and <b>prejudice</b> .	... notre recherche d'un monde exempt d'injustice, de discrimination et de <b>préjudice</b> . (10/12/91)
(iii) Since the passage of that <b>infamous</b> bill in 1987, Bill C-22 ...	Depuis l'adoption de l' <b>infâme</b> projet de loi C-22 en 1987 ... (17/11/92)
(iv) In this country ... with our reputation for <b>affluence</b> , with our reputation for humanity ...	... dans notre pays reconnu ... pour son <b>affluence</b> et son empressément à aider ... (10/05/93)
(v) Also, the polls say that regardless of the position of the Tories in England before and during the campaign – and it was up and down and <b>neck and neck</b> for a while ...	... quelle qu'ait été la position du Parti Conservateur en Angleterre avant et durant la campagne, et on peut dire que la position des partis a beaucoup fluctué, que ces derniers ont été <b>nez à nez</b> pendant un certain temps ... (10/04/92)
(vi) They ... drive their children to practices and <b>games</b> .	Ils ... conduisent leurs enfants à des pratiques et à des <b>joutes</b> . (12/02/93)
(vii) It is about \$190 going over \$200 <b>depending upon</b> how much money they earned ...	Il est d'environ 190 \$ et peut dépasser les 200 \$, <b>dépendant</b> du montant gagné ... (15/02/93)
(viii) ... the destruction of private property cannot be tolerated under <b>any circumstances</b> .	... la destruction de la propriété privée ne peut être tolérée <b>en aucun temps</b> . (5/05/92)
(ix) ... profits from a <b>used</b> equipment sale ...	... recettes d'une vente d'équipement de <b>seconde main</b> ... (4-06-90)
(x) ... his interest rate projections ... are on a <b>calendar year basis</b> .	... les taux d'intérêt sont basés sur l'année de <b>calendrier</b> (4/04/90)
(xi) It is also an <b>essential condition</b> for the confidence ... that will ensure a strong recovery.	Il s'agit également d'un <b>prérequis</b> au retour de la confiance ... (25/02/92)
(xii) You are resting on <b>technicalities</b> .	Vous vous arrêtez à des <b>technicalités</b> . (12/03/92)
(xiii) ... many of whom were historic <b>supporters</b> of the Conservative party ...	... dont beaucoup ont toujours été des <b>supporteurs</b> du Parti conservateur ... (21/03/91)
(xiv) A real <b>rip-off</b> occurred during the long weekend.	Lorsqu'il y a des longues fins de semaine, on assiste à un véritable <b>racket</b> ... (25/05/93)
(xv) Canada Employment has recently announced a section 25 <b>grant</b> to help it upgrade its facilities ...	Emploi Canada a récemment annoncé l' <b>octroi</b> à cette entreprise d'une subvention en conformité de l'article 25, afin de l'aider ... (27/04/93)
(xvi) The European Economic Community's research and development project has a budget of <b>\$17 million</b> spread over four years ...	Le projet de recherche et de développement de la Communauté économique européenne bénéficie d'un budget de <b>70 millions de dollars</b> réparti sur quatre ans ... (17/04/86)

TABLEAU 2

Corpus	Nombre total d'erreurs repérées	Erreurs véritables	Bruit	Précision
Débats 1	63	60	3	95 %
Débats 2	33	28	5	85 %
Débats 3	50	48	2	96 %
MDN 1	76	52	24	68 %
MDN 2	60	31	29	52 %

# TransCheck : recopie d'écran I

```

emacs@mesange.citi.doc.ca
Buffers  File  Edit  Help
Taking away authority from the police is not the way to achieve this goal. Diminuer le pouvoir de la police n'est pas le moyen de rendre le Canada plus sûr.

If this legislation is going to do that then I would have to oppose the bill. Si c'est ce que vise le projet de loi, alors je m'y oppose.

In conclusion I want answers to the following questions: How many police officers have died in this country in the line of duty as compared to the lives of criminals lost? Je terminerai en demandant qu'on réponde aux questions suivantes : Combien d'agents de police sont morts en devoir au Canada et combien de criminels sont morts au cours d'opérations policières?

If more police have died then I would suggest this legislation must be geared toward protecting them. Si le nombre de policiers est plus élevé que le nombre de criminels, alors il faut que la loi les protège mieux.

Is that the case? Qu'en est-il?

--%%-Emacs: *TransCheck Text* (TransCheck)--60%-----
Compiled query: F(en devoir) f(~(en devoir de | du | d' | des))
Français incorrect: en devoir BUT_NOT en devoir de
Anglais: on duty
Français correct: de service ; en service ; de garde
Type d'erreur: expression syntagme prépositionnel impropre
Source: Multi
</Error>

<Text>
CoupleInfo: Couple ID = 2554,
            Document = "text",
            Source = English
tokens[L1] = {}
tokens[L2] = {72493, 72494}
In:          PP          Je:          Pron-ps-genI-sing-p\
conclusion:  NC          1
I:          PN          terminerai:  Verb-IndFut-sing-p1
want:       VB          en:         Prép
answers:    VB          demandant:  Verb-ParPré
to:         TO          qu':       ConS
the:        DT          on:        Pron-ps-aasc-sing-p\
following:  AJ          3
questions:  NC          réponde:   Verb-SubPré-sing-p3

--%%-Emacs: *TransCheck Info* (TransCheck Narrow)-- 1%-----

```

# TransCheck : recopie d'écran II

```

emacs@mesange.citi.doc.ca
Buffers  File  Edit  Help
The first, "The Mighty River"
directed by Frédéric Back and
produced by Hubert Tison at the
Société Radio-Canada, was nominated
as Best Animated Short film.
Le premier, Le fleuve aux grandes
eaux, dirigé par Frédéric Back et
réalisé par Hubert Tison, de la
Société Radio-Canada, est en lice
comme meilleur court métrage
d'animation.

The second, "The Broadcast Tapes of
Doctor Peter", produced by David
Papernay and Arthur Ginsberg at the
Canadian Broadcasting Corporation in
Vancouver, was nominated as Best
Feature-Length Documentary Film.
Le deuxième, The Broadcast Tapes of
Doctor Peter, réalisé par David
Papernay et Arthur Ginsberg, du
réseau anglais de la Société Radio-
Canada, à Vancouver, a été nominé
comme meilleur long métrage
documentaire.

I would like to congratulate our
nominees and recognize the quality
and dedication of our public
broadcasters.
Je tiens à féliciter nos nominés et à
reconnaître la qualité et le
dévouement de nos radiodiffuseurs
publics.

They deserve our continued support.
Ils méritent notre appui total.

Multicultural Week
La Semaine multiculturelle

--%%-Emacs: *TransCheck Text* (TransCheck Compiling)--90%-----
Français incorrect: nominer+/Verb
Anglais:          nominate+/VB
Français correct: mettre en nomination; recevoir une nomination; sélectionner
Type d'erreur:    emprunt
Source:           Multi
%% emprunt: le verbe n'existe pas en français (R&C)
</Error>

<Text>
CoupleInfo: Couple ID = 3905,
             Document = "text",
             Source = English
tokens[L1] = {}
tokens[L2] = {109798}
The:        DT          Le:          Déte-ddéf-masc-sing
second:     OD          deuxième:    Ordi-masc-sing
:;         PC          .:         Punc-pcca
The:        DT          Broadcast:   NonP
Broadcast:  NC          Tapes:      Verb-SubPré-sing-p2
Tapes:     NC          of:         Verb
of:        PP          Doctor:     NonP
Doctor:    NC          Peter:     NonP-masc-sing
--%%-Emacs: *TransCheck Info* (TransCheck Compiling Narrow)--13%-----

```

# *TransSearch* : Un concordancier bilingue pour la traduction et la terminologie

**L**es tableaux qui suivent ont servi à étayer la présentation que j'ai faite lors de l'Atelier sur les outils de gestion terminologique, tenu à Ottawa le 5 avril 1995. Cet atelier a été organisé par le Comité consultatif canadien du Comité technique terminologie de l'Organisation internationale de normalisation (CCC/ISO/TC37) et le Comité technique sur la terminologie de l'Association canadienne de normalisation (CSA).

## Les traductions préexistantes : une ressource à exploiter

Les traductions préexistantes contiennent davantage de solutions à davantage de problèmes de traduction que toute autre ressource !

Les archives sont à l'heure actuelle inutilisables faute d'outils appropriés.

Solution : outils capables de construire automatiquement des bases de données bitextuelles.

Dans une base de données bitextuelle, les segments de S sont liés à ceux de T

⇨ par conséquent, les segments « problème » sont liés à des segments « solution ».

Alignement automatique des phrases : une première approximation déjà réalisable.

## *TransSearch*

Nouveau concordancier bilingue du Centre d'innovation en technologies de l'information (Citi).

Conçu à partir d'un analyseur qui apparie les phrases des deux textes. Les requêtes peuvent porter sur l'une ou l'autre langue, ou sur les deux langues pour :

- Trouver des exemples de l'anglais *out to lunch*;
- Trouver des exemples du français *passer un sapin*;
- Trouver des exemples du terme *commitment* traduit par *attachement*.

*Automatic morphological expansion of queries*<sup>(1)</sup> : passer ⇨ passe, passes, passons, etc.

Rapide : indexation plein-texte sur les deux langues

Interface utilisateur *X-Windows*

## Intérêt pour la terminologie

Le concordancier permet de vérifier l'usage des traducteurs.

Usage ≠ norme

Mais l'usage n'est-il pas un facteur à prendre en compte dans la définition d'une norme ?

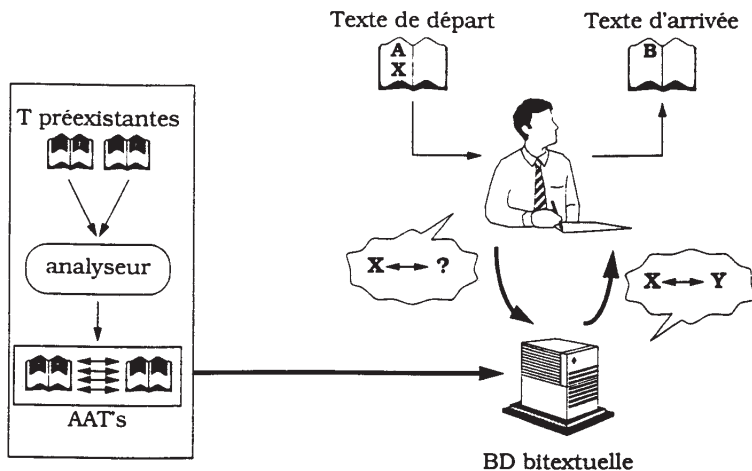
Vérifier dans quelle mesure une norme est effectivement respectée.

Récupérer le travail terminologique « ponctuel » qui n'aurait pas été versé dans la mémoire terminologique d'entreprise.

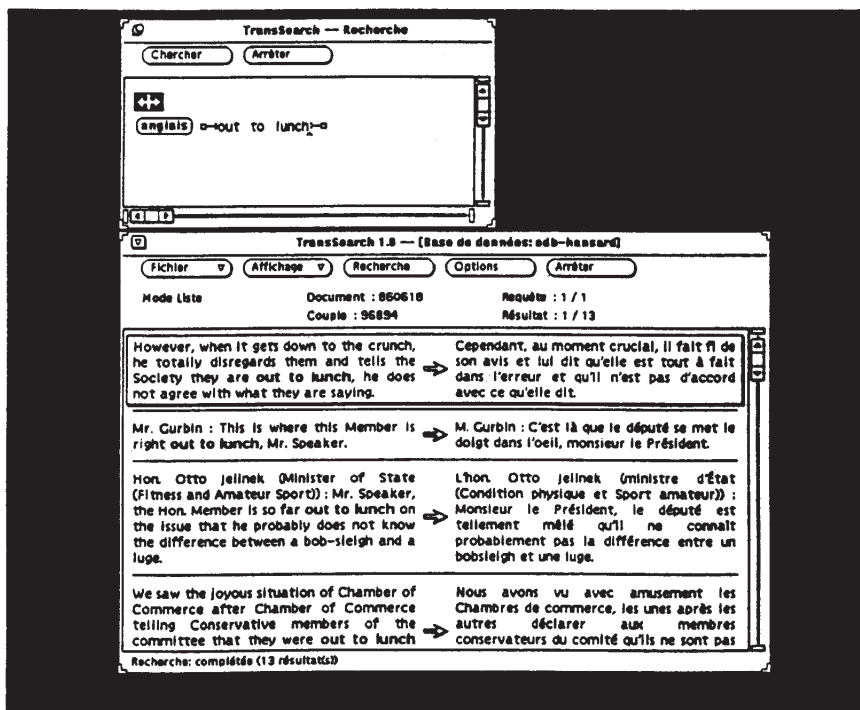
*Pierre Isabelle,  
Centre d'innovation en technologies  
de l'information,  
Laval (Québec),  
Canada.*

(1) Note de la rédaction :  
*prise en compte automatique de  
la flexion.*

## Utilisation du concordancier bilingue



## TransSearch: Écran I



# La base de données liégeoise du *Robert & Collins*: un outil pour le lexicographe, le linguiste ou le traducteur

Cet article traite de la transformation du dictionnaire bilingue *Robert & Collins* (anglais-français) en une base de données lexico-sémantiques offrant une grande diversité de chemins d'accès. Le modèle théorique sous-jacent est celui du linguiste russe Igor Mel'cuk dont la théorie des fonctions lexicales est utilisée pour modéliser les relations collocationnelles et paradigmatiques découvertes au sein de la microstructure des entrées du dictionnaire bilingue. Le but est de fournir aux lexicographes, aux linguistes ou aux traducteurs un outil souple permettant d'accéder facilement aux collocations par le biais de la base, du collocatif, des catégories syntaxiques ou des relations lexico-sémantiques. Les programmes d'application permettant l'extraction de données linguistiquement pertinentes sont décrits en détail.

Termes-clés :  
collocation; lexicographie  
computationnelle; dictionnaire  
informatisé; bases de données  
lexicales; dictionnaire bilingue;  
fonction lexicale.

## 1 Introduction

**D**epuis une quinzaine d'années environ, la communauté scientifique et, dans une certaine mesure, le monde industriel, s'intéressent à l'exploitation et à la construction de dictionnaires informatisés. Les raisons de cet intérêt sont multiples. Confrontée à une explosion du volume de textes à traduire et à traiter, notre société de l'information se voit contrainte de développer toute une série d'outils destinés à faciliter la tâche du rédacteur, du traducteur ou du documentaliste. La qualité des systèmes de traduction assistée par ordinateurs et des systèmes de recherche documentaire dépend en grande partie de la taille et du raffinement linguistique du module lexical qui les accompagne. La construction de ces lexiques (bilingues ou monolingues) est cependant une tâche coûteuse et de longue haleine, qui réclame des compétences lexicographiques et linguistiques encore trop peu répandues. Il n'est dès lors pas surprenant que les chercheurs se soient penchés sur les possibilités de réutilisation et d'exploitation offertes par certains dictionnaires

commerciaux disponibles sur support informatique. La tradition lexicographique anglo-saxonne a en effet produit une série de dictionnaires d'apprenants dont la richesse potentielle a alléché bon nombre de linguistes désespérément en quête de données lexicales susceptibles de nourrir leurs applications computationnelles (voir entre autres les dictionnaires des maisons Longman, Oxford University Press ou HarperCollins - Procter 1978, Cowie 1989, Sinclair 1987; voir aussi Boguraev & Briscoe 1989 pour un panorama de la lexicographie computationnelle et une description détaillée de recherches sur le *Longman Dictionary of Contemporary English*).

Les dictionnaires bilingues ont généralement moins attiré l'attention des chercheurs. Il est vrai que les maisons d'édition se sont montrées moins enclines à distribuer les bandes magnétiques de leurs ouvrages bilingues, souvent par peur d'une utilisation malhonnête de leurs données. Une raison scientifique de ce faible intérêt réside dans le fait que les dictionnaires bilingues existant sous format informatique se prêtaient moins facilement à une structuration en base de données que leurs équivalents monolingues anglais. La structure de ces derniers rendait plus facile l'identification de divers types d'information d'une entrée lexicale

Communauté française de

# Belgique

(partie du discours, codes syntaxiques, définitions, exemples, propriétés morphologiques, etc.) alors que les quelques dictionnaires bilingues disponibles n'étaient rien d'autre que les bandes magnétiques ayant servi à la photocomposition de l'ouvrage. Un énorme travail d'interprétation et de décodage des entrées était dès lors nécessaire afin de transformer un simple dictionnaire lisible par machine en une véritable base de données lexicales permettant un accès rapide, aisé et diversifié aux informations lexicales. C'est exactement ce que nous avons tenté de réaliser dans le cadre d'un doctorat en linguistique anglaise à l'Université de Liège (Fontenelle 1995). Le dictionnaire utilisé dans cette recherche était le dictionnaire anglais-français *Robert & Collins* (Atkins & Duval 1978) dont la version magnétique avait été mise à notre disposition dans le courant des années quatre-vingts, sous contrat de recherche avec les éditeurs (les Dictionnaires Le Robert à Paris et HarperCollins Publishers à Glasgow). Les recherches ont pu être menées à bonne fin grâce au concours précieux de Jacques Jansen et Luc Alexandre, informaticiens à l'Université de Liège. Le premier a conçu la structure de la base de données relationnelle et a analysé les bandes magnétiques du dictionnaire. Le second a participé, à mes côtés, à la création de toute une série de programmes d'application permettant d'interroger la base de données et d'en extraire de l'information.

## 2 Une base de données lexico-sémantiques

Les lecteurs familiers des dictionnaires *Robert & Collins* savent que les lexicographes ont appliqué un système cohérent permettant de spécifier les restrictions de sélections,

les collocations restreintes et les indicateurs de champs sémantiques dans les entrées. Ce système de codage est basé sur la typographie, ainsi que l'illustrent les entrées suivantes:

**grip 4 vi** [*wheels*] adhérer, mordre; [*screw, vice, brakes*] mordre; [*anchor*] crocher (sur le fond)

**jam 4 b** (*become stuck*) [*brake*] se bloquer; [*gun*] s'enrayer; [*door, switch, lever*] se coincer

**line vt clothes** doubler (*with de*); [*bird*] *nest* garnir, tapisser; (*Tech*)

revêtir, chemiser; *brakes* garnir  
**squeal 2 vi a** [*person, animal*] pousser un *or* des cri(s) aigu(s) *or* perçant(s); [*brakes*] grincer, hurler; [*tyres*] crisser  
**squeal 1 n** [*person, animal*] cri aigu *or* perçant; [*brakes*] grincement *m*, hurlement *m*, [*tyres*] crissement *m*

Comme on peut le constater, l'appareil métalinguistique en italiques couvre toute une série d'informations cruciales pour la désambiguïsation et la sélection de la traduction correcte dans un contexte donné. Les informations en italiques vont de la spécification des parties du discours (n, vt, adj, vi...) aux codes matières (Bio, Tech, Scol...), en passant par les collocations et autres restrictions de sélection. Il est clair que, dans les exemples ci-dessus, les mots *person* et *animal* (s.v. *squeal*) font référence à une restriction d'ordre sémantique (le sujet du verbe doit être +HUMAIN ou +ANIMAL) alors que les diverses occurrences de *brake* (= fr *frein*) ou *tyre* (= *pneu*) dénotent une restriction d'ordre collocationnel. On conçoit aisément l'importance de ce genre d'information pour le traitement automatique de la langue et, plus spécifiquement, pour la traduction assistée par ordinateur où la détermination du sens d'un mot dépend des informations contextuelles dont dispose le système. L'approche utilisée par les lexicographes du *Robert & Collins*

pour le codage de ces contraintes peut être résumée comme suit:

- Les noms sujets typiques d'un verbe apparaissent entre crochets;
- Les noms utilisés comme compléments d'un autre nom apparaissent aussi entre crochets;
- Les objets directs d'un verbe et les noms modifiés par un adjectif apparaissent à côté de la partie du discours (pas de crochets ni de parenthèses);
- Les adjectifs, verbes ou adverbes modifiés par un adverbe apparaissent sans crochets ni parenthèses.

Les entrées ci-dessus appellent plusieurs commentaires. Le verbe *squeal* est traduit comme «pousser un cri *or* des cris aigu(s) *or* perçant(s)». La conjonction *or* est utilisée ici par le lexicographe pour gagner de la place mais il est évident qu'une machine ne pourra pas se servir de ce genre de traduction. Un décompactage est absolument nécessaire afin de permettre à la machine de reconnaître quatre traductions (*pousser un cri perçant, pousser des cris perçants, pousser un cri aigu, pousser des cris aigus*). Ce travail de décompactage est difficilement automatisable, mais nécessaire si l'on veut utiliser le champ des traductions comme clé d'accès (et chercher, par exemple, les verbes anglais correspondant à la traduction *pousser un cri perçant*). Ce travail de bénédictin a été mené à bien par Jacques Jansen grâce à un programme interactif qui émettait des hypothèses sur la reconstitution des traductions.

On peut également remarquer que, pour un utilisateur francophone, la partie anglais-français du *Robert & Collins* est un dictionnaire de décodage puisqu'elle permet d'interpréter le texte anglais et de choisir la traduction française correcte en fonction des éléments contextuels de l'appareil métalinguistique. Dans le même temps, on note que les différentes collocations correspondent au modèle de combinaison polaire

développé par Hausmann (1979) puisqu'elles comportent une base et un collocatif. Dans ce modèle bien connu, la base est responsable de la sélection du collocatif dont le sens est en quelque sorte adapté en fonction du contexte. Le problème de l'accès à ces collocations est cependant crucial, ainsi qu'en témoignent les quelques dictionnaires de collocations influencés par les théories de Hausmann, comme par exemple le *BB1* anglais (Benson *et al.* 1986) ou le *Kontextwörterbuch* français-allemand (Ilgenfritz *et al.* 1989). Ces dictionnaires offrent un accès aux collocations par le biais de leurs bases, ce qui répond à un besoin dans une perspective d'encodage. Il faut malheureusement constater que, dans cette optique, les entrées ci-dessus ne sont que d'une faible utilité si le lecteur souhaite partir d'une base quelconque et retrouver les collocatifs potentiels. L'organisation purement alphabétique du dictionnaire ne permet en effet l'accès que via les collocatifs (les entrées proprement dites) alors que les bases correspondent aux éléments en italiques, inaccessibles en tant que clé d'accès primaire dans la version imprimée.

C'est ici que l'informatique apporte une aide précieuse au linguiste ou au traducteur qui a recours au dictionnaire. Dans sa version développée à Liège, la base de données est organisée de façon à permettre l'accès à n'importe quel type d'information, qu'il s'agisse de l'entrée, de la partie du discours, de la base (à l'origine en italiques), de la traduction française, des registres de langue ou des codes matières. Cela signifie qu'il est par exemple possible d'extraire les verbes prenant le nom *brake* comme objet direct (*jam, line, operate, reliné*) ou les verbes intransitifs dont *brake* peut être sujet (*drag, fail, grip, jam, scream, screech, squeal*) ou les traductions françaises respectives de ces verbes (*frotter,*

*lâcher, mordre, se bloquer, hurler, grincer*). On imagine aisément les applications potentielles de ce type de base de données puisqu'il est possible d'accéder en quelques secondes à des données implicites dans la version «papier» du dictionnaire. Pour obtenir le même résultat, il faudrait parcourir les quelque 800 pages du dictionnaire afin d'y repérer, au sein des entrées, les occurrences de *brake* ou de toute autre base en italiques. Les perspectives dans le domaine de l'enseignement de la traduction sont tout aussi évidentes puisqu'il s'agit en fait de développer un dictionnaire en ligne permettant de fournir à l'étudiant ou au traducteur chevronné du matériau linguistique lui permettant d'opérer les choix de traductions les plus judicieux en fonction d'un contexte donné. Or, les recherches sur l'organisation mentale du lexique ont montré que notre cerveau fonctionne par associations d'idées et que nous partons généralement d'un mot ou d'un concept qui, en quelque sorte, active un réseau de mots ou concepts partenaires avec lesquels il entretient une série de relations dites lexico-sémantiques.

### 3 Enrichissement et accès par des fonctions lexicales

La notion de relation lexico-sémantique évoquée dans le paragraphe précédent est une notion clé dans l'approche relationnelle qui sous-tend nos travaux. Si une manipulation informatique du dictionnaire *Robert & Collins* a permis de générer les réseaux sémantiques d'environ 10 000 bases (à l'origine en italiques), ce qui représente un peu plus de 70 000 paires de «partenaires» potentiels (*brakes drag, fail, jam, scream; les freins frottent, lâchent, se bloquent, hurlent...*), il n'en reste pas moins que les liens qui unissent les

diverses composantes de ces collocations sont de nature diverse. Ainsi, des verbes comme *drag-frotter* évoquent une idée d'«obstruction» alors qu'un verbe comme *fail (lâcher)*, en combinaison avec *brake*, signifie que les freins cessent de fonctionner correctement (fin d'un processus). Des verbes comme *scream, screech* ou *squeal*, quant à eux, désignent le son typique émis par les freins (on notera par ailleurs la récurrence des onomatopées à l'initiale en anglais alors que les équivalents français – hurler, grincer – ne présentent pas les mêmes régularités). La diversité des relations dans notre base de données nous a convaincus qu'il serait utile d'explicitier ces liens lexico-sémantiques en les étiquetant pour permettre une modélisation plus fine des réseaux sémantiques ainsi produits. Cette tâche de longue haleine allait ainsi nous permettre d'enrichir le dictionnaire en ajoutant à la base de données une clé d'accès nouvelle, à savoir la nature du lien sémantique unissant l'élément en italiques dans la version papier et l'entrée sous laquelle cet élément apparaît. Cette approche était destinée à permettre à l'utilisateur de la base de données lexicales de poser des questions sémantiquement beaucoup plus fines comme par exemple :

- Y a-t-il des verbes signifiant qu'un frein ne fonctionne pas comme il le devrait?
- Quels verbes peuvent être utilisés pour désigner le bruit/son typique des freins?
- Y a-t-il des verbes transitifs/causatifs signifiant que quelque chose rend les freins inopérants?
- Y a-t-il des substantifs désignant des parties de freins (sabots, patins, garnitures...)?

Les questions ci-dessus évoquent immanquablement le développement de dictionnaires spécialisés d'un genre nouveau, à savoir des dictionnaires

semi-techniques où l'accent est mis sur la combinatoire lexicale restreinte (voir entre autres Frawley 1988). Il ne s'agit certes pas de terminologie à proprement parler, mais il ne s'agit plus non plus tout à fait de langue générale et même un dictionnaire *a priori* non technique comme le *Robert & Collins* s'avère receler une richesse impressionnante en collocations restreintes pour la langue générale et le langage semi-technique.

La question se posait, au début du projet, de pouvoir déterminer un ensemble d'étiquettes permettant d'identifier tous ces liens lexico-sémantiques. Comme nous n'avions pas affaire à un glossaire terminologique nécessitant la détection et le codage de liens hyper-spécialisés (v. Blampain 1993, Blampain *et al.* 1992), nous nous sommes tournés vers la théorie Sens ↔ Texte d'Igor Mel'čuk et plus particulièrement vers son système complexe de fonctions lexicales. Ce système est suffisamment connu, du moins dans les grandes lignes, pour m'en épargner une nouvelle description (v. Mel'čuk *et al.* 1984, Mel'čuk, Clas et Polguère 1995). Je rappellerai tout simplement que le terme *fonction lexicale* désigne une relation de sens assez abstraite telle que l'expression linguistique de cette fonction dépend du lexème auquel elle vient se joindre. En se basant sur le modèle mathématique traditionnel, on utilise la notation  $f(x)=y$ , où  $f$  est la fonction lexicale (FL),  $x$  est le mot clé (la «base» chez Hausmann) et  $y$  la valeur de la fonction. Ainsi, si l'on postule qu'il existe une fonction «très» dénotant l'intensité, on aura:

«très» (malade)= très, gravement (\*grièvement)

«très» (blessé)= grièvement, gravement (mais \*très)

Les fonctions lexicales chez Mel'čuk sont souvent désignées à l'aide de mots d'origine latine. Ainsi, la fonction «très» est mieux connue sous le nom de Magn (Magn (prix)=

haut, élevé...). On distingue une cinquantaine de fonctions, comme par exemple:

– **Oper** (qui codifie les verbes supports, c'est-à-dire presque vides de sens et «supportant» le mot clé): Oper<sub>1</sub> (attention)= faire; Oper<sub>1</sub> (bêvue)= commettre;

– **Liqu** (codifie les verbes exprimant la destruction de quelque chose): Liqu (maladie)= éradiquer; Liqu (loi)= abroger, abolir; Liqu (fichier)= détruire, effacer;

– **Son** (codifie le verbe exprimant le son typique): Son (plancher)= craquer; Son (cloche)= sonner; Son (éléphant)= barrir;

– **Fact**<sub>0</sub> (codifie les verbes exprimant la «réalisation» du mot clé): Fact<sub>0</sub> (rêve)= se réaliser;

– **Mult** (désigne un ensemble régulier): Mult (clé)= trousseau; Mult (abeille)= essaim; Mult (poisson)= banc;

– **Sing** (désigne une portion ou unité): Sing (poussière)= grain; Sing (herbe)= brin; Sing (paille)= fêtu.

Les fonctions lexicales peuvent se combiner pour former des fonctions complexes. Ainsi, IncepOper, dénote le verbe exprimant le début d'un processus (Incep< inchoatif) pour un verbe support: IncepOper<sub>1</sub> (habitude)= prendre. De la même façon, S<sub>0</sub>Son générera le substantif (S<sub>0</sub>) exprimant le son typique d'un mot clé: S<sub>0</sub>Son (éléphant)= barrissement.

Les listes complètes des fonctions lexicales et de leurs propriétés sont données dans Mel'čuk (1984, 1988, 1992) et Mel'čuk *et al.* (1995), entre autres.

Dans le cas de la base de données du *Robert & Collins*, nous avons décidé de coder de façon systématique la fonction lexicale unissant l'élément en italiques et son entrée. Bien sûr, dans un certain nombre de cas (18000 sur un peu plus de 71000 enregistrements), le champ que nous avons créé dans la base de données pour y loger la FL

est resté vide. L'idée était, en effet, de coder la fonction lexicale pour autant que la relation corresponde à une des fonctions paradigmatiques ou syntagmatiques standard du modèle mel'čukien. Nous n'avons donc pas essayé de «forcer» les données dans un moule théorique. Nous n'avons cependant pas voulu non plus adhérer de façon trop flexible au modèle Sens ↔ Texte et nous avons ajouté une série de relations sémantiques non prévues par la théorie des FL. Par exemple, ayant noté que la relation partie-tout était largement représentée dans les données du *Robert & Collins*, nous avons décidé d'utiliser le modèle des fonctions lexicales en créant la fonction *Part* pour rendre cette relation accessible par le biais des programmes d'application, même si Mel'čuk la relègue au rang d'information encyclopédique, et donc non lexicale (*Part* (brake/frein)= drag/sabot, lining/garniture).

Un programme a été écrit afin de faciliter la tâche de l'encodeur et lui permettre de sélectionner la fonction appropriée à partir de menus déroulant émettant des hypothèses basées sur la partie du discours des composantes de la collocation et la nature typographique de l'élément en italiques (selon qu'il apparaissait entre parenthèses, entre crochets, hors crochets, etc.). L'enrichissement de la base de données et l'encodage proprement dit ont réclamé de nombreux mois de travail. Le résultat en valait cependant la peine puisque les programmes d'application permettent maintenant d'accéder à l'information via l'une des clés suivantes:

- Base (à l'origine, l'élément métalinguistique en italiques);
- «Collocatif» (à l'origine, l'entrée vedette);
- Partie du discours;
- Traduction de l'entrée-vedette;
- Traduction de la base;
- Fonction lexicale;

Les deux dernières clés d'accès correspondent à l'information qui a été ajoutée au dictionnaire. Ainsi, les éléments en italiques (par ex. *brake*) ont été systématiquement traduits afin de créer une véritable base de données bilingue. Il est donc possible de demander quels sont les verbes transitifs anglais se combinant avec le mot *brake* ou, sans passer par l'anglais, poser la même question pour le mot français *frein*. Comme on le verra plus bas, il est possible de combiner les clés d'accès ou de les utiliser séparément. Les programmes d'application permettent également d'extraire l'information, de l'exporter ou de mettre la base de données à jour sans en connaître la structure interne (1).

(1) La base de données est interrogeable sur PC et sur station UNIX (Sun SparcStation). Les programmes d'application ont été écrits en C par Luc Alexandre et en Clipper par moi-même. La structure proprement dite de la base de données relationnelle et les possibilités d'exploitation de celle-ci sont décrites en détail dans Fontenelle (1995).

On remarque que la structure des données imite le format  $f(x)=y$ . L'item appartenant à la métalangue en italiques apparaît entre parenthèses (*brake*) avec sa traduction, alors que la vedette contenant dans sa macrostructure une référence à *brake* apparaît après le signe =. La fonction lexicale (part, degrad, fact0, causobstr...) permet de regrouper des mots partageant un certain sens en combinaison avec une base de données. On notera également que certaines fonctions sont en fait des relations 1→n puisque les valeurs qui y sont associées ne sont pas nécessairement synonymes (part indique les différentes parties des freins; preparfact0 génère un ensemble de verbes indiquant les opérations à effectuer pour assurer le fonctionnement normal des freins, même si *line* et *reline* ne sont pas synonymes du tout).

Finalement, l'information entre crochets à la fin de chaque ligne fait référence au contenu d'un champ de la base de données qui indique la

nature typographique de l'item en italiques dans la version imprimée du dictionnaire. Ainsi, [C] signifie que ce terme apparaissait à l'origine entre crochets; [P] signifie qu'il apparaissait entre parenthèses et [S] qu'il apparaissait au niveau de «surface», c'est-à-dire sans parenthèses ni crochets. Comme on l'a vu, l'association de ce type d'information à la spécification de la partie du discours permet de déterminer si on a affaire à une collocation Verbe+Objet, Sujet+Verbe, Adjectif+Nom, Nom+Nom, etc.

Les programmes d'application ont été conçus de façon à pouvoir interroger la base de données souplement en offrant la possibilité de combiner les clés d'accès. Ainsi, le programme d'interrogation *Robcol*, conçu par notre collègue Luc Alexandre, permet, par le biais d'une ligne de commande, de spécifier les conditions et contraintes que l'utilisateur souhaite formuler. En utilisant la notation suivante, l'utilisateur peut sélectionner les champs à interroger en faisant précéder leur abréviation d'un trait d'union:

- i : item à l'origine en italiques;
- h : vedette/entrée (headword);
- pos : partie du discours (part of speech);
- lex : fonction lexicale.

La commande suivante, par exemple, permet d'extraire de la base de données les verbes prenant le mot *law* comme objet direct et signifiant 'liquider', 'détruire', 'éradiquer' (fonction lexicale = liqu):  
robcol -i law -lex liqu.

La base de données contient les informations suivantes (qui peuvent être présentées sous d'autres formes, en ne sélectionnant que certains champs, par exemple, ce qui nous permet de réutiliser ces données pour interroger automatiquement un corpus de textes et valider les informations du dictionnaire):

L'exemple suivant illustre le type d'informations contenues dans la base de données pour le mot *brake*:

part ( <i>brake</i> / frein )	= drag (sabot <m> de frein) [P]
degrad ( <i>brake</i> / frein )	= drag (frotter) [C]
finfact0 ( <i>brake</i> / frein )	= fail (lâcher) [C]
magnfact0 ( <i>brake</i> / frein )	= grip (mordre) [C]
causobstr ( <i>brake</i> / frein )	= jam (bloquer) [S]
obstr ( <i>brake</i> / frein )	= jam (se bloquer) [C]
preparfact0 ( <i>brake</i> / frein )	= line (garnir) [S]
part ( <i>brake</i> / frein )	= lining (garniture <f>) [C]
antifact0 ( <i>brake</i> / frein )	= to be off (être desserré) [C]
fact0 ( <i>brake</i> / frein )	= to be on (être serré) [C]
causfact0 ( <i>brake</i> / frein )	= operate (faire marcher) [S]
a1excesst° ( <i>brake</i> / frein )	= overheated (qui chauffe) [S]
s0antireal1 ( <i>brake</i> / frein )	= release (dégagement <m>) [C]
preparfact0 ( <i>brake</i> / frein )	= reline (changer la garniture de) [S]
son ( <i>brake</i> / frein )	= scream (hurler) [C]
s0son ( <i>brake</i> / frein )	= screech (grincement <m>) [C]
son ( <i>brake</i> / frein )	= screech (grincer) [C]
s0son ( <i>brake</i> / frein )	= sound (bruit <m>) [C]
s0son ( <i>brake</i> / frein )	= squeal (grincement <m>) [C]
son ( <i>brake</i> / frein )	= squeal (grincer) [C]
( <i>brake</i> / frein )	= to tamper with (toucher à (<sans> <permission>)) [S]

abolish (vt) : ~law~  
 → abroger (loi,liqu)  
 annul (vt) : ~law~  
 → abroger (loi,liqu)  
 do away with (vt fus) : ~law~  
 → supprimer (loi,liqu)  
 repeal (vt) : ~law~  
 → abroger (loi,liqu)  
 rescind (vt) : ~law~  
 → abroger (loi,liqu)  
 revoke (vt) : ~law~  
 → rapporter (loi,liqu)

La flexibilité de l'outil permet de sélectionner les verbes transitifs (vt) jouant le rôle de verbe support sémantiquement appauvri (fonction lexicale = oper1) pour des items commençant par la lettre a, ce qui se formule comme suit:

rbocol -i «a\*» -pos «vt\*» -lex oper1

L'utilisation de l'astérisque après «vt» permet d'extraire tous les types de verbes transitifs, même les verbes à particules (phrasal verbs) si fréquents en anglais et codés «vt sep» ou «vt fus» (*bring forward, put forward, put in...*).

## 4 Conclusions

Il n'est pas possible de détailler ici tous les types d'application de la base de données bilingue que nous avons développée. Il va de soi qu'elle recèle des informations très précieuses pour le traducteur ou l'apprenant à la recherche du mot juste. Les possibilités de structuration sémantique offertes par l'enrichissement du dictionnaire à l'aide de fonctions lexicales permettent à l'utilisateur de poser des questions sémantiquement très fines en courant moins le risque de se voir submergé par des données non pertinentes. Nous avons en fait voulu apporter une dimension thésaurique à notre dictionnaire informatisé dans le but de créer une nouvelle génération d'outils lexicographiques. Plus fondamentalement, nous avons également tenté d'utiliser cette base de données enrichie pour des études de sémantique lexicale où l'accès à

une grande quantité de données est crucial pour pouvoir dégager des généralisations, par exemple en se concentrant sur la réalisation, à travers tout le lexique, d'une fonction lexicale particulière (voir par exemple Fontenelle 1994 sur la relation entre les fonctions lexicales et certains types de métaphores). Dans une autre optique, notre équipe compte exploiter ce dictionnaire dans une perspective de désambiguïsation automatique et de sélection de la traduction appropriée. Nous sommes convaincus que les recherches en linguistique computationnelle et en lexicographie vont, de plus en plus, nécessiter le développement de grandes bases de données lexicales, mono- et multilingues. Le *Robert & Collins* liégeois et les réseaux lexico-sémantiques qu'il contient n'offrent probablement qu'une parcelle de solution au problème complexe de la désambiguïsation. Nous avons cependant la faiblesse de croire qu'associé à d'autres outils, il pourrait jouer un rôle non négligeable.

*Thierry Fontenelle,  
 Département d'anglais,  
 Faculté de Philosophie et lettres,  
 Université de Liège.*

Échantillon de la liste:

lay (vt) : ~accusation~  
 sling (vt) : ~accusation~

fling (vt) : ~accusation~

enjoy (vt) : ~advantage~

give (vt) : ~answer~

give (vt) : ~answer~

proffer (vt) : ~apology~

present (vt) : ~apology~

tender (vt) : ~apology~

offer (vt) : ~apology~

put in (vt sep) : ~application~

bring forward (vt sep) : ~argument~

develop (vt) : ~argument~

put forward (vt sep) : ~argument~

pull out (vt sep) : ~argument~

prefer (vt) : ~argument~

pose (vt) : ~argument~

exert (vt) : ~authority~

exercise (vt) : ~authority~

wield (vt) : ~authority~

→ porter (accusation,oper1)

→ lancer (<at> <sb> à qn)

(accusation,oper1)

→ lancer (<at> <sb> à qn)

(accusation,oper1)

→ jouir de (avantage,oper1)

→ faire (réponse,oper1)

→ donner (réponse,oper1)

→ offrir (excuse,oper1)

→ présenter (<to> à) (excuse,oper1)

→ offrir (excuse,oper1)

→ offrir (excuse,oper1)

→ faire (candidature,oper1)

→ avancer (argument,oper1)

→ développer (argument,oper1)

→ avancer (argument,oper1)

→ sortir [informal] (argument,oper1)

→ présenter (argument,oper1)

→ présenter (argument,oper1)

→ exercer (autorité,oper1)

→ exercer (autorité,oper1)

→ exercer (autorité,oper1)

## Remerciements

Les recherches décrites dans cet article s'inscrivent dans le cadre du projet Decide (*Designing and Evaluating Extraction Tools for Collocations in Dictionaries and Corpora - Multilingual Action Plan - MLAP 93/19*) partiellement financé par la Commission européenne et clôturé en janvier 1996. Outre l'Université de Liège (coordinateur du projet), le consortium comprenait l'Institut für maschinelle Sprachverarbeitung (Université de Stuttgart) et le Rank Xerox Research Centre de Grenoble.

Outre les éditeurs (HarperCollins Publishers, Glasgow et les Dictionnaires Le Robert, Paris), qui ont gracieusement mis à notre disposition les versions magnétiques du *Robert & Collins*, il me faut remercier ici Messieurs Jacques Jansen et Luc Alexandre dont les compétences informatiques ont permis l'aboutissement de ce travail.

## Bibliographie

- Atkins (B.T.S.) & Duval (A.), 1978: *Robert & Collins Dictionnaire anglais-français, français-anglais*, Paris, Le Robert, Glasgow, Collins.
- Benson (M.), Benson (E.) & Ilson (R.), 1986: *The BBI Combinatory Dictionary of English*, Amsterdam & Philadelphia, John Benjamins.
- Blampain (D.), 1993: «Notions et phraséologie. Une nouvelle alliance?», dans *Terminologies Nouvelles*, 10, décembre 1993, pp.43-49.
- Blampain (D.), Petrusa (P.) & Van Campenhoudt (M.), 1992: «À la recherche d'écosystèmes terminologiques», dans *L'Environnement traductionnel - la station de travail du traducteur de l'an 2001; Journées scientifiques du Réseau thématique de recherche Lexicologie, terminologie et traduction*, Actes du colloque (Mons, 25-27 avril 1991), Presses de l'Université du Québec et Aupelf-Uref, pp.273-282.
- Boguraev (B.) & Briscoe (T.), 1989: *Computational Lexicography for Natural Language Processing*, Longman & New York, Longman.
- Cowie (A.P.), 1989: *Oxford Advanced Learner's Dictionary*, Oxford University Press.
- Fontenelle (Th.), 1994: «Towards the construction of a collocational database for translation students», dans *Meta*, vol. 39, n° 1, pp.47-56.
- Fontenelle (Th.), 1994: «Using lexical functions to discover metaphors», dans Martin *et al.* (éd.) *Euralex'94 Proceedings*, Vrije Universiteit Amsterdam, pp.271-278.

Fontenelle (Th.), 1995: «Réseaux sémantiques et dictionnaires bilingues électroniques», à paraître dans les *Actes des 4es Journées scientifiques du Réseau Lexicologie, terminologie et traduction: dictionnaire et lexicomatique*, Aupelf-Uref (Lyon, septembre 1995).

Fontenelle (Th.), 1995: *Turning a bilingual dictionary into a lexical-semantic database*, thèse de doctorat, Université de Liège.

Frawley (R.), 1988: «New forms of specialized dictionaries», dans *International Journal of Lexicography*, vol. 1, n° 3, pp.189-213.

Hausmann (F.), 1979: «Un dictionnaire des collocations est-il possible?», dans *Travaux de Linguistique et de Littérature*, vol. 17, n° 1, pp.187-195.

