

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

24
décembre 2004

Actes de la vitrine-forum sur l'informatisation des langues

(Montréal, 13 et 14 juin 2003)

Cahiers du Rifal

Réseau
international
francophone
d'aménagement
linguistique

Actes de la vitrine-forum sur l'informatisation des langues

Cahiers du Rifal 24

3

Introduction

13

Interventions officielles

14 Intervention de
Mme Michèle Gendreau-Massaloux,
rectrice de l'Agence universitaire de la
Francophonie

16 Allocution d'ouverture de
M. Isidore Ndaywel, directeur des
langues et de l'écrit à l'Agence
intergouvernementale de la
Francophonie

19

L'état du traitement
informatique des langues en
Francophonie

20 Stéphane Chaudiron :
*Le traitement informatique des langues en
France*

25 Réjean Roy :
*Le traitement informatique des langues au
Québec*

33 André Manseau :
*Le traitement informatique des langues au
Canada*

36 Sylvie Wallez :
*Le traitement informatique des langues
dans la communauté Wallonie-Bruxelles*

44 Sandra Manzi :
*Le traitement informatique des langues en
Suisse*

46 Marcel Diki-Kidiri :
Les langues africaines sur la Toile

49

Conférence principale

50 Jean-Claude Guédon :
*Vers une nouvelle écologie mondiale des
langues*

55

Atelier sur la recherche

56 Marie-Claude L'Homme :
*Formation et recherche dans le domaine du
traitement automatique des langues en
contexte universitaire*

61 Mokhtar Ben Henda :
*Traitement informatisé des langues, langues
nationales et partenariat francophone*

65 Marcel Diki-Kidiri :
*La recherche sur l'informatisation des
langues africaines*

66 Robert Chaudenson :
La graphisation des langues africaines

69

Atelier sur l'industrie

70 Claude Rigault :
*Les enjeux industriels dans le secteur du
TIL : l'expérience d'un spin-off
universitaire*

74 Jean-Patrice Isaac :
*Le secteur du traitement informatique des
langues : un marché frileux*

76 Compte-rendu :
*Traitement informatique des langues et
propriété intellectuelle*

79

Atelier sur les voies de
l'avenir

80 Stéphane Chaudiron :
*L'avancement des travaux est affaire
d'intégration*

81 Velina Slavova :
*Sciences cognitives et traitement
informatique des langues*

84 Guy Lapalme :
*Les voies d'amélioration des outils de
traitement de l'écrit*

89 Laurence Devillers :

*Les défis en matière de traitement de la
parole*

95

Présentations effectuées lors
de la vitrine des produits et
des prototypes

Oceanie.org

Terminotix

Oralys

Multicorpora

Éditerm

Socatra

OTIL

Termdat

Sylex

Le Grand dictionnaire terminologique

Betel

WizArt

109

Déclaration finale
et proposition
d'un plan d'action

Déclaration finale

Proposition d'un plan d'action

114

En bref - Nouvelles du Rifal

Louis-Jean Rousseau :

Suivi de l'Assemblée générale du Rifal.

Activités du Rifal.

Remerciements

L'Agence intergouvernementale de la Francophonie et l'Agence universitaire de la Francophonie remercient Patrimoine canadien, le ministère des Relations internationales du Québec et l'Office québécois de la langue française pour leur appui financier ou logistique à l'organisation de la Vitrine-forum 2003.

Introduction

C'est en octobre 1995, à Bruxelles, que s'est tenu le dernier grand événement francophone en matière d'ingénierie linguistique. Organisé par l'ACCT et l'Aupelf-Uref, avec l'appui de la Communauté française de Belgique et de la Région wallonne, le forum de l'Observatoire du traitement informatique du français et des langues nationales partenaires – l'OTAF – visait, peu de temps après la création du Web, à souligner l'importance du traitement informatique du français et des langues partenaires pour l'avenir de la Francophonie.

Plusieurs années ont passé depuis ce rendez-vous. Comment les travaux universitaires francophones en génie linguistique ont-ils évolué depuis? Le français est-il maintenant bien outillé pour faire face aux défis que pose la transformation de nos sociétés? La production d'outils intelligents de deuxième ou de troisième génération est-elle avancée? Possède-t-on actuellement les dictionnaires électroniques et autres ressources linguistiques nécessaires à la construction de passerelles solides entre le français et les langues partenaires? Le français et les langues du Sud occupent-ils la place qui leur revient sur Internet?

La tenue, en juin 2003, à Montréal, d'une deuxième manifestation sur le traitement informatique des langues (TIL) a permis de répondre à ce genre de questions. Plus important encore, elle a donné à plus de 80 industriels et universitaires, actifs dans le secteur, une occasion unique de discuter des obstacles à l'accroissement de la présence du français et des langues partenaires sur l'inforoute, et à la mise au point d'outils de TIL toujours plus performants.

M. Stéphane Chaudiron, chargé de mission *Information scientifique et technique – Ingénierie linguistique*, au ministère français de la Recherche, a ouvert la Vitrine-forum par une description des enjeux de l'informatisation des langues et une présentation de l'action *Technolangue*, laquelle vise à renforcer la place de la France dans le secteur du TIL. M. Chaudiron a conclu sa présentation en soulignant que si le projet *Technolangue* est actuellement franco-français, il n'a pas à le demeurer.

M^{mes} Sandra Manzi, chercheure à ISSCO, et Sylvie Wallez, responsable de l'Observatoire du Traitement Informatique des Langues et de l'Inforoute (OTIL); MM. Réjean Roy, consultant, André Manseau, directeur de portefeuille au Conseil national de recherches du Canada

(CNRC) et Marcel Diki-Kidiri, chargé de recherche au CNRS, ont ensuite fait le portrait du secteur du TIL en Suisse, dans la Communauté Wallonie-Bruxelles de Belgique, au Québec, au Canada et en Afrique.

Lors d'un déjeuner-conférence intitulé *Vers une nouvelle écologie mondiale des langues*, M. Jean-Claude Guédon, professeur à l'Université de Montréal, a ensuite avancé l'idée que les francophones devraient chercher à profiter de l'occasion présentée par l'explosion des logiciels libres pour renforcer la place du français et des langues partenaires au XXI^e siècle.

L'intervention de M. Guédon a pavé la voie à une table ronde sur les solutions susceptibles de mener à une intensification des travaux de recherche universitaires dans le domaine du TIL et à un accroissement de leur qualité. Présidée par M. André Clas, professeur à l'Université de Montréal, cette table réunissait M^{me} Marie-Claude L'Homme, professeure à l'Université de Montréal; M. Mokhtar Ben Henda, secrétaire général de l'Association internationale des écoles des sciences de l'information (AIESI); M. Marcel Diki-Kidiri, chargé de recherche au CNRS, et M. Robert Chaudenson, directeur de l'Institut de la Francophonie. Les présentations des participants ont permis de constater que les problèmes de recherche éprouvés par les chercheurs du Nord sont bien différents, dans le domaine du TIL, de ceux rencontrés par leurs collègues du Sud.

La deuxième journée de la Vitrine-forum a débuté par la tenue d'une table ronde sur les solutions susceptibles de mener à l'épanouissement des entreprises de génie linguistique et à une hausse de l'offre d'outils de TIL. Présidée par M^{me} Sylvie Wallez, la table a permis à M^{me} Claude Rigault, vice-présidente de Nomino (compagnie québécoise), et à M. Jean-Patrice Isaac, directeur du bureau Canada-États-Unis de la société européenne WizArt, de faire état des défis nombreux que doivent relever les PME actives dans l'industrie. Les interventions des membres de la table ont de plus mené à de nombreuses réactions de la part des participants dans la salle, notamment sur le thème de la propriété intellectuelle.

Lors d'une dernière table ronde, présidée par M. Stéphane Chaudiron, M^{mes} Velina Slavova, professeure à la Nouvelle Université Bulgare de Sofia, et Laurence Devillers, maître de conférences à l'Université de

Paris-Sud et membre du LIMSI, et M. Guy Lapalme, professeur-chercheur au centre RALI de l'Université de Montréal, ont discuté des avenues de recherche de demain dans le secteur du TIL. Leurs présentations ont notamment aidé les participants à se faire une idée de l'état de l'art en matière de traitement de l'écrit et de traitement de la voix.

La rencontre s'est terminée par la production, par les participants, d'une déclaration finale reconnaissant la nécessité, pour « les organismes et les gouvernements de la Francophonie [d']intensifier leurs efforts pour soutenir les chercheurs, particulièrement ceux du Sud, et les industriels actifs dans le domaine du traitement informatique du français et des langues partenaires ». Une proposition de plan d'action a aussi été soumise par l'auditoire.

Quelles seront les retombées de la rencontre? S'il n'en tient qu'aux personnes ayant assisté à la Vitrine-forum, il ne faudra pas attendre longtemps pour le savoir. Celles-ci ont en effet émis le souhait que l'Agence intergouvernementale de la Francophonie et l'Agence universitaire de la Francophonie rééditent l'événement tous les deux ans, de manière à favoriser la cartographie du secteur et le partage de renseignements entre les participants.

Réjean Roy,
consultant et responsable de l'organisation de la Vitrine-forum

Les enjeux de la vitrine-forum¹

Lorsque l'on discute d'informatisation des langues, on parle généralement, en 2003, de deux choses, c'est-à-dire :

- 1) des différentes manœuvres entreprises pour favoriser l'utilisation du français et des langues partenaires, dans l'univers des technologies de l'information et de la communication, inforoute incluse ;
- 2) de la mise au point, par les spécialistes du génie linguistique, d'outils logiciels capables de traiter le français et les langues partenaires un peu comme le fait l'être humain, à l'oral ou à l'écrit.

Français, langues partenaires et nouvelles technologies

L'une des preuves les plus significatives du passage de nos sociétés à l'ère informationnelle est la pénétration grandissante des technologies de l'information dans toutes les sphères de l'activité humaine. Sous une forme ou sous une autre, les ordinateurs et ce qui leur permet de fonctionner, les logiciels, sont de plus en plus présents dans l'entreprise et dans les foyers.

Une telle ubiquité peut avoir des conséquences importantes, sur les plans linguistique, culturel et économique, si elle n'est pas bien gérée. En effet, si la chèvre dont Monsieur Seguin tirait son revenu n'avait pas de langue, pas plus que la pelle des mineurs observés par Frederick Taylor, les instruments que doivent de plus en plus employer les francophones pour gagner leur pain – les produits de traitement de texte, les tableurs, les logiciels de courrier électronique, les outils de programmation, la *tuyauterie* de l'Internet, etc. –, eux, en ont une. C'est en utilisant cette langue que l'utilisateur peut donner à la machine, ou recevoir d'elle, les instructions qui lui permettent d'être productif.

Mais la question linguistique ne se pose pas seulement sous l'angle du *contenant* ; elle se pose aussi sous celui du *contenu*. Par exemple, comme le soulignaient les auteurs du *Plan d'action pour la mise en œuvre des inforoutes* du gouvernement du Québec : « Le niveau d'intérêt [des francophones] pour l'autoroute de l'information sera directement proportionnel à la qualité et à l'utilité du contenu qu'[on] y trouvera. »

1. Présentation tirée des documents de préparation de la rencontre.

Pour faire des ordinateurs et de l'inforoute des leviers économiques, culturels et sociaux dans la Francophonie, il importera donc d'assurer que le français et les langues partenaires *peuvent servir* intégralement dans le secteur informatique et sur le Web, et que ce potentiel est pleinement exploité.

Français, langues partenaires et génie linguistique

On peut définir les entreprises du secteur de l'ingénierie linguistique de la manière suivante : ces sociétés offrent des produits ou des services capables d'effectuer, à des niveaux divers, un traitement automatique ou semi-automatique de la langue, autant parlée qu'écrite.

La production d'outils performants de traitement de la langue est, pour plusieurs raisons, un enjeu majeur pour les francophones. Premièrement, l'augmentation de la capacité des ordinateurs à traiter la langue comme nous le faisons est nécessaire, parce que c'est à l'information qu'est attribuable, dans la nouvelle économie, la majeure partie du prix de vente des produits et des services lancés sur le marché², et parce que l'essentiel de l'information est produit à l'aide de la langue. Autrement dit, il faut traiter l'information de plus en plus efficacement et, pour ce faire, il faut absolument traiter la langue. Des outils comme les moteurs de recherche, les logiciels de dictée automatique pour salles d'urgence, les aides à la rédaction en langue désambiguïsée, et ainsi de suite, traitent la langue et facilitent la gestion que nous faisons de l'information.

Deuxièmement, l'essor de certains produits et services dépend de la mise au point d'outils de traitement de la langue. Ainsi, les logiciels permettant de taper un mot donné à l'aide d'un nombre minimal de chiffres (ex. : « merci » = 6-3-7-2-4) accélèrent l'envoi de messages textuels par le truchement de téléphones mobiles³. De même, la miniaturisation et la baisse des coûts des appareils

2. Selon Michael Dertouzos, professeur au MIT, 50 à 60 % du PNB des États modernes est imputable à des activités de traitement de l'information. Rapporté dans Gates, Bill (1999), *The Speed of Thought*, New York, Warner Books, p. 14.

3. Voir à cet égard le site www.t9.com, où *America Online* présente la technologie T9 mise au point par la société Tegic. T9 fonctionne en français.

d'accès à Internet reposent sur la mise au point de systèmes de reconnaissance vocale de plus en plus robustes⁴.

Troisièmement, l'utilisation des outils de traitement de la langue permet aux organisations qui y recourent d'être plus productives. Par exemple, Bexley, une banlieue londonienne de 200 000 habitants, a créé une base de connaissances administratives pour permettre aux entrepreneurs d'obtenir, sans aide, les renseignements courants dont ils ont besoin lors du démarrage d'une société. Résultat : les fonctionnaires spécialisés de la municipalité ont désormais deux fois plus de temps à consacrer aux dossiers réellement complexes qui leur sont présentés.

Quatrièmement, les outils du génie linguistique sont la composante grâce à laquelle il sera possible dans l'avenir d'intégrer l'ensemble de la population, plutôt qu'une partie de celle-ci, à la société de l'information. La hausse du niveau d'intelligence de la machine permettra, par exemple, aux utilisateurs moins aguerris de communiquer avec celle-ci en français courant plutôt qu'en langage technique, à l'aide de la voix plutôt que du clavier. Faire en sorte que l'ordinateur s'adapte aux besoins de l'utilisateur, et non pas le contraire, revêt une importance critique, étant donné que les recherches récentes ont démontré que l'attitude vis-à-vis des technologies [demeure] le principal déterminant de l'emploi des ordinateurs⁵.

Cinquièmement, les technologies linguistiques, particulièrement les traducteurs automatiques, les aides à la traduction et les dictionnaires électroniques bilingues, rendront possible l'érection de passerelles entre le français et les autres langues. Cela est évidemment crucial sur le plan culturel. En effet, si personne n'était capable ou désireux de traduire les romans des auteurs belges ou africains vers l'italien ou l'anglais, le rayonnement de la littérature francophone en Italie ou dans le monde anglo-saxon serait nul. De même, l'œuvre de Margaret Atwood, de Naguib Mahfouz ou d'Umberto Eco ne serait pas venue enrichir

l'imaginaire des francophones si ceux-ci n'avaient pu avoir accès à sa traduction.

Finalement, il ne faudrait pas oublier que les francophones gagneront à maximiser leur part des ventes mondiales d'outils du génie linguistique. En effet, la création d'un Microsoft français, québécois, tunisien ou autre dans le secteur de l'informatique des langues aurait des effets majeurs en matière de création d'emploi et de richesse.

Problématique

L'accroissement de la valeur économique du savoir et de l'information et la mondialisation de nos économies sont des phénomènes dont les effets seront bouleversants sur le plan linguistique. D'une part, la baisse de productivité – et, donc, de richesse collective – qui résulterait de l'incapacité des francophones à passer avec succès de l'ère industrielle à l'ère informationnelle aurait des répercussions négatives sur le prestige de leur langue pivot. D'autre part, la mondialisation, en entraînant la multiplication des contacts interlinguistiques, accroîtra à un niveau sans précédent la concurrence que se livrent les langues aux quatre coins du monde. En Afrique francophone, par exemple, elle pourra avoir pour effet d'augmenter considérablement la pression à l'utilisation de l'anglais comme langue véhiculaire.

Pour réussir la prise de ces deux virages, les francophones devront réussir à augmenter la maîtrise que l'ordinateur – principal instrument de traitement de l'information – a de la langue française et des langues partenaires. Cela signifie d'abord qu'ils devront avoir accès à du matériel informatique, à des logiciels et à des contenus électroniques de qualité dans leur langue, sur Internet et ailleurs. Cela signifie ensuite qu'ils devront pouvoir tirer profit de l'existence d'outils de plus en plus performants de traitement automatique du français et des langues partenaires.

Il n'y a pas de doute que le traitement informatique du français se soit déroulé à un rythme extrêmement rapide ces dernières années et que la situation soit aujourd'hui meilleure qu'elle ne l'était au milieu des années 1990. Ainsi, le retard des premières années a été comblé et la Francophonie compte actuellement parmi les communautés linguistiques les plus branchées au monde ; en Europe, le

4. Aux États-Unis, il existe maintenant des mobiles jetables. Ces appareils sont vendus sans clavier, ce qui permet d'en réduire le coût. La composition du numéro se fait par reconnaissance vocale.

5. Modahl, Mary (2000), *Now or Never. How Companies Must Change Today to Win the Battle for Internet Consumers*, New York, HarperBusiness, p. xxi.

français est non seulement présent sur les sites Web des organisations francophones, mais, aussi très souvent, sur ceux des entreprises et des organismes étrangers; au Québec, plusieurs acteurs industriels du génie linguistique incapables de trouver du financement par le passé ont reçu ces 24 derniers mois des injections en capital de plusieurs millions de dollars; il ne fait pas de doute maintenant que les moteurs de recherche francophones disponibles sur le Web soient aussi performants que leurs équivalents de langue anglaise, etc.

Malheureusement, certains problèmes demeurent. D'une part, il arrive encore que les internautes francophones ne puissent trouver dans leur langue l'information qu'ils souhaiteraient avoir, ce qui les amène à se tourner vers des sites de langue anglaise⁶. Au Sud, il est clair, selon Marcel Diki Kidiri, que « les langues africaines sont des langues dont on parle abondamment sur la Toile, mais [qu'] elles ne sont pas des langues dans lesquelles on communique sur cette même Toile⁷ ». D'autre part, des défis importants se posent encore dans le secteur du génie linguistique. Ainsi, il existe encore relativement peu de modèles de traducteurs automatiques ayant le français comme langue source ou langue cible; les travaux européens portent le plus souvent sur la traduction de l'anglais vers d'autres langues et des autres langues vers l'anglais. De plus, les recherches sur le traitement du *sens* de la langue en sont encore à leurs débuts, en Francophonie comme ailleurs, et beaucoup de travail reste à faire. Au Sud, le traitement informatique des langues africaines a par ailleurs accumulé « un retard considérable qu'il sera difficile de combler, soutiennent Émile Camara et autres⁸, sans des actions incitatives de grande ampleur » dans le domaine.

Pour résoudre les problèmes qui demeurent et, surtout, relever les défis de demain sur le plan du traitement informatique des langues, diverses mesures devront être

adoptées. Celles-ci se répartissent en différents champs d'intervention.

Normes et standards

En principe, aucun élément d'infrastructure ne s'oppose actuellement à ce que le français soit employé aussi facilement que l'anglais en informatique ou sur Internet. Dans les faits cependant, l'utilisation intégrale, sur Internet et ailleurs, de la langue pivot de la Francophonie – tout comme celle, entre autres, de l'arabe, du russe ou des langues subsahariennes – pose encore des difficultés à cause de problèmes historiques⁹. Il serait important de résoudre ces problèmes une fois pour toutes.

Plus important encore cependant, la Francophonie devra s'intéresser aux normes et standards nouveaux dont l'adoption aura des effets sur la productivité de la langue française et des langues partenaires dans l'univers électronique.

Expertise

Dans les domaines de l'informatique, de la production de contenus électroniques ou du génie linguistique, aucune ressource n'est plus précieuse que le capital humain. En effet, la valeur des idées incorporées dans la conception et la commercialisation d'un logiciel, l'élaboration d'une page Web ou d'un portail, ou encore la création d'un outil de reconnaissance des caractères explique la quasi-totalité de son coût.

Malheureusement, la main-d'œuvre et les chercheurs spécialisés dans ces questions sont, encore, souvent rares.

6. Voir notamment à ce sujet Roy, Réjean et Pierre Georgeault (2003), *Enquête sur la langue des sites Web des sociétés grand public actives au Québec*, Québec, CLF.

7. Diki-Kidiri, Marcel (s.d.), « Les langues africaines dans la société de l'information », intervention présentée dans le cadre du Colloque sur le plurilinguisme dans la société de l'information, Paris, Unesco, consulté sur le Web le 11/11/01, <http://www.unesco.org/comnat/france/colloquemars3.htm#Diki>.

8. Camara, Émile (1995), *Traitement informatique des langues africaines: problèmes et perspectives*, s.l., Centre national de la recherche scientifique, consulté sur le Web le 11/11/01, <http://www.lpl.univ-aix.fr/projects/alaf/ALA1.html>.

9. Ces problèmes ont deux causes principales. D'abord, l'ordinateur et Internet ont été inventés aux États-Unis, et les normes et les standards de l'informatique tenaient seulement compte, au départ, des besoins des utilisateurs anglophones. Ensuite, s'il est relativement facile d'émettre de nouvelles règles favorisant l'usage du français et des autres langues dans le secteur des technologies de l'information, il est beaucoup plus difficile de les mettre en vigueur.

Ainsi, les compagnies québécoises ont de la difficulté à recruter des travailleurs compétents dans le champ de l'ingénierie linguistique. Autre exemple: on retrouve d'excellents techniciens dans les pays du Sud, mais ceux-ci sont encore trop peu nombreux, en partie parce que l'offre de formation en français y est faible.

Cela dit, la question la plus importante demeure la suivante: l'expertise dont la Francophonie a besoin pour travailler au perfectionnement des technologies actuelles et aux outils de traitement du sens de demain existe-t-elle?

Financement

Comme c'est souvent le cas, les questions d'argent expliquent une partie des problèmes éprouvés par les francophones, quand vient le temps de créer des contenus électroniques en français ou dans les langues partenaires. Ainsi, certains sites Web ne sont pas mis sur pied, parce qu'aux yeux des bailleurs de fonds, il serait impossible de les rentabiliser. D'autres sites sont quant à eux laissés en friche en cours de route, parce que l'argent est venu à manquer.

Dans le secteur du génie linguistique, le manque de fonds nuit aussi souvent à la mise au point d'outils de langue française. Par exemple, la grande majorité des entreprises et des centres de recherche québécois actifs dans le domaine souffrent de sous-capitalisation chronique. Cela tient notamment au fait qu'au Québec et au Canada, contrairement à la situation prévalant en Europe, il n'existe actuellement aucune aide étatique spécifique au développement de l'industrie.

Manque d'information

Les intervenants du secteur de l'informatique des langues doivent disposer d'information sur les tendances scientifiques, techniques et commerciales susceptibles d'affecter leur positionnement à connaître le succès. Par exemple, ils doivent connaître: les percées réalisées dans les centres de recherche nationaux et étrangers, les besoins liés à l'apparition de nouveaux standards (ex.: XML) et de nouvelles tendances (ex.: commerce mobile), les créneaux les plus porteurs (traduction automatique ou outils d'aide à la traduction, correcteurs grand public ou correcteurs journalistiques, etc.), et ainsi de suite. Force est

malheureusement d'admettre que l'information mettant en lumière les besoins, les occasions et les risques dans le secteur est souvent difficile à obtenir pour les PME et les petits centres de recherche.

Faiblesse de la demande

En Francophonie, la demande pour des contenus Web de qualité ou des logiciels performants de traitement de la langue est plus faible, toute proportion gardée, que la demande américaine. Cela tient à plusieurs facteurs, dont: la taille absolue du marché francophone, le sous-branchement relatif des internautes francophones; la prise de conscience plus ou moins rapide, de la part des organisations utilisatrices, des mérites du Web et des outils du génie linguistique.

La faiblesse de la demande a évidemment des effets importants sur l'épanouissement des entreprises et des organismes actifs dans le domaine du traitement informatique des langues. Elle fait notamment en sorte que les entrepreneurs locaux ne trouvent pas sur leur marché premier les débouchés favorisant leur croissance rapide et se tournent vers d'autres marchés linguistiques.

Des transferts déficients

La présence sur un territoire de centres de recherche universitaires, gouvernementaux ou privés de haut niveau débouche fréquemment sur la génération d'économies externes, lesquelles peuvent avoir un effet positif sur l'innovation. L'importance des transferts technologiques réalisés par ces centres ne dépend pas seulement de la qualité de leurs recherches, mais aussi des politiques industrielles dont ils se dotent. Ainsi, la force de Silicon Valley s'expliquerait en bonne partie par le fait que l'Université Stanford a bien réussi à échafauder des programmes promouvant les contacts université-entreprises.

À ce niveau, la Francophonie éprouve actuellement certaines difficultés. Par exemple, selon le Groupe de travail sur le traitement informatique du langage, quoique les chercheurs et les laboratoires de France soient «de tout premier plan par rapport à la concurrence internationale, le passage de la recherche fondamentale aux applications

industrielles est le maillon faible [du] dispositif» de génie linguistique¹⁰.

En résumé...

Comme nous venons de le voir, différents problèmes marquent actuellement le départ de la Francophonie dans le secteur de l'informatique des langues. Heureusement, ces problèmes pourront être résolus, en tout et en partie, par la concertation des francophones et l'élaboration des stratégies appropriées.

Quelles seront ces stratégies? Voilà la question à laquelle les participants de la Vitrine-forum sur l'informatique des langues ont été appelés à répondre.

10. Voir à ce sujet le document de la DGLF présentant les activités réalisées en 2000 par le Conseil supérieur de la langue française dans le champ des technologies de l'information et de la communication, consulté sur le Web le 11/11/01, <http://www.culture.fr/culture/dglf/politique-langue/CSLF-rap/Annexe1-part3.html>.

Programme final de la vitrine-forum

Première journée

Accueil des participants

M^{me} Nicole René, présidente-directrice générale de l'Office québécois de la langue française, secrétaire générale du RIFAL

M. Isidore Ndaywel E. Nziem, directeur des langues et de l'écrit à l'Agence intergouvernementale de la Francophonie

M^{me} Michèle Gendreau-Massaloux, rectrice de l'AUF

Conférence d'ouverture sur les enjeux du développement du secteur du traitement informatique des langues (TIL)

M. Stéphane Chaudiron, chargé de mission *Information scientifique et technique - Ingénierie linguistique*, ministère de la Recherche, France

Présentation de l'état du TIL en Amérique du Nord, en Europe et en Afrique

M. Réjean Roy, consultant (Québec-Canada) (coordonnateur)

M. André Manseau, CNRC, directeur de portefeuille (Québec-Canada)

M^{me} Sylvie Wallez, responsable, OTIL (Belgique et inventaire francophone)

M^{me} Sandra Manzi, chercheuse ISSCO (Suisse)

M. Marcel Diki-Kidiri, chargé de recherche au CNRS (Afrique)

Déjeuner. Présentation des enjeux du multilinguisme sur l'inforoute

M. Jean-Claude Guédon, professeur, Université de Montréal

Table ronde sur les solutions susceptibles de mener à une intensification des travaux de recherche universitaires dans le domaine du TIL et à un accroissement de leur qualité

M. André Clas, professeur, Université de Montréal (coordonnateur)

M^{me} Marie-Claude L'Homme, professeure, Université de Montréal

M. Mokhtar Ben Henda, secrétaire général de l'Association internationale des écoles des sciences de l'information (AIESI)

M. Marcel Diki-Kidiri, chargé de recherche au CNRS

M. Robert Chaudenson, directeur, Institut de la Francophonie

Cocktail, discours d'accueil au Québec et visite de la vitrine

M. René Leduc, ministère des Relations internationales du Québec

Deuxième journée

Table ronde sur les solutions susceptibles de mener à l'épanouissement des entreprises actives dans le domaine du TIL et à une hausse de l'offre d'outils

M^{me} Sylvie Wallez, responsable,
OTIL (coordonnatrice)¹¹
M^{me} Claude Rigault, vice-présidente,
Recherche et développement, Nomino
M. André Manseau, CNRC, directeur de portefeuille
M. Jean-Patrice Isaac, directeur du bureau
Canada-États-Unis, WizArt

Table ronde sur les avenues de recherche de demain en IL

M. Stéphane Chaudiron, chargé de mission
Information scientifique et technique - Ingénierie linguistique,
ministère de la Recherche, France
M^{me} Velina Slavova, professeure,
Nouvelle Université Bulgare, Sofia, Bulgarie
M^{me} Laurence Devillers, maître de conférences
à l'Université de Paris-Sud et membre du LIMSI
M. Guy Lapalme, professeur-chercheur,
centre RALI, Université de Montréal

Présentation et adoption d'une ébauche de plan d'action
et de déclaration finale

11. M. François Lim, directeur de l'Institut de linguistique appliquée de l'Université de Bangui, en Centrafrique, devait intervenir lors de cet atelier, mais n'a pu être présent.



Interventions officielles

Intervention de Mme Michèle Gendreau-Massaloux, rectrice de l'Agence universitaire de la Francophonie

Madame la Présidente-directrice générale de l'Office québécois de la langue française, Secrétaire générale du RIFAL,

Monsieur le Directeur des langues de l'Agence intergouvernementale de la Francophonie,
Mesdames et Messieurs les experts,

Tous ceux qui ont travaillé à l'organisation de cette manifestation partagent une conviction : il est urgent d'affirmer l'importance du traitement informatique des langues en Francophonie.

Cette réunion a lieu en un moment où l'outil informatique doit devenir un des secteurs productifs de la Francophonie comme de l'ensemble des langues du monde ; il s'agit aussi d'encourager la possibilité, pour d'autres langues, de créer cet outil. Et si j'ai choisi d'être présente parmi vous aujourd'hui, c'est parce que cette rencontre illustre de façon exemplaire la complémentarité du travail des deux Agences, l'Agence intergouvernementale de la Francophonie (AIF) et son réseau, le RIFAL, et l'Agence universitaire de la Francophonie (AUF) et ses réseaux de chercheurs, et aussi parce qu'il me semble que cette rencontre doit permettre de remédier à des obstacles et d'éclairer une difficulté.

Les obstacles : ils sont linguistiques, mais aussi techniques, comme en témoigne, par exemple, le nombre de variables à rassembler pour parvenir à la reconnaissance vocale d'une langue ou à l'élaboration de procédures capables de produire des logiciels d'aide à la traduction.

La difficulté : nous sommes ici pour dire que nous souhaitons voir le traitement informatique des langues devenir un objet de demande sociale et une priorité institutionnelle. Mais pour parvenir à ce résultat il convient d'abord de dissiper ce que j'estime être un trouble.

Quand l'homme pense aux machines, il y a du trouble. Il y a du trouble à Montréal quand les magnifiques machines automobiles concourent dans le Grand Prix. Les machines ont toujours provoqué non seulement des passions, des projections, mais aussi des peurs. L'ère de la machine au 19^e siècle a été une ère de mythologies optimistes – la machine allait faire le bonheur de l'homme – mais aussi de cauchemars terrifiants : la machine allait dévorer l'homme. Les phantasmes concernant la machine me semblent s'expliquer en partie par le fait que l'homme n'aime pas se reconnaître lui-même comme une machine.

(Pourtant, les porteurs de stimulateurs cardiaques le savent bien, de nombreuses fonctions du corps sont des fonctions machinales, ou machiniques.) Dans ce trouble se conjuguent la compétition entre l'homme et la machine, et la peur d'une fin de l'homme, que pourrait préfigurer l'extension des fonctions de la machine, ainsi que la réduction de l'organisme vivant à un ensemble de machines.

Un second trouble intervient lorsqu'il s'agit de langues. Les machines sont particulièrement suspectes quand elles s'attaquent aux langues, parce que la langue, réputée être le propre de l'homme, se trouve souvent investie d'une dimension métaphysique.

Or la langue aussi est une sorte de machine. Dans sa leçon inaugurale au Collège de France, Roland Barthes, utilisant une autre métaphore, politique celle-ci, pour décrire le même aspect, qualifiait la langue de « fasciste », parce qu'il reconnaissait à tout système linguistique un caractère implacable, hérissé de contraintes dont le sujet ne peut pas se libérer. À Paris, le Prix Édouard Glissant vient d'être remis à un écrivain d'origine grecque qui a fait l'effort d'apprendre une langue d'Afrique, le sango. Ce que Vassilis Alexakis a trouvé de plus remarquable en sango, c'est que là où le français, dans une phrase négative comme par exemple « je ne pourrai pas assister à vos travaux, j'ai d'autres obligations », est dans l'impossibilité de dire autre chose que « je ne pourrai pas » – d'employer les deux termes de la négation qui entourent le verbe et forment avec lui un ensemble indissociable –, le sango, lui, dit l'équivalent de « je pourrai assister à vos travaux ne pas », ce qui est non seulement très intéressant, mais d'une portée considérable : en sango, la pensée du sens reste suspendue jusqu'au bout de la phrase qui « se développe à l'ombre d'un doute »¹². Un système d'informatisation du sango devra évidemment, en termes de logiciels, imaginer des composantes qui tiennent compte de la place de la négation dans le surgissement du sens, et qui seront différentes de celles du français ou de l'anglais. Autrement dit, la langue est un système où l'on peut repérer des éléments mécaniques, ce qui met en évidence, de façon parfois troublante, que dans chaque langue il y a des procédures figées.

12. Vassilis Alexakis, *Les Mots étrangers*, Paris, Stock, 2002.

Une des conséquences les plus importantes de ce trait essentiel, qui en termes industriels a de nombreuses applications – sans doute exposées dans cette vitrine forum –, concerne par exemple les logiciels de traitement des procédures pour la création de petites entreprises en fonction de la réglementation de l'État ou de la province, ou pour répondre à un appel d'offre, etc. Ces procédures sont exprimées en codes, qui reproduisent de manière automatique un jeu de phrases simples, et qui peuvent être utilisés par des machines. Dépasser le trouble de la langue-machine, c'est voir la langue à la fois comme un outil de création, parce qu'elle permet l'invention, la liberté individuelle, l'expression spontanée de l'idiosyncrasie originale que chacun porte en lui, et comme une production régie par des lois formalisables et reproductibles.

Le travail auquel nous sommes conviés est d'autant plus délicat que, comme l'écrit un remarquable chercheur de l'Université Laval, Justin K. Bisanswa, la langue française, en particulier, n'a jamais cessé d'être représentée elle-même de façon hypostasiée. La diversité interne « a été refoulée dans les consciences par une manœuvre de construction » que Justin Bisanswa appelle un « discours unitariste ou essentialiste : un discours qui vise à rendre monolithique aux consciences ce qui n'est objectivement qu'un conglomerat de variétés linguistiques ». ¹³ Ce simple trait justifierait, s'il en était besoin, les travaux que conduit l'AUF avec l'AIF, et l'AIF de son propre côté avec le RIFAL, sur les langues partenaires. Une Francophonie repliée sur une norme unique et sur le refus de son osmose avec des langues partenaires, susceptibles de transformer et d'enrichir la langue française, serait une menace pour la langue française elle-même.

Et si la langue française vit de sa diversité, les machines peuvent et doivent en rendre compte. Vous êtes les mieux placés, vous qui êtes chercheurs, experts, linguistes, techniciens, et percevez quel rôle peuvent tenir les machines, pour affirmer que toute langue vivante, objet de plaisir et de création, peut bénéficier du traitement

informatique des langues. Vous êtes les mieux placés pour comprendre et pour faire comprendre que, mieux l'on connaît, mieux l'on agit et mieux l'on sert. Mieux l'on dissèquera les mécanismes et les techniques, meilleures et plus performantes seront les machines de l'avenir, et mieux on fera pénétrer dans les milieux sociaux les plus divers la capacité des langues à se parler et à se traduire par l'intermédiaire des machines, à s'enrichir des apports extérieurs, et mieux l'on préservera, de ces langues, la capacité d'exubérance et de prolifération. Toute machine ne décrit-elle pas d'elle-même, à la fois les limites de la machine, mais aussi la possibilité de leur dépassement ?

Pari sur la machine suppose de l'audace. Il faut encore de l'audace pour aller jusqu'au bout de ce pari, par la parole mais aussi par l'action, et en particulier en suscitant l'action des investisseurs. Nous ne pourrions réellement faire œuvre utile que si notre rencontre s'assortit d'un projet de recherche-action – tel est bien le projet du RIFAL, comme celui auquel les chercheurs de l'AUF sont invités – et si demain, avec nous, les gouvernements et les États invitent les machines que sont les langues et les machines qui procèdent au traitement informatisé des langues à devenir des outils propres aux hommes, pour les hommes, des outils maîtrisés qui non seulement ne fassent plus peur mais qui soient même capables d'agrandir l'horizon des hommes.