Heid (U.), 1994: «On Ways Words Work Together - Topics in Lexical Combinatorics», dans Martin *et al.* (éd.) *Euralex'94 Proceedings*, Vrije Universiteit Amsterdam, pp.226-257.

Ilgenfritz (P.), Stephan-Gabinel (N.) & Schneider (G.), 1989: *Langenscheidts Kontextwörterbuch Französisch-Deutsch*, Langenscheidt, Berlin-München.

Mel'čuk (I.) *et al.*, 1984/1988/1992: *Dictionnaire explicatif et combinatoire du français contemporain: Recherches lexicosémantiques I, II, III*, Presses de l'Université de Montréal.

Mel'čuk (I.), Clas (A.) & Polguere (A.), 1995: *Introduction à la lexicologie explicative et combinatoire*, Aupelf-Uref & Duculot (Universités francophones).

Procter (P.), 1978: *Longman Dictionary of Contemporary English*, Longman, Harlow.

Sinclair (J.), 1987: *Collins Cobuild English Language Dictionary*, HarperCollins Publishers.

## *Trader*, aux frontières du néologisme

Un incident financier du début de l'année 1995 (la faillite de la banque Barings) a fait (ré)apparaître dans la presse un terme (*trader*) qui appartenait depuis plusieurs années à la terminologie des milieux économique-financiers. *Trader* est actuellement accompagné des «attributs» qui signalent les néologismes (guillemets, italiques, etc.). Cet exemple soulève des questions d'ordre théorique touchant à la définition même du concept de *néologisme* et des questions d'ordre plus pratique et méthodologique.

Termes-clés :  
emprunt, néologisme de forme,  
néologisme de sens, situation de  
communication, statut néologique.

**L**a faillite de la banque Barings, au début de l'année 1995, a donné lieu à des pratiques linguistiquement intéressantes. À cette occasion les médias ont utilisé le terme *trader* pour désigner l'activité professionnelle de la personne à l'origine de cette faillite. C'est le statut de cette unité qui pose des questions de différents ordres. D'une part, *trader* (emprunt à l'anglais) est employé, depuis plusieurs années, dans les milieux de la finance et de l'économie. Aussi, d'un point de vue purement terminologique et à ne s'en tenir qu'au signifiant, il ne peut être question de considérer *trader* comme un néologisme. Mais d'autre part, cet «incident» financier a élargi son champ d'utilisation qui passe d'un emploi technique, limité à un milieu professionnel, à un emploi plus généralisé.

Dans ce contexte la question du caractère néologique de l'unité concernée se pose en d'autres termes et, pour y répondre, plusieurs paramètres doivent être pris en considération. La typographie (utilisation ou non d'italiques ou de guillemets), le contexte (présence éventuelle d'énoncés définitoires, de synonymes, etc.), le sémantisme (sens fluent ou fixé), la prononciation, etc.,

sont des éléments déterminants qui permettront de préciser le statut de cette unité.

Nous avons effectué cette étude essentiellement à partir de deux journaux, un hebdomadaire économique-financier : *L'Expansion*, et un quotidien «généraliste» : *Le Monde* (dont le CD-ROM facilite la consultation et rend confortable l'interrogation). Quand nécessaire pour étayer notre argumentation, nous avons élargi notre corpus à d'autres journaux ou revues et nous nous sommes autorisé quelques incursions dans la presse parlée. À ce corpus «journalistique» nous avons adjoint :

- 1) Des dictionnaires de spécialité :
  - a. Dictionnaires bilingues à double entrée (anglais-français, français-anglais) : *Glossaire de la Banque mondiale* (1989), *la langue des affaires*, *Dictionnaire commercial et économique bilingue* (1989).
  - b. Dictionnaires monolingues français, *Lexique de l'économie et des finances*, Ahmed Silem, Jean-Marie Albertini (1989, 3<sup>e</sup> éd.); *Dictionnaire du marketing et des affaires*, Rémi Van Dick, 1991).
- 2) Des dictionnaires généraux : le *Petit Larousse illustré*, 1993 (PLI), le *Trésor de la langue française* (TLF), le *Grand Robert*, 1986 (GR), le *Nouveau Petit Robert*, 1993 (NPR).

# France

## 1 *Trader*, des premières attestations jusqu'au début des années 1990

On rencontre assez tôt des attestations de *trader* dans la presse spécialisée. M. Chansou, dans un article du numéro spécial de *La banque des mots* (Chansou 1990), date l'apparition du terme dans la presse française de 1987. Mais dès 1974 nous trouvons plusieurs occurrences de *trader*<sup>(1)</sup>, dans le numéro de mai de *Réalités*: «Pryde a vécu au milieu des Esquimaux de 1958 à 1970 comme acheteur de fourrures pour la Compagnie de la Baie d'Hudson. C'est donc un témoignage de première main [son livre], passionné et passionnant sur les populations de l'Arctique, mais aussi sur les «traders» qui protègent leur existence, et qui peu à peu influencent leur mode de vie de façon non négligeable», ainsi que dans une annonce, parue dans *L'Express* du 4 février de la même année: «Assistant «Oil and Bunkers department» PARIS pour courtier pétrolier et «trader» international». Il faudra attendre 1977 pour rencontrer une nouvelle attestation<sup>(2)</sup> de *trader*: «La fièvre monte sur le marché mondial du café. Mercredi 17 août, en plein cœur de la City de Londres, le «coffee market» de la Mark Lane Street est en ébullition. Dans la salle surchauffée, de jeunes «traders» en manches de chemise s'interpellent: les vendeurs, peu nombreux, mènent la vie dure aux acheteurs.» (Agnès Rebattet, «café: un grain de folie», dans *L'Express*, 22-28 août.)

(1) Momentanément nous ne considérerons que la forme de l'unité, sans nous occuper des différents sens que celle-ci a pu avoir et a encore en français.

(2) Ces exemples proviennent des dépouillements effectués par l'Observatoire du français contemporain, du CNRS.

Pour les dictionnaires de spécialité consultés, *trader* n'apparaît, dans les bilingues, qu'à la macrostructure de la partie anglaise et les traductions en français correspondent aux sens de la langue générale en usage en anglais: «*trader*, 1. commerçant, marchand, négociant [...]; 2. trafiquant (nuance péjorative)» (*La langue des affaires*, 1989). Le *Glossaire de la banque mondiale* (1989) ajoute «opérateur [Bourse]». Les dictionnaires monolingues en français ne retiennent que le sens du domaine de la finance, qui seul nous intéresse ici: «*Trader*. – Fin. – Opérateur sur marché à terme. Le terme est également utilisé pour désigner ceux qui prennent des positions purement spéculatives», (*Lexique de l'économie*, 1989); «*trader* (mot anglais). Spéculateur sur les marchés à terme. Le trader intervient pour réaliser des profits, au contraire du «*merger*» ou «*adosseur*», qui se couvre contre les variations de cours», (*Dictionnaire du marketing et des affaires*, 1991). C'est aussi le sens retenu par le NPR (1993): «*TRADER* [tɾɛdœr] n.m.—v. 1980; mot angl. «marchand» ♦ ANGLIC. Opérateur de marchés financiers. → *broker*». Les autres dictionnaires généraux n'enregistrent pas le terme.

Les exemples tout comme les définitions des dictionnaires appellent plusieurs remarques. *Trader*, à l'origine, est employé en français à l'occasion de relations commerciales qui s'effectuent dans un contexte spécifiquement anglo-saxon. C'est ce contexte qui induit l'utilisation du mot anglais, l'équivalent français aurait très bien pu lui être substitué, sans que le sens de l'énoncé en soit modifié, sans introduire un déficit informationnel dans la communication. La prononciation, quand elle est précisée, reste strictement anglophone. Aux «sens de base» de *trader* («commerçant», «marchand», etc.) s'ajoute un sens qui s'inscrit dans une terminologie

précise: celle de la finance, de la bourse et c'est cette acception que retiendra par la suite le français.

## 2 *Trader*, dans la presse après 1990

On pourrait s'attendre à ce qu'un emprunt attesté depuis au moins 1974 soit suffisamment intégré au système lexical et ait acquis une stabilité de sens et d'usage qui dispense son utilisateur de signaler une certaine «distance» par rapport à cette unité, comme c'est le cas pour *walkman*, par exemple, qui a été emprunté à l'anglais sensiblement à la même date que *trader*. Les deux termes évolueront par la suite différemment. Dès le 18 février 1983, le *Journal officiel* proposait pour «*walkman*, qui est une marque déposée, [et] ne doit pas être utilisé» (*Dictionnaire des termes officiels de la langue française*) un équivalent français *baladeur* qui est le seul terme en usage dans les textes officiels.

Or, en 1995, *trader* continue à être accompagné de ces «marques [qui] constituent autant de clins d'œil à l'interlocuteur» (L. Guilbert, article *Néologie*, dans le GLLF) et qui ont une fonction néologique (Guilbert 1973). Par exemple, dans *L'Expansion* du 20 mars toutes les occurrences de *trader* (8 au total) sont en italiques quel que soit le rédacteur des articles (Gilles Pouzin, «Chaos monétaire: il faut changer les règles», p.14-16, et Philippe Chalmin, «Barings, un drame en deux actes», p.98-100).

Pour des raisons conjoncturelles, que nous avons rappelées au début de cette étude, *trader* a une fréquence d'emploi beaucoup plus importante en 1995. La consultation du CD-ROM du *Monde* pour les années 1993 à 1995, met bien en évidence cette variation (8 occurrences de *trader(s)* en 1993, 13 en 1994 et 36 pour le seul mois de mars 1995). La majorité

des attestations (≈68%) sont accompagnées de ces «marques» que L. Guilbert évoquait et qui signalent le statut néologique de l'unité concernée. Les auteurs des articles qui utilisent ces «marques» adoptent différentes stratégies: *trader* est soit guillemeté (12 occurrences), soit accompagné d'énoncés définitoires: «les professionnels de la spéculation (traders)» (François Renard, «Purge sur le dollar», 2-05-1994), ou d'«explicitation du contenu sous forme d'une traduction par un mot déjà connu» (Guilbert 1973). Dans plusieurs articles, *trader* commute avec des unités du même champ sémantique: «Nick Leeson l'opérateur indélicat, est recherché aujourd'hui par Interpol [...]. Le trader a pu conspirer avec un criminel» (Marc Roche, 1-03-1995), «[...] les traders décompressent. Un cambiste lit, les pieds sur la table, le Canard enchaîné» (Arnaud Le Parmentier, 10-03-1995), «Le profil des professionnels du Forex [...] n'a rien à voir avec celui d'un golden boy terni comme Nick Leeson, le trader fou de Singapour» (*id.*) etc.

Pour la période considérée (1990-1995), l'avis des auteurs sur le statut de *trader* paraît partagé. Pour certains c'est une unité qui est intégrée au stock lexical de tous les locuteurs francophones et elle a perdu tout caractère néologique, ce que semblerait confirmer la presse parlée qui adapte la prononciation de *trader* pour l'intégrer au système phonétique français. À plusieurs reprises, en mars 1995, il a été question lors des informations télévisées ou des informations radiophoniques du [tRadɛR] Nick Leeson. Pour d'autres, au contraire, *trader* reste un néologisme qui doit être marqué comme tel, qui est encore suffisamment instable pour qu'il soit nécessaire d'en rappeler ou d'en préciser le sens. Curieusement ce sont surtout les spécialistes des domaines

financier et économique qui adoptent (dans une démarche pédagogique?) cette option. Philippe Chalmin, conseiller économique de la Sfac et professeur à Paris IX-Dauphine, utilise toujours l'italique dans son article de *L'Expansion* auquel nous nous sommes référé. Dans *Le Monde*, ce sont généralement les spécialistes des rubriques financières du journal qui guillemettent le plus fréquemment *trader*.

### 3 La délicate étude des néologismes

*Trader* concentre plusieurs problèmes que l'on rencontre assez régulièrement dès que l'on se livre à l'étude, à l'analyse un tant soit peu rigoureuse des néologismes et rend assez bien compte de la complexité méthodologique et théorique que rencontre tout néologue. Nous n'aborderons, bien évidemment, que quelques-unes des questions que soulève la néologie. Nous ne poserons pas, par exemple la question théorique de la définition du concept même de néologisme, nous nous situerons dans une optique beaucoup plus pratique.

Si l'on adopte pour définition de néologisme celle que rappelle J. Humbley dans le numéro spécial de *La banque des mots* de 1993: «Le néologisme est défini comme toute forme ou tout emploi non répertorié, formé régulièrement ou irrégulièrement en français, ou emprunté à une autre langue» (Humbley 1993: 68), il faut porter la même attention à la forme qu'au sens de l'unité nouvelle et toujours garder en mémoire qu'un néologisme de forme peut donner lieu, dans une synchronie plus ou moins rapprochée à un néologisme de sens. Il nous semble bien que c'est en partie ce qui s'est produit pour *trader*. Les premiers emplois que nous avons rencontrés

sont assez strictement synonymes de «commerçant, marchand, négociant», et le domaine du commerce peut être tout aussi bien celui des fourrures que celui du pétrole. Pour certaines attestations (postérieures à 1990) que nous avons signalées comme n'étant pas accompagnées des «attributs» néologiques traditionnels, c'est dans ce sens que *trader* est employé (*trader en produits alimentaires*). Si le signifiant est le même, le signifié, lui, est différent, un nouveau sens qui s'inscrit dans une terminologie bien précise est apparu, et là réside une des causes de la persistance du terme en français. M. Chansou (1990) dans son article de *La banque des mots* montre bien que si *trader* subsiste, c'est qu'il occupe une case dans le système qui n'est pas couverte par les quasi-synonymes.

Un autre point nous paraît déterminant dans toute démarche néologique: «Une des principales difficultés de l'observation de la néologie n'est pas seulement le repérage d'un sens nouveau, mais plutôt la reconnaissance et la description notionnelle et linguistique de ce qu'on a repéré» (Humbley 1993: 67). L'analyse linguistique passe, pour les langues de spécialité, par la prise en compte des «situations de communication». Le vocabulaire utilisé n'est pas le même selon que la communication s'établit entre «spécialistes d'un domaine», entre «spécialistes et un public averti mais non spécialiste» et entre «journaliste spécialisé et public plus large» (Löffler-Laurian 1991). *Trader* apparaît d'abord dans des situations de communication bien précises: c'est un terme de finance que les spécialistes utilisent et qui a pour eux une définition bien précise que nous avons rappelée précédemment et qui inscrit «le nouveau concept dans un système hiérarchisé de notions préexistantes» (Humbley 1993: 67) et à ce niveau *trader* n'est synonyme ni de *courtier*, ni de *cambiste*, ni

d'*opérateur*. En 1995, nous sommes dans une autre situation de communication. Les spécialistes, les journalistes spécialisés s'adressent à un public large: les lecteurs de *L'Expansion*, du *Monde* qui ne sont pas forcément familiers avec la terminologie financière. Les auteurs des articles signalent par la typographie le caractère néologique du mot, ils donnent les éléments définitoires qui permettront aux lecteurs d'en connaître approximativement le sens («professionnel de la spéculation») et de le situer dans un champ sémantique: *trader* est associé dans un même énoncé à *golden boy*, à *cambiste*, à *opérateur*, etc. *Trader* est intégré au stock lexical de la langue générale, ce qui fait aussi bouger le sémantisme de l'unité. Entre ces deux moments, un temps plus ou moins long peut s'écouler. Que *trader*, vingt ans après sa date «d'apparition» dans la langue française, soit encore considéré et traité comme un néologisme n'est donc pas étonnant. Dans un autre domaine de spécialité, *cœlacanthe* en est un bon exemple: «tel terme technique-scientifique ancien peut-être considéré à bon droit comme néologisme à un autre niveau [...]. Il ne faut pas confondre les nouveautés dans une classe de locuteurs avec la nouveauté dans une situation de communication» (Rey 1974).

À ne pas tenir compte de ces données on risque de dater une première attestation en passant à côté de la complexité du sémantisme du néologisme.

*Chantal Girardin,*  
*Centre de terminologie et de néologie,*  
*Laboratoire de linguistique*  
*informatique,*  
*CNRS, Ura 1576,*  
*France.*

## Bibliographie

Aitonhuen (Emmanuel), 1994: «La tension entre l'emprunt et la néologie de souche française», dans *Initiales*, vol. 14, p. 33-40.

Chansou (Michel), 1990: «Termes de la finance et de la bourse, Essai d'organisation d'une terminologie», dans *La banque des mots*, n° spécial, p. 35-43.

Guilbert (Louis), 1973: «Les Vocabulaires techniques et scientifiques», dans *Langue française*, n° 17, Paris, Larousse.

*Dictionnaire des termes officiels de la langue française*, 1994: Paris, Délégation générale à la langue française.

Guilbert (Louis), 1974: «La néologie lexicale», dans *Langage*, n° 36, Paris, Didier-Larousse.

Humbley (John), 1993: «L'observation de la néologie terminologique, l'expérience du CTN», dans *La banque des mots*, n° spécial 5, p. 65-73.

Kocourek (Rostislav), 1991: *La langue française de la technique et de la science*, 2<sup>e</sup> édit., Wiesbaden, Brandstetter Verlag.

Loffler-Laurian (Anne-Marie), 1991: «La relation émetteur récepteur dans le discours de l'exposition scientifique», dans *Alfa*, vol. 3/4, p. 145-146.

Rey (Alain), 1974: «Essai de définition du concept de néologie», dans *L'aménagement de la néologie, Actes du colloque international de terminologie*, Québec, Lévis, 29 sept-20 oct 1974.

Rey (Alain), 1992: *La terminologie, noms et notions*, 2<sup>e</sup> édit., Paris, PUF (Que sais-je).

# Analyse de textes spécialisés pour le recueil de données terminologiques

Les nouveaux besoins en terminologie demandent que soient définis de nouveaux modèles et de nouvelles méthodologies. Trois caractéristiques principales que devraient respecter ces nouveaux modèles de données sont présentées : distinction des données linguistiques et conceptuelles, mise en réseau des relations conceptuelles, renseignements précis sur l'usage. La meilleure façon d'obtenir ces informations consiste à travailler sur des corpus. Il faut définir des méthodes assistées par des outils informatiques pour repérer ces données dans les textes. L'article propose une première description des tâches à effectuer et des outils permettant de les mener à bien.

Termes-clés : analyse de corpus, bases de connaissances terminologiques, outils informatiques, sémantique textuelle.

## 1 Introduction

**D**e nombreux projets associés à l'amélioration de la communication en entreprise (cf. *Terminologies nouvelles*, n° 13) ou à la veille technologique font apparaître de nouveaux besoins concernant la terminologie. Dans un article à paraître dans *Terminogramme*, nous avons défini trois classes d'éléments qui devraient caractériser les besoins pour les futures bases de données terminologiques. La prise en compte de ces éléments permet de constituer ce qu'on appelle désormais les *Bases de connaissances terminologiques (BCT)* (cette appellation venant se substituer à celle de *Base de données terminologiques (BDT)*). Le recueil de ces nouvelles données terminologiques nécessite la mise en œuvre de nouvelles méthodes de travail et même d'analyse des données. En particulier, afin de rendre compte au plus juste du fonctionnement de la terminologie, il est indispensable de travailler sur des textes réellement produits dans les entreprises, considérés comme représentatifs de la langue de l'entreprise. Le travail à effectuer se rapproche alors de celui de l'analyse de corpus, mais les tâches et les moyens de les effectuer, en particulier en s'aidant d'outils informatiques, restent encore à définir.

L'article présente le modèle de Base de connaissances terminologiques (BCT) que nous

avons défini. Puis, il propose le détail des tâches à mener à bien pour recueillir les données terminologiques dans les textes. Enfin il s'essaie à une première description de la mise en œuvre d'outils pour assister ces tâches.

## 2 Modèle de BCT

D'abord dans le cadre d'Aramihs (Action recherche et application Matra-Irit en interface homme-système) puis dans celui de l'ERSS (Équipe de recherche en syntaxe et sémantique), nous avons mis au point un modèle simple de BCT qui a été mis en œuvre pour différents types d'application, d'abord à Matra Marconi Space (MMS) puis à Aérospatiale et au Centre national d'études spatiales (CNES). Nous reprenons les grandes lignes de ce modèle ici. Pour plus de précisions, voir Condamines & Amsili (1993) et Condamines (1994).

### 2.1 Présentation du modèle

Le modèle est organisé en trois champs:

- Un champ «terme» (T) qui contient les informations linguistiques (formes du signifiant, nature, genre, type de complémentation... );
- Un champ «lien terme/concept» (L T/C) qui concerne les conditions d'usage, c'est-à-dire les éléments qui, dans la situation de communication, influent sur l'établissement de(s) lien(s) entre un (des) terme(s) et un (des) concept(s);

– Un champ «concept» qui permet d'enregistrer les données concernant le concept dénommé par le terme; ces concepts sont mis en réseau par l'intermédiaire de relations qui ont comme particularité d'être consensuelles (partagées par l'ensemble des locuteurs) et réutilisables quelles que soient les applications ultérieures.

Les relations entre termes sont calculées. Ainsi, deux termes seront totalement équivalents si :

- Ils sont reliés au même concept;
- Ils ont la même nature grammaticale; en effet, nous considérons par exemple que, du point de vue conceptuel, un verbe et un nom, ou un adjectif et un nom... peuvent désigner le même concept, par exemple, *gel d'un test* et *geler un test* renvoient au même concept; mais, si on parle d'équivalence linguistique, alors, l'appartenance à la même catégorie grammaticale est requise;

- Les valeurs respectives des attributs sont identiques; si ces valeurs ne sont pas identiques, l'équivalence n'est pas totale.

Dans le schéma ci-dessous, T1 et T2 (qui renvoient au même concept) sont identiques si la catégorie grammaticale de T1 est la même que la catégorie grammaticale de T2 et que LT1/T2 et LT2/C1 sont identiques.

On peut, de la même façon, calculer l'équivalence interlinguistique en ajoutant une restriction sur la langue.

Les concepts C1, C2, C3, C4 et C5 sont reliés par des relations conceptuelles «est-un», «partie-de», «éventuellement cause», «conséquence»... Chaque concept se manifeste par une réalisation linguistique (T) ou par plusieurs comme le concept C1. Les liens LT/C rendent compte des conditions d'usage.

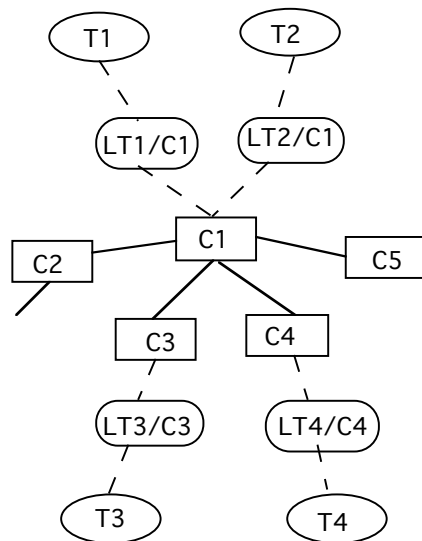


figure 1 : exemple de BCT

## 2.2 Les trois éléments principaux

Trois caractéristiques de ce modèle nous semblent indispensables pour répondre aux nouveaux besoins en matière de terminologie :

### 2.2.1 Distinction du champ concernant les données linguistiques et du champ concernant les données conceptuelles

On sait que le terme fonctionne à la fois comme signe linguistique et comme dénomination de concept. Les deux types d'informations, linguistiques et conceptuelles, sont généralement enregistrées dans les bases de données terminologiques (BDT) mais elles ne sont pas forcément bien identifiées et distinguées. Or, il convient de ne pas confondre ces données, d'une part parce que, théoriquement, la confusion des deux niveaux conduit à une identification du sémantique avec le conceptuel, ce qui est erroné; d'autre part, en fonction des utilisations, il se peut que les besoins

d'enrichissement concernent soit les données linguistiques, soit les données conceptuelles. En effet, si la BCT est conçue comme une ressource de base, consensuelle et réutilisable, sa mise en œuvre va nécessiter, dans la plupart des cas, que les données soient complétées. Une utilisation de type Traitement automatique de la langue entraînera un enrichissement des données linguistiques et une utilisation de type Représentation des connaissances, un enrichissement des données conceptuelles.

### 2.2.2 Mise en réseau des concepts

Les données conceptuelles apparaissent généralement dans les BDT sous une forme rédigée, dans la définition. Il est clair que parce qu'il s'agit de termes, la définition sert ici à situer l'élément étudié dans le système conceptuel (contrairement à un mot que la définition devrait servir à situer dans le système linguistique). Or, s'il y a un système conceptuel, il doit être possible de définir, à partir des formes linguistiques présentes dans les définitions, quelles relations conceptuelles se manifestent. Les recherches n'étant pas encore suffisamment avancées, il semble prématuré, à l'heure actuelle, de représenter uniquement sous forme de réseau toutes les définitions. Pour les relations les plus fréquentes, cela est toutefois possible tout en conservant, pour le moment, la définition rédigée. Dans le cas où il n'y a pas de définition déjà faite, il est préférable de chercher directement dans les corpus ces relations conceptuelles. (cf. 3.3).

### 2.2.3 Informations précises sur l'usage

Pour les besoins qui commencent à se faire sentir en entreprise, ce sont là des informations qui jouent un rôle très important. En effet, les entreprises qui se lancent dans la maîtrise de leur

communication en interne ont besoin non seulement d'informations sur la terminologie du domaine dans lequel elles évoluent mais aussi d'informations sur l'usage réel dans leur entreprise. Pour reprendre la comparaison avec la langue générale, les entreprises ont besoin d'informations non seulement sur la compétence mais aussi sur la performance. Qui utilise quel terme pour parler de quel élément ? Telles sont les questions qui doivent recevoir une réponse grâce au travail du terminologue-linguiste. Jusqu'à présent, pour des raisons évidentes de coût, ces informations étaient peu recherchées. Or, à la fois parce que les besoins évoluent et parce que le travail est de plus en plus assisté par des machines, on peut envisager de donner des réponses à ces questions et les intégrer dans la BCT.

Comment ces différents types d'informations peuvent être retrouvés dans les corpus et comment ce travail peut être assisté par des outils ? Nous essayons de répondre à ces questions dans les prochains paragraphes.

### 3 Les tâches à effectuer pour retrouver les données terminologiques dans les corpus

Différentes tâches, toutes basées sur l'analyse de corpus, doivent être effectuées pour déterminer les données terminologiques correspondant au modèle défini en 2.

#### 3.1 Analyse de corpus

L'analyse de corpus représentatifs du domaine (textes écrits ou retranscriptions d'entretiens, d'exposés...) permet de rendre compte de données conformes à l'usage réel qui est fait par les locuteurs du

domaine voire d'une entreprise du domaine. En réalité, il s'agit de mettre en évidence les caractéristiques non seulement de la langue spécialisée utilisée mais aussi des discours spécialisés, c'est-à-dire, de rendre compte des compétences du locuteur manifestées à travers l'actualisation de la langue spécialisée dans des discours, «car c'est bien dans le texte que la signification s'inscrit dans la réalité de la communication» (Descamps 1992 : 334).

Il s'agit de caractériser le comportement linguistique des formes à l'étude, à partir de l'examen de leurs occurrences en corpus. L'analyse s'appuie sur les contextes d'apparition de chaque forme afin de déterminer les parentés, le plus souvent sémantiques, qui peuvent être établies par généralisation à partir de manifestations lexico-sémantiques différentes. L'objectif est d'utiliser des connaissances que l'on a sur le fonctionnement de la langue générale afin d'opérer des classifications de contextes en langue spécialisée. Si l'analyse est suffisamment poussée, on peut définir des patrons lexico-syntaxiques qui définissent ces classes de fonctionnement des contextes (cf. exemples en 3.3).

Notons que nous faisons l'hypothèse, que nous discutons ailleurs (à paraître dans Condaminis 1995), qu'une continuité s'établit entre signifié (auquel nous accédons par l'analyse linguistique) et concept. Nous parlerons désormais de concept en lieu et place de signifié et de relations conceptuelles en lieu et place de relations sémantiques.

Prenons le cas où deux termes X et Y ont été repérés et pour lesquels on souhaite définir les données terminologiques conformément au modèle décrit en 2.

Trois séries de tâches sont à effectuer : les premières sont centrées sur les variantes de formes linguistiques (terme (T)), les

deuxièmes sur les relations terme/concept (lien terme/concept (L T/C)); les troisièmes sur les relations entre concepts (C).

#### 3.2 Tâches centrées sur le repérage des variantes de signifiant, analyse des variantes d'une même forme linguistique

Des formes morfo-lexicalement apparentées à X ou Y apparaissent dans des contextes sémantiquement identiques à ceux dans lesquels apparaissent X ou Y. L'analyse des contextes ne faisant apparaître qu'une seule classe sémantique, on peut en déduire qu'il n'y a qu'un concept identifiable sous des formes différentes. Ces formes sont soit des sigles – *Scao* pour *Système de contrôle d'attitude et d'orbite* – soit des variantes linguistiques (ellipses ou autres) – par exemple *segment moyen de l'IVA* peut apparaître sous la forme *IVA moyenne*.

#### 3.3 Tâches centrées sur la relation terme/concept

##### 3.3.1 Analyse des dénominations alternatives : 2 termes (non apparentés morphologiquement)/un seul concept

On a affaire à une seule classe de contextes, c'est-à-dire un seul type de fonctionnement sémantique, entourant des formes n'ayant pas de parentés morphologiques. Comme dans le cas précédent, il n'y a qu'un seul concept mais, cette fois, deux signifiants non apparentés morphologiquement. Par exemple, dans la terminologie de Matra Marconi Space, *senseur* et *capteur* sont équivalents. Dans ce cas, deux termes sont créés, chacun étant en relation avec le même concept.

### 3.3.2 Analyse des polyacceptions : 1 seul terme/plusieurs classes de contextes

Une même forme linguistique apparaît dans des contextes qui se généralisent en au moins deux classes. Dans ce cas, au moins deux cas de figure peuvent se présenter. Soit la forme est homonymique, soit la forme est polysémique.

Si la forme est homonymique, deux ou plusieurs classes de contextes apparaissent sans parentés sémantiques; deux termes sont créés, chacun étant en relation avec un concept particulier. Si la forme est polysémique, deux ou plusieurs classes de contextes apparaissent qui entretiennent des parentés sémantiques évidentes; un seul terme est créé, qui est en relation avec plusieurs concepts, eux-mêmes entretenant des relations. Ce cas de figure apparaît souvent avec les déverbaux (les noms reliés à un verbe), fréquents dans les langues de spécialité; ces noms dénomment fréquemment soit l'action, soit le résultat de l'action : un achat peut correspondre soit à l'action d'acheter («au moment de l'achat de sa voiture...») soit au résultat de cette action («pose tes achats sur la table»).

Dans certains corpus, la polyacception (un terme/plusieurs concepts), concerne un seul référent (au sens d'élément de la réalité), central pour le domaine. Dans ce cas-là, nous considérons que nous avons affaire à plusieurs «points de vue» sur un objet. On peut souvent relier un point de vue à un métier, une compétence particulière, c'est-à-dire que l'on trouve un point de vue dominant (une classe de contextes dominante) dans les textes issus d'un seul et même métier. Ainsi, une étude réalisée au CNES et rapportée dans Rébeyrolle (1995) a mis en évidence sept acceptions pour le terme *satellite*, chaque acception étant caractéristique d'un métier concernant le produit

satellite. Si nous ne considérons que deux de ces acceptions, le satellite vu comme un «corps artificiel» et le satellite vu comme un «véhicule», on peut montrer que le premier point de vue est repérable grâce aux contextes correspondant au patron :

**V de type *projeter dans l'espace* + dét + N1** <sup>(1)</sup>

[1] *lancer des satellites en orbite*

**V support de type *faire* + dét + (déverbal) DVB de type *projeter dans l'espace* + de + dét + N1**

[2] *effectuer le lancement du satellite*

[3] *effectuer le tir de SPOT4*

et la seconde acception grâce aux contextes correspondant au patron :

**V de type *placer* + dét + N2 + sur + dét + N1**

[4] *placer un système sur le satellite*

[5] *placer un senseur stellaire sur le satellite*

[6] *poser des fixations sur le satellite*

**V support de type *faire* + dét + DVB *placer* + de + N2 + sur + dét + N1**

[7] *réaliser l'embarquement du prototype sur le satellite SPOT4*

[8] *faire l'intégration des instruments sur les satellites porteurs*

[9] *faire l'intégration des essais électriques, essais mécaniques sur le satellite*

[10] *réaliser l'embarquement du navigateur autonome sur le satellite*

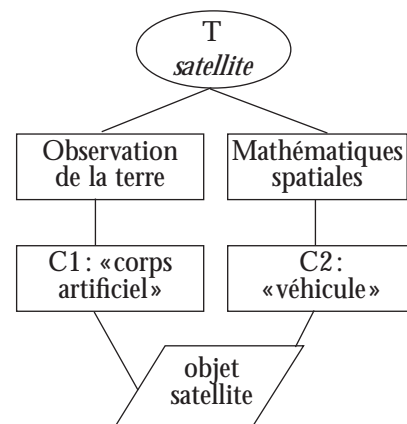
Le premier type d'acception apparaît surtout dans le département Observation de la terre et le second dans le département Mathématiques spatiales. Cette différence d'acception en fonction du métier (qui manifeste des points de vue différents) est à situer dans la question de l'usage dont nous traitons ci-dessous.

(1) Remarque : N1 = *satellite* ou *satellite* + adjectif.

### 3.3.3 Analyse de l'usage en fonction de la provenance du document : influence sur le choix de l'une ou l'autre forme linguistique ou influence sur le lien terme/concept

Deux cas se présentent. Dans le premier, nous avons opté pour consigner l'information dans la partie Terme; il s'agit des cas où une forme, sigle ou variante, est préférée dans l'un ou l'autre métier, il peut être alors indiqué, dans le champ Terme : «forme préférée dans tel métier».

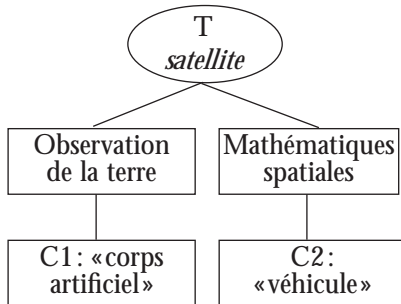
Dans le deuxième cas, l'information est indiquée dans le champ terme/concept. Il s'agit des cas où le métier a une influence sur le choix d'un terme pour dénommer un concept (cas des termes équivalents : plusieurs termes pour un seul concept) ou bien sur le point de vue (cas où un terme dénomme plusieurs concepts apparentés : un terme/plusieurs concepts). Par exemple, les deux points de vue évoqués dans l'étude du CNES pourraient être représentés de l'une des façons suivantes :



Première façon de représenter deux points de vue pour *satellite*

Dans ce cas, il y a création d'un nouveau champ qui permet d'enregistrer l'existence d'un référent unique. Notons que «corps artificiel» et «véhicule» ne sont ici que des

étiquettes de concepts. Le premier concept devrait être en relation avec le terme *corps artificiel* et le second avec le terme *véhicule*. Le champ lien terme/concept contient ici l'information concernant le métier dans lequel est privilégié un usage de satellite.



Deuxième façon de représenter deux points de vue pour *satellite*

Dans ce cas, l'existence d'une parenté entre les concepts, ici étiquetés «corps artificiel» et «véhicule», est indiquée par la mise en relation de ces deux concepts.

### 3.4 Tâches centrées sur l'analyse des relations entre concepts

#### 3.4.1 Déterminations de paradigmes taxonomiques : plusieurs termes/plusieurs concepts taxonomiquement reliés

Cette étape a pour objet d'aider l'étape suivante d'identification des relations conceptuelles.

En effet, des termes représentant des concepts taxonomiquement apparentés à X ou Y peuvent apparaître en position sémantique identique à celle de X ou Y dans les phrases, c'est-à-dire peuvent permettre de repérer une relation conceptuelle «basique» entre X et Y. En effet, si on a :

X *est-un* Z et Z *R* U, on peut avoir X *R* U.

Or, dans les cas où il y a peu d'occurrences de X, il peut être intéressant de tester la relation sur Z et de faire l'hypothèse qu'elle est valable aussi pour X, puisqu'en principe un élément hérite des traits de fonctionnement des éléments supérieurs dans la taxonomie. Par exemple, si Z *a-pour-symptôme* U et si X *est-un* Z, on peut penser que X *a-pour-symptôme* U; Si une infection *a-pour-symptôme* de la fièvre et si la grippe *est-une* infection alors on peut penser que la grippe *a-pour-symptôme* de la fièvre. Dans les corpus, au moment de travailler sur les relations qu'entretient l'hyperonyme (*infection* dans l'exemple), on pourra aussi travailler à partir des hyponymes (*grippe* dans l'exemple), en particulier dans les cas où la forme étudiée apparaît peu fréquemment.

Cette étape constitue un premier type d'analyse de relations. Il faut chercher les contextes dans lesquels X et Y sont cooccurrents et apparaissent avec des marqueurs de relations taxonomiques, par exemple un contexte comme :

«de tous les X, Y est le seul qui...» qui permet de repérer que Y *est-un* X..

#### 3.4.2 Analyse de la (des) relation(s) conceptuelle(s) entre X et Y

L'hypothèse de départ la plus simple est qu'il y a une seule relation sémantique entre X et Y qui s'actualise par des manifestations lexico-syntaxiques différentes. S'il y a plusieurs relations, il faut évaluer quels sont les différents points de vue mis en jeu.

Pour effectuer l'analyse en corpus de la façon la plus efficace, on peut mettre en œuvre une méthodologie en plusieurs étapes:

##### • Étape A

Pour chaque terme (pour nous, X et Y), constitution de classes :

– Classe 1 des variantes, sigles ou variantes de formes;

– Classe 2 des équivalents (dénominations alternatives);

– Classe 3 des hyponymes.

##### • Étape B

Recherche des cooccurrences de X et Y apparaissant dans le même contexte<sup>(2)</sup>, en faisant intervenir, selon les besoins (cooccurrences insuffisantes ou au contraire trop nombreuses) les classes définies en A.

##### • Étape C

Interprétation des résultats;

Soit il n'y a aucun contexte d'apparition simultanée de X (ou ses différentes classes) et de Y (ou ses différentes classes). Dans ce cas, ou bien le corpus n'est pas représentatif du domaine, ou bien, il n'existe pas de relation entre les concepts dénommés par X et par Y.

Soit il y a un contexte d'apparition simultanée de X (ou ses classes) et de Y (ou ses classes). Dans ce cas, on peut retrouver plusieurs cas de figure.

– Premier cas de figure.

Le contexte peut être insignifiant, c'est-à-dire qu'on ne peut pas lui attribuer une pertinence par rapport à une relation conceptuelle. Cette insignifiance peut être due au fait que la relation qui existe est trop implicite pour être mise au jour dans le contexte à l'étude; elle peut être due au fait que la relation conceptuelle qui existe n'est pas pertinente dans l'assertion linguistique qui sous-tend le contexte; elle peut être due enfin au fait que l'interprétant (le linguiste-terminologue) manque de compétences linguistiques (il ne connaît pas certains mots ou tournures apparaissant dans le

(2) La notion de contexte mériterait ici un plus long développement. En effet, l'expérience nous a montré que se contenter d'une phrase comme contexte était insuffisant. Nous avons décidé de limiter arbitrairement le contexte de façon numérique et avons opté pour un contexte de 35 chaînes de caractères.

contexte) ou techniques (il ne connaît pas la nature de la relation conceptuelle qui apparaît).

– Deuxième cas de figure.

Le contexte est signifiant d'une seule façon : on peut trouver, à travers les divers contextes une régularité sémantique qui manifeste l'existence d'une relation conceptuelle unique.

– Troisième cas de figure.

Les divers contextes dans lesquels apparaissent X et Y permettent d'identifier plusieurs classes de fonctionnement sémantique, donc plusieurs relations conceptuelles; ce cas semble associé aux cas de polyacceptions en lien avec des points de vue en fonction de types de raisonnement différents (par exemple, les relations cause et conséquence peuvent être inversées).

L'approche que nous préconisons pour repérer les relations conceptuelles constitue une abstraction par rapport à ce qui est réellement fait. En effet, on ne teste pas tous les couples de termes, ce qui serait extrêmement long et fastidieux mais seulement les couples dont nous pensons, pour diverses raisons (compétence linguistique, intervention d'un expert), qu'ils peuvent entretenir des relations. Par ailleurs, au fur et à mesure que le travail progresse, les résultats sur les relations conceptuelles se cumulent et rendent la recherche plus efficace.

Il est clair que les différentes étapes que nous proposons font appel à un examen minutieux des corpus qui demande des connaissances sur le fonctionnement de la langue et une grande capacité à généraliser. Par ailleurs, les résultats que nous obtenons constituent des hypothèses linguistiques qui doivent être soumises à un, voire plusieurs, expert(s). Le fait que chaque résultat puisse être argumenté par des portions de textes rend la discussion à la fois plus distanciée (puisque'elle se base sur un texte écrit par un tiers) et plus fructueuse.

Ce type d'approche, qu'il paraissait inconcevable de mettre en œuvre il y a quelques années deviendra sans doute incontournable. En effet, d'une part, les besoins en données terminologiques qui rendent compte au plus près du fonctionnement linguistique à l'intérieur des entreprises sont de plus en plus pressants; d'autre part, et heureusement, de plus en plus d'outils viennent faciliter ce travail d'analyse de corpus. Dans le prochain paragraphe, nous essayons de faire un panorama rapide des types d'outils qui pourraient être utilisés pour ces tâches.

## 4 Quels outils d'aide pour ces tâches

La rapide présentation que nous proposons ici a essentiellement pour objectif de montrer combien les besoins émergents en ce qui concerne l'analyse de corpus, pour en extraire des données terminologiques, vont de pair avec la définition de nouveaux outils d'analyse de textes, pas toujours dédiés à la terminologie mais qui pourraient fort bien être utilisés dans ce cadre. Nous n'avons utilisé que certains des outils que nous évoquons; pour les autres, nous avons assisté à une démonstration ou bien nous n'en avons connaissance que par les articles qui les décrivent. Nous avons voulu nous placer du côté de l'utilisateur potentiel qui, ayant défini ses besoins, souhaiterait voir son travail assisté par des outils.

### 4.1 Outils de recherche de concordances

Les concordanciers permettent d'accéder à toutes les occurrences de formes et à leur contexte d'apparition. Cet accès peut être affiné si, comme dans *Sato* (Daoust 1990), une base de données lexicale permet d'attribuer à

chaque forme sa ou ses catégories grammaticales. Il reste que s'il n'y a pas d'analyse syntaxique, une forme est traitée comme ayant toutes les valeurs qui lui sont attribuées dans la base lexicale. Ainsi, *porte* est à la fois verbe et nom et conserve cette étiquette quel que soit son contexte d'apparition. Il est évident que le couplage avec un analyseur syntaxique rendrait ces outils beaucoup plus efficaces. *Sato* permet aussi d'attribuer des propriétés aux chaînes de caractères. Il est possible d'utiliser cette fonctionnalité pour constituer les classes de variantes ou d'hyponymes telles que nous les avons présentées en 3.3. Ce type d'outils est l'outil de base pour l'exploration des corpus.

### 4.2 Outils d'extraction de termes et de relations candidats

De nombreux laboratoires de recherche français (nous connaissons moins bien la situation ailleurs) travaillent à la définition d'outils d'extraction de termes candidats ou de relations candidates.

Les outils d'extraction de termes fonctionnent sur deux principes. Soit ils fonctionnent sur un principe statistique (Enguehard et Pantéra 1995), soit ils fonctionnent à partir d'une analyse linguistique poussée (cf. *Lexter* : Bourigault 1994 et *Termino* : David & Plante 1990), soit ils fonctionnent sur un principe mixte couplant une approche statistique et linguistique (Daille 1995).

Dans tous les cas, ces outils fournissent une liste de termes candidats, à charge pour le linguiste/terminologue d'effectuer un tri.

Les outils d'extraction de relations candidates ne sont pas conçus, dans un premier temps pour assister le travail du terminologue, mais celui du cogniticien, c'est-à-dire de la personne chargée de recueillir les connaissances (en l'occurrence à partir

de corpus) qui vont lui permettre de construire un système à base de connaissances. Or, ainsi que nous l'avons montré ailleurs (Condamines 1995), les relations mises en place dans un système à base de connaissances sont proches des relations conceptuelles qu'on peut trouver dans une terminologie. Il est donc possible d'utiliser ces outils dans le cadre de la terminologie. Deux types d'outils peuvent exister en fonction du principe de base mis en œuvre. Dans le premier, on utilise des connaissances sur la langue afin de repérer des relations conceptuelles dans les langues de spécialité; par exemple pour *Seek* (Jouis 1995) une liste des marqueurs linguistiques de la relation *est-un et partie-de* (*est constitué de, a pour partie...*) a été faite. Lorsque le système rencontre ces formes dans le corpus, il suggère à l'utilisateur l'existence d'une relation conceptuelle. Dans le deuxième type d'outils (Frath *et al.* 1995), le système, après avoir lemmatisé le corpus, repère des formes récurrentes apparaissant entre deux formes dont il présume qu'elles sont des termes. L'idée est ici qu'il peut y avoir des marqueurs de relations voire des relations qui sont dépendants du domaine. Lorsque l'utilisateur est mis en présence de ces contextes récurrents, il peut travailler à leur généralisation pour définir la nature de la relation mise en œuvre.

### 4.3 Outils d'analyse statistique

Nous ne parlerons ici que de *Alceste* (Reinert 1991), que nous connaissons mais il est vraisemblable que tous les outils à base de calculs statistiques un peu poussés permettraient le même type d'utilisation.

Ces outils visent à donner du texte une classification en termes de contenu sur la base de calculs d'occurrences lexicales. Un texte se

trouve ainsi organisé en plusieurs classes auxquelles l'utilisateur donne une interprétation en fonction de sa problématique. Par exemple, on devrait pouvoir identifier des «points de vue» émergents à partir des résultats fournis par un tel outil.

L'expérimentation que nous avons faite de *Alceste* (Rébeyrolle 1995), nous a permis de mettre au point une méthodologie qui rend compte de classes de fonctionnement de contextes autour d'un terme. Or, nous l'avons vu, la mise en évidence de plusieurs classes de contextes nous permet de présumer un phénomène de type homonymique ou polysémique. L'utilisation d'un tel outil a donc présenté un grand avantage pour nous (cf. Condamines & Rébeyrolle, à paraître) puisqu'il nous a permis de travailler à la mise en évidence de points de vue à travers l'étude de contextes à partir des résultats fournis par *Alceste*.

### 4.4 Quelles perspectives pour l'utilisation de ces outils

Si le travail sur corpus devient un des éléments importants de la terminologie – mais aussi d'autres types de travaux comme l'extraction de connaissances à partir de textes – on devrait assister à un développement important des outils. Mieux encore, ces outils devraient être intégrés pour constituer de véritables stations d'analyse de textes. C'est déjà ce qui se met en place dans certaines grandes entreprises comme à IBM ou à EDF (cf. Herviou *et al.* 1995).

## 5 Conclusion

La terminologie est actuellement le champ de changements importants tant théoriques qu'appliqués. Les besoins, en particulier dans les

entreprises, ont évolué; les modèles de données se sont adaptés; les méthodologies, grâce au développement d'outils informatiques, convergent vers une prise en compte plus grande des corpus qui, seuls, manifestent l'usage réel des termes par une communauté de locuteurs. L'analyse de corpus est actuellement faite de façon souvent intuitive. Il convient de définir des méthodes adaptées aux besoins spécifiques de la terminologie, aussi bien en ce qui concerne la façon d'utiliser les outils et, plus particulièrement d'utiliser les résultats fournis par un outil en entrée d'un autre outil, que sur l'interprétation que l'on peut faire des résultats obtenus. Afin de rendre ces méthodes les plus efficaces et reproductibles possibles, nous avons proposé d'organiser le recueil de données terminologiques en tâches que nous avons définies par rapport à un modèle de BCT. Ce travail devrait se poursuivre pour permettre un affinement de la méthodologie et de la prise en compte d'outils.

*Anne Condamines,  
Équipe de recherches en syntaxe et  
sémantique, Ura 1033, CNRS,  
Université de Toulouse le Mirail,  
France.*

## Bibliographie

Auger (Pierre), 1994: «Les outils de la terminotique», dans *Terminologies nouvelles*, n° 11, p. 46-52.

Bourigault (Didier), 1995: «Conception et exploitation d'un logiciel d'extraction de termes en entreprise. Problèmes méthodologiques et théoriques», dans Clas (André) *Actes des 4<sup>es</sup> Journées scientifiques du réseau Lexicologie, terminologie, traduction*, Lyon 28-30 septembre 1995, à paraître.

Bourigault (Didier) et Condamines (Anne), 1995: «Réflexions sur le concept de base de connaissances

- terminologiques», dans Actes des 5 Journées du PRC IA, Nancy 1-3 février 1995, Toulouse, Teknea, p.425-444.
- Condamines (Anne) et Amsili (Pascal), 1993 : «Terminology between Language and Knowledge : An example of Terminological Knowledge Base», dans Schmitz (Klaus-Dirk), TKE93 (*Terminology and Knowledge Engineering*), Frankfurt, Indeks Verlag, p. 316-323.
- Condamines (Anne), 1995: «Terminology: new needs, new perspectives», dans *Terminology*, vol. 2, n°2, p.219-238.
- David (Sophie), Plante (Pierre), 1990: *Termino version 1.0*, Rapport du Centre Ato, *Analyse de textes par ordinateur*, Université du Québec à Montréal.
- Daille (Béatrice), 1995: «ACTA: Une maquette d'aide à la construction automatique de banques terminologiques monolingues et bilingues», dans Clas (André), *Actes des IV<sup>es</sup> Journées Scientifiques du réseau LTT : lexicomatique et dictionnaires*, Lyon, 28-30 septembre 1995, à paraître.
- Daoust (François), 1990 : «L'informaticien, le lecteur et le texte. L'approche Sato», dans *ICO Québec* (Intelligence artificielle et sciences cognitives au Québec), vol. 2, n°3, p. 55-60.
- Descamps (J.L), Mochet (M.A), Lewin (T), Lamizet (B) et Costes (D.), 1992: *Sémantique et concordances*, Paris, Klincksieck, (Publication de l'Inalf, Collection «St-Cloud»).
- Enguehard (Chantal) et Pantéra (Laurent), 1995: «Automatic Natural Acquisition of a Terminology», dans *Journal of quantitative linguistics*, vol. 2, n°1, p. 27-32.
- Frath (Pierre), Oueslati (Rochdi) et Rousselot (François), 1995: «Identification de relations sémantiques par repérage et analyse de cooccurrences de signes linguistiques». dans *Actes des Journées Acquisition, validation, apprentissage du PRC-GDR-IA du CNRS*, Grenoble: 5-7 avril, Le Chesnay, France, Inria, p.173-186.
- Herviou (Marie-Luce), Ogonowski (Antoine) et Dauphin (Eva), 1994: «Tools for extracting and structuring Knowledge from Texts», dans *Proceedings of Computational Linguistics, Coling*, Japan : Kyoto, vol. 2, p. 1049-1053.
- Jacquemin (Christian), 1994 : «Quelques mécanismes spécifiques d'une grammaire d'unification adaptée à l'extraction terminologique», dans *Actes du 9<sup>e</sup> congrès Reconnaissance des formes et intelligence artificielle* (RFIA'94), Paris, AFCET, p. 385-396.
- Jouis (Christophe), 1995 : «Seek, un logiciel d'acquisition des connaissances utilisant un savoir linguistique sans employer de connaissances sur le monde externe», dans *Actes des Journées Acquisition, validation, apprentissage du PRC-GDR-IA du CNRS*, Grenoble : 5-7 avril, Le Chesnay, France, Inria, p. 159-172.
- Kocourek (Rostislav), 1991 : «Textes et termes», dans *Meta*, vol. 36, n°1, p. 71-76.
- Meyer (Ingrid), Bowker (Lyne) et Eck (Karen), 1992 : «Cogniterm: An experiment in Building a Terminological Knowledge Base»; dans Tomola (H.), Varantola (K.) et Salimi-Tolonen (T.), *Euralex'92 Proceedings I-II*, Tampere, Finland : University of Tampere, p. 159-172.
- Rastier (François), 1991 : *Sémantique et recherche cognitive*, Paris, Puf.
- Rebeyrolle (Josette), 1995 : *Apports de la terminologie à l'étude des points de vue*, Mémoire de fin de stage de DEA au CNES, Université Toulouse le Mirail, Équipe de recherches en syntaxe et sémantique.
- Reinert (Max), 1991 : «Les mondes lexicaux et leur "logique" à travers l'analyse statistique d'un corpus de récits de cauchemars», dans *Langage et société*, n°66, p. 5-39.

# Pour une analyse linguistique des termes dans la phrase

Les vocabulaires scientifiques ou techniques sont consignés dans des nomenclatures multilingues, des dictionnaires spécialisés et des terminologies. L'objet de ces recueils n'est pas de décrire les langues de spécialité dans leur usage mais d'apporter des outils terminologiques fiables aux étudiants et aux experts de chaque secteur d'activité.

Un projet ambitieux, celui d'un observatoire qui permettrait de dégager non plus seulement des mots mais un ensemble de collocations grammaticales susceptibles de rendre compte de la grammaire et de la sémantique de ces unités dans la phrase, renouvelle la tradition terminologique jusqu'alors enclavée dans l'étude de sous-systèmes linguistiques.

Termes-clés :  
collocations grammaticales ;  
sémantique des termes dans la phrase.

Ce bref article qui présente une étude linguistique des termes dans la phrase offre un prolongement au numéro thématique de *Terminologies nouvelles* n° 10 consacré à la phraséologie. Il s'agit, en fait, d'une approche particulière de la phraséologie telle que la conçoit la norme Iso 1087-[Iso 89] dans la définition «Étude scientifique des notions et des termes en usage dans les langues de spécialité». Étude *scientifique*, certes, puisqu'il s'agit d'une approche linguistique, *notionnelle* dans la mesure d'une sous-catégorisation en traits syntactico-sémantiques que nous développerons plus bas, et *en usage* du fait du corpus de référence.

Cette approche linguistique des langues spécialisées aborde les unités par leur fonctionnement au sein des phrases, car la sémantique de ces mots, bien que relevant d'univers particuliers de la connaissance, dépend aussi de l'interprétation des relations grammaticales. Ce qui change de la langue générale dans les langues spécialisées, c'est la phraséologie. Il s'agit donc de proposer une étude qui, partant d'une syntaxe de position, permettra de rendre compte du style statistiquement dominant de tel ou tel corpus. Les descriptions offertes par les opérateurs linguistiques permettent de sélectionner des arguments par domaines d'activité.

Un opérateur peut avoir trois formes : verbe, nom ou adjectif :

- Un verbe dans *Luc est vacciné*;
- Un nom dans *La vaccination de Luc*;

- Un adjectif dans *Luc est vaccinable*.

Un verbe donné est interprété en fonction de son sujet et de son complément, ce qui permet de désambiguïser les phrases :

- *Abaïsser* dans *L'étudiant abaisse le store*; est différent de *abaïsser* en géométrie, dans *L'étudiant abaisse une perpendiculaire*;

- *Coder* est transitif dans *Jean code un message*;

- *Coder* est intransitif en biologie, dans *Un gène code pour une protéine*.

L'opérateur est donc défini par l'ensemble de ses arguments.

Les noms prédicatifs sont actualisés par des verbes supports qui ont la même fonction que les désinences verbales.

Dans l'exemple *Le chirurgien procède à l'amputation*, *amputation* est actualisé par *procède*.

Dans *Le navire exécute ses mouvements*, *mouvements* est actualisé par *exécute*.

Un adjectif comme *alcoolique* présente des significations différentes suivant qu'il s'agit :

- *D'une solution alcoolique* «qui par nature contient de l'alcool»;

- *D'une fermentation alcoolique* «particulier à l'alcool éthylique»;

- *D'un délire alcoolique* «qui résulte de l'alcoolisme»;

- *D'un homme alcoolique* «qui s'adonne à l'alcoolisme».

La procédure de traitement (base de données) s'effectue suivant des paramètres ponctuels, préalablement définis suivant chaque sphère d'activité. La réalisation peut se résumer globalement ainsi :

1. Le recensement des vedettes;
2. Le domaine d'emploi;
3. La description des emplois.

## 1 Le recensement des vedettes

Les ouvrages spécialisés par domaines permettent de recueillir des candidats termes. La présence des unités terminologiques dans les dictionnaires spécialisés confirme l'appartenance des candidats termes à un secteur d'activité donné. Parfois, le recours aux experts s'impose pour des domaines récents qui ne connaissent par de dictionnaires (c'est le cas de la physionique) et pour ceux dont les nombreuses innovations ne permettent pas de disposer de recueils lexicographiques à jour (le secteur de l'informatique par exemple).

Par exemple en partant d'un corpus technique des moyens de transport ferroviaire propres à la Régie autonome des transports parisiens (la RATP), les vedettes et leurs collocations pour les termes *matériel roulant*, *MF 77*, *matériel ferroviaire 1977*, *MS 61*, *matériel suburbain 1961* sont les suivantes :

[1] [...] *le MS-61 (Matériel Suburbain 1961) est mis en circulation en 1967. [...] il circule exclusivement sur la ligne A [...].*

[2] *Traitement des informations à bord du train pour optimiser l'adaptabilité aux différents matériels roulants circulant sur la même ligne.*

[3] *Une nouvelle génération de matériel roulant [...] commençait à circuler sur les lignes régionales [...].*

[4] *Le MF 77, constitué en trains de cinq voitures d'une capacité [...] de 574 voyageurs, circule sur les lignes 7, 8 et 13.*

On a sélectionné pour ces phrases le prédicat verbal *circuler* à partir de leur structure :  
circuler/N0:mt-f/sur N1:loc<voie>  
où N0 = sujet (matériel de transport ferroviaire) et N1 = locatif (voie de communication).

## 2 Le domaine d'emploi

La sous-catégorisation en traits syntactico-sémantiques est applicable à l'ensemble des opérateurs décrits au sein d'une classe d'objets. La notion de trait permet de coder sujets et objets spécifiques pour tel opérateur aussi bien en langue générale qu'en langue spécialisée.

[1] *Un pipe-line immergé traversera la Manche pour alimenter les véhicules en carburant.*

[2] *Les trains sont alimentés en courant continu 750 volts [...]*

[3] *Les véhicules étaient désormais alimentés à l'alcool, au gaz de ville ou au gaz de charbon de bois, les fameux «gazogènes».*

1.2.3 = être alimenté/N0:mt-(f+n+tm)/en N1:inc<énergie>  
où in = «inanimé», c = concret. Les traits syntaxiques sont des outils qui permettent d'apporter des informations complémentaires pour préciser la structure de certains types de phrases et seulement celles-là.

La notion de classe d'objets, développée par le Laboratoire de linguistique informatique (LLI, à l'Université Paris-Nord), permet de prendre en charge le traitement de collocations dans leur milieu contextuel. Elle est chargée d'indiquer le domaine d'emploi (ici mt : moyen de transport) précisé le cas échéant par une rubrique (dans cet exemple : f = transport ferroviaire, n = nautique, tm = terrestre à moteur).

## 3 La description des emplois

Dans cette étude les phrases sont traitées comme des jeux de prédicats et d'arguments. L'étude syntaxique et sémantique consiste à rendre compte des emplois des unités recensées. Ce type de caractérisation présente des articulations entre syntaxe et

sémantique où le caractère grammatical de l'étude dépend des contraintes de sélection entre un opérateur et son argument nominal.

Le domaine d'emploi du verbe dépend des traits sémantiques révélés par ses arguments.

[1] *L'avion piqua du nez.*  
(langage technique);

[2] *Paul piqua du nez.* (langue courante).

Dans la langue générale, les verbes monosémiques sont rares. Dans les langues spécialisées, la sphère d'activité sélectionne le plus souvent un seul emploi, pour un ensemble donné d'arguments, par verbe approprié. Au sein du domaine général des transports on distingue ainsi deux emplois de *descendre* suivant la sphère d'activité (mt = moyen de transport, n = nautique, a = aérien) malgré une structure quasi identique où seulement l'un des deux verbes est transitif. D'autre part la classe d'objets dissocie également les deux emplois de *descendre*.

1. descendre/N0:mt-a/sur N1:loc.<voie>

2. descendre/N0:mt-n/N1:loc.<voie>

Pour les contextes :

[1] *L'aéronef [...] descendit sur la piste [...].*

[2] *Ils virent ainsi plusieurs bateaux qui montaient et descendaient péniblement le fleuve.*

On peut sélectionner les verbes régissant des noms pertinents pour un domaine d'étude. Si nous choisissons celui des moyens de transport aérien, l'argument *avion* invite certains opérateurs. Pour chaque occurrence de terme, on retrouve soit le verbe, soit le prédicat approprié. Ce sont ces opérateurs appropriés qui définissent les classes d'objets. Il s'agit d'extraire des contextes les prédicats qui fonctionnent et de les coder. Considérons ici une extraction de quelques verbes des moyens de transport :

verbes prédicatifs	corpus	références
atterrir/N0:mt-a/sur N1:loc<voie>	Cet avion [...] est censé décoller et atterrir sur des pistes très courtes.	<i>La Recherche</i> , sept. 94 : 875
circuler/ N0:mt-mtc/sur N2:loc<voie>	C'est fin 1906 [...] que l'omnibus automobile -l'autobus- circula [...] sur la ligne Montmartre/Saint-Germain-des-Prés.	RATP, «Bus - Le matériel roulant» janvier 1990, np
circuler/N0:mt-f/sur N1:loc <voie>	Les premiers trains MS 61 du RER circulaient sur la ligne de Sceaux [...]	RATP, «La RATP a 40 ans», décembre 1988 : 3
circuler/N0:mt-f/sur N1:loc<voie>	Le MF 77, constitué en trains de cinq voitures d'une capacité [...] de 574 voyageurs, circule sur les lignes 7, 8 et 13.	RATP, «La RATP a 40 ans», décembre 1988 : 5
circuler/N0:mt-tmc/sur N1:loc	Les véhicules ont circulé sur l'itinéraire du 94 [...]	RATP, <i>Les transports de surface des origines</i> : 19
circuler/N0:mt-tmc/sur N1:loc<voie>	C'est ainsi que des autobus bleus circulèrent sur la ligne 82 [...]	RATP, <i>Les transports de surface des origines</i> : 18
conduire/N0:hum/N1:mt-tmc	- les machinistes qui conduisent les bus [...]	RATP, «Les années 1990» : 22
être alimenté/N0:mt-f/en N1:inc<énergie>	Les trains sont alimentés en courant continu 750 volts [...]	RATP, «Une journée sur la 1», octobre 1990, np
être alimenté/N0:mt-f/en N1:inc<énergie>/par N2:inc	Le MS 61. Il se présente sous forme d'un élément constitué de 3 voitures : deux motrices, alimentées en courant de traction 1500 volts continu par caténaire [...]	RATP, «RER, le matériel roulant», septembre 1989, np
être alimenté/N0:mt-mt/à N1:inc<énergie>	Les véhicules étaient désormais alimentés à l'alcool, au gaz de ville ou au gaz de charbon de bois, les fameux «gazogènes».	RATP, «De l'omnibus à l'autobus», novembre 1991 : 8
être mis en service/N0:mt-mtc/sur N1:loc<voie>	Juillet 1981 : mise en service sur la ligne 57, d'une voiture prototype [...]	RATP, «Les transports de surface des origines à nos jours», 1985 : 24
être mis en service/N0:mt-tmc/sur N1:loc <voie>/Prép N3:temps	En 1981, fut mis en service sur la ligne 304 un autobus standard [...]	RATP, <i>Les transports de surface des origines</i> : 19
être mis en service/N0:mt-tmc/sur N1:loc<voie>	Novembre 1982 : mise en service sur la ligne 91 du premier autobus articulé dit «superbus», d'une capacité de 127 places.	RATP, «Les transports de surface des origines à nos jours», 1985 : 24
être piloté/N0:mt-tm/par N1:inc<automate>	Étape suivante : concevoir des véhicules susceptibles d'être entièrement pilotés par ordinateur.	<i>La Recherche</i> , nov. 94 : 1104
faire le plein/de N0:mt-tm/de N1:inc<énergie>	Le plein de gasoil du véhicule est fait à la rentrée, l'aspirateur passé.	RATP, «Une journée sur le 38», mai 1989, np

fonctionner/N0:mt-tm/à N1:inc<énergie>	Catalyseur pour pots d'échappement destiné pour véhicules fonctionnant au gaz naturel.	<i>La Recherche</i> , avr. 95 : 373
rouler à/N0:mt-tm/N1:inc<énergie>	un autobus roule à l'hydrogène.	<i>La Recherche</i> , oct. 94 : 984
rouler/N0:mt-tm/à N1:inc<énergie>	[...] plus d'un million de véhicules roulent au gaz naturel	<i>La Recherche</i> , avr. 95 : 372
rouler/N0:mt-tm/entre N1:loc/et N2:loc	[...] un engin à traction autonome faisait son apparition, qu'on appela automobile. En 1905, le Salon de l'Automobile présenta les engins de ce type, qui roulèrent entre la Bourse et le Grand Palais [...]	RATP, «De l'omnibus à l'autobus», novembre 1991 : 4
La suite la plus longue peut être également définie par l'ensemble des arguments des prédicats nominaux. Ces derniers admettent parfois un verbe support (codé W:) et peuvent parfois être à l'origine d'un verbe. manœuvre/N0:hum/N1:mt-n/W:faire <i>faire une manœuvre = manœuvrer</i> (un navire)	stationnement/N0:mt-tm/W:être en être en stationnement = stationner vol/N0:mt-a/W:faire L'avion fait un vol = l'avion vole [...] <i>Le louvoyage rapide du sardinier.</i> Louvoyage est un opérateur nominal à un argument : <i>sardinier</i> [...] <i>les conducteurs chargés de la conduite des métros ou de la</i>	<i>surveillance de leur bonne marche en pilotage automatique.</i> Conduite est un opérateur nominal à deux arguments : <i>conducteur</i> et <i>métro</i> et offre la structure : conduite/N0:hum/N1:mt-f-
<b>termes</b>	<b>noms prédicatifs</b>	<b>corpus</b>
autobus	circulation/N0:mt-tmc/sur N1:loc<voie>	Pour tenter de résoudre le problème, toujours actuel et pressant, de la circulation des autobus [...]
barque de passage	liaison/N0:mt-n/entre N1:loc/et N2:loc/W:effectuer	Les liaisons entre l'île et les rives du fleuve nécessiteront l'établissement de barques de passage qui constitueront ainsi le premier moyen de transport public par eau sur la Seine.
bateau	manœuvre/N0:hum/N1: mt-n/W:faire	Le redressement brutal du bateau [...] présagea de sa grande stabilité. Les essais en mer permirent de le vérifier après que la manœuvre ait dû être réinventée en s'appuyant sur les souvenirs de quelques marins.
bus	conducteur/N0:hum/N1:mt-tm	Des conducteurs de bus ou de métro aux ingénieurs [...] la liste des métiers exercés par le personnel de la RATP est longue [...]
matériel fer	circulation/N0:mt-f/sur N1:loc <voie>	Première circulation en 1967 sur la ligne 3 du nouveau matériel fer MF 67 [...]

méto	conduite/N0:hum/N1:mt-f	[...] en ligne, les conducteurs chargés de la conduite des métros ou de la surveillance de leur bonne marche en pilotage automatique.
MF 77	consommation/N0:mt-f/N1:inc<énergie>	Mais surtout le MF 77 utilisa pour la première fois un «hacheur de courant» à thyristors (késar) permettant une plus grande souplesse de conduite tout en réduisant la consommation de courant et en autorisant le freinage par récupération [...]
navette	vol/N0:mt-e/W:effectuer	Vingt-huit de ces vols seront effectués par navette.
rame	évacuation/de N0:mt-f/par N1:hum	Les mesures immédiates : elles peuvent consister en une évacuation de rame par ses voyageurs, la rentrée en station d'un train immobilisé sous tunnel [...]
rame	garage/N1:mt-f	À partir de 18h30, ont lieu les premiers garages de rames [...]
sardinier	louvoyage/N0:mt-n	Le louvoyage rapide du sardinier a donc étonné à demi.
train	circulation/card N0:mt-f/Prép N1:temps	À partir de 1983, la véritable interconnexion devenait effective avec la circulation de 8 trains à l'heure de bout en bout sur la ligne B [...]
train	roulement/N0:mt-f/sur N1:loc<voie>	[...] le roulement des trains s'effectue sur des pistes [...]
voilier	embarquement/N0:hum/sur N1:mt-n	Ils sont aujourd'hui intégrés au Port-Musée et participent aux rassemblements de voiliers traditionnels. Le plus ancien a transporté des milliers de personnes [...] et de nombreuses classes de mer et du patrimoine y trouvent un embarquement [...]
<p>Ce projet s'inscrit dans les recherches menées par le Laboratoire de linguistique informatique de Paris XIII qui réalise depuis plusieurs années un dictionnaire électronique de la langue française. Pour ma part, partant du constat que les langues de spécialité utilisent de façon particulière la langue générale comme support nécessaire à la communication, une approche linguistique des langues spécialisées</p>		<p>abordant les unités dans les phrases pour rendre compte des régularités combinatoires constitue un nouvel axe de recherche. Il se situe au confluent de la terminologie et des méthodes pratiquées en linguistique d'analyse de la phrase par la description de la syntaxe et de la sémantique des termes en contexte. Cette approche permet de dégager non plus seulement des mots, mais un ensemble de collocations</p> <p>grammaticales dans les phrases par une étude linguistique des langues de spécialité où la sémantique des termes n'est pas indépendante de leur fonctionnement syntaxique.</p> <p><i>Jacques Boissy, Centre de terminologie et de néologie, Laboratoire de linguistique informatique, Inal-CNRS Ura 1576, Université Paris XIII.</i></p>

# *Balnéo*: un projet de réseau informatique pour la veille néologique

Le contrôle de l'évolution des langues suppose des moyens d'observation rapides et interactifs dans le domaine clef qu'est la néologie scientifique et technique. La francophonie, qui a une longue expérience dans ce domaine, a donc souhaité se doter d'un réseau informatique international utilisant Internet pour la collecte et la diffusion des «attestations de néologismes». Les principes d'organisation du réseau informatique *Balnéo* commandité par le Rint et réalisé par l'Université de Rennes II sont présentés à travers les différents acteurs qu'il fédère, la structure de l'information qu'il collecte (l'attestation) et les outils de saisie et de consultation qu'il utilise.

Termes-clés :  
néologie scientifique et technique;  
attestation de néologismes; Rint;  
réseau informatique; Internet.

**L**es langues évoluent à la fin du XX<sup>e</sup> siècle de façon prodigieuse, sous la poussée notamment des nouvelles techniques et de la révolution informationnelle en cours. Pour évoluer de façon harmonieuse, voire pour continuer d'exister, il leur faut s'informatiser et utiliser à leur profit les formidables outils de communication qui relient la planète.

Cette grande aventure des langues constitue un pari, mais aussi un risque.

Un pari car le véritable avenir des langues réside dans le plurilinguisme. Un risque car les maintenir au niveau des enjeux en cours exige de l'invention, des moyens, et une vision à long terme.

Il n'existait pas encore de moyen véritablement efficace pour observer les évolutions des langues dans le domaine de la néologie. Il est impossible à partir de là de pouvoir évaluer la vitalité réelle de leur lexique, et de répondre aux besoins de création de formes et de définitions. Une politique terminologique reste dans ce cas relativement aveugle, même si elle peut par ailleurs s'appuyer sur des indices ou des intuitions qui ne sont pas forcément à négliger.

Or, la politique, particulièrement le responsable d'une politique linguistique, doit nécessairement avoir une vision de la langue ou des langues à aménager, et définir des objectifs à atteindre. En ce qui concerne une langue comme le français, l'un des objectifs que se sont fixés les francophones est aujourd'hui d'observer l'évolution de ses

vocabulaires, particulièrement dans les domaines scientifiques et techniques. Sans vision d'ensemble de la néologie du français, en effet, il est vain de vouloir prévoir de façon rigoureuse toute évolution.

Disposer d'une base de données d'attestations de néologismes peut permettre en effet de faire des choix pertinents dans le cadre par exemple de la normalisation terminologique (fixation des termes), d'harmonisation terminologique (mise en correspondance de désignations autour de concepts), et plus généralement en matière de traduction de termes.

## 1 Les enjeux

La préoccupation de la francophonie a été depuis plusieurs années de résoudre la question de la néologie scientifique et technique. Pour faire évoluer la langue française, il est en effet nécessaire d'observer et d'analyser les néologismes qui apparaissent dans cette langue, soit par glissement de sens, soit par création de mots nouveaux.

À cette fin, des relevés d'attestations de néologismes, c'est-à-dire des formes repérées dans des contextes textuels (revues, journaux, etc.), ont été effectués par différentes équipes de recherche. L'une des expériences les plus représentatives est celle de la collection de fascicules *Néologie en marche* qui a lié pendant plus de dix années les équipes de l'Office de la langue française du Québec à celles du Haut Comité (puis du Commissariat général) de la langue française en France. Plus

récemment, le Centre de néologie et de terminologie du CNRS a élaboré, avec le soutien de la Délégation générale à la langue française, plusieurs «Cahiers de termes nouveaux», qui regroupent les relevés qu'il a lui-même réalisés. Dans la francophonie, des relevés sont également faits, notamment au Canada et au Québec.

Mais à l'évidence, ces expériences ne sont pas satisfaisantes, car elles demeurent trop ponctuelles, et ne parviennent pas à répondre aux besoins d'observation de la langue française aujourd'hui. En effet, il est indispensable d'avoir une vue large et précise de son évolution, que révèle la néologie, pour pouvoir intervenir à bon escient, cerner les priorités, et travailler de façon convergente avec les francophones.

Il faut pour cela globaliser le problème, travailler en temps réel et en mode interactif.

D'autant qu'il ne s'agit pas simplement de relever des néologismes français, mais aussi de faire un travail sur les néologismes de langue anglaise qui sont susceptibles de s'implanter en français. Il est également possible d'imaginer à terme que cette expérience puisse s'élargir à d'autres communautés linguistiques (hispanophonie, lusophonie, germanophonie), et aux langues partenaires de la francophonie.

## 2 Définition des objectifs

Afin de résoudre cette question du relevé des attestations de néologismes, il a été décidé par le Réseau international de néologie et de terminologie francophone (Rint), d'unir dans un premier temps les efforts de chacun des partenaires qui le composent en exploitant trois axes de travail :

- Localiser, dans chacun des pays francophones, les organismes et

personnes ayant par leur travail, à faire des relevés d'attestations de néologismes;

- Se répartir le travail de dépouillement en ciblant les domaines les plus porteurs de néologismes (techniques de pointe, informatique, communication, etc.);

- Élaborer un outil de travail qui puisse concentrer et diffuser l'ensemble des relevés effectués par les pays francophones.

C'est sur cette décision qu'a été lancé en 1994 le projet de base de données d'attestations *Balnéo*. La révolution technologique que constitue la constitution de réseaux mondiaux de communication de type Internet autorise aujourd'hui à penser l'organisation du travail en commun de façon radicalement différente. Dans ce cadre, il a été décidé de créer un système informatique appelée *Balnéo*, qui permette :

- D'établir un cadre de relevé commun;
- De stabiliser un format d'échanges de données commun;
- De diffuser sur Internet des fiches d'attestations de néologismes;
- D'en permettre la consultation;
- De les gérer dans une base de données;
- D'en effectuer la mise à jour;
- De travailler en temps réel ou en temps différé;
- De permettre un mode d'échanges d'information interactif.

Cet outil, conçu par l'Université de Rennes II, permettra de ce fait de recueillir les relevés d'attestations de néologismes effectués par les partenaires francophones du Rint qui ont accès au réseau Internet et de leur rediffuser l'ensemble des relevés réalisés.

Il sera ainsi possible à la fois d'alimenter et de puiser dans ce fonds d'attestations de néologismes. C'est donc ainsi une véritable banque interactive de données d'attestations de néologismes qui est proposée.

L'un des grands intérêts de cet outil est qu'il permet d'envisager le problème de la néologie du français de façon plus globale en répartissant les tâches de dépouillement entre plusieurs acteurs de la francophonie et au-delà. Son caractère interactif et partagé évite notamment que des domaines restent par inadvertance ignorés ou que deux partenaires dépouillent ou relèvent les mêmes choses. Il est possible d'envisager que puissent alimenter et avoir accès à cette base d'autres intervenants, comme la demande en est d'ores et déjà formulée (par le CNES par exemple). Cela aurait l'avantage de démultiplier les efforts des équipes du Rint, et d'enrichir ce fonds d'attestations néologiques dans une perspective encore inimaginable il y a quelques mois.

## 3 Principes d'organisation

### 3.1 Les principales fonctions

Le but du réseau *Balnéo* est de collecter rapidement et de diffuser largement au moyen d'un réseau informatique des attestations de néologismes lus ou entendus par des observateurs francophones. Dans la première phase de lancement il faut pour cela que :

- Le réseau informatique se calque sur la structure hiérarchique du réseau commanditaire (le Rint) qui est composé de modules nationaux qui travaillent avec des organismes dans lesquels des rédacteurs sont chargés de la production des «attestations»;
- La gestion du réseau informatisé soit automatisée au maximum : après son inscription (par une instance supérieure) sur le réseau, tout membre ajoute, consulte et commente librement les attestations sans intervention d'une tierce personne (il n'y a ni censeur ni «modérateur»);

– Des principes méthodologiques communs soient adoptés au sein du Rint pour la description d'un néologisme et que le réseau informatique collecte et fasse circuler les «fiches» ainsi définies (l'information échangée sur le réseau est purement textuelle, et le jeu de caractères utilisé est l'Iso-latin);

– Le réseau informatique garantit que toute information diffusée en mentionne le rédacteur, le titulaire des droits de réutilisation, et informe sur les conditions de réexploitation et de cession;

– Seul le rédacteur d'une information puisse la modifier; mais une instance supérieure peut se substituer à lui en cas d'abus ou de défaillance;

– Tout membre du réseau puisse annoter une information produite par un autre membre. Les annotations sont jointes à l'information originale lors de sa diffusion. Elles ne peuvent être modifiées ou supprimées que par leur auteur.

### 3.2 Les usagers

– Les membres du réseau sont très dispersés (au moins en Europe, en Amérique et en Afrique) mais il sont supposés avoir un accès à Internet à partir de micro-ordinateurs sous *Windows*.

– Le travail de saisie et de mise à jour de l'information est long, il doit s'effectuer «hors ligne» avec exportation périodique des données vers la base centrale.

– La consultation de cette base et la rédaction d'annotations doivent pouvoir se faire «hors ligne» et «en ligne».

– Les rédacteurs sont essentiellement des utilisateurs de traitement de texte. Les outils locaux de saisie et les procédures d'import/export via Internet doivent donc être simples et conformes aux règles ergonomiques de *Windows* qu'ils connaissent déjà.

## 4 Le réseau informatique *Balnéo*

Développé dans le cadre très évolutif (voire mouvant) d'Internet, le système *Balnéo* a connu deux étapes bien distinctes. De juin à décembre 1994, un prototype a permis de tester la faisabilité d'un tel réseau. Il a été utilisé par les premiers modules nationaux qui avaient accès à Internet, notamment le module belge. Les observations et commentaires ont été utilisés pour concevoir le produit final qui devrait voir le jour pendant le premier semestre de 1996.

### 4.1 Principes de la première version (décembre 1994) : échanges en différé par messagerie

#### 4.1.1 La production de l'information

La première version de *Balnéo* utilise uniquement une base de données locale (dite «base éditoriale») sous *Windows* pour la saisie des néologismes et la génération des messages administratifs. Cette base de données développée à partir d'*Access* de Microsoft permet la saisie des néologismes, leur recherche, leur mise à jour etc. et la formulation de requêtes.

#### 4.1.2 L'envoi de l'information

Une fonction d'exportation exporte la base de données ou la requête dans un fichier ou dans le presse-papiers de *Windows*. Le résultat de cette exportation est transmis par un logiciel de messagerie (Mail, Eudora...) à une boîte aux lettres située sur un serveur central.

#### 4.1.3 Le traitement des messages

Sur le serveur central un programme capte le message (selon le principe des «serveurs de liste» qui a inspiré cette première version). Il identifie l'émetteur et interprète le contenu du message. Si c'est une commande d'ajout, il enregistre les données jointes au message dans la base de données relationnelle du système central. S'il s'agit d'une requête (question), il interroge cette même base centrale et renvoie les réponses dans la boîte aux lettres de l'émetteur du message. Celui-ci lit son courrier, extrait la réponse dans un fichier ou le presse-papiers puis l'importe dans sa base éditoriale locale où il peut lire, annoter, éditer...

#### 4.1.4 Avantages et inconvénients de cette méthode

Le procédé est simple à mettre en œuvre. Il se prête bien à la production des données (travail long sur la base locale puis envoi périodique en masse) mais il est lent et lourd pour les interrogations ponctuelles de la base centrale (pour vérifier par exemple si un néologisme a déjà été enregistré).

Le défaut majeur de la messagerie réside dans le procédé d'identification de l'émetteur : les logiciels de messagerie sous *Windows* ne contrôlent ni le nom ni l'«adresse Internet» de l'utilisateur, il est alors pratiquement impossible d'authentifier l'émetteur d'un message et donc de connaître son statut et ses droits. Sur ce point une expérimentation est en cours, qui permettra d'authentifier tout utilisateur entrant sur le réseau informatique (voir 4.2).

### 4.2 Les évolutions (mars 1996) : l'accès par le *World Wide Web* (*W3*)

Peu après la conception de cette première version, le standard HTML

2+ qui est à la base des services W3 (connus aussi sous le nom de «Web» ou, en français «cyberboile») a apporté un nouvel élément pour la réalisation de services interactifs : les formulaires. Grâce à eux, l'utilisateur d'un client W3 peut transmettre des informations à un serveur W3 où elles sont immédiatement traitées par un programme. Des développements informatiques spécifiques ont abouti par ailleurs à une solution satisfaisante dans le domaine de l'authentification de l'utilisateur (Le Meur 1995). Une nouvelle version de *Balnéo* plus interactive et plus «sécurisée» est donc en cours de développement.

Tout utilisateur reçoit lors de son inscription sur le réseau un mot de passe. Par la suite, quand il se connecte au moyen d'un client W3, il fournit ce code confidentiel. L'utilisateur authentifié peut procéder à toutes les opérations (administration, alimentation du réseau, annotation, interrogation) selon le statut et les privilèges qu'il a dans le réseau.

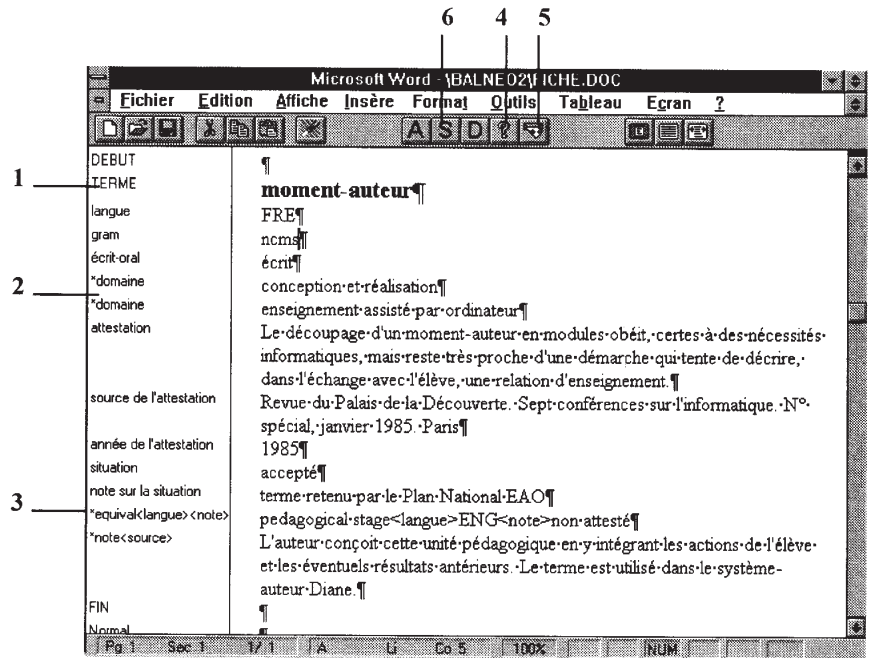
Quand il lit une fiche, il peut taper une annotation qui sera immédiatement ajoutée à cette fiche.

Un autre formulaire permet une saisie rapide et en ligne d'une attestation mais les limites de l'éditeur des clients W3 (*Netscape* ou *Mosaïc*) rendent l'opération difficile pour certains types de champs complexes et répétitifs.

Autre évolution importante : la première version de *Balnéo* ne disposait que de la base éditoriale *Access* pour la saisie des attestations.

Les utilisateurs ont souhaité disposer d'outils plus proches de leur logiciel favori : le traitement de texte. L'exemple ci-dessous illustre la saisie d'un néologisme telle qu'elle pourra s'effectuer sous *Word 2* pour *Windows*:

Les données saisies ou importées dans la base éditoriale *Access* peuvent être exportées en format SGML, RTF ou Iso 2709 pour une réutilisation par une base de données documentaire, un



1. Chaque élément d'information est constitué d'un paragraphe dont le style (*Terme* par exemple) indique la nature;
2. Un champ peut être répétitif : il y a ici deux domaines;
3. Un champ est parfois complexe : un équivalent est suivi de sa langue (obligatoire) et peut être suivi d'une note;
4. Cette icône lance un programme de vérification de la fiche en cours;
5. Cette icône vérifie toutes les fiches et les dépose dans le presse-papiers en vue d'une exportation;
6. Les icônes A et S *Ajoutent* ou *Suppriment* une fiche;
7. L'icône D permet de définir des «valeurs par défaut» afin d'éviter les saisies répétitives.

logiciel de PAO, une base diffusable sur CD-Rom, etc.

Sur le site central, la gestion des données administratives et néologiques est assurée par un SGBD relationnel (Oracle 7 en l'occurrence). Ce type de base est nécessaire pour la gestion courante (gestion des usagers, des codes confidentiels, des ajouts de fiches, des mises à jour, des suppressions, des annotations) mais il

présente des limites dans la consultation des néologismes (par exemple la recherche des composants d'un terme). Les néologismes seront donc transférés périodiquement dans une base WAIS (Wide Area Information Service). Un formulaire de recherche permettra de formuler des équations de recherche plus fines à partir du client W3.

### 4.3 Perspectives et échéances

Lors de son dernier Comité d'orientation en janvier 1996, le Rint a fait le point sur le développement de *Balnéo* et a pris plusieurs décisions qui engagent son avenir.

– La structure d'une attestation a été définitivement stabilisée. La liste des 17 éléments constitutifs a été fixée. Il ne reste qu'à établir des listes fermées de valeurs pour certains champs comme celui des catégories grammaticales.

– Le modèle d'une attestation formalisé en SGML va être mis à la disposition des modules et des organismes qui souhaitent alimenter *Balnéo* à partir de leur propre module de saisie.