Voilà pourquoi je vous remercie de m'avoir invitée, et en particulier vous, Madame la Présidente-directrice générale de l'Office québécois de la langue française, Secrétaire générale du RIFAL, qui avez ici tenu le rôle de pionnière.

Merci.

13. Justin K. Bisanswa, « La francophonie: mémoire de la langue, langue des mémoires », séminaire virtuel « Mémoires historiques d'ici et d'ailleurs: regards croisés » à l'École des Hautes Études en Sciences Sociales, cf.: http://www.fl.ulaval.ca/celat/histoire.memoire/histoire/coll_grandeshommes_herosnat_Paris.htm.

Allocution d'ouverture de M. Isidore Ndaywel, directeur des langues et de l'écrit à l'Agence intergouvernementale de la Francophonie

Madame la Rectrice de l'Agence universitaire de la Francophonie,

Madame la Présidente-directrice générale de l'Office québécois de la langue française,

Mesdames, Messieurs,

Au nom de l'Administrateur général de l'Agence intergouvernementale de la Francophonie (AIF), M. Roger Dehaybe, je vous souhaite la bienvenue à cette manifestation organisée conjointement par l'AIF, par l'intermédiaire du RIFAL, réseau de l'AIF, et par l'Agence universitaire de la Francophonie (AUF).

Je tiens à remercier l'AUF d'avoir accepté d'apporter son concours à l'organisation et à la tenue de la manifestation, ce qui illustre la capacité des deux opérateurs de la Francophonie d'unir leurs efforts en vue de résultats plus riches et plus significatifs, dans une perspective qui nous tient tous à cœur, la promotion du français et des langues partenaires, par leur traitement informatique.

Je remercie également l'Office québécois de la langue française qui a assuré la conception et l'organisation de cette manifestation.

L'informatisation des langues constitue pour l'AIF une préoccupation importante et figure en bonne place dans sa programmation 2003-2004.

L'AIF s'intéresse à l'ingénierie linguistique depuis de nombreuses années par les travaux du RIFAL, héritier d'un ancien réseau de l'Agence, le RIOFIL¹⁴. On se souviendra notamment de la première grande rencontre en matière d'industries de la langue, l'OTAF (Observatoire du traitement avancé du français et des langues partenaires), qui s'est tenue à Bruxelles en octobre 1995.

La Francophonie considère la diversité linguistique comme un enjeu capital. En effet, en 2001, les ministres de la Culture des États et gouvernements de la Francophonie ont adopté, lors de leur réunion, la Déclaration et le Plan de Cotonou sur la diversité culturelle et linguistique, qui invitent la Francophonie à se pencher sur la question du développement des langues, le français et les langues partenaires, dans un environnement numérique. Cette orientation a été confirmée par le Sommet de Beyrouth, tant

en ce qui a trait à la diversité linguistique qu'au développement des technologies de l'information.

À l'Agence, nous estimons qu'à côté des actions visant à maintenir et à améliorer le statut du français dans le monde, il convient d'assurer sa présence dans le monde des technologies de l'information et de la communication, et de soutenir le développement des outils informatiques qui traitent ou qui utilisent le français, notamment dans les domaines suivants:

- l'informatique documentaire ;
- la création de systèmes experts ;
- les outils d'analyse et de génération de textes ;
- les outils pour la traduction assistée par ordinateur, la lexicomatique et la terminotique ;
- les outils d'aide à la rédaction et à l'édition de textes ;
- les outils de saisie automatique de l'écrit ;
- les outils de traitement automatique de la parole ;
- la localisation ;
- les outils inforoutiers (navigateurs, moteurs de recherche, etc.).

De plus, nous estimons, s'agissant du développement des langues partenaires du français, qui fait également partie des priorités de l'AIF, qu'il est important de travailler à lever les obstacles linguistiques et techniques à l'informatisation des langues africaines et créoles, dans la mesure où l'on peut affirmer aujourd'hui que l'informatisation d'une langue est devenue un facteur déterminant de son usage et de sa survie, comme ce fut jadis le cas de l'imprimerie.

Il existe en effet des obstacles linguistiques, telle l'absence de standardisation de l'orthographe de certaines langues, mais aussi, et surtout, des obstacles techniques, comme la présence de caractères spéciaux pour lesquels il n'existe pas de polices de caractères dans les normes techniques existantes. L'étape suivante consistera à soutenir le développement de contenus informationnels sur support électronique et le développement d'outils d'ingénierie linguistique traitant et utilisant les langues partenaires, dans les mêmes domaines que ceux que je viens d'énumérer pour le français.

Toutes ces préoccupations seront présentes dans la programmation de l'AIF pour le prochain biennium 2004-2005, conformément aux souhaits formulés par tous les acteurs de la Francophonie et des États et gouvernements membres.

14. Réseau international des Observatoires francophones des industries de la langue, créé en 1986 et remplacé, en 2000, par le RIFAL.

Il me reste à souhaiter que cette manifestation fasse l'objet de suivi et qu'elle constitue la première d'une série de rencontres entre les acteurs de l'ingénierie linguistique de la Francophonie, de manière à créer un forum permanent dans ce domaine.

Je vous remercie.



L'état du traitement informatique des langues en Francophonie

Le traitement informatique des langues en France

Je souhaiterais tout d'abord remercier les organisateurs de cette manifestation de me donner l'occasion de présenter le programme français d'ingénierie linguistique *Technolangue* et de présenter les raisons pour lesquelles une telle action a été lancée.

Le gouvernement français a lancé le programme *Technolangue* dans le but de favoriser le développement du secteur français du traitement informatique des langues (TIL). L'essor de l'industrie permettra à la France de relever différents enjeux que je rappellerai brièvement.

Un enjeu de société

Le premier de ces enjeux est politique. Comme vous le savez, les Européens travaillent activement, depuis 50 ans, à la construction de l'Union européenne. Celle-ci compte actuellement 11 langues officielles – dont, bien sûr, le français –, mais à côté de ces langues, on trouve des langues régionales comme le catalan, le basque ou le gallois, et des langues d'immigration comme le turc, l'ourdou ou l'arabe maghrébin. D'ici un an, le nombre de langues dont il faudra tenir compte sera encore plus important. Ainsi, avec l'adhésion de plusieurs nouveaux pays prévue en mai 2004, plus de 20 langues officielles coexisteront sur le territoire de l'Union européenne.

Cela dit, le multilinguisme n'est pas une question qui touche seulement l'Union européenne. Ainsi, une étude récente a révélé qu'en France, cinq millions de personnes parlent une langue régionale en plus du français et trois millions sont issues de l'immigration. Lorsqu'on tient compte de ces données et de la connaissance que les Français ont de langues comme l'espagnol, l'allemand ou l'anglais, on constate que 47% de la population française parle deux langues ou plus.

En théorie, toutes les langues d'Europe devraient jouir du même respect. Cependant, en pratique, certains problèmes importants demeurent. Ainsi, les différentes communautés linguistiques d'Europe n'ont pas toutes un accès égal à des objets communicants aussi importants de nos jours que les téléphones portables, les ordinateurs, et ainsi de suite. Pour corriger la situation, il conviendrait donc de développer des interfaces de plus en plus puissantes et conviviales, c'est-à-dire, forcément, des interfaces tenant compte de la langue.

Un enjeu économique

Le deuxième enjeu lié au développement de l'ingénierie linguistique française est économique. Les chiffres avancés par les consultants doivent toujours être manipulés avec circonspection, notamment parce que chaque cabinet utilise sa propre méthodologie de calcul, mais comme leurs estimations donnent une idée de l'intérêt suscité par un secteur, en voici quelques-uns: selon les prédictions des experts-conseils, les ventes d'outils de traitement de la parole atteindront un milliard d'euros en 2003, tandis que celles de produits et de services de traitement de l'écrit se chiffreront à 400 millions d'euros (il y a une dizaine d'années, les deux montants étaient inversés). Pour sa part, le marché de la traduction humaine, automatique et semi-automatique devrait avoir une taille de 16 milliards d'euros en 2004.

Ces chiffres sont intéressants en soi, mais ils ne disent pas tout. En effet, les outils servant au TIL sont des technologies *diffusantes*, c'est-à-dire que, dans bien des cas, ils ne sont pas vendus en tant que produits ou services finis; ils sont plutôt incorporés, en tant que modules, à d'autres produits ou services. Par exemple, la plupart des logiciels de traitement de texte comprennent aujourd'hui des fonctionnalités de correction orthographique et les «cockpits» d'avions de chasse renferment des modules permettant au pilote de commander par la voix à l'appareil. De même, ce sont les outils linguistiques qui permettent au marché de l'information électronique professionnelle, un marché de 3,5 milliards d'euros en Europe seulement, de fonctionner adéquatement. Sans moteur de recherche de qualité, il serait en effet difficile, voire impossible, de mener des fouilles de données hautement performantes. D'autres domaines d'activité utilisent des logiciels qui intègrent des technologies linguistiques, comme le commerce électronique, le télétravail ou l'intelligence économique.

Quelle est la valeur des implants linguistiques?

Pas toujours facile à dire, estime Stéphane Chaudiron. Ainsi, quand on achète un logiciel de traitement de texte, on ne sait pas vraiment quelle part de son coût est imputable aux fonctionnalités de correction qu'on y retrouve. De même, on ne sait pas toujours quelle est leur importance aux yeux du client.

La mise au point de solutions de TIL est donc un enjeu social et économique important. Force est donc d'admettre que différents défis devront être relevés dans les années à venir. Il est vrai que, depuis trois ou quatre ans, de plus en plus d'outils de TIL sont vendus comme tels, dans les grandes surfaces et ailleurs. Ainsi, consommateurs et entreprises peuvent facilement se procurer des systèmes de dictée vocale, des outils d'aide à la traduction, des logiciels de fouille de données, etc. Cependant, il ne faudrait pas déduire de cela qu'il n'existe plus de verrous technologiques à la mise au point de produits performants. La question du traitement informatique des langues est loin d'avoir été réglée, contrairement à ce qu'on entend parfois dans les milieux politiques – en tout cas dans les milieux politiques français. Il importe donc de ne pas ralentir les efforts déjà entrepris en termes de recherche et de développement.

Mais parallèlement à l'effort de RD qui doit être maintenu, voire amplifié quand on le compare à celui que consentent par exemple les États-Unis, d'autres pistes doivent être explorées pour tenir compte des nouveaux besoins des professionnels (industriels et académiques) et des usagers.

Se pose en premier lieu le problème de la *disponibilité des ressources linguistiques*. Il faudrait en effet d'abord veiller à ce que les acteurs du domaine, aussi bien les acteurs industriels qu'académiques, disposent facilement – où *facilement* ne signifie pas nécessairement *gratuitement* – de ressources linguistiques de bonne qualité. Ces ressources devraient être réutilisables et conformes à des normes claires. Quand on parle de ressources linguistiques, une question très difficile se pose : celles-ci devraient-elles être disponibles pour toutes les langues et pour tous les domaines ? Il y a en gros trois manières de répondre à cette question.

Selon un premier point de vue que je qualifierais de *sympathique*, il faudrait que pour toutes les langues du monde, la couverture soit complète : dans chaque langue, il faudrait des dictionnaires généraux, puis des dictionnaires dans les différents domaines de spécialité, l'aéronautique, le spatial, le nucléaire, etc. Selon un deuxième point de vue – un point de vue sectaire selon moi –, il conviendrait au contraire de limiter le développement de ressources linguistiques à quelques grandes langues véhiculaires. Enfin, il existe un troisième point de vue, celui des gens qui préconisent qu'on y aille d'une couverture plus ciblée.

Comme il est très coûteux de constituer des banques de données terminologiques, des dictionnaires, des glossaires et des nomenclatures spécialisées, il vaudrait mieux produire moins de ressources, mais produire celles qui comptent vraiment pour chaque groupe de locuteurs. Selon moi, cette dernière approche devrait prévaloir.

Toujours dans le domaine des ressources, il faudrait également que tous les acteurs aient facilement accès à des composants logiciels de base. Ces composants constituent un préalable indispensable au développement d'applications à haute valeur ajoutée. Ces modules *génériques* sont des outils permettant la manipulation ou la production de données linguistiques, par exemple des lemmatiseurs, des étiqueteurs, des phonétiseurs, des analyseurs syntaxiques, des outils de traitement du signal, etc.

Évidemment, il convient de veiller à ce que les ressources et les logiciels de base mis au point répondent à des normes et des standards favorisant leur échange et leur réutilisation. Par le passé, les questions de normalisation et de standardisation ont été négligées – souvent par ignorance –, avec pour résultats que même les outils émergents de bons projets pouvaient difficilement être réemployés. Actuellement, des pratiques exemplaires se répandent dans la communauté scientifique, mais moins dans le domaine industriel. Il faudrait corriger ce problème.

Corollaire du point précédent, tout État concerné par le développement du secteur du TIL doit veiller à l'élaboration de normes, de standards et de pratiques favorables à l'informatisation des langues dont il veut favoriser l'épanouissement. Actuellement, plusieurs consortiums ou groupes plus ou moins formels travaillent à l'élaboration de telles règles, mais, malheureusement, les acteurs de la Francophonie en étaient fréquemment absents.

Ces différents modules pourraient être utilisés par les nouvelles entreprises du secteur, de manière à ce qu'elles n'aient pas à « réinventer la roue ». Ils pourraient aussi être employés dans les écoles de formation en ingénierie linguistique, en ingénierie documentaire, en documentation, en veille économique, etc., pour favoriser la réalisation de travaux de recherche.

Enfin, une autre piste pour innover dans le secteur du TIL concerne la mise en place de campagnes d'évaluation qui permettent à la fois de mesurer l'état d'avancement des travaux et de stimuler l'activité de RD. Pendant très longtemps, la France a négligé cette question : peu d'acteurs

réfléchissaient à la manière dont les bancs d'essai devraient se dérouler, aux méthodes d'évaluation, aux protocoles ; ils se contentaient plutôt de participer aux campagnes d'évaluation organisées par les États-Unis (notamment dans le cadre des campagnes TREC – *Text Retrieval Evaluation Conference*) ou, plus récemment, à la campagne CLEF (*Cross Language Evaluation Forum*), financée par la Commission européenne. Il est tout d'abord incontestable que la participation à ces campagnes a été extrêmement positive, en permettant aux équipes françaises d'évaluer leurs résultats en les comparant avec d'autres résultats au plan international. Mais cette participation peut par ailleurs poser quelques problèmes, en particulier sur le plan de la confidentialité des résultats. Il n'est pas certain, non plus, que les protocoles utilisés correspondent bien aux objectifs industriels, économiques et commerciaux français. Enfin, les protocoles utilisés sont, la plupart du temps, *techno-centrés* et ne rendent pas vraiment compte d'autres paramètres, tels que la pertinence des résultats pour l'utilisateur.

Il importe quand même de souligner que l'évaluation des technologies linguistiques est une piste reconnue depuis un certain temps dans le cadre de la Francophonie. En effet, entre 1996 et 1998, l'Aupelf-Uref (l'actuelle Agence universitaire de la Francophonie) avait lancé différentes campagnes d'évaluation dans les domaines de la parole et de l'écrit et permis ainsi de mobiliser la communauté sur ces questions. Le ministère chargé de la Recherche avait également participé à l'organisation de la campagne *Amaryllis*, consacrée à l'évaluation des systèmes de recherche d'information textuelle. Outre des résultats intrinsèquement très positifs, qui ont été présentés dans les Journées des 15 et 16 avril 1997, ces actions ont préparé le volet *Évaluation* du programme *Technolangue*.

La réalisation d'évaluations francophones aura au moins deux effets. Premièrement, elle permettra d'augmenter le niveau de confiance des utilisateurs. Par le passé, certains producteurs y sont allés d'affirmations publicitaires nettement exagérées. Je crois que les usagers auraient une attitude plus sereine face au marché s'ils avaient une meilleure idée des capacités *réelles* des outils de TIL. Deuxièmement, la création de bancs d'essai permettra aux entreprises d'avoir une meilleure idée de la qualité relative de leurs propres outils, cela à un coût moins élevé qu'actuellement, alors que chacun fait cavalier seul.

C'est pour tenir compte des questions précédentes que la France, comme suite à la publication en 2000 du rapport d'André Danzin sur la situation du secteur français du TIL, a décidé de lancer le programme *Technolangue* par le truchement du ministère de la Recherche¹⁵.

Le programme *Technolangue*

Technolangue est un programme coopératif, c'est-à-dire que le ministère de la Recherche exige que chaque projet implique au moins trois partenaires provenant du monde des universités et du monde des entreprises. Cependant, *Technolangue* n'est pas un programme de RD. D'autres structures, en particulier les Réseaux d'innovation, soutiennent en effet les projets de ce type.

Le programme *Technolangue* est structuré en quatre volets. Dans un premier volet, *Technolangue* cherche à stimuler la production et la diffusion de ressources linguistiques et logicielles. Plus précisément, le programme tente de provoquer l'émergence d'une « boîte à outils » contenant les ressources linguistiques et logicielles minimales – corpus oraux, corpus textuels, lemmatiseurs, analyseurs, etc. – nécessaires à l'informatisation du français¹⁶.

Dans un deuxième volet, *Technolangue* met l'accent sur l'évaluation. Plus précisément, le programme vise à créer une infrastructure d'évaluation des outils de traitement de la langue parlée et de la langue écrite – qu'il s'agisse de produits finis ou de modules intégrés à des applications ou des services – et à favoriser une réflexion théorique sur les techniques d'évaluation et de mesure des performances de systèmes de TIL, une réflexion qui permettra de répondre à des questions comme : « Quels sont les protocoles les plus appropriés quand on cherche à évaluer un correcteur ? » Cette réflexion s'impose, puisque les paradigmes actuellement dominants, notamment en système de recherche d'information, ne nous conviennent qu'à moitié.

15. D'autres programmes favoriseront aussi l'essor du domaine. Ainsi, le ministère de la Culture et le ministère chargé de la réforme de l'État français ont été chargés de veiller au développement d'outils de TIL utilisables par l'administration. La mission de former les professionnels français de l'ingénierie documentaire a pour sa part été confiée au ministère de l'Éducation nationale.

16. En anglais, on parle de BLARK, pour *Basic Language Resources Kit*. Voir à ce sujet : www.elda.fr/blark.

Dans un troisième volet, *Technolangue* cherche à aider les organisations à prendre connaissance des normes et des standards développés dans le domaine du TIL, de même qu'à soutenir la participation française à leur constitution. Il nous a paru important d'appuyer financièrement la présence française aux tables internationales de normalisation et de standardisation, présence traditionnellement faible. Il nous a aussi paru important de veiller à ce que les résultats des négociations menées aux tables soient directement transférés chez les acteurs, pour éviter que ceux-ci ne développent des outils ou des ressources sur la base de spécifications non normalisées.

Dans un quatrième volet, *Technolangue* vise à appuyer la mise en place d'un dispositif de veille qui aide à faire connaître l'état de l'art dans le domaine du TIL et l'usage qui peut être fait des applications de traitement des langues. Le dispositif associe l'ensemble des acteurs, notamment les acteurs associatifs (le syndicat professionnel APIIL et les sociétés savantes AFCP et ATALA), à l'effort de veille. Les résultats de la veille sont en bout de ligne diffusés par Internet sur le portail *Technolangue.Net*.

Le premier appel *Technolangue* s'est terminé en 2002. Un comité de pilotage, composé de 15 personnes provenant du monde de l'écrit et du monde de l'oral, de l'univers universitaire et de l'univers industriel, a été chargé de retenir les meilleures propositions de projets.

Sur 52 propositions déposées devant le comité, 27 ont été acceptées. Ces propositions se répartissent ainsi :

- neuf projets portent sur la création de ressources linguistiques. Par exemple, les projets reçus visent la mise au point de dictionnaires spécialisés dans le domaine de l'aéronautique ou de l'automobile, l'élaboration de détecteurs automatiques de noms propres, la conception de corpus de textes alignés, la mise sur pied de corpus de voix d'enfants, etc. Dans chaque projet, le français est pris en compte, mais la plupart du temps, les chercheurs et industriels travaillent en contexte multilingue, en français comme en allemand, en anglais, en italien ou en arabe ;
- cinq projets visent la création d'outils de base. Ces outils respectent le modèle économique du logiciel libre et seront diffusés à la fin du projet ;
- deux grands projets portent sur les normes et les standards (ex. : Normalangue, projet parapluie) ;
- un projet a été accepté dans le créneau Veille ;

- dix projets concernent les questions d'évaluation. De ce nombre, neuf sont regroupés sous le chapeau d'une plateforme appelée Evalda. Evalda a pour objectif la création d'une infrastructure pérenne d'évaluation des technologies de la langue en France, pour la langue française. Par exemple, la campagne CESART visera l'évaluation de systèmes d'extraction de terminologie, entre autres des outils d'extraction d'ontologies et de relations sémantiques, tandis que CESTA visera l'évaluation de systèmes de traduction automatique ayant le français comme langue pivot.

Le ministère de la Recherche et le ministère de l'Industrie investiront chacun cinq millions d'euros dans les projets *Technolangue*. Ces 10 millions d'euros représenteront environ 50% du budget total des 27 projets sélectionnés. Comme on le voit, *Technolangue* n'aurait pas pu fonctionner sans la participation active des industriels. On peut également relever que le programme mobilise plus de 90 participants différents, y compris quelques participants non français qui participent à leurs frais aux différentes actions.

Technolangue est un programme de trois ans. Le ministère de la Recherche espère en effet qu'au bout de trois ans, la viabilité économique des projets d'infrastructure aura été assurée – même s'il est certain qu'il aura toujours à intervenir sur certains points.

En bout de ligne, je voulais en arriver à ceci : le projet *Technolangue* est actuellement un projet essentiellement franco-français, mais il n'a pas à le demeurer.

Déjà, nous avons proposé à un certain nombre de nos partenaires européens de participer à nos efforts dans le cadre d'actions bilatérales. Plusieurs sont intéressés à le faire, comme l'Allemagne, l'Italie, la Norvège et possiblement la Finlande, puisqu'ils ont les mêmes préoccupations linguistiques et économiques que nous. Un tel partenariat est rendu possible par l'existence, au sein du sixième cadre européen de recherche et de développement, d'une nouvelle action (ERA) qui nous donne la possibilité, en tant que ministère, de monter des projets de coopération avec des agences gouvernementales ou des ministères étrangers ayant leur propre programme d'appui au génie linguistique.

Cela dit, la coopération française pourrait dépasser le simple cadre européen. En fait, notre vœu est de travailler de concert avec les acteurs de l'ensemble de la Francophonie. Cette collaboration pourrait porter sur les

quatre thématiques de *Technolangue*, mais aussi sur d'autres thématiques.

Dans certains cas, la collaboration francophone pourrait se faire de manière bilatérale. Par exemple, grâce aux ententes internationales existantes, le Canada pourrait collaborer avec la France à des actions bilatérales financées par la Commission européenne. Dans le cadre du programme de coopération franco-qubécoise, le Québec pourrait, pour sa part, continuer les travaux entamés avec la France dans le secteur du génie linguistique.

Cela dit, il serait intéressant de mettre en branle des programmes multilatéraux si cela représentait un quelconque intérêt pour les acteurs de la Francophonie. Ces programmes pourraient certes viser la réalisation d'activités de recherche transnationales (ex. : lancement d'appels de propositions conjoints), mais ils pourraient également servir à favoriser l'échange d'information et de pratiques exemplaires (ex. : pratiques d'évaluation, pratiques de gestion de programmes), la réalisation d'activités stratégiques (ex. : identification de barrières au développement du secteur) et la mise en branle d'activités communes (ex. : coordination de projets, réalisation d'évaluations communes).

Dans ces différents cas de figure, *Technolangue* pourrait servir de point d'appui. Avant de conclure, je me permets de vous indiquer que l'ensemble des renseignements concernant le programme et, en particulier, les états d'avancement des projets, seront disponibles sur le futur portail *Technolangue.Net*. En attendant, ces renseignements sont accessibles sur le site du ministère à l'adresse <http://www.recherche.gouv.fr/technolangue>.

Je vous remercie de votre attention.

Stéphane Chaudiron,
chargé de mission Information scientifique et technique – Ingénierie
linguistique, ministère de la Recherche, France.

Le traitement informatique des langues au Québec¹⁷

Nous vivons dans un monde où les mets libanais, les programmes informatiques indiens, les émissions de télévision américaines, les textiles chinois, les travailleurs de la ferme mexicains, les musiques créoles ou les moteurs d'avion québécois circulent plus librement que jamais.

L'enchevêtrement croissant des économies de la planète a évidemment des effets considérables sur le plan linguistique. Par exemple, la Commission européenne demandait récemment à la France de cesser d'imposer l'utilisation du français pour l'étiquetage des denrées alimentaires vendues sur son territoire¹⁸. Dans une veine différente, certaines sociétés transnationales imposent désormais à leurs employés non anglophones d'utiliser l'anglais dans leurs communications internes.

Dans une économie de plus en plus ouverte, le combat pour la promotion du français passera par l'adoption de toutes nouvelles stratégies. Entre autres, l'appui de l'État à l'industrie du traitement informatique des langues (TIL) pourra constituer, à moyen ou à long terme, une façon efficace de renforcer la place du français au Québec.

Mais quelle est la situation actuelle du secteur du TIL au Québec? Le secteur a-t-il évolué depuis le dernier portrait qui en a été tracé au milieu des années 1990 par l'Observatoire québécois des industries de la langue? C'est ce que nous chercherons à voir ici.

Grille d'analyse

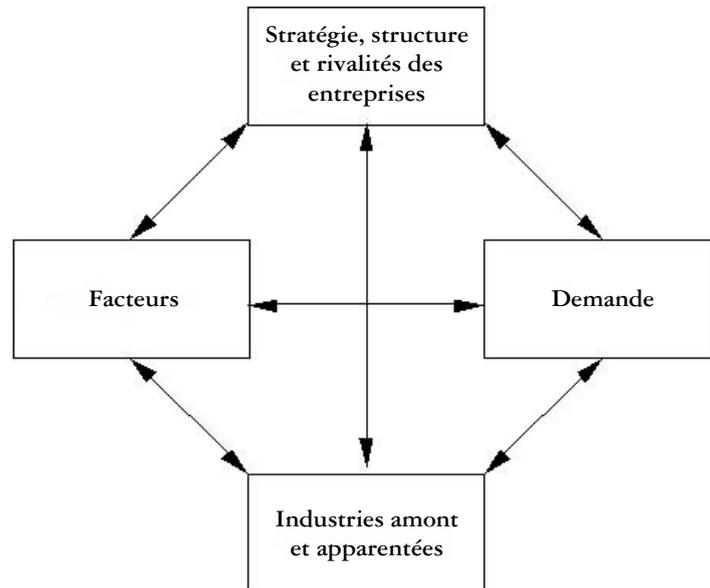
Afin de vérifier l'état de l'industrie québécoise du TIL, nous aurons recours à une version légèrement adaptée de la grille mise de l'avant par le professeur Michael Porter, dans l'ouvrage *L'avantage concurrentiel des nations*¹⁹. Conformément à cette approche, nous décrirons de quelle façon les

conditions de développement présentées dans le schéma 1 se retrouvent au Québec. Nous examinerons aussi les moyens gouvernementaux mis en œuvre pour appuyer l'épanouissement des IL.

Selon le professeur Porter, l'avantage concurrentiel d'une industrie nationale dépend essentiellement de quatre éléments: des *facteurs de production* auxquels ses entreprises ont accès; de l'état de la *demande* pour leurs produits; de la force des industries placées *en amont* et des industries *apparentées*; et de sa *structure*, de la *rivalité* qui y prévaut et des *stratégies* des acteurs.

Pour connaître l'état de ces différents éléments au Québec, nous avons d'abord réalisé une revue des écrits disponibles. Nous avons ensuite mené une série de dix entretiens, de 60 à 90 minutes, auprès de chercheurs, d'industriels et d'investisseurs dans le secteur du TIL²⁰. Voici les résultats de cette enquête.

Schéma 1



17. Cet article est une synthèse du rapport de Roy, Réjean et Pierre Georgeault (2003), *Le traitement informatique des langues au Québec*, Québec, Conseil supérieur de la langue française.

18. Voir notamment Anonyme (2002), «Les consommateurs tiennent à l'étiquetage en français de leurs produits», *La Charente libre*, 30 juillet, consulté le 20 novembre 2002 à l'adresse http://www.langue-francaise.org/Articles_Dossiers/Etiquetage_francais_produit_charente.html.

19. 1993, Paris, Dunod.

20. Neuf des dix personnes interviewées l'ont été en personne (la dixième l'a été par téléphone). Les entretiens se sont déroulés au printemps 2002. Compte tenu de la taille réduite du secteur, il est possible d'affirmer que les personnes rencontrées provenaient d'un échantillon représentatif des entreprises et des centres de recherche québécois actifs en industries de la langue.

Facteurs de production

Ressources humaines

En 1995, les auteurs du rapport *Panorama des industries de la langue*²¹ relevaient qu'il était difficile, pour les entreprises, de recruter les travailleurs dont elles avaient besoin. Depuis cette étude, la situation en matière de ressources humaines ne semble pas s'être améliorée. En effet, les industriels rencontrés continuent de déplorer qu'il leur soit très difficile de trouver sur le marché de l'emploi des travailleurs possédant *simultanément* le profil informatique et le profil linguistique. Conséquence : « Nous devons former le personnel nous-mêmes », affirme un haut dirigeant.

Le problème vient d'abord de ce qu'il n'existe toujours pas de programme interdisciplinaire en TIL au Québec. Il n'est donc pas possible pour un étudiant d'acquérir l'expertise *à la fois* informatique et linguistique dont il a besoin pour réussir dans le secteur. Le problème vient ensuite de ce que relativement peu de personnes sortent chaque année des universités avec le bagage académique requis pour rapidement avoir des répercussions en entreprise. En 2002, on pouvait estimer que seulement une vingtaine de personnes étudiaient dans le secteur du TIL au niveau de la maîtrise ou du doctorat. Heureusement, les étudiants étrangers constituent un apport substantiel pour le secteur. Ainsi, on retrouve plusieurs personnes formées en France dans les entreprises du Québec en industries de la langue.

Savoir scientifique et technique

Le nombre d'entreprises québécoises actives dans le secteur du TIL demeure réduit plus d'une décennie après sa naissance. En fait, en incluant les sociétés étrangères actives au Québec, une douzaine ou quinzaine de PME seulement étaient actives dans le domaine en 2002. Celles-ci incluaient notamment Alis (traduction automatique), Convera (repérage translingue), Delphes (repérage), Druides (correction automatique), John Chandieux experts-conseils (terminologie), Lingua Technologies (aides à la traduction), Locus Dialogue (reconnaissance de la voix) et Nomino (repérage et relation client).

Le Québec compte sur moins d'une dizaine de pôles de recherche en TIL. La majorité des équipes québécoises travaillent à l'informatisation du français. Beaucoup œuvrent aussi au traitement informatique de l'anglais et d'autres langues. Les pôles québécois vont du Centre de recherche informatique de Montréal (CRIM, centre actif dans le domaine de la reconnaissance vocale) au Laboratoire de recherche appliquée en linguistique informatique (RALI, centre affilié à l'Université de Montréal qui se préoccupe notamment de traduction assistée).

Fiset et autres ont bien montré, dans une étude récente, les nombreux obstacles qui se posent, au Québec, à la collaboration des entreprises et des centres de recherche :

Concernant les facteurs qui constituent un frein à l'implication des universités dans des relations avec l'entreprise, les documents des intervenants québécois pointent de manière presque unanime en direction des *différences culturelles*. Pour tous les intervenants, l'arrimage des deux cultures est faible et c'est là un obstacle majeur à des relations fréquentes et fructueuses²².

Comme le prouvent les commentaires que nous avons recueillis, il est clair que les acteurs du secteur du TIL éprouvent l'ensemble de ces difficultés. À titre d'exemple, comme l'a souligné un industriel québécois, « les centres de recherche québécois font souvent des trucs fantastiques, mais le problème, c'est qu'ils travaillent à un rythme bien différent du nôtre. Les résultats de leurs travaux se font seulement sentir à long terme, alors que nous devons produire à court terme ».

La *gourmandise* des centres de recherche serait un autre obstacle à la collaboration des entreprises avec les universités. Dans les mots d'un entrepreneur, « les chercheurs manquent d'argent, alors ils voient en nous une source de financement et nous en demandent plus que nous ne pouvons en investir. Tant que nous serons perçus comme une solution aux ennuis financiers des centres de recherche, l'arrimage sera difficile ». Évidemment, il y a un autre côté

21. Roy, Réjean et Pierre Georgeault (1995), *Panorama des industries de la langue*, Québec, Observatoire québécois des industries de la langue.

22. Fiset, Liette et autres (1999), *Institutions du savoir et PME : synthèse des écrits et bilan des problèmes et besoins des intervenants québécois*, INRS-Urbanisation, rapport d'étape présenté à Développement économique Canada, p. 19.

à la médaille. Par exemple, comme l'a mentionné un chercheur, « tant qu'il ne s'agit que d'investir du temps, les entreprises sont intéressées à collaborer avec nous; elles sont cependant beaucoup plus réfractaires quand il s'agit de mettre de l'argent sur la table ».

Dans certains cas, il serait aussi difficile, pour des motifs juridiques, de transférer la propriété intellectuelle des universités à l'entreprise. Il est clair que les entreprises et les centres de recherche québécois ont moins d'expérience à cet égard que leurs contreparties américaines. Cela explique que, dans certains cas, la création de *spin-offs* chargés d'exploiter commercialement les travaux des chercheurs soit plus complexe et plus longue à réaliser que souhaitable.

Cela dit, il est indéniable que les travaux de recherche réalisés au Québec ces quinze dernières années ont mené à d'importants transferts de connaissances vers l'entreprise. Ainsi, les travaux du CRIM en reconnaissance vocale ont mené à la création d'un important *spin-off*, Locus Dialogue; ceux du Laboratoire en informatique cognitive et environnements de formation (LICEF) ont débouché sur la fondation de Nomino; les travaux du professeur Jacques Ladouceur ont mené à la conclusion d'une importante entente de transfert de propriété intellectuelle entre l'Université Laval et la société nStein, spécialiste de l'indexation assistée par ordinateur; etc.²³

Capital

En matière de recherche, la baisse du financement des universités représente un problème crucial. Par exemple, un centre de pointe comme le RALI doit presque entièrement s'autofinancer²⁴. Pour ce faire, il peut emprunter l'une de trois voies. Il peut chercher à obtenir des fonds auprès d'organismes subventionnaires québécois ou canadiens tels le Conseil de recherche en sciences humaines (CRSH) et le Conseil de recherches en sciences naturelles et en génie du

Canada (CRSNG); d'organismes subventionnaires étrangers et de clients pour le compte desquels il mène des recherches ou auxquels il cède des éléments de propriété intellectuelle.

Chacune de ces avenues comporte des embûches. Ainsi:

- les chercheurs déplorent que les recherches dans le champ du TIL tombent souvent *entre deux chaises*: par exemple, pour une institution comme le CRSH, les travaux en linguistique informatique relèveraient plutôt des organismes subventionnaires actifs dans le domaine des sciences pures, tel le CRSNG, tandis que, pour ces derniers, ils tombent bien sous l'autorité du CRSH;
- les centres de recherche québécois peuvent parfois participer à des projets soutenus financièrement par des gouvernements étrangers, à condition d'investir eux-mêmes de l'argent dans l'opération; en théorie, certains programmes québécois et canadiens prévoient l'accord de subventions aux participants à de tels projets; dans la pratique, ces subventions seraient toutefois difficiles à obtenir;
- lors du montage de projets, il arrive que l'obtention d'une subvention dépende de ce que le centre de recherche trouve des partenaires industriels capables d'y aller d'une contribution en argent sonnante; il serait fréquent que les entreprises sollicitées ne puissent ni ne veuillent le faire.

Le financement des entreprises dans le secteur du TIL est aussi un enjeu, mais celui-ci est un peu différent.

Ces dernières années, les compagnies du secteur ont eu relativement peu de problèmes à amasser des sommes importantes pour mettre au point leurs prototypes ou leurs premiers produits. Par exemple, à la fin des années 1990, Nomino recevait plus de 3 millions de dollars en capital de risque de fonds comme T2C2, pour transformer un prototype de moteur de recherche en produit fini. En trois ans, Semantix a pour sa part reçu plus de 10 millions de dollars d'investisseurs institutionnels, comme la Caisse de dépôt et de placement du Québec.

Par contre, les PME ont eu accès plus difficilement aux fonds leur permettant de passer à haute vitesse au stade de la commercialisation. Comme l'exprimait un financier fortement familiarisé avec le secteur: « Nous avons fait du bon travail en ce qui concerne le financement initial des PME de l'industrie, mais l'enjeu est maintenant de les aider à croître, c'est-à-dire à prendre pied sur les marchés étrangers et à multiplier leur chiffre d'affaires. Et de ce côté, la bataille est extrêmement difficile ».