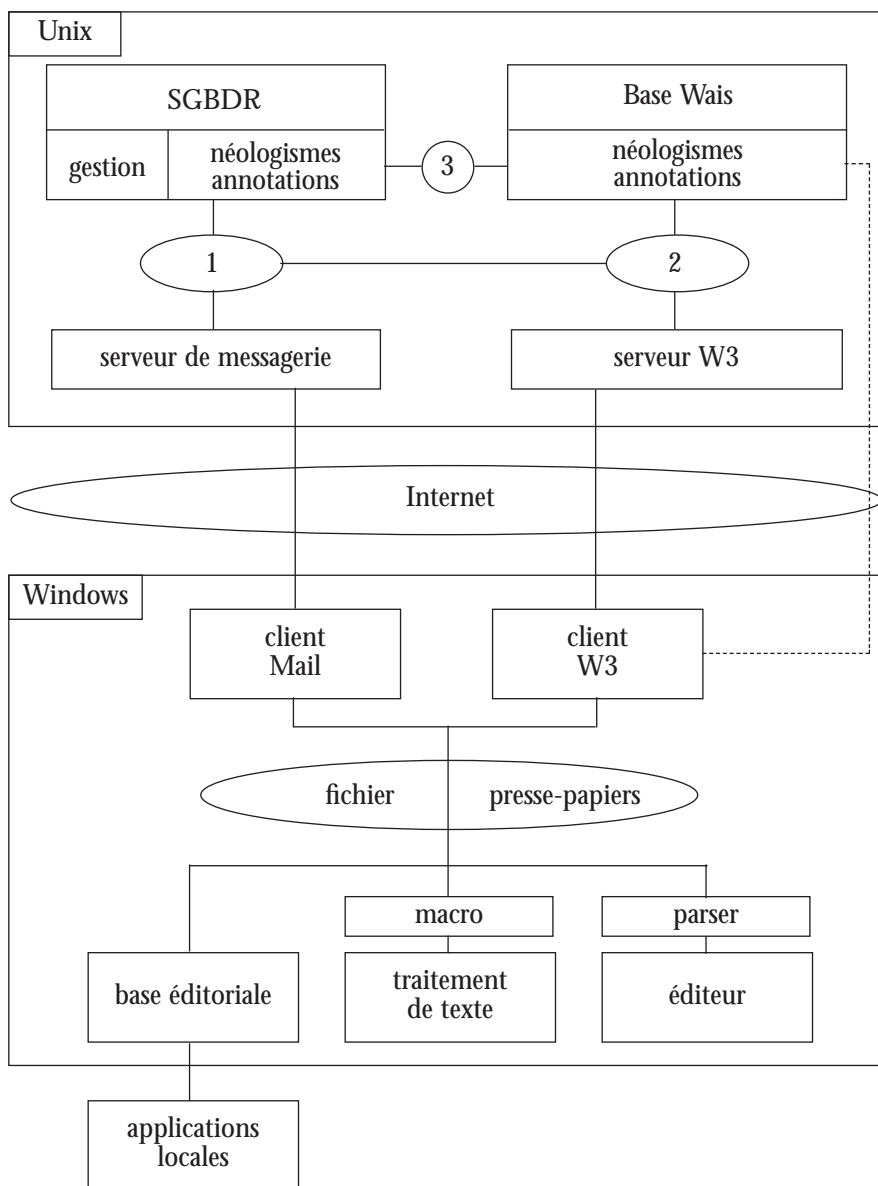
– Le système central, qui a été développé au Centre de ressources informatiques de l'Université de Rennes I, va être «porté» sur les machines de l'«exploitant» : le CRI de Rennes II (mars-avril 1996)

L'ensemble du dispositif devrait donc être opérationnel à la fin de ce premier semestre 1996. Il constituera un des éléments de la présence du Rint sur Internet.

Pour la première fois, il est possible de gérer selon un plan global les attestations de néologismes d'une langue, le français. Ce cadre sera étendu dans un premier temps aux langues partenaires de la francophonie, puis à d'autres langues.

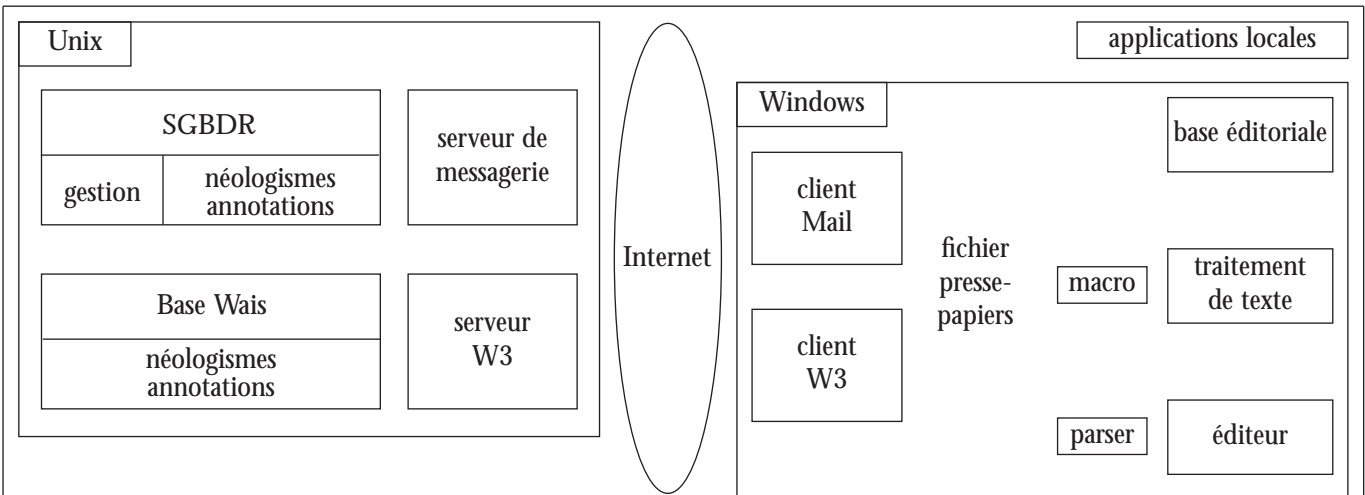
Parmi les autres développements envisagés figure une intégration plus poussée des relevés d'attestation de néologismes. Il devient envisageable en effet d'automatiser le repérage, le traitement, et la mise en forme de fiches d'attestation de néologismes à partir de banques de données textuelles, en plein essor aujourd'hui.

### *Balnéo* : vue d'ensemble



Sur le site central trois applications ont été programmées :

- (1) reçoit, authentifie et traite les messages provenant de la boîte aux lettres ou du programme (2). Il met à jour la base de données relationnelle, renvoie un message à la boîte aux lettres du demandeur ou rend un message à (2).
- (2) authentifie un demandeur, transmet ses demandes à (1), rend les messages de retour au client W3.
- (3) convertit la base de néologismes au format WAIS et produit une base interrogeable par (2).



## 5 Quelques éléments de réflexion

Pour faciliter les développements informatiques et la maintenance du réseau, la structure de toutes les informations (administratives ou néologiques) est définie dans un modèle unique, une DTD SGML. Outre les avantages d'un formalisme unique, ce choix ouvre le réseau sur les outils standard développés autour de cette norme. On peut ainsi construire librement des éditeurs de messages *Balnéo* et utiliser les «parsers» pour vérifier leur conformité au modèle ou utiliser les langages de manipulation de données tels que *Balise* de la société AIS, *Omnimark* de Exoterica pour produire des dictionnaires, des CD-Rom, etc.

*André Le Meur,*  
*Université de Rennes II,*  
*Loïc Depecker,*  
*Délégation générale à la langue française,*  
*Paris.*

## Bibliographie

Le Meur (André), 1995 : «Méthodes et outils pour un travail de groupe sur Internet. Une application à la veille néologique», dans *Actes JRES95*, Comité réseau des universités, Unité réseau du CNRS, Chambéry novembre 1995 (à paraître).

# Une autre façon de mesurer l'utilisation du français écrit dans les entreprises

Ce texte présente les principaux résultats d'une étude empirique commandée par l'Office de la langue française afin de valider deux échelles de mesure du français écrit, telles qu'elles sont conçues par ce dernier. À partir d'une méthode développée par Christopher McAll et son équipe, soit la reconstruction de la journée de travail, l'auteur en arrive à dresser un tableau assez précis de l'utilisation du français écrit dans quatre entreprises réparties selon les secteurs de la production, de la distribution et des services.

Termes-clés :

choix linguistique; alternance de codes; reconstruction de la journée de travail; pratiques linguistiques; matériaux écrits.

*«Le texte écrit est une réalité complexe : il est à la fois un objet de communication, produit d'une pratique sociale et culturelle, ainsi qu'un lieu de production et de circulation du sens.»*

H. Boyer

## 1 Introduction

**D**epuis les années 70, l'Office de la langue française intervient dans les domaines des communications, du travail, du commerce et des affaires tant du côté public que privé, afin d'élargir l'utilisation du français au Québec, et ce, tel qu'il est prescrit par la *Charte de la langue française*. Or, au cours de ces dernières années, l'Office, alimenté en cela par une réflexion critique sur l'état d'avancement de la francisation, s'est trouvé devant la nécessité d'affiner les points de repère de ses interventions, soit les évaluations touchant à l'utilisation du français, puisqu'il lui a été démontré que l'obtention d'un certificat de francisation par les entreprises ne signifiait pas nécessairement un usage correspondant du français, comme on serait pourtant tenté de le croire (Plourde 1988). Aussi, devant le manque de fiabilité de ce vecteur linguistique qu'est le taux de certification, l'Office s'est fixé comme objectif d'élaborer une mesure d'évaluation simple du degré d'utilisation du français écrit et oral en milieu de travail<sup>(1)</sup>.

S'inscrivant dans cette démarche, ce texte présente les principaux résultats d'une étude empirique

(Teixeira 1995) menée dans quatre entreprises. Le but de cette recherche était de valider qualitativement auprès d'une vingtaine d'interviewés, à raison de cinq par entreprise, deux échelles de mesure de l'utilisation du français dans les matériaux écrits, l'une ayant trait au choix linguistique (degré de présence du français) et l'autre à l'alternance de codes (degré de présence de termes anglais).

(1) Pour atteindre cet objectif, l'Office a formé un comité *ad hoc* appelé *Groupe de travail sur les paramètres*. Par ailleurs, l'élaboration de cette mesure a été inspirée par la définition qu'a donnée C. Loubier (1993 : 65) de l'implantation du français comme étant à la fois un processus et un résultat. Toutefois, lors de la construction de la mesure, le Groupe de travail sur les paramètres a choisi de ne retenir que l'aspect résultat de cette définition, car cela «semblait mieux coller au mandat confié [au groupe] d'élaborer une mesure permettant de positionner les entreprises sur un continuum (ou plusieurs continuums, s'il y a lieu) d'utilisation réelle du français» (Bouchard 1994b : 3), ce qui du même coup rendait possible l'évaluation des entreprises inscrites ou non dans le processus officiel d'implantation du français (processus faisant suite à l'adoption de la *Charte de la langue française*).

# Québec

## 2 La méthodologie

### 2.1 Une problématique à trois temps

Pour expérimenter sur le terrain la validité de l'échelle de mesure du français écrit, nous avons mis au point une problématique à trois temps. Ainsi, en premier lieu nous avons demandé aux interviewés d'évaluer le degré du choix linguistique des documents à partir d'une échelle à choix multiples<sup>(2)</sup>. Dans un deuxième temps, en reprenant les mêmes documents nous avons cherché à évaluer le degré de l'alternance de codes, et ce, également à partir d'une échelle constituée de plusieurs niveaux<sup>(3)</sup>. Pour compléter l'ensemble, les interviewés ont enfin été amenés à se prononcer sur leur

(2) Cette échelle est l'application méthodologique de la définition conceptuelle du choix linguistique. Elle a été aussi développée par le Groupe de travail sur les paramètres (Bouchard 1994c).

Concrètement, celle-ci est la suivante : 6) en français uniquement; 5) beaucoup plus en français qu'en anglais; 4) un peu plus en français qu'en anglais; 3) bilingues (ou versions distinctes); 2) un peu plus en anglais qu'en français; 1) beaucoup plus en anglais qu'en français; 0) en anglais uniquement.

(3) Elle a été aussi élaborée par le Groupe de travail sur les paramètres (Bouchard 1994c). Elle se présente comme suit : 6) aucun emprunt; 5) beaucoup moins d'emprunts que de termes français; 4) un peu moins d'emprunts que de termes français; 3) autant d'emprunts que de termes français; 2) un peu plus d'emprunts que de termes français; 1) beaucoup plus d'emprunts que de termes anglais; 0) uniquement en anglais.

perception générale quant à leur utilisation du français écrit et sur les raisons expliquant leur pratique linguistique. De cette manière, il nous apparaissait possible de contre-vérifier, jusqu'à un certain point, la véracité des informations recueillies par les deux échelles de mesure.

Pour mener à terme une telle problématique, en particulier en raison de la diversité et du nombre des matériaux écrits circulant dans les entreprises, nous avons suivi la classification des formes d'écritures du Groupe de travail sur les paramètres (GTP 1993). Ainsi, nous avons regroupé les matériaux écrits selon les catégories suivantes : 1) les formulaires<sup>(4)</sup>; 2) les autres documents<sup>(5)</sup>; 3) les autres écrits<sup>(6)</sup>. De plus, il faut ajouter que nous avons distingué auprès des formulaires l'aspect imprimé<sup>(7)</sup> de l'aspect rempli<sup>(8)</sup>. Bref, la mesure d'évaluation du français écrit s'est appliquée à l'aspect imprimé de tous les matériaux et aussi à l'aspect rempli des formulaires. Par la suite, lors de

(4) Ce sont des documents constitués par un aspect imprimé et un aspect rempli comme les demandes de matériel et les bons de commande.

(5) Ce sont des documents qui n'ont qu'un cadre imprimé comme les extraits d'ordinateur et les rapports comptables.

(6) Ce sont des manifestations d'écriture qui se situent en dehors des documents, comme les messages d'un écran d'ordinateur et les inscriptions sur des panneaux de commande.

(7) Il désigne le contenu d'impression de tout support d'écriture.

(8) Il désigne l'écrit qu'appose manuellement (crayon, stylo, etc.) ou mécaniquement (machine à écrire, imprimante, etc.) une personne au cadre d'impression d'un formulaire.

l'analyse des résultats, nous avons divisé les matériaux en trois catégories.

### 2.2 La reconstruction de la journée de travail

Les outils conceptuels de cueillette des données utilisés ici se trouvant être à mi-chemin entre les procédures qualitative et quantitative sans pour autant en embrasser l'une ou l'autre complètement, il nous est apparu pertinent d'adopter une stratégie développée par C. McAll et son équipe de recherche en sociologie du langage. Cette méthode repose sur l'idée de reconstruire en temps différé la journée de travail de la personne sélectionnée (McAll, Montgomery et Tremblay 1994) lors de la recherche. Elle consiste donc à rencontrer les interviewés dans les premières heures suivant la fin de leur période de travail et à leur faire relater cette dernière.

Dans le cas de notre étude, l'informateur devait reconstituer son quart de travail en précisant les matériaux écrits utilisés. Ensuite, il était invité à évaluer l'utilisation de la langue. En étant capable de rattacher les matériaux écrits à des pratiques concrètes et récentes de travail, les personnes rencontrées pouvaient se montrer plus précises dans leurs évaluations langagières (dénombrement des matériaux écrits utilisés) et linguistiques (mesure d'évaluation du français écrit). Résultat, cette méthode nous a permis d'analyser près de 196 types de matériaux écrits distribués, il est vrai, surtout parmi les catégories des autres documents et formulaires.

### 2.3 Le choix des entreprises et des postes de travail

Au lieu d'utiliser des critères comme l'ancienneté de l'obtention du certificat de francisation ou le nombre d'employés, nous avons décidé de

choisir les entreprises selon leur activité économique, car une récente étude avait révélé le poids important de ce facteur sur la nature des matériaux écrits utilisés au travail (Bouchard 1994a : 2). Donc, des quatre entreprises choisies, deux se situaient respectivement dans le secteur de la production<sup>(9)</sup> et dans celui de la distribution<sup>(10)</sup>. En raison de la diversité et de l'importance du secteur des services, nous avons décidé d'y sélectionner les deux autres, soit plus précisément l'une dans le sous-secteur des institutions financières<sup>(11)</sup> et l'autre dans celui de l'hébergement et de la restauration<sup>(12)</sup>. En termes d'effectifs, toutes ces entreprises avaient plus de 100 employés au Québec, sauf celle du sous-secteur de la distribution, qui en comptait entre 50 et 100.

En plus d'avoir accès aux grands secteurs économiques, il nous est apparu important de tenir compte de la division des pratiques langagières<sup>(13)</sup> (McAll 1992) qui prévaut dans les entreprises. Notre étude ne s'est donc pas restreinte à une ou deux catégories professionnelles ou occupationnelles dont le travail serait plutôt de type

(9) Cette entreprise est située à Montréal et travaille entre autres dans les domaines de la conception et de la fabrication des câbles de puissance et de communication.

(10) Cette entreprise se situe à Mont-Royal et travaille dans la fabrication, la vente et l'entretien d'équipement informatique de bureau.

(11) C'est une banque à charte, qui offre des services financiers aux particuliers et aux entreprises. La succursale que l'on a visitée est située à Laval.

(12) Cette entreprise est située dans le centre-ville de Montréal et évolue dans les domaines de l'hébergement et de la restauration.

«langagier»<sup>(14)</sup> (McAll, Montgomery et Tremblay 1994 : 5). Ce faisant, les informateurs ont été sélectionnés à partir d'une grille conceptuelle à cinq branches comprenant les cadres, les «professionnels», les employés du soutien technique<sup>(15)</sup>, les employés du soutien technique et de service<sup>(16)</sup> et les employés du soutien de service<sup>(17)</sup>.

(13) Comme il est suggéré par C. McAll, cette répartition structurée des pratiques langagières fonctionnerait principalement selon trois axes : complexité/simplicité; originalité/non-originalité; fréquence/absence. Par ailleurs, il est à noter que le poids de cette division sociale s'est fait particulièrement sentir dans le cas de trois interviewés (l'opérateur, le mécanicien et l'agent de sécurité) au point que nous avons dû élargir notre champ d'investigation, en intégrant aussi les autres tâches non effectuées au cours de leur période de travail décrite, de manière à obtenir un nombre suffisant de documents à analyser.

(14) C'est-à-dire un travail où le langage joue le rôle de matière première et non pas simplement celui d'accessoire ou d'outil servant, au cours de la tâche, à la transformation des autres matériaux comme le métal et le plastique. Bref, une activité où le langage est à la fois le tenant et l'aboutissant même du travail.

(15) Ce sont des employés qui n'ont pas de contacts quotidiens avec la clientèle interne et externe de l'entreprise en rapport avec leurs tâches.

(16) Ces employés ont des contacts moyennement réguliers avec la clientèle interne et externe de l'entreprise lors de la réalisation de leurs tâches.

(17) Ces employés ont des contacts continuels avec la clientèle interne et externe de l'entreprise lors de la réalisation de leurs tâches.

Et d'ailleurs, la présentation des résultats suivra cet ordre pour chacune des quatre entreprises.

## 2.4 Les contraintes méthodologiques

Avant de terminer la présentation de cette première partie, il convient de mentionner les deux principales contraintes rencontrées en cours de route. D'abord, pour que les employeurs puissent nous laisser interviewer un nombre suffisant de répondants, nous avons dû réduire dans certains cas le temps des entrevues et interroger plusieurs personnes avant la fin de leur période normale de travail. Aussi, la majeure partie des résultats de cette recherche ne peuvent prétendre couvrir l'ensemble des matériaux écrits qui auraient été utilisés dans une journée de travail habituelle. Enfin, la distinction imprimé/rempli dans le cas des formulaires n'a pas été considérée intégralement dans toutes les entrevues suite à une erreur de manipulation. Il en résulte que, dans certaines d'entre elles, seul l'aspect imprimé ou l'aspect rempli a été considéré. Toutefois, comme cet impair ne touche que les formulaires, soit seulement une partie des données, nous croyons que l'analyse restera valable dans son ensemble.

## 3 Une entreprise du secteur de la production

En recourant à la reconstruction de la journée de travail, nous avons pu obtenir, pour les cinq informateurs retenus, les résultats suivants en matière d'utilisation linguistique à l'écrit.

Dans le cas du contremaître, il semble qu'il travaille dans un espace exclusivement français. L'ensemble des matériaux écrits (dont l'aspect

imprimé des formulaires) utilisés durant sa journée de travail étaient en français. Et ces documents étaient, à une exception près, exempts de tout emprunt. Pour l'intervu, cette utilisation maximale du français s'expliquerait par trois faits :

- 1) L'ensemble des matériaux écrits circulant dans l'aire de la production sont en français;
- 2) Beaucoup d'opérateurs ne connaissent pas l'anglais;
- 3) L'entreprise appartient à des intérêts français.

Du côté du contrôleur adjoint, la situation s'avère être l'inverse de celle du contremaître, puisqu'ici il y a une présence importante de l'anglais même s'il n'y a pas perception d'utilisation d'emprunts. En effet, les divers documents utilisés sont uniquement en anglais, alors que les formulaires (imprimés) sont pour la plupart bilingues ou francisés. Selon ce contrôleur, cette tendance lourde vers l'unilinguisme anglais s'expliquerait dans son cas par le fait qu'il produit et reçoit beaucoup de matériaux écrits à destination et en provenance du siège social de l'entreprise et des succursales, situés à l'extérieur du Québec. Aussi, on comprend pourquoi il a plutôt l'impression de travailler en anglais à l'écrit.

Quant au technicien, son lieu de travail est en partie comparable à celui du contremaître. De fait, nous sommes ici devant une «situation totale» d'unilinguisme français pour l'ensemble des matériaux écrits (comprenant l'aspect imprimé des formulaires), soit une présence maximale du français accompagnée d'une absence d'emprunts. Selon l'intervu, une telle situation s'expliquerait par l'importance accordée, à l'interne, à l'usage des matériaux rédigés en français et par la «petitesse» de l'emploi de ceux écrits en anglais. Dès lors, il n'est pas étonnant de constater que ce

technicien a l'impression de travailler en français à l'écrit.

Pour ce qui est de l'opérateur, son environnement écrit est orienté vers le français, malgré une présence non négligeable de l'anglais. En effet, si certains matériaux mentionnés sont en anglais uniquement, la plupart sont en français uniquement ou beaucoup plus en français qu'en anglais. Toutefois, il y a ici absence d'usage d'emprunts. Pour ce travailleur, cette situation favorable au français s'expliquerait par le fait qu'il est en quelque sorte la norme linguistique de son lieu de travail tant du point de vue de son rayon d'action que de son intensité d'application. Aussi a-t-il l'impression de travailler entièrement en français à l'écrit.

Enfin, du côté de l'acheteur, on dénote une plus grande présence de l'anglais que du français et une utilisation plus ou moins importante d'emprunts. En effet, si les formulaires (imprimés) utilisés sont en français uniquement, il en est autrement des autres documents et autres écrits, qui, selon le cas, vont du bilinguisme à l'unilinguisme anglais. Par conséquent, cet acheteur a l'impression de travailler plus en anglais qu'en français et ce en raison de ses nombreux contacts professionnels avec des fournisseurs situés à l'extérieur du Québec, l'amenant du coup à utiliser l'anglais.

En répartissant les lieux de travail de cette entreprise en deux territoires organisationnels distincts, il nous est possible d'observer que l'anglais est beaucoup plus présent du côté de l'administration (avec le contrôleur-adjoint et l'acheteur) que du côté de la production (avec le contremaître, le technicien et l'opérateur), tant pour le choix linguistique que pour l'alternance de codes. Il en résulte que cette entreprise a une situation à l'écrit où l'on rencontre autant le français que l'anglais, mais dans le cadre d'un bilinguisme articulé selon deux

espaces qui tendent vers un unilinguisme spécifique et différent – l'anglais pour l'administration et le français pour la production.

## 4 Une entreprise du secteur de la distribution

À partir des cinq entrevues réalisées au sein de cette entreprise nous avons découvert que, de façon générale, la situation linguistique y était légèrement en faveur de l'anglais. Mais qu'en est-il plus exactement des différents postes de travail observés?

D'abord du côté du directeur régional, les pratiques linguistiques à l'écrit (y compris l'aspect imprimé des formulaires) sont assez partagées entre le français et l'anglais. De fait, la situation va du français uniquement à l'anglais uniquement en passant par une utilisation de documents en versions bilingues ou distinctes. Nous avons constaté également que l'intervu recourait de façon prononcée aux emprunts. Dès lors, il n'est guère surprenant que cet informateur émette un avis nuancé quant à son impression de travailler en français à l'écrit. Ainsi, il a affirmé qu'il utilisait le français avec les clients et les employés situés au Québec, mais l'anglais avec le siège social, les employés et les clients situés en dehors du Québec. Bref, il emploie presque autant l'anglais que le français et avoue qu'il ne pourrait travailler plus en français en raison de ses contacts professionnels.

Quant à l'univers écrit (dont l'aspect imprimé des formulaires) du contrôleur, il est pour sa part traversé par l'anglais avec toutefois un usage limité d'emprunts pour les matériaux rédigés totalement ou en partie en français. Ainsi, tandis qu'une majorité d'entre eux sont en français uniquement, beaucoup plus en français qu'en anglais ou bilingues, une minorité est écrite en anglais

uniquement. Par conséquent, ce contrôleur a véritablement l'impression de travailler en français à l'écrit. Et cette perception apparaît vraisemblable dans la mesure où il précise qu'il travaille en français dans 55 % du temps, le reste étant en anglais.

Quant à la secrétaire, son lieu de travail (comprenant l'aspect rempli des formulaires) est un lieu où se côtoient le français et l'anglais. Le français semble être cependant un peu plus présent et le recours aux emprunts y est généralement faible. Ainsi, une grande partie des documents mentionnés sont en versions distinctes ou en français uniquement, tandis que l'autre partie est davantage orientée vers l'anglais et parfois vers l'unilinguisme anglais. Malgré la place quelque peu prépondérante accordée au français, l'interviewée n'a pas l'impression de travailler en français à l'écrit. En effet, elle affirme passer 60 % de son temps au travail en anglais. Selon elle, cela découlerait de ses nombreux contacts avec le siège social de l'entreprise situé, nous l'avons vu, à l'extérieur du Québec.

Pour ce qui est du technicien supérieur, sa situation linguistique (y compris l'aspect rempli des formulaires) apparaît à l'inverse de celle de la secrétaire. De fait, il utilise des matériaux écrits où l'anglais est légèrement plus présent que le français, même s'il n'y a pas de présence d'emprunts du côté des documents français. Toutefois, il est surprenant de remarquer que, dans un tel contexte, cet informateur ait malgré tout l'impression de travailler en français. Selon lui, ceci s'expliquerait par le fait que la très grande majorité des clients avec qui il a des contacts sont au Québec.

Enfin du côté du technicien, son univers écrit (dont l'aspect rempli des formulaires) est dominé par l'anglais au point que même le seul document mentionné en français comporte des emprunts. En effet, tous les matériaux

présentés, sauf dans le cas d'un formulaire, sont en anglais uniquement. Aussi est-on surpris de constater que l'interviewé a malgré tout l'impression de travailler en français, du moins plus qu'avant. En effet, à l'oral il se sert davantage du français maintenant, mais du côté de la documentation technique ce n'est jamais le cas. Les propos de l'interviewé sur l'écrit finissent donc par rejoindre les résultats recueillis dans le cadre de sa journée de travail.

Si nous répartissons les cinq cas retenus dans cette entreprise en deux territoires organisationnels, soit l'administration (avec le directeur régional, le contrôleur et la secrétaire) et l'atelier (avec le technicien supérieur et le technicien), on s'aperçoit alors que le français et l'anglais sont autant présents l'un que l'autre du côté de l'administration, tandis que l'anglais occupe l'essentiel de l'espace du côté de l'atelier. Par ailleurs, quel que soit le lieu d'observation, les matériaux écrits mentionnés ne comptent pas ou peu d'emprunts. En bout de ligne, l'univers écrit de cette entreprise de distribution est donc fortement partagé entre le français et l'anglais, et même légèrement en faveur de l'anglais.

## 5 Le cas d'une institution financière

En analysant les témoignages des cinq personnes rencontrées pour notre recherche, nous avons pu noter que la situation linguistique à l'écrit de cette entreprise était nettement en faveur du français au point que l'on peut véritablement parler de situation linguistique hégémonique pour ce dernier.

Ainsi, du côté de la directrice des services financiers (division des prêts), tous les matériaux écrits (comprenant l'aspect imprimé/rempli des formulaires) sont en français

uniquement, sauf son ordinateur personnel qui est en versions distinctes. Quant aux emprunts, ils sont très rarement mentionnés. Aussi, il n'est pas étonnant de voir que cette informatrice a l'impression de travailler en français à l'écrit. Selon elle, cela s'expliquerait par le fait que son milieu de travail (comprenant les employés et les clients) est francophone à 99,9 % et que la majorité des employés ont de la difficulté à utiliser l'anglais.

La même situation linguistique (y compris l'aspect imprimé/rempli des formulaires) prévaut aussi chez le directeur des services financiers (division placement et dépôt). Dans un contexte d'unilinguisme aussi grand, en ce qui touche la place du français et l'absence de tout emprunt, là encore il n'est pas surprenant de constater que l'interviewé a l'impression de travailler en français à l'écrit. Les éléments explicatifs présentés par ce dernier vont également dans le même sens que ceux de la directrice.

Quant à l'assistante-responsable du service à la clientèle, même si elle signale la présence de quelques emprunts pour l'aspect rempli de certains formulaires, la situation linguistique (dont aussi l'aspect imprimé des formulaires) n'y est pas vraiment différente de celle des deux précédents informateurs. De plus, elle a aussi l'impression de travailler en français à l'écrit. D'ailleurs, pour expliquer cette situation, elle affirme qu'il n'y a pas plus francophone que son milieu de travail tant du point de vue de l'emplacement géographique que de la succursale que de l'homogénéité ethnolinguistique de ses collègues de travail.

Il en est de même pour la secrétaire de direction et pour la caissière, que nous avons interrogées par la suite. Ainsi, les matériaux écrits (comprenant l'aspect imprimé/rempli des formulaires) utilisés par ces dernières sont en français

uniquement, sauf dans le cas de l'ordinateur de traitement de texte de la secrétaire, qui était en version bilingue et dans celui du terminal administratif de la caissière, qui était de son côté un peu plus en français qu'en anglais. Bien plus, elles n'ont mentionné aucune présence d'emprunts. Après avoir dit qu'elles avaient l'impression de travailler en français à l'écrit, elles ont apporté des facteurs explicatifs qui allaient dans le sens des témoignages déjà présentés, à savoir qu'elles évoluent dans un environnement d'expression francophone.

En divisant les résultats recueillis pour cette entreprise en deux espaces organisationnels distincts, à savoir l'administration (avec la directrice de la division du prêt, le directeur de la division du placement et dépôt et la secrétaire de direction) et le comptoir (avec l'assistante-responsable du service à la clientèle et la caissière), il apparaît que le français est omniprésent autant dans l'un que dans l'autre. Par contre, il nous a été possible d'observer un nombre un peu plus élevé d'emprunts au comptoir qu'à l'administration, même, si, il faut le dire, ce nombre n'est pas significatif par rapport à l'ensemble des matériaux écrits utilisés dans chacun des deux territoires. Dès lors, il semble que le français écrit est en «situation hégémonique» dans cette institution financière, l'anglais y étant négligeable.

## 6 Le cas d'une entreprise d'hébergement et restauration

Du côté de cette entreprise, la situation linguistique à l'écrit varie d'un territoire organisationnel à un autre. C'est pourquoi nous sommes ici en présence d'une situation hétérogène qui contraste avec celle de l'institution financière.

En analysant les données relatives à la journée de travail (y compris l'aspect imprimé des formulaires) de la directrice adjointe, on découvre que le français y est plus présent que l'anglais. Ainsi, un grand nombre de formulaires sont en français uniquement ou beaucoup plus en français qu'en anglais, tandis qu'une minorité est en versions distinctes ou beaucoup plus en anglais qu'en français. Il en est de même pour la catégorie des autres documents où on n'en relève en fait que deux en faveur de l'anglais. Quant aux emprunts, ils sont surtout présents du côté des formulaires, mais en nombre restreint. Somme toute, l'interviewée a l'impression de travailler en français à l'écrit, d'autant plus que selon elle son milieu de travail est composé à 80 % de francophones et qu'elle favorise la promotion d'un français de qualité auprès de ses employés.

Quant au mécanicien, sa situation linguistique (dont l'aspect rempli des formulaires) à l'écrit se révèle être orientée vers le français, même si elle est quelque peu ponctuée par un «usage minimal» d'emprunts auprès des formulaires. Malgré qu'une très petite minorité de documents soit en anglais, il a l'impression, tout de même, de travailler en anglais, car les termes techniques qu'il emploie sont dans cette langue, et ce, même si les documents sont pour la plupart en français.

Pour ce qui est du magasinier, on remarque une prédominance du français avec une forte présence d'emprunts. Bref, une situation linguistique semblable à celle du mécanicien, mais qui s'exprime en des termes plus prononcés. La grande majorité des matériaux écrits (comprenant l'aspect imprimé des formulaires) sont en français uniquement. Quant à la minorité, elle est composée de formulaires en version bilingue ou en anglais uniquement. Du côté des emprunts, même si la situation est beaucoup

plus diversifiée, la concentration des emprunts auprès des trois types de matériaux écrits ne dépasse le niveau du : «autant d'emprunts que de termes français». Malgré cette présence non négligeable de l'anglais, l'informateur a l'impression de travailler en français à l'écrit, d'autant plus qu'à la suite d'une francisation des termes techniques où les termes anglais ne figuraient plus dans les documents, il s'est mis à utiliser le vocabulaire francisé.

En ce qui concerne l'agent de sécurité, l'espace linguistique (y compris l'aspect rempli des formulaires) y est nettement partagé entre le français et l'anglais avec toutefois une petite prédominance du français, même si les emprunts sont inexistantes. Aussi, au niveau de l'écrit, l'interviewé a l'impression de travailler à demi-temps en français et à demi-temps en anglais. Cela s'expliquerait, selon lui, par le fait que le français est utilisé pour les documents demeurant à l'hôtel, tandis que l'anglais l'est pour les documents destinés au siège social de l'entreprise, situé à l'extérieur du Québec.

Enfin, dans le cas de la serveuse, la situation linguistique penche en faveur du français. Cependant, il y a une présence non négligeable de l'anglais, tant du côté du choix linguistique que de l'alternance des codes. Malgré cet état de fait, cette informatrice a l'impression de travailler en français à l'écrit, parce que selon elle les documents sont en français et lorsqu'ils sont envoyés en version anglaise à son service, ils sont retournés à la direction sans avoir été lus. Bref, la présence de l'anglais n'est pas assez forte pour qu'elle ait l'impression d'être dans un environnement écrit bilingue.

En répartissant les résultats de cette entreprise selon les territoires organisationnels, on s'aperçoit que le français écrit occupe une place importante dans la section de l'administration (avec la directrice adjointe) et dans celle du restaurant

(avec la serveuse). Par contre, il a une place moins grande dans les «voies de circulation»<sup>(18)</sup> (avec le mécanicien, le magasinier et l'agent de sécurité). De façon générale, le français écrit est donc plus présent que l'anglais, même si celui-ci occupe une place quelque peu appréciable.

## 7 Les pratiques linguistiques des quatre entreprises

### 7.1 Une typologie des quatre entreprises selon le choix linguistique et l'alternance de codes

Après avoir présenté les résultats selon les postes de travail, il serait opportun maintenant de les regrouper afin de situer les entreprises sur les continuums du choix linguistique et de l'alternance des codes.

Pour ce faire, nous avons élaboré deux indicateurs<sup>(19)</sup> : un taux de présence du français et un taux de présence d'emprunts. Ces derniers vont nous permettre de tirer certaines conclusions. Une analyse des formulaires nous indique, d'abord, que le français occupe un espace très important dans l'institution financière (avec un taux de 99,8 %) et dans l'entreprise de production (91 %), alors qu'il n'atteint que 81 % dans l'entreprise d'hébergement et restauration et 42,4 % dans l'entreprise de distribution. Quant au

(18) C'est-à-dire un espace où les personnes rencontrées sont amenées à circuler à travers l'entreprise. C'est en raison de l'instabilité territoriale de leur poste de travail que nous avons apposé ce qualificatif de «voie de circulation».

(19) Pour connaître les modalités de calcul de ces deux indicateurs, nous invitons le lecteur à se référer directement au rapport final de cette recherche.

taux de présence d'emprunts pour les formulaires, l'entreprise d'hébergement et restauration se démarque avec un taux de 11 %, tandis qu'il n'atteint que 7,6 % dans l'entreprise de distribution, 4 % dans l'entreprise de production et 1,6 % dans l'institution financière.

En ce qui a trait à la catégorie des autres documents, les mêmes tendances en termes d'ordonnancement des entreprises ont pu y être observés<sup>(20)</sup>, mais avec cette nuance que le français est moins présent dans cette catégorie que dans celle des formulaires. Par surcroît, le taux de présence d'emprunts y est plus élevé<sup>(21)</sup>, exception faite de l'institution financière.

En ce qui touche à la catégorie des autres écrits, le taux de présence du français y est sensiblement comparable à ce que l'on a observé dans les deux premiers types de matériaux écrits<sup>(22)</sup>. Cependant, cette catégorie se démarque nettement dans le cas du taux de présence d'emprunts, puisqu'ici nous n'en retrouvons que dans l'entreprise d'hébergement et restauration (16,5 %)<sup>(23)</sup>.

En intégrant l'ensemble de ces données, tous matériaux écrits confondus, nous sommes maintenant en mesure de situer dans l'ordre

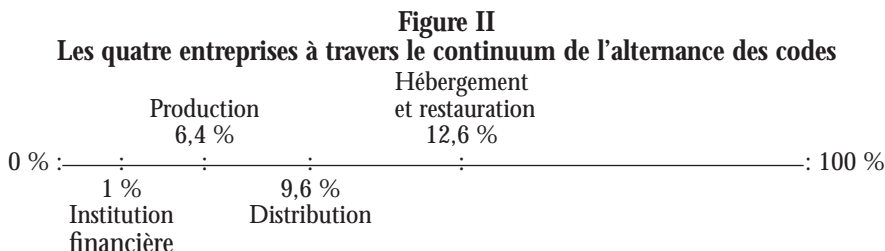
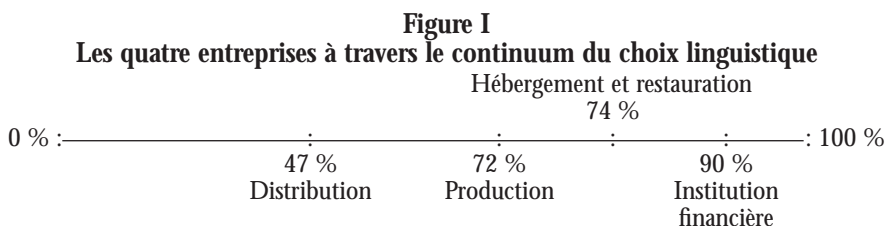
définitif de classement les quatre entreprises selon le choix linguistique et l'alternance des codes. On découvre alors que c'est dans l'institution financière que le français occupe la plus grande place à l'écrit (90 %), suivie de l'entreprise

(20) Soit l'institution financière avec un taux de 100 %, suivie de l'entreprise de production avec 66,6 %, de l'entreprise d'hébergement et restauration avec un taux de 58,6 %, et de l'entreprise de distribution avec 47,8 %.

(21) Soit de l'ordre de 12,6 % pour l'entreprise d'hébergement et restauration, de 12,3 % pour l'entreprise de distribution, de 8,5 % pour celle de la production et de 0 % pour l'institution financière.

(22) Ainsi, nous avons dans l'ordre l'entreprise d'hébergement et restauration avec 100 %, suivie de l'institution financière avec 71,6 %, de l'entreprise de production avec 53,6 % et de l'entreprise de distribution avec 50 %.

(23) C'est dire que pour le reste des trois entreprises ce taux se révèle être de 0 %.



d'hébergement et restauration (74 %), de l'entreprise de production (72 %) et loin derrière de l'entreprise de distribution (47 %).

Dans le cas du taux de présence d'emprunts, on trouve que l'entreprise d'hébergement et restauration obtient le taux le plus élevé (12,6 %), suivie par l'entreprise de distribution (9,6 %), par celle de la production (6,4 %) et par l'institution financière où le taux ne dépasse pas le 1 %.

En nous permettant ainsi d'atteindre une certaine finesse d'analyse, l'échelle du choix linguistique et dans une moindre mesure celle de l'alternance des codes pourraient s'avérer être un outil d'investigation utile pour l'Office. Cependant, compte tenu qu'elles ne constituent pas une panacée méthodologique, il serait pertinent d'indiquer leurs limites, d'où le prochain sous-thème.

## 7.2 Une réflexion sur le choix linguistique et l'alternance des codes

En effet, au cours de la réalisation de cette recherche, quelques constats nous sont apparus. Nous les présentons ici comme pistes éventuelles de réflexion.

Premièrement, les interviewés semblent avoir bien compris dans l'ensemble ce que l'on entendait par choix linguistique. Il en fut autrement pour l'alternance des codes : en plus d'une compréhension plus ou moins confuse de la question, il semble que les interviewés ne percevaient pas certains emprunts comme tels.

À la suite des difficultés d'application de cette mesure de l'alternance des codes, cette échelle ne nous apparaît donc pas comme un outil suffisamment fiable et fonctionnel pour être considéré comme un indicateur valable de

l'utilisation du français écrit en milieu de travail.

Deuxièmement, une certaine confusion est ressortie entre les expressions *en français uniquement* et *bilingues (ou versions distinctes)*, lorsque la version anglaise n'est pas utilisée tout en étant disponible. Il apparaît dès lors important de préciser la différence entre ce qui existe officiellement (formel) et ce qui est finalement appliqué dans l'entreprise (l'informel). Mais au-delà de ce problème mineur, notre recherche confirme la pertinence du choix linguistique comme mesure de l'utilisation du français écrit.

Et troisièmement, se situant plus en aval de cette réflexion, nous aimerions ici ajouter que lorsque nous avons tenté de mesurer la réalité linguistique à l'écrit, nous avons fait face à la langue parlée. Ainsi, là où l'on tente de saisir l'écrit se présente l'oral. Nous avons pu surtout le remarquer dans certains commentaires des interviewés sur leur impression de travailler en français ou non. Dans un contexte où l'Office tente, parallèlement à ce qui a été présenté dans cet article, de concevoir une mesure du français oral, il nous apparaît pertinent de souligner l'importance de comprendre les liens qui unissent et qui probablement à l'occasion séparent ces deux modes de communication que sont la langue parlée et la langue écrite.

## 8 Conclusion

À l'heure où l'Office s'interroge sur la fiabilité des évaluations linguistiques disponibles et où certaines pratiques de l'aménagement linguistique sont remises en question (Loubier 1994), on comprend toute l'importance de disposer d'outils de mesure pouvant aider à départager ce qui est de l'ordre du réel et du prescrit ou du supposé en matière d'utilisation du français – ici le

français écrit. En constatant les résultats de notre analyse, il nous semble que le choix linguistique pourrait adéquatement participer à ce renouvellement des méthodes d'investigation en matière de langue. Cependant, l'alternance des codes, dans les conditions actuelles, ne semble pas être un indicateur suffisamment sûr pour que l'on puisse s'y fier efficacement. Quant à la méthode de la reconstruction de la journée de travail, face au besoin de tenir davantage compte, lors des évaluations, de l'aspect social des pratiques linguistiques en milieu de travail (Loubier 1994), elle pourrait être une méthodologie particulière à laquelle on peut recourir pour mieux saisir la réalité sociale à l'étude<sup>(24)</sup>. Ceci permettrait éventuellement à l'Office de pouvoir améliorer ses interventions, eu égard à son rôle de planificateur de l'aménagement linguistique au Québec.

*Carl A. Teixeira,  
Université de Montréal,  
Québec.*

(24) Même si nous n'avons pas mis ici l'accent sur le contexte social concret (description de la journée de travail) dans lequel s'est inscrite l'utilisation des matériaux écrits (parce que notre objectif était autre), cette reconstitution minutieuse contient des éléments sociologiques qui pourraient nous aider à mieux comprendre l'utilisation du français.

## Bibliographie

Bouchard (Pierre), 1994a :

*Détermination des documents les plus représentatifs. Validation du questionnaire. Groupe de travail sur les paramètres*, Service de la recherche, Office de la langue française.

Bouchard (Pierre), 1994b : *Élaboration d'une mesure du degré d'implantation du français dans les documents en circulation dans divers milieux de travail*,

Communication présentée dans le cadre du Colloque de l'Acsalf, Document de travail, Service de la recherche, Office de la langue française.

Bouchard (Pierre), 1994c : *Travailler en français : confrontation de la définition du groupe de travail aux comportements des travailleurs*, Service de la recherche, Office de la langue française, inédit.

GTP, 1993 : *Questionnaire. Utilisation du français en entreprise*, (Groupe de travail sur les paramètres), Service de la recherche, Office de la langue française.