23. nStein a depuis décidé – unilatéralement – de mettre fin à cette entente.

24. L'Université de Montréal met toutefois des locaux et de l'équipement à la disposition du centre de recherche.

Le problème a plusieurs sources :

1. Il est clair que, de manière générale, les investisseurs, y compris ceux qui investissent du capital de risque, sont plus frileux qu'ils ne l'étaient il y a quelques années ;

2. Souvent, les investissements réalisés dans le secteur n'ont pas donné les résultats escomptés ; par exemple, un géant du domaine comme Lernout & Hauspie s'est effondré complètement, tandis que la capitalisation de grandes sociétés, comme Nuance ou Speechworks, a fondu au soleil. Cela a eu pour effet d'échauder les investisseurs et de les amener à regarder ailleurs.

3. Pour certains, le Québec a le mauvais réflexe de *saupoudrer* les fonds disponibles. « Plutôt que de donner un peu d'argent à tout le monde, il faudrait en donner beaucoup aux joueurs les plus prometteurs », soutient un financier. « Pour sept millions de personnes, nous comptons sur trois ou quatre producteurs de moteurs de recherche. Peut-être qu'il y en a un ou deux de trop là-dedans ».

Cela dit, certaines entreprises ont récemment réussi à mettre la main sur des sommes relativement élevées pour intensifier leurs efforts en matière de commercialisation. Ainsi, en mars 2002, le Fonds de solidarité de la FTQ annonçait un investissement de 10 millions de dollars dans la société nStein, compagnie publique²⁵.

Information de marché

Malheureusement, l'information en français mettant en lumière les besoins, les occasions et les risques dans le secteur du TIL est souvent difficile à obtenir pour les entreprises. L'arrêt complet des activités du Centre d'expertise et de veille inforoutes et langues (CEVEIL), en 2001, signifie notamment qu'aucun organisme public ou parapublic québécois n'a actuellement pour mission de faire de la veille permanente et de la sensibilisation dans le domaine.

Selon nos rencontres, on peut toutefois penser que les producteurs québécois ne souffrent pas de l'absence

25. Plus précisément, le Fonds acquérait pour 6 millions de dollars d'actions ordinaires de la compagnie. Il s'engageait également à acheter pour 4 millions de dollars d'actions additionnelles de l'entreprise, moyennant l'atteinte, par celle-ci, de certains objectifs financiers.

d'information en français sur le secteur du TIL. Celle-ci est disponible en anglais, ce qui semble leur suffire. S'il y a un manque, il se situe probablement davantage du côté de l'information neutre vulgarisée à l'intention des utilisateurs potentiels.

Matériaux langagiers

Selon un chercheur, la *disponibilité* des matériaux langagiers (ex. : analyseurs morphologiques) et des ressources informatiques (ex. : filtres PDF), nécessaires à la mise au point d'outils de traitement du français, fait parfois problème. « Souvent, ce dont j'ai besoin existe, mais m'est inaccessible. Au niveau de la recherche, il faudrait que les chercheurs partagent davantage. Il en coûte trop cher de travailler chacun dans son coin ».

D'autre part, il arrive que les matériaux requis *n'existent pas*. Par exemple, un chercheur souligne l'absence quasi complète de corpus oraux en langue québécoise, laquelle est un frein sérieux à la mise au point d'outils de reconnaissance ou de synthèse du français parlé ici. « Si ces banques ne sont pas mises à notre disposition, les logiciels dont les Québécois ont besoin ne seront pas créés ou, encore, ils le seront par des compagnies étrangères – des compagnies avec les moyens et la volonté de produire les corpus requis. »

La demande

Il est toujours plus facile pour un entrepreneur de réussir lorsqu'il trouve près de chez lui un marché réceptif. Par exemple, avant de se lancer à la conquête du monde, le Cirque du Soleil a reçu un appui enthousiaste du public québécois. Malheureusement, la demande locale pour les outils de traitement des langues produits par les entreprises québécoises paraît représenter un obstacle substantiel à l'essor des entreprises québécoises du secteur des IL. Ainsi :

- les outils de TIL sont souvent destinés davantage aux entreprises d'envergure qu'aux consommateurs ou aux PME. De plus, le nombre de compagnies québécoises de grande taille, particulièrement celui des grandes sociétés requérant des outils de traitement du français, est relativement limité. « Il n'y a pas de doute : le gros de notre clientèle potentielle se trouve et continuera de se

trouver à l'extérieur du Québec», nous a confié un industriel;

- les utilisateurs québécois ne sont pas toujours conscients qu'ils ont un besoin que les produits de TIL sont en mesure de combler. « Il faut faire beaucoup d'évangélisation, affirme un entrepreneur. Je n'ai pas seulement à convaincre les acheteurs que mon produit est le meilleur de sa catégorie; il faut également que je les persuade qu'il leur servira à résoudre adéquatement le problème qu'ils ont »;
- dans certains cas, la demande pour les produits des industriels québécois du secteur de la langue est faible parce que certains utilisateurs potentiels n'ont pas de problème réel à résoudre. À titre d'exemple, il existe sur le marché des robots virtuels capables de répondre de manière automatisée à un courriel. Cependant, comme peu d'organisations québécoises font face à un nombre élevé de courriels dans une journée, la demande locale pour ce type d'outils demeure limitée;
- du côté des administrations publiques, la demande s'est aussi manifestée *timidement*. À ce jour, celles-ci se sont à peu près contentées d'acquiescer pour leurs sites Web ou leurs intranets. Les autres produits de TIL sont peu utilisés par les gouvernements du Québec et du Canada, ou encore par les municipalités. Ainsi, selon un industriel, « la demande pour nos outils ne s'est jamais vraiment manifestée ici, malgré l'existence de besoins importants du côté des ministères. Nous réalisons la presque totalité de notre chiffre d'affaires auprès du gouvernement américain ».

Secteurs en amont et secteurs apparentés

Dans les années 1980 et 1990, la force du secteur téléphonique au Québec a eu des effets positifs sur l'essor du secteur de la reconnaissance vocale. Les premiers travaux dans ce domaine ont effectivement été menés par les unités de RD de compagnies comme Bell. De même, la prospérité de l'industrie québécoise de la traduction humaine a eu des retombées sur le développement du sous-secteur québécois de la traduction automatique et de la traduction assistée par ordinateur. Cependant, on peut peut-être déplorer que les grands cabinets d'informatique-conseil québécois se soient toujours relativement peu préoccupés du développement du secteur national du TIL.

Dans un autre ordre d'idées, certaines normes mises en place dans des secteurs apparentés – XML et VoiceXML par exemple – devront être prises en considération par les entreprises québécoises actives dans le secteur du TIL. Cependant, le Québec et les autres pays francophones ne sont pas toujours présents aux tables où se discute l'avenir de ces normes.

Structure, stratégie et rivalité

L'industrie québécoise du TIL est constituée d'un nombre réduit de PME et de petits centres de recherche. Dans l'ensemble, ces joueurs se concertent lorsque cela peut rapporter. Par exemple, Semantix collabore avec Alis, parce que l'offre des deux entreprises en matière de traduction est complémentaire plutôt que concurrente.

Sur le plan de la stratégie, on peut noter les éléments suivants:

- les entreprises québécoises dans le secteur du TIL travaillent, pour la plupart, à l'informatisation du français et de l'anglais; de manière générale, le marché francophone est perçu comme étant trop petit pour assurer la survie d'un producteur;
- de plus en plus, les entreprises québécoises du secteur du TIL offrent des services linguistiques traditionnels (ex.: traduction humaine) à leur clientèle, *en plus* ou *au lieu* de produire des outils de traitement automatique des langues. Elles adoptent cette mesure pour des raisons essentiellement financières: dans de nombreux cas, leur direction ou conseil d'administration perçoit que la vente de produits de TIL ne suffira pas à assurer leur avenir à court ou moyen terme, et que d'autres moyens de produire des liquidités doivent être prévus;
- la majorité des entreprises et des centres de recherche québécois travaillent au traitement de la langue écrite plutôt qu'à celui de la langue parlée, bien que le secteur de la reconnaissance et de la synthèse de la voix soit actuellement en pleine expansion, en raison, notamment, de l'impopularité des solutions de type clavier (« Pour avoir des renseignements sur votre facture, faites le 8 »);
- la plupart des entreprises québécoises commercialisent leurs outils linguistiques auprès d'utilisateurs finaux (entreprises ou consommateurs), plutôt qu'auprès de sociétés intéressées à les intégrer à d'autres logiciels

(ex. : intégration d'un correcteur à un logiciel de traitement de texte).

L'action du gouvernement du Québec

Jusqu'à 1995-1996, le gouvernement du Québec a employé diverses méthodes pour appuyer le développement des IL. Par exemple, il a permis aux entreprises du secteur du TIL – comme aux autres entreprises de logiciel – de déduire une part importante de leurs dépenses salariales en RD. Dans certains cas, il a aussi utilisé son pouvoir d'achat : c'est, par exemple, grâce à une commande initiale (Exploratexte) du ministère de l'Éducation du Québec qu'a pu être développé le Correcteur 101 de la firme Machina Sapiens.

Depuis, ses principales actions dans le domaine ont inclus l'offre, par le Fonds de l'autoroute de l'information (FAI) du Québec, d'une subvention de 400 000 dollars pour la production d'un moteur de recherche francophone de *deuxième génération* (cette subvention a mené à la création de la société Nomino), l'utilisation *très occasionnelle* de son pouvoir d'achat (ex. : le ministère des Ressources naturelles a placé des correcteurs sur chaque poste de travail pour améliorer la qualité de rédaction des contenus déposés dans l'intranet ministériel) et la création, par l'Office québécois de la langue française (OQLF), de bancs d'essai visant à mesurer le niveau d'adaptation des logiciels commerciaux aux attentes des francophones.

Discussions

Nous venons de le voir, la situation du secteur du TIL n'est pas rose au Québec. Des centres de recherche sous-financés ont de la difficulté à avoir accès aux matériaux langagiers dont ils ont besoin ; bien souvent, des producteurs de petite taille n'ont pas les ressources pour développer et commercialiser leurs produits et services convenablement dans un marché d'acheteurs frileux ou peu sensibilisés, ce qui les force à investir des créneaux plus ordinaires pour boucler les fins de mois ; etc.

Diverses solutions seraient susceptibles de favoriser le développement du secteur au Québec. Certaines de celles-ci sont de nature financière, mais plusieurs visent

essentiellement la création de *conditions propices* à la mise au point et au lancement d'outils de traitement de la langue. Voici ces solutions.

Assurer la croissance du bassin de travailleurs

Il ne fait pas de doute que la principale façon pour le Québec de conquérir un avantage concurrentiel dans le domaine du TIL sera le développement d'une main-d'œuvre de qualité, la rétention de cette main-d'œuvre et l'attraction de cerveaux étrangers. Nous avons vu que les compagnies québécoises ont souvent de la difficulté à recruter des ressources humaines compétentes dans le champ des industries de la langue.

Pour pallier ce problème, le gouvernement du Québec pourrait notamment instaurer un programme visant l'offre de bourses d'études spécialisées aux étudiants se consacrant à des travaux de recherche dans le secteur de l'informatique des langues. Il pourrait aussi accorder des contrats de recherche aux professeurs d'université actifs dans le secteur du TIL, notamment en raison des retombées de tels contrats sur la formation et les choix de carrière futurs des étudiants.

Veiller au financement adéquat des acteurs

La presque totalité des entreprises québécoises actives dans le secteur de l'informatique des langues sont de petite taille et manquent de fonds pour mener des efforts de commercialisation d'envergure au Québec et à l'étranger. Pour corriger cette situation, le gouvernement pourrait chercher à orienter ses programmes d'aide et les efforts des sociétés de capital de risque qu'il appuie, de manière à ce que davantage d'argent soit accordé à un nombre plus réduit de compagnies.

Cela dit, le gros des efforts de l'État devrait sans doute viser à assurer aux centres de recherche en TIL un financement adéquat de leurs efforts. Par conséquent, en plus des mesures précisées dans la section précédente, le gouvernement devrait particulièrement faire de la sensibilisation et des représentations pour que les organismes québécois et canadiens concernés par les questions de recherche multidisciplinaire créent des programmes d'aide auxquels les centres de recherche seront admissibles.

L'État québécois devrait aussi tirer profit de l'ouverture manifestée par l'Union européenne à l'égard de la participation de centres de recherche étrangers à des projets communautaires et appuyer financièrement les centres de recherche désireux de participer à des projets de recherche internationaux en matière de TIL.

Assurer la disponibilité des matériaux langagiers

Nous avons vu que les chercheurs et industriels québécois dans le domaine du TIL n'ont pas toujours à leur disposition les matériaux langagiers dont ils auraient besoin pour leurs travaux. Comme ces matériaux sont souvent coûteux et peuvent servir à l'ensemble de la communauté, le gouvernement du Québec devrait appuyer financièrement la production de matériaux langagiers pour le français québécois et assurer que les ressources langagières précommercialisables développées avec des fonds publics sont mises à la disposition de l'ensemble des acteurs.

Stimuler la demande pour les produits de TIL

Au Québec, la demande pour les produits et les services conçus par les entreprises du secteur du TIL est plus faible, toutes proportions gardées, que la demande américaine ou canadienne-anglaise. Cela tient à plusieurs facteurs, dont le sous-branchement relatif des internautes québécois, la prise de conscience plus ou moins rapide, de la part des entreprises québécoises traditionnelles, des mérites des outils de traitement de l'information et la faiblesse de la demande gouvernementale.

Pour favoriser un accroissement de la demande gouvernementale pour les produits des IL, de même qu'une hausse de sa sophistication, le gouvernement pourrait stimuler l'adoption, dans l'ensemble de l'appareil gouvernemental, de standards de qualité élevés en matière de service à la clientèle. À titre d'exemple, l'adoption d'une règle voulant que « tout ministère ou organisme doit répondre dans les quatre heures à la demande d'information produite par courriel par une entreprise ou un citoyen » pourra amener les ministères et les organismes à se doter de technologies permettant la répartition et le traitement automatiques des messages électroniques. Plus directement, le gouvernement pourrait aussi obliger les ministères et les

organismes à consacrer un pourcentage des sommes allouées à la mise en branle de projets informatiques à l'acquisition de produits du TIL.

Dans une veine différente, la mise en vigueur de lois et de règlements visant à augmenter la sécurité des Québécois et l'accessibilité de l'information produite par les organisations faisant affaire au Québec pourrait déboucher, indirectement, sur le recours accru aux solutions de TIL. De telles réglementations existent aux États-Unis.

Le Québec pourrait imiter l'Union européenne, les problèmes du secteur étant aussi imputables à la mauvaise connaissance qu'ont les utilisateurs des produits du génie linguistique, et renouveler son appui à la réalisation de travaux périodiques de veille, visant à étudier les cas étrangers, publics et privés, les plus intéressants en matière d'utilisation des outils de traitement de la langue, et diffuser ces travaux au sein de l'administration publique et du secteur privé.

Infléchir les stratégies des acteurs du domaine

Finalement, les intervenants du secteur du TIL émettent fréquemment le souhait de pouvoir participer à de brèves rencontres réunissant un éventail aussi large que possible des acteurs du milieu. Le gouvernement pourrait appuyer la tenue de rencontres nationales, lesquelles auraient pour avantage, entre autres, de favoriser la circulation d'information et de rassurer les intervenants sur l'attention portée par l'État à leur industrie.

Conclusion

Nous avons vu que le secteur du TIL pourrait mieux se porter au Québec. La situation, qui serait déplorable ailleurs, l'est encore plus ici, étant donné le contexte linguistique particulier dans lequel la nation québécoise se trouve, alors même que s'accroît l'ouverture sur le monde de son économie. Une telle situation devrait être corrigée, en raison des effets positifs que la mise au point d'outils de traitement du français – plus précisément de logiciels destinés à augmenter la productivité de notre langue – pourrait avoir sur notre devenir collectif.

Peu d'écrits témoignent mieux, selon nous, de l'importance pour l'État québécois d'appuyer fortement l'informatisation du français que le passage suivant, tiré du mémoire que Guy Bertrand, de l'Université du Québec, présentait en 2001 devant la Commission des États généraux sur la situation et l'avenir de la langue française au Québec :

« Au chapitre des nouvelles technologies de l'information et des communications, nous manquons d'innovations qui promeuvent les contenus en français dans les domaines du travail et de la vie personnelle. Il nous faut plus d'effervescence. La concurrence est forte. Le gouvernement américain investit massivement (subventions, contrats, concertation, etc.) en vue d'encourager l'effervescence et de conserver son avance. En particulier, il s'est doté, depuis 1958, d'un instrument très efficace, l'*Advanced Research Projects Agency*, qui soutient fortement le développement technologique aux États-Unis, outil dont le but premier est d'inventer les armes de demain. L'effervescence dans le domaine des technologies militaires a une incidence sur le développement technologique et économique dans son ensemble. L'ARPA a fortement soutenu le développement initial d'Internet aux États-Unis, ce qui a largement contribué à assurer la position dominante des entreprises américaines en matière de nouvelles technologies de l'information et des communications.

Bien sûr, continue Guy Bertrand, le rôle du Québec n'est pas de développer les armes de demain. Mais ce que l'ARPA fait pour l'armement aux États-Unis, un "organisme pour le développement du traitement avancé des contenus et services en français" ne pourrait-il pas, toutes proportions gardées, le faire au Québec pour l'épanouissement de la diversité culturelle et linguistique? Le dynamisme de notre culture et de notre langue n'est-il pas aussi important pour nous que l'est la puissance militaire pour les Américains? De plus, comme eux, nous en tirerions des profits commerciaux et économiques²⁶. »

Réjean Roy, consultant.

26. Consulté le 20 novembre 2002 à l'adresse
http://www.etsatsgeneraux.gouv.qc.ca/pdf_journees/Bertrand_G.pdf.

Le traitement informatique des langues au Canada

Comme annoncé dans la courte biographie qu'on a présentée aux participants, je ne suis pas un spécialiste du traitement informatique des langues (TIL). Je suis plutôt spécialisé dans la mise en place de programmes scientifiques et technologiques. C'est donc à titre de gestionnaire que, pour le compte du Conseil national de recherches du Canada (CNRC) et du Secrétariat du Conseil du Trésor, je vais aujourd'hui vous parler de la situation des industries de la langue au Canada.

Depuis deux ans, différents ministères fédéraux – le Bureau du Conseil privé, le Secrétariat du Conseil du Trésor, Développement des ressources humaines Canada, Industrie Canada, Patrimoine canadien, Travaux publics et Services gouvernementaux – travaillent de concert, formellement et informellement, au développement des industries de la langue (IL) au Canada.

Comment ces intervenants définissent-ils le secteur des IL? Leur réflexion se poursuit toujours, mais disons qu'ils y incluent pour l'instant la traduction, la formation linguistique, la production d'outils de traitement informatique des langues (TIL) – ex. : aides à la traduction, logiciels d'apprentissage des langues – et celle de produits incorporant des fonctionnalités de TIL – ex. : certains outils facilitent la gestion de la relation client par téléphone ou par l'entremise du Web.

Au Canada, on exagère à peine quand on dit que les industries de la langue sont nées avec l'adoption, dans les années 1960, de la *Loi sur les langues officielles*. En requérant des ministères et des organismes fédéraux qu'ils fonctionnent en français et en anglais et en exigeant des entreprises faisant affaire au Canada qu'elles communiquent dans les deux langues avec les consommateurs, cette loi a mené à une explosion de la demande pour des services de traduction.

Ces dernières années, l'utilisation de plus en plus massive d'Internet et la mondialisation ont eu pour effet une augmentation fulgurante de la taille du marché des IL.

D'une part, le gouvernement fédéral, sous le leadership du Secrétariat du Conseil du Trésor, a élaboré une politique globale visant à permettre à l'ensemble des citoyens du Canada – anglophones et francophones – d'accéder à des services en ligne de qualité. Cette politique l'a amené à appuyer la création d'infrastructures inforoutières avancées, l'achat de matériel informatique par les écoles, le

développement de sites Web communautaires, etc. Elle l'a aussi amené à prendre le virage du gouvernement électronique. Le Canada a réussi ce virage. Selon Accenture, société-conseil internationale qui tient compte de critères comme la qualité de l'accès et des contenus et le niveau de convivialité, le Canada est depuis trois ans le pays le plus avancé au monde en matière d'administration en ligne.

D'autre part, l'informatisation de nos sociétés et la mondialisation de nos économies ont eu pour effet l'augmentation des besoins des organisations canadiennes en matière d'adaptation linguistique et culturelle ou, en jargon informatique, de *localisation* et d'*internationalisation*. En effet, pour faire affaire sur les marchés internationaux, les entreprises doivent être en mesure de produire leurs sites Web en plusieurs versions.

Malheureusement, l'augmentation fulgurante de la demande en matière de traduction ou de la localisation n'a pas été accompagnée d'une hausse équivalente de l'offre. Par exemple, le Canada aurait besoin de plusieurs milliers de nouveaux traducteurs professionnels par année, alors qu'il n'en forme que 300. Le pays aurait aussi besoin de plus de professeurs de langue pour faire face à la hausse de la demande pour des cours de français, d'anglais et autres – une demande constituée à 40% d'étudiants étrangers.

Dans ce contexte, on comprend que le Canada fonde de grands espoirs dans le développement d'outils de TIL.

Mais la situation n'est pas rose de ce côté. En effet, la plupart des producteurs de logiciels sont de petite taille. Ils ont peu de ressources pour mettre au point de nouveaux produits et faire face aux défis qui se présentent à eux. De plus, leur secteur est grandement méconnu des acheteurs, que l'on parle du gouvernement ou des entreprises.

Pour corriger ce portrait, le Canada vient de se donner un plan d'action en IL. L'orientation de ce plan est *d'abord et avant tout* commerciale et industrielle. L'objectif du gouvernement du Canada est en effet de stimuler l'industrie de manière à en faire un *leader* incontesté à l'échelle internationale.

La stratégie canadienne comporte plusieurs grands volets.

Premièrement, elle vise à favoriser la mise en réseau des acteurs du secteur, c'est-à-dire des entreprises, des institutions universitaires et des centres de recherche. Seul le regroupement des forces des intervenants pourra en effet permettre au Canada de se tailler une place de choix à

l'échelle internationale. À cette fin, le gouvernement fédéral a décidé d'appuyer la fondation d'une association industrielle en IL, AILIA (Association en industries de la langue *Language Industries Association*). Industrie Canada investira 1 million de dollars par année pendant cinq ans pour soutenir les activités de l'association et assurer le réseautage et la coordination des membres de l'industrie.

Dans une deuxième étape, nous essayerons de voir quelles sont les forces réelles du Canada dans le secteur, celles sur lesquelles il devrait miser, dans les années à venir, pour percer les marchés extérieurs. À cette fin, nous avons entrepris une série de consultations qui se termineront par la production de ce que nous avons baptisé une *carte routière technologique*.

Le Canada entend aussi faire connaître le secteur aux acheteurs potentiels. Pour ce faire, il mettra en place des programmes de soutien à la mise en marché et à l'élaboration d'une image de marque. Cinq millions de

dollars, répartis sur cinq ans, seront affectés à ces efforts de *branding*.

Finalement, un centre de recherche en IL, créé en partenariat par le CNRC, l'Université du Québec en Outaouais et le Bureau de la traduction du Canada, pourra aider les entreprises à se positionner avantageusement au niveau technologique. On peut penser que le centre pourra temporairement accueillir des entreprises poursuivant des travaux de RD et leur fournir de l'encadrement.

Cet encadrement pourra prendre diverses formes. Par exemple, le centre pourra développer des plates-formes informatiques communes pour les entreprises et soutenir la formation de personnes qualifiées en industries de la langue. Il pourra aussi déployer une infrastructure d'évaluation des outils de TIL et mener des activités de veille.

Comme vous le voyez par mon vocabulaire et par cette présentation, l'optique du Canada en matière de TIL est vraiment commerciale.

Les défis du Canada dans le secteur du TIL

Cible	Recommandations
Formation	<ul style="list-style-type: none"> Favoriser le multilinguisme Encourager la multidisciplinarité (traduction – technologie) Faciliter l'accès aux programmes (à distance, formation continue, etc.) Incorporer l'entrepreneuriat / commercialisation Former de nouveaux professeurs et chercheurs Soutenir et faciliter l'intégration au marché du travail (stage, mentorat, etc.) Mettre en priorité les industries de la langue dans la stratégie nationale de formation
Relève	<ul style="list-style-type: none"> Créer des programmes d'études supérieures, principalement en technologie Faciliter l'accès à la formation (Internet, à distance, formation continue, etc.) Sensibiliser les jeunes et les professionnels aux professions de la langue Attirer les membres des communautés culturelles vers les professions Faciliter l'immigration de professionnels, de chercheurs et de formateurs Attirer des étudiants étrangers et inciter les diplômés à demeurer au pays
Grappe industrielle	<ul style="list-style-type: none"> Développer la connaissance des industries (carte routière technologique, étude de marché, recensement de compétences, inventaires des ressources, profils types d'entreprise, etc.) Intégrer les industries aux programmes existants (<i>Branding Canada</i>, <i>Highly Qualified People</i>, etc.) Évaluer la faisabilité d'une grappe industrielle (centre d'excellence ou autres appellations / définitions)

Cible	Recommandations
Financement RD	Établir un fonds de recherche Créer des outils de diffusion des innovations langagières Simplifier l'accès aux sources actuelles (responsabilités claires) Incorporer le volet langagier dans l'application des subventions
Commercialisation	Mettre en action une stratégie de <i>branding</i> Établir une vitrine internationale (comme l'industrie du tourisme) Intégrer les industries à la représentation canadienne à l'étranger (ambassades – consulats, promotion internationale, missions internationales) Définir de façon commerciale les industries (comme l'industrie de l'environnement) Établir un fonds de commercialisation
Le secteur public : client modèle et levier économique pour les industries de la langue	Rééquilibrer l'offre et la demande en utilisant plus et autrement le secteur privé Favoriser les technologies langagières dans les politiques d'achats Promouvoir des projets d'envergure nationale

Source: Canada (2002), *Les industries de la langue au Canada*, rapport produit dans le cadre des travaux d'élaboration de la stratégie d'innovation du Canada, document électronique consulté le 15 octobre 2003 à l'adresse : www.innovationstrategy.gc.ca/cmb/innovation.nsf/RapportsSectoriels/Langue.

André Manseau, directeur de portefeuille. Conseil national de recherches du Canada.

Le traitement informatique des langues dans la communauté Wallonie-Bruxelles

L'état de situation que je vais vous présenter aujourd'hui se base sur les données que nous avons rassemblées par le truchement d'enquêtes quantitatives et qualitatives menées en Belgique francophone²⁷.

Ces enquêtes permettent, d'une part, de tenir à jour un répertoire électronique des intervenants wallons-bruxellois du secteur du traitement informatique des langues (TIL) et de leurs travaux. Elles servent, d'autre part, à suivre les tendances qui influencent l'évolution du domaine et à repérer les obstacles qui y freinent la recherche et le développement (RD).

Les intervenants ont communiqué les renseignements les concernant par l'entremise de questionnaires électroniques placés sur le site Web de l'Observatoire du traitement informatique des langues et de l'inforoute (OTIL). Pour garder les données à jour, nous avons incité les acteurs à les actualiser. Nous avons aussi fait de la veille technologique.

La dernière mise à jour des données de la banque a été faite à la fin de septembre 2002. Cette mise à jour a été suivie, en novembre 2002, par l'analyse des données et par la publication des résultats.

Pour construire la banque de données, nous avons utilisé une définition *large* du concept de TIL.

Plus précisément, pour être pris en considération dans la banque :

- l'intervenant devait être une société, une institution, une association ou un centre de recherche universitaire (ou assimilé) localisé dans la Région wallonne ou la Région de Bruxelles-Capitale;
- il devait participer à des projets de recherche ou de développement d'outils de TIL se déroulant, pour une phase au moins, dans la Région wallonne ou la Région de Bruxelles-Capitale (ou il devait créer des services de TIL faisant appel à l'expertise wallonne ou bruxelloise);

27. En Belgique, il n'existe pas de statistiques propres au secteur du traitement informatique des langues (TIL). La Belgique dispose seulement de statistiques officielles dans le domaine des technologies de l'information et de la communication (TIC), auquel le secteur des TIL appartient. En 2002, les organisations du domaine des TIC ont réalisé un chiffre d'affaires de 30 milliards d'euros et créé pour 11 milliards d'euros de valeur ajoutée. Environ 10% de cette valeur ajoutée est réalisée dans le sous-secteur du développement et de la fabrication de produits TIC.

- avoir des activités dans au moins un domaine TIL, que l'outil soit utilisable en linguistique, médecine, économie, formation, finance, gestion ou autre.

L'analyse complète des données présentées dans cette synthèse est disponible sur le site Web de l'Observatoire du traitement informatique des langues et de l'inforoute Wallonie-Bruxelles (www.otil.org).

La définition du TIL

La manière dont on définit l'expression *traitement informatique des langues* dépend des intervenants et fluctue en fonction des progrès technologiques. Selon notre conception, la chaîne d'activité en TIL s'étend de la recherche en ingénierie linguistique à la production et à la fourniture de produits et de services intégrant des fonctionnalités de traitement du langage naturel.

Les organisations du secteur du TIL créent des outils qui donnent de la valeur ajoutée aux technologies de l'information et de la communication. En effet, ces applications simplifient l'accès des usagers à l'information en facilitant la recherche de renseignements à partir de sources multilingues, en permettant l'interrogation des ordinateurs en langue naturelle, et ainsi de suite. Elles favorisent également une meilleure utilisation de l'information en accélérant et en améliorant l'analyse de textes oraux ou de textes écrits, en rendant possible l'élaboration de systèmes-experts, etc.

Les technologies TIL permettent aussi de repousser progressivement les limites de l'utilisation que nous faisons de la langue. Par exemple, les systèmes de reconnaissance vocale sont appelés, à moyen ou à long terme, à faire partie de notre vie quotidienne: ils serviront en effet à commander les appareils qui nous entourent, à la maison ou au travail, dans notre voiture comme dans les lieux publics²⁸.

Cela dit, dès la constitution de son répertoire, en 1998, l'OTIL a choisi d'associer les questions d'inforoute aux questions de TIL. Il lui paraissait en effet important de

28. Ingénierie du langage, Exploiter toutes les ressources du langage - LINGLINK / Anite Systems - <http://www.linglink.lu/le/en/whatis.html>

tenir compte dans sa banque des méthodes et des environnements qui servent de toile de fond au développement et à l'implantation des outils de traitement des langues. Cette approche *large* du secteur TIL permet, selon nous, de mieux tenir compte de son évolution technologique et des enjeux sociétaux de son développement.

L'adoption d'une définition large a amené l'OTIL à diviser le champ du TIL en 11 domaines d'activité différents: informatique documentaire; traitement de la parole; systèmes-experts; éducation/formation; analyse et génération de textes; évaluation; rédaction et édition de textes; normalisation et localisation; lexicographie, terminotique et traduction (LTT); outils inforoutiers et, enfin, traitement de l'écrit.

Les résultats de l'enquête

Nombre d'intervenants

En 2002, 136 organisations wallonnes ou bruxelloises étaient actives dans le champ du TIL. De ce nombre, 67 étaient des organismes de type « société » ou « institution » et 69, des centres de recherche. Près de 60 % des organismes répertoriés dans la banque de l'OTIL, soit 40 sur 67, avaient le statut juridique de société anonyme.

Nous avons constaté différents mouvements dans le répertoire TIL, même si le nombre total d'organismes a peu varié depuis la dernière enquête menée par l'OTIL (2000). Ainsi, certaines sociétés ont disparu, que ce soit parce qu'elles ont déposé leur bilan à cause du repli économique de la période 2001-2002, parce qu'elles ont été reprises par une autre organisation ou parce qu'elles ont abandonné leurs activités de traitement des langues pour se concentrer sur leurs forces. Certains centres de recherche ont aussi été victimes de la restructuration des facultés universitaires. Évidemment, la perte de certains intervenants a été contrebalancée par l'émergence timide de nouvelles entreprises, l'implantation de succursales d'organismes étrangers dans l'espace Wallonie-Bruxelles, etc.

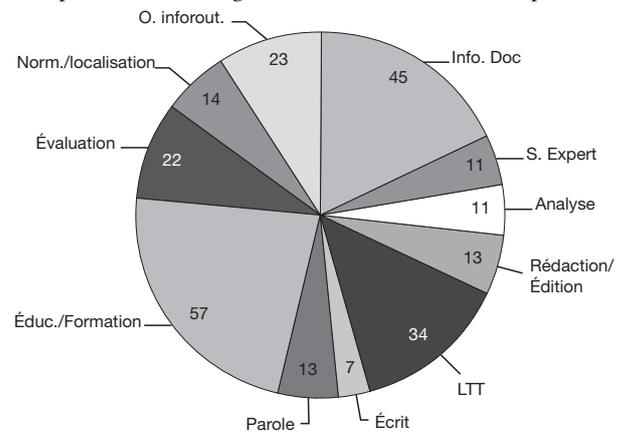
Dans l'ensemble, on peut estimer que le secteur du TIL s'en est bien tiré. La situation pourrait cependant se détériorer si les sociétés de financement en venaient à désertier définitivement la sphère des nouvelles technologies,

comme pourrait le laisser craindre la décision de Technicom de ne pas financer la société 2Ci, fournisseur de services applicatifs²⁹.

Caractéristiques des organismes TIL

Le graphique 1 montre quels sont les champs d'activité de l'ensemble des organisations actives dans le champ du TIL. Comme on le voit, c'est dans le domaine éducation/formation qu'est actif le plus grand nombre d'entreprises et de centres de recherche (57). Le champ de l'informatique documentaire est le deuxième secteur le mieux investi par les intervenants wallons et bruxellois.

Graphique 1
les champs d'activité des organisations actives dans le champ du TIL



Lorsque l'on jette un regard séparé sur les activités des centres de recherche et des sociétés, on se rend cependant compte que le portrait est un peu différent. Ainsi, pendant que les centres de recherche s'affairent principalement à développer des outils dans le champ éducation/formation et dans celui de la lexicographie, de la terminotique et de la traduction, les entreprises œuvrent surtout dans le domaine de l'informatique documentaire et dans celui de l'éducation/formation. Fait à noter, un seul centre de recherche se dit actuellement spécialisé dans le traitement de l'écrit.

29. Voir De Clerck, J.-P., « La Wallonie fait tomber son ASP », 12 août 2002, consulté sur <http://www.internetaddict.be>.

Par rapport à l'analyse menée en 2000, nous constatons que les champs d'activité des organisations ont peu varié. Par exemple, en traitement de la parole, on compte deux sociétés de plus en 2002 qu'en 2000; en normalisation/localisation, on retrouve aujourd'hui deux centres de recherche de plus qu'il y a 24 mois.

Dans certains secteurs, on retrouve beaucoup de PME, tandis que dans d'autres, on en rencontre peu. Par exemple, les PME (SA et SPRL) représentent 75 % des sociétés actives en informatique documentaire, mais seulement 43 % de celles recensées dans le champ éducation/formation. La taille des entreprises varie aussi d'un secteur à l'autre. Ainsi, c'est dans le domaine de l'informatique documentaire que l'on dénombre le plus de sociétés comptant de 21 à 100 employés.

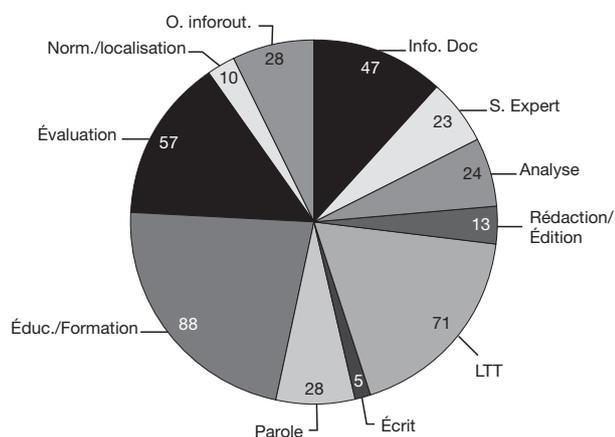
Projets de recherche

Les activités TIL répertoriées sont essentiellement de trois ordres: projets de recherche, mise au point de produits et mise au point de services.

Sans surprise, les centres de recherche sont responsables de 80 % des projets de recherche en TIL. Entre 2000 et aujourd'hui, le nombre de PME impliquées dans des projets de recherche a cependant augmenté de 10 à 13.

On recense actuellement 237 projets de recherche dans la Communauté Wallonie-Bruxelles. Le graphique 2 illustre la manière dont ils sont répartis.

Graphique 2
les thèmes des projets de recherche



Par rapport à 2000, nous relevons une stagnation du nombre de projets de recherche entrepris dans le domaine du traitement de l'écrit, une légère augmentation du nombre de projets en cours dans le domaine du traitement de la parole (deux projets de plus qu'il y a trois ans) et un accroissement du nombre de projets en lexicographie, terminotique et traduction (cinq projets de plus).

Sur les 184 projets de recherche dont la source de financement est connue, 56 sont soutenus par des fonds européens. Le nombre de projets appuyés par la Région wallonne a, pour sa part, grimpé de 22, en 1999, à 33, aujourd'hui. En outre, seulement neuf des projets menés dans des centres de recherche ont reçu l'aide de fonds privés belges ou étrangers.

L'importance de la catégorie *Fonds européens* s'explique du fait que les nouvelles technologies sont une priorité pour l'Union européenne. Cette dernière cherche évidemment à influencer la manière dont les projets sont menés. Les centres de recherche qui veulent s'octroyer un maximum de chances de participer aux programmes européens doivent en effet créer des équipes interuniversitaires et pluridisciplinaires.

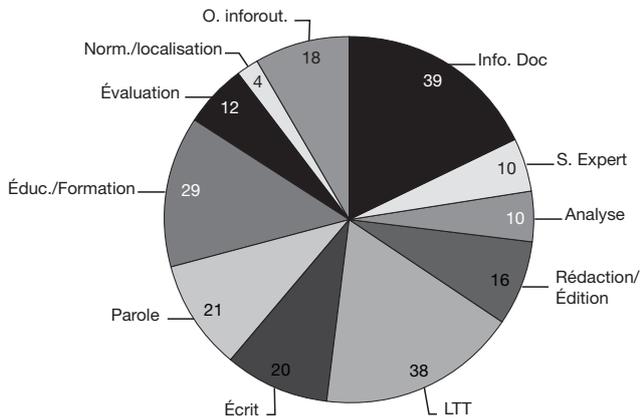
Le faible taux de participation des entreprises belges ou étrangères au financement de projets dénote certainement un manque de synergie entre le secteur privé et le monde académique. Il met aussi en lumière le fait que l'obtention de l'aide à l'échelle régionale ou provinciale suppose des démarches lourdes. Finalement, le fait qu'une seule PME de grande taille (plus de 100 employés) peut revendiquer avoir participé à un projet européen de TIL met en exergue la réticence ou la difficulté des PME à participer à des projets internationaux, en dépit de l'existence de nombreuses initiatives³⁰ visant à les soutenir à cet égard.

Produits de TIL

Les données de l'enquête révèlent que les entreprises de la Communauté Wallonie-Bruxelles ont mis au point 144 produits de TIL. Ces produits tombent surtout dans les catégories *Information documentaire* et *Lexicographie, terminotique et traduction* (voir graphique 3).

30. Helpdesk de l'Union européenne, CRAFT (*Cooperative Research Action for Technology*), programme *Horizon* de la Région wallonne, projet ETI-BICS issu d'une initiative du CEEI, le soutien des Centres relais innovation (CRI), etc.

Graphique 3
les catégories des produits de TIL



N. B. Un produit peut appartenir à plus d'une catégorie, ce qui explique que le total des nombres de la tarte soit de 217.

Par rapport à 2000, notons une hausse du nombre de produits de traitement de la parole (+ 8)³¹, de produits de traitement de l'écrit (+ 3), de produits d'analyse et de rédaction de texte (+ 4) et d'outils de lexicographie, terminotique et traduction (+ 6).

Les deux tiers (68%) des produits ont été créés par 34 sociétés. Dix-huit centres de recherche ont mis au point le tiers restant. Ce dernier tiers est surtout constitué de produits appartenant à la catégorie éducation/formation.

Douze p. cent des produits mis au point appartiennent à la catégorie des ressources linguistiques. Ces outils – que l'on parle de dictionnaires électroniques, de lexiques, de corpus ou autres – servent à créer des applications de TIL (88% des produits).

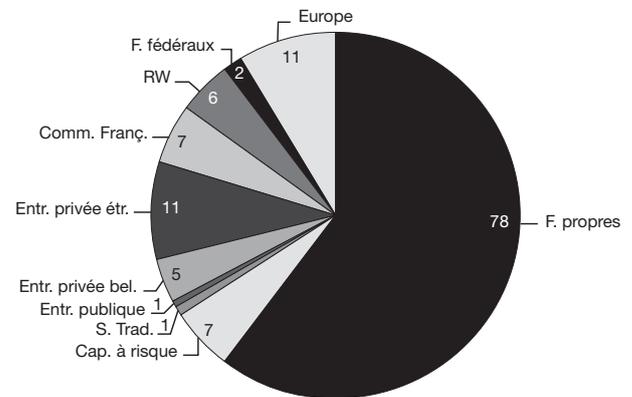
La majorité (80%) des technologies TIL développées dans la Communauté Wallonie-Bruxelles sont incorporées à des applications destinées aux professionnels plutôt qu'au grand public.

De nombreux produits doivent leur existence ou leur popularité croissante à Internet. L'enquête de l'Agence wallonne des télécommunications (AWT), sur les usages technologiques des PME wallonnes, révèle que 77% des PME disposent d'un accès Internet et que 61% possèdent un réseau local. En progression

31. Pour mémoire, ces technologies figurent sur une liste de « 40 technologies-clés pour la Wallonie ».

constante, ces chiffres laissent prévoir une augmentation rapide de la demande pour des outils de gestion informatisée des documents, de création automatique des contenus, etc. Il faudra toutefois informer les utilisateurs sur le potentiel des outils de TIL dans leurs activités professionnelles.

Graphique 4
le financement des produits de TIL



De manière générale, le financement reste un sujet délicat pour les entreprises, qui ne dévoilent pas facilement de renseignements à ce sujet, surtout en ce qui concerne leurs sources de capital. Une seule société anonyme revendique une participation européenne dans le financement de son produit. On peut cependant relever que certains produits sont basés sur une technologie développée dans le cadre de projets européens, mais avec le soutien de fonds propres (graphique 4).

Les organisations recourent relativement peu à la collaboration pour le développement de produits. Notons cependant que la collaboration sociétés-centres de recherche a mené au développement de 15 produits, que la coopération sociétés-sociétés a permis de créer 11 produits et qu'en travaillant de concert, les centres de recherche ont élaboré 16 produits.

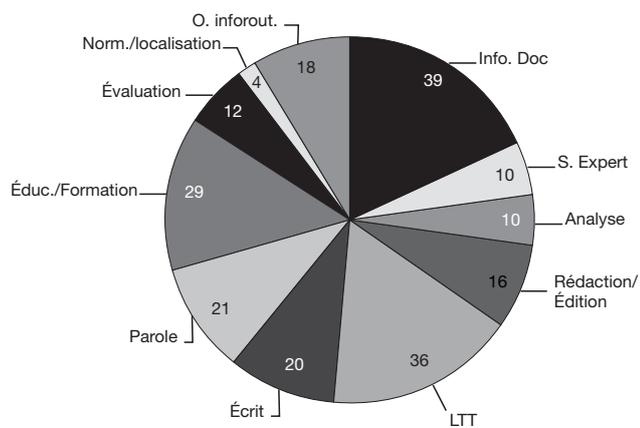
Services de TIL

Les services proposés dans le secteur du TIL ont pour but d'adapter les technologies ou les applications de TIL

pour qu'elles répondent aux besoins spécifiques des utilisateurs : par exemple, certains fournisseurs se spécialisent dans l'ajout de moteurs de recherche plein texte sur les sites Web de leurs clients, développent des thésaurus pour optimiser la gestion du contenu de centres de documentation, organisent des séances de formation, font de la localisation de logiciels, etc.

Actuellement, 47 sociétés et 23 centres de recherche commercialisent 118 services de TIL. Par rapport à 2000, il s'agit d'une augmentation du nombre de prestataires. Les services sont le plus souvent offerts dans les catégories *Information documentaire*, *Éducation/formation* et *Lexicographie, terminotique et traduction* (graphique 5).

Graphique 5
les catégories des services de TIL



N.B. Un service peut appartenir à plus d'une catégorie, ce qui explique que le total des nombres de la tarte soit de 205.

Quelques constats

Le secteur

Les activités de TIL ne peuvent pas toutes être abordées avec une même approche. On ne peut pas traiter de la même manière les ressources linguistiques et les logiciels de type industriel qui sont destinés à être intégrés à d'autres produits ou services, et les applications de TIL à *consommation immédiate* – correcteurs, logiciels de navigation avec interface vocale, etc. – destinées aux professionnels et au grand public.

Les organisations

- Les organismes sont principalement de faible structure, quoiqu'ils aient relativement bien résisté à l'essaiage, en 2000-2001, des sociétés technologiques;
- la collaboration entre acteurs se produit surtout lors de la réalisation de projets de recherche européens ou autres, ou lors de la signature de contrats de sous-traitance relatifs à l'élaboration de produits ou de services;
- l'information sur les résultats de recherche est lacunaire et difficile à exploiter. Cependant, les centres de recherche impliqués dans des projets européens ou dans des prestations de service communiquent plus volontiers sur leurs activités;
- les centres de recherche n'ont pas de politique de gestion de la qualité; les sociétés en ont rarement;
- on note la constitution de centres de compétences et d'excellence.

Les projets

- Il y a des projets intéressants dans certains domaines de base (analyse de texte, lexicographie, systèmes-experts, traitement de la parole et de l'écrit), mais ceux-ci sont trop peu nombreux pour permettre un véritable essor des technologies TIL;
- il n'existe pas de véritable symbiose entre la recherche fondamentale et la recherche appliquée;
- la participation des entreprises de la Communauté Wallonie-Bruxelles aux projets européens est nettement trop faible; cela s'explique en partie par une mauvaise connaissance des rouages européens et un manque de temps, d'argent ou de ressources humaines pour constituer un dossier.

Les produits

- Les applications de TIL développées par les acteurs du domaine sont principalement destinées aux professionnels;
- les sociétés sont principalement actives dans des domaines où l'on retrouve des applications de TIL (informatique documentaire, OCR);

- on constate l'émergence de produits spécifiquement développés pour le Web et non plus simplement adaptés à ce support ;
- par rapport à l'analyse menée en 2000, on relève une augmentation de la participation des sociétés de capital à risque et de la Région wallonne au financement des produits.

Les services

- La livraison d'un produit destiné aux professionnels est généralement soutenue par la prestation d'un service (dans le secteur du TIL, la dichotomie entre biens et services tend à devenir floue) ;
- la prestation de services constitue l'activité principale de certains organismes ; les fournisseurs s'appuient sur une technologie développée à l'interne ou acquise auprès d'une autre société (nationale ou non) ;
- le service de TIL est souvent intégré dans un ensemble de services qui visent à fournir au client un produit fini ;
- les services de formation ou autres représentent une source de financement pour certains centres de recherche en TIL, centres dont le nombre est en hausse.

Pistes de réflexion

Les réflexions et les recommandations suivantes ne s'éloignent pas de celles émises dans les analyses de 1999 et 2000. Cependant, la situation a évolué. Les intervenants et les décideurs ont pris conscience des difficultés actuelles éprouvées au niveau de la RD dans le domaine du TIL.

Dans le cadre du projet Prométhée/Région wallonne, diverses actions (ex. : classification des secteurs-clés en Wallonie, constitution de grappes technologiques) ont été proposées et diverses pistes de réflexion (ex. : centraliser l'information relative à la RD, instaurer un dialogue et une véritable collaboration interuniversitaire) ont été avancées. Elles favorisent l'émergence d'un climat propice à l'innovation technologique. Devant cette évolution du climat de la RD, nous nous contenterons de souligner certains points cruciaux pour le secteur TIL.

Il est regrettable que les divers programmes relatifs à l'innovation ne fassent pas l'objet d'une évaluation

systématique. Outre des éléments utilisables pour un éventuel réajustement de l'aide, ces évaluations fourniraient des données supplémentaires à la connaissance du domaine des TIC.

Pistes sur le plan de la recherche

Sur le plan de la recherche, il conviendrait de :

Définir une politique de RD à moyen terme

Abandonner le devoir d'innovation à la seule responsabilité des chercheurs serait une erreur. La recherche est au service de l'innovation. La Région wallonne a dressé une liste des 40 technologies-clés pour la Wallonie, parmi lesquelles on retrouve deux domaines TIL : Traitement de la parole et Éducation/Formation.

La recherche dans ces domaines gagnerait en pertinence si elle s'inscrivait dans une politique élargie de l'innovation (anticiper les besoins du marché et y déceler les niches qualitativement et quantitativement les plus profitables, accorder un financement ciblé, etc.). De plus, définir des niches d'activité permettrait l'apparition de projets rassemblant divers pôles de compétences (TIL et non TIL).

Encourager la recherche fondamentale et appliquée

La pérennité des recherches et des produits de TIL nécessite un investissement soutenu dans la recherche (analyse et génération de textes, lexicographie, terminotique et traduction, systèmes-experts intégrant des éléments de TIL). Ces investissements peuvent s'inscrire dans un cadre international, vu la complexité de ces domaines.

Rationaliser la recherche

L'instauration d'incitants à la création de centres d'excellence regroupant divers centres de recherche contribuerait à la rationalisation de la recherche. Cette recommandation se fonde sur le constat de la pléthore de recherches dans certains domaines TIL (éducation/formation, évaluation, etc.).

Favoriser le transfert technologique entre recherche fondamentale et appliquée

Tant que les divers niveaux de financement de la recherche (État fédéral, Communauté française, Région

wallonne) ne mettront pas en œuvre une politique commune de recherche, de nombreux résultats de la recherche fondamentale continueront «à dormir au fond du tiroir». Ce constat est également valable pour le secteur TIL.

Valoriser le chercheur et son travail

Outre les occupations de recherche, les activités de prospection et de promotion doivent être intégrées dans le statut du chercheur. De plus, une valorisation du statut (nomination, rémunération...) stimulerait le chercheur à adopter une attitude proactive.

Évaluer la pertinence économique d'un projet de recherche dès sa définition

Des études du marché des technologies, des besoins des entreprises devraient accompagner tout projet. Elles permettraient aux chercheurs :

- de savoir si des recherches similaires ou apparentées ont déjà été effectuées ;
- de réutiliser des ressources linguistiques ;
- de proposer des services, prototypes ou produits ayant des chances de survie significatives ;
- d'inscrire leurs recherches dans la chaîne des activités TIL ;
- de repérer les partenaires susceptibles d'être intéressés par ce type de recherche ;
- de déterminer de nouveaux créneaux de recherche.

La gestion stratégique d'un projet doit être prise en charge par des spécialistes en gestion de la valorisation de la recherche ET initiés aux exigences propres des divers domaines du secteur TIL. À ce titre, une mise en réseau des services d'interface des universités favoriserait l'émergence d'une politique de marketing adaptée à la RD.

Pistes sur le plan industriel

Promouvoir et mieux cibler les organismes de soutien à l'innovation

Très peu de PME TIL font appel aux institutions proposant leur soutien à l'innovation (développement d'un concept ou produit, participation à un programme européen), malgré leur pléthore. Ce constat ne devrait-il pas exhorter les organismes à revoir leur organisation

(promotion, communication basée sur la transparence, simplification des démarches...)?

Simplifier l'accès à l'aide à l'innovation

Les nombreuses démarches relatives à la demande d'une aide à l'innovation, ainsi que les délais d'octroi sont des freins pour les PME actives dans le domaine des technologies.

Favoriser la mise en réseau des compétences

Dans la situation actuelle de concurrence, tant au niveau des universités qu'au niveau des sociétés, la collaboration ne s'impose pas. Elle doit découler d'une prise de conscience, par les organismes, des bénéfices qu'elle générerait. Cependant, les pouvoirs publics peuvent fournir des outils, incitants et programmes d'aide propices à l'émergence d'un tel état d'esprit.

Améliorer la pertinence et la visibilité de l'information scientifique

La publication de l'information scientifique et technique doit s'inscrire dans le mécanisme de la valorisation de la recherche. Elle devrait être obligatoire. On pourrait envisager que tout projet de recherche comprendrait un cahier des charges qui définirait la forme de cette information à caractère «valorisant» et qui respecterait le désir de discrétion des centres de recherche et les besoins des entreprises. De plus, ces publications permettraient aux divers intervenants de la RD de réaliser des évaluations.

Actuellement, les intervenants TIL (et non TIL) peuvent prendre connaissance des activités TIL sur le site de l'OTIL. Cependant, la chaîne des activités TIL nécessite des compétences non TIL (informatique, linguistique, psychologie...). À ce titre, nous soutenons la proposition du projet Prométhée relative à la mise en place d'une base de données reprenant les diverses recherches.

Proposer des programmes d'aide basés sur la collaboration

Des initiatives actuelles introduisent déjà le concept de collaboration dans les conditions de financement : le programme *Wallonie-Université-Développement*, les mandats

First Entreprises qui permettent la réalisation d'un doctorat au sein d'une société, la constitution d'une grappe technologique (projet Prométhée), etc.

Instaurer un système de qualité

Les recherches et les produits TIL sont souvent intégrés à d'autres applications. En conséquence, l'instauration d'un système de qualité, tant pour la recherche que pour la production, faciliterait les échanges interdisciplinaires et la synergie entreprises/centres de recherche. Le projet Prométhée propose diverses interventions de la RW pour favoriser la démarche de qualité dans les universités. Vu le peu d'entreprises certifiées, il nous semble qu'un mécanisme d'aide similaire pourrait être envisagé pour les sociétés.

Encourager la constitution de centres technologiques

Actuellement, deux centres sont concernés par le TIL, Multitel et le CEDITI. Ces centres sont liés aux universités, mais orientés vers la recherche appliquée, le transfert de technologie et le service aux entreprises.

Inciter les partenariats internationaux

Eu égard à la complexité des domaines TIL, l'instauration de réseaux de collaboration internationale doit être encouragée (aide à la recherche de partenaires, facilitation de la participation à des missions économiques à l'étranger...

En conclusion

Cette analyse comporte de nombreuses interrogations restées sans réponse, mais elle a permis de mettre en évidence deux points cruciaux d'intervention pour le plein essor du secteur :

- il faut assurer la mise en réseau des compétences, des connaissances et de l'information. En effet, le traitement informatique des langues n'existe pas par et pour lui-même; ses effets se font sentir dans de nombreuses applications relevant de plusieurs domaines;
- il faudra adapter les structures de soutien pour qu'elles répondent aux besoins des organismes.

Il serait en outre primordial de créer une meilleure image du secteur, que ce soit auprès du grand public, des professionnels ou des décideurs. Il est regrettable qu'en 2002, trop peu de sociétés faisant le TIL soient impliquées dans l'informatisation de l'administration ou de l'enseignement. Promouvoir le secteur permettrait également d'attirer des investisseurs, de susciter l'intérêt des jeunes universitaires, d'encourager les intervenants du TIL à entreprendre, etc.

Sylvie Wallez,
responsable, Observatoire du traitement informatique des langues et de l'inforoute – Wallonie et Bruxelles, Communauté française de Belgique.

Le traitement informatique des langues en Suisse

On m'a demandé de dresser un portrait de la situation des industries de la langue en Suisse. Si vous me le permettez, de manière à ne pas être prise de court à cause des contraintes de temps, je vais commencer par les conclusions: la situation n'est pas catastrophique, mais le pays pourrait mieux faire.

Rappelons, pour mettre en contexte, que la Suisse est un pays multilingue qui subit à la fois la pression de l'anglais, mais, aussi, dans le cas des cantons romands, celle de l'allemand. L'allemand est en effet la langue de près de 75 % des habitants de la Suisse, alors que la part du français n'est que de 20 %. Comme vous le voyez, si le marché québécois du TIL est petit, le marché suisse romand, lui, est minuscule. Il est donc tout à fait vital, pour une organisation helvète désireuse de mettre au point ou de commercialiser des outils de traitement du français, de collaborer avec l'étranger et d'y développer de bonnes antennes.

À cause du caractère cosmopolite de la Suisse et de son positionnement géographique et politique, les organisations actives dans le secteur du TIL sont particulièrement préoccupées par les questions de multilinguisme. Cela est vrai, que l'on parle des travaux menés dans le secteur financier, pour accélérer le traitement des documents écrits, ou encore des efforts réalisés dans des domaines comme la médecine, pour améliorer la productivité du personnel hospitalier (ex.: dictée automatique de bilans médicaux).

En Suisse, le financement de la recherche provient essentiellement du Fonds national suisse de la recherche scientifique (FNRS) et de l'Office fédéral de l'éducation et de la science (OFES). Comme l'indique le site de l'organisme, «le FNRS est responsable de plusieurs programmes de recherche au niveau national et international. Il s'agit d'efforts de recherche coordonnés, limités dans le temps et affectés à des buts clairement définis. La collaboration avec des partenaires extérieurs au milieu académique, le transfert de savoir et de savoir-faire dans l'enseignement et la pratique, ainsi que la mise en valeur des résultats de recherche pour de futurs utilisatrices et utilisateurs et auprès d'un public intéressé sont des caractéristiques essentielles de ces programmes de recherche». De son côté, l'OFES joue essentiellement le rôle

de lien, sur le plan de la recherche, entre la Suisse et l'Union européenne.

Le travail de l'OFES est important, puisqu'il permet à la Suisse, par le truchement d'accords bilatéraux, de participer activement aux projets de recherche de l'Union européenne. Quant au FNRS, ses actions ont débouché, voilà deux ans environ, sur la création de centres d'excellence en industries de la langue.

Ces actions sont très importantes, mais on peut déplorer leur caractère limité dans le temps. Par exemple, les centres d'excellence dont je viens de parler ont été créés pour une période de cinq années. Cela signifie que l'étudiant qui arrive à la licence, qui se dit: «Ce secteur m'intéresse» et qui continue tout de suite son parcours académique risque de finir son doctorat au moment même où le plan quinquennal prend fin. Comme il peut se passer plusieurs années avant qu'un nouveau programme ou projet ne débute, l'étudiant peut se retrouver coincé. Et imaginez que l'on parle maintenant, dans certains cas, de plans biennaux! C'est beaucoup trop court pour mener des projets dans le domaine.

J'aimerais maintenant dire un mot au sujet de la demande. Comme vous le savez, on retrouve un grand nombre d'organisations internationales en Suisse et à Genève. Ces institutions ont d'énormes besoins en traduction, mais elles en ont aussi dans des domaines comme la recherche documentaire ou la terminologie. Habituellement, elles essayent de trouver seules, chacune de leur côté, les solutions qui leur permettront de résoudre leurs problèmes, les outils qui les aideront à traduire leurs documents, et ainsi de suite. On devine que ce processus est assez lourd et coûteux.

Il me semble qu'il serait possible d'améliorer cet état de choses. Par exemple, il me paraîtrait utile, pour reprendre les idées exprimées plus tôt par Stéphane Chaudiron, de travailler à la création de banques de ressources linguistiques communes, au déploiement de bancs d'évaluation dont les résultats seraient utilisables par tous, et ainsi de suite.

*Sandra Manzi,
chercheuse, ISSCO, Université de Genève, Suisse.*

L'Observatoire suisse des industries de la langue (OSIL) présente sur le site www.osil.ch/liste.html un choix de liens utiles pour le traitement informatique du français et des autres langues de la Suisse.

Un nouvel accord sur la recherche entre la Suisse et l'Union européenne a été paraphé le 5 septembre 2003. Le nouvel accord assure la participation intégrale de la Suisse comme « pays associé » au 6^e programme-cadre de l'Union européenne, lequel prendra fin en 2006.

Même si les procédures de ratification de l'accord vont durer jusqu'en 2004, la Suisse et l'Union européenne se sont mises d'accord pour appliquer le nouvel accord provisoirement.

Dès le 1^{er} janvier 2004, les chercheurs suisses auront les mêmes droits que leurs partenaires de l'Union européenne et, comme ces derniers, ils recevront directement leurs subsides de la Commission européenne³².

32. À ce sujet, voir <http://www.bbw.admin.ch/html/pages/forschung/eu/abkommen-f.html>.

Les langues africaines sur la Toile³³

Il existe environ 2000 langues en Afrique selon les linguistes. De ce nombre, 113 peuvent être considérées comme *majeures*, parce qu'elles sont répandues dans beaucoup de pays ou parce qu'elles ont un statut spécifique dans les pays où elles sont parlées.

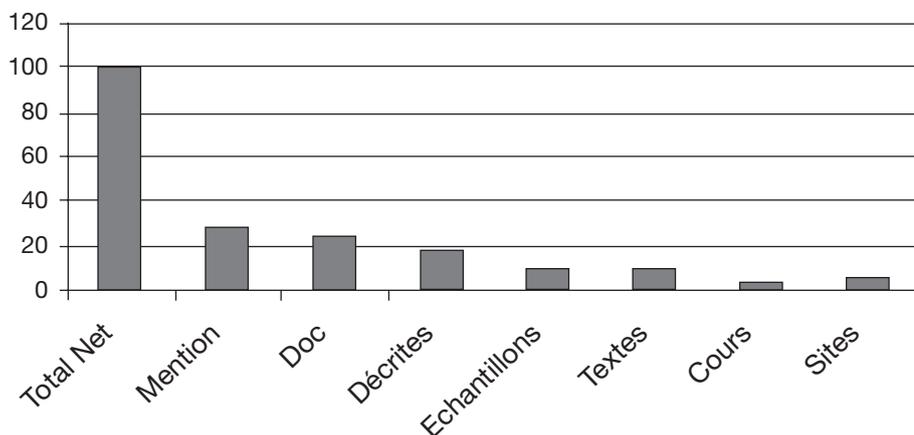
Nous avons mené une recherche sur la Toile pour voir quelle y est la situation des langues les plus parlées de ce groupe, que ces langues aient une grande diffusion ou non. En bout de ligne, notre étude a porté sur la présence de 64 langues.

Pour mener notre étude, nous avons fonctionné de la manière suivante. Nous avons pris le nom de chacune des 64 langues retenues et l'avons utilisé, sous toutes ses variantes – ex. : Hausa peut s'écrire H-a-u-s-a ou H-a-o-u-s-s-a ou H-a-w-s-a – comme requête du moteur de recherche Copernic 2001 Plus.

Cette procédure nous a emmenés à visiter 3332 sites. De ce nombre, seulement 1374 sites traitaient des langues africaines. Les autres parlaient plutôt des cultures de l'Afrique ou de ses populations, et ont été rejetés.

Les 1374 sites retenus ont été ventilés selon leur contenu. Certains ne mentionnent que le nom d'une langue, sont dédiés à des ressources documentaires sur cette langue ou, encore, contiennent de l'information sociolinguistique ou linguistique à son propos, comme le nombre de personnes qui la parlent. D'autres sites livrent des échantillons plus ou moins importants de la langue – par exemple des citations ou des textes –, contiennent des cours interactifs pour l'enseigner ou l'utilisent effectivement comme moyen de communication avec le visiteur.

Graphique 1
l'utilisation des langues africaines

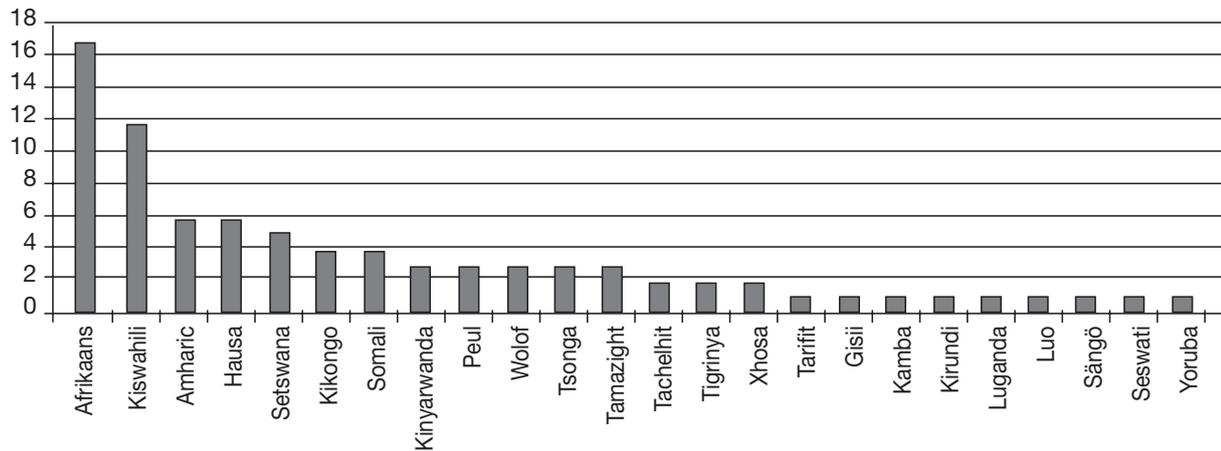


Le graphique 1 montre bien que sur l'ensemble des sites relevés, la proportion de ceux qui mentionnent seulement une langue africaine ou bien qui se limitent à donner de la documentation sur une langue africaine est bien supérieure à celle des sites en langue africaine.

En fait, les sites qui utilisent les langues africaines pour communiquer avec l'internaute (voir le graphique 2 pour leur distribution), que ce soit à quelques endroits – par exemple sur les boutons de navigation, dans les *news* ou dans les forums de discussion – ou encore majoritairement ou complètement ne représentent que 3% de l'ensemble du corpus examiné.

33. Cette présentation est basée sur l'article suivant : Diki-Kidiri, Marcel (2003), «Les langues africaines sur la Toile», *Cahiers du Rifal*, numéro 23.

Graphique 2
les sites en langues africaines



Quelle est la place du français dans l'ensemble des sites examinés? On pourrait croire que les langues parlées dans les pays d'Afrique francophone – par exemple le bambara et le lingala – sont traitées dans des sites utilisant largement le français comme langue de communication. Ce n'est pas le cas. En fait, l'anglais est la langue utilisée pour communiquer avec l'internaute dans 80 % des cas. Le français vient ensuite (8 %). Les langues africaines et les autres langues européennes – l'allemand, l'espagnol, l'italien, le polonais et le néerlandais – se placent encore plus loin derrière.

Comme on le voit, le nombre de sites en langues africaines est faible actuellement. Pour le multiplier, il faudrait d'abord essayer de créer et de diffuser des cours de langues africaines sur la Toile, et appuyer le développement de cybercommunautés utilisant ces dernières. En effet, s'il n'y a pas assez de sites en langues africaines, c'est qu'il n'y a pas assez de gens qui les utilisent pour communiquer sur Internet. Et s'il n'y a pas assez de gens qui les utilisent sur Internet, c'est parce que le nombre de personnes compétentes dans leur langue à l'écrit est trop faible.

On pourrait aussi facilement multiplier le nombre de sites en français sur les langues africaines. En effet, beaucoup de recherches faites en français sur les langues africaines n'ont jamais été mises sur la Toile. Elles ont plutôt été laissées dans le tiroir.

Cela dit, nous parlons aujourd'hui des difficultés des langues africaines sur Internet, alors que d'autres problèmes touchent peut-être davantage les populations d'Afrique. Par exemple, il n'y a pratiquement pas de présence des langues africaines dans la presse. Les publications en langues africaines sur la bureautique n'existent pas non plus.

Pourquoi? Parce qu'il est inconcevable pour le moment d'imaginer que l'on puisse produire de tels documents en langues africaines. Par exemple, des revues comme *L'Expansion*, *News Time*, *Vogue* ou *Jeune Afrique* utilisent chacune une dizaine de polices différentes par page – polices qui n'existent pas ou ne sont pas disponibles en quantité suffisante pour les langues africaines.

Cet exposé donne une idée sommaire de l'immense travail qui reste à faire en matière de traitement informatique des langues. Il est bien de mener des travaux de recherche dans les universités et ailleurs,

mais il faudrait absolument que ceux-ci aient éventuellement des débouchés pour le grand public. Par exemple, il faudrait alphabétiser les populations et leur donner les outils informatiques de base dont elles ont besoin pour leur travail, mais rien n'a encore été fait en ce sens.

*Marcel Diki-Kidiri,
chargé de recherche, CNRS, Villejuif, France.*



Conférence principale

Vers une nouvelle écologie mondiale des langues

Une conférence se tient actuellement à Paris sur l'avenir du français. Les conférenciers sont appelés à se demander si le français demeurera une langue mondiale, à réfléchir à son évolution.

Selon moi, la question n'est pas de savoir si le français va survivre. On compte en effet un nombre bien suffisant de locuteurs francophones pour garantir l'essor du français pendant de très longues années. Cependant, on peut se demander si le français, dans les années à venir, connaîtra le sort enviable de l'espagnol, dont la croissance est fulgurante à tous points de vue, ou, à l'inverse, celui du néerlandais, dont le rayon d'action se limite désormais aux Pays-Bas, à la Flandre, à l'Afrique du Sud – si l'on assimile l'afrikaans à la langue néerlandaise, ce qui pose quelques problèmes malgré tout – et à quelques rares centres de l'Indonésie.

Le niveau où le français se situera à l'intérieur de cette alternative dépendra de nombreux facteurs. Il dépendra notamment de la capacité des francophones à maîtriser la technique. En effet, quoique l'appropriation des technologies par un groupe de locuteurs ne suffise pas, à elle seule, à garantir l'essor d'une langue, l'alliance entre langue et technologie est indéniablement un facteur important de développement sur le plan linguistique. La mutation numérique actuellement en cours donne une importance accrue au facteur technique et nous impose donc d'y prêter une attention toute particulière.

Par exemple, Robert Chaudenson pourrait vous expliquer comment la radio peut favoriser l'épanouissement d'une langue, comment la stratégie de promotion de la langue française par ce moyen de communication traditionnel pourrait s'avérer extrêmement intéressante dans des régions, voire des pays, où l'oralité domine fortement, comme c'est souvent le cas en Afrique. Autre exemple : il est indéniable que l'apparition d'Internet aura des effets majeurs sur le rapport existant entre les langues du monde.

C'est de cette dernière question que j'aimerais discuter dans la vingtaine de minutes qui suivent.

Pourquoi l'inforoute a-t-elle des répercussions sur l'équilibre des langues ? Selon moi, son importance tient au fait qu'elle permet la création de *communautés*. Je pourrais passer beaucoup de temps à définir ce dernier terme, mais disons simplement qu'il renvoie à des groupes de personnes maintenant un contact vivant, un contact qui permet aux langues de vivre à travers la distance et le temps.

Avec Internet, on peut *désynchroniser* les liens de communication et, par cet assouplissement des moyens de communication, assurer plus aisément que les utilisateurs d'une langue distribuée aux quatre coins de la planète puissent maintenir le contact. Avec Internet, on peut donc espérer, pour la première fois, transformer la Francophonie en véritable espace ou, comme diraient les géographes, en territoire. Rappelons en effet qu'un territoire, selon ces derniers, c'est un espace structuré par des moyens et des modalités de communication.

Une chance unique s'offre donc à la Francophonie. Cependant, celle-ci est-elle en mesure de la saisir ? La réponse à cette question dépendra surtout de la présence ou non de certaines mesures et décisions d'ordre politique, social et économique.

Premièrement, il faudra veiller à ce qu'Internet soit accessible partout en Francophonie. Actuellement, la capacité dont disposent deux personnes pour communiquer entre elles varie énormément selon que celles-ci se trouvent à un endroit ou un autre de la Francophonie. Par exemple, envoyer un courriel de Montréal à Paris, à Bruxelles ou dans plusieurs autres villes de la Francophonie riche ne pose pas vraiment de problème. En revanche, il sera beaucoup plus difficile, voire impossible, d'envoyer ce même message dans un pays d'Afrique. La faiblesse des infrastructures, la rareté des points d'accès, le manque de maîtrise des nouveaux outils, le coût d'accès, sans parler de facteurs d'une bien plus grande banalité, telle la capacité même de lire ou d'écrire, se conjuguent pour s'opposer aux désirs et aux besoins des utilisateurs.

Évidemment, on ne peut guère parler de communauté quand l'infrastructure technologique n'est pas là. C'est un peu comme parler de commerce dans une région sans routes, où seuls serpentent quelques sentiers incertains et dangereux sur lesquels les marchandises passent si lentement et difficilement que le marché ne peut pas se développer : les marchandises se perdent, se détériorent ou bien encore ne parviennent pas à destination dans la période des besoins.

Cela dit, je ne suis pas en train de prétendre que les limites rencontrées ne sont que techniques. Mais il faut bien comprendre aussi qu'établir de bonnes liaisons inforoutières d'un point à l'autre de la Francophonie ne coûterait pas si cher et ne serait pas très compliqué à réaliser sur le plan de l'ingénierie des réseaux. À l'occasion

d'une réunion à laquelle j'ai eu le privilège de participer à Bamako, il y a deux ou trois ans, j'ai pu constater que l'on avait installé des soucoupes pour ainsi établir un lien par satellite dans le Palais des congrès de cette capitale. De manière inattendue, mais finalement, à bien y penser, pas si surprenante, le lien Internet ainsi établi était extrêmement rapide. Hélas, il n'a duré que ce que durent les roses, surtout dans le climat désertique du Mali, l'espace d'un congrès. Celui-ci terminé, on a déboulonné les soucoupes et l'isolement communicationnel a retrouvé tous ses droits habituels en restituant toutes les formes de pouvoir aux rares personnes qui parviennent à en reculer un peu les limites. Faire de la communication une denrée rare, c'est lui donner localement une grande valeur; celle-ci, évidemment, se concentre entre quelques mains de l'élite politique et économique du pays qui assoit ainsi son pouvoir et assure ses revenus.