Loubier (Christiane), 1994 :

« L'implantation du français : du microcosme au macrocosme », dans *Terminologies nouvelles*, n° 12, p. 20-39.

Loubier (Christiane), 1993 :

« L'implantation du français comme langue de travail : vers un processus de changement linguistique planifié », dans Martin (André) et Loubier (Christiane) *L'implantation du français. Actualisation d'un changement linguistique planifié*, Québec, Office de la langue française, p. 59-133.

McAll (Christopher), Montgomery (Catherine) et Tremblay (Louise), 1994 : « Utilisation du langage et des langues au travail : la reconstruction de la journée de travail et la cartographie sociolinguistique d'entreprise », dans *Terminogramme*, n° 74, p. 1-7.

McAll (Christopher), 1992 : « Langues et silence : les travailleurs immigrés au Québec et la sociologie du langage », dans *Sociologie et société*, vol. XXIV, n° 2, p. 117-130.

Plourde (Michel), 1988 : *La politique linguistique du Québec 1977-1987*, Québec, Institut québécois de recherche sur la culture.

Teixeira (Carl A.), 1995 : *Le français écrit et la perception linguistique auprès de quatre entreprises. Une validation d'une mesure d'évaluation du degré d'utilisation du français écrit*, Rapport de recherche soumis à l'Office de la langue française.

## La terminologie en germanophonie : vent nouveau sur la collaboration

**S**i on essaie d'avoir une vue d'ensemble des initiatives lancées ces dernières années pour renforcer la coopération en matière de terminologie dans l'espace germanophone, on ne tarde pas à constater que la liste des institutions concernées est plus courte et l'agenda des réunions moins chargé que dans l'aire francophone. Une organisation internationale germanophone, qui se voue expressément à favoriser les échanges et la coopération par-dessus les frontières, à savoir le Conseil de la terminologie germanophone, CoTG (*Rat für Deutschsprachige Terminologie, RaDT*) n'existe que depuis un an.

On peut être surpris du retard avec lequel le Conseil de la terminologie germanophone<sup>(1)</sup> a été créé, si on compare la date de sa fondation à celle du Rint. Mais cela s'explique : ce n'est que depuis récemment que la terminologie est étudiée systématiquement dans un cadre suffisamment large en germanophonie en dehors de l'Allemagne, et les travaux de terminologie entrepris dans les autres pays germanophones ne pouvaient s'appuyer, il y a encore quelques années, que sur quelques relais à caractère surtout international, comme par exemple Infoterm; il est vrai que cette institution a toujours eu un champ d'activités mondial et

n'a jamais axé son travail plus spécialement sur l'allemand, mais ayant son siège à Vienne, ville germanophone où Wüster a enseigné jusqu'en 1974, elle a toujours eu des rapports particulièrement étroits avec les terminologues de langue allemande. Il ne faut également pas oublier le Service de terminologie de la Commission de l'Union européenne à Luxembourg, où la terminologie germanophone a eu une grande importance dès le début, ni la Conférence des services de traduction des États européens, dont le Groupe de travail Terminologie et documentation a publié les *Recommandations relatives à la terminologie*<sup>(2)</sup>. La conférence, qui encourage l'échange et la coopération entre les services qui lui sont affiliés, avait reconnu l'importance qu'a notamment la terminologie, avec la traduction, pour la communication entre spécialistes à travers les frontières; elle entendait établir une base pour la coopération en matière de terminologie par la publication de ses recommandations qui ont d'abord paru en allemand et se fondaient surtout sur des travaux en langue allemande.

Il n'en reste pas moins qu'il n'existait pas, en dehors de l'Allemagne, mis à part quelques exceptions, de partenaires engagés qui auraient travaillé à assurer la collaboration entre les institutions

(1) Les adresses des organismes cités sont fournies en annexe.

(2) *Recommandations relatives à la terminologie*. Berne : Chancellerie de la Confédération suisse; version allemande (1990/92); version française (1990); version anglaise (1990); version italienne, (1994); version espagnole, (1996 à paraître).

dans le secteur de la terminologie. Alors que l'on assistait à la fondation en Allemagne de sociétés telles que le *Deutscher Terminologie-Tag, DTT* (en 1985) et à celle de l'Association pour la terminologie et le transfert des connaissances (*Gesellschaft für Terminologie und Wissenstransfer, GTW*, en 1987), il n'existait pas dans les autres pays de langue allemande, d'organisation spécialisée pour défendre les intérêts de ce secteur et en coordonner les activités. Ce n'est qu'au début des années nonante que s'est constituée en Suisse une association lâche, issue d'une initiative privée, *Computerm*, à laquelle participent les institutions de terminologie qui étudient les quatre langues nationales. Le tableau que nous venons de brosser n'a guère changé, à l'exception de quelques améliorations minimales. Le rapport final concernant le projet de la Commission européenne connu sous la dénomination de *projet Pointer* fait d'ailleurs clairement ressortir la faiblesse de l'infrastructure des institutions de terminologie en germanophonie hors d'Allemagne ; le second pays de langue allemande, l'Autriche, n'y étant même pas mentionné. Chercher les raisons de cet état de choses peu encourageant permettrait certainement d'en déceler une : une politique linguistique incitative à la base de bien des projets de terminologie dans l'aire francophone fait défaut dans la plupart des régions de langue allemande.

Si une certaine collaboration a malgré tout pu se faire sur le plan terminologique en germanophonie, cela est dû d'une part aux associations professionnelles de traduction et aux institutions de formation, surtout les universités offrant une formation en traduction et, d'autre part, aux organisations mentionnées qui en organisant des rencontres d'information et d'étude, ont créé le cadre dans lequel des contacts utiles

et fructueux ont pu avoir lieu. C'est notamment le cas des colloques organisés depuis 1985 par le *Deutscher Terminologie-Tag*. Assidûment fréquentés par un grand nombre de personnes s'intéressant à la terminologie et venues de toute l'aire germanophone, entièrement axés sur la pratique et notamment sur l'échange d'informations et la communication des résultats de travaux, ces colloques du DTT sont la preuve que l'on recherche la collaboration et que les travaux de terminologie en langue allemande ont entre-temps atteint un niveau honorable. Par ses initiatives de ces dernières années, le DTT est devenu le véritable moteur de la coopération sur ce plan. Il a créé en 1995, avec l'Institut allemand de terminologie (*Deutsches Institut für Terminologie*), une Association pour la terminologie et le transfert des connaissances (*Dienstleistungs- und Kompetenzzentrum*) dont la circulaire multilingue et à caractère interrégional, publiée depuis deux ans, est une source particulièrement riche d'informations sur tous les aspects de la terminologie. Ce n'est pas pour rien que le mot *coopération* (*Kooperationen*) est écrit en grandes lettres sur cette publication du DTT (3).

On ne se trompe pas si on pense que la fondation du Conseil de la terminologie germanophone est due aux pionniers de cette collaboration que l'on trouve au DTT, à Infoterm et au GTW. C'est sur leur initiative que les commissions nationales de l'Unesco d'Allemagne et d'Autriche, auxquelles la commission nationale suisse s'est jointe après coup, ont lancé l'initiative en faveur de la terminologie germanophone. C'est donc de nouveau l'Unesco qui, après avoir favorisé la fondation d'Infoterm et l'avoir ensuite soutenu durant des

(3) DTT-Rundbrief, Nr. 4/1995, S. 3 und 35.

années, a encouragé un projet de collaboration en matière de terminologie, cette fois, il est vrai, par le biais des commissions nationales de ses membres de langue allemande. Les principaux objectifs de l'initiative, qui se fonde sur une décision de la 26<sup>e</sup> conférence générale de l'Unesco (de 1991) et qui insiste sur la nécessité d'une coopération en matière de terminologie, tant sur le plan national que sur le plan international, correspondent dans une large mesure, *mutatis mutandis*, aux objectifs du Rint.

Entre-temps, l'initiative a permis la fondation du Conseil de la terminologie germanophone CoTG qui s'est constitué le 14 novembre 1994 au siège de la commission nationale allemande pour l'Unesco, à Bonn. Ce conseil comprend quelque 25 membres; il s'agit d'experts représentant des organisations et des institutions de l'économie, de la science, de l'administration et de l'éducation de toute la germanophonie, qui traitent de questions de terminologie, de ses applications et de son enseignement(4). Le CoTG veut notamment :

- Stimuler l'intérêt pour la terminologie en germanophonie et favoriser la coopération dans ce domaine;
- Coordonner et soutenir les activités en matière de terminologie;
- Élaborer des lignes directrices pour une politique et une stratégie en matière de terminologie, les faire connaître et collaborer à leur mise en application(5).

Ainsi, la fondation du Conseil de la terminologie germanophone

(4) D'après *Geschäftsordnung des Rates für Deutschsprachige Terminologie*, § 2: *Aufgaben und Ziele*, version du 1<sup>er</sup> décembre 1995.

(5) *Idem*.

constitue un pas important vers la collaboration internationale dans le domaine de la terminologie germanophone, car presque toutes les régions de langue allemande d'Europe y sont représentées. La fondation de ce conseil est peut-être le pas décisif pour le développement de la terminologie, tout particulièrement pour les petites collectivités germanophones. En outre, on espère que le CoTG pourra tenir compte de la situation particulière des pays plurilingues et qu'il intégrera la terminologie plurilingue dans ses travaux. Il deviendrait ainsi l'interlocuteur des réseaux terminologiques existant dans d'autres aires linguistiques, ce qui pourrait donner une impulsion nouvelle à la collaboration internationale dans le domaine de la terminologie.

*Rolf Moos,  
Section de terminologie,  
Services linguistiques centraux,  
Chancellerie de la Confédération suisse,  
Berne.*

*Traduction : Rolf Haegler  
(Section française)*

## Adresses

Conseil de la terminologie  
germanophone/Rat für Deutschsprachige  
Terminologie  
Prof. Dr. K.-D.Schmitz, Fachhochschule  
Köln, Fachbereich Sprachen  
Mainzer Strasse 5, D-50678 Köln  
tél. : +49.221.8275.3272;  
télé. : +49.221.8275.3312  
Internet : kdschmitz@fh-koeln.de

Deutscher Terminologie-Tag  
Frau Ursula Reisen, An der Alten Post 18,  
D-50859 Köln  
tél. : +49.2234.76.826;  
télé. : +49.2234.76.826

Deutsches Institut für Terminologie  
Frau Deborah Fry, Feldbergstr. 28,  
D-55118 Mainz  
tél. : +49.6134.22.504;  
télé. : +49.6134.22.860  
Internet :  
100433,3552@compuserve.com

Computerm  
Jean-Pierre Neuhaus, chef du service  
linguistique, c/o Ascom SA  
Belpstrasse 37, CH-3000 Bern 14  
tél. : +41.31.999.27.71;  
télé. : +41.31.999.20.94

Association pour la terminologie et le  
transfert des connaissances

Gesellschaft für Terminologie und  
Wissenstransfer e. V.  
FR 8.6, Universität des Saarlandes,  
Postfach 15 11 50,  
D-66041 Saarbrücken  
tél. : +49.681.302.2929;  
télé. : +49.681.302.4440

Commission de l'Union européenne,  
SdT, Terminologie  
Bâtiment Jean Monnet,  
L-2920 Luxembourg

Conférence des services de traduction des  
États européens  
Groupe de travail terminologie et  
documentation  
Werner Hauck, président  
Section de langue allemande, Services  
linguistiques centraux, Chancellerie de la  
Confédération suisse,  
CH-3003 Berne  
tél. : +41.31.324.11.08;  
télé. : +41.31.324.11.02

Infoterm, Centre international  
d'Information pour la terminologie  
c/o Österreichisches Normungsinstitut,  
Heinestrasse 38, Postfach 130,  
A-1021 Wien  
tél. : +43.1.213.00.307;  
télé. : +43.1.216.32.72  
Internet: 100423.2307@compuserve.com

# Vers la planification linguistique du portugais : Portugal, Brésil, Afrique, quelques réflexions (1)

Du point de vue linguistique, la notion d'aménagement est directement liée à celle de jugement de valeur. Ces jugements sont calqués sur une idéologie politique et ils existent dans la nature prescriptible de la normalisation.

En élargissant le sens de *normalisation linguistique*, nous pouvons affirmer que cette expression implique des changements linguistiques. Les changements ne seront effectués que s'ils découlent d'un consensus entre les citoyens qui parlent cette langue et la politique linguistique institutionnalisée, puisque le changement linguistique est étroitement lié au changement social.

Dans ce contexte, nos objectifs sont :

- 1) De commenter la fonctionnalité du portugais dans les trois pays où il est la langue officielle;
- 2) D'analyser la situation du portugais dans le cadre politique, c'est-à-dire langue versus pouvoir;
- 3) De vérifier le processus de diffusion du portugais au sein de la Communauté européenne, qui développe, entre autres, une fonction d'institution normalisatrice des langues.

Enfin, en compilant des documents provenant de quelques commissions de la Communauté européenne, nous cherchons à déceler les interventions que la langue portugaise a subies pour s'ajuster au régime juridique communautaire.

Termes-clés :  
planification linguistique;  
portugais; lusophonie; Portugal; Brésil;  
Angola; Cap-Vert; Guinée-Bissau;  
Mozambique; São Tomé et Príncipe;  
Communauté européenne.

## 1 La notion d'aménagement linguistique

**L**e concept d'aménagement linguistique s'appuie sur un projet linguistique collectif. Il vise l'harmonisation linguistique et, pour réussir cette harmonisation, l'aménagement devra reposer sur un consensus social.

Il y a deux façons d'aménager les langues, selon les propres définitions d'aménagement. En premier lieu, l'*aménagement linguistique* se définit comme une intervention par laquelle l'État fait un «effort à moyen et à long terme pour mieux tirer parti d'une ressource collective, la ou les langues, en fonction des besoins et des intérêts de la nation» (Corbeil 1980 : 9). Dans ce sens, nous pouvons dire que le concept conduit une langue cible au processus d'harmonisation interne, dans un espace géographique déterminé. En ce qui concerne le portugais, c'est justement en Afrique lusophone que le processus d'aménagement trouve une grande répercussion, une fois qu'il y a une forte coexistence et moyenne concurrence entre le portugais - langue officielle - et plusieurs langues nationales, dans un même espace géographique.

Corbeil (1993 : 17) affirme : «lorsqu'on met en place un plan d'aménagement linguistique, on

provoque, dans les faits, la confrontation de deux séries de forces sociales, plus ou moins contraires, plus ou moins convergentes». Et il continue : «La première est constituée de l'ensemble des facteurs qui définissent la situation sociolinguistique des langues en présence. Ces facteurs ne jouent pas toujours en faveur de la même langue, d'où le phénomène de concurrence linguistique dont le résultat global est d'assurer à l'une des langues le statut de langue dominante.» Pour lui, il s'agit de forces sociales «constantes et à long terme». Nous pouvons illustrer ces mots avec la situation linguistique des pays africains où, à côté du portugais comme langue officielle, il y a une autre langue en concurrence, le créole. Ainsi, le Cap-Vert, la Guinée-Bissau et São Tomé et Príncipe ont développé un type de créole comme langue de communication interne.

La seconde série de facteurs a pour objectif de modifier l'essor de la concurrence en faveur de la langue

(1) Une synthèse de cet article, sous le titre *L'aménagement linguistique du portugais : Portugal, Brésil et Afrique* a été présentée au 62<sup>e</sup> congrès de l'Association canadienne française pour l'avancement des sciences (Acfas), du 16 au 20 mai 1994, Université du Québec à Montréal - Uqam, Montréal.

dominée et d'en favoriser l'usage pour des motifs sociaux connus et soutenus par les locuteurs de cette langue. «Les forces sociales de cette série sont soumises aux fluctuations des relations entre les groupes linguistiques en présence : les courants de l'opinion publique se modifient, ce qui peut, à terme, se traduire par des changements plus ou moins radicaux...» Dans ce cadre de changements, il convient de discuter les tentatives d'unifier le portugais écrit par les sept pays lusophones et aussi la création d'une Communauté des pays lusophones; ces deux actions sont dues à des initiatives du Brésil.

De même, pour les besoins de cet exposé, nous nous référerons à une deuxième notion d'aménagement linguistique. Celle-ci se situe sur un plan opérationnel et favorisera l'organisation de notre pensée sur la diffusion de la langue portugaise au sein de la Communauté européenne.

Labrie (1993 : 28), en citant Cooper (1989 : 45) qui a créé une définition de l'aménagement linguistique, explicite le besoin de réécrire cette définition, en ayant pour objet une définition opérationnelle pour l'étude de l'aménagement linguistique dans la Communauté européenne : «L'aménagement linguistique fait référence à des efforts délibérés visant à influencer, ou ayant pour effet d'influencer, le comportement des autres, en ce qui concerne l'acquisition, la structure et la répartition fonctionnelle de leurs codes linguistiques.»

Ce sont justement «ces efforts délibérés ayant pour effet d'influencer» qui nous intéressent ici, ceux qui concernent la langue portugaise, une des douze langues de la Communauté européenne.

## 2 La langue portugaise dans le monde

La langue portugaise est la langue officielle du Portugal, du

Brésil, de l'Angola, du Mozambique, de São Tomé et Príncipe, de la Guinée-Bissau et du Cap-Vert. Ces sont les «Sept» qui constituent le monde de la lusophonie. En considérant le nombre total de la population de tous ces pays, nous pouvons dire qu'il y a, approximativement, 190 millions d'habitants qui parlent le portugais. Au Portugal et au Brésil, le statut du portugais, langue officielle et nationale, ne cause pas de problèmes; il en est différemment dans les pays d'Afrique où l'officialisation de cette langue rencontre des difficultés à cause de la présence de plusieurs autres langues nationales.

## 3 La langue du Portugal

Le Portugal, province romaine qu'on appelait alors *Lusitanie*, d'où l'expression *lusophone* pour désigner ses locuteurs, a comme langue officielle le portugais. Situé en Europe, il compte 10,4 millions d'habitants qui parlent le portugais.

Pays à vocation conquérante, il a diffusé sa langue au gré de ses découvertes, sur plusieurs continents. Ainsi, en Asie, le portugais subsiste dans des noyaux de lusophones dans les anciens territoires indiens de Goa, Damão et Diu, à Malaka dans la presqu'île du même nom (Malaysia); à Timor, une des îles de la Sonde (Indonésie) et au Macao en Chine.

En Afrique, les cinq pays cités plus haut utilisent le portugais comme langue officielle. Mais c'est au Brésil, pays qui occupe la moitié de l'Amérique du Sud, que le portugais occupe la plus grande place. Langue officielle du pays, c'est aussi la langue maternelle de la presque totalité de ses habitants.

Cunha et Cintra (1985 : 9) déclarent que «Dans cette aire très étendue et discontinue dans laquelle le portugais est parlé, cette langue se comporte comme n'importe quelle

langue vivante : elle se divise en des variétés plus au moins éloignées l'une de l'autre en ce qui concerne la prononciation, la grammaire et le vocabulaire<sup>(2)</sup>.» Les auteurs disent encore que «le concept linguistique de norme, dans la langue portugaise, implique une notion de libéralisme grammatical». Et ils confirment ce qu'a déjà dit Serafim da Silva Neto : «unité n'est pas égalité»<sup>(3)</sup>.

Entre le portugais européen et le portugais sud-américain, il y a des variations propres, à cause de l'usage de la langue dans des contextes culturels différents. En observant le «droit à la différence», Teyssier (1983 : 50) dit qu'«il vaudrait la peine d'étudier de près les différences qui séparent la norme brésilienne de la norme portugaise. Ces différences touchent tous les aspects de la langue, mais à des degrés variables : très accentuées au niveau de la phonétique, encore assez importantes dans le vocabulaire et moins fortes sur le plan de la syntaxe. C'est à partir de cette réalité que devra être élaboré ce que nous pouvons appeler une politique de la langue.»<sup>(4)</sup>

(2) «Na área vastíssima e descontínua em que é falado, o português apresenta-se, como qualquer língua viva, internamente diferenciado em variedades que divergem de maneira mais ou menos acentuada quanto à pronúncia, à gramática e ao vocabulário.»

(3) «Unidade não é igualdade.»

(4) «Valeria a pena estudar de perto as diferenças que separam a norma brasileira da norma portuguesa. Essas diferenças abrangem todos os aspectos da língua, mas em graus variáveis : bastante acentuadas ao nível da fonética, ainda importantes no caso do vocabulário, são mais superficiais quanto à sintaxe. E é a partir desta realidade que deve ser elaborado o que podemos chamar uma política da língua.»

Quand nous parlons de la langue portugaise en Europe, nous ne devons pas oublier de nous référer à une certaine unité historique entre le portugais et le galicien, une des langues officielles de la Galice, à côté du castillan. Les contacts entre le galicien, au nord-ouest de l'Espagne, et le portugais ont produit une langue commune, le «galicien-portugais». Mais la frontière politique qui sépare la Galice du Portugal a divisé cet «idiome» et aujourd'hui, il ne reste plus que quelques possibilités d'intercommunication entre le peuple galicien et le peuple portugais.

La loi de normalisation, signée par le Parlement autonome de Galice, le 16 juin 1983, a émancipé le galicien, à côté du castillan. Vázques Cuesta (1983 : 121) se prononce, tout de même, ardemment : «Les relations typologiques et historiques existant entre le galicien et le portugais ont fait couler beaucoup d'encre en Galice ces dernières années. – Êtes-vous partisane ou non du «*lusismo*»? [...] Naturellement que je suis partisane du «*lusismo*», si, par là, on entend la conviction sérieuse que le galicien et le portugais d'aujourd'hui constituent des réalisations d'un même système linguistique [...]; qu'ils sont nés dans un même berceau - l'ancienne *Gallaecia romaine*, intégrée par notre terre et par la province lusitane d'Entre-Douro-et-Minho [...]»<sup>(5)</sup>.

(5) «As relacións tipolóxicas e históricas existentes entre galego e portugués tenen feito correr rios de tinta estes últimos anos na Galiza. «Vostede é ou non partidária do «*lusismo*»? [...] Naturalmente que eu son partidária do «*lusismo*», se se entender como tal a convicción non gratuita de que o galego e o portugués de hoxe constituem realizacións dun mesmo sistema lingüístico [...]; de que naceron no mesmo berce - a antiga Gaellaecia romana, integrada pola nosa terra e a provincia lusa de Entre-Douro-e-Minho [...]».

Malgré l'esprit d'identité, le portugais garde toujours ses particularités, en Europe, comme en Amérique du Sud. Voyons une déclaration au *Jornal de Letras*, sur un correcteur d'orthographe pour ordinateur, appelé *yeux de Lynx* : «C'est Carlos Amaral qui avance qu'il existe déjà des logiciels semblables au *Lynx* disponibles sur le marché portugais, mais que ce nouveau logiciel possède un avantage déterminant : il est réalisé vraiment en portugais, ou, comme on dit, en «portugais européen» et fait au Portugal; il n'a rien en commun avec des produits similaires originaires des États-Unis, dans un portugais brésilien.»<sup>(6)</sup>

## 4 Les langue de l'Afrique appelée *lusophone*

Avant de parler spécifiquement de chaque pays de l'Afrique lusophone, il convient de dire, en quelques mots, que l'indépendance politique de l'Afrique a réveillé un sentiment d'indépendance linguistique. Ainsi, l'aménagement du portugais provoque le développement systématique autant du créole que des langues nationales. À ce propos, Perl affirme (1989 : 26) :

«La mise en œuvre de la politique linguistique s'effectue d'abord indirectement dans les

(6) Article de Rodrigues da Silva paru dans *Jornal de Letras*, 29 mars 1994, p. 31. «E é Carlos Amaral que adianta existirem ja programas semelhantes ao Lince disponíveis no nosso mercado, só que este novo possui uma vantagem e determinante : é em português mesmo, ou, como soi dizer-se, em «português europeu» e feito cá, o que não tem nada a ver com produtos similares oriundos dos Estados Unidos, num português abasileirado.»

journaux et revues par des discussions dirigées qui portent sur le développement de la langue. Les particularités nationales y sont toujours jugées de manière positive et on assiste à une polémique dirigée contre la supériorité de certaines variantes linguistiques. [...] Il est cependant nécessaire que les divers pays d'expression portugaise trouvent un accord. Le développement et l'enseignement des registres stylistiques dépendent encore essentiellement de la maîtrise de la langue. C'est la raison pour laquelle la compétence communicative des locuteurs ne sera atteinte que peu à peu et sur la base d'une baisse du nombre des illettrés et de l'augmentation parallèle de la connaissance du portugais.»

### 4.1 Planification linguistique en Angola

La population d'Angola est estimée à 10 284 000 habitants<sup>(7)</sup>. La capitale du pays est Luanda et la langue officielle est le portugais. Mais il y a une cooccurrence entre le portugais et plusieurs langues nationales, nées au sein des groupes ethnolinguistiques. La cooccurrence se caractérise par le contact entre le portugais et une langue nationale courante, ce qui provoque l'apparition des unités parallèles dans le discours de l'une ou de l'autre langue.

Ainsi, à cause de sa pluriethnie et du multilinguisme, la République populaire d'Angola (RPA) a décidé la meilleure orientation «politico-linguistique» pour l'État. Elle a donc résolu de maintenir le portugais comme langue officielle et de faire

(7) Les estimations des nombres d'habitants des pays africains ont été extraites de l'*Almanaque Abril*, São Paulo, Editora Abril, 1993.

avancer les travaux de systématisation des langues nationales.

Endruschat (1989 : 71) commente qu'il (nous) semble, cependant, injustifié d'appeler le portugais langue officielle, langue d'État ou langue nationale d'Angola. En Angola on utilise toujours une autre désignation, langue véhiculaire (*língua veicular*) terme qui [à son avis] reflète de manière adéquate la fonction et le statut du portugais en Angola». Et elle dit de plus que «La politique de langues de la RPA a pour objet d'implanter dans la population un bilinguisme fonctionnel, c'est-à-dire de rendre chaque Angolais capable d'utiliser le portugais et sa langue maternelle, dans des situations de communication fixées respectivement.» (*op.cit.* :72.)

Les langues nationales d'Angola peuvent être classifiées en dix groupes ethnolinguistiques, selon la définition de Valdés Bernal (1990 : 45-49) :

1. Umbundu (le groupe majoritaire);
2. Kimbundu;
3. Kikoongo;
4. Ulunda-kichowe;
5. Ganguela;
6. Olunyaneka-lunkumbi;
7. Kwanyama;
8. Herero;
9. Khoisan ;
10. Manyengo.

Les efforts de planification des langues nationales en Angola ont déjà connu des résultats positifs. Pour les langues kikoongo, kimbundu, cokwe (du groupe ulunda-kichowe), umbundu, mbunda, et kwanyama, des projets d'alphabets ont été élaborés, à partir des descriptions des systèmes phonologiques de ces langues, et des ouvrages d'alphabétisation écrits en portugais ont été traduits en chacune des langues nationales.

Depuis que la RPA s'est engagée dans une voie de développement d'orientation socialiste, les conditions de vie ont changé et, ainsi, une communication plus efficace est

devenue indispensable. La fonction du portugais s'est transformée de langue de communication d'élite en langue véhiculaire pour tout un peuple. Le contact entre la langue officielle et les langues nationales a transformé le portugais en langue prêteuse, ainsi qu'en langue cible pour des emprunts. Le résultat est que les changements, sur le plan de la norme d'usage, donnent place à des innovations sur le plan du système de la langue.

## 4.2 Planification linguistique au Cap-Vert

La population du Cap-Vert est composée d'environ 341 000 habitants. La capitale du pays est Praia et la langue officielle est le portugais.

Au Cap-Vert, deux langues bien distinctes vivent en concurrence au milieu de la population scolarisée : la langue officielle, le portugais, et la langue maternelle, le créole, qui est considéré comme la vraie langue de communication. Le portugais, cependant, y est important à cause de trois types d'emploi : langue vernaculaire, celle de la population lettrée; langue régionale, variété qui a reçu beaucoup d'influences du créole; et langue rudimentaire, qui est utilisée en des moments déterminés.

Sous condition d'aménagement, la langue portugaise devient une langue utile au Cap-Vert, parce que, sur le plan politique, elle est reconnue comme une «langue de grande diffusion internationale»<sup>(8)</sup>. Elle est donc la langue du domaine de l'écriture, de l'école, et de l'administration : en somme, elle est la langue de la culture.

Les particularités géographiques des îles cap-verdiennes influencent directement le processus linguistique.

(8) Dulce Duarte, citée par Belmiro Ramos (1983 :230).

La République du Cap-Vert est constituée de 9 îles et de 5 petites îles qui se divisent en deux groupes, appelés Barlavento et Sotavento. Ces deux groupes géographiques divisent le créole aussi en deux variantes dialectales selon la dénomination des propres groupes. Mais, malgré toutes les difficultés économiques qui retardent l'implantation d'une politique linguistique bien planifiée, le pays ne refuse pas la langue du peuple colonisateur - le portugais - et défend en même temps le créole comme la langue de l'identité nationale. Pour réussir une planification linguistique, il faudrait élaborer un plan pédagogique qui permette, entre autres, une investigation linguistique. Cette investigation aura pour but la description de la langue créole sur le plan grammatical, lexicographique et littéraire.

## 4.3 Planification linguistique en Guinée-Bissau

La Guinée-Bissau compte près de 944 000 habitants. La capitale du pays est Bissau et la langue officielle est le portugais.

Dans ce pays de composition ethnique complexe, la situation linguistique refléchet aussi cette complexité. La problématique linguistique s'explique par la multiplicité des langues qui vivent en concurrence et en cooccurrence. La situation de bilinguisme fait partie de la pratique langagière du peuple guinéen.

Il y a un créole, assez proche du créole de Cap-Vert, qui fonctionne comme langue véhiculaire. À côté de cette langue et de la langue portugaise, existent plusieurs langues africaines qui se divisent en deux grands groupes : le groupe mandé et le groupe d'ouest-atlantique.

Au groupe mandé appartiennent le mandiga, une des langues les plus

importantes du pays, et le jacanga. Au groupe d'ouest-atlantique appartiennent, entre autres, le fula et le balanta. Ce dernier est une langue de grande diffusion, parlée par 39% de la population.

La politique linguistique de la Guinée-Bissau se développe en accord avec les intentions du *Partido Africano pela Independência de Guiné-Bissau e Cabo Verde* (PAIGC). Ce parti politique a reconnu que la diversité linguistique était une menace à l'unité du mouvement même : le PAIGC a opté pour le portugais comme langue officielle, puisqu'il pouvait fonctionner comme un moyen de communication interethnique. Le choix de cette langue se justifie par la nécessité de gagner du temps. Ainsi, la première action du gouvernement, après la conquête de l'indépendance du pays a été la fondation d'écoles dont l'enseignement se donne en portugais. Néanmoins, la Guinée-Bissau est un pays où l'oralité prédomine. Effectivement, le peuple transmet sa façon de penser, de sentir et de vivre, avant tout, oralement.

Le portugais est la langue utilisée dans les documents écrits. Voilà pourquoi il reste peu connu dans le pays. Comme l'affirme Henriques (1983 : 236) : «la population africaine de notre pays regarde le monde de façon très différente»<sup>(9)</sup> [par rapport aux Portugais].

La situation linguistique réelle de la Guinée-Bissau est la suivante : on parle le créole partout dans le pays, en milieu familial, entre amis et en public, c'est la véritable langue de communication; les langues nationales sont parlées en famille. Malgré son statut de langue officielle, le portugais continue à être une langue étrangère. C'est néanmoins la langue de communication et du savoir scientifique.

(9) «a população africana do nosso país vê o mundo um bocado diferente».

Cette position fragile du portugais, face à toutes les autres langues nationales et au créole guinéen, change le caractère du bilinguisme en Guinée-Bissau.

#### 4.4 Planification linguistique au Mozambique

Le Mozambique a environ 14 628 000 habitants. La capitale du pays est Maputo et la langue officielle est le portugais.

La décision de prendre le portugais comme langue officielle venait de la nécessité de trouver un instrument qui permît l'unité de la population auprès de la *Frente de Libertação de Moçambique* (Frelimo). Ainsi, le portugais surgit comme langue opérationnelle de consensus.

La grande famille linguistique au Mozambique est le bantou. Ce grand groupe comprend de vingt-cinq à trente langues qui se divisent, elles-mêmes, encore en plusieurs dialectes. Leiste (1989 : 59), en citant Rzewuski (1978), présente une typologie comprenant neuf groupes linguistiques :

1. Groupe *makua-lomwe* (langues makua, lomwe, ngulu, koti, cwabo);
2. Groupe *tswa-ronga* (langues tswa, tsonga, ronga);
3. Groupe *nsenga-sena* (langues kunda, nsenga, nyungere, sena);
4. Groupe *shona* (langue shona);
5. Groupe *nyanja* (langue nyanja);
6. Groupe *copi* (langues copi et gitonga);
7. Groupe *yao* (langues yao et makonde);
8. Groupe *swahili* (langues swahili et mwani);
9. Groupe *nguni* (langues zulu et swazi).

À côté de l'usage de près de 30 langues maternelles, le portugais a gagné le statut de langue seconde. Cela implique le développement d'un bilinguisme au Mozambique. Néanmoins, en ce qui se réfère à

l'enseignement du portugais aux étudiants qui maîtrisent bien une langue nationale, le problème qui se pose est de savoir quelle est la norme du portugais qui devra être enseignée. Celle du portugais de Portugal? Celle du portugais mozambicain? Dans ce dernier cas, quelle est la norme du portugais du Mozambique, si elle n'est pas définie?

La politique linguistique entreprise par le Frelimo se réalise au niveau national, avec des objectifs bien définis. L'un de ces objectifs est la création d'un bilinguisme fonctionnel, c'est-à-dire que les langues nationales devront jouer un rôle de même valeur que le portugais dans l'enseignement. En 1983, Gonçalves reconnaît que, au Mozambique, de nombreuses perspectives d'investigation s'ouvrent sur le plan de l'analyse linguistique, mais qu'il y a un manque de ressources techniques et humaines (Gonçalves 1983 : 246).<sup>(10)</sup> Elle justifie ensuite que, malgré la problématique du bilinguisme, il n'existait pas encore de conditions pour le développement d'une étude contrastée et systématique entre le portugais et une langue bantoue (Gonçalves 1983 : 247).<sup>(11)</sup>

Effectivement, le portugais fonctionne comme le moyen de communication supraethnique entre les langues nationales mozambicaines.

(10) «À análise linguística abrem-se inúmeras perspectivas de investigação que não estamos ainda em condições de empreender por falta de recursos técnicos e humanos disponíveis.» (Gonçalves 1983 : 246.)

(11) «apesar da problemática do bilinguismo, não houve ainda condições para um estudo contrastivo sistemático do Português relativamente a, pelo menos, uma língua bantu.» (Gonçalves 1983 : 247.)

## 4.5 Planification linguistique à São Tomé et Príncipe

São Tomé et Príncipe sont deux îles qui comptent environ 123 000 habitants. La capitale du pays est Saint-Thomas et la langue officielle est le portugais.

À côté de la langue officielle, il y a deux langues créoles - le forro et le monco - et un dialecte appelé *angolar*. Ces langues et le dialecte fonctionnent en cooccurrence. Le créole forro est une langue pratiquée dans les milieux littéraires, mais il ne possède pas de description grammaticale normative ni de dictionnaire. L'absence de ces documentations donne une place à la diffusion du portugais dans les divers domaines de la vie são toméenne.

À ce propos, Espírito Santo (1983 : 258) déclare que «Par le contact avec de nouvelles réalités et idées, (la langue portugaise) se garnit d'un nouveau lexique varié, et elle met en valeur toutes les altérations éprouvées sur le plan de la grammaire. Cela signifie que les actes linguistiques, directement liés à des impressions et à des expériences, sont des actes de création.<sup>(12)</sup> De même, le créole a subi des influences du portugais et du créole cap-verdien aussi.

La politique linguistique de São Tomé et Príncipe donne au portugais le rôle de langue d'alphabétisation qui se développe dans le pays, dès lors qu'il y est le principal véhicule de communication.

(12) «Em contacto com as novas realidades e ideias (a língua portuguesa) apetrecha-se de un novo e variado léxico realçando-se as profundas alterações que sofre a nível da gramática, o que significa que os actos linguísticos, por estarem directamente vinculados às impressões e vivências, são actos de criação.»

## 5 La langue du Brésil

La dernière constitution brésilienne, dénommée *Nouvelle Constitution du Brésil (Nova Constituição do Brasil)* et promulguée le 5 octobre 1988, réaffirme : «La langue portugaise est l'idiome officiel de la République fédérative du Brésil» (Titre II, chap. III, art.13).

Aujourd'hui, le Brésil est considéré comme la mère patrie de la langue portugaise, avec une population de 146 154 502 habitants; c'est un pays composé de 26 États fédérés, où tout le monde parle la même langue. Nous pouvons affirmer qu'il s'agit d'un pays monolingue.

Même en considérant le «caractère monolingue» du Brésil, il faut observer la pluralité linguistique courante en Amazonie à cause des langues amérindiennes.

Selon Rodrigues (1992 : 367) : «Même avec la réduction à 250 langues, l'Amazonie représente encore une des plus importantes réserves de variété linguistique au monde, surtout à cause de la grande diversité génétique qui la caractérise.<sup>(13)</sup>»

L'Amazonie occupe une énorme surface au nord de l'Amérique du Sud. Là sont reconnues «19 familles génétiques» (*ibid.*) qui, ayant de nombreux groupes au Brésil, se divisent entre les pays voisins, à savoir la Bolivie, le Pérou, L'Équateur, la Colombie, le Venezuela, la Guyane, le Surinam et la Guyane française. Après avoir fait des études sur les espaces dépeuplés, Rodrigues remarque que «La quantité de langues autochtones antérieure à la pénétration européenne était d'environ 500

(13) «mesmo com a redução a 250 línguas, a Amazônia ainda representa uma das mais importantes reservas de variedade linguística do mundo, sobretudo devido à grande diversidade genética que a caracteriza.»

langues.<sup>(14)</sup>» (*ibid.*) Aujourd'hui la situation peut être considérée comme grave, une fois que la diversité linguistique se montre très réduite.

Dans cet article de Rodrigues, nous trouvons des préoccupations qui pourront coïncider avec un projet de planification pour «l'étude, la connaissance, la préservation et le développement des langues autochtones de l'Amazonie». Il affirme : «Il est nécessaire et urgent que se créent, dans les pays d'Amazonie, des centres de recherche qui puissent recevoir des équipes de linguistes agissant par périodes variables, avec des ressources pour l'investigation et pour le travail sur place et avec la participation de membres des communautés indigènes.<sup>(15)</sup>» (*op.cit.* : 369.)

Nous notons ces faits parce qu'ils permettent de reprendre quelques remarques sur l'histoire de la langue du Brésil. Jusqu'au XVIII<sup>e</sup> siècle, le parler brésilien est demeuré fidèle à la prononciation du Portugal, mais la langue a subi des influences amérindiennes et africaines qui se font sentir avant tout sur le plan du lexique. Plus tard, l'immigration européenne, installée au Centre-Sud donne aussi son empreinte au portugais. Donc, éloignée du Portugal, la langue du Brésil n'a pas accompagné les changements du parler de Lisbonne. Cela explique les différentes modalités phonétiques qui se déroulent à l'intérieur de ce vaste territoire, si bien que la langue garde

(14) «a quantidade de línguas nativas anterior à penetração europeia [era] cerca de 500».

(15) «é necessário e urgente que se criem nos países amazônicos centros de pesquisa que possam manter equipes de linguistas atuando por prazos de extensão variável, com recursos para trabalho de campo e trabalho no centro, inclusive com a participação de membros das comunidades indígenas.»

une singulière unité. C'est l'unité dans la variété.

En 1922, un mouvement littéraire, appelé *Semaine d'art moderne* (*Semana de arte moderna*) avait pour objectif de briser les modèles traditionnels du Portugal et de privilégier les singularités du parler brésilien.

En 1958, le projet de Nomenclature grammaticale brésilienne - *NGB* (*Nomenclatura Gramatical Brasileira*) d'une commission formée par des philologues et des linguistes brésiliens, est né et est entré en vigueur en 1959. L'arrêté du ministre d'État de l'Éducation et de la culture, datant du 28 janvier 1959, recommande, dans son article premier, «l'adoption de la Nomenclature grammaticale brésilienne dans l'enseignement de la langue portugaise et dans les activités qui visent l'apprentissage, etc.».

En 1971, la première édition du *Nouveau dictionnaire Aurélio de la langue portugaise* (*Novo Dicionário Aurélio da Língua Portuguesa*) est publié. Le lexicographe Aurélio Buarque de Hollanda Ferreira et une équipe de collaborateurs ont écrit la principale œuvre lexicographique de la langue portugaise courante de ces dernières années. Ce dictionnaire en est maintenant à sa deuxième édition, actualisée, révisée et amplifiée du tiers de ses articles. Il est aussi disponible en disque optique.

En revenant à la question «d'une même langue», nous devons reconnaître que l'adoption de la Nomenclature grammaticale brésilienne, autant que l'édition du dictionnaire *Aurélio*, fixent l'utilisation d'une norme propre du portugais au Brésil. C'est le «standard» du portugais sud-américain qui prévaudra dans tous les milieux de communication.

Cependant, il existe des efforts pour que les normes du Brésil et du Portugal, ainsi que l'usage,

maintiennent une proximité normative, même divisées entre deux cultures très différentes.

Ainsi, en 1984, est publiée au Portugal la *Nouvelle grammaire du portugais contemporain* (*Nova Gramática do Português Contemporâneo*), écrite par les philologues Celso Cunha, du Brésil, et Luis Felipe Lindley Cintra, du Portugal. Cette grammaire qui avait promis de changer le sens courant de norme, de correct et d'incorrect n'est rien de plus qu'encore une grammaire traditionnelle, avec en plus des observations sur la différence entre la variété du Portugal et celle du Brésil. Cette œuvre est sortie au Brésil en 1985 seulement.

Néanmoins, au cœur du mouvement pour une unité linguistique, il y a une divergence qui montre que, entre le Portugal et le Brésil, les relations linguistiques ne sont pas toujours harmonisées. Un des principaux problèmes de la normalisation linguistique se situe sur le plan de l'orthographe : il s'agit des accords ou «désaccords» orthographiques qui sont survenus au cours de l'histoire de la langue portugaise.

Bref, dès le XVII<sup>e</sup> siècle, on tente l'unification des orthographes portugaise et brésilienne. Mais c'est seulement en 1911 que le gouvernement du Portugal a approuvé un projet de réforme entrepris en 1904. Cette nouvelle orthographe arrive au Brésil en 1931 et est refoulée par la Constitution brésilienne de 1934. Puis intervient un nouvel accord entre les deux pays : le projet est revigoré au Brésil en 1943. Le Portugal a adopté l'orthographe de ce dernier accord déjà en 1945; mais le Brésil en adopte une autre.

Vers 1975, l'Académie des sciences de Lisbonne et l'Académie brésilienne des lettres ont conclu un «accord de principe» pour l'unification orthographique de la

langue portugaise, inspiré des travaux d'Antonio Houaiss, lexicographe, philologue et professeur brésilien. Coordonnateur de tous les travaux, Houaiss a obtenu comme résultat le Projet d'orthographe unifiée (*Projeto de Ortografia Unificada*). Il s'agit d'un accord signé, le 16 décembre 1990, par les ministres de l'Éducation et de la culture des sept pays d'expression portugaise : Angola, Brésil, Cap-Vert, Guinée-Bissau, Mozambique, São Tomé et Príncipe et Portugal.