Le problème, on le voit, est donc plutôt politique. En effet, dans plusieurs pays de la planète, les compagnies de communication qui contrôlent la tuyauterie et les moyens de transmission entretiennent des rapports, disons, « *spéciaux* » avec les gouvernements: je simplifie, question de faire vite, mais disons qu'en échange du maintien d'un certain type de régime de communication profitable pour les élites nationales – régime qui, surtout, ne va pas remettre en cause leurs capacités de contrôle –, ces compagnies peuvent continuer à travailler localement, c'est-à-dire à faire de l'argent. Dans tout ce joli calcul, évidemment, l'intérêt général passe au dernier rang.

Revenons à cette série d'obstacles qu'il s'agit de lever. Une fois que l'on aurait déployé l'infrastructure inforoutière requise, une autre action s'imposerait. Il faudrait créer un bassin de gens compétents, former des individus capables de faire fonctionner Internet dans une région, ou encore susceptibles de faire se déployer ces technologies nouvelles sur de nouveaux territoires. Vous le savez, un certain nombre d'agences de la Francophonie travaillent actuellement sur ce dossier, mais elles butent sur une série de problèmes, hélas trop bien connus dans la Francophonie: les coûts de formation sont élevés, l'appareillage dont le personnel a besoin est cher, il faut trouver des moyens de diffuser l'information, et ainsi de suite. On constate que de nombreux pays du Sud sont incapables, actuellement, de prendre le virage Internet, faute de moyens suffisants et en

l'absence d'une réelle volonté politique susceptible de dépasser la simple quête de profits.

Une autre grande question devra être abordée pour qu'Internet apporte les fruits escomptés à la Francophonie: celle de la propriété intellectuelle.

Les choses, de ce côté, ne se passent pas exactement comme elles le devraient. Comme vous le savez sans doute, la *Convention de Berne pour la protection des œuvres littéraires et artistiques* et les ententes conclues depuis à l'Organisation mondiale du commerce (OMC) ou à l'Organisation mondiale de la propriété intellectuelle (OMPI) ont grandement augmenté les possibilités offertes aux créateurs de contrôler la propriété intellectuelle. Ainsi, le droit d'auteur est maintenant protégé un peu partout (sauf au Canada, pour le moment du moins) pendant la vie de l'auteur et les 70 ans suivant sa mort³⁴.

Pour replacer les choses dans leur contexte, rappelons qu'il y a 300 ans, en Grande-Bretagne, la protection accordée par le *copyright* à l'auteur ne durait que... 14 années et n'était renouvelable qu'une fois. L'objectif de la Reine était alors d'atteindre un équilibre entre les attentes de l'auteur (celui-ci voulait voir sa créativité récompensée) et les besoins de la société (d'une part, il fallait veiller à stimuler la créativité des membres de la collectivité; d'autre part, il fallait assurer la constitution d'un fonds commun d'œuvres scientifiques, littéraires, théâtrales ou musicales, un fonds, par conséquent, appartenant à tous).

Rapidement, cette pratique essaima dans d'autres pays, dont la France, où la République reconnaissante décida d'offrir au citoyen créateur une protection temporaire, une récompense, un privilège lui permettant de bénéficier

34. Soixante-dix ans, sauf dans certains pays comme le Canada, où la protection est toujours de 50 ans. Cette différence de 20 ans n'est pas sans provoquer certains conflits. Je vous donne un exemple. Jean-Marie Tremblay, professeur du Cégep de Chicoutimi, a récemment décidé d'enrôler une bande d'élèves dynamiques pour numériser des ouvrages de sociologie. Croyez-le ou non, à ce jour, la petite équipe de M. Tremblay a scanné 600 classiques tombés dans le domaine public au Québec et au Canada. Le problème, c'est qu'une bonne part de ces ouvrages sont toujours protégés par le droit d'auteur en France et que les Presses universitaires de France (PUF) ont menacé de poursuivre le Don Quichotte de la sociologie québécoise. Certains d'entre nous ont tenté de faire comprendre aux PUF qu'il vaudrait mieux ne pas aller sur ce terrain. Mais au cas où l'affaire déboucherait sur des accusations, nous avons alerté diverses associations, comme celle des universités du Canada, pour faire une collecte de fonds et payer un avocat à M. Tremblay.

temporairement de la jouissance exclusive de son œuvre. Assez rapidement, cette dernière devait cependant retomber dans le domaine public.

Pendant des décennies, un bon équilibre a donc existé entre les droits de l'auteur et ceux de la société. Cependant, la balance penche maintenant clairement du côté des créateurs. Le nouveau régime de propriété intellectuelle bénéficie particulièrement aux entreprises américaines et britanniques, puisque celles-ci disposent actuellement des moyens de production et de distribution les plus forts au monde (il n'est donc pas étonnant que les États-Unis aient été le principal moteur de changement dans ce domaine, mais l'Europe ne s'est pas montrée particulièrement timide ou timorée à cet égard : la directive européenne sur les bases de données, pour ne prendre que cet exemple épouvantable, en fait foi).

Il y a là un danger énorme pour la Francophonie, ainsi d'ailleurs que pour les locuteurs d'autres langues que l'anglais (à l'exception peut-être des Chinois). Cependant, je crois possible de mobiliser les nouvelles technologies dans le but de changer les règles du jeu. Pour illustrer cette thèse (ou prétention), prenons un autre exemple de domaine où les règles du jeu, en matière de propriété intellectuelle, avantagent présentement les anglophones : celui du logiciel.

Personne ne l'ignore. Dans le secteur du logiciel, Microsoft domine tellement les divers marchés du monde que, récemment encore, le mieux, pour un consommateur ou une compagnie utilisatrice, semblait être de se résigner à employer *Windows*, *Microsoft Office*, etc. De même, les concurrents de Microsoft semblaient devoir se résigner à baisser les bras.

En effet, il est clair que, malgré l'existence de règles strictes en matière de concurrence – de règles forçant Microsoft à dévoiler l'essentiel des changements apportés à son système d'exploitation –, les producteurs de logiciels de traitement de texte, de tableurs ou autres ont toute la misère du monde à concurrencer la compagnie de Bill Gates. Ils reçoivent de l'information, certes, mais ils la reçoivent plus tard que les responsables de *Microsoft Word* ou *Microsoft Excel*. De plus, certains renseignements stratégiques ne sortent jamais de chez Microsoft.

C'est un peu comme si, dans le monde du transfert ferroviaire, un fabricant de wagons possédait les voies ferrées et, de ce fait, contrôlait l'écartement des rails. Dans

de telles conditions, il deviendrait difficile de parler de marché sain ; entre ce fabricant et ses concurrents, le jeu ne serait plus égal. Si le monde du chemin de fer fonctionne bien, c'est en partie parce que les fabricants de wagons ont tous accès à la même information et parce que l'écartement des rails obéit à un standard fixe. Le monde de l'informatique fonctionne moins bien parce qu'une entreprise a la possibilité de changer les règles du jeu facilement et rapidement.

Pour autant, l'arrivée des logiciels libres a engendré un changement de donne. Ainsi que leur nom l'indique, les logiciels libres donnent un accès sans entraves à leur code source. Ceci ne veut pas dire que le logiciel libre n'appartient à personne ; au contraire, les licences d'usage des logiciels libres, par leur simple existence, montrent clairement qu'il ne s'agit pas là de domaines publics ; au contraire, les cadres juridiques de la propriété intellectuelle sont utilisés à fond pour assurer un effet de propagation « virale » : toute personne s'appuyant sur un logiciel libre pour créer un objet nouveau, peut-être commercial, doit aussi relâcher le code source de ses propres ajouts.

Souvent créés par des communautés de programmeurs travaillant de façon répartie ou distribuée – et virtuellement – à leur amélioration, mais aussi par des sociétés privées, les logiciels libres ne sont pas toujours gratuits, à l'inverse de ce que beaucoup pensent, mais ils ne sont jamais très chers et, surtout, ils offrent une possibilité de contrôle et d'adaptation aux nécessités locales que les logiciels non libres – on dit souvent « propriétaires » - ne peuvent guère concurrencer. De surcroît, ils se révèlent souvent en mesure de remplacer avantageusement les outils de Microsoft. Par exemple, *Linux* et un bureau comme KDE ou Gnome remplacent aisément *Windows* tandis qu'*Open Office* peut se substituer à *Microsoft Office*.

Il y a là une leçon importante à tirer. En effet, le succès du logiciel libre ne repose pas sur la force financière de ses créateurs, mais bien sur leur capacité à travailler de concert. *Linux*, *Open Office*, etc. évoluent parce qu'une synergie forte existe entre des personnes dispersées sur la planète, des personnes compétentes qui ont accès à toute l'information.

Pourquoi ces personnes contribuent-elles au développement du logiciel libre ? Pour deux raisons. Premièrement, comme c'est le cas dans l'univers de la publication scientifique, elles font don des résultats de leur travail à la communauté parce que cela les fera connaître,

leur permettra d'établir leur réputation. Deuxièmement, elles contribuent à l'élaboration des logiciels libres parce qu'elles ont l'espoir de créer des produits ou des services commerciaux qui pourront s'y arrimer. Il ne s'agit donc pas d'un calcul utopique ou idéaliste, mais bien d'un calcul très intelligent et, à ce titre, parfaitement viable sur le plan économique, institutionnel et social, tout aussi viable que l'activité scientifique depuis trois siècles.

Je viens de faire un long détour par le logiciel libre pour vous expliquer qu'il faut penser à la langue anglaise – notez que je dis ceci sans hostilité aucune envers cette langue que je pratique et que j'aime beaucoup – *comme s'il s'agissait d'une sorte de Microsoft linguistique*. Quand on utilise cette grille, on se rend compte que, pour améliorer le positionnement de la langue française, il faudra utiliser des stratégies à la mesure de nos moyens et de nos possibilités. Des stratégies qui ressemblent à celle du logiciel libre.

Sommes-nous prêts à ce virage? Je ne suis pas sûr. En effet, en matière de protection de la propriété intellectuelle, le monde francophone a évolué plus vite encore que le monde anglophone, peut-être à cause de la tradition particulière du droit d'auteur qui met davantage l'accent sur l'auteur que le *copyright* ne le faisait à l'origine. En fait, on ressent, en France et ailleurs, une énorme crispation chez les acteurs du commerce culturel qui se sentent fortement menacés par tout relâchement des droits de l'auteur³⁵.

Pourtant, il va falloir, pour tenter de sauver le rôle mondial de la langue française, lui reconquérir une place de choix; pour ce faire, je propose de mobiliser Internet tout comme Robert Chaudenson veut mobiliser la radio pour restaurer l'importance du français dans le pays à forte raison (et à raison de préconiser cette solution).

Il ne s'agit pas de mobiliser Internet par le truchement des grands diffuseurs culturels sur lesquels la France fondait tant d'espoir il n'y a pas si longtemps. Si l'on se limite à employer les méthodes traditionnelles de projection des cultures francophones, à espérer que les grands groupes médiatiques francophones feront le travail, le français connaîtra, je pense, le sort du néerlandais. La vraie Francophonie ne se fera pas au travers des Vivendi ou des

France Télécom de ce monde, et le Québec n'y arrivera pas non plus en s'appuyant, par exemple, sur Quebecor.

De même, il ne s'agit pas de demander aux États de jouer aux maîtres d'œuvre de la production francophone à côté des grands éditeurs commerciaux. D'abord, ceci nous entraînerait vers des querelles interminables entre les gouvernements et l'entreprise privée, des disputes du genre: «Ceci n'est pas le rôle du public, c'est aux entreprises de faire ce genre de choses.» Ensuite, le centralisme, dans le champ culturel, ne marche pas.

En revanche, si l'on se met à considérer que la langue française nous appartient à tous et que, par conséquent, *tous* nous devons être impliqués dans la création; si l'on en arrive à imiter dans le secteur linguistique ce qui est en train d'être accompli dans le domaine logiciel, le français, je le crois fermement, pourra continuer de s'épanouir dans le monde.

Concrètement, qu'est-ce que ceci signifie? Je vous donne un exemple. J'ai souvent dit qu'une des choses que nous pourrions faire au Québec, étant donné la structure sociale de ce pays, la façon dont les gens se raccordent les uns aux autres, serait de créer une encyclopédie *par la base*. Cette encyclopédie serait produite, par le truchement d'Internet, par les habitants de nos villes, de nos villages, de nos quartiers. Elle traiterait de ce qui les intéresse. Elle s'échafauderait petit à petit, de jour en jour, d'année en année. En bout de ligne, ce travail collectif et distribué donnerait une idée de ce qu'est le Québec et de ce qu'il pourrait être.

De même, j'ai proposé que l'on utilise le modèle du logiciel libre lors de la numérisation rétrospective des revues savantes en France.

Pourquoi ne pas étendre ces projets à l'échelle de la Francophonie? Pourquoi ne pas lancer un mouvement en ce sens? Mais il faut aller vite: Wikipedia offre déjà 500 000 rubriques encyclopédiques en plusieurs douzaines de langues. L'anglais, à lui tout seul, en concentre plus de la moitié, tandis que le français ne dispose que d'environ 25 000 rubriques, soit à peine 5% du total.

Voilà donc, en gros, les deux thèmes que je voulais développer aujourd'hui. Les technologies, dont Internet, sont en train de modifier le rapport entre les langues, mais elles pourraient servir à la relance de la langue française dans le monde. En effet, en créant un mode de production, de circulation et de stockage des contenus en français

35. Les éditeurs récupèrent le droit d'auteur au moment de la publication de l'ouvrage d'un auteur. Celui-ci renonce en effet à tous ses droits, sauf les droits dits moraux, comme le droit de repentir (le droit de repentir permet à l'auteur de demander que son œuvre soit retirée du marché s'il n'est pas satisfait de ses effets sur la société).

fonctionnant sur le modèle du logiciel libre, nous pourrions rêver à une présence francophone forte. En d'autres mots, l'avenir du français repose sur une appropriation révolutionnaire d'une révolution technique, rien de moins.

*Jean-Claude Guédon,
professeur, Université de Montréal, Québec.*

Si les logiciels commerciaux ne parlent pas votre langue, les logiciels libres le pourront peut-être

« Les logiciels libres sont attirants à de nombreux égards. Leur adoption peut mener à des réductions de coût, la résolution de problèmes de sécurité et une diminution de la dépendance face à un fournisseur. Mais ces outils présentent un autre avantage : parce qu'ils peuvent être modifiés librement, ils sont plus faciles à traduire, ou localiser, dans une langue donnée.

« Les grands fournisseurs ont peu de raisons de soutenir les petites langues du monde. Microsoft, par exemple, commercialise 24 versions linguistiques de Windows 2000, 33 de Windows XP et 20 de la suite Office. Cependant, dans le cas de nombreuses langues, les éditeurs concluent souvent qu'il n'est pas viable économiquement de produire la version localisée d'un produit.

« Les programmeurs qui travaillent à la localisation de logiciels libres fonctionnent cependant selon d'autres règles. Ensemble, les interfaces les plus populaires du système [... d'exploitation] Linux – KDE et GNOME – sont disponibles en deux fois plus de langues que Windows. KDE a été localisé en 42 langues et 46 nouvelles versions sont en route. Mozilla, [...] fureteur libre, parle maintenant 65 langues ; bientôt, il en saura 34 de plus. OpenOffice [...] est disponible en 31 langues dont le slovène, le basque, le galicien et des langues indiennes comme le gujrati, le devanagari, le kannada et le malayalam. Quarante-quatre langues, dont l'islandais, le lao, le latvien, le gallois et le yiddish seront bientôt traitées.

« Localiser un logiciel est un boulot d'envergure, mais certaines personnes sont assez passionnées pour recourir à des mesures inhabituelles. La traduction hongroise d'OpenOffice allait trop lentement au goût de Janos Noll, fondateur de la Fondation hongroise pour le logiciel libre. Il a donc élaboré des outils Web et organisé une fête dans la salle informatique de l'Université technique de Budapest. Une douzaine de personnes y ont travaillé pendant que quelque 100 Hongrois soumettaient leur travail par le truchement du Web. Le gros du travail [...] fut complété en trois jours ».

Tiré de Anonyme (2003), « Open Source's Local Heroes », *The Economist Technology Quarterly*, 6 décembre, pp. 3-4.



Atelier sur la recherche

Formation et recherche dans le domaine du traitement automatique des langues en contexte universitaire

Dans cette intervention, nous nous pencherons sur le rôle spécifique que jouent ou que peuvent jouer les institutions universitaires dans la mise au point de ressources liées au traitement automatique des langues (TAL).

En contexte universitaire, la recherche et le développement dans n'importe quelle discipline sont étroitement liés à l'importance accordée à la formation dans ce domaine. En effet, l'intérêt porté à un secteur se traduit par l'embauche de spécialistes (chercheurs et professeurs), la création de nouveaux cours, l'encadrement d'étudiants aux cycles supérieurs, la tenue de séminaires, etc. Ainsi, la plus grande partie de notre propos s'articulera autour des différentes activités liées à la formation en traitement automatique des langues.

Formation et universités

Aujourd'hui, la plupart des universités offrent des formations en TAL ou dans des champs connexes qui font appel à ses techniques. Nous distinguons ici deux types de formation. Les universités peuvent offrir des cours portant spécifiquement sur le traitement automatique des langues (ou la linguistique informatique), dans lesquels les étudiants se familiarisent avec les techniques symboliques et statistiques conçues pour traiter différentes composantes de la langue (phonologie, morphologie, syntaxe, sémantique et lexique). On y trouve également des cours portant sur d'autres thématiques, mais qui abordent des notions liées au TAL. Nous pensons ici à des cours sur la recherche d'information, les outils d'aide à la traduction, la terminologie ou à des cours de linguistique plus classiques, comme la lexicographie. Ces cours ne portent pas spécifiquement sur le traitement automatique des langues, mais font appel à des notions relevant de cette discipline pour expliquer des applications.

Toutefois, les modalités varient d'un établissement d'enseignement à l'autre. Nous avons choisi de décrire la situation qui règne dans un département de linguistique et de traduction, à savoir celui de l'Université de Montréal. D'une part, il s'agit de la situation que nous connaissons le mieux. D'autre part, ce département est sans conteste représentatif d'autres départements nord-américains où on

enseigne la traduction ou la linguistique. Deux précisions s'imposent toutefois :

1. La formation et la recherche en TAL sont rarement confinées à un seul département universitaire. Souvent, les départements d'informatique hébergent des chercheurs intéressés directement ou indirectement par le TAL (par exemple, la traduction automatique, ou encore la recherche d'information (RI) qui intègre un nombre croissant de traitements morphosyntaxiques et sémantiques). On peut également citer d'autres secteurs qui comportent des volets qui peuvent être associés au TAL : des travaux en enseignement des langues qui visent à mettre au point des didacticiels intégrant des analyses linguistiques (par exemple, des règles de grammaire ou des représentations sémantiques) ; des travaux en sciences de l'information (par exemple, indexation automatique, génération semi-automatique de thésaurus). Par ailleurs, des collaborations peuvent s'établir entre des chercheurs affiliés à ces divers départements ;
2. La configuration du Département de linguistique et de traduction est un peu particulière en ce sens qu'elle offre deux séries de programmes bien distincts. Dans d'autres universités, la traduction et la linguistique peuvent être offertes par deux entités autonomes ou encore être rattachées à d'autres disciplines. Par exemple, la traduction est souvent offerte dans des départements de littérature comparée. La situation un peu particulière de l'Université de Montréal nous forcera à aborder les liens avec le TAL à partir des deux disciplines.

Formation TAL en linguistique

En linguistique, le traitement automatique des langues est souvent présenté comme un des domaines présentant les meilleures perspectives d'emploi. À ce titre, Montréal jouit d'une situation privilégiée, puisqu'on y trouve des entreprises travaillant dans des domaines compatibles avec le TAL : traitement automatique de la parole, traduction, correction automatique, etc. Les départements universitaires offrant des formations en linguistique ont tout intérêt à s'adapter à cette réalité. Par ailleurs, même dans des champs classiques, l'apprentissage d'applications informatiques est désormais obligatoire. Par exemple, on s'attend désormais d'un lexicographe qu'il maîtrise toute une série d'outils

logiciels, qu'il puisse manipuler des corpus de grande taille, les annoter et les interroger.

Parmi les domaines du TAL dans lesquels il sera important de former les futurs diplômés en linguistique, citons le traitement automatique de la parole (synthèse ou reconnaissance), la morphologie et la syntaxe computationnelles (les étudiants devraient pouvoir mettre au point des grammaires et des analyseurs morphologiques), la lexicographie informatisée (apprendre à enrichir ou à élaborer des dictionnaires électroniques, ou des bases de données lexicales directement utilisables dans des programmes de traitement automatique). En outre, les étudiants devraient être en mesure de faire des descriptions fines dans des langues moins souvent décrites, des langues pour lesquelles on ne trouve pas de ressources déjà construites (comme des grammaires ou des dictionnaires). Une formation en statistique linguistique semble également incontournable, même si cette dernière formation est sans doute plus approfondie dans les départements d'informatique. L'encadré 1 montre quels sont les cours offerts dans les programmes de linguistique de l'Université

ENCADRÉ 1
COURS DE LINGUISTIQUE AVEC UNE ORIENTATION TAL
(UNIVERSITÉ DE MONTRÉAL)

- 1^{er} cycle (cours à option): *Introduction aux langages formels, Grammaires formelles, Atelier de programmation linguistique, Traitement automatique du langage, Introduction à la lexicométrie*, quelques cours d'informatique;
- 2^e cycle (cours à option): *Linguistique informatique, Traduction automatique, Terminologie et ordinateur* (option linguistique appliquée – Terminologie);
- 3^e cycle (Ph. D. en linguistique ou option Intelligence artificielle): *Phonologie computationnelle, Morphologie computationnelle, Syntaxe computationnelle, Sémantique computationnelle*.

Quatre des 18 professeurs de linguistique du département se partagent ces cours et forment les étudiants de deuxième et de troisième cycles dans le secteur du TAL³⁶.

de Montréal. Il convient de souligner qu'actuellement, les cours de TAL sont optionnels au premier cycle.

Évidemment, cette initiation aux multiples facettes du TAL repose nécessairement sur une formation fondamentale en linguistique. La difficulté consiste alors à trouver un équilibre entre les domaines fondamentaux de la linguistique (phonétique, phonologie, morphologie, syntaxe, lexicologie, sémantique) et les applications spécifiques qu'on en fait en traitement automatique des langues.

En outre, les formateurs doivent s'adapter aux multiples changements que subissent les domaines du traitement automatique des langues. Par exemple, ces dernières années, les chercheurs semblent accorder une plus grande priorité à la mise au point et à la diffusion de ressources linguistiques en format électronique, c'est-à-dire des corpus textuels, des dictionnaires enrichis d'information morphosyntaxique et sémantique, au détriment d'une formation plus classique aux théories et aux formalismes linguistiques ou logiques.

Enfin, les formateurs en linguistique doivent prévoir des contenus de cours qui se démarquent des formations en TAL données dans d'autres disciplines (en informatique ou en sciences de l'information) et trouver des manières de valoriser une formation qui intègre une composante forte en linguistique.

Formation TAL en traduction

En traduction, les besoins de formation se présentent différemment. Les traducteurs sont avant tout des utilisateurs d'applications informatiques et rares sont ceux qui deviendront des concepteurs véritables. Le rôle des formateurs consiste alors à les initier à la multitude d'outils d'aide à la traduction qui orneront éventuellement leur poste de travail (bases de données, logiciels de terminologie, mémoires de traduction, aligneurs de phrases, correcteurs automatisés, fonctions avancées du traitement de texte, concordanciers, etc.).

À l'heure actuelle, toutes les universités offrant des programmes de traduction sont tenues de rendre obligatoires des cours portant sur les outils d'aide à la traduction. Les employeurs s'attendent en effet de leurs futurs employés qu'ils soient en mesure d'en faire une

36. On trouvera d'autres détails, notamment la description des cours, dans le site Web du département: <http://www.ling.umontreal.ca>

utilisation efficace dès l'embauche. On offre également, depuis quelques années, des programmes de formation en localisation dans lesquels les étudiants sont appelés à se familiariser avec différentes techniques de traduction de logiciels ou de pages Web. L'Université de Montréal ne fait pas exception et offre un microprogramme en localisation qui s'adresse à des traducteurs en exercice désireux de se familiariser avec ces nouvelles techniques.

L'encadré 2 montre quels sont les cours offerts dans les programmes de traduction de l'Université de Montréal. Le cours de premier cycle intitulé *Outils informatiques des langagiers* est obligatoire ainsi que les cours de localisation (pour les étudiants inscrits au microprogramme). Les autres cours, toutefois, sont optionnels.

ENCADRÉ 2

COURS DE TRADUCTION AVEC UNE ORIENTATION TAL (UNIVERSITÉ DE MONTRÉAL)

- 1^{er} cycle: *Outils informatiques des langagiers* (option: Initiation à la localisation et quelques cours d'informatique);
- 2^e cycle: (option: *Traductique*, possibilité de suivre quelques cours de linguistique);
DESS en traduction: *Outils informatiques des langagiers*
Microprogramme en localisation: *Initiation à la localisation, Atelier de localisation, Projet personnel en localisation* (option: *Traductique*, cours de bibliothéconomie).
- 3^e cycle: (option: *Traductique*, possibilité de suivre des cours de linguistique).

Deux des 11 professeurs de traduction du département se partagent ces cours et forment les étudiants de deuxième et de troisième cycles dans le secteur du TAL³⁷.

En traduction, le défi principal des formateurs consiste à intégrer dans les cours quelques notions de traitement automatique des langues et à se détacher d'un enseignement qui ressemble plutôt à une série de modes d'emploi de logiciels. Ici aussi, les formateurs doivent suivre l'évolution

dans le domaine et sont sans cesse appelés à moduler le contenu des cours.

Les départements peuvent tenter d'encourager une partie des étudiants à poursuivre des études de deuxième et de troisième cycles en traductique, même si les programmes de traduction ont une forte orientation professionnelle. Les étudiants intéressés à poursuivre leurs études complètent généralement leur formation de premier cycle par des cours de linguistique formelle. Ceux qui ont fait le choix de se spécialiser en traductique peuvent suivre différents cours à option comme *Linguistique informatique* ou *Traduction automatique*. Ils ont aussi la possibilité de suivre certains cours d'informatique.

Souvent, dans les programmes de traduction, la terminologie constitue la porte d'entrée vers ces secteurs compatibles avec le TAL. Les cours de terminologie sont ceux où les étudiants commencent à entendre parler des dictionnaires électroniques, de corpus électroniques et à se familiariser avec les notions de base de la lexicologie.

Recherche en TAL en linguistique et traduction

En contexte universitaire, la recherche dans le domaine du TAL s'organise de différentes manières. D'abord, des laboratoires peuvent consacrer la totalité ou une partie de leurs activités au traitement automatique des langues. Par ailleurs, d'autres chercheurs peuvent mener des recherches individuelles sur un aspect ou un autre du TAL. Enfin, comme nous l'avons déjà dit plus haut, les collaborations peuvent s'instaurer entre les chercheurs d'un département de linguistique et de traduction et d'autres départements, ou encore avec des groupes extérieurs à l'université (des entreprises privées, d'autres universités canadiennes ou des universités étrangères).

À l'Université de Montréal, tous ces cas de figure sont représentés. Un premier groupe de recherche, l'Observatoire de linguistique Sens-Texte (OLST)³⁸, consacre une partie de ses activités à des applications relevant du TAL. Il met un

37. On trouvera d'autres détails, notamment la description des cours, dans le site Web du département: <http://www.ling.umontreal.ca>

38. L'adresse du site Web du groupe OLST est la suivante: <http://www.fas.umontreal.ca/LING/olst/>.

accent particulier sur les modélisations lexicales et la sémantique formelle. Il consacre par ailleurs une partie de ses activités à la confection de ressources électroniques, comme des corpus généraux et spécialisés, des concordanciers pour les interroger, des corpus annotés, des ressources lexicales et terminologiques. Dans d'autres groupes de recherche, on mène des recherches dans des domaines de la linguistique compatibles avec certains domaines du TAL, comme la phonétique ou la syntaxe formelle. Enfin, des liens de collaboration sont établis entre le Laboratoire de Recherche Appliquée en Linguistique Informatique (RALI)³⁹ du Département d'informatique et de recherche opérationnelle sous la forme de séminaires hebdomadaires auxquels sont conviés étudiants de 2^e et de 3^e cycles, chercheurs universitaires et entrepreneurs⁴⁰.

Le financement de ces activités de recherche repose principalement sur des fonds octroyés par des organismes gouvernementaux, comme le Conseil de recherche en sciences humaines du Canada (CRSH), le Fonds québécois de la recherche sur la société et la culture (FQRSC), le Conseil de recherches en sciences naturelles et génie du Canada (CRSNG). Il est à noter que le gouvernement fédéral du Canada a fait du vaste domaine des industries de la langue, un domaine de développement prioritaire en 2001. Les chercheurs peuvent également obtenir des fonds auprès d'organismes internationaux comme l'Agence universitaire de la Francophonie (AUF) ou en concluant des ententes avec des entreprises privées.

La présence de groupes de recherche et l'instauration de collaborations avec des entreprises a un effet direct sur l'intérêt des étudiants pour le domaine du TAL. Les données des tableaux 1 et 2 donnent une idée de l'attrait que présente ce secteur pour la relève en linguistique et en traduction. Le tableau 1 montre que le cinquième (22 %) des travaux menés dans un département de linguistique comme le nôtre le sont dans le domaine du TAL. En traduction, la proportion est moins importante, à 10 % (tableau 2). Ce faible taux s'explique notamment au fait que la formation à l'utilisation des outils informatiques est assez récente dans notre département et dans d'autres départements québécois ou canadiens.

39. L'adresse du site Web du groupe RALI est la suivante : <http://www-rali.iro.umontreal.ca/>.

40. Le programme des séminaires est affiché à l'adresse suivante : <http://www-rali.iro.umontreal.ca/seminaires.html>.

Tableau 1
répartition des mémoires et des thèses en linguistique (1996-2002)⁴¹

	Mémoires	Thèses
TAL	4	5
Compatibles avec TAL	8	8
Non TAL	46	42

Tableau 2
répartition des mémoires et des thèses en traduction (1996-2002)⁴²

	Mémoires	Thèses
TAL	6	–
Compatibles avec TAL	2	1
Non TAL	53	30

Quelques pistes pour stimuler les recherches en TAL dans les universités

Nous croyons que les universités ont un rôle à jouer afin de stimuler l'intérêt des étudiants pour le domaine du traitement automatique des langues (TAL). Les programmes de linguistique et de traduction représentent souvent le premier contact des étudiants avec ce secteur.

Un premier moyen permettant de stimuler les recherches dans le domaine du TAL consisterait à donner aux étudiants de premier cycle – au Québec, les étudiants du baccalauréat – un certain nombre de cours obligatoires dans le domaine du TAL. En linguistique, la formation gagnerait à être plus *générale* : elle porte parfois sur des modèles élaborés pour régler des problèmes linguistiques hautement pointus, sans toujours prévoir un enseignement des techniques de base, comme les méthodes de traitement de chaînes de caractères ou d'élaboration de dictionnaires électroniques. Les encadrés 3 et 4 résument les quelques pistes que nous venons d'évoquer. En traduction, les formateurs pourraient intégrer davantage de notions fondamentales lorsqu'ils abordent les aides à la traduction.

41. Thèses et mémoires terminés ou en cours.

42. Thèses et mémoires terminés ou en cours.

Souvent, on a tendance à focaliser sur les applications en tant que telles, sans les placer dans un contexte plus global. L'étudiant peut devenir un expert dans la manipulation d'un produit commercial particulier, mais n'est pas toujours en mesure de généraliser des modes de fonctionnement à des familles d'applications⁴³.

Il conviendrait également de mieux baliser les enseignements donnés aux futurs acteurs du secteur du TAL. Souvent, la linguistique, la traduction et l'informatique semblent être des disciplines concurrentes. Il faudrait assurer, plutôt, que les formations reçues par les linguistes, les traducteurs et les informaticiens sont complémentaires.

Les suggestions faites pour bonifier la formation dans le domaine du TAL sont résumées dans les encadrés 3 et 4.

ENCADRÉ 3
FORMATION DE PREMIER CYCLE:
QUELQUES PISTES INTÉRESSANTES

- Rendre certains cours en TAL obligatoires au premier cycle;
- Enseigner certaines techniques de base en TAL aux étudiants;
- Mieux définir le rôle des linguistes et des traducteurs dans la conception d'outils de TAL;
- Souligner le rôle central des applications informatiques dans les cours plus traditionnels (ex.: traduction, terminologie, lexicographie);
- Rappeler les liens existant entre la *formation traditionnelle* et l'intégration d'applications informatiques;
- Encourager la spécialisation aux deuxième et troisième cycles seulement.

ENCADRÉ 4
FORMATION DE 2^e et 3^e CYCLES:
QUELQUES PISTES INTÉRESSANTES

- Former les étudiants à l'utilisation de programmes de traitement des données linguistiques (corpus, dictionnaires);
- Fournir des plates-formes de développement de ressources linguistiques formelles (analyseurs, grammaires ou entrées lexicales);
- Faire participer les étudiants à des descriptions linguistiques;
- Définir des projets de recherche débouchant sur des applications concrètes.

En recherche, quelques suggestions pourraient stimuler les activités liées au TAL. Il importerait d'abord de bien définir le rôle du chercheur universitaire en linguistique, en traduction ou en informatique. Actuellement, certains programmes de financement exigent de la part des chercheurs – pour qu'ils soient admissibles – qu'ils s'engagent à livrer un produit à la fin de leur projet. Il faut se demander si l'université est vraiment l'endroit pour faire ce genre de travail, puisqu'elle concurrence ainsi directement le secteur privé. On peut penser que le rôle du secteur universitaire consiste davantage à œuvrer en amont de l'entreprise – à contribuer à l'élaboration d'applications dans une optique à long terme, en travaillant sur des questions de recherche fondamentale.

En terminant, il importerait pour les laboratoires de faire un meilleur partage de leurs ressources. Trop souvent, les laboratoires travaillent isolément, rechignent à partager les données linguistiques dont ils disposent avec le voisin ou créent des ressources incompatibles avec les siennes.

Marie-Claude L'Homme,
Observatoire de linguistique Sens-Texte (OLST),
Université de Montréal, Québec.

43. Quelques solutions à ce problème sont proposées dans le cadre du cours Outils informatiques des langagiers. Voir Robichaud, Benoît et Marie-Claude L'Homme (2003), « Teaching the Automation of the Translation Process to Future Translators », dans Workshop on Teaching Translation, Technologies and Tools, XV Machine Translation Summit, 23-27 septembre 2003 et L'Homme, Marie-Claude (2003), « Traduction et outils informatiques: mise en forme d'un cours d'initiation », dans Mareschal, Geneviève et autres (éd.), La formation à la traduction professionnelle, Ottawa: Les Presses de l'Université d'Ottawa, pp. 177-197.

Traitement informatisé des langues, langues nationales et partenariat francophone

Je participe à cette manifestation avec plusieurs casquettes à la fois: je porte celle de chercheur à l'Université de la Manouba, Tunisie, celle de membre du Groupe de recherche expérimentale sur les systèmes informatisés de la communication de l'Université Bordeaux 3 (GRESIC) et celle de secrétaire général de l'Association internationale des écoles des sciences de l'information. Chacun de ces profils m'a conduit, au fil des ans, à faire des recherches sur les questions du multilinguisme, l'un des axes fondamentaux des travaux menés dans le secteur du traitement informatisé des langues (TIL).

Je commence toujours ce genre d'intervention en faisant un recours à l'exemple mythologique de la tour de Babel. Comme vous le savez, le mythe de Babel prétend que les hommes avaient décidé un jour de construire une énorme tour pour se rapprocher du Ciel. Devant cette arrogance humaine, Dieu décida de punir leur vanité en faisant en sorte qu'ils ne puissent plus parler la même langue et, de la sorte, ne plus se comprendre. L'édifice resta inachevé. Et la Terre se transforma en un univers multilingue et multiculturel.

Quels seront les effets des technologies de l'information (TI) dans un contexte pareil? Renforceront-elles les partenariats linguistiques et les échanges interculturels pour essayer de pallier la conséquence de ce « péché originel »? Ou au contraire, amplifieront-elles les inégalités entre les langues des groupes majoritaires (dominantes) et celles des communautés minoritaires (soumises), comme les Basques, les Kurdes ou les Berbères?

Je crois qu'une approche multilingue pourrait assurer que la première réponse prévaut. Je qualifie cette approche de *multilinguisme numérique du Sud*. Permettez-moi d'expliquer ce que j'entends par cette idée.