Avant la divulgation officielle du Projet, le Portugal a protesté contre son contenu. Mais le 4 juin 1991, le Parlement portugais a ratifié le projet et, après que la commission chargée de la rédaction l'eut rendu officiel, le gouvernement du Portugal l'a signé, espérant une réponse positive du Brésil. Selon l'accord entre les sept pays, la «réforme» devait avoir force de loi en 1994. Mais jusqu'à aujourd'hui, le Brésil ne l'a pas entériné. En défense de l'unification, Houaiss (1994 : 29) a récemment écrit : «La lusophonie, c'est l'unique phonie où il y a, avec le même alphabet, deux orthographes officielles : - la portugaise, acceptée par les pays africains qui ont comme langue officielle le portugais, - et la brésilienne, dans une situation qui a contribué à la méconnaissance progressive mutuelle des intégrants du complexe linguistique et culturel lusophonique.<sup>(16)</sup>»

Il existe toutefois des efforts pour que la langue du Portugal, du Brésil et d'Afrique forment une unité dans

(16) «A lusofonia é a única fonia que, com o mesmo alfabeto, tem duas ortografias oficiais - a portuguesa, aceita pelos países africanos de língua oficial portuguesa, e a brasileira - numa situação que tem contribuído para o progressivo desconhecimento recíproco dos integrantes do complexo linguístico e cultural lusofónico.»

le monde de la lusophonie. Pour cela, le Brésil a proposé la formation d'une Communauté des pays de langue portugaise. Nous aborderons ce sujet au prochain point.

## 6 La Communauté des pays de langue portugaise

La création de la Communauté des pays de langue portugaise a fait l'objet de nombreuses réunions au Brésil, au Portugal et en Afrique. Ayant comme principal animateur l'ambassadeur du Brésil au Portugal, José Aparecido de Oliveira, qui pontifie sur l'importance du projet, par rapport à la formation de la Communauté européenne, depuis le 1<sup>er</sup> janvier 1994 : «Les pays s'organisent en blocs par définition géographique ou par intérêts économiques. Pourquoi ne pas créer un bloc par affinités socio-culturelles?(17)» Nous pouvons spéculer en disant que cette Communauté sort d'une proposition politico-diplomatique pour la promotion des associations de management et de projets socio-économiques. À ce propos, nous ajoutons l'opinion d'Oliveira : «L'idée est de consolider une coopération culturelle, autant que d'assurer une présence politique.(18)»

À grands traits, nous allons poser les jalons de la Communauté(19).

(17) Article paru dans le *Jornal do Brasil*, 26 octobre 1993. p. 7 : «Os países estão se organizando em blocos por definição geográfica ou interesses econômicos. Por que não criar um bloco por afinidades sócio-culturais?»

(18) *Ibid.* : «A idéia é consolidar uma cooperação cultural e assegurar a presença política.»

(19) La plupart des informations sont extraites du *Jornal de Letras (JL)*, de Lisboa.

Le 6 décembre 1993, se déroule à Lisbonne un débat sur la constitution de la Communauté des pays de langue portugaise. Les principaux sujets traités ont été :

1. Le parlement des peuples de langue portugaise et les perspectives de coopération politique et diplomatique intergouvernementale.

Quelques résultats du débat suscitent des mots d'enthousiasme de la députée portugaise Manuela Aguiar qui dit que cette Communauté «est tout ce que la Communauté européenne ne peut être. Celle-ci est une fatalité géographique, l'autre doit être une communauté de la langue, du sang et de tendresse(20).»

Également enthousiaste, le député angolais João Melo parle avec emphase «du développement de chacun des pays africains(21) et du rôle que le Portugal et le Brésil pourront assumer dans ce contexte spécifique.

Ensuite, le journaliste brésilien Hermano Alves, en opposition aux autres, a posé des questions relatives à l'institution du Parlement. Il dit que le Portugal, jusqu'à maintenant, a fait des restrictions à l'entrée des étrangers au pays, ce qui touche le citoyen brésilien(22).

2. L'institut international de la langue portugaise comme instrument de promotion et de diffusion de l'idiome.

Sans perdre de vue la condition de pauvreté de l'Afrique, le journaliste mozambicain Hilário Matusse

(20) Dans *JL*, 14 décembre 1993. p. 15 : «esta Comunidade é tudo que o que a Comunidade Europeia não pode ser. Esta é uma fatalidade geográfica, a outra há-de ser uma comunidade de língua, sangue, afecto».

(21) *Ibid.* : «o desenvolvimento de cada um dos países africanos».

(22) *Ibid.*

attribue au Portugal et au Brésil l'autonomie du projet, en affirmant que «les frères plus âgés pourront assumer le rôle déterminant, dans les sphères de la formation et des investissements»(23). Une autre personnalité africaine de São Tomé et Príncipe, Armino Vaz de Almeida, a cherché le consensus, en déclarant que «C'est au milieu de la Société civile qu'on vit l'esprit de communauté.(24)»

3. L'Université des sept et la coopération pour le développement.

Sur ce thème, le député David Hopffer Almada du Cap-Vert a dit que l'objectif est de «former du personnel qualifié»(25). Comme ex-ministre de la culture de son pays, il se garde des projets mégalomaniaques. Ensuite, sur le même thème, le professeur et historien brésilien Caio Boschi opine : à son avis, l'Université des sept devra assumer «une structure flexible et dynamique parce que l'histoire est faite avec les pays et non pour les pays»(26).

4. Le dernier sujet traité est l'Institutionnalisation et la structure de la Communauté des pays de langue portugaise.

Tous les participants de cette table ronde ont ressenti le besoin d'une institution qui soit ouverte au monde de la Communauté, sans hégémonie, sans rivalités, et avec des

(23) *Ibid.* : «irmãos mais velhos poderão assumir papel determinante, nos âmbitos da formação e dos investimentos».

(24) *Op. cit.*, p. 16 : «É na sociedade civil que se tem de viver o espírito de comunidade.»

(25) *Ibid.* : «O objetivo é formar quadros capazes».

(26) *Ibid.* : «uma estrutura não rígida e dinâmica porque a História também o é, feita com os países e não para os países».

articulations. Un autre aspect qui a mérité beaucoup d'attention et de sollicitations est la réunion avec les chefs d'État des sept pays pour signer le traité d'institutionnalisation de la Communauté. Cette réunion aurait dû se dérouler la dernière semaine de juin, à Lisbonne. Comme l'a bien remarqué José Carlos de Vasconcelos, directeur du *Jornal de Letras*, demain il faudra «passer des beaux mots aux actes»<sup>(27)</sup>.

Dans l'optique de José Aparecido de Oliveira, «le monde qui arrive sera un monde dans lequel se confirment des identités ethniques; en même temps, on définit les grandes places pour l'action politique et économique. Tous profitent des avantages du voisinage, de la continuité géographique [...]. Dans notre cas, nous pouvons affirmer notre vigueur sur la scène internationale par l'usage d'un même portugais en Europe, en Afrique et en Amérique.»<sup>(28)</sup>

## 7 Quelques remarques sur la langue portugaise dans la Communauté européenne

La langue portugaise est une des douze langues de la Communauté européenne (CE). Nous savons que

(27) *Op. cit.*, p. 15 : «passar das belas palavras aos actos».

(28) *Op. cit.*, p. 17 : «o mundo que chega será um mundo no qual se confirmam os traços da identidade étnica, mas, ao mesmo tempo, estão sendo definidos os grandes espaços de acção política e econômica. Todos aproveitam as vantagens da vizinhança, da continuidade geográfica [...]. No nosso caso, estamos afirmando nos mapas comunitários a nossa força na cena internacional pelo uso de uma mesma língua na Europa, na África e na América.»

l'entrée du portugais au sein de la Communauté est due au Portugal qui fait partie, depuis le 1<sup>er</sup> janvier 1986, de ce marché unique.

Ainsi, au cours de notre article, nous présentons les perspectives d'une unification de la langue, considérée comme la même au Portugal, au Brésil et en Afrique. À cet égard, il convient de s'interroger si l'usage du portugais européen au sein de la Communauté contribuera à le rapprocher de la norme brésilienne ou si, au contraire, cet usage servira à éloigner de nouveau la norme européenne de la norme sud-américaine et de l'usage africain.

Dans un premier temps, notre propos sera de réfléchir sur cette question, en observant le principe d'aménagement courant dans la Communauté européenne. Ensuite, nous ferons quelques remarques sur l'aménagement du portugais.

En revenant à Labrie (1993), cité au début de cet exposé, nous pouvons rappeler le sens de sa définition de l'aménagement linguistique, en mettant l'accent sur le but de cet aménagement où il existe des efforts délibérés qui atteindront directement le comportement linguistique des autres. Or, pour aménager une langue, elle doit avoir été normalisée, parce qu'une telle décision exerce un impact, même indirect sur la structure de la langue.

Dès lors que la normalisation linguistique est «destinée à modeler le comportement langagier des locuteurs d'une langue» (Auger 1993), il est juste que des interventions humaines conscientes existent; que ce soit une intervention sur le code de la langue, qui touchera le système d'écriture ou même l'alphabet, la rédaction des grammaires, des dictionnaires, des manuels d'enseignement, des œuvres littéraires, scientifiques ou techniques etc., ou une intervention, au regard d'Auger (1993), strictement «limitée aux vocabulaires de spécialité pour une langue donnée».

Or, la mise en place de 12 langues officielles dans la CE a forgé un multilinguisme par lequel s'est développée la politique du pluralisme linguistique. Ainsi, tout acte juridique communautaire est publié dans chaque langue officielle des États membres, surtout pour des raisons juridiques.

La CE traite les douze langues sur un plan d'égalité. C'est ce qui explique pourquoi la traduction, l'interprétation et la terminologie sont des outils fondamentaux. Dans ce cadre, le travail linguistique de la CE se structure sur les langues normalisées, et la normalisation ne s'effectue pas isolément. C'est en vue d'une concertation en matière d'aménagement linguistique que le Groupe interinstitutionnel de terminologie-documentation (GIIT) a été créé.

Afin de ne pas prolonger ce sujet par des généralisations, nous parlerons brièvement du portugais dans la CE. Nous citerons des exemples appropriés.

Un article publié en 1992, sur *L'emploi des capitales, des majuscules et des minuscules, dans les actes établis dans les neuf langues officielles des institutions communautaires*, présente des commentaires intéressants sur l'aménagement de l'orthographe. L'auteur, Gérard Losson, (1992 : 154) dit que «Dans le creuset communautaire, caractérisé dans tous les domaines par un effort continu de rapprochement et d'unification, les langues constituent cependant une exception dans la mesure où tout est mis en œuvre pour sauvegarder la diversité linguistique, considérée comme une richesse exceptionnelle de même qu'un moyen de respecter la liberté d'expression des peuples, par ailleurs prêts à vivre politiquement et économiquement en communauté.»

Losson a poursuivi une étude comparative «pour faire ressortir les convergences et les divergences marquant les langues officielles des

institutions sur un point spécifique» - l'emploi des capitales, des majuscules et des minuscules. Comme résultat, l'auteur obtient que les langues peuvent se diviser en trois grands groupes. L'allemand, tout seul, forme un groupe à part. Comme on sait, l'emploi des majuscules a un rôle véritablement grammatical, «consistant à distinguer méthodiquement par une majuscule les substantifs (noms communs et noms propres) de tous les autres mots» (*op. cit.* : 163). Le deuxième groupe comprend «le danois et le grec, pour lesquels des commissions nationales ont fixé les règles à suivre» (*op. cit.* : 167). Dans le dernier groupe «se rangent les autres langues, à savoir celles pour qui la question des majuscules et des minuscules est régie par la tradition et... par la pratique «imposée» par les rédacteurs de textes».

Et il continue : «Ce groupe peut se subdiviser à son tour en trois sous-groupes : le premier inclut le français et l'italien, qui tendent tous deux à limiter autant que possible l'emploi de la majuscule aux noms propres et aux dénominations assimilées à ces derniers. Le français apparaît comme le plus économe en majuscules (on observe une recrudescence de majuscules, souvent sous l'influence de l'anglais et de rédacteurs dont la langue maternelle n'est pas le français). Le deuxième sous-groupe réunit plus ou moins l'anglais, le néerlandais et le portugais, tous marqués par une longue tradition fidèlement suivie. L'espagnol constitue le troisième sous-groupe; il écrit les dénominations de pays et d'actes avec une majuscule pour les substantifs et les adjectifs, mais applique, comme le danois, la minuscule aux noms de populations.» (*op. cit.* : 167-168.)

Finalement, pour confirmer notre remarque sur le fait qu'il y a des interventions dans les langues européennes, Losson maintient :

«C'est pour circonscrire les méfaits de la «majusculite», au moins dans les actes officiels, que le formulaire des actes du Conseil a édicté un certain nombre de règles minimales. Et, grâce au filtre que constitue la phase dite de la mise au point juridique et linguistique des actes, qui précède directement leur adoption, ainsi que le rôle modérateur joué par le *Journal officiel des Communautés européennes*, il est possible d'assurer une relative uniformité de la pratique en la matière.» (*op. cit.* : 170.)

En lisant les *Règles d'emploi des majuscules et des minuscules en langue portugaise*, transcrites par Losson (1992 : 236-243), nous observons que ce sujet a reçu un traitement très minutieux dans la CE. Pour cette sorte de normalisation, l'auteur présente dix règles d'emploi de minuscules, avec 21 subdivisions, suivies d'exemples, et dix règles d'emploi des majuscules, avec 20 subdivisions et aussi des exemples. Il s'agit d'un travail de normalisation fait exclusivement par la Communauté, puisqu'aucune grammaire normative de la langue portugaise n'a porté une semblable attention à ce sujet.

## 8 Conclusion

Les mondes lusophones européen et sud-américain, sont des mondes qui ne subissent pas la menace d'un bilinguisme institué. Le Portugal diffuse sa langue au sein de l'Europe et il apprend d'autres langues à travers le projet *Lingua* de la Communauté européenne. Le Brésil suit un parcours inédit, en découvrant les pays hispanophones d'Amérique du Sud.

En profitant de leur voisinage, le Portugal est aujourd'hui un des douze membres de la CE; le Brésil marche en direction du Mercosur, là où la langue portugaise aura comme

partenaire l'espagnol. Ce marché du Sud exigera une nouvelle interprétation de l'aménagement linguistique du portugais.

Placé dans des univers linguistiques et culturels différents, le portugais du Brésil n'est pas le même que celui d'Europe; l'espagnol de l'Amérique du Sud n'est pas le même que celui d'Europe. Afin que le dialogue des partenaires s'effectue de façon satisfaisante, il faut que les principes d'aménagement linguistique soient appliqués minutieusement.

En Afrique lusophone, le statut du portugais comme langue officielle mérite une réflexion. Il faut établir un projet sociolinguistique où, à côté d'une planification pour la diffusion du portugais, on décrive aussi une planification pour les langues nationales. De plus, un plan d'aménagement des langues d'Afrique lusophone ne doit pas ignorer non plus le prestige du français et de l'anglais dans les milieux scolaires et sociaux comme langues de communication internationales, une place occupée maintenant par le portugais.

*Enilde Faulstich,  
Département de linguistique,  
langues classiques et vernaculaires,  
Institut des lettres,  
Université de Brasília,  
Brésil.*

## Bibliographie

Auger (Pierre), 1993 : *Norme - Normalisation - Normalisation terminologique*, 12 p. (Notes de cours).

Auger (Pierre), 1993 : *Présentation de l'aménagement terminologique*, 9 p. (Notes de cours).

Corbeil (Jean-Claude), 1980 : *L'aménagement linguistique du Québec*, Montréal, Guérin (Langue et société).

Corbeil (Jean-Claude), 1993 : «Dynamique de l'aménagement linguistique au Québec», dans *Actes du colloque sur la problématique de l'aménagement linguistique (enjeux théoriques et pratiques), 5-7 mai 1993*, Université du Québec à Chicoutimi, Québec, Office de la langue française & Université du Québec à Chicoutimi, p. 17-33.

Cunha (Celso) & Lindley Cintra (Luis Felipe), 1985 : *Nova gramática do português contemporâneo*, Rio de Janeiro, Nova Fronteira.

Endruschat (Annette), 1989 : «Création lexicale en portugais parlé dans la République populaire d'Angola », dans *La langue portugaise en Afrique, Études portugaises et brésiliennes. Travaux du Centre d'études hispaniques, hispano-américaines, portugaises, brésiliennes et de l'Afrique d'expression portugaise*, Rennes, Université de Haute Bretagne, p 69-86.

Espírito Santo (Carlos), 1983 : «Situação actual da Língua Portuguesa em Cabo Verde», dans *Congresso sobre a situação actual da Língua Portuguesa no Mundo, Actas*, 2<sup>e</sup> éd., Lisboa, Dialogo, vol. 1, p. 253-263.

Gonçalves (Perpétua), 1983 : «Situação actual da Língua Portuguesa em Moçambique», dans *Congresso sobre a situação actual da Língua Portuguesa no Mundo, Actas*, 2<sup>e</sup> éd., Lisboa, Dialogo, vol. 1, p. 243-252.

Henriques (Maria Augusta), 1983 : « Situação e perspectivas do Português na Guiné-Bissau », dans *Congresso sobre a situação actual da Língua Portuguesa no Mundo, Actas*, 2<sup>e</sup> éd., Lisboa, Dialogo, vol 1, p. 234-242.

Houaiss (Antonio), 1994 : «Decálogo do Acordo Ortográfico Unificado», dans *Jornal de Letras Ideias*, Lisboa, 27 avril 1994, p. 29.

Labrie (Normand), 1993 : *La construction linguistique de la Communauté européenne*, Paris, Honoré Champion.

Leiste (Doris), 1989 : «Quelques réflexions sur la situation linguistique en République populaire du Mozambique», dans *La langue portugaise en Afrique, Études portugaises et brésiliennes. Travaux du Centre d'études hispaniques, hispano-américaines, portugaises, brésiliennes et de l'Afrique d'expression portugaise*, Rennes, Université de Haute Bretagne, p. 55-67.

Losson (Gérard), 1992 : «L'emploi des capitales, des majuscules et des minuscules dans les actes établis dans les neuf langues officielles des institutions communautaires », dans *Terminologie & traduction*, 1992 (1), p. 153sv.

Perl (Matthias), 1989 : «Le portugais et le créole portugais en Afrique», dans *La langue portugaise en Afrique, Études portugaises et brésiliennes. Travaux du Centre d'études hispaniques, hispano-américaines, portugaises, brésiliennes et de l'Afrique d'expression portugaise*, Rennes, Université de Haute Bretagne, p. 9-27.

Ramos (Belmiro), 1983 : «Situação actual da Língua Portuguesa em Cabo Verde», dans *Congresso sobre a situação actual da Língua Portuguesa no Mundo, Actas*, 2<sup>e</sup> éd., Lisboa, Dialogo, vol. 1, p. 230.

Rodrigues (Aryon Dall'Igna), 1992 : «Diversidade linguística na Amazônia», dans *Seminário Internacional sobre Meio Ambiente, Pobreza e Desenvolvimento da Amazônia - Simdamazônia : Anais*, Brasil, Belém, Propeda, p. 367-370.

Teyssier (Paul), 1983 : «Lição final», dans *Congresso sobre a situação actual da Língua Portuguesa no Mundo, Actas*, 2<sup>e</sup> éd., Lisboa, Dialogo, vol. 1, p. 45-55.

Valdés Bernal (Sergio), 1990 : *Visión lingüística del Africa al sur del Sahara*, La Habana, Editorial de Ciencias Sociales.

Vazques Cuesta (Pilar), 1983 : «Situacion e perspectivas do Galego», dans *Congresso sobre a situação actual da Língua Portuguesa no Mundo, Actas*, 2<sup>e</sup> éd., Lisboa, Dialogo, vol. 1, p. 109-135.

# Développement harmonisé et distribution des ressources terminologiques: le Réseau panlatin de terminologie – Realiter

**A**u mois de novembre 1991, lors d'un Colloque national pour la terminologie organisé par l'Union latine (Rome, Palazzo Barberini), naissait en Italie l'Associazione Italiana per la Terminologia (Assiterm). L'un des principaux buts de cette association était de relier entre eux les différents secteurs intéressés par l'étude et la diffusion des langues de spécialité : l'administration publique, l'université et les organismes de recherche, l'édition, les secteurs économiques et financiers, les moyens de communication, le monde de l'entreprise et de l'industrie, les écoles supérieures d'interprètes et de traducteurs, les professionnels de la traduction et de la documentation. Il s'agissait, en effet, de répondre à une exigence qui, depuis plusieurs années déjà, s'était présentée en Italie, ainsi que de suivre les progrès effectués au plan international dans le domaine de la terminologie technique et scientifique.

Lors de la réunion préparatoire pour la création d'une commission panlatine de terminologie, organisée à Paris les 13 et 14 décembre 1993, Severina Parodi - qui intervenait au nom de l'Accademia della Crusca et d'Assiterm - avait exprimé la volonté italienne de collaborer à la naissance de la nouvelle initiative. Étaient prises en compte, toutefois, la distance qui séparait l'Italie de la France et des pays francophones, qui - suite à une politique linguistique sérieuse - ont consolidé une position éminente dans l'étude des problèmes liés à la création, à la traduction et à la

diffusion des terminologies techniques et scientifiques, mais également la distance séparant l'Italie des pays ibériques d'Europe et du continent américain, où l'activité terminologique connaît - depuis déjà plusieurs années - une période de grande expansion, avec son entrée de plein droit dans la formation scolaire secondaire et universitaire (bien souvent, aussi, dans les facultés scientifiques), et grâce au développement spécifique de l'étude de la formation des néologismes dans chaque langue.

Ce n'est pas un hasard si, depuis déjà plusieurs années, dans ces aires linguistiques, s'étaient affirmés des espaces privilégiés pour l'élaboration et l'échange des terminologies techniques et scientifiques, comme le Rint et Riterm.

L'attention croissante dont fait l'objet la terminologie technique et scientifique en tant qu'instrument d'étude de la connaissance, ou, mieux, de l'organisation et de la transmission de la connaissance scientifique, avait déjà poussé l'Union latine (organisation intergouvernementale créée en 1954 par la Convention de Madrid, regroupant actuellement trente États de langue néolatine) à promouvoir, en 1972, un Fonds international de terminologies romanes, le Fitro, qui n'avait pas survécu, faute de structure à même de centraliser l'intérêt opérationnel démontré par plusieurs spécialistes et personnalités de nationalités différentes. Cependant, la nécessité d'harmoniser le développement des langues néolatines était restée vivante, surtout dans le secteur de la formation lexicale, qui

met en relation leurs traits communs dans l'utilisation de modes de création et de formants très proches.

C'est dans ce but, et grâce à la collaboration entre la Délégation générale à la langue française et l'Union latine, que le Réseau panlatin de terminologie était créé en décembre 1993. Au printemps 1995, a été également inclus dans la dénomination l'acronyme *Realiter*, qui, s'il peut, par doute ou malveillance, faire penser à une affirmation superflue de son existence, rappelle l'origine linguistique commune latine et fait allusion au phénomène de la réalisation ou exécution linguistique, par laquelle a été désignée la capacité concrète du locuteur à utiliser la connaissance des règles qui régissent sa langue. De plus, il exprime et confirme la raison essentielle de sa propre constitution, celle d'un réseau d'harmonisation linguistique et terminologique des langues néolatines.

À plusieurs reprises a été signalée, au plan international, la coexistence de deux exigences contradictoires (Rey 1992 : 108-120 et Cabré 1993 : 66-68). D'une part, la volonté - devenue désormais une impérative nécessité - d'une communication scientifique et technique de plus en plus dense, qui a trouvé son expression naturelle en une sorte de nouvelle *koinè* ou *lingua franca*, l'anglais, qui commence à montrer les premiers signes d'une évolution ou d'une émancipation pouvant sembler accélérées par rapport aux rythmes morphologiques, et peut-être aussi syntactiques, originaires. Pour justifier l'expansion rapide de cette langue, il est souvent fait appel à sa simplicité et à sa capacité expressive, même si je préfère retenir que la raison principale réside dans le fait qu'elle est le véhicule de diffusion de la grande innovation scientifique et technologique qui s'est produite aux États-Unis au cours de

ce siècle. D'autre part, et parallèlement, s'est déterminé un processus plus articulé et plus complexe, tantôt inspiré du principe d'un « plurilinguisme international », tantôt traversé par des considérations puristes dictées par un « nationalisme linguistique (ou interlinguistique) ».

Avec finesse, Maria Teresa Cabré définit ces deux tendances seulement « apparemment contradictoires », en rappelant que les langues de spécialité présentent des éléments d'uniformité (réduction maximale de la subjectivité et nécessité de communication internationale) et des éléments de diversification, car les termes - même dans des proportions moins importantes que les mots d'une langue - « reflètent une certaine vision du monde » et naissent en tant qu'« expression authentique de la culture d'un peuple ».

Concurremment à ces deux exigences, et certainement du fait de la nécessité croissante de dénommer les concepts et les objets produits par l'incessant progrès scientifique et technique, au cours de ce siècle, ont surgi des initiatives, plus ou moins officielles, destinées à régler les flux de la production terminologique, qui, parfois, prend des chemins chaotiques : il s'agit d'organismes scientifiques et techniques, d'associations ou d'unions de spécialistes, d'institutions administratives nationales ou internationales. Du point de vue méthodologique, s'est développée une gamme très vaste et diversifiée d'interventions : proposition de termes nouveaux; critique de termes créés ou adoptés; collecte, classification ou élaboration de systèmes conceptuels et terminologiques; constitution de banques de données terminologiques plurilingues; prescription, normalisation et harmonisation des terminologies. Il est peut-être superflu de rappeler le rôle important que les spécialistes de chaque discipline et les traducteurs scientifiques ont dû

assumer, souvent dépourvus aussi bien d'instruments de consultation efficaces et dignes de foi que de répertoires des formants et de règles pratiques pour la création néologique. Les lexicographes et les lexicologues ont dû remplacer la figure professionnelle du terminologue, qui a commencé à être constituée seulement récemment, en la dotant des éléments nécessaires de compétence spécifique. Par ailleurs, les spécialistes de linguistique n'ont pas toujours démontré leur penchant ou leur intérêt pour le travail de définition et de codification de ce secteur émergent, même si, dans les dernières années, le débat scientifique s'est ranimé.

Parmi les propositions avancées au cours des dernières décennies pour régler l'évolution des langues, celle de l'aménagement linguistique, qui s'est constituée au Québec à la fin des années soixante et a vu en Jean-Claude Corbeil (1973) un défenseur remarquable, semble trouver un accueil de plus en plus important. Il s'agit d'un type d'intervention, emprunté à l'organisation du territoire, qui vise à favoriser la coexistence et le développement des différentes langues, surtout dans les situations où l'identité culturelle et linguistique ne coïncide pas avec la forme d'organisation de l'État. Cette formule, qui a déterminé le dépassement partiel de méthodes d'intervention plus radicales - comme le dirigisme linguistique, soutenu en France dans les années soixante-dix, et le *language planning* lui-même, diffusé à partir des États-Unis -, s'associe au développement de l'industrie de la langue, visant à promouvoir la constitution de ressources et d'instruments pour soutenir un développement correct des potentiels linguistiques, et à assurer un échange efficace et systématique avec les autres langues. Il n'est pas difficile de relever des préoccupations de type politique,

économique et social qui intéressent plutôt le domaine de la sociolinguistique. Le principe inspirateur a été toutefois accepté par le Réseau panlatin de terminologie et, à l'inverse, il en constitue son objectif principal : favoriser un développement harmonieux des langues néolatines, en tenant compte de leur origine commune et du fait qu'elles utilisent des modalités semblables dans la formation des termes et des formants similaires. De cette façon, on affirme la volonté de valoriser et de préserver un patrimoine commun, dans le respect d'identités historiques et culturelles différentes.

Compte tenu des langues parlées dans les pays représentés dans le Réseau (catalan, espagnol, français, italien, portugais et roumain), les travaux de terminologie menés au sein du Réseau lui-même devront refléter cette diversité linguistique, en accordant aux différentes langues et à chacune de leurs variantes un statut d'égalité dans chaque projet<sup>(1)</sup>.

## 1 Structure du Réseau panlatin de terminologie

Les normes qui établissent le fonctionnement et les activités du Réseau sont réunies dans un *Règlement* (qui a été en grande partie approuvé définitivement lors de la réunion de Barcelone<sup>(2)</sup> et, pour certains aspects, confié à la délibération d'un comité), où sont

(1) Cf. point 1.1 : Principes méthodologiques.

(2) Les réunions de coordination ont eu jusqu'à maintenant le calendrier suivant : Paris, 17 mai 1994; Bruxelles, 8-9 décembre 1994; Lisbonne, 27-29 avril 1995; Barcelone, 14-16 décembre 1995.

présentés les objectifs spécifiques du Réseau :

- Établir des principes méthodologiques communs applicables à la réalisation des produits élaborés conjointement;
- Mener des recherches en commun et créer des outils susceptibles de servir le développement des langues latines;
- Mener des travaux terminologiques conjoints multilingues dans des domaines d'intérêt commun touchant la société;
- Mettre en commun les matériaux de référence documentaires;
- Favoriser la formation réciproque à travers les échanges de formateurs, d'experts, d'étudiants et de matériaux didactiques.

L'organigramme prévoit un secrétariat, confié à l'Union latine; un comité, en exercice pendant trois ans et composé d'au moins neuf membres (un représentant pour chaque groupe linguistique et deux représentants pour les groupes linguistiques intercontinentaux : espagnol, français, portugais); et un secrétaire général, qui est élu tous les trois ans par le comité, dont il fait partie, et assume la représentation du Réseau. Le premier comité, élu lors de la réunion de Barcelone, est constitué de Maria Teresa Cabré, Sofia Alvarez, Leticia Leduc, Loïc Depecker, Louis-Jean Rousseau, Giovanni Adamo, Maria Teresa Lino, Enilde Faulstich et un représentant roumain à désigner. Sont prévues la constitution de groupes de travail responsables des différents projets et la définition éventuelle de modules nationaux de coordination. Peuvent faire partie du Réseau, en qualité de membres, les individus ou les institutions présentés par au moins deux membres du Réseau et admis par le comité à la majorité simple : les demandes devront être adressées au secrétariat. En ce qui concerne la périodicité des réunions, il a été décidé d'organiser une réunion plénière annuelle, à laquelle il est

souhaitable que soient présents tous les membres réunis en assemblée, et une réunion semestrielle de coordination, avec la participation du secrétaire général, du comité, du secrétariat et des responsables des groupes de travail. Autant que possible, sera respectée une alternance dans le choix des lieux des réunions, en rapport avec la représentativité linguistique et géographique des membres. D'un point de vue d'économie générale, seront pris en considération les déplacements programmés de certains membres (par exemple, les réunions du Rint et de Riterm). Le financement des activités du Réseau continuera à être assuré, tant que des subventions des États membres ou des institutions internationales concernées ne seront pas trouvées, par les institutions qui participent actuellement aux groupes de travail.

Enfin, il est important de signaler que l'Université Pompeu Fabra a mis en place un site Internet (*Web*) dédié au Réseau panlatin de terminologie<sup>(3)</sup>. Ce site complète les instruments de communication et de diffusion disponibles dans le Réseau, en particulier la conférence électronique, en fonction depuis 1994 et gérée par l'Union latine, avec accès réservé aux membres.

## 2 Projets de travail du Réseau panlatin de terminologie

### 2.1 Principes méthodologiques

L'une des premières décisions prises par le Réseau a été d'établir des principes méthodologiques communs applicables à la réalisation des

(3) <http://www.iula.upf.es/cpt/cptfr.html> (pour les usagers francophones).

produits élaborés conjointement. À cette fin, une enquête a été menée dans le but d'adopter ces principes méthodologiques communs. Ses conclusions seront rassemblées dans une publication destinée à des techniciens et à des experts non terminologiques. La responsabilité du groupe de travail a été confiée à Louis-Jean Rousseau, de l'Office de la langue française du Québec. Le document, adopté lors de la réunion de Barcelone et publié ci-dessus, rassemble les principes généraux du travail terminologique dans lesquels se reconnaissent les membres du Réseau. Une version italienne a déjà été rédigée, et les traductions en langue espagnole et en langue portugaise sont en cours. En effet, le besoin s'est fait sentir de répondre, en particulier, à la demande du Brésil d'un instrument destiné à guider le travail des non-spécialistes.

Lors de la réunion qui a eu lieu à Lisbonne, est apparue la nécessité de donner vie à un nouveau projet qui constituera un développement d'application du document sur les principes de base du Réseau. Il s'agit de l'élaboration d'un guide méthodologique pour le travail terminographique, sous la responsabilité de Louis-Jean Rousseau, en collaboration avec Loïc Depecker, de la Délégation générale à la langue française, et Maria Teresa Cabré, de l'Université Pompeu Fabra de Barcelone. Ce guide, ou recueil de textes, fournira des indications pratiques pour l'activité terminologique, et pourra être adapté à des usagers et à des produits différents.

## 2.2 Matériaux documentaires de référence

Il a également été décidé de constituer un patrimoine commun de matériaux de référence, y compris le

matériau relatif aux cours de formation.

Le premier travail, qui a été terminé et présenté lors de la réunion de Barcelone, consiste en une table multilingue d'équivalence des formants, avec l'indication de la valeur sémantique, de l'étymologie et d'exemples relatifs à chaque langue. Sous la responsabilité de Maria Teresa Cabré, les données ont été collectées en une base consultable sur le serveur *WWW* de Realiter.

L'Union latine garde constamment à jour un tableau récapitulatif des enquêtes sur la terminologie dans les différents États membres ainsi que dans d'autres pays et réunit les questionnaires sur les ressources terminologiques utilisées pour ces enquêtes, ainsi que leurs résultats, dans la mesure du possible, qu'elle met à la disposition des membres intéressés.

## 2.3 Outils pour le développement des langues néolatines

Dans le but de promouvoir des recherches communes dans le domaine de la constitution d'outils pour le développement des langues néolatines, a été repéré, lors de la réunion de Lisbonne, un premier secteur d'activités, qui consiste en l'élaboration d'outils informatiques pour le traitement de la table multilingue des formants, sous la responsabilité d'Henri Zinglé, de l'Université de Nice, qui, lors de la réunion de Barcelone, a présenté un prototype de programme d'application.

## 2.4 Travaux terminologiques multilingues conjoints

L'élaboration de travaux terminologiques multilingues conjoints dans les secteurs d'intérêt

commun, avec une attention particulière envers les nécessités sociales de chaque pays, constitue l'une des principales occupations du Réseau.

Alain Fantapié, de l'Université de Paris III, a coordonné la rédaction d'un *Dictionnaire multilingue de l'informatique*. Une première section de la nomenclature (environ 200 termes) a été définie en langue française, et les tableaux d'équivalence ont été rédigés dans les autres langues. La diffusion du document est prévue pour 1996, après la révision nécessaire de la part de chaque collaborateur.

Pour le *Vocabulaire multilingue de l'environnement*, dont la responsable est Maria Teresa Lino, de l'Universidade Nova de Lisbonne, une première ébauche de nomenclature, qui devrait être complétée et structurée du point de vue conceptuel, a été rédigée.

Lors de la réunion de Barcelone, la proposition de rédaction d'un *Glossaire panlatin du secteur agro-alimentaire* a été acceptée. Le groupe de travail, coordonné par Massimo Ragucci, du Ministero delle Risorse Agricole, Alimentari e Forestali d'Italie, abordera, dans une première phase, le vocabulaire de la céréaliculture.

Un nouveau projet envisage la constitution d'un *Vocabulaire panlatin de la terminologie*. Le travail s'articule en deux phases : la première prévoit une structuration conceptuelle du secteur, et la deuxième la rédaction du vocabulaire. La coordination du groupe est confiée à Amelia de Irazazábal, de l'Université de Comillas (Espagne), et à Gigliola Negrini, de l'Istituto di Studi sulla Ricerca e Documentazione Scientifica du CNR italien. Il est important de signaler aussi la contribution qui pourrait être apportée aux travaux du Comité technique 37 de l'Iso, qui s'occupe actuellement de la révision des normes terminologiques.

## 2.5 Formation

L'un des secteurs d'activité les plus délicats et les plus importants concerne la formation de terminologues, à travers l'échange de formateurs, d'experts et d'étudiants.

Les responsables du groupe de travail, l'Université Pompeu Fabra et le Colegio de México, sont en train de préparer un *Inventaire des ressources pour la formation*, qui tient compte aussi de la possibilité d'accueil de boursiers de la part des diverses institutions qui adhèrent au Réseau. Un recensement des cours de formation est constamment mis à jour.

Caroline de Schaetzen et Hugo Marquant, de l'Institut libre Marie Haps, de Bruxelles, sont en train d'effectuer un recensement des mémoires de maîtrise qui présentent un contenu terminographique ou terminologique, afin de permettre une circulation plus ample des sujets communs de recherche. Un questionnaire conçu par l'Institut libre Marie Haps et l'Union latine a

été envoyé. Les réponses obtenues de la part des membres du Réseau semblent encourager l'expérimentation de mémoires rédigés conjointement par plusieurs étudiants, de nationalité ou de langue différentes, sur des sujets d'intérêt commun.

Une nouvelle activité prévoit l'élaboration d'un projet de formation à distance en terminologie, confié à l'Université de Paris III. Ce projet sera présenté lors de la prochaine réunion plénière du Réseau, qui se tiendra à Mexico les 11 et 12 novembre 1996, à la suite du 5<sup>e</sup> Symposium ibéro-américain de terminologie, organisé par le Réseau ibéro-américain de terminologie (Riterm). Préalablement, une réunion de coordination de Realiter se tiendra à Nice, dans la deuxième moitié du mois de juin 1996.

*Giovanni Adamo,  
Lessico Intellettuale Europeo - CNR,  
Rome,  
Italie.*

## Bibliographie

Cabré (Maria Teresa), 1993 : *La terminología. Teoría, metodología, aplicaciones*, Barcelona, Editorial Antártida-Empúries.

Corbeil (Jean Claude), 1973 : «Éléments d'une théorie de l'aménagement linguistique», dans *La banque des mots*, vol. 5.

Rey (Alain), 1992 : *La terminologie : noms et notions*, Paris, Puf.

# Principes méthodologiques du travail terminologique au sein du Réseau de terminologie Realiter

Ce texte correspond au projet présenté par Louis-Jean Rousseau (Office de la langue française, Québec) et adopté par les membres du Réseau présents à la IV<sup>e</sup> Réunion de coordination du Réseau panlatin de terminologie (Barcelone, 14-16 décembre 1995).

Termes-clés :  
méthodologie; terminologie; Realiter.

(1) On entend par *harmonisation* la mise en correspondance des termes, phrasèmes et autres unités, les uns avec les autres au sein d'une même langue, et entre langues, en prenant en considération les phénomènes de variation (synonymie, niveaux de langue, variantes géographiques, etc.). En matière de néologie, l'harmonisation peut aller jusqu'à l'utilisation des mêmes racines et des mêmes formants.

(2) On entend par *technolecte* un sous-système linguistique utilisé dans un champ d'expérience particulier et caractérisé par une terminologie spécifique et par d'autres moyens linguistiques tels le style et la phraséologie. Cette notion est le plus souvent nommée *langue de spécialité*, mais cet usage est critiqué (considéré comme impropre).

Attendu que les experts représentant des organismes de terminologie de différents pays réunis à Paris en décembre 1993 ont formulé les objectifs suivants à mettre en œuvre par le Réseau panlatin de terminologie :

- Favoriser le développement harmonisé des langues néolatines, compte tenu de leur origine commune et du fait qu'elles ont recours à des modes de formation lexicale voisins et qu'elles utilisent des formants semblables;
- Établir des principes méthodologiques communs applicables à la réalisation des produits élaborés conjointement;
- Mener des travaux terminologiques conjoints multilingues dans des domaines d'intérêt commun touchant la société.

Attendu que le Réseau panlatin de terminologie s'est déjà donné un programme d'établissement de vocabulaires.

Les membres du Réseau panlatin de terminologie adoptent les principes méthodologiques suivants.

## 1 Principes généraux

### 1.1 Diversité et égalité linguistiques

Compte tenu de la diversité des langues parlées dans les pays représentés dans le Réseau, les travaux terminologiques menés au sein du

Réseau doivent refléter cette diversité linguistique, notamment en donnant aux différentes langues visées et à chacune de leurs variantes un statut d'égalité dans chaque projet.

### 1.2 Approche variationniste

Les travaux terminologiques menés au sein du Réseau, tenant compte des principes socioterminologiques qui se dégagent des tendances actuelles de la terminologie, privilégient l'approche variationniste. Cette approche consiste à refléter la variété des usages et des traditions terminologiques des différentes aires géographiques propres à chaque langue qui fait l'objet des travaux.

D'une manière générale, cette démarche doit également contribuer au développement harmonisé<sup>(1)</sup> des technolectes<sup>(2)</sup>.

### 1.3 Approche systémique

Les travaux terminologiques du Réseau se fondent sur l'approche systémique, c'est-à-dire sur la structuration des notions en fonction des types de relations établies ou observées entre les notions traitées et sur leur définition systémique, laquelle doit refléter le système adopté.

### 1.4 Partenariat

Les travaux terminologiques menés au sein du Réseau privilégient toute forme de partenariat et, au premier chef, le partenariat entre

terminologiques<sup>(3)</sup> de langues et de pays différents, spécialistes des domaines visés par les travaux et usagers de la terminologie.

Ce partenariat, indispensable à la qualité des produits de la terminologie, doit exister à toutes les étapes du travail terminologique.

## 1.5 Qualité

Le Réseau panlatin de terminologie adhère aux principes généraux de qualité qui, s'agissant de terminographie, peuvent être définis, du point de vue de l'utilisateur, de la façon suivante :

– L'accessibilité : l'utilisateur doit pouvoir trouver rapidement la terminologie appropriée à ses besoins et adaptée au niveau et à la situation de communication;

– L'actualité : quel que soit le domaine ou le champ d'expérience, l'utilisateur doit pouvoir utiliser une terminologie à jour;

– La fiabilité : l'utilisateur doit utiliser une terminologie reconnue par ses pairs et par le milieu professionnel auquel il s'adresse, s'il veut communiquer efficacement.

## 2 Principes relatifs aux pratiques terminographiques

### 2.1 Système de notions

L'établissement ou l'adoption d'un système de notions doit se faire selon les principes et les méthodes établis, notamment par l'Iso, en tenant compte de l'objet visé par le travail terminologique et des besoins des usagers. Un système de notions peut être du type hiérarchique (générique

ou partitif), non hiérarchique (séquentiel ou pragmatique) ou mixte.

### 2.2 Nomenclature

La liste des notions à traiter doit être établie selon les besoins observés ou manifestés par les usagers visés. Elle peut être établie par le dépouillement d'un corpus de texte qui doit répondre aux exigences de représentativité, d'exhaustivité relative et d'actualité.

### 2.3 Définition des notions

La définition terminologique doit être systémique, c'est-à-dire qu'elle doit refléter le type de système ou de micro-système dans lequel la notion définie s'inscrit. La définition, qui est constituée d'un énoncé d'une seule phrase, doit comprendre, outre la mention des caractères distinctifs pertinents, la formulation des relations établies entre ces caractères. Ainsi, et sans exclure les autres formes possibles, dans un système générique, la définition comprendra la mention de la notion superordonnée et celle des caractères spécifiques, alors que dans un système partitif, la définition comprendra la mention d'un définisseur tel « partie de » (ou une formule équivalente) suivie, par exemple, de la mention des caractères relatifs à la fonction de l'objet. Les renseignements notionnels supplémentaires peuvent être consignés dans des notes techniques.

Le niveau de technicité et la formulation des définitions doivent être adaptés aux besoins des usagers.

### 2.4 Choix des termes

Le choix des termes doit se faire pour chaque langue visée en tenant compte du niveau de communication, et en tenant compte de l'usage observé, soit l'usage officialisé lorsqu'il existe, soit l'usage dominant parmi le

public auquel le produit terminologique s'adresse. Les variantes nationales ou régionales (il s'agit ici des variantes topolectales) pour chacune des langues doivent être consignées en utilisant les indicatifs des pays ou des régions. On notera que la mention de ces variantes peut être dans certains domaines un élément clé du travail terminologique. Selon les besoins, on pourra également consigner les variantes socioprofessionnelles qu'il conviendra alors de marquer.

Les relations d'équivalence terminologique entre les langues concernées sont fondées sur le parallélisme entre les relations terme(s) - notions établies pour chacune des langues traitées.

Les remarques terminologiques, métaterminologiques ou linguistiques sont consignées dans des notes explicatives. Ces remarques peuvent porter sur un usage particulier, sur le degré de synonymie ou d'équivalence des termes, sur les marques d'usage, etc.

### 2.5 Données terminologiques

Le choix des données terminologiques à présenter au public dépend des besoins de ce public et des objectifs poursuivis par les travaux. Cependant les données habituellement consignées en terminologie sont les suivantes (les données marquées d'un astérisque sont obligatoires et constituent normalement le contenu minimum d'un vocabulaire).