Actuellement, on remarque que des langues comme l'arabe, l'hébreu ou l'ourdou sont des langues statistiquement importantes sur le plan démographique, mais qu'elles sont *technologiquement minoritaires*, en ce sens que leur traitement informatique n'est pas très avancé par comparaison avec les acquis des langues latines.

D'abord, lorsque ces langues sont prises en compte, elles ne sont jamais les seules composantes d'une application. Par exemple, lorsqu'un outil traite l'arabe, c'est qu'il a l'anglais ou le français comme langue pivot. La situation est bien différente dans le cas de ces deux

dernières langues: la nature unilingue (française ou anglaise) quasi dominante du marché mondial du logiciel atteste de ce déséquilibre linguistique et de l'autonomie de certaines langues devant la dépendance d'autres.

On note également que les développeurs de logiciels et les fabricants de matériel informatique ne tiennent pas toujours bien compte, dans leurs produits, de certains aspects culturels, sociologiques ou même cognitifs importants pour les locuteurs du Sud. Par exemple, les logiciels d'origines *latines*, adaptés pour la langue arabe, présentent souvent un nombre assez important d'incohérences aux yeux de l'utilisateur arabophone type. Ces incohérences peuvent avoir de grandes répercussions sur l'acceptabilité, l'utilisabilité et l'efficacité même de l'interface.

La source du problème d'interface émane du fait qu'il existe deux types de multilinguisme: un multilinguisme *souple*, très répandu et couvert par la recherche, et un multilinguisme *lourd*, peu documenté et moins connu. Le multilinguisme souple renvoie à la coexistence de deux ou plusieurs langues latines ou germaniques (i.e. français, anglais) dans un seul espace applicatif ou logiciel donné. Il est relativement facile à gérer par les développeurs et à appréhender par l'utilisateur.

Le multilinguisme lourd, par contre, oppose deux ou plusieurs familles de langues dans le même contexte logiciel ou applicatif sur la base des deux points majeurs: le rendu visuel de la forme graphique des caractères (i.e. graphie latine ou arabe) et la directionnalité de l'acte graphique (directionnalité Gauche-Droite ou Droite-Gauche). Ce cas de figure est nettement plus difficile à gérer, ce qui explique les difficultés toujours enregistrées au niveau de la gestion d'interface homme-machine et des systèmes internes de traitement des données codées.

Ce phénomène de bidirectionnalité, par exemple, est très manifeste dans les cultures et les langues à directionnalité opposée par rapport à ce qui est courant dans les cultures occidentales à directionnalité Gauche-Droite. Dans la culture arabe, on prononce en théorie les nombres en passant des unités aux dizaines, aux centaines, aux milliers, etc., évoluant dans une directionnalité Droite-Gauche. En principe, on écrit ces mêmes nombres également de droite à gauche. Par exemple, le chiffre 1988 se prononce « huit quatre-vingt neuf cent mille » et s'écrit dans la même orientation Droite-Gauche « 1988 », ce qui

donne un état de conformité entre l'acte d'écriture, le mode de prononciation et le modèle de calculabilité arithmétique. Cette pratique constitue encore de nos jours l'une des marques de distinction de la bonne maîtrise de la langue arabe.

Pourtant, dans toutes les solutions multilingues proposées par les grandes instances internationales de normalisation, les chiffres sont exclusivement traités en partant du plus gros au plus petit, de gauche à droite. En français (et autres langues latines), le chiffre « 1988 » est prononcé « mille neuf cent quatre-vingt-huit » et rédigé dans une logique directionnelle Gauche-Droite « 1988 ». Ceci engendre un état de non-conformité entre l'acte graphique et le mode de prononciation d'une part (G-D), et le mode de calculabilité arithmétique de l'autre (D-G). Il est certain que les personnes d'origine arabe éduquées dans les écoles occidentales s'adaptent relativement bien à ce système directionnel Gauche-Droite; sauf que ces personnes ne représentent qu'une infime portion de la population arabophone. Il faudrait, pour généraliser l'utilisation de l'ordinateur et des technologies numériques par cette dernière, s'adapter à ses attentes linguistiques et cognitives.

En somme, un certain nombre de questionnements se pose actuellement devant un déséquilibre évident entre les langues universelles :

- comment gérer l'aspect linguistique intensif du contenu informationnel sur le réseau (codage des langues)?
- comment ont été abordés les deux principes de l'internationalisation (I18n) et de la localisation (L10n) dans les systèmes et les ressources d'information multilingues (langues complexes)?
- quels sont les apports de l'I18n pour les langues « complexes » (c'est-à-dire arabe) au sein des systèmes d'information multilingues?
- quelle forme de partenariat peut exister entre le français comme langue pivot et l'arabe comme sa première langue partenaire du Sud?

Il est certes évident que le traitement automatique des langues du Sud n'est pas au même seuil d'importance que les langues du Nord. Il suffit de constater l'assise linguistique des produits logiciels dans le marché pour conclure que les langues dites « complexes » n'ont pas une présence massive et, si elles sont prises en compte, elles ne le sont pas toujours adéquatement. La Francophonie devrait

s'efforcer de corriger ces lacunes. Quelle devrait être sa stratégie pour ce faire?

Pour favoriser l'informatisation des langues du Sud, il faudra d'abord régler certains problèmes terminologiques. En effet, les langues partenaires sont habituellement plus pauvres à ce chapitre que celles du Nord. Bien sûr, il y a des causes endogènes à cela, mais il y a aussi des causes extérieures.

Il conviendrait aussi à la Francophonie, pour combattre l'hégémonie des grandes compagnies de logiciels, d'appuyer les initiatives de développement de solutions locales par le renforcement des initiatives d'internationalisation (I18n) et localisation (L10n). Ces initiatives devraient viser le traitement des langues médianes du Sud. Les langues médianes représentent un assemblage entre les langues officielles et les dialectes. Ces langues transversales s'adressent à tout le monde; elles sont de plus en plus utilisées dans les journaux, dans les revues, etc.

Cependant, je suis réaliste et je me rends compte que la plupart des logiciels continueront de provenir des pays du Nord, en particulier des États-Unis. Je me permets cependant d'espérer que la localisation de ces outils sera faite, à l'avenir, par des personnes-ressources vivant dans le marché linguistique et culturel visé, plutôt que par des experts n'ayant qu'une connaissance théorique de la langue cible.

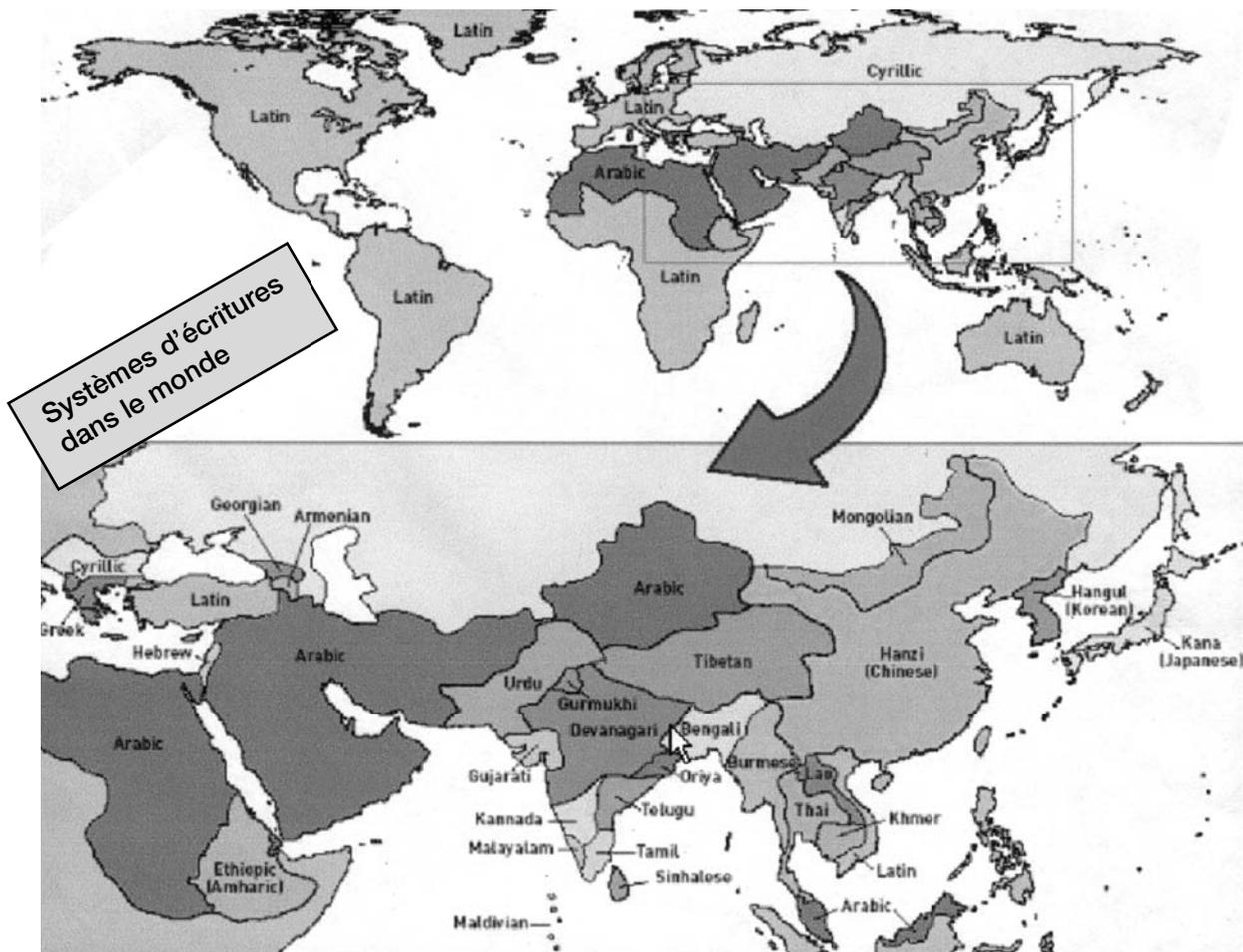
Enfin, le français devra chercher à occuper le terrain comme langue pivot, sinon l'anglais le fera à sa place. Par exemple, pendant plusieurs années, les arabophones ont disposé de versions arabe/français des systèmes d'exploitation de Microsoft, tel *Windows 95*. Cependant, *Windows 98* fut seulement offert en version anglais/arabe. Résultat: rapidement, on a remarqué que les étudiants se sont mis à utiliser une terminologie anglophone. Avec la pénétration d'Unicode, les choses se sont corrigées un peu, mais il reste des choses à faire.

Dans une veine un peu différente, il arrive souvent qu'il faille d'abord, pour traduire un texte depuis l'arabe, le traduire en anglais, parce qu'aucune passerelle directe n'existe entre l'arabe et le français.

Rien de cela ne sera possible sans l'intensification de la collaboration internationale entre les équipes de chercheurs et les équipes de développeurs, pour assurer que l'on tient bien compte des attentes linguistiques et culturelles des

locuteurs du Sud, au niveau phonétique, syntaxique, sémantique, et ainsi de suite.

Illustration 1
le contexte linguistique mondial



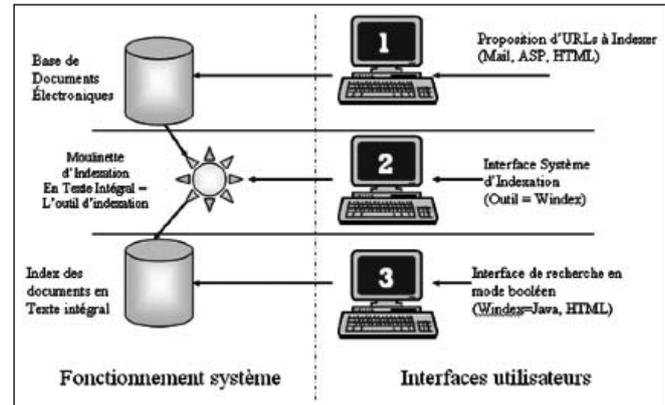
Je termine cette intervention par la citation d'un projet conduit dans le contexte de la Francophonie dans lequel on essaye de renforcer le partenariat entre le français et la langue arabe. Il s'agit du projet Cyberprof, développé sous forme d'une application pédagogique pour les enseignants en information et en communication dans les écoles des sciences de l'information francophone du Sud. Cette application a pour but d'enseigner la matière de la

recherche de l'information sur intranet en langue française. Il s'agit d'une application qui consiste à développer un corpus textuel francophone sur lequel est greffée une couche logicielle permettant d'indexer et de rechercher des données en utilisant les techniques des stratégies de recherche booléenne. C'est un outil qui permet aux enseignants de l'informatique documentaire de disposer d'un outil pédagogique pour conduire des travaux pratiques de

recherche d'information. Dans une deuxième phase, il est question d'incorporer des textes dans les langues nationales, de les indexer pour que ce soit un support pédagogique multilingue conforme aux règles de l'internationalisation et de la localisation. En effet, il est actuellement question de collaborer avec les concepteurs français de l'outil d'indexation pour couvrir une plus grande diversité de langues par le robot indexeur. Le partenariat francophone avec les diverses langues partenaires aura ainsi abouti à quelque chose de concret au profit de la communauté scientifique francophone du Sud.

*Mokhtar Ben Henda,
CEM-GRESIC, Université de Bordeaux 3, France, et
ISD, Université La Manouba-Tunisie.*

Illustration 2
le schéma fonctionnel de Cyberprof



La recherche sur l'informatisation des langues africaines

Comme vous le savez, le partenariat entre les langues africaines et le français s'inscrit très largement dans l'histoire. Dans plusieurs pays d'Afrique, le français a aujourd'hui le statut de langue officielle ou co-officielle. De plus, sur le plan sociologique, il occupe souvent une place si importante que personne ne pense à le remplacer, mais malgré cela, on ne peut pas se passer des langues africaines quand on veut vraiment œuvrer au développement des populations de nos pays.

En effet, même si les cadres et les techniciens de la santé, de la culture, de l'industrie, etc. sont formés en français dans des écoles francophones, dans leurs activités professionnelles quotidiennes, ils ont à côtoyer des gens qui ne connaissent pas le français. Par exemple, les vétérinaires qui travaillent dans le milieu des éleveurs ou les techniciens affectés au développement des milieux ruraux doivent transmettre leurs connaissances à des gens qui ne parlent pas le français.

Malheureusement, ces spécialistes n'ont pas appris à exprimer leur savoir dans les langues africaines. Il faut donc les y aider en rendant les langues africaines capables de supporter et de véhiculer les connaissances modernes. Voilà l'enjeu de l'aménagement linguistique en Afrique.

Pour arriver à ce résultat dans le monde informatique, il est d'abord nécessaire de faire la description linguistique des langues les plus parlées d'Afrique et de stabiliser leur système orthographique. Ensuite, il faut se préoccuper de la question de l'encodage des caractères africains.

Les deux premières questions, dans de nombreux cas, ont été résolues, mais la troisième, fréquemment, demeure. En effet, les logiciels capables de prendre les caractères africains en considération sont encore assez rares, même s'il en existe, comme ceux de la société québécoise BPI. Sur le plan international, il faudra aussi travailler à résoudre les nombreux problèmes techniques qui handicapent la normalisation, par le consortium Unicode, des caractères spéciaux nécessaires à l'écriture des langues africaines.

Voilà le premier verrou qu'il faudrait arriver à faire sauter, mais d'autres travaux seront aussi absolument nécessaires, comme la production ou la localisation de logiciels *de base* en langues africaines. Par exemple, dans le Sud comme dans le Nord, on a aujourd'hui besoin de suites bureautiques pour être productifs. Cependant, ces outils indispensables ne sont pas, le plus souvent, disponibles dans

la langue des personnes appelées à les utiliser pour produire de l'information.

Des équipes sont prêtes à travailler à la localisation en langues africaines des logiciels libres utiles aux gens du Sud. Il faudra seulement mettre en place les liens de collaboration nécessaires au succès de telles entreprises.

À côté de cela, il reste évidemment du travail de recherche à faire pour mettre au point des dictionnaires, des correcteurs et des lemmatiseurs pour les langues africaines. Certaines universités ont commencé à se pencher sur ces questions, mais leurs travaux sont peu avancés ou n'ont pas vraiment fait l'objet de diffusion.

Si l'on veut que les langues africaines réussissent à prendre leur place sur Internet, on devra aussi disposer d'outils permettant aux internautes de les utiliser pour faire de la navigation, publier des documents Web, correspondre par courrier électronique, etc. Or, quels sont les navigateurs en exercice aujourd'hui capables de reconnaître les caractères spéciaux des langues africaines? Quels sont les logiciels de messagerie capables d'envoyer un courriel en caractères éthiopiens? Il y en a très peu, sinon pas du tout.

Dans une veine différente, il serait intéressant de tenir compte, lors des travaux d'informatisation des langues africaines, des écritures qui ont été utilisées pour rédiger certains des textes anciens de l'Afrique⁴⁴, comme le tifinar, le masaba, le nsibidi, et ainsi de suite. On ne le fait pas actuellement, avec pour résultat qu'un linguiste étudiant ces écritures ne peut pas les saisir dans ses documents de recherche. Il doit se contenter de faire des copier-coller d'images.

En conclusion, je parlais plus tôt de la nécessité de localiser les logiciels en tenant compte des langues africaines. À cet égard, il faut saluer une initiative réalisée au sein du RIFAL. En effet, l'Office québécois de la langue française a développé une base de données terminologiques et textuelles multilingue et multiplate-forme, la BTML, qui tiendra compte des spécificités du tri alphabétique, tonal et structurel du lingala et du hausa. Cela illustre bien, selon moi, les bienfaits de la collaboration francophone en matière de traitement informatique des langues.

Je vous remercie.

Marcel Diki-Kidiri, chargé de recherche, CNRS, Villejuif, France.

44. On pense toujours à l'Afrique comme le continent de l'oralité, mais on y a écrit énormément de choses aussi.

La graphisation des langues africaines

Je me situerai aujourd'hui en amont des approches qui ont été proposées ici en ce sens que, avant même que de discuter du problème de l'informatisation des langues, je parlerai du problème de leur graphisation. Marcel Diki-Kidiri a bien montré ce matin que quatre langues seulement occupent la quasi-totalité de l'espace de la Toile. À l'exception de l'afrikaans, rejeton du néerlandais, les langues de l'Afrique subsaharienne en sont presque totalement absentes.

Quelles sont les raisons de l'absence de ces langues sur la Toile? Je crois pour ma part que celle-ci s'explique, en partie, du fait des codes graphiques utilisés pour les représenter par écrit. Ce handicap sera de plus en plus lourd pour toutes ces langues.

Lorsque l'on parle de la graphisation des langues, je pense toujours à cette fable d'Ésope, l'esclave phrygien; lorsque son maître lui ordonne de lui servir la meilleure chose du monde, il lui sert de la langue; le lendemain, quand son maître lui commande de servir la pire chose du monde, il lui sert encore de la langue!

Créer le code graphique d'une langue est à la fois la chose la plus facile du monde et la plus difficile du monde. La plus facile, parce que beaucoup de systèmes graphiques ont été créés par des gens qui n'avaient aucune compétence dans ce domaine. N'importe qui peut s'installer à son bureau et concevoir un système graphique.

La plus difficile, parce que dans la graphie d'une langue, le vrai problème n'est pas le **comment**, mais le **pour quoi?** Si on ne sait pas « pour quoi? » on le fait, on ne peut savoir comment le faire. Voici trois petits exemples pour l'illustrer.

Ésope et la langue

Un certain jour de marché, Xantus qui avait dessein de régaler quelques-uns de ses Amis, lui commanda d'acheter ce qu'il y aurait de meilleur, et rien d'autre chose. Je t'apprendrai, dit en soi-même le Phrygien, à spécifier ce que tu souhaites, sans t'en remettre à la discrétion d'un esclave. Il n'acheta que des langues, lesquelles il fit accommoder à toutes les sauces. L'Entrée, le Second, l'Entremets, tout ne fut que langues. Les Convies louèrent d'abord le choix de ces mets; à la fin ils s'en dégoûtèrent. Ne t'ai-je pas

commandé, dit Xantus, d'acheter ce qu'il y aurait de meilleur? - Et qu'y a-t-il de meilleur que la langue? reprit Ésope. C'est le lien de la vie civile, la Clef des Sciences, l'Organe de la Vérité et de la Raison. Par elle on bâtit les Villes et on les police; on instruit; on persuade; on règne dans les Assemblées; on s'acquitte du premier de tous les devoirs, qui est de louer les Dieux. - Et bien (dit Xantus, qui prétendait l'attraper), achète-moi demain ce qui est de pire: ces mêmes personnes viendront chez moi; et je veux diversifier. Le lendemain Ésope ne fit servir que le même mets, disant que la Langue est la pire chose qui soit au monde. C'est la Mère de tous débats, la Nourrice des procès, la source des divisions et des guerres. Si l'on dit qu'elle est l'organe de la Vérité, c'est aussi celui de l'erreur, et qui pis est, de la Calomnie. Par elle on détruit les Villes, on persuade de méchantes choses. Si d'un côté elle loue les Dieux, de l'autre elle profère des blasphèmes contre leur puissance. Quelqu'un de la compagnie dit à Xantus que véritablement ce Valet lui était fort nécessaire; car il savait le mieux du monde exercer la patience d'un Philosophe.

Source: Jean de la Fontaine, cité dans: www.mythorama.com/_contes/indexfr.php?tid=298, consulté le 1^{er} décembre 2003.

Premier exemple: celui des créoles. Depuis le XVIII^e siècle, on écrit tous les créoles français sans aucun problème, en utilisant la graphie du français, parce que les créoles lui ressemblent un peu, pour le lexique en particulier. On se tire donc d'affaire avec une graphie vaguement étymologisante. Cependant, on pourrait très bien adopter l'orthographe cyrillique pour arriver au même résultat.

Deuxième exemple: celui du malgache, langue malayo-polynésienne qu'on écrit avec des caractères latins, mais qu'on a aussi écrit, dans le passé, avec des caractères arabes.

Troisième exemple, celui des langues de plusieurs ex-républiques asiatiques de l'Union soviétique, qui ont changé de systèmes graphiques trois ou quatre fois en un siècle, à cause de calculs politiques qui se sont tous trouvés plus faux les uns que les autres.

On peut donc écrire n'importe quelle langue n'importe comment. En fait, le créateur d'une graphie pourrait décider d'utiliser des symboles monétaires pour traduire

certaines sons – les locuteurs d'un pays d'Asie centrale ont d'ailleurs décidé de recourir au symbole « £ » pour régler un différend de ce type –, on peut former les lettres à partir de combinaisons de chiffres, etc. Le texte que j'écris en ce moment s'écrit dans la mémoire de l'ordinateur avec deux chiffres, 1 et 0. La graphisation d'une langue est toujours une succession de choix arbitraires.

En revanche, tenir compte, quand on choisit une graphie, des structures paradigmatiques, lexicales ou morphosyntaxiques de la langue est beaucoup plus difficile. Les problèmes ne surgissent généralement que beaucoup plus tard.

Prenons le cas du français. Aucune langue n'a une orthographe plus absurde, plus aberrante ou plus incohérente, ce qui a deux conséquences majeures.

D'abord, l'orthographe du français fonctionne comme un système de sélection sociale. Ensuite, elle complique grandement la numérisation du français, comme l'attestent ces messages électroniques indéchiffrables où les carrés, les « Y », les chapeaux pointus, etc. ont pris la place des « ç », « è », « i » ou autres lettres pourvues de signes diacritiques.

Dans l'univers numérique, les signes diacritiques du français, comme ceux de toute autre langue, sont une véritable catastrophe. Ce qui fait le malheur de notre langue fera peut-être le bonheur de plusieurs langues africaines non dotées, pour le moment, de systèmes d'écriture. En effet, les Africains pourraient tirer profit de ce vide scriptural pour créer des systèmes propres à faciliter la numérisation de leurs langues.

Avant de penser à de la formation sur l'informatisation et la numérisation des langues africaines, je crois donc qu'il faudrait reprendre une chose que l'on avait faite il y a 10 ans, à Bordeaux, à l'ACCT, dans un tout autre esprit : il faudrait essayer de créer ou d'aménager les systèmes des langues africaines de l'espace francophone dans l'intention première de maximiser leur compatibilité avec l'ordinateur.

L'enjeu est majeur. En effet, comme l'a écrit Jean-Claude Guédon en 1995, « on assiste à un transfert massif de toutes les cultures du monde dans les systèmes numériques, au point que dans 20 ou 30 ans au plus, tout ce qui ne sera pas codable en "0" et "1" sera complètement marginalisé au point d'être oublié comme les manuscrits médiévaux ».

Il faut donc réfléchir aux questions de graphisation, parce que seule la graphisation adéquate des langues

africaines garantira leur utilisation dans l'univers numérique et, par conséquent, la préservation de cette diversité culturelle à laquelle les gouvernements attachent tant d'importance actuellement, dans le discours du moins.

Il ne faudrait pas reproduire l'erreur de Mico, exemple intéressant à étudier parce qu'il montre bien les effets pervers des choix de graphie. Inventé par M. T. Canté, docteur en droit islamique plutôt que linguiste, le système Mico fait un retour au Mali, en Côte d'Ivoire et ailleurs.

M. Canté a fabriqué son alphabet en s'inspirant à la fois de l'alphabet latin et de l'alphabet arabe. Son système était également original en ce qu'il prescrivait que l'on écrive dans le même sens que l'arabe, c'est-à-dire de droite à gauche. De plus, M. Canté avait fabriqué son propre système de chiffres parce que les chiffres arabes ne lui suffisaient pas.

Tout cela est fort ingénieux et amusant, mais fort peu commode. Comme vous le voyez, il est possible d'être fantaisiste quand on crée un système d'écriture, mais cette fantaisie a un énorme prix. Elle peut compliquer ou même empêcher la numérisation d'une langue, et la condamner à l'avance sur la scène internationale.

Certains pourront dire que la simplicité aussi a un prix et qu'elle rend difficile la reproduction de la phonétique d'une langue. Je crois qu'ils se trompent totalement. En effet, ce n'est pas parce qu'un francophone du Sud de la France prononce le mot « grammaire » comme « grand-mère » qu'il l'écrit avec trois « m ». Les Toulousains prononcent le mot « foin » « fouan », mais ils l'écrivent comme les autres Français avec « i » et non pas avec « a », etc. La graphie d'une langue est sa représentation abstraite, non pas la reproduction fidèle de sa prononciation, qui n'est d'ailleurs jamais unique pour une langue donnée, mais présente toujours une variation dialectale.

En conclusion, je suggérerais que nous commençons à percevoir le retard des langues africaines sur la Toile comme une chance et une opportunité plutôt qu'un manque et une lacune – une occasion extraordinaire de favoriser leur numérisation en les pourvoyant de graphies très simples. Si l'on a comme objectif de favoriser la numérisation et l'informatisation des langues africaines, il s'agit là d'une étape obligée.

*Robert Chaudenson,
directeur, Institut de la Francophonie, France.*

C'est à la technique de s'adapter à l'être humain, pas le contraire

La technique de la numérisation, les technologies informatiques ont été inventées par l'homme et peuvent évoluer selon les besoins de l'homme. Est-ce que nous devons nous borner à nous adapter à une technique datant d'une autre époque ou est-ce que nous devons chercher à adapter la technique à nos besoins? Les Coréens, entre autres, n'ont pas hésité à choisir la deuxième option.

L'écriture coréenne a été inventée au XV^e siècle et sa prise en compte en informatique était beaucoup plus compliquée au départ que celle du français. Pourtant les Coréens n'ont jamais songé à remplacer le Hangul par l'alphabet latin. En fait, ils se sont tellement battus pour une solution favorisant l'utilisation de leurs caractères et de leur graphie, que le consortium Unicode, que Microsoft, que l'ISO ont dû accepter de faire des changements dans les normes établies. Les Coréens ne voulaient pas renoncer à satisfaire leurs besoins culturels et linguistiques seulement parce que les techniques de codage du moment ne jouaient pas en leur faveur. Les pays du Sud n'ont pas à le faire davantage.

Intervention de Jian Yang, conseiller, Office québécois de la langue française.



Atelier sur l'industrie

Les enjeux industriels dans le secteur du TIL : l'expérience d'un *spin-off* universitaire

O n m'a demandé de parler de mon *vécu* dans le domaine du TIL et de certains des problèmes que j'ai pu rencontrer. Pour commencer, permettez-moi de présenter l'entreprise pour laquelle je travaille : Nomino Technologies. Nomino, PME montréalaise d'une quinzaine d'employés, est née à la fin des années 1990. La compagnie est un *spin-off* du LICEF, centre de recherche universitaire. Elle met au point des applications de TIL utilisables dans le domaine du commerce électronique, plus particulièrement du service à la clientèle.

Le sort de Nomino est enviable à plusieurs égards. Ainsi, la société a des brevets en instance, elle a reçu l'appui de financiers du capital de risque, elle profite de crédits d'impôt avantageux⁴⁵ et certains de ses projets sont réalisés avec le soutien de partenaires des secteurs public et parapublic. Cependant, tout n'est pas rose.

La première chose que nous avons apprise en tant que compagnie, c'est qu'il ne faut jamais confondre prototype et produit fini. Je me rappelle notre excitation peu avant la création de Nomino, alors que nous avions déjà en main un prototype comprenant toutes les fonctionnalités attendues d'un moteur de recherche et répondant aux attentes de notre client *bêta*. Nous étions alors bien naïfs. En fait, à sa naissance, Nomino possédait une technologie prometteuse. Pour créer réellement un produit, il a fallu se préoccuper d'une série de questions dont les centres de recherche ne devraient pas avoir à se préoccuper. En effet, le marché s'attend à un grand nombre de choses d'un produit : il doit pouvoir fonctionner dans des environnements informatiques multiples, il doit supporter une montée en charge importante, il doit être très robuste de manière à ce que le client ne soit pas lésé – et de manière à ce que le producteur n'ait pas à dépenser une fortune pour assurer le service client après-vente –, il doit être convivial aux yeux de différentes clientèles, il doit fonctionner sur plusieurs plates-formes informatiques, etc. Les laboratoires de recherche ne sont pas conçus pour assurer ces caractéristiques.

Mais il y a plus. Même quand on a un produit, on n'a pas tout. Un produit peut rester sur les tablettes longtemps s'il n'a pas l'*emballage* qu'il faut, s'il n'est pas vendable, si l'on ne se préoccupe que du développement. Quand on a un produit, il faut encore faire sa mise en marché et de la représentation. tre en entreprise, c'est accepter que la recherche, ce n'est pas si important : le marketing, la finance, la vente et l'entretien comptent autant ou davantage.

Il aurait été important pour nous de bien assimiler cela dès le départ. Malheureusement, personne n'était là pour nous fournir le mode d'emploi. Le livre *Spin-off pour les nuls* n'existait pas. Quelques organismes appuient bien les jeunes entrepreneurs, mais moins sur le plan technologique que sur les plans financiers et proprement commerciaux. Nous avons reçu du capital de risque, mais pas de mentorat pour nous aider à franchir le pas entre le laboratoire et l'entreprise. D'ailleurs, les financiers et les industriels comprenaient-ils bien ce que nous ne savions pas ? Ils avaient certainement des préjugés concernant les chercheurs, mais nous tenions tous, sans le savoir, un dialogue de sourds. Le secret pour mieux sélectionner, préparer et soutenir les équipes de recherche qui démarrent des compagnies me semble encore à trouver et cela prend du temps.

Cela m'amène à un deuxième point : beaucoup sous-estiment le temps nécessaire à la création d'une entreprise de logiciel viable. Tous les secteurs de l'industrie ne souffrent pas également de ce problème. Dans le domaine pharmaceutique, par exemple, tout le monde – y compris le capital de risque – sait qu'il existe un fossé ou plutôt un cheminement balisé entre une découverte et la création d'un médicament. Il faudra des tests, des essais cliniques, des bilans, de nouveaux tests, etc. Éventuellement, avec de la chance, un remède sera mis en marché et réellement commercialisé à terme. Pour les logiciels de traitement de la langue, la situation est assez semblable. Créer un outil de TIL intelligent, un logiciel de traitement de la voix ou de l'écrit, demande des moyens et du temps.

Malheureusement, les bailleurs de fonds confondent souvent tout ce qui touche à Internet, par exemple : un moteur de recherche, un système de questions/réponses, un site Web. Dans le secteur du traitement de l'information, ils s'attendent généralement à ce que leur investissement rapporte très rapidement : dans un court laps de temps,

45. L'obtention de ces crédits est encore moins évidente que la recherche de subventions dans le monde universitaire. C'est pourquoi de nombreuses PME y renoncent. Il serait sans doute souhaitable de simplifier les procédures administratives en la matière.

l'entreprise devra avoir connu le succès avec ses produits, avoir fusionné ou avoir trouvé un acheteur. Cependant, les sommes investies sont souvent insuffisantes pour, à la fois, compléter le cycle de développement d'un produit et le commercialiser. Quand on réussit à obtenir des capitaux de risque pour la première fois, on ne se rend pas toujours compte que l'on ne va pas loin avec quelques millions de dollars ou d'euros dans le domaine du génie linguistique. Encore une fois, transformer une technologie en produit coûte cher. Et percer le marché requiert des ressources encore plus considérables.

Par ailleurs, les attentes des financiers sont importantes. Au Québec, il faut nécessairement s'engager à sortir des frontières du marché local et viser le marché américain – la perspective de ventes d'un million de dollars par an n'excitant guère le capital de risque. Mais il faut déboursier une fortune pour prendre pied aux États-Unis. Il ne suffit donc pas de passer à travers une première ronde de financement. Il faut se rendre à la deuxième. Mais voilà où le cercle devient vicieux. Pour recevoir des fonds de commercialisation en quantité suffisante à une sérieuse implantation, il faut déjà avoir réalisé des ventes d'importance.

Ce problème tient en partie à la nature du capital de risque au Québec et au Canada. Les financiers y risquent des fonds en bonne partie publics plutôt que privés. C'est peut-être ce qui les rend frileux ou ce qui les amène, fréquemment, à saupoudrer les sommes dont ils disposent. En comparaison, les entreprises de logiciel qui démarrent aux États-Unis avec l'appui de capital de risque privé reçoivent dès le départ des sommes allant de 20 à 50 millions de dollars, soit plusieurs fois la mise québécoise.

La différence ne s'arrête pas là. Aux États-Unis encore, les ministères et les organismes constituent aussi une source importante de financement pour les entreprises du secteur du TIL et pour les centres de recherche. Par exemple, nos collègues américains reçoivent des sommes considérables de la *Defense Advanced Research Projects Agency* (DARPA) pour mener des travaux de développement de systèmes de questions/réponses. De notre côté, l'appui est nettement moins considérable, même s'il existe. Nous essayons de suivre, malgré tout, et je dirais que nous y arrivons assez bien, compte tenu de nos ressources.

Finalement, force est de reconnaître que certains problèmes existent en *amont* de l'industrie. D'une part, il ne

se fait pas assez de recherche fondamentale dans plusieurs domaines de la linguistique, comme la sémantique ou la pragmatique. Cela tient en grande part au sous-financement des activités de recherche dans le secteur universitaire, en particulier en recherche fondamentale. Ce manque de fonds est dramatique, parce qu'il provoque souvent la désagrégation de bonnes équipes montées lentement, avec soin. Un des miroirs aux alouettes, pour ceux qui sont plus près de la recherche-développement, consiste en la création de *spin-offs*, comme nous l'avons fait. Cette avenue est séduisante du point de vue des universitaires, elle paraît aussi intéressante du point de vue des gouvernements. En fait, pour eux, la solution est magique: les *spin-offs* leur évitent de s'impliquer dans les fonds de recherche, puisque les gens se débrouillent tout simplement par eux-mêmes. Cependant, nous l'avons vu, cette solution ne résout pas facilement le problème économique et elle rend encore plus périlleuse, bien souvent, la situation du secteur de la recherche dans un univers qui ne peut survivre que de la vente des produits.

Un autre problème gangrène la recherche dans le domaine du TIL, c'est le manque de matériaux génériques disponibles. Des lacunes majeures existent actuellement, que l'on parle de ressources linguistiques comme des bases de données terminologiques, des banques de textes écrits ou des banques de textes traduits, ou encore des ressources informatiques comme des banques d'objets logiciels. Il existe actuellement, au Canada à tout le moins, du côté du téléapprentissage et du commerce électronique, des projets forts intéressants de mise en place de banques d'objets qui incluront des outils de base que l'ensemble des chercheurs et des industriels pourront utiliser – ce qui leur évitera de réinventer la roue à chaque fois. Ces outils seront accessibles selon des modalités variables: dans certains cas, il s'agira de gratuits; dans d'autres cas, les utilisateurs paieront des redevances aux producteurs. On devrait penser à créer un modèle semblable dans le domaine du TIL.

D'autres problèmes sont encore cause de retard, pour les PME en particulier qui ne peuvent que difficilement, faute de moyens suffisants, suivre ce qui se passe dans les organismes de normalisation et de standardisation. On pourrait aussi gloser sur les faibles ressources, hors des États-Unis, en matière d'évaluation, mais le problème le plus criant reste encore celui de la formation: formation première et formation continue.

Il est souvent difficile de trouver des étudiants formés dans des champs de la linguistique où les besoins sont très pressants, tels que la sémantique ou, mieux encore, la pragmatique. Dans le petit monde universitaire franco-montréalais, par exemple, un seul professeur enseigne actuellement cette discipline. Et plutôt que de tourner entre les universités pour maximiser le nombre d'étudiants formés, cette personne ne peut offrir son cours qu'occasionnellement et au sein de sa seule institution.

Il est encore plus difficile d'embaucher des gens qui possèdent à la fois une formation en linguistique, ou dans un domaine proche des sciences cognitives, et une formation en informatique. On aimerait bien que les gens qui arrivent chez nous maîtrisent chaque aspect de la question, qu'il existe davantage de ces bêtes curieuses formées dans les deux domaines. Malheureusement, les programmes multidisciplinaires sont encore rares ou inexistantes. Notons cependant qu'à Montréal les choses s'améliorent puisqu'un doctorat en informatique cognitive est désormais offert par l'UQAM et la TÉLUQ. Cela a demandé sept ans d'efforts à une large communauté multidisciplinaire de professeurs.

Finalement, il est évidemment ardu de trouver la perle rare en mesure de se faire valoir *dès son entrée* dans l'entreprise. Une solution susceptible de régler ce problème serait la création de programmes coopératifs du genre de ceux qu'offre depuis déjà 40 ans l'Université de Sherbrooke, avec succès et dans divers domaines. Dans des secteurs comme le génie ou l'administration des affaires, il est ainsi possible aux étudiants de recevoir une excellente formation universitaire, tout en acquérant de bonnes expériences de travail. Les étudiants font en effet des allers-retours entre l'université et le monde du travail, sans que leur formation académique en soit le moindrement entachée.

Mon dernier point concerne la formation continue. Il n'est pas toujours facile, dans les PME, d'accorder à cette question toute l'attention qu'elle mérite. Quand l'économie tourne au ralenti, on scrute toute sortie d'argent et on *coupe* tout ce qui ne semble pas servir l'entreprise à court terme. Comme les dépenses de formation rapportent surtout à moyen ou à long terme, elles comptent souvent parmi les premières victimes. Pourtant, la formation continue est absolument nécessaire. Il serait fantastique que l'on puisse créer de bons produits sans recherche, sans formation, mais ce n'est pas le cas. Par conséquent, *forcer* les entreprises à consacrer une partie de leur budget à la formation continue, comme on le faisait au Québec jusqu'à récemment⁴⁶,

constitue une saine mesure de sauvegarde non seulement des cerveaux, mais aussi de l'économie.

*Claude Rigault,
vice-présidente, Recherche et développement, Nomino, Montréal,
Québec.*

Comment vendre quelque chose que le client estime devoir obtenir gratis ?

Qu'une large part d'un secteur industriel ait du mal à vendre ses produits, alors que le besoin sociétal est bien présent, peut paraître étonnant. Malheureusement, dans le secteur du traitement de l'information textuelle, nous ne souffrons pas seulement du ralentissement économique, nous devons également résoudre un deuxième problème : comment faire pour que les acheteurs éventuels ne perçoivent pas nos outils comme de simples commodités, c'est-à-dire comme des produits à faible valeur ajoutée qui devraient presque être donnés plutôt que vendus !

Un progiciel comme SAP ne sera jamais perçu comme une commodité. Mais les logiciels que nous fabriquons – pensons aux traducteurs assistés, aux outils de correction, aux moteurs de recherche, aux applications de reconnaissance de la voix ou aux systèmes de gestion automatique de dialogue questions/réponses – le sont bien souvent, malgré toute leur complexité. C'est qu'ils ne sont pas perçus comme tels et, aux yeux de nombreux clients, il ne devrait pas être nécessaire de les acheter. Ces instruments devraient venir gratuitement avec des logiciels qui apparaissent aux yeux du public comme beaucoup plus complexes (par exemple, la suite Microsoft Office) et qui, eux, seraient payés au prix fort. Ne retrouve-t-on pas toujours, attaché au TIL, le syndrome de M. Jourdain ? La langue parlée par tous a tendance à être banalisée et, avec elle, tout ce qui la concerne.

Intervention de Claude Rigault, vice-présidente RD, Nomino, lors de l'atelier *Industrie*.

46. La règle dite « du 1 % » forçait en effet les entreprises à consacrer au moins 1 % de leur masse salariale à des activités de formation ; elle a été abrogée partiellement par le gouvernement québécois élu au printemps 2003.

Il faut sortir des sentiers battus

Quels sont les obstacles à l'essor de l'industrie francophone du TIL? Certains, pendant cet atelier, ont déploré que les entreprises étrangères recourent parfois au dumping pour accroître leur part de marché dans le champ du TIL. Malheureusement, que voulez-vous, c'est comme ça que la partie se joue dans l'univers capitaliste.

Certains ont de leur côté déploré la quantité de paperasse gouvernementale qu'il faut remplir pour démarrer une entreprise. Cependant, quand j'entends dire, comme hier, que ceux qui se mettent à la tâche reçoivent en crédits d'impôt des sommes égales à 80 % de leurs dépenses de RD, je me dis que le jeu en vaut vraiment la chandelle.

Finalement, il est sûr, comme on l'a dit plus tôt, que la générosité des capitalistes-risqueurs a grandement diminué récemment, mais c'est un peu normal. Il y a deux ou trois ans, on pouvait trouver, dans n'importe quel secteur du logiciel, 25 entrepreneurs convaincus, dans leur plan d'affaires, de pouvoir prendre chacun 25 % du marché. Il était clair que les mathématiques les rattraperaient un jour.

De mon côté, je vois deux problèmes réels et deux solutions. D'abord, l'offre, dans le secteur francophone du TIL, est extrêmement fragmentée. Et les organisations sont petites. Dans ce contexte, la collaboration entre entreprises, entre centres de recherche ou entre entreprises et centres de recherche n'est pas un luxe : c'est un impératif.

Quand j'entends Claude Rigault dire qu'elle a passé un été à s'occuper de propriété intellectuelle plutôt que de RD, je me dis que quelque chose cloche sérieusement. Quand je vois qu'il me faudrait deux ans et des ressources que je n'ai pas pour former un agent

de valorisation, c'est-à-dire quelqu'un capable de faire en sorte que mes recherches trouvent des débouchés dans le monde de l'entreprise, je me dis que ça ne va pas.

La solution, dans ce contexte, c'est de mettre nos ressources en commun. Je ne peux m'offrir un agent de valorisation, mais ensemble, les centres de recherche québécois en TIL pourraient s'en payer un. Claude Rigault ne peut pas se payer un expert en propriété intellectuelle, mais toutes ensemble, les entreprises québécoises pourraient s'en offrir un.

Dans cette optique, je vois d'un bon œil la création de centres spécialisés en valorisation comme Valorisation-Recherche Québec ou celle d'organismes dédiés à la recherche en génie linguistique comme le Centre de recherche en technologies langagières. Ce dernier pourrait par exemple s'occuper de développer les matériaux linguistiques génériques dont tous ont besoin et dont tous déplorent l'absence depuis hier.

Le deuxième problème que je vois, c'est que nous ne faisons pas toujours assez attention à la concurrence. Trop souvent, les chercheurs universitaires et leurs étudiants se lancent dans un champ de recherche sans avoir vérifié préalablement s'il y existe déjà des brevets. Trop souvent les entreprises se lancent dans la mise au point d'un produit sans avoir tenté de savoir si ce dernier existe déjà. Après, il ne faut pas se surprendre que les capitalistes-risqueurs soient peu enthousiastes. En Francophonie, nous sommes condamnés à trouver des niches inoccupées.

En conclusion, il nous faudra travailler de concert, les uns avec les autres, et éviter les sentiers battus. C'est à ces conditions que le secteur francophone du TIL pourra progresser.

Intervention de Pierre Dumouchel, vice-président RD, CRIM, lors de l'atelier *Industrie*.

Le secteur du traitement informatique des langues : un marché frileux

Permettez-moi de vous présenter ma vision de ce qui se passe dans le secteur du TIL – la vision d'un entrepreneur sur le terrain. Il y a quelques années, j'ai reçu le mandat de créer la filiale nord-américaine de la société européenne WizArt.

WizArt se spécialise dans la production d'outils de localisation. Ces outils permettent aux organisations québécoises, canadiennes, américaines ou autres de traduire les applications Windows, Web ou Java qu'elles utilisent – dans le jargon du métier, de les *localiser*.

Du point de vue d'un producteur de logiciels, la première question qui se pose actuellement est : Est-ce que les acheteurs sont prêts à accueillir de nouvelles technologies ? Ce n'est pas sûr. Après tout, la *bulle* technologique a explosé et l'économie mondiale fonctionne de plus en plus au ralenti. Par exemple, l'Allemagne est officiellement en récession, tandis que la France devrait en vivre une d'ici six mois.

Il y a quelques années, WizArt est passée à deux signatures de se retrouver cotée en bourse.

Dieu merci, notre opération Nasdaq a échoué !

La compagnie serait morte et enterrée aujourd'hui !

Le pendant de ce ralentissement, c'est l'augmentation de la frilosité du capital de risque. Il y a quelques années, il était extrêmement facile à n'importe quelle entreprise d'obtenir du financement. Le paradigme de la boîte dorée, *mais vide*, prévalait. Lors d'une première ronde, une compagnie pouvait obtenir 20 millions de dollars US sur simple présentation d'une idée ou d'un plan d'affaires. Lors d'une deuxième ronde, armée seulement d'une maquette, la compagnie pouvait empocher 15 millions de dollars de plus.

La situation a changé radicalement. Le capital de risque est aujourd'hui inexistant en Allemagne et très difficile à trouver en France. En Amérique du Nord, les capitalistes-risqueurs sont devenus beaucoup plus prudents. Ils ont renoncé, dans de nombreux cas, à investir dans de nouvelles compagnies, pour plutôt injecter leurs fonds dans des entreprises déjà dotées d'un marché. De plus, ils souhaitent aujourd'hui multiplier leur mise par 20 ou 30 à l'intérieur de deux années, plutôt que de cinq comme par le passé.

D'un côté, les PME de mon secteur ont de la difficulté à obtenir les fonds dont elles ont besoin ; d'un autre, elles

souffrent de la concurrence de plus en plus forte des grandes sociétés de traduction et de localisation. En effet, ces dernières n'hésitent pas à arriver avec des soumissions déloyalement basses – à faire du *dumping* – pour accroître leurs parts de marché au détriment des petits joueurs.

Cela m'amène à faire trois constats. Premièrement, à court ou moyen terme, il sera dur à de nouvelles PME et à de nouveaux projets de démarrer avec succès dans le domaine du TIL. Même les entreprises bien établies connaîtront des jours difficiles. Par exemple, WizArt aura de la difficulté à progresser si elle ne trouve pas de financement dans les 24 mois qui viennent ; elle devra possiblement se contenter de vivre.

Deuxièmement, il serait important de chercher à augmenter notre masse critique dans le domaine et à y conclure des alliances de tous genres, ou celui-ci pourrait bien nous échapper. Par exemple, à la suite d'une longue série d'acquisitions, trois compagnies non francophones, SDL International, Bowne et Lionbridge, se divisent actuellement 80 % du marché mondial de la localisation (éventuellement, ces entreprises pourraient elles-mêmes être achetées par des géants qui, comme Microsoft, Oracle ou IBM, pourraient être intéressés à offrir l'ensemble des services et des technologies de localisation requis par les entreprises). En restant chacune dans leur coin, les entreprises francophones pourront-elles vraiment tirer leur épingle du jeu ?

Troisièmement, il faudrait faire des efforts pour éveiller les organisations à l'importance de tenir compte de la mondialisation des marchés et de localiser leurs applications. Il est certain que les mentalités évoluent peu à peu à cet égard. Ainsi, certaines organisations ont commencé à se doter d'un *Chief Global Officer* ou *CGO* – en français, d'un directeur de la mondialisation. Par le passé, cette question tombait souvent sous la responsabilité des gens de marketing qui ne lui accordaient qu'une partie de leur attention. De nos jours, les compagnies se rendent compte que la mondialisation mérite d'être traitée autrement que comme une arrière-pensée, qu'elle doit être perçue comme un enjeu stratégique. Il conviendrait de veiller à ce que cette tendance se poursuive, au sein des grandes entreprises comme des petites.

Quatrièmement, je pense que les producteurs d'outils faisant le traitement informatique des langues devront recourir à de nouvelles méthodes de commercialisation. Par

exemple, nous avons commencé à vendre nos produits différemment, les clients potentiels de WizArt ayant beaucoup moins d'argent à dépenser qu'ils n'en avaient il y a deux ou trois ans. En fait, nous les leur prêtons maintenant pendant quelques semaines, de manière à ce qu'ils traduisent quatre, cinq ou dix éléments essentiels de leurs logiciels ou progiciels. Ceci leur permet de faire des tests de marché avec leurs propres clients et partenaires. Avant d'acheter notre technologie, les entreprises peuvent maintenant voir concrètement si son utilisation leur rapportera!

*Jean-Patrice Isaac,
directeur du bureau Canada-États-Unis, WizArt, Westmount,
Canada.*

Attentes déçues

La faiblesse de la demande a certes pour cause le ralentissement économique, mais elle tient aussi, pense John Chandioux, président de John Chandioux experts-conseils, au fait que la clientèle a souvent été échaudée par le passé! «L'industrie du TIL est partiellement responsable de son sort, souligne M. Chandioux, parce que plusieurs entreprises y sont allées, au fil des ans, de fausses promesses. Par exemple, certains producteurs de correcteurs affirmaient, dans les années 1990, que leurs outils permettaient d'améliorer le français écrit des utilisateurs, alors que l'on ne peut pas utiliser ces applications quand on ne maîtrise pas la langue au départ. Il faut avoir un sens critique développé pour évaluer les suggestions faites par la machine!»

«De même, continue M. Chandioux, les producteurs de traducteurs automatiques grand public n'ont jamais réussi à tenir leurs promesses.

Je pense par exemple à Global Translator, l'inventeur du jetticiel, [...] logiciel pitoyable, mais suffisamment bon marché pour que tout le monde l'essaye et que personne ne s'en serve. Le PDG de l'entreprise disait lui-même: 'Pourquoi me priverais-je de vendre ce logiciel même s'il ne vaut rien si je peux faire 10 millions de dollars par an de chiffre d'affaires avec? »

Traitement informatique des langues et propriété intellectuelle

Un des problèmes qui se posent dans l'industrie du TIL, comme dans d'autres industries logicielles, c'est celui du piratage. En effet, de nombreux clients n'achètent pas les applications qu'ils utilisent. À la place, ils les copient.

Comme le relève John Chandioix, président de John Chandioix experts-conseils, le problème tient en bonne partie au fait que la culture Internet ambiante encourage le non-respect de la propriété intellectuelle. « On le voit notamment dans le domaine de l'enseignement. Certains professeurs justifient le piratage de logiciels en disant qu'ils n'ont pas les budgets requis pour les acheter. Ils ne pensent pas un instant qu'ils sont censés donner l'exemple aux jeunes! »

« Par exemple, continue M. Chandioix, on me racontait récemment l'histoire d'une enseignante rencontrant une collègue: 'Connais-tu le logiciel X?', demande la première. La deuxième répond affirmativement. 'Je pensais d'ailleurs l'acheter', ajoute-t-elle. 'Te fatigue pas, rétorque l'autre, j'en ai fait une centaine de copies en fin de semaine!' Vous vous imaginez ce que cela fera au petit éditeur québécois qui éditait ce logiciel? Il aura un manque à gagner énorme ».

La même situation prévaut dans le domaine de la traduction, soutient John Chandioix. « Un client m'a récemment demandé de sillonner le Canada pour visiter des traducteurs pigistes. Croyez-moi! Je suis resté sidéré de voir le nombre de personnes recourant à des copies piratées du *Grand Robert*, de *Windows*, etc. En fait, 50% des ordinateurs que j'ai examinés comptaient une majorité de produits piratés! »

Les brevets et les autres méthodes de protection de la propriété intellectuelle (ex.: les droits d'auteur) constituent des solutions à ce problème. Cependant, comme le rappelle Sylvie Wallez, responsable de l'OTIL, en citant l'Organisation mondiale de la propriété intellectuelle (OMPI), « une information insuffisante sur l'importance de la propriété intellectuelle dans la conduite quotidienne des affaires, les coûts élevés associés à l'obtention et à la sanction des droits de propriété intellectuelle, l'idée selon laquelle le système de propriété intellectuelle est ésotérique, trop contraignant et exige trop de temps figurent parmi les raisons pour lesquelles de nombreuses PME montrent parfois peu d'empressement pour protéger leurs actifs de propriété intellectuelle ».

L'OMPI, comme le rappelle Sylvie Wallez, est une organisation intergouvernementale et internationale qui se consacre à la promotion, à la création, à l'utilisation et à la protection de la propriété intellectuelle. L'OMPI fait profiter les États qui en sont membres de son expérience et de sa compétence, de manière à leur permettre de mieux tirer parti du système de propriété intellectuelle mondial. Pour l'OMPI, l'utilisation efficace, par les PME, des avoirs de propriété intellectuelle est un facteur essentiel du développement économique.

Il est indéniable, comme le souligne l'OMPI, que beaucoup d'organisations n'ont pas les ressources financières nécessaires à l'obtention de droits de propriété intellectuelle. Il faut avoir les reins solides pour se lancer dans l'aventure, les règles en la matière variant souvent d'un pays à l'autre. Par exemple, Jean-Patrice Isaac, responsable de la filiale nord-américaine de WizArt, souligne qu'« il en aurait coûté entre un million et trois millions de dollars à son entreprise pour obtenir un brevet mondial. Il est clair qu'une telle somme est hors de notre portée ». Même un organisme spécialisé en recherche comme le CNRC, qui dispose d'une équipe d'une dizaine de personnes dans le champ de la propriété intellectuelle, devrait en compter le double, pense André Manseau.

Il est aussi vrai que, pour veiller à la protection de leurs droits de propriété intellectuelle, les organisations doivent négliger des activités plus centrales à leur essor. Par exemple, Claude Rigault, responsable du développement de produits chez Nomino, souligne qu'elle a dû passer un été complet à discuter droits d'auteur avec des avocats au lieu de faire de la RD. Comme Nomino allait commercialiser des produits mis au point dans les laboratoires du Licef, centre de recherche universitaire, il fallait déterminer quels droits revenaient aux chercheurs, à l'institution universitaire et à l'entreprise.

Cela dit, quelques solutions pourraient permettre de résoudre ces problèmes partiellement. Pour Claude Rigault, l'Association en industries de la langue au Canada (AILIA) et le CNRC⁴⁷, par le truchement du nouveau Centre de recherche en technologies langagières de Gatineau,

47. Selon André Manseau, le CNRC possède déjà une équipe dédiée à la protection de ses actifs intellectuels.

pourraient offrir des services d'aide à la protection de la propriété intellectuelle des entreprises en démarrage.

Éventuellement, l'harmonisation des règles nationales relatives à la propriété intellectuelle pourrait simplifier la vie des organisations francophones. Certains efforts récents ont été faits dans ce domaine, relève Sylvie Wallez. « Ainsi, la Commission européenne a récemment proposé la mise en place d'une directive qui ferait en sorte que les logiciels créés dans un pays donné seraient protégés, sur le plan intellectuel, dans l'ensemble de l'Europe. Pour nous Européens, c'est une première étape. Éventuellement, il faudrait en arriver à une brevetabilité internationale ».

Stéphane Chaudiron n'est cependant pas d'accord avec cette idée. Selon le responsable français, « dans le contexte international comme dans celui de la Francophonie, breveter les outils de TIL et leurs composantes serait une grave erreur. Cela, en effet, équivaldrait à limiter la possibilité pour les pays du Sud d'entrer dans la joute et de se doter de logiciels de traitement des langues ».

Une petite histoire d'horreur

On ne peut pas parler d'innovation sans parler de propriété intellectuelle. Les deux sont indissociables. Malheureusement, la propriété intellectuelle d'une entreprise est coûteuse à protéger. Comme la plupart des PME n'ont pas les moyens de le faire, il faudrait que les pouvoirs publics les y aident. Je crois que l'anecdote suivante vous convaincra de cela.

Comme vous le savez peut-être, je m'appelle John Chandioux. Cela fait 25 ans que mon entreprise, John Chandioux experts-conseils, fait la pluie et le beau temps au Canada. Je suis en effet le propriétaire du système de traduction automatique METEO, logiciel de TIL utilisé par le gouvernement du Canada pour

traduire l'ensemble des bulletins météorologiques produits au pays.

Il y a deux ans, METEO a été copié par un employé du gouvernement canadien. Cet employé s'est vu accorder le contrat de traduction automatique des bulletins météorologiques après avoir présenté une soumission plus basse que la nôtre. Pour faire valoir nos droits, nous avons présenté notre cas à la Gendarmerie royale du Canada (GRC). La GRC a étudié le dossier et l'a trouvé suffisamment documenté et persuasif pour déclencher une enquête criminelle.

Nous venons d'apprendre que le Procureur général du Canada a décidé de mettre fin à l'enquête, sous prétexte que ce type de problème ne relève pas de la GRC. On nous a plutôt conseillé de poursuivre les parties concernées au civil. Cependant, le copieur de notre technologie n'a pas d'actifs – il a créé sa compagnie au simple fait d'emporter le contrat fédéral –: en le poursuivant, nous pourrions avoir la satisfaction de gagner, mais nous ne pourrions pas récupérer les sommes perdues. Nous pourrions aussi poursuivre le client, c'est-à-dire le gouvernement, mais notre lutte ressemblerait à celle opposant le pot de terre au pot de fer. En effet, faute de moyens, nous ne pourrions pas, comme lui, nous doter d'une batterie d'avocats.

La seule solution que nous avons alors, c'est celle que j'utilise aujourd'hui: j'amène le dossier sur la place publique – ce que j'ai évité de faire ces deux dernières années –, en espérant que je finirai par être entendu. Qu'on se le dise: il n'y a pas d'innovation sans propriété intellectuelle. Il faudrait peut-être que le gouvernement donne l'exemple.

Compte-rendu d'une discussion tenue pendant la table ronde *Industrie*.



Atelier sur les voies de l'avenir

L'avancement des travaux est affaire d'intégration

Dans le domaine du traitement informatique des langues, l'avancement des travaux est affaire d'intégration.

Sur le plan scientifique, afin de faire progresser les résultats de la science, il est nécessaire d'intégrer les connaissances développées dans des disciplines diverses comme la linguistique, l'informatique, les sciences cognitives et les sciences de l'information.

Sur le plan technologique, il faut accepter l'idée que les solutions aux problèmes ne sont pas seulement linguistiques, mais qu'elles peuvent provenir d'autres approches, statistiques, probabilistes, neuronales ou connexionnistes. Pour résoudre les difficultés éprouvées dans le traitement d'une langue donnée, il importe également de recourir aux approches multilingues, en plus d'employer les approches monolingues.

Enfin, sur le plan industriel, il convient, pour créer des produits performants et conviviaux, de combiner les habiletés techniques de spécialistes de la linguistique à celles d'experts en informatique, et de spécialistes en ergonomie et dans le domaine des interfaces.

Comme on le voit, l'avancement du secteur dépendra donc de l'intégration des efforts d'acteurs appartenant à différentes écoles ou disciplines, et de la complémentarité entre des technologies souvent perçues, à tort, comme concurrentes. Dans le domaine des industries de la langue, on a longtemps considéré que la linguistique était l'unique approche possible pour résoudre les problèmes inhérents au traitement automatique des langues naturelles ; il est temps de se rendre compte que, seule, l'hybridation des technologies donnera des résultats conformes aux attentes des usagers. On voit notamment les résultats de l'intégration des connaissances et des expertises dans le champ du traitement de l'information vers lequel convergent des spécialistes issus du monde de la bibliothéconomie, de la structuration du document électronique (SGML, XML, etc.) et de l'ingénierie linguistique.

L'intégration ou, au moins, la coordination des efforts des organismes publics sera aussi indispensable. À plusieurs reprises, on a parlé, pendant ce colloque, de projets visant à recenser, à valoriser et à rendre disponibles des ressources linguistiques, des corpus, des annuaires de compétence, des listes de projets, etc. Si on ne peut qu'admirer la multiplicité des projets en cours, en même temps, il faut

regretter le fait que lorsqu'un acteur francophone désire savoir ce qui est vraiment disponible, c'est quasiment mission impossible. De plus, s'il y a foisonnement d'initiatives nationales, régionales, multilatérales ou bilatérales, il y a aussi, dans de nombreux cas, des redoublements que l'on pourrait peut-être éviter.

Cela dit, différentes idées devraient guider les travaux des chercheurs du secteur du TIL dans les années à venir :

- premièrement, les interfaces de demain seront multimodales, c'est-à-dire qu'elles comporteront à la fois des composantes de traitement du texte, de la voix, de l'image, etc. Par conséquent, on ne peut plus travailler sur chaque question individuellement, en vase clos. Il faut plutôt penser en termes de complémentarité ;
- deuxièmement, les technologies linguistiques de demain devront être auto-apprenantes. Par exemple, le système permettant à l'acheteuse d'un téléphone cellulaire de dicter un numéro vocalement devrait être capable de s'habituer à elle. Il pourrait ainsi la reconnaître en toutes circonstances, par exemple même lorsqu'elle est enrhumée ;
- troisièmement, l'heure est au multilinguisme. Il faudra par exemple développer des outils permettant d'interroger une banque de données en une langue pour repérer un document écrit en une autre langue ;
- quatrièmement, il faudra continuer de travailler au traitement sémantique des connaissances. En effet, on passe progressivement d'une informatique de *contenant* à une informatique de *contenu*. Autrement dit, on ne peut plus traiter seulement le signal, par exemple les caractères composant un mot ; on doit aussi tenir compte du sens de ce dernier ;
- cinquièmement, les travaux devraient être menés en tenant compte des usages. Trop souvent, les développements techniques réalisés dans le domaine du traitement de la parole ou de l'écrit tiennent peu compte de la manière dont les gens utiliseront les outils mis au point. Il faudrait avoir une attitude plus modeste à l'égard de la technologie, admettre que les gens ne l'utilisent pas toujours comme il a été prévu.

*Stéphane Chaudiron,
chargé de mission Information scientifique et technique - Ingénierie linguistique, ministère de la Recherche, France.*

Sciences cognitives et traitement informatique des langues

Plusieurs d'entre vous ne savent pas que la Bulgarie est un pays francophone. Elle l'est, non pas parce que la langue bulgare est d'origine latine – c'est une langue slave –, mais bien parce qu'il existe dans mon pays une tradition très forte d'apprentissage du français. Il y a des lycées bilingues en Bulgarie, des étudiants qui obtiennent des baccalauréats français et des antennes de l'AUF où l'enseignement est entièrement donné en français. L'appartenance de la Bulgarie à l'espace francophone explique que je fasse partie de cet événement.

En vous écoutant, je me suis rendu compte que les problèmes que vous avez ici n'ont rien à voir avec les nôtres. Parce que la Bulgarie est actuellement en pleine transition économique, les industries de la langue n'y sont pas du tout développées. Il y a bien, ici et là, de petites entreprises qui essayent de faire quelque chose dans le secteur, mais leur impact est marginal. Actuellement sur le plan industriel, nous sommes, comme il est connu, plus forts en matière de piratage et de production de virus qu'en génie linguistique.

Dans le champ des sciences fondamentales, la situation est bien meilleure, mais n'est pas toujours idéale non plus. Ainsi, le bulgare est une langue partenaire pour laquelle il n'y a toujours pas de corpus, on ne peut pas suivre de formation en linguistique computationnelle en Bulgarie – pas plus d'ailleurs qu'ailleurs dans notre région –, etc.

Cela dit, il existe des raisons d'être optimistes. Ainsi, un centre reconnu de recherches en sciences cognitives, le Centre de l'Europe Centrale et Orientale, se trouve au sein de l'université où je suis active. J'y suis active et j'y ai un peu travaillé sur la modélisation des langues naturelles.

Les sciences cognitives sont une discipline interdisciplinaire assez obscure. On peut en trouver autant de définitions du secteur qu'il y a par exemple, de sites sur la question, alors je vais me limiter à une des définitions possibles : l'expression *cognition* désigne « l'ensemble des formes de connaissance dont est doué l'être humain, pris individuellement ou collectivement : non seulement la connaissance abstraite (les contenus des sciences, le raisonnement sous ses multiples formes par exemple), mais aussi la perception, la mémoire, l'action, le langage, la conscience, etc. »

Dans cette définition, l'activité proprement humaine est exactement le langage – quelque chose qui nous est propre et qui est *fusionné* à ce que nous sommes. À partir de

deux ans et demi, l'être humain se distingue du chimpanzé par sa capacité à utiliser la langue.

En sciences cognitives, plusieurs chercheurs issus de différentes disciplines – il y a des psychologues, des linguistes, des mathématiciens, des physiciens, des informaticiens –, essayent de comprendre comment la langue fonctionne, pour faire en sorte qu'il soit éventuellement possible de créer des outils qui permettront à la machine de *parler*.

La question qui intéresse le plus les experts en sciences cognitives – et qui leur rend la vie difficile –, c'est celle de la sémantique, du sens. Les neuropsychologues et les médecins cliniciens ont récemment découvert que l'espace cognitif appelé à traiter des tâches liées uniquement à la langue et celui destiné à traiter les concepts sont des zones complètement différentes.

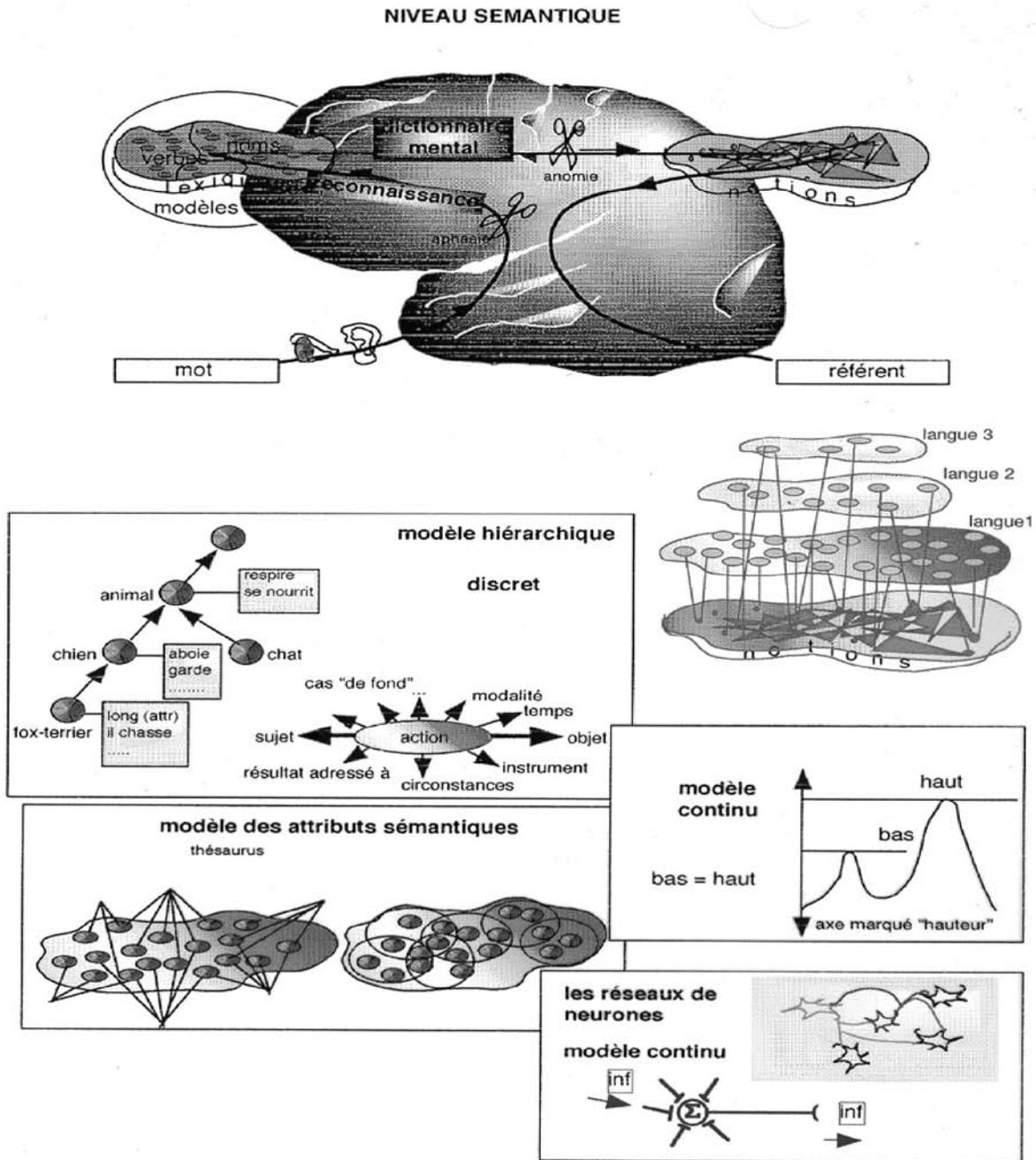
Dans le schéma 1, vous voyez que les questions sémantiques sont prises en compte à un endroit donné du cerveau – dans le néocortex, on ne sait encore pas tout à fait comment – et que les différentes langues d'une personne – sa langue maternelle, ses langues secondes – sont traitées ailleurs. Le système où le sujet se représente sémantiquement le monde est donc différent de celui qu'il utilise pour articuler des mots.

Cette division des tâches, à l'intérieur du cerveau, explique pourquoi les experts en sciences cognitives pourraient aider les linguistes informatiques à créer des outils de TIL plus performants.

Voici encore un exemple qui aide à comprendre ce que je dis. Comme vous le savez peut-être, il a été prouvé, lors de plusieurs expériences, que quand on écoute quelqu'un parler, on ne capte pas tous les phonèmes qui ont été prononcés, même quand il n'y a pas de bruit. Nos capacités auditives ne sont tout simplement pas assez bonnes pour cela. C'est grâce à l'intervention active de notre cerveau, grâce aux modèles qu'il a de la langue et du monde, que les « trous sont bouchés » et le contenu capté du message est reconstruit pour former ensuite une image interne de ce qui a été dit.

De même, les logiciels de reconnaissance de la parole éprouvent de la difficulté à fonctionner adéquatement dans des endroits bruyants, parce qu'ils ratent certains phonèmes. Dans un cas comme celui-là, on devine que les spécialistes de la cognition pourront aider les linguistes informatiques à améliorer les performances de la machine.

Schéma 1



Je crois qu'on pourrait aussi dire que les percées en sciences cognitives aideront à améliorer la performance des logiciels de traduction automatique. En effet, on ne peut pas comprendre un texte et le traduire adéquatement sans avoir une image sémantique intérieure de ce qui a été dit.

Je conclus en vous disant un mot d'un projet auquel je participe partiellement, un projet mené par la partie flamande de l'Université libre de Bruxelles (VUB) et par l'École polytechnique de Chine. Dans ce projet, on tente d'animer les lèvres d'un être artificiel en traduisant des phonèmes en visèmes (voir l'illustration 1).

Est-ce qu'on pourrait utiliser l'approche inverse pour améliorer les capacités de reconnaissance de la voix des logiciels? Je crois que oui. En effet, la mémoire visuelle et

les modèles phonétiques, chez l'humain, sont stockés et fonctionnent différemment. D'un point de vue technologique, le recours à un téléphone avec caméra vidéo pourrait donc pallier la présence de bruit dans le signal sonore – et *vice versa*.

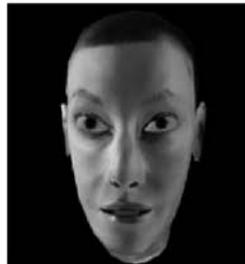
Finalement, il me paraît nécessaire pour les spécialistes de la linguistique informatique de travailler de concert, dans les années à venir, avec les cognitivistes. Les Américains dépensent déjà des sommes importantes sur la question. Il faudrait leur emboîter le pas.

*Velina Slavova,
professeure, Nouvelle Université Bulgare, Sofia, Bulgarie.*

Illustration 1
Phonèmes et visèmes



vi : /o -ou/



vj : /#/



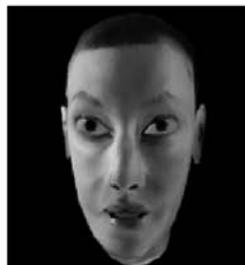
vk : /b, p, m/



vl : /f, v/



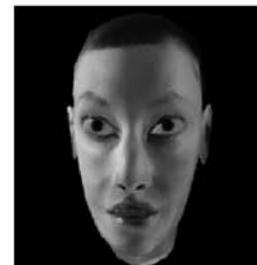
vm : /t, d, s, z, th, dh/



vn : /w, r/



vo : /ch, jh, sh, zh/



vp : /k, g, n, l, ng, h, y/

Les voies d'amélioration des outils de traitement de l'écrit

Je travaille pour un groupe de recherche appelé le RALI, pour « Recherche appliquée en linguistique informatique ». Le RALI mène des travaux dans plusieurs domaines, mais surtout dans celui de l'écrit. La particularité du laboratoire, c'est qu'il examine la question d'un angle très *ingénieur*, avec la volonté de faire des applications pratiques.