Pour chaque langue considérée :

- \* Terme principal
- Synonyme(s)
- \* Variante(s)
- Dérivés
- \* Indicatif de langue
- \* Indicatif de pays ou de région (s'il y a des variantes)
- Abréviation
- \* Domaine
- Marque d'usage
- \* Indicatif de grammaire

(3) Il s'agit ici de spécialistes de la terminologie ayant une formation linguistique, par opposition aux spécialistes des domaines étudiés.

- \* Définition
- Contexte
- Données phraséologiques
- Note
- \* Source (pour chaque donnée citée).

## 2.6 Présentation des ouvrages terminologiques

### 2.6.1 Paratexte

Le paratexte est constitué :

- D'une préface d'une personne faisant autorité dans le domaine concerné;
- D'un avant-propos exposant les objectifs de l'ouvrage, indiquant le public cible (niveau de technicité, catégorie socioprofessionnelle, etc);
- De la mention des sources premières utilisées pour établir la nomenclature;
- De la mention des collaborations (voir 1.4 Partenariat);
- Des noms et qualités du ou des auteurs, environnement de travail, etc.;
- De la mention de l'aval d'un ou de plusieurs organismes faisant autorité en terminologie ou dans le domaine concerné (organisme de normalisation, organisation professionnelle, organisme à vocation terminologique, etc.);
- Des index jugés nécessaires à la consultation de l'ouvrage;
- D'une bibliographie des ouvrages consultés présentée selon les règles de l'art.

### 2.6.2 Constitution des articles

L'article terminologique est constitué de l'ensemble des données terminologiques (voir 2.5) jugées nécessaires en fonction des besoins du public cible. Cependant, les données cataloguées comme obligatoires doivent être effectivement présentes dans les ouvrages terminologiques élaborés au sein du Réseau.

### 2.6.3 Arrangement des articles

L'arrangement des articles doit être conçu en fonction des modes de consultation du public cible. Ainsi, il peut être jugé préférable de présenter les articles dans l'ordre alphabétique de l'une des langues traitées, ce qui rend nécessaire la présence d'index alphabétiques pour toutes les autres langues. Par contre, si l'on s'adresse surtout à des spécialistes, on pourra préférer la présentation des articles dans un ordre systématique, ce qui a pour double avantage de ne privilégier aucune des langues traitées et de respecter la structure des notions étudiées. Toutefois, des index alphabétiques demeurent nécessaires pour la recherche rapide d'un terme.

La diffusion des ouvrages terminologiques sur support informatisé élimine tous les inconvénients et permet l'exploitation maximale des données terminologiques.

### Bibliographie<sup>(4)</sup>

- Auger (Pierre) et Rousseau (Louis-Jean), 1978 : *Méthodologie de la recherche terminologique*, Québec, Office de la langue française (Études, recherches et documentation).
- Cabre (M. Teresa), 1992 : *La terminologia, la teoria, els mètodes, les aplicacions*, Barcelona, Editorial Empúries.
- Delage (Gisèle) et Lavoie (Jocelyn), 1992 : *La terminologie dans les documents normatifs - guide de rédaction et de présentation*, Québec, Bureau de normalisation du Québec [Fascicule de documentation].
- Depecker (Loïc), 1994 : *Projet de charte d'attribution du label Rint aux dictionnaires ou glossaires du Réseau international de*

*néologie et de terminologie* [document inédit].

Office de la langue française, 1985 : *Vocabulaire systématique de la terminologie*, Québec, Les Publications du Québec.

Organisation internationale de normalisation, 1987 : Iso 704 : *Principes et méthodes de la terminologie*.

Organisation internationale de normalisation, Iso 860 : (DIS) *Harmonisation internationale des notions et des termes*.

Organisation internationale de normalisation, 1990 : Iso 1087 : *Vocabulaire de la terminologie*.

Organisation internationale de normalisation, Iso 10241 : *Normes terminologiques : élaboration et présentation*.

Réseau international de néologie et de terminologie, « Harmonisation des méthodes en terminologie, actes du séminaire », dans *Terminologies nouvelles*, n° 3, juin 1990.

Rousseau (Louis-Jean), 1994 : « Vers une méthodologie de la définition terminologique », dans *Langues et sociétés en contact*, coll. « Sonderdruck aus Canadiana Romanica 8 », Tübingen, Max Niemeyer Verlag, p. 169-174.

Schaetzen (Caroline de), 1993 : *La terminologie au service de la didactique des sciences*. Rapport au ministère de l'Éducation de la Communauté française, Bruxelles.

(4) Cette bibliographie se limite aux ouvrages consultés pour la préparation des principes méthodologiques.

# Gestion assistée des liens entre concepts

Depuis trente ans, l'ordinateur assiste le stockage et la consultation des dictionnaires, des bases et banques de termes. Par contre, il n'y a pas longtemps qu'il intervient dans le travail des terminographes<sup>(1)</sup> eux-mêmes. Le présent article dresse le bilan d'un volet important de cette terminotique systématique, les services des micro-ordinateurs lors de la description des champs notionnels.

Termes-clés :  
terminotique; concept;  
renvoi; lien.

## 1 Arborescences de domaines et renvois

### 1.1 Clé des macro et microstructures

**L**es terminologues appellent «arbre du domaine» ou «champ notionnel» ce que les lexicographes nomment «champ sémantique». Qu'il soit élaboré, consigné ou qu'il reste implicite dans l'esprit du terminographe, ce champ notionnel est la clé des dictionnaires, glossaires, lexiques ou vocabulaires exhaustifs.

Tout d'abord, un champ sémantique de notions spécialisées est beaucoup plus structuré qu'un champ sémantique de concepts courants. S'il est bien décrit, il permet dès lors de vérifier l'exhaustivité de la macrostructure du dictionnaire en voie d'élaboration (pour reprendre la terminologie des lexicographes). La macrostructure est l'ensemble des vedettes (dans le jargon des terminographes alimentant des banques de termes) ou des entrées (vocabulaire utilisé par les lexicographes).

Ensuite, le champ notionnel favorise la rédaction de la micro-structure, c'est-à-dire des articles de dictionnaires, le contenu des

rubriques portant sur les différentes entrées de ce dictionnaire («contenu de la fiche», dans la terminologie des responsables des banques terminologiques). Le champ rend les définitions cohérentes d'une vedette-entrée à l'autre. Grâce à l'arborescence des notions, le terminologue fera par exemple débiter par leur hyperonyme la définition par genre prochain et différence spécifique d'absolument tous les hyponymes.

En outre, le champ notionnel facilite la confection de renvois consistants d'un article à l'autre (rubrique de dictionnaire, toujours pour respecter la terminologie des lexicographes) ou d'une fiche (jargon des terminographes travaillant pour les banques de termes) à l'autre. Si les renvois ne sont pas utiles pour les notions génériques, ils évitent les confusions entre des notions proches lorsqu'elles sont très spécifiques. Le jeu des renvois peut éclairer les degrés d'équivalence d'une langue à l'autre, tels que les citent Aitchison et Gilchrist (1987) : équivalence exacte, intersection, équivalence partielle, équivalence «un à plusieurs», absence d'équivalence. Or, dans les dictionnaires imprimés, et électroniques, ces renvois sont souvent inconsistants ou circulaires.

Enfin, voir le champ notionnel lui-même n'est pas sans utilité pédagogique. Les dictionnaires encyclopédiques, notamment,

(1) Nous n'avons pas renoncé à convaincre la communauté des spécialistes du terme de la pertinence de la distinction établie par G. Lurquin il y a 20 ans déjà entre *terminographie* (confection de dictionnaires de vocabules de spécialités) et *terminologie* (science des termes ou collection des termes d'une spécialité), en parallèle à celle qui existe entre *lexicographie* (confection de dictionnaires de la langue générale) et *lexicologie* (science des mots).

exposent un domaine en même temps que ses termes. D'autres terminologies servent à la fois de dictionnaires, de catalogues et de thesaurus (comme c'est le cas au sein de la société Krupp). Les interrogations de ces dictionnaires multi-usage et multi-cible s'appuient sur la structure logique du champ sémantique du domaine.

L'assistance à la génération des champs sémantiques est dès lors une nécessité, pas un raffinement.

## 1.2 Représentations graphiques

H. Felber (1986) a montré très tôt la plupart des représentations graphiques susceptibles de traduire, avec précision et clarté, la structure variable des champs notionnels :

- Arborescence hiérarchique stricte;
- Arborescence en réseau;
- Cases emboîtées;
- Liste d'items numérotés selon le système de numération américaine, assortie de mises en retrait qui croissent au fur et à mesure que baisse le niveau hiérarchique des notions;
- Thesaurus fléché, un mode de représentation emprunté aux documentalistes;
- Thesaurus circulaire, fléché ou non, également en usage dans les sciences documentaires.

Ajoutons-y le «terminogramme», un autre mode de représentation graphique des thesaurus, adapté aux thesaurus multilingues et imaginé par la société Van Dijk.

Sur la base d'un ou plusieurs de ces modèles, une série d'établissements enseignant la terminologie ont dès lors pris l'heureuse initiative de demander à leurs étudiants de représenter sous forme graphique le champ du sous-domaine décrit par leurs fiches unilingues, bilingues ou multilingues. Cet exercice s'inscrit notamment dans le cadre de mémoires ou de thèses de terminographie.

## 1.3 Renvois et renvois inverses

Dans les dictionnaires, les références croisées ou renvois sont des expressions de type *voir* ou *voir aussi* qui, dans un article de dictionnaire, invitent le lecteur à lire l'article d'une autre entrée. Ce second article comporte des informations additionnelles sur l'entrée depuis laquelle est fait le renvoi. C'est le cas pour les renvois latéraux, c'est-à-dire renvois à des quasi-synonymes, des synonymes, des variantes, des cohyponymes ou coméronymes. Ce l'est également pour les renvois verticaux, qui sont des renvois des génériques à leurs spécifiques et vice-versa.

Les renvois ou références inverses sont des renvois symétriques de renvois simples. Par exemple, les références d'un terme plurilexématique vers un de ses lexèmes-composants. Autre renvoi inverse, le renvoi depuis un terme à un terme qui lui est lié et à partir duquel une référence avait déjà été faite vers ce premier terme.

Les terminographes ne peuvent se souvenir de toute la chaîne logique liant la macrostructure d'un sous-domaine. L'illogisme des renvois est donc la règle dans les dictionnaires. L'utilisateur sera plus gêné par cette incohérence dans un dictionnaire électronique que dans un dictionnaire imprimé. Des études ont montré que l'utilisateur humain était perdu dans une chaîne de plus de trois références consécutives mais ces observations portaient sur les dictionnaires publiés. Ces chiffres doivent être revus pour les ouvrages électroniques, car la rapidité des déplacements dans les logiciels sous Windows incite l'utilisateur à suivre une chaîne de renvois beaucoup plus longue. C'est d'autant plus vrai que ces sauts sont mémorisés par la machine. En pressant quelques touches, le lecteur peut donc les suivre en sens inverse

jusqu'à l'entrée de dictionnaire dont il était parti.

## 2 Tableur pour concepts

Pour dresser et consulter un arbre notionnel, une des assistances les plus efficaces de l'informatique a été imaginée par l'Université d'Ottawa.

### 2.1 Concepts, propriétés

Originellement utilisé en intelligence artificielle, comme ses homologues, le gestionnaire de réseaux sémantiques *code* (*Conceptually-Oriented Design Environment*) a été adapté aux exigences terminographiques. Dans ce «tableur intelligent pour concepts», les concepts sont considérés dans l'optique de l'IA, concepts de classes (*ville*) ou concepts d'instantiation de ces classes (*Paris*). Comme les «frames» de l'IA et les traits sémantiques de la lexicologie componentielle, les propriétés sont quant à elles des unités d'informations minimales décrivant un concept.

D. Skuce (1990) ajoute que ces propriétés peuvent se transmettre automatiquement aux sous-concepts de l'arborescence notionnelle et aux instances de ces sous-concepts. Cet héritage dépendra du type de propriété (il y en a dix) et du type de lien d'héritage (au nombre de quatre). L'utilisateur peut spécifier tous les concepts ou les relations requis mais la relation générique-spécifique peut être spécifiée pour tous les concepts (un élément vide peut être introduit à certains niveaux hiérarchiques s'il le faut). Dans la plupart des cas, tous les sous-concepts héritent de toutes les propriétés mais l'héritage peut être paramétré: lorsqu'une propriété est modifiée à un certain niveau de la hiérarchie ou lorsque le

terminographe déplace un concept d'une branche du réseau à une autre, le terminographe doit dire au programme si cette modification doit se transmettre à tous les sous-concepts. Ces sources potentielles d'inconsistance, comme d'autres encore, sont vérifiées automatiquement. Aussi le terminographe peut-il imaginer et représenter rapidement plusieurs versions de l'arbre notionnel.

## 2.2 Facettes

Les documentalistes appellent très justement «facette» un point de vue adopté pour la structuration d'un champ sémantique. Selon la facette adoptée, un concept peut avoir plusieurs génériques. Les facettes sont vitales en médecine, pour n'en donner qu'un exemple avec Lewalle (1993). Indépendamment de la prudence terminologique que doivent entraîner les sensibilités politiques, les statisticiens, dit-il, groupent les désordres sans incidence statistique sous NEC (*Not Elsewhere Classified*). Les épidémiologistes ont quant à eux adopté la classification internationale des maladies, critiquée par les cliniciens. Les pathologistes recourent à une nomenclature internationale des maladies recommandant des noms préférés, listant tous les synonymes connus et donnant une «description-parapluie», terminologiquement correcte mais cliniquement insuffisante et partiellement pertinente statistiquement. Les cliniciens, enfin, veulent une nomenclature médicale reflétant l'état des connaissances sur les maladies, avec des liens à d'autres maladies non reliées étiologiquement.

Pour une facette précise, certaines propriétés peuvent être bloquées par *Code* et certains concepts, inactivés, dit Skuce (*op. cit.*). Les terminographe peuvent

travailler simultanément sur toutes les facettes.

## 2.3 Affichage du graphe

*Code* peut afficher en tous temps un graphe fléché (arborescence stricte) du domaine en cours de structuration. Les arcs de cette arborescence stricte indiquent les liens verticaux et latéraux. Comme dans tous les réseaux utilisés en intelligence artificielle, ces liens peuvent être soit (co)hypéronymiques soit (co)méronymiques. Des zooms verticaux ou latéraux peuvent être effectués sur les graphes de grande dimension.

Certains nœuds ou sous-arborescences peuvent être temporairement cachés ou déplacés d'un endroit à l'autre de l'arbre. *Code* peut afficher plusieurs graphes à la fois, pour les comparaisons d'un système notionnel dans plusieurs langues et celles des arborescences pour plusieurs facettes.

## 2.4 Fiches

À chaque concept du graphe correspond une fiche, qui traduit en mode textuel la position de ce concept sur le graphe. Ces fiches sont créées et mises à jour automatiquement. Elles décrivent les concepts en détail (propriétés, liens avec les concepts liés). Des notes de taille libre peuvent être rédigées à propos de ces concepts (définitions, caractéristiques linguistiques). Grâce au lien entre arbre et fiches, les modifications qu'apporte le terminographe dans les fiches sont répercutées automatiquement sur l'arbre et vice-versa.

Les déplacements entre concepts et propriétés sur le graphe, ceux qui s'opèrent d'une fiche à l'autre, sont hypertextuels. Grâce au multi-fenêtrage, les graphes, sous-graphes,

dénominations, document-fiche relatifs à un concept peuvent être vus simultanément.

## 3 Gestionnaires de thesaurus

À notre connaissance, *Code* n'est pas disponible sur le marché et il ne tourne pas sur micro-ordinateur. Il existe peut-être d'autres gestionnaires de réseaux sémantiques pour micro-ordinateurs, qui gèrent par héritage propriétés et instances. Nous ignorons si ces gestionnaires sont couplés à des «mini-» gestionnaires de fiches et s'ils permettent les déplacements hypertextuels au sein des graphes.

Les gestionnaires de thesaurus peuvent rendre un service différent de celui que fournissent les gestionnaires de réseaux sémantiques comme *Code*. Ils présentent sur ces gestionnaires l'intérêt d'être disponibles à faible prix, pour *Dos* comme pour *Windows*.

### 3.1 Thesaurus et terminographie

Les gestionnaires de thesaurus peuvent servir au travail terminographique. Un thesaurus est une liste de descripteurs ou mots-clés, c'est-à-dire de termes utilisés dans la recherche d'informations bibliographiques au sein des bibliothèques. L'utilisateur désireux de trouver un ouvrage sur tel sujet, par exemple le traitement de texte, consultera d'abord le thesaurus des mots-clés. Cette consultation lui indiquera si la référence bibliographique des ouvrages portant sur ce sujet figure sur les fiches bibliographiques intitulées *traitement de texte* ou sur celles dont l'onglet porte l'appellation *texteur*. Si c'est sur les fiches intitulées *texteur* qu'ont été notés auteur, éditeur, année et lieu d'édition des ouvrages traitant du traitement de texte, l'utilisateur verra dans le thesaurus une expression *U:* (*use*), signifiant «*traitement de texte*,

voir les fiches relatives à *texteur*». Parallèlement à l'utilisateur, les documentalistes utilisent le même thesaurus et pour connaître eux aussi les descripteurs. Il leur faut connaître les descripteurs licites pour signifier que tel livre à inclure et classer dans leur bibliothèque traite de texteurs (dans notre cas, *traitement de texte* et non *texteur*). Comme le vocabulaire d'une spécialité, un thesaurus est une macrostructure exhaustive et structurée. Il porte soit sur un sous-domaine, si la bibliothèque comporte des ouvrages d'une spécialité, soit sur un sous-ensemble *ad hoc*, si le fonds de la bibliothèque recoupe plusieurs domaines dans lesquels travaille l'entité qui l'héberge. Un gros thesaurus a la taille d'un dictionnaire de spécialité.

Un thesaurus doit donc permettre à l'utilisateur de retrouver les livres dans la bibliothèque. Or, le thème des publications qu'il cherche est parfois extrêmement ciblé. Il souhaite que les livres ou articles traitent par exemple d'une ou plusieurs matières précises; ou bien de telle, telle et telle matière, ou encore, de telle matière mais pas de telle autre et, en même temps, qu'ils aient été publiés après telle année et par tel auteur. Cet utilisateur ne pensera peut-être pas non plus à l'intitulé utilisé par le documentaliste pour qualifier ces matières mais à un synonyme, voire un méronyme ou un hyperonyme. Documentaliste et utilisateur doivent donc déployer une stratégie, de description du contenu des ouvrages pour le premier, de consultation des fiches bibliographiques pour le second. Pour faciliter les cheminements d'une notion à l'autre qu'impliquent ces tâches symétriques, les thesaurus sont dès lors le plus souvent structurés comme un champ notionnel.

Il existe des logiciels destinés aux concepteurs de thesaurus, c'est-à-dire aux documentalistes qui choisissent, tapent et structurent

hiérarchiquement les descripteurs d'un ou plusieurs thesaurus. Beaucoup de gestionnaires de thesaurus sont commercialisés en même temps que les autres modules d'un gestionnaire de bibliothèque (gestionnaire des prêts, des commandes de livres, notamment). Mais certains sont vendus isolément. Parmi les gestionnaires de thesaurus qui peuvent être achetés séparément d'un logiciel documentaire pour les micro-ordinateurs, citons *Cormorant*, *Micro Stride*, *Minisis*, *PC Index*, *Root*, *Term Manager (Cardbox)* et *Thesaurus (Calm)*.

## 3.2 Gestion des liens

Les gestionnaires de thesaurus facilitent la structuration d'un champ notionnel et surtout, la création, la copie, la suppression, les changements de liens entre descripteurs ou vedettes. Dans les gestionnaires de thesaurus, les liens sont explicites. Comme ils importent et exportent des fichiers Ascii, les gestionnaires de thesaurus peuvent donc être utilisés au moment où le terminographe note les renvois et renvois inverses dans son dictionnaire.

### 3.2.1 Génération des liens inverses

Les gestionnaires génèrent automatiquement les liens inverses. Si j'ajoute au thesaurus le descripteur *texteur*, qui, pour un dictionnaire, sera un terme, un logiciel comme *Term Manager* notera à la droite de *texteur* *U (use, à utiliser, sous-entendu comme descripteur)*. Si je demande à ce logiciel de relier ce descripteur-terme par un lien de type «terme plus étroit» (*NT: narrower term*) au terme *texteur scientifique*, de le relier par un lien de type «terme plus large» (*BT: broader term*) à *logiciel de bureautique* et par un lien de type «terme lié» (*RT: related term*) à *tableur*, *grapheur*, *dessineur*, *Term Manager* reliera à *texteur* respectivement *logiciel de*

*bureautique*, par un lien *BT*, *texteur scientifique*, par un lien *NT*, *tableur*, *grapheur*, *dessineur*, par un lien *RT*. Si j'ajoute un lien «utiliser plutôt» (*UF: use for it*) pour lier *texteur* à *traitement de texte*, le logiciel créera un lien inverse appelé «utiliser» (*U: to use*) pour relier *texteur* à *traitement de texte*.

Il en découle une série d'écritures par le gestionnaire, qui évitent de fastidieuses corrections. Si je modifie le descripteur *scientific word processor* en *scientific word processor*, le gestionnaire changera ce terme dans tous les renvois qui le mentionnent. De la même manière, la modification d'un lien se répercute sur les liens qui en découlent. Ainsi, si je supprime le lien entre *advertising* et *advertising expenditure*, le lien inverse entre ce dernier et *advertising* disparaîtra lui aussi. Et si, dans une liste des termes spécifiques de *bureautique*, je supprime d'une touche le lien *BT* qui relie *bureautique* à *texteur*, *tableur*, *grapheur*, *dessineur*, le lien *NT* de spécifique à générique qui liait ces quatre termes à *bureautique* disparaîtra lui aussi, de même que le lien *RT* qui les liait entre eux. En deux pressions de touche, je peux transporter de *bureautique* à *informatique administrative* les liens *BT* qui viennent d'être cités. Un lien *NT* sera alors respectivement créé entre *informatique administrative* et *texteur*, *tableur*, *grapheur*, *dessineur*. Un lien *RT* sera créé entre ces quatre termes, en même temps que sera supprimé le lien *NT* qui reliait *bureautique* à ses quatre spécifiques.

### 3.2.2 Autres fonctions de gestion des liens

Si elles ne sont pas basées sur l'héritage de propriétés comme dans *Code*, les vérifications de consistance sont, elles aussi, assistées par un gestionnaire de thesaurus.

L'utilisateur peut ponctuellement bloquer la génération de liens inverses d'un champ à un autre.

Des masques de saisie peuvent être conçus pour la frappe de descripteurs et de données à leur propos qui sont de même type. Ces masques peuvent donc accélérer la frappe d'entrées et d'articles de dictionnaires.

La recherche, l'impression, la correction, la copie dans un fichier séparé, et l'effacement de descripteurs peuvent se faire par le truchement d'une formule booléenne ou par pointage au sein d'une liste de descripteurs.

Les sauts logiques d'un descripteur à l'autre se font par pression d'une touche (saut au début ou à la fin du thesaurus, au descripteur suivant selon l'ordre alphabétique, saut à la page-écran précédente et suivante). Des filtres à l'affichage facilitent également la cohérence du champ notionnel: affichage des seuls génériques, spécifiques ou collatéraux ou variantes non «normalisées» d'un descripteur assez haut placé dans la hiérarchie...

### 3.2.3 Variété et ouverture des liens

Inconnue de *Code*, une fonction des meilleurs gestionnaires de thesaurus fait varier le type des liens reliant les descripteurs, donc, pour le terminographe, sous-tendant les renvois croisés. Un gestionnaire puissant permet l'inclusion des liens énumérés par C. Ritzler (1991). Dans la liste ci-après, le second élément des paires de liens est souvent symétrique du premier. Il peut donc être généré par l'ordinateur puisqu'il est un inverse:

– Variante non préférée du descripteur. Ce type de lien peut être utilisé en terminographie pour les termes à éviter (archaïsmes ou jargon de firme, par exemple), les dénominations qui ne sont pas

normalisées (banques de termes normalisés), ou les abréviations;

– Variante d'usage équivalent, utilisable pour les synonymes vrais, les quasi-synonymes, les alternatives (acception identique mais vue d'une autre facette);

– Composé ou syntagme créé avec le descripteur, auquel cas ce dernier est évidemment un terme unilématique;

– Exemple, une zone dans laquelle le terminographe peut noter un contexte;

– Descripteur générique et descripteurs spécifiques. Plusieurs niveaux de spécificité peuvent être indiqués;

– Méronyme et holonyme;

– Cohyponyme et coméronyme;

– Cause et effet;

– Domaine, sous-domaine;

– Prédécesseur et successeur;

– Contraire;

– En opposition avec, un lien utilisable en terminographie pour souligner les confusions fréquentes;

– Producteur et produit;

– Matériau et objet;

– Descripteur en langue-source et en langue-cible, pour indiquer l'équivalent du terme dans une ou plusieurs langues, si le gestionnaire de thesaurus est multilingue;

– Descripteur général et instance (relation d'appartenance);

– Homonyme.

D'autres gestionnaires permettent au concepteur de thesaurus, donc au terminographe, d'ajouter des liens, qu'il libelle à sa guise. L'intérêt d'un choix des liens est que les liens varient d'un domaine à l'autre. De plus, les relations terminologiques ne sont pas uniquement logiques. En voici deux exemples. Dans le gestionnaire de glossaires *Phenix*, la société Site a prévu l'assignation, à tous les termes des glossaires techniques (*pompe à huile*, par exemple), d'un champ intitulé «secteur d'activité» (*automobile*, pour l'exemple ci-

dessus), d'un champ «firme» (*Renault*), d'un champ «division» de cette firme (*R25*) et «composant» (*V6*). Ces champs s'ajoutent à celui du domaine (*hydraulique*). Autre illustration, Molholt et G. Goldbogen (1990) ont choisi 17 paires de descripteurs, dans la centaine que répertorie Nutter (1989). Nutter avait lui-même extrait sa liste d'un thesaurus de 23 mono-hiérarchies et 4000 descripteurs, portant sur le bâtiment et la construction. Selon Molholt et G. Goldbogen, ces 17 couples suffiraient pour structurer une série d'autres domaines techniques. Quelques liens intéressants, dans la sous-liste de Molholt et G. Goldbogen:

– Action, ou processus, et résultat;

– Processus, ou opération, et agent ou instrument;

– Agent et contre-agent (comme dans la paire *peinture* et *solvant*);

– Matériau et usage;

– Objet, ou matériau, et propriété qui lui est étroitement liée (*béton* et *précontraint*);

– Activité, ou action, et participants ou bénéficiaires;

– Action et étapes;

– Produit, ou objet, et propriétés réciproques (liant *toxicité* et *poison*);

– Produit, ou objet, et propriétés non réciproques (*élastique* et *caoutchouc*);

– Cardinalité;

– État et conditions, y compris l'antinomie (*première* et *traitée*, pour *matière*).

D'autres rubriques peuvent être prévues pour chaque descripteur, dans les gestionnaires de thesaurus:

– Des champs administratifs, dont certains sont remplis automatiquement, comme le nom de l'auteur du thesaurus, la date de création et des mises à jour de ce thesaurus, ses sources bibliographiques;

– Des notes optionnelles, de taille indéfinie et de plusieurs types (définitions, notes sur l'usage, notes

diachroniques). Ces notes peuvent être rédigées en mode «plein texte», permettant les recherches de chaînes de caractères en leur sein.

Par des impressions partielles, prêtes à cliquer s'il le faut, la cohérence de la macrostructure du dictionnaire et celle des liens sont vérifiables :

- Liste alphabétique ou systématique des descripteurs, avec ou sans leurs liens;
- Liste de descripteurs, filtrée selon certains critères;
- Thésaurus partiels;
- Liste des synonymes;
- Listes hiérarchiques de descripteurs (descripteurs plus génériques ou plus spécifiques qu'un descripteur précis).

## 4 Postes de travail du lexicographe

Un programme avait été dédié à la confection électronique du dictionnaire *Cobuild*. Pour chaque article de ce dictionnaire, une routine de ce programme pouvait imprimer les mots liés logiquement, syntactiquement ou par l'usage à son entrée-vedette. En tous temps, le gestionnaire pouvait également imprimer la section du dictionnaire relative à ces mots apparentés.

Certains postes de travail du lexicographe sont sur le marché. Il s'agit de logiciel gérant toutes les étapes, notamment la mise en page et la typographie, des dictionnaires destinés à être publiés. La création des renvois y est également assistée. Dans *Compulexis*, le terminographe peut vérifier directement à quoi est fait le renvoi. Le poste de travail l'avertit si l'endroit référencé est en cours de modification à ce moment-là. L'utilisateur peut ainsi vérifier si le renvoi doit être modifié lui aussi. Enfin, *Compulexis* déplace le renvoi lorsque le terme ou le concept auquel il est fait référence est déplacé.

## 5 Générateurs d'hypertextes

L'hypertexte est un ensemble de blocs d'informations textuelles, de passages de texte, stockés sur support électronique, hiérarchisés et délimités, appelés nœuds, pourvus de liens, c'est-à-dire de relations entre passages (Hustache-Godinet, 1993). Les liens, explique-t-il, sont physiquement ancrés à des zones, par exemple à un mot ou une phrase, du texte; ces zones sont appelées «boutons». Un des sous-statuts des boutons est celui de lien implicite, c'est-à-dire que son activation transporte l'utilisateur à un autre endroit du document.

### 5.1 Renvois implicites

Comme dans les dictionnaires diffusés sur papier, les renvois directs et indirects peuvent être notés en toutes lettres dans les dictionnaires sur support électronique (CD-Rom, disquette, disque rigide, Internet). La présentation de ces renvois sera alors identique à celle des ouvrages imprimés, par exemple «voir aussi *tableur*» (à la fin de la définition de *texteur*). Depuis l'invention de l'hypertexte, un renvoi peut également être implicite. Dans ce cas, il ne consiste pas en une expression comme *voir*; *voir aussi*, *cf.* Le saut se fait via des endroits «sensibles» de l'écran. Lors des renvois implicites, c'est l'ordinateur, non l'utilisateur, qui se reporte à l'entrée du terme apparenté.

### 5.2 Renvois simples

L'hypertexte est l'outil idéal pour noter les références implicites. *Multiterm* est un gestionnaire de glossaires, à savoir un gestionnaire de fichier spécialisé dans la création et la consultation de glossaires et vocabulaires multilingues. Il a été mis sur le marché à l'intention des

traducteurs et permet la création de renvois hypertextuels. Le concepteur du glossaire ajoute un accent circonflexe devant le mot ou l'expression à l'article duquel il veut renvoyer son futur lecteur. Lors de la consultation du glossaire, le mot au statut hypertextuel s'affiche en vert dans l'article du glossaire. Un «clic» sur ce mot fait afficher l'article relatif à cette entrée (il faut bien sûr que le glossaire comporte également un article sur cette expression).

Le mot hypertextuel peut figurer à côté d'une expression explicite de renvoi comme *Voyez aussi*. Mais le terminographe peut aussi donner le statut d'hypertexte au générique figurant au début d'une définition par genre prochain et différence spécifique. Une combinaison de renvois explicites et implicites est également imaginable. Les renvois explicites peuvent par exemple être faits aux termes avec lesquels des confusions sont fréquentes, tandis que les renvois implicites reportent le lecteur aux termes, également définis dans le glossaire, qui figurent dans des contextes.

### 5.3 Renvois inverses

Dans les logiciels dont les références peuvent être consultées en mode hypertextuel, les références inverses sont implicites; elles ne sont ni écrites dans l'article ni même visibles par un statut de bouton hypertextuel conféré à un mot de l'article.

Lorsque l'utilisateur d'un vocabulaire ou dictionnaire créé avec *Multiterm* clique, au sein d'une définition, d'un contexte ou d'une note, sur un terme auquel le concepteur du dictionnaire a donné le statut de bouton, *Multiterm* affiche la fiche ou article relatif à ce terme. En cliquant sur un des bords précis de fenêtre, l'utilisateur peut ensuite revenir à l'article du terme apparenté

qu'il avait quitté en activant ce lien hypertextuel. La référence inverse, qui est ce mouvement de retour vers la fiche d'origine, est donc préprogrammée dans le logiciel lui-même.

Dans le *Grand Robert électronique*, cette préprogrammation revêt une autre forme. Une zone de l'écran affiche en permanence l'historique, c'est-à-dire la liste des derniers mots demandés par l'utilisateur. Celui-ci peut cliquer sur l'un d'eux pour sauter à son article de dictionnaire.

## 6 Ergonomie des logiciels commerciaux

La finition des logiciels très répandus et anciens est soignée. C'est le cas des gestionnaires de thesaurus. Ils présentent des caractères ergonomiques qui peuvent les rendre plus attrayants, au bout du compte, que des logiciels spécialisés comme *Code*, indépendamment du fait que ces derniers ne tournent que sur des mini-ordinateurs et au sein d'un seul organisme.

### 6.1 Paramétrisation

C. Ritzler (*op. cit.*) énumère les adaptations qu'offre le logiciel de gestion de thesaurus *Index*:

- Attribution possible de mots de passe et d'autres restrictions d'accès;
- Affichage en bas de l'écran, en tous temps, du menu de toutes les commandes disponibles;
- Paramétrisation des menus, de la couleur, des messages du système et des messages d'erreurs;
- Affichage, choix et changement des paramètres d'impression;
- Paramétrisation de l'ordre de saisie des données et des libellés de liens.

### 6.2 Autres «services» confortables

- Citons d'autres fonctions qui, dans les gestionnaires de thesaurus, font gagner du temps:
- Utilisation de masques pour les recherches de descripteurs (avec *Term Manager*; *agrochemical*? fera afficher la fiche de *agrochemical* mais aussi de *agrochemicals*, si c'est sous cette orthographe que le terme ou descripteur avait été encodé. *Adv+* fera lister tous les descripteurs commençant par *adv*);
  - Routines de sauvegardes automatiques;
  - Transferts rapides de thesaurus aux traitements de texte ou aux systèmes d'exploitation, sans sortir du gestionnaire, éventuellement en arrière-plan, que ce soit pour des thesaurus structurés ou non structurés. Les va-et-vient entre générateur de concordances et gestionnaire de thesaurus ou entre gestionnaire de thesaurus et traitement de texte en sont donc eux aussi facilités;
  - Affichage de la structure statistique.

## 7 L'utilisateur peut être exigeant

Les quelques exemples donnés ci-dessus illustrent la phase d'automatisation actuelle de la terminographie systématique. Ils sont loin d'être les seuls. Citons encore le puissant gestionnaire de liens *Termisti*, programmé par l'Institut supérieur de traducteurs et interprètes ISTI, Bruxelles), et qui a fait l'objet d'une description dans un numéro antérieur de *Terminologies nouvelles*.

Les assistants décrits dans cet article améliorent spectaculairement la cohérence, l'exhaustivité, des dictionnaires de spécialité. Dans deux à trois ans, l'utilisateur pourra dès lors se montrer beaucoup plus

intransigeant sur la qualité des vocabulaires, lexiques, dictionnaires, banques de terminologie, réalisés avec eux.

Caroline de Schaetzen  
Centre de Terminologie de Bruxelles  
Institut Libre Marie Haps

## Bibliographie

- Aitchison (J.) & Gilchrist (A.), 1987: *Thesaurus construction. A practical manual*, London, Aslib.
- Felber (H.), 1986: *Manual of terminology*, Paris, Infoterm-Unisist.
- Hustache-Godinet (H.), 1993: «L'hypertexte au lycée», dans *Cahiers pédagogiques*, n° 311, Paris, Crap.
- Lewalle (P.), 1993: «Terminological and related data interchange problems and practice at WHO», dans *Terminology and Knowledge Engineering*, Frankfurt, Ed. K.-D. Schmitz, Indeks Verlag.
- Molholt (P.) & Goldbogen (G.), 1990: «The Use of Inter-Concept Relationships for the Enhancement of Semantic Networks and Hierarchically Structured Vocabularies», dans *Electronic Text Research. Proceedings of the Sixth Conference of the UW Centre for the New Oxford Dictionary and Text Research*, Waterloo, University of Waterloo, UW Centre for the New OED and Text Research.
- Nutter (J.T.), 1989: *A Lexical Relation Hierarchy. Department of Computer Science Technical Report TR 89-6*, Virginia, Polytechnic Institute.
- Ritzler (C.), 1991: «Comparative study of PC-supported thesaurus software», dans *Nisko '91. Proceedings of the International Conference on Knowledge Organization, Terminology & Information Access Management*, Bratislava, Nisko, Scientific Technical Society.
- Skuce (D.) & Meyer (I.), 1990: «Computer-assisted conceptual analysis: an essential component of a terminologist's workstation», dans Czup (H.) et Nedobity (W.), éd., *TKE '90: Terminology and Knowledge Engineering*, vol. 1, Frankfurt, Indeks Verlag.

## Bibliographie de la néologie

SABLAYROLLES (Jean-François), 1996 : *Les néologismes du français contemporain. Traitement théorique et analyse de données*, Thèse de doctorat, Université de Paris VIII, Paris.

La *Bibliographie de la néologie* de ce numéro prend une forme exceptionnelle dans la mesure où elle ne comporte qu'une seule mais volumineuse notice, et que celle-ci concerne un ouvrage presque introuvable, puisqu'il s'agit d'une thèse encore inédite. En plus, l'ouvrage en question porte sur le français général, ce qui, à première vue, l'éloigne des préoccupations du Rint, tournées vers les vocabulaires spécialisés. Depuis la création de cette rubrique, cependant, nous n'avons pas encore eu l'occasion de signaler une étude portant exclusivement sur les fondements théoriques et pratiques de la néologie, ni de la langue générale ni des langues de spécialités, et il nous a semblé important de faire connaître certains résultats, en attendant une publication définitive.

Comme toute bonne thèse, celle-ci commence par une présentation des travaux antérieurs sur le sujet, ce qui est particulièrement utile dans un domaine où les études se font en ordre dispersé. L'auteur présente d'abord la définition de la néologie (et des néologismes) relevée dans des dictionnaires français de 1855 à 1995, puis dans les études consacrées

à ce sujet. Il en ressort une évolution dans le temps, car les préoccupations changent ; au XIX<sup>e</sup> siècle, les néologismes suscitaient encore une profonde inquiétude, dont on trouve encore des traces de nos jours. L'opposition néologisme/archaïsme, encore courante au XIX<sup>e</sup> siècle, semble avoir disparu, mais J.-Fr. Sablayrolles revient sur cette question, et démontre le lien entre archaïsme et résurrection de formes anciennes. Les analyses contemporaines de la néologie sont extrêmement disparates, surtout parce qu'elles se font dans des buts très variés. Elles mettent néanmoins en lumière différents aspects des processus de création lexicale interne ou externe (lorsqu'on inclut dans la néologie les différents phénomènes d'interférence, dont les emprunts), ainsi que les résultats. On peut déplorer que *néologisme* désigne à la fois le processus et le résultat, mais l'auteur met à profit cette ambiguïté, et distingue la néologie considérée comme outil, qui permet d'identifier et d'analyser la diachronie en marche, et la néologie considérée comme objet d'analyse d'une manifestation de la compétence linguistique des locuteurs. La grille de lecture des études de néologie qui ressort de cette comparaison permet enfin non seulement une vue d'ensemble, mais aussi la traduction d'un système

d'analyse néologique dans les termes d'un autre.

Après avoir examiné la définition de la néologie dans les dictionnaires, l'auteur analyse la place qu'elle occupe dans différentes théories linguistiques contemporaines. Ce n'est une surprise pour personne de constater que celle-ci est réduite à la portion congrue, la plupart des modèles synchroniques laissant peu de place à l'évolution du système. Le traitement de la néologie dans les travaux de Danielle Corbin fait l'objet d'une analyse toute particulière, et non seulement parce que ce chercheur minimise l'importance des attestations. Il reste néanmoins à expliquer pourquoi certains mots sont effectivement réalisés à partir de l'ensemble des mots possibles, et pourquoi certains néologismes transgressent les règles énoncées. J.-Fr. Sablayrolles démontre que le néologisme pourrait bien constituer un test qui mesure la puissance d'une théorie générale, et fournit ainsi une preuve de sa pertinence pour la linguistique en général.

L'auteur, pour sa part, se fonde pour la démonstration de sa thèse sur des attestations, puisées dans six corpus différents, illustrant chacun un aspect de l'activité terminologique. Le premier corpus est consciemment aléatoire, comme celui du Gril de l'Université Paris VII, déjà décrit dans *Terminologies nouvelles* (n°11, p. 26).

Bibliographie de la

# néologie

Les autres corpus, plus systématiques, proviennent d'extraits homogènes du *Monde*, de la presse féminine (peu néologique), d'un auteur particulièrement fécond en créations lexicales, R. Jorif, de la chronique radiophonique de Philippe Meyer, sans oublier un dernier corpus plus étonnant mais en réalité tout à fait justifié, celui des copies d'élèves de lycée. Il en dégage douze types d'informations qui conditionnent l'emploi de néologismes, ce qui met bien en lumière la complexité et l'hétérogénéité du processus néologique.

Il en ressort que la néologie est une partie constitutive de la vie d'une langue et que les créations néologiques sont bien plus nombreuses que ce que l'on trouve même dans les dictionnaires de mots nouveaux (avis donc aux alimentateurs de *Balnéd*). En effet, la majorité des items de son corpus est composée d'hapax, ce qui n'infirmes pas du tout sa thèse. Il démontre, au moyen d'analyses statistiques, que la structure morphologique des néologismes n'obéit pas toujours à des règles préexistantes de la langue. Le sentiment de nouveauté, évoqué dans de nombreux écrits, reçoit ici, enfin, une définition opératoire, qui, paradoxalement, remet quelque peu en cause la possibilité de bien classer les phénomènes néologiques. Il se sert également du concept de prototype pour expliquer la formation de paradigmes néologiques.

La thèse de Jean-François Sablayrolles exclut expressément la néologie terminologique. Ceci constituerait le sujet d'une autre thèse. On peut néanmoins pointer, en creux, quelques différences par rapport à la néologie de la langue générale. On note d'abord le caractère très spontané des néologismes de langue, très ancrés dans le discours, et par ailleurs très souvent motivés par un besoin purement syntaxique. La néologie terminologique, en revanche,

serait bien plus réfléchie, consciente, contrôlée, qu'elle soit le fait de scientifiques, de techniciens ou d'aménageurs. La néologie de la langue, telle qu'on nous la présente ici, est relativement peu «dénominateur»; on exprime davantage un sentiment, on met en valeur, on n'essaie pas de trouver un nom pour un nouveau concept. Le fait que les néologismes de langue sont souvent accompagnés de remarques métalinguistiques de la part du locuteur/scripteur n'infirmes pas cette constatation, ni, *a contrario*, la part de la néologie de discours dans les langues de spécialité, qui est peut-être plus importante que l'on ne pense. Troisième différence, plus spécifique au français : la part de l'emprunt, et de l'interférence en général, est minime. Les membres du jury s'étonnaient de relever si peu d'exemples de l'influence de l'anglais dans un aussi vaste corpus de langue française. La néologie terminologique, relève souvent, en revanche, de ce que J.C. Sager appelle la création terminologique secondaire.

Les études de néologie terminologiques profiteront non seulement de cette présentation du fonctionnement de la néologie en français, mais pourraient aussi s'inspirer de certaines méthodes pratiquées dans la thèse, en particulier un fort ancrage de l'étude des néologismes dans le discours, d'où l'importance des contextes des banques de néologie qui commencent à voir le jour.

*John Humbley,  
Centre de néologie et de terminologie,  
Laboratoire de linguistique  
informatique,  
Inalf - Ura 1576,  
Université Paris XIII,  
France.*

Disponible sous forme de microfiches :  
Atelier national de reproduction  
des thèses  
Université de Lille III  
9 rue A. Angellier  
59046 Lille, France

## Inventaire des travaux en cours et des projets de terminologie

**D**epuis la décision du Rint de diffuser l'inventaire des travaux en cours et des projets de terminologie sous forme de chronique dans *Terminologies nouvelles*, une première chronique a été publiée dans le numéro 13. Voici la deuxième chronique : elle regroupe les travaux en cours qui nous ont été signalés depuis six mois.

La rédaction de la mention a été faite à partir des informations qui nous ont été transmises. Les rectifications, s'il y a lieu, peuvent être adressées au responsable de votre module.

Nous remercions toutes les personnes qui ont collaboré à cet inventaire tant à l'intérieur qu'à l'extérieur du Rint, les participants à l'enquête qui nous ont communiqué les renseignements.

ABRANTES VENTURA GARCIA,  
Maria de Lurdes.

### *Terminologia da senologia.*

Responsable : Maria Teresa Rijo Fonseca Lino (professeur).

Glossaire portugais-anglais-français avec abréviations, arbre du domaine, contextes, définitions, domaines d'emploi, exemples, indicatifs de grammaire, notes, renvois, sources, synonymes, références bibliographiques, index. 150 p.

Échéance du travail : 1997.

Thèse de maîtrise -  
Universidade Nova de Lisboa.

Vedette-matière : sein.

Adresse : Universidade Nova de Lisboa,  
Av. de Berna, 26 C - 1000 Lisboa, Portugal.

BOMBARDIER. GROUPE CANADAIR.

### *Aéromot : base de données.*

Mise à jour permanente.

Responsable : Lise Genest.

Vocabulaire français-anglais avec abréviations, contextes, définitions, domaines d'emploi, formes fautives, indicatifs de grammaire, notes, renvois, sigles, sources, synonymes, références bibliographiques.

Publication prévue pour décembre 1995.

Document sur support informatique.

Vedettes-matière : industrie aéronautique; avions - conception et construction; aéronautique.

Adresse : Bombardier Inc., Groupe Canadair, 400, chemin de la Côte-Vertu, Dorval (Québec) Canada, H4S 1Y9.

Tél. : (+1 /14) 855-7273,

télééc. : (+1 /14) 855-7404.

# Inventaire

CENTRE DE LINGUISTIQUE  
APPLIQUÉE DE DAKAR (CLAD).*Terminologie des mathématiques,  
du préscolaire au CM2.*

Responsable : Chérif Mbodj;  
collab. : Aliou Ngoné Seck.  
Lexique français-wolof avec index français  
et wolof.  
Échéance du travail : novembre 1996,  
publication prévue pour mars 1997.  
Vedette-matière : mathématiques.

*Vocabulaire des élections.*

Responsable : Chérif Mbodj;  
collab. : Aliou Ngoné Seck.  
Lexique français-wolof avec index français  
et wolof.  
Échéance du travail : novembre 1996,  
publication prévue pour mars 1997.  
Vedette-matière : élections.  
Adresse : Centre de linguistique appliquée  
de Dakar, Université Cheikh Anta Diop,  
Dakar, Sénégal.  
Tél. : (+221) 23 01 26,  
télé. : (+221) 24 71 47.

## CENTRE FRANÇOIS-CHARON.

*Dictionnaire des termes de la  
réadaptation. T. 2. Vocabulaire des termes  
d'intervention en réadaptation, en  
adaptation et en intégration sociale.*

Responsable : Maurice Blouin;  
collab. : Caroline Bergeron.  
Dictionnaire français avec définitions,  
sources, références bibliographiques,  
index anglais, espagnol, portugais,  
allemand. Environ 2 000 entrées.  
Travail sur support informatique.  
Vedettes-matière : handicapés -  
réadaptation; réadaptation; intégration  
sociale; handicapés - intégration.  
Adresse : Centre François-Charon,  
525, boul. Hamel, Québec (Québec)  
G1M 2S8.  
Tél. : (+1 418) 529-9141,  
poste 6396,  
télé. : (+1 418) 529-3699.

CONSEIL DE L'EUROPE.  
BUREAU DE LA TERMINOLOGIE.*Population.*

Responsable : Jacques Vandeloek.  
Lexique français-anglais avec abréviations,  
références bibliographiques. 500 p.  
Publication prévue par les Éditions du  
Conseil de l'Europe pour 1996.  
Vedettes-matière : démographie;  
population.  
Adresse : Conseil de l'Europe, F-6700  
Strasbourg, France.  
Tél. : (+33) 88 41 25 40,  
télé. : (+33) 88 41 27 01.

## DE ILLANES, N.

*Elsevier's Dictionary of Drug Traffic  
Terms : in English, French, German,  
Spanish and Portuguese.*

Lexique anglais-français-allemand-  
espagnol-portugais. Environ 6 000  
termes et 500 p.  
Publication prévue par Elsevier pour  
1996.  
ISBN 0444819371.

Vedette-matière : drogues - trafic.  
Adresse : Elsevier Science B.V., P.O. Box  
1991, 1000 BZ Amsterdam, Pays-Bas.  
Tél. : (+31) 20-4852 462,  
télé. : (+31) 20-4852 623.

*Dictionnaire anglais des  
télécommunications = French Dictionary of  
Telecommunications.*

Dictionnaire anglais-français.  
Publication prévue par Routledge pour  
août 1996.  
Vedette-matière : télécommunications.  
Adresse : Routledge, Reference Marketing  
Department, 29 West, 35th Street,  
New York, NY 10001-2299, États-Unis.  
Tél. : (+1 212) 224-3336,  
télé. : (+1 212) 563-2269.

## DUGUAY, Christine.

*La terminologie du typage génétique.*

Responsable : Geneviève Mareschal  
(directrice).  
Vocabulaire français-anglais avec  
définitions, indicatifs de grammaire,  
notes, synonymes, index. 100 termes.  
Échéance du travail : 1995.

Thèse - Université d'Ottawa, École de  
traduction et d'interprétation.  
Vedette-matière : génétique.  
Adresse : Université d'Ottawa, École de  
traduction et d'interprétation,  
52, rue de l'Université, B.P. 450, succ. A,  
Ottawa (Ontario) Canada, K1N 6N5.  
Tél. : (613) 564-2346,  
télé. : (613) 564-2959.

## EUREKA TRANSLATIONS.

*Lexique bilingue pour l'industrie des  
transports.*

Mise à jour permanente.  
Responsable : Pascaline Denblyden.  
Lexique français-anglais avec abréviations.  
Environ 300 termes.  
Vedette-matière : transport.  
Adresse : Eureka Translations, 21,  
Devondale Avenue, Willowdale (Ontario)  
Canada, M2R 2C9.  
Tél. : (+1 416) 733-8547,  
télé. : (+1 416) 226-2895.

## FARJI, Débora.

*Étude comparée de la terminologie  
néologique en français et en espagnol dans le  
domaine de la télédétection en  
agrométéorologie.*

Responsable : John Humbley  
(directeur de thèse).  
Vocabulaire français-espagnol avec  
abréviations, contextes, indicatifs de  
grammaire, notes, sources, synonymes,  
références bibliographiques.  
Échéance du travail : 1997.  
Thèse de doctorat en sciences du langage  
- Université Paris-Nord.  
Vedettes-matière : météorologie agricole -  
télédétection; agriculture; télédétection.  
Adresse : Université Paris-Nord,  
avenue Jean-Baptiste-Clément,  
93430 Villetaneuse, France.  
Tél. : (+33) 1 49 40 38 56,  
télé. : (+33) 1 48 26 30 19,  
courrier électronique :  
humbley@d.univ-Paris13.fr.

GAZZOLA, Beth.

*Terminology of Human Gene Mapping Using Somatic Cell Hybridization.*

Responsable : Geneviève Mareschal (directrice).

Vocabulaire anglais-français avec définitions, indicatifs de grammaire, notes, synonymes, index. 50 notions, 110 termes.

Échéance du travail : 1995.

Thèse - Université d'Ottawa, École de traduction et d'interprétation.

Vedettes-matière : cartes chromosomiques humaines; cellules - hybridation; génétique.

Adresse : Université d'Ottawa, École de traduction et d'interprétation, 52, rue de l'Université, B.P. 450, succ. A, Ottawa (Ontario) Canada, K1N 6N5. Tél. : (+1 613) 564-2346, téléc. : (+1 613) 564-2959.

GENDARMERIE ROYALE DU CANADA.

*Vocabulaire de la police de protection.*

Responsable : Lucie Guertin.

Vocabulaire français-anglais, anglais-français.

Publication prévue par la GRC pour décembre 1995.

Vedettes-matière : police; protection civile.

Adresse : Gendarmerie royale du Canada, 1200, promenade Vanier, A-400, Ottawa (Ontario) Canada, K1A 0R2. Tél. : (+1 613) 991-0609.

GOFFER, Z.

*Elsevier Dictionary of Archaeometry and Archaeological Materials : in English, German, Spanish, French, Italian and Portuguese.*

Lexique anglais-allemand-espagnol-français-italien-portugais avec quelques définitions anglaises. Environ 2 000 termes et 400 p.

Publication prévue par Elsevier pour 1995. ISBN 0444819495.

Vedettes-matière : archéométrie - instruments; archéologie - instruments.

Adresse : Elsevier Science B.V., P.O. Box 1991, 1000 BZ Amsterdam, Pays-Bas.

Tél. : (+31) 20-4852 462, téléc. : (+31) 20-4852 623.

HOEPELMAN, J.; MAYER, R.

*Elsevier's Dictionary of Information Technology : in English, German, French, Italian, Spanish and Dutch.*

Lexique anglais-allemand-français-italien-espagnol-néerlandais avec abréviations, acronymes, quelques définitions anglaises, synonymes. Environ 2 300 termes et 500 p.

Publication prévue par Elsevier pour 1995. ISBN 0444884106.

Vedette-matière : technologie de l'information.

Adresse : Elsevier Science B.V., P.O. Box 1991, 1000 BZ Amsterdam, Pays-Bas.

Tél. : (+31) 20-4852 462, téléc. : (+31) 20-4852 623.

INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (France). UNITÉ CENTRALE DE DOCUMENTATION (Jouy-en-Josas).

*Vocabulaire de la cuisson-extrusion.*

Responsables : Annick Bouroche et Michèle Le Bars.

Vocabulaire français-anglais-allemand avec abréviations, définitions et notes en français, contextes et équivalents en anglais et en allemand, indicatifs de grammaire, renvois, synonymes, index anglais-français et allemand-français, références bibliographiques. 132 notions. Publication prévue pour 1995. Vedette-matière : aliments - industrie et commerce. Descripteurs libres : cuisson-extrusion; agro-alimentaire.

Adresse : Institut national de la recherche agronomique, 147, rue de l'Université, 75341 Paris Cedex 07, France.

Tél. : (+33) 1 42 75 90 00, téléc. : (+33) 1 47 05 99 66.

INSTITUTO DE ACONDICIONAMIENTO AMBIENTAS.

*Terminología sobre : energía en arquitectura.*

Responsable : Guillermo Enrique Gonzalo.

Dictionnaire espagnol-français-anglais-allemand avec abréviations, définitions, notes, sources, références bibliographiques. 1 200 termes et 300 p. Échéance du travail : décembre 1995.

Vedettes-matière : architecture; architecture et économies d'énergie. Adresse : Instituto de Acondicionamiento Ambientas, Av. Solano Vera 1086 (4107) Yerba Buena Tucuman, Argentine. Tél. : (+54) 81 25 25 72, téléc. : (+54) 81 31 03 87.

LABORATOIRE D'INGÉNIEURIE LINGUISTIQUE (LILLA).

*Vade-mecum du médecin en chirurgie digestive.*

Responsable : Henri Zingle, directeur du LILLA.

[Lexique] espagnol-portugais-français-italien.

Vedette-matière : digestif, appareil - chirurgie.

Adresse : Laboratoire d'ingénierie linguistique, 98, boul. Édouard-Herriot, BP 209 F 06204, Nice Cedex 3, France.

Tél. : (+33) 93 37 54 39, téléc. : (+33) 93 37 55 36, courrier électronique : henri@lilla.unice.fr.

MAD.

*Diccionario de computación/ciencia de la información.*

Responsable : Andrés Cruz Paz; collab. : M. Molares, D. Betancourt et O. Piñero.

Dictionnaire espagnol-français-roumain-anglais-allemand avec abréviations, définitions, domaines d'emploi, notes, renvois, synonymes, index.

Vedettes-matière : informatique; sciences de l'information; télécommunications.

Adresse : MAD, Concepcion 158, Lawton, La Habana Codigo 10700, Cuba. Tél. : (+1 537) 99-1395, téléc. : (+1 537) 62-6501.

## ONTARIO-HYDRO.

*Banque de données terminologiques.*

Responsable : Janice Guthrie.  
Banque de terminologie française-anglaise avec abréviations, contextes, domaines d'emploi, indicatifs de grammaire, notes, sigles, sources, synonymes, références bibliographiques.

Vedette-matière : électricité.

Adresse : Ontario Hydro, 700, University Avenue, Toronto (Ontario) Canada, M5G 1X6.

Tél. : (+1 416) 592-5892,  
télé. : (+1 416) 592-9751.

ORGANISATION  
DE COOPÉRATION  
ET DE DÉVELOPPEMENT  
ÉCONOMIQUES.

*Glossaire de l'agriculture : anglais-français.* 2<sup>e</sup> éd. ref. et augm.

Lexique anglais-français avec liste des noms scientifiques latins. Plus de 12 000 termes.

Publication prévue par les Éditions de l'OCDE pour début 1996.

Vedette-matière : agriculture.

*Glossaire des difficultés de traduction : anglais-français.*

Lexique anglais-français.

Publication prévue par les Éditions de l'OCDE pour 1997.

Vedette-matière : traduction.

Adresse : OCDE, 2, rue André-Pascal, 75775 Paris Cedex 16, France.

OTTAWA-CARLETON (ONT.),  
GREFFE RÉGIONAL.*Dictionnaire.*

Mise à jour permanente.

Responsable : Dianne Charron.

Dictionnaire anglais-français avec abréviations, contextes, formes fautives, notes, renvois, sigles, sources, synonymes.

Vedette-matière : administration municipale.

Adresse : Municipalité régionale d'Ottawa-Carleton, Greffe régional, 111, rue Lisgar, Ottawa (Ontario) Canada, K2P 2L7.

Tél. : (+1 613) 560-2058,  
télé. : (+1 613) 560-2115.

## PODZIMEK, Lenka.

*The Parks Canada Visual Dictionary: A Commented Bilingual Terminology File of Musical Instruments Terms.*

Responsable : Geneviève Mareschal (directrice).

Vocabulaire anglais-français avec définitions, indicatifs de grammaire, notes, synonymes, index. 100 termes.

Échéance du travail : 1995.

Thèse - Université d'Ottawa, École de traduction et d'interprétation.

Vedette-matière : instruments de musique.

Adresse : Université d'Ottawa, École de traduction et d'interprétation, 52, rue de l'Université, B.P. 450, succ. A, Ottawa (Ontario) Canada, K1N 6N5.

Tél. : (+1 613) 564-2346,  
télé. : (+1 613) 564-2959.

QUÉBEC. COMMISSION  
DE TOPONYMIE.*Terminologie des parcs et espaces verts associés.*

Responsable : Jean-Yves Dugas.

Vocabulaire français avec contextes, définitions, indicatifs de grammaire, notes, renvois, sources, synonymes, références bibliographiques. Environ 30 termes.

Échéance du travail : février 1996.

Document à diffusion restreinte.

Vedettes-matière : parcs; espaces verts.

Adresse : Commission de toponymie du Québec, Édifice Marie-Guyart, Aile René-Lévesque, 4<sup>e</sup> étage, 1060 rue Conroy, Québec (Québec), G1R 5V8.

Tél. : (+1 418) 643-0768,  
télé. : (+1 418) 644-9466,  
courrier électronique : Alain Vallieres-  
@infopuq.quebec.ca.

QUÉBEC. OFFICE DE LA LANGUE  
FRANÇAISE.*Le tour [affiche, micro-lexique].*

Responsable : Dominique Dos Ghali.

Lexique anglais-français, français-anglais. Environ 30 termes.

Publication prévue pour la fin de 1995.

Vedettes-matière : machines-outils; tour (outil).

Adresse : Office de la langue française, 200, chemin Sainte-Foy, 4<sup>e</sup> étage, Québec (Québec), G1R 5S4.  
Tél. : (+1 418) 646-9592,  
télé. : (+1 418) 643-3210.

RAPENNE, Bothelo,  
ANNETTE, Pierette.*Terminologia e tradução do secretariado.*

Responsable : Maria Teresa Rijo da Fonseca Lino (professeur).

Vocabulaire portugais-français avec abréviations, arbre du domaine, contextes, définitions, domaines d'emploi, exemples, indicatifs de grammaire, notes, renvois, sources, synonymes, références bibliographiques, index. 150 p.

Échéance du travail : 1997.

Thèse de maîtrise - Universidade Nova de Lisboa.

Vedette-matière : travail de bureau.

Descripteur libre : secrétariat.

Adresse : Universidade Nova de Lisboa, Avenue de Berna, 26 C -1000 Lisboa, Portugal.

## SAUVAGE, Jean-Louis.

*Vocabulaire anglais de la presse utile à un lecteur francophone.*

Vocabulaire français-anglais. Environ 18 000 termes.

Publication prévue par les PUM pour le début de 1996.

Vedette-matière : presse - Grande-Bretagne.

Adresse : Université de Mons-Hainaut, Faculté Warocqué, 17, Place Warocqué, 7000 Mons, Belgique.

Tél. : (+32) 65 37 32 06,  
télé. : (+32) 65 37 30 54.

SERVIÇO NACIONAL DE APRENDIZAGEM INDUSTRIAL. CENTRO TECNOLÓGICO DO MOBILIÁRIO.

*Glossário mobiliário e madeiro.*

Vocabulaire portugais-espagnol-français-italien-anglais-allemand avec définitions, synonymes, références bibliographiques. 800 termes.

Publication prévue par CETEMO/SENAI/IBICT.

Vedettes-matière : meubles; bois, travail du.

Adresse : Serviço nacional de aprendizagem industrial. Centro tecnológico do mobiliário, Av. Presidente Costa e Silva, 571 Bento Gonçalves - RS CEP 95700-000, Brésil.  
Tél. : (054) 451 91 66,  
télé. : (054) 451 35 85.

THIRY, Bernard.

*Terminología bilingüe de la responsabilidad civil extracontractual en los derechos español y belge.*

Dictionnaire espagnol-français avec arbre du domaine, abréviations, contextes, définitions, domaines d'emploi, indicatifs de grammaire, notes, renvois, sources, synonymes. 200 entrées.

Vedettes-matière : droit civil - Belgique; responsabilité civile - Belgique; droit civil - Espagne; responsabilité civile - Espagne.  
Adresse : École des hautes études commerciales,  
24, rue Louvrex, Liège, Belgique.  
Tél. : (+32) 41 32 72 11,  
télé. : (+32) 41 32 72 40.

TORRES CAMPOS PERDIGOTO, Maria Teresa.

*Terminologia da cartografia.*

Responsable : Maria Teresa Rijo Fonseca Lino (professeur).

Vocabulaire portugais-français-anglais avec abréviations, arbre du domaine, contextes, définitions, domaines d'emploi, exemples, indicatifs de grammaire, notes, renvois, sources, synonymes, références bibliographiques, index. 150 p.

Échéance du travail : 1997.

Thèse de maîtrise - Universidade Nova de Lisboa.

Vedette-matière : cartographie.

Adresse : Universidade Nova de Lisboa, Avenue de Berna, 26 C -1000 Lisboa, Portugal.

UNIVERSITÉ DE MONCTON. CENTRE DE RECHERCHE EN LINGUISTIQUE APPLIQUÉE.

*Aquaculture.*

Responsable : Claire Leblanc.

Vocabulaire français-anglais avec abréviations, définitions, domaines d'emploi, formes fautives, indicatifs de grammaire, notes, renvois, sigles, synonymes, références bibliographiques, index anglais et français. 1 300 termes.

Publication prévue par le CRLA.  
Échéance du travail : décembre 1995.

Document sur support informatique.  
Vedettes-matière : aquiculture; pisciculture; conchyliculture; algues - culture; crustacés - élevage.

*Vocabulaire de l'arpentage.*

Responsable : Jeannette Leblanc.

Vocabulaire français-anglais avec abréviations, définitions, indicatifs de grammaire, notes, sigles, synonymes, références bibliographiques. Environ 1200 termes.

Publication prévue par le CRLA à une date indéterminée.

Document sur support informatique.  
Vedettes-matière : arpentage; topométrie; topographie; photogrammétrie; astronomie; géodésique; géomatique.  
Adresse : Université de Moncton, Centre de recherche en linguistique appliquée, Moncton (Nouveau-Brunswick) Canada, E1A 3E9.  
Tél. : (+1 506) 858-4057 ou (+1 506) 858-4415,  
télé. : (+1 506) 858-4166,  
courrier électronique : termin@umoncton.ca.

VAN HOOFF, D.

*Elsevier's Legal Dictionary : in English, French, German, Spanish and Dutch.*

Lexique anglais-français-allemand-espagnol-néerlandais. Environ 25 000 termes et 1500 p.

Publication prévue par Elsevier pour 1995. ISBN 0444817859.

Vedette-matière : droit.

Adresse : Elsevier Science B.V., P.O. Box 1991, 1000 BZ Amsterdam, Pays-Bas.  
Tél. : (+31) 20-4852 462,  
télé. : (+31) 20-4852 623.

ZIZÉTE MACHADO CÊA MARCO, Virginia.

*Terminologia da PAC.*

Responsable : Maria Teresa Rijo da Fonseca Lino (professeur).

Vocabulaire portugais-français avec abréviations, arbre du domaine, contextes, domaines d'emploi, exemples, indicatifs de grammaire, notes, renvois, sources, synonymes, références bibliographiques, index. 150 p.

Échéance du travail : 1997.

Thèse de maîtrise - Universidade Nova de Lisboa.

Vedette-matière : politique agricole - pays de la Communauté économique européenne.

Adresse : Universidade Nova de Lisboa, Avenue de Berna, 26 C -1000 Lisboa, Portugal.

*Marjolaine Gagnon,  
Micheline Gagnon,  
Office de la langue française,  
Québec.*

## Publications

### *Les dictionnaires de langue : méthodes et contenus*

Tous les ans, l'Université de Cergy-Pontoise, sous l'impulsion de Jean Pruvost, organise une Journée des dictionnaires, manifestation composée d'une série de conférences sur un thème donné accompagnée d'une exposition de production dictionnaire. Les actes de la première journée viennent de paraître. Deux sous-thèmes sont abordés : « Des mots aux dictionnaires », et « Diversité de l'information et richesse des procédures dans les dictionnaires de langue ».

Dans la première catégorie, on compte les communications suivantes :

- Entretiens sur le thème : « Des mots aux dictionnaires » (Bernard Quémada);
- Des mots aux dictionnaires : domaines scientifiques et techniques (Danielle Candel);
- Les marques d'usage dans les dictionnaires français monolingues (Michel Glatigny);
- La neuvième édition du Dictionnaire de l'Académie française (Denis Rieu).

La seconde catégorie comporte les communications suivantes :

- Entretiens sur le thème : « Diversité de l'information et richesse de procédures dans les dictionnaires de langue » (Josette Rey-Debove);

- L'articulation entre la pratique lexicographique, la métalexigraphie et la linguistique dans la formation (Pierre Corbin);

- Entretiens sur le thème : « Diversité de l'information et richesse de procédures dans les dictionnaires de langue » (Claude Kannas);
- La composante culturelle de l'exemple dans le dictionnaire de langue (Alise Lehmann).

Chaque communication est accompagnée d'une transcription de la discussion qui a suivi sa présentation.

La journée du dictionnaire 1996 sera consacrée aux dictionnaires spécialisés et aux dictionnaires de spécialité. Pour tout renseignement, s'adresser à Jean Pruvost, Université de Cergy-Pontoise, F-95011 Cergy-Pontoise Cedex.

Pruvost (Jean) (éd.),  
*La journée des dictionnaires 1994*,  
Centre de recherches texte/histoire,  
Université de Cergy-Pontoise.

### *Construire le pluralisme linguistique : quelles initiatives dans l'Union européenne ?*

Sous l'égide du ministère de l'Enseignement supérieur et de la Recherche, un collectif de linguistes a organisé une rencontre sur le thème de l'avenir linguistique de l'Europe. L'enjeu est la sauvegarde et la promotion du patrimoine plurilingue

de la nouvelle Union face aux menaces qui se profilent à l'horizon. Le présent volume est le pré-rapport des experts, qui a été remis aux participants. Les thèmes abordés sont les suivants :

- Pratiques de diversification linguistique et enseignement;
- Pratiques de diversification linguistique et ingénierie linguistique;
- Un instrument fondamental pour la diversification linguistique : la terminologie.

La dernière rubrique comporte des contributions de Maria Teresa Cabré, Christian Galinski, Anita Nuopponen, Rita Temmerman, Philippe Thoiron et Paul Wijnands.

Des exemplaires de ce dossier sont disponibles auprès des organisateurs (Claude Truchot, Université de Franche-Comté, UFR Lettres, 25030 Besançon Cedex, téléc. (+33)81.66.53.00.

Un des résultats de cette rencontre est la décision de créer un observatoire des pratiques linguistiques en Europe, et une association est en train d'être constituée. La date de lancement prévue est 1996. Les renseignements sur cet observatoire peuvent être obtenus auprès des organisateurs du colloque.

Truchot (Claude) (éd.), *Construire le pluralisme linguistique : quelles initiatives dans l'Union européenne ?* Besançon, 1995.

*Pratiques professionnelles, pratiques*

Le présent volume, dédié à la mémoire de Louis Guespin, constitue les actes du premier colloque des jeunes chercheurs en sciences du langage, et se regroupe autour de quatre thématiques :

- Langage documentaire;
- Interactions professionnelles;
- Interaction homme-machine;
- Rapport au public.

Il est disponible au prix de 80 F auprès de Michelle Van Hooland, URA CNRS 1164, IRED Université de Rouen, 76821 Mont-Saint-Aignan Cedex.

Tsekos (Nicolas), Van Hooland (Michelle), Wable (Thierry) (éd.), *Pratiques professionnelles, pratiques langagières*, Ailleur/Ura CNRS 1164, 1995.

*Manuel de lexicographie spécialisée*

Persévérant dans sa politique d'édition de monographies de qualité en matière de langue spécialisée, l'éditeur John Benjamins nous annonce la parution de cet intéressant ouvrage dont nous publierons prochainement un compte rendu.

Bergenholtz (Henning) et Tar (Sven), éd., *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries*, Amsterdam, John Benjamins, 1995.

*Introduction à la lexicologie explicative et combinatoire*

Cet ouvrage d'Igor Mel'cuk, André Clas et Alain Polguère est une coédition de l'Aupelf-Uref et des éditions Duculot et s'inscrit à la fois

dans la collection «Universités francophones» (Aupelf-Uref) et dans la collection «Champs linguistiques» (Duculot). L'*Introduction à la lexicologie explicative et combinatoire* est un manuel destiné à tous ceux qui s'intéressent à la lexicologie, à la lexicographie et, de façon plus générale, à tous ceux qui ont une passion pour les questions traitant du lexique. Ce livre se veut une systématique des descriptions et des présentations lexicologiques et lexicographiques. On y montre bien que la connaissance explicative des faits linguistiques acquiert sa validité dans un cadre théorique précis, où le crible analyseur reste stable et logiquement explicite. Ces ouvrage présente les concepts clés de la lexicologie et s'inscrit dans la théorie sens-texte.

Mel'cuk (Igor), Clas (André) et Polguère (Alain), *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve, Aupelf-Uref et Duculot, 1995, 256 p. (Universités francophones et Champs linguistiques).

*Traduction et terminologie. Répertoire biographique internationale*

Annoncé dans le dernier numéro de *Terminologies nouvelles*, le répertoire biographique coédité par l'Union latine, Infoterm, Praetorius et International Where+How vient de paraître. Il comporte quelque 2 000 notices consacrées à des personnalités qui se sont distinguées dans le domaine de la traduction et de la terminologie. Une structure normalisée, un glossaire détaillé et des index proposés sur une disquette facilitent la consultation.

L'ouvrage peut être commandé au prix de 660 FRF (soit 105 ECU ou 126 USD) auprès de La Maison du dictionnaire, 98 Bd Montparnasse, F-75014 Paris, Tél. (+33.1)43.22.12.93, téléc. (+33.1)43.22.01.77.

*Realiter sur Internet*

Toute l'information sur Realiter, le Réseau panlatin de terminologie évoqué dans l'article de Giovanni Adamo, est accessible sur Internet. Les pages de Realiter sont hébergées sur le serveur de l'Université Pompeu Fabra, à Barcelone. On peut notamment y consulter les principes terminologiques décrits dans ce numéro.

Adresse : <http://www.iula.upf.es/cpt/cptfr.html>

*Le secrétariat général du Rint sur Internet*

À l'initiative de l'Office de la langue française (Québec), de nombreux renseignements sur le Réseau international de néologie et de terminologie sont disponibles sur Internet.

Adresse : <http://www.olf.gouv.qc.ca/techno/pages/f08.htm>

*Le centre Termisti sur Internet*

Le centre Termisti, centre de recherche terminologique de l'Institut supérieur de traducteurs (Isti, Bruxelles), est désormais présent sur le réseau Internet, où il bénéficie de l'hospitalité du serveur de la francophonie (Refer). Outre des renseignements sur les activités du

centre, ce nouveau site propose la consultation de microglossaires multilingues accompagnés de réseaux notionnels. On y trouve aussi un vaste «infoport» offrant de nombreux renvois vers d'autres sites et un cours intitulé *Abrégé de terminologie multilingue*, illustré de nombreux schémas et tableaux : il s'agit assurément du premier cours de terminologie disponible en français sur Internet.

Adresse : <http://www.refer.fr/termisti/termisti.htm>

# Congrès, colloques, séminaires

• *3<sup>es</sup> Journées internationales de terminologie (Bruxelles, 19-20 avril 1996)*

Thème : Interdisciplinarité et terminologie. Étude des problèmes de communication dans la recherche interdisciplinaire.

Dans le monde de la recherche, on estime que les projets interdisciplinaires ne sont souvent qu'une juxtaposition d'approches, sans articulation aucune.

Ce manque d'intégration ne provient pas uniquement de problèmes d'organisation, de stratégies conflictuelles ou de perspectives propres aux différentes disciplines. Il serait lié pour une grande partie à des problèmes de communication et notamment de vocabulaire. Les chercheurs signalent une synonymie envahissante, une désynonymisation irrégulière, des emprunts entre disciplines, assortis de glissements sémantiques, une polysémie confuse non actée, des néologismes de sens difficilement identifiables.

Ils sont confrontés la fois à un phénomène d'érosion terminologique et à un phénomène de surspécialisation. Le recours à la normalisation pure et simple ne semble pas résoudre les problèmes rencontrés.

En effet, le concept de domaine scientifique ou technique bien délimité, qui sert d'axe pour une structuration du système notionnel d'une discipline, ne peut servir de fil conducteur pour circonscrire le vocabulaire circulant à l'intérieur d'équipes multidisciplinaires. D'autres solutions doivent être élaborées.

Étudier et gérer ce phénomène relève de l'étude des vocabulaires de

spécialité qu'est la terminologie. Organisation : Association européenne des linguistes et des professeurs de langues et Centre de terminologie de l'Institut libre Marie Haps.

Les personnes qui souhaiteraient présenter une communication sont priées de faire parvenir le titre ainsi qu'un résumé d'une page à M. Adrien Hermans, CTB, rue d'Arlon 11, B-1040 Bruxelles, téléc. : +32.(0)2.511.98.37.

Pour tout renseignement complémentaire s'adresser à l'AELPL, Interdisciplinarité et terminologie, 23 rue Cavendish, F-15019 Paris, téléc. : +33.1.42.41.81.96.

• *13<sup>e</sup> École d'été en terminologie (Vienne, 12-16 août 1996)*

Organisée en anglais par Termnet et l'International Institute for Terminology avec la collaboration de H. Picht (Danemark) et de G. Budin (Autriche). Cette année encore, l'école d'été propose un cours de base sur la terminologie et ses applications dans la pratique du travail terminologique et terminographique.

Informations : Termnet, Grüngasse 9/17, A-1050 Vienne, Autriche, tél. : (+44.1)586.77.63, téléc. : (+44.1)586.77.64.

• *Quelle formation pour le traducteur de demain ? (Paris, 6-8 juin 1996)*

Thèmes des séances :  
– Quel cadre théorique pour la formation?

- Texte, contextes;
- Quels outils pour le traducteur?
- Terminologie et traduction : le traducteur a-t-il besoin d'une formation terminologique?
- Langue des signes et interprétation : une formation spécifique?
- La réexpression relève-t-elle d'une pédagogie?

Organisation : École supérieure de traducteurs et d'interprètes (Ésit), Université Paris III - Sorbonne nouvelle.

Inscriptions et renseignements :  
Ésit - colloque international,  
Mme J. Tharaud, Centre universitaire Dauphine, F-75116 Paris,  
tél. (+33.1)44.05.42.12,  
téléc. : (+33.1)44.05.41.13

• *Informatique et formation des «langagiers» (Rennes, 6-7 septembre 1996)*

Thèmes :

- Quelle informatique pour le futur traducteur, terminologue, rédacteur?
- Objectifs et modalités d'exploitation de l'informatique pour les langues;
- Informatique dans les cursus : contenus et agencements;
- Configurations matérielles pour la pédagogie en milieu universitaire;
- Applications du multimédia aux formations du langagier.

Lieu : Université de Rennes II.  
Informations : D. Gouadec,  
6 avenue Gaston Berger, F-35043  
Rennes Cedex  
tél. (+33)99.33.13.37 ou 99.14.16.06  
ou 99.14.17.48.  
téléc.: 99.54.01.99 ou 99.14.16.05.  
Internet : Daniel.Gouadec@uhb.fr

• *3<sup>e</sup> université d'été en traduction (Rennes, 10-12 septembre 1996)*

Réunira les professionnels, les organismes représentatifs, les donneurs d'ouvrage, les formateurs et les étudiants pour faire le point, en sessions plénières, sur l'évolution de la profession, des outils, des méthodes et procédures, et des contextes.

Proposera une formation en ateliers sur les thèmes suivants :

- Le poste de travail du traducteur, les «aides», les «mémoires de traduction»;
- La formation des traducteurs;
- L'assurance de qualité;
- Le contrôle de qualité, la relecture et la révision;
- La pratique de la traduction juridique;
- La pratique de la traduction littéraire;
- La gestion et la conduite de la traduction technique.

Sera suivie d'un colloque international sur la traduction (voir ci-après).

• *Stratégie de gestion et d'exécution des traductions, stratégie de formation des traducteurs (Rennes, 13-14 septembre)*

Thèmes :

- État des lieux et avancées en traduction générale, littéraire, juridique, commerciale, technique;
- Assurance de qualité en milieu professionnel et en milieu universitaire;
- Contrôle de qualité en milieu professionnel et en milieu universitaire;
- Outils de gestion de la traduction;

- Les automates : assistance ou asservissement?
- Profils de formation des traducteurs;
- Objectifs et stratégies de formation des traducteurs;
- Coursus : contenus et organisation;
- Formation à distance des traducteurs;
- Formations universitaires et validation;
- Statut du traducteur.

Lieu : Université de Rennes II, Bâtiment des Langues.

Comité scientifique : J.R. Ladmiral (Paris), Ph. Thoiron (Lyon), P. Briend (ENSCR), D. Gouadec (Rennes), D. Toudic (Rennes), H. Marquant (Bruxelles), E. Lebreton (LCI-France).

Informations : D. Gouadec, 6 avenue Gaston Berger, F-35043 Rennes Cedex

tél. (+33)99.33.13.37 ou 99.14.16.06 ou 99.14.17.48.

télec. : 99.54.01.99 ou 99.14.16.05.

Internet : Daniel.Gouadec@uhb.fr.

• *5<sup>e</sup> Université d'automne en terminologie et phraséologie (Rennes, 17-19 septembre 1996)*

Réunira les professionnels, les organisations représentatives, les donneurs d'ouvrage, les formateurs et les étudiants pour faire le point en sessions plénières sur la production terminologique/phraséologique : acteurs, outils, méthodes, procédures, mises en œuvre.

Proposera une formation en ateliers sur les thèmes suivants :

- Définition d'un cahier des charges pour la terminologie et la phraséologie;
- Méthodologie de la terminographie systématique;
- Méthodologie et outils de la terminographie ponctuelle (pour traduction, rédaction);
- Outils d'aide à la gestion et à la production terminographique;
- Production et gestion systématiques de répertoires, fichiers, bases de données;
- Assurance de qualité en terminologie;
- Systèmes exploitant les terminologies;
- Aménagement terminologique et normalisation (entreprises et institutions).

Sera suivie d'un colloque international sur la terminologie et la phraséologie (voir ci-après).

*Terminologie et phraséologie (Rennes, 20-21 septembre 1996)*

Thèmes :

- Nature des données;
  - Normalisation, standardisation, protection des données;
  - Supports et outils (réseaux, multimédia, bases de données, etc.);
  - Terminologie et formation et/ou formations à la terminologie/terminographie;
  - Enseignements et formations à distance en terminologie (coursus et outils).
- Comité scientifique : N. Kalfon (Madrid), D. Gouadec (Rennes), D. Blampain (Bruxelles), H. Sorensen (Copenhague).

Informations : D. Gouadec,  
6 avenue Gaston Berger, F-35043  
Rennes Cedex  
tél. (+33)99.33.13.37 ou 99.14.16.06  
ou 99.14.17.48.  
téléc.: 99.54.01.99 ou 99.14.16.05.  
Internet : Daniel.Gouadec@uhb.fr.

• *3<sup>e</sup> Université d'automne en rédaction et documentation (Rennes, 24-26 septembre 1996)*

Proposera une formation complète en ateliers sur les thèmes suivants :

- Gestion électronique de documents;
- Outils, standards, modalités;
- Rédaction technique et spécialisée en français;
- Rédaction scientifique et spécialisée en anglais.

Informations : D. Gouadec,  
6 avenue Gaston Berger, F-35043  
Rennes Cedex  
tél. (+33)99.33.13.37 ou 99.14.16.06  
ou 99.14.17.48.  
téléc.: 99.54.01.99 ou 99.14.16.05.  
Internet : Daniel.Gouadec@uhb.fr.

• *6<sup>e</sup> symposium linguistique international : Comparaisons, contrastes, correspondances : le français et l'anglais en terminologie et en langue de spécialité (Halifax, 3-4 octobre 1996)*

Thèmes :

- Problèmes descriptifs, théoriques, pragmatiques de comparaison de terminologie, de traduction, de terminographie, de terminotique, de socioterminologie, d'aménagement terminologique; emprunts, calques,

faux amis, néologie et francisation; emploi du passif, du participe passé, de la forme en *-ing*, syntagmes nominaux complexes, longueur des phrases, phrases complexes, concision, impersonnalité, mots de liaison, lisibilité, isotopie, cohérence et cohésion, ponctuation, capitalisation; autres sujets connexes.

Propositions de communication à soumettre avant le 15 mars 1996.  
Langue du symposium : le français et l'anglais.

Durée des communications : environ 25 minutes plus discussion.  
Publication des actes dans le volume 9 de la revue *Alfa*.

Frais d'inscription : 50 \$ canadiens.

Organisation : Département de français

Université Dalhousie  
Halifax, Nouvelle Écosse  
Canada, B3H 3J5

tél. : (+1)902.494-2430  
téléc. : (+1)902.494-1626

• *La terminologie au service des entreprises (Mons, 24 et 25 octobre 1996)*

Titre : *Au commencement était le terme : la terminologie au service des entreprises.*

Lieu : Université de Mons-Hainaut, École d'interprètes internationaux.

Dates : 24 et 25 octobre 1996.

Langues : anglais, allemand et français.

Thèmes proposés :

- Le rôle de la terminologie dans l'entreprise;
- Développements récents en terminologie;
- Les nouveaux outils terminologiques;

- Exploitation et diffusion des ressources existantes;
- La terminologie, outil de communication des entreprises;
- Réseaux universitaires et formation/spécialisation en terminologie.

Comité scientifique : Margaret Rogers (University of Surrey, Guilford), Teresa Cabré (Universitat Pompeu Fabra, Barcelona), Daniel Gouadec (Université de Haute-Bretagne, Rennes II), Ursula Reisen (DTT, Köln), Jean Klein (Université de Mons-Hainaut), Geoffrey Kingscott (Language International, Nottingham), John D. Graham (Mannesmann + BDÜ, Duisburg), Reiner Arntz (Europäische Akademie, Bolzano), Jean-Marie Dufrasne (Parlement européen, Luxembourg), Kees van der Horst (Commission européenne, Luxembourg), Klaus-Dirk Schmitz (Fachhochschule Köln), Ton Beerden (Université de Mons-Hainaut), Txomin Martinez (École d'interprètes internationaux, Mons), Louis-Yvon Chaballe (Chaballe Communication, Mons)

Organisation : Université de Mons-Hainaut (UMH), École d'interprètes internationaux (EII), Association des traducteurs et interprètes diplômés de l'UMH (Atim).

Résumé de la communication (maximum 1 page) à envoyer pour le 29 février 1996 au plus tard.

Adresse : Prof. Dr. Jean Klein  
École d'interprètes internationaux  
Université de Mons-Hainaut

Av. du Champ de Mars  
B-7000 Mons, Belgique

tél. (+32.65)37.36.21

téléc. (+32.65)37.30.54

Internet : klein@umh.ac.be.

• *TIA-97 : 2<sup>es</sup> rencontres «Terminologie et intelligence artificielle» (Toulouse-le Mirail, 3 et 4 avril 1997)*

Le groupe Terminologie et intelligence artificielle (TIA) organise une rencontre de deux jours (3 & 4 avril 1997) sur le thème «terminologie et intelligence artificielle» et lance un appel à communication. Les problématiques communes à la terminologie et à l'IA concernent au moins quatre disciplines :

- La terminologie;
- La linguistique sur corpus;
- Le traitement automatique des langues;
- L'acquisition et la représentation des connaissances en IA.

Chacune de ces disciplines peut trouver un apport dans la confrontation et la réflexion commune avec les autres disciplines.

Les laboratoires de recherche et les entreprises sont intéressés par des méthodes de constitution de données terminologiques et par les diverses applications qui les mettent en œuvre (traduction, aide à la rédaction, indexation, capitalisation des connaissances, mémoires d'entreprises...).

La participation importante aux premières rencontres TIA, en 95 (plus de 200 personnes), a montré qu'il existe une demande forte dans le sens de cette interdisciplinarité. Afin de mieux répondre à cette attente, TIA-97 sera organisé en deux temps et comprendra :

- Des tutoriels visant à donner les éléments indispensables pour que chaque participant puisse comprendre les objectifs et les méthodes des autres disciplines;
- Des exposés sélectionnés par le Comité de programme.

Le présent appel à communication concerne ces exposés. Les articles retenus devront rendre compte de travaux interdisciplinaires ou de l'intérêt des travaux présentés pour une ou plusieurs autres disciplines.

Thèmes concernés par les rencontres :

- Fondements épistémologiques;
- Terminologie et ontologies;
- Analyse de corpus pour la constitution de ressources terminologiques : approches linguistiques, approches semi-automatiques;
- Modèles de données terminologiques;

- Terminologie et modélisation des connaissances;
- Terminologie et acquisition des connaissances;
- Bases de Connaissances Terminologiques;
- Utilisation des ressources terminologiques : gestion de la documentation, accès à la documentation, mémoire technique.

Calendrier :

- Réception des articles : 11 octobre 1996;
- Réponses aux auteurs : 11 décembre 1996;
- Réception de la version finale des articles corrigés et au format adéquat : 31 janvier 1997.

Modalités de soumission : Les articles devront comporter au maximum 5 000 mots et 14 pages (figures, bibliographie et annexes comprises).

Adresse de l'envoi :

Anne Condamines  
 ERSS, Maison de la Recherche  
 5, Allée Antonio Machado  
 F-31058 Toulouse Cedex, France  
 Internet : [erss@cict.fr](mailto:erss@cict.fr)  
 tél. : (+33) 61 50 36 08  
 téléc. : (+33) 61 50 46 77

# 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, DG TSL, 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<sup>e</sup> 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<sup>e</sup> 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<sup>e</sup> 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<sup>e</sup> 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)