Je parlerai aujourd'hui d'approches et de voies de recherche susceptibles de favoriser l'amélioration des outils de traitement de l'écrit. Avant de commencer, il est cependant utile de parler un peu des techniques employées actuellement.

Dans certains cas, les spécialistes traitent la langue écrite en recourant à des approches symboliques, lesquelles supposent l'expression des connaissances à l'aide de règles. Les approches symboliques présentent un avantage important : elles permettent de travailler sur des bases très solides. En effet, les informaticiens n'ont pas inventé le traitement des langues ; il y a longtemps que les grammairiens ont développé une connaissance fine de celles-ci.

Les approches symboliques comportent cependant un désavantage important : il faut faire de l'adaptation manuelle. Dans bien des cas, cette stratégie est malgré tout la plus efficace qui soit. Par exemple, on sait que les bons correcteurs de fautes d'orthographe fonctionnent essentiellement sur la base d'approches symboliques.

À côté des approches symboliques, il existe des approches probabilistes. Le recours à ces dernières est fructueux quand on manque de connaissances et qu'il n'est pas requis d'articuler des règles pour permettre à l'ordinateur de faire le travail demandé. Par exemple, il n'est pas nécessaire, le plus souvent, d'utiliser une approche symbolique pour aider la machine à établir la langue d'un texte, de manière à déterminer quel correcteur ou traducteur devrait être employé pour le corriger ou le traduire. Il est plus simple de fournir une masse de textes à l'ordinateur et de lui dire : « Voici de l'anglais », « Voici du français » ou « Voici du chinois ».

Dans les années à venir, il faudra perfectionner ces approches pour améliorer les capacités de la machine en matière d'extraction, de synthèse et de traduction de l'information.

Extraction de l'information

On dispose actuellement de bons outils pour retracer l'information dans des bases de données. Cependant, l'information requise *ne se trouve pas toujours* dans les bases de données. Elle réside souvent dans des textes ordinaires. Par exemple, dans les hôpitaux, elle se trouve dans les rapports des médecins ; dans le domaine juridique, elle réside dans les jugements des juges ; etc.

Quand l'information se trouve dans des textes normaux, on est encore relativement mal équipés pour la repérer. Actuellement, ce que l'on a sous la main, ce sont essentiellement des outils de recherche par mot-clé. Ces outils sont performants. Cependant, s'ils permettent fréquemment d'aller chercher le bon document, ils ne permettent pas de dépasser le niveau de la perception et d'en comprendre le sens.

Le RALI travaille actuellement aux questions d'extraction d'information. Plus particulièrement, il s'intéresse de près à l'analyse des renseignements contenus dans les courriels que les clients d'une entreprise envoient à cette dernière. Cela pourra permettre d'y percevoir des tendances.

« L'extraction d'information consiste à identifier de l'information bien précise d'un texte en langue naturelle et à la représenter sous forme structurée. Par exemple, à partir d'un rapport sur un accident automobile, un système d'extraction d'information sera capable d'identifier la date et le lieu de l'accident, le type d'incident ainsi que les victimes. Ces informations pourront ensuite être stockées dans une base de données pour y effectuer des recherches ultérieures ou être utilisées comme base à la génération automatique de résumés ».

Source : Page Web du RALI

Résumés

La production automatique de résumés présente un grand intérêt commercial. En effet, comme nous sommes submergés d'information, il nous est difficile de faire un tri rapide entre les textes intéressants qui nous tombent sous la

main et les autres. Les résumés peuvent aider les lecteurs à vite se faire une idée du propos d'un document et, par conséquent, à gagner un temps précieux.

Les chercheurs travaillent beaucoup, actuellement, sur les questions de synthèse d'articles de journaux. Il s'agit d'un domaine d'application où le recours aux technologies donne de bons résultats, parce que la machine reçoit un bon coup de main du rédacteur. En effet, les journalistes sont entraînés à résumer leur article dans son premier paragraphe...

Cela dit, les chercheurs commencent à s'intéresser à des domaines beaucoup plus spécialisés et difficiles que celui de l'écrit journalistique. Par exemple, ils travaillent actuellement à produire des résumés automatiques d'articles scientifiques. Cette question présente un intérêt certain, même si les rédacteurs de ce genre de documents en font déjà un résumé. En effet, l'auteur est la pire personne disponible pour faire un bon résumé. Le lecteur d'un texte, en effet, y cherche généralement quelque chose de différent de ce que le rédacteur voulait y mettre. Par ailleurs, la demande pour des résumés de textes juridiques est aussi élevée.

Les chercheurs ont aussi commencé à mettre au point des résumés capables de s'attaquer automatiquement à *plusieurs* documents pour en faire la synthèse. Par exemple, un outil de ce type pourrait appréhender plusieurs textes journalistiques publiés sur une même question et produire automatiquement une chronologie.

La génération de manchettes journalistiques est une question qui intéresse également les chercheurs. Ces derniers tentent de produire l'en-tête d'un article à partir de son contenu.

Traduction

La question de la traduction automatique a donné naissance au domaine de la linguistique informatique. Après des années d'hibernation, elle reprend actuellement de la vigueur, surtout aux États-Unis. Pour assurer la sécurité du territoire, les Américains veulent mieux comprendre ce que les non-anglophones disent et trament. Je reviens d'ailleurs de la conférence nord-américaine de l'*Association for Computational Linguistics*. Pendant des

années, personne n'y a parlé de traduction; cette année le quart des publications portaient sur ce sujet.

Les logiciels de traduction automatique fonctionnent plus ou moins bien actuellement. S'ils sont utiles, c'est essentiellement lorsqu'on désire avoir une idée très approximative d'un document produit en langue étrangère.

Les travaux sur la traduction assistée par ordinateur (TAO) paraissent plus prometteurs. De notre côté, nous travaillons à la création de ce genre d'outils en recourant à des approches statistiques. L'idée, quand on emploie cette méthode, est de partir de grands corpus. D'un côté, on a le texte source; de l'autre, sa traduction. Cette mise en parallèle permet de dégager des règles et des constantes qui permettent de produire de nouvelles traductions ou de créer des outils d'aide à la traduction.

L'illustration 1 présente un système que nous avons développé en laboratoire: TSRali. TSRali permet de trouver la traduction française juste d'expressions en anglais et *vice versa*. Le logiciel vise à pallier les faiblesses de la traduction par chaînes de caractères. Par exemple, on ne peut traduire l'énoncé « tirer à boulet rouge sur quelqu'un » par « *to pull red bullets on someone* ». Cependant, c'est ce que font les logiciels de traduction par chaînes de caractères.

TSRali fonctionne plutôt en s'appuyant sur une mémoire de traduction, une mémoire constituée à l'aide des traductions humaines réalisées pour constituer le *Journal de la Chambre des communes* du Canada. En utilisant TSRali, le traducteur apprend qu'il est par exemple possible de traduire « tirer à boulets rouges » par « fire on ».

Le RALI mène aussi le projet Transtype. Transtype est un aide à la traduction interactif. Il suffit au traducteur de commencer à taper pour se voir proposer des traductions en temps réel.

Éventuellement, il serait intéressant que la machine puisse établir elle-même le niveau de confiance qu'elle a dans une traduction. L'ordinateur suspicieux devant sa propre performance pourrait alors faire part de ses doutes à l'utilisateur. La machine confiante pourrait faire son travail automatiquement, sans demander l'intervention de l'utilisateur. Toutefois, en attendant la création d'un tel mécanisme d'autocritique, le traducteur a le contrôle. Lorsqu'une traduction fait son bonheur, il l'accepte; sinon il continue à taper.

Dans un domaine connexe, nous travaillons actuellement aux questions de repérage translingue.

Autrement dit, nous essayons de créer des outils qui permettront aux usagers de produire une requête dans une langue – par exemple en français – et de repérer des textes dans une autre – par exemple en anglais. On ne parle pas ici de traduire la requête – souvent celle-ci est trop ambiguë –, mais bien de la jumeler à un texte dans une autre langue. Ce genre d'applications est intéressant, parce qu'il arrive souvent qu'une personne ait seulement une connaissance passive d'une autre langue.

«Après avoir confondu bien des sceptiques, le logiciel de traduction Transtype, mis au point au [RALI] passe la vitesse supérieure et devient une affaire internationale. Transtype 2 sera [...] bientôt utilisé non seulement par les traducteurs pour passer de l'anglais au français et inversement, mais aussi par ceux qui travaillent avec l'espagnol ou l'allemand.

«Notre premier logiciel a montré que la traduction assistée par ordinateur était possible; la seconde phase permet d'associer des partenaires européens à notre aventure et d'étendre les possibilités de l'appareil», a dit le professeur Guy Lapalme à l'occasion du lancement de Transtype 2, le 29 septembre dernier, peu avant la première réunion des partenaires en sol canadien.

C'est grâce à un financement du Conseil de recherches en sciences naturelles et en génie du Canada et à la participation financière d'un consortium européen que ce nouveau départ est possible. Au Canada, ce sont 750 000\$ qui seront investis en trois ans dans ce projet, permettant la création d'une vingtaine d'emplois, alors que les partenaires européens, universitaires et industriels, y injecteront quelque 2,6M\$. Au total, l'équipe canado-européenne recevra 3,4M\$.

Mathieu-Robert Sauvé, «Les logiciels de traduction passent à la deuxième génération», *Forum*, 20 octobre 2003, consulté le 5 novembre 2003 à l'adresse <http://www.iro.umontreal.ca/~lapalme>.

Besoins

Différentes mesures permettraient de favoriser les travaux francophones dans le domaine de l'écrit.

Premièrement, il serait important d'accentuer nos efforts en matière d'évaluation.

On y arrive parfois assez bien. Par exemple, en matière de recherche d'information, il n'est pas si difficile de savoir si les documents retracés par l'ordinateur sont pertinents ou non. Il suffit essentiellement de tenir compte des mots-clés employés à des fins de repérage, puis de compter le nombre de documents pertinents et non pertinents repérés pour avoir des statistiques en matière de précision et de rappel.

Dans d'autres domaines, mener des évaluations sera toutefois plus difficile. À titre d'exemple, les choses se gâtent lorsqu'on tente de mesurer la performance d'un système de traduction. En effet, il est difficile de déterminer avec exactitude ce qu'est une bonne ou, encore, une mauvaise traduction. Deux traductions peuvent être bien différentes l'une de l'autre – sur le plan de la longueur notamment – et être de qualité égale.

Actuellement, les gens qui travaillent au traitement du français écrit n'ont pas, souvent, pour priorité de se comparer les uns aux autres. Il faudrait corriger cette situation. Les résultats seront parfois rassurants. Après tout, comme le dit le dicton: «Quand je me regarde, je me désole; quand je me compare, je me console.»

Deuxièmement, il faudra mettre au point des matériaux langagiers.

D'abord, il ne fait pas de doute que la question des corpus est importante dans le secteur de l'écrit. Pendant longtemps, il a été difficile de mettre la main sur des corpus représentatifs. Les textes numérisés, en effet, étaient rares, en partie parce qu'il était difficile de saisir les textes papier. Maintenant le problème a changé, puisque les textes sont généralement créés directement à l'aide de la machine. L'arrivée du Web a aussi facilité l'accès à des documents intéressants.

La question du codage des textes demeure toutefois. Ainsi, on commence à s'entendre sur l'utilisation d'Unicode pour représenter les documents écrits, mais tous les documents ne sont pas encodés selon cette norme. De même, XML, standard de balisage, suscite beaucoup d'intérêt, mais il ne suffit pas de mettre des balises dans un document: il faut s'entendre sur leur signification et les former adéquatement.

Sur le plan des dictionnaires électroniques, il existe aussi des problèmes. Ainsi, il est sûr qu'il en existe en français. Cependant, ces outils sont moins disponibles en

français qu'ils ne le sont en anglais. Le problème vient de ce que les organismes qui les ont développés, souvent, refusent de les partager ou les vendent à un prix déraisonnable, comme un dollar ou un euro le mot. Si les gens ne jurent que par *WordNet*, dictionnaire anglophone, c'est parce que *WordNet* est relativement gratuit. *WordNet* a plein de défauts, mais il est bien fait, utile et d'accès très économique.

Je n'ai pas toutes les réponses aux problèmes de matériaux langagiers, mais, selon moi, une avenue intéressante pour les francophones serait de travailler de manière distribuée à la création de dictionnaires ou d'ontologies. Il faut absolument profiter de ce qu'Internet nous permet de collaborer les uns avec les autres. Il existe un précédent à cela : le projet *Papillon* vise en effet la

création d'un dictionnaire orienté traduction. Chaque personne peut contribuer à la création de l'outil en fournissant certaines entrées. Ces dernières sont ensuite validées à l'aide d'un procédé de validation bien robuste.

Conclusion

En conclusion, des progrès majeurs ont été enregistrés dans le domaine de l'écrit ces dernières années. Cependant, beaucoup de travail reste encore à faire, particulièrement en langue française.

Guy Lapalme,
professeur-chercheur, centre RALI, Université de Montréal, Québec.

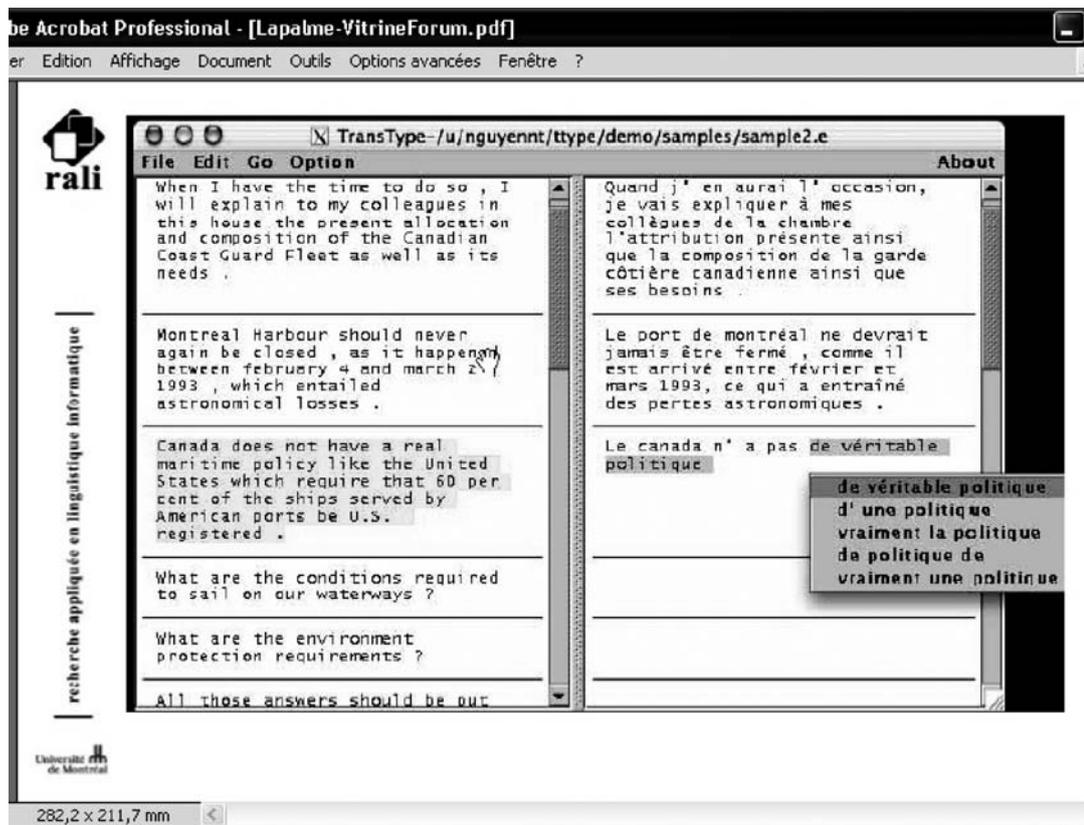
Illustration 1
Le système TSRALI

The screenshot shows the TSRALI web interface. At the top, the user is identified as 'utilisateur: lapalme'. Navigation links include 'Requêtes', 'Mon compte', 'Préférences', 'Aide', and 'Quitter'. A search bar shows 'Collection de documents : Hansard canadien (1986-2002)' and 'Expression : tiré... boulet+'. A 'Chercher' button is present. The search results are displayed in a table with three entries:

Id	Texte (Français)	Texte (Anglais)
1	Je puis dire en toute sincérité que j'ai été choqué l'autre jour quand la ministre a tiré à boulet rouge sur une politique qui proposait simplement que, lorsque des gens arrivent ici après avoir déchiré leurs documents, les avoir fait disparaître dans la toilette de l'avion pour cacher leur identité, nous devrions faire comme les autres pays sûrs et les placer en détention.	I can say in all sincerity that I was shocked when the other day the minister fired on a policy that simply says if people have arrived here having torn up their documents, having flushed them down the toilet on the plane and hidden their identity, we would do as other safe countries are doing and detain them.
2	On a tiré sur le messager à boulets rouges pour le faire taire.	They shot the messenger with red hot bullets to shut him up.
3	Je suis très surpris de voir que lorsqu'ils étaient dans l'opposition, et je termine là-dessus, ils s'attaquaient aux changements que proposait le gouvernement conservateur du temps à propos des 2249 pensions de vieillesse. Aujourd'hui, ce parti au pouvoir tire à boulets rouges sur les pensions de vieillesse.	In closing, I am very surprised to see that when they were in opposition, they attacked the changes to old age pensions proposed by the Tory government, but now that they are in office, they do not hesitate to assail old age pensions.

At the bottom left, the dimensions '282,2 x 211,7 mm' and a back arrow are visible.

Illustration 2
TransType



Les défis en matière de traitement de la parole

Le but de cet exposé est de donner un état de l'art du traitement automatique de la parole et, plus particulièrement, de la reconnaissance de la parole, et de présenter les grands axes de recherche innovants dans le domaine. Actuellement, les axes de recherche porteurs sont l'accès à l'information par le contenu audio et les systèmes d'interaction homme-machine.

Présentation du conférencier

Mon nom est Laurence Devillers, je suis maître de conférences à l'Université de Paris-Sud, en France, et membre du LIMSI-CNRS. Le LIMSI (Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur) possède un département de recherche sur la communication homme-machine. Une des unités du département s'intéresse plus particulièrement au traitement de la langue parlée. Ses activités portent notamment sur la reconnaissance vocale, l'indexation audio et la création de systèmes de dialogue homme-machine. Au sein de ce groupe, je participe à diverses activités de recherche, tant au niveau acoustique que linguistique : décodage acoustico-phonétique, modèles de compréhension de la parole, systèmes de dialogue et, enfin, détection des émotions à partir d'indices prosodiques et lexicaux. J'ai aussi beaucoup d'intérêt pour les projets de création de matériaux langagiers et d'évaluation.

Je collabore notamment à une action « Corpus » qui vise à mettre en place une plate-forme de stockage des corpus provenant de différentes universités, qu'il s'agisse de corpus produits par des linguistes, des spécialistes du geste ou de l'oral, etc. Cet instrument commun fonctionnera à l'aide de standards d'annotation reconnus. Cela permettra aux acteurs de synchroniser leurs efforts et de réutiliser les matériaux linguistiques disponibles.

En matière d'évaluation, j'ai participé aux activités de l'action de recherches concertées ARC B2 de l'Aupelf-Uref relatives à la définition de métriques d'évaluation des systèmes de dialogue. J'ai aussi participé au montage du projet français *Technolangue* MEDIA/EVALDA qui a débuté en janvier 2003. MEDIA porte sur l'évaluation de la compréhension *bors* et *en* contexte du dialogue. Le projet a pour but de mettre en place une méthodologie d'évaluation pérenne et de mener une campagne

d'évaluation auprès des principaux acteurs académiques et industriels français du domaine.

Des connaissances en progression rapide

Les premiers systèmes élaborés dans le secteur des technologies de reconnaissance de la parole ont été des systèmes de commande et de contrôle qui ne nécessitaient la reconnaissance que d'un petit nombre de mots isolés. Ils ne pouvaient fonctionner que dans un environnement très calme.

La dictée vocale, adaptée à la voix d'une personne, a été popularisée par des compagnies comme IBM (avec *ViaVoice*) ou Dragon Systems (avec *NaturallySpeaking*). Actuellement, les applications les plus importantes dans le domaine sont l'indexation des documents par le contenu audio et les systèmes de dialogue.

La reconnaissance de la parole est un domaine intrinsèquement pluridisciplinaire où interviennent des spécialistes de disciplines comme le traitement du signal, l'acoustique, la phonétique, la linguistique, l'intelligence artificielle et les sciences cognitives. Les modélisations les plus performantes en ce moment reposent sur des modèles statistiques produits à partir de corpus de données réelles.

Le signal de la parole est complexe. Cette complexité est due à plusieurs facteurs : à la coarticulation des sons avec les sons voisins (pas de coupure entre les mots), à des problèmes d'homomorphie (pour une même prononciation d'un mot plusieurs écritures sont possibles, par exemple en français de nombreuses lettres sont muettes), quatre axes principaux de variabilité du signal sont dus aux locuteurs, aux styles de parole, au langage de l'application et à l'environnement de transmission.

De nombreux progrès ont été faits depuis 20 ans. Les recherches sur la reconnaissance de la parole portent maintenant sur des tâches de plus en plus difficiles comme la parole conversationnelle (conversations téléphoniques par exemple) ou sur les programmes radiotélévisés : fictions, documentaires ou journaux télévisés. Les enregistrements traités peuvent être obtenus dans des environnements de transmission sonore difficiles

(enregistrements sur des lieux publics, au téléphone). Les systèmes de reconnaissance de la parole sont capables de prendre en compte de plus en plus de variabilité interlocuteur comme les accents régionaux ou étrangers.

Des modèles spécifiques sont développés pour chaque type d'application. Par exemple, reconnaître des textes de journaux télévisés n'a rien de commun avec reconnaître le contenu d'un texte de dictée vocale. Le traitement de la langue spontanée nécessite de disposer de modèles différents, principalement pour le modèle de langage et le lexique. Par exemple, le lexique de reconnaissance doit tenir compte des interjections et hésitations qui n'existent pas dans les modèles de dictée. Il doit aussi contenir des formes agglutinées et fragmentées, puisqu'en mode de parole spontanée (ou relâchée), tous les phonèmes qui constituent une phrase ne sont pas toujours prononcés. Par exemple : « J'drais un billet d'train demain. » Il faut donc pouvoir prendre en compte les variantes de prononciation des mots « je voudrais » en « j'drais », « de train » en « d'train ». Pour chaque application, il faut également des modèles acoustiques adaptés au canal de transmission, par exemple pour le canal téléphonique. Des modèles spécifiques sont aussi en général développés pour les locuteurs féminins et masculins.

Évaluation, corpus et outils

Les progrès réalisés, ces dernières années, en reconnaissance de la parole sont principalement dus à l'existence de campagnes d'évaluation et à la disponibilité de corpus.

Les programmes d'évaluation de l'agence militaire américaine DARPA, programmes auxquels le LIMSI participe depuis le début des années 1990, permettent d'obtenir énormément de données et de confronter les technologies françaises à celles de laboratoires américains prestigieux. En 1996, DARPA classait le système du LIMSI au premier rang des logiciels de reconnaissance de la parole. En 1997 et 1998, il était classé parmi les trois meilleurs systèmes. De 1999 à 2002, DARPA classait de nouveau le système du LIMSI au premier rang. Les campagnes d'évaluation ont beaucoup évolué au fil des ans. Ainsi, l'enjeu, de 1992 à 1995, consistait principalement à dicter des textes de journaux contenant des vocabulaires de plus

en plus grands. Depuis 1995, il s'agit de transcrire des émissions d'informations radiotélévisées de type HUB4 et des conversations téléphoniques de type HUB5.

Actuellement, les évaluations portent sur la capacité des systèmes à faire des transcriptions enrichies. L'enjeu n'est plus seulement de reconnaître ce qui a été dit, mais aussi de reconnaître qui l'a dit, dans quelle langue et dans quel style.

Dans un système de transcription de la parole, deux traitements sont utilisés. Le premier fait une analyse acoustique du signal audio et le transforme en une suite de vecteurs de paramètres acoustiques. Le second est le module de reconnaissance, il transforme la suite de vecteurs acoustiques en une suite de mots. Il utilise trois types de modèles : des modèles acoustiques, un dictionnaire de prononciation et des modèles linguistiques. Les modèles acoustiques sont des modèles de Markov cachés qui permettent de construire des modèles de sons à partir de la succession des vecteurs acoustiques. Ces modèles de sons tiennent compte des distorsions temporelles du signal de parole et utilisent des densités de probabilité multigaussiennes par état. Le dictionnaire est une ressource qui contient toutes les variantes de prononciation, à chaque mot du système de reconnaissance sont associées plusieurs formes phonétiques. Les modèles linguistiques sont des modèles de probabilité de succession des mots, des modèles statistiques n-grammes (bigrammes, trigrammes, quadrigrammes, etc.).

L'évaluation est aussi un sujet important en France. Citons l'action Aupelf-B1 en 1997, qui portait sur l'évaluation de la capacité des systèmes à faire la transcription de journaux lus et, actuellement, la campagne ESTER, qui va se prolonger jusqu'en 2005. ESTER a pour but l'évaluation de la capacité de différents systèmes à déterminer le contenu et les langues d'émissions d'informations radiophoniques et télévisuelles, de même que les types de locuteurs prenant part aux conversations. Ce projet fait intervenir plusieurs laboratoires français.

Voici quelques performances indicatives des systèmes de reconnaissance sur des tests en anglais américain. Ces performances sont mesurées en taux d'erreur sur les mots

à partir de la comparaison entre le texte reconnu et le corpus de référence transcrit manuellement. Les systèmes font peu d'erreurs quand il s'agit de reconnaître des nombres (0,7 %). Le taux d'erreur est de 7 % (0,5 % pour les performances humaines) dans le cas de journaux dont le contenu est lu à haute voix. Dans le cas des journaux télévisés, le taux grimpe à 20 % d'erreur. Pour une conversation téléphonique, il est de l'ordre de 35 % contre 4 % pour un être humain. La machine est encore loin des performances humaines.

L'évaluation des systèmes de dialogue est beaucoup plus difficile que l'évaluation des systèmes de reconnaissance de la parole. Actuellement, il n'y a pas de consensus sur la manière d'évaluer un système : on fonctionne beaucoup avec des critères subjectifs. Il n'y a pas de standard d'annotation non plus. En France, l'objectif de la campagne MEDIA du projet Technolanguage est d'amener les grands laboratoires et les entreprises du secteur à s'entendre sur des standards d'annotation et des façons d'évaluer la qualité des systèmes de dialogue.

Les modélisations les plus performantes en reconnaissance de la parole reposent sur des modèles statistiques qui nécessitent de grands corpus. On peut donner une idée de la taille des corpus utilisés pour la reconnaissance de grands vocabulaires. Le corpus audio pour entraîner les modèles acoustiques en américain contient 200 heures de documents sonores radiophoniques et télévisuels (ABC, CNN, ...) soit un total de 1,9 million de mots issus de transcriptions fines, c'est-à-dire qu'il est possible de faire un bon alignement entre le signal audio et la transcription linguistique. Pour construire des modèles acoustiques en français et en allemand, des sources radio et télévisées ont été enregistrées (Arte, TF1, A2, FrInfo, FrInter). Le *Linguistic Data Consortium* (LDC) fournit aussi des données dans d'autres langues que l'anglais américain comme le mandarin. Pour entraîner les modèles de langage, des données textuelles issues de journaux ou des transcriptions commerciales sont utilisées. Pour donner un ordre d'idées, en américain, 790 millions de mots issus de journaux et 240 millions de mots de transcriptions commerciales ont été utilisés.

Verrous technologiques

Il n'existe pas de système universel de reconnaissance de la parole. Le développement d'une application et l'entraînement des modèles nécessitent toujours le recours à des données réelles. Cependant, l'annotation de ces données est très coûteuse. Le défi consiste donc à favoriser l'adaptation rapide des modèles acoustiques et des modèles de langage dont nous disposons, à permettre à la machine d'apprendre automatiquement des prononciations. Il faudra aller dans cette direction si on veut être capable de produire des modèles performants à moindre coût. Pour illustrer ce propos : lorsqu'on entraîne un modèle acoustique à l'aide d'un corpus de 500 heures transcrites automatiquement et d'un corpus d'une heure transcrite manuellement, on obtient les mêmes performances que quand on a recours à un corpus de 200 heures transcrites manuellement.

Le « challenge » est donc de trouver des technologies génériques et de minimiser la collecte de corpus transcrits.

Au-delà de la reconnaissance de la parole

Les travaux en reconnaissance automatique de la parole s'étendent aussi à l'identification de la langue et du locuteur. Les mêmes technologies sont utilisées. Les difficultés sont différentes suivant les langues et les locuteurs et varient grandement selon le type de parole. Comme nous l'avons déjà souligné, il est difficile de transcrire automatiquement des journaux télévisés, mais il est encore plus difficile de transcrire automatiquement des conversations téléphoniques de parole spontanée.

Pour donner un ordre des performances de reconnaissance du locuteur, la machine ne se trompe que dans 1 % des cas quand elle a eu deux minutes pour se familiariser avec la voix de chacun des locuteurs appartenant à un groupe de 630 personnes enregistrées dans d'excellentes conditions en studios. Son taux d'erreur est nettement plus élevé, à 40 %, dans le cas d'enregistrements téléphoniques spontanés. Cette hausse importante du taux d'erreur vient en partie du fait que la machine ne peut pas reconnaître deux personnes qui parlent en même temps, phénomène fréquent dans l'oral spontané. En fait, on estime que les gens parlent en même temps jusqu'à 30 % du

temps, souvent pour maintenir la conversation en donnant des signes d'accord.

Pour donner un ordre de performance en identification de la langue sur le corpus OGI de parole téléphonique en 11 langues (anglais, farsi, français, allemand, hindi, japonais, coréen, mandarin, espagnol, tamoul, vietnamien), à 10 secondes, le système a un taux d'erreur d'identification de la langue de 20 %, à 45 secondes de parole, son taux d'erreur chute à 10 % seulement, le corpus d'apprentissage étant de 24 heures d'enregistrement de bonne qualité, de 1 heure 30 minutes à 4 heures 30 minutes par langue.

Quelques enjeux et applications innovantes

L'indexation audio représente un enjeu considérable dans le domaine de la reconnaissance de la parole. En effet, chaque pays produit et stocke chaque année des centaines de documents audiovisuels qu'il faudrait arriver à structurer et à organiser à moindre coût. Une des manières d'aborder la question consiste à indexer les documents audiovisuels de manière automatique, à l'aide de la reconnaissance de la parole. Les applications de cette technologie seront multiples : elle facilitera la recherche documentaire, permettra de produire de l'information à la demande, facilitera la surveillance des médias, etc.

Il faut d'abord, pour faire de l'indexation automatique, structurer le signal audio en tours de parole, segments musicaux, etc. Ensuite, le contenu audio est transcrit automatiquement avec un système de reconnaissance de la parole. Finalement, l'indexation peut se faire. Ceci signifie que les textes doivent avoir été annotés et balisés par des thèmes de référence. Le procédé consiste à utiliser des lemmatiseurs et des filtres sur les mots les plus connus, de manière à produire des modèles thématiques. La recherche de document se fait ensuite par une requête en langage naturel. Voici un exemple de la manière dont fonctionne la recherche de document. La requête de l'utilisateur subit un pré-traitement. Par exemple, la question : « Quelles sont les mesures prises contre la maladie de la vache folle ? » devient après lemmatisation « Quelle mesure prendre maladie vache folle ? » Cette requête est ensuite enrichie par l'ajout des termes les plus fréquemment retrouvés dans les documents du corpus les plus pertinents à la requête. Une deuxième requête ainsi constituée est lancée. Le moteur de recherche

se sert de l'index pour extraire les documents audiovisuels les plus pertinents, c'est-à-dire ceux qui ont une similarité entre les termes de la requête et ceux des segments indexés.

Le LIMSI a mis au point un système d'indexation audio pour une application d'alerte : Audiosurf. Ce système est un prototype de laboratoire. Une personne s'inscrit à ce service en indiquant les thématiques qui l'intéressent. Tous les jours, Audiosurf indexe les journaux télévisés diffusés sur plusieurs chaînes de télévisions nationales (Fr2, Arte...). La base contient actuellement une centaine d'heures de journaux télévisés. Quand le système détecte la présence d'un thème, il envoie un courriel au client avec l'extrait de la vidéo du journal détecté. Par exemple, le client intéressé au thème *météorite* se verrait envoyer l'extrait du document vidéo correspondant au bulletin sonore suivant :

« Des scientifiques sont en train d'arriver sur place. Romuald Bonnant : « Un objet lumineux fonçant dans le ciel à plus de 200 000 kilomètres/heure et puis, quelques secondes plus tard, le sol qui tremble. Voilà ce que les habitants de deux villages sibériens ont vu et ressenti lorsque tout à l'heure, en pleine nuit, une météorite a heurté la terre à pourtant plus de 100 kilomètres de chez eux. Une importante pierre céleste... » Le document détecté correspond bien au thème recherché, malgré 20 % d'erreurs de reconnaissance dans ce type d'enregistrement dues en majorité à des fautes d'accord grammatical, des homophones et des noms propres. Il a été montré que 20 % d'erreurs de transcription est la limite pour ne pas dégrader les performances de détection des documents pertinents.

Un autre champ d'application intéressant de la reconnaissance de la parole est celui de la traduction parole-parole. Le défi consiste à faire de la reconnaissance de conversations dans des domaines très différents. Ce type de système est évidemment difficile à réaliser. Une difficulté qui s'ajoute aux difficultés de traduction vient du fait que l'on part de documents oraux où il y a des erreurs de reconnaissance. On peut, par exemple, imaginer qu'un tel système de traduction fonctionnerait sur le mode du dialogue. Il serait possible à l'utilisateur de discuter avec le système pour raffiner la traduction qu'il veut obtenir. Le groupe TLP au LIMSI participe à un projet de ce type dans le cadre d'un projet européen du FP6.

Un autre axe très porteur de la reconnaissance de la parole est celui des systèmes d'interaction homme-machine.

Où désirez-vous aller?

Voici un exemple d'un système de dialogue développé au LIMSI dans le cadre d'un programme européen. Il s'agit d'un système utilisé pour donner des renseignements sur les horaires de train de la SNCF. Le lexique de ce genre d'application multilocuteurs est assez restreint. Il compte seulement 2 000 mots.

Système: Quel renseignement désirez-vous? Locuteur: Je voudrais aller de Paris à Avignon demain à 14 h 30. Vous avez un train direct de Paris à Avignon à 15 h 30 arrivant à 18 h 50 cécans le 12 octobre. Qu'en pensez-vous? Voulez-vous réserver? Quel est le prix en première classe non-fumeur avec une réduction Carissimo? Le prix du billet Paris-Avignon est de 301 francs en première classe avec une réduction Carissimo. Souhaitez-vous une autre information? Je souhaiterais aller de Grenoble à Clermont-Ferrand à Noël. Jeudi 25 décembre. À quel moment de la journée voulez-vous partir? À 14 h 35. Pardon je ne vous ai pas compris, pourriez-vous répéter svp? À 14 h 30. Vous avez un train partant de Grenoble à 14 h 08, arrivant à Clermont-Ferrand à 20 h 53 avec un changement à Paris, gare de Lyon, le 25 décembre.

Une version de ce système d'information sur les horaires de train de la SNCF est maintenant accessible au grand public par téléphone. Une autre fonctionne à partir d'un kiosque multimodal qui a été évalué à la gare Saint-Lazare auprès de vrais voyageurs, mais reste encore un prototype de laboratoire. Pour le protocole de test, chaque sujet devait effectuer quatre transactions. Quatre-vingt-treize p. cent de celles-ci ont été réalisées avec succès. La perception du public par rapport au logiciel s'est avérée excellente: 87 % des sujets ont dit trouver plus agréable de transiger avec une borne compréhensive et agréable à l'oreille plutôt qu'avec les bornes actuelles.

Un autre axe innovant du domaine est la détection des émotions. Par exemple, il serait intéressant que la machine conversant avec le client de l'extrait suivant puisse discerner son inquiétude: « Voilà, voilà, mon problème est le suivant: ça m'indique "compte verrouillé". » Cela pourrait l'amener à changer de stratégie de communication ou, tout simplement, à transmettre l'appel à un opérateur humain.

Parmi les applications innovantes dans le domaine de l'interaction orale, il faut citer aussi les systèmes de communication homme-homme médiatisée par la machine. Ce type de système pourrait être capable d'observer, de reconnaître et de comprendre des communications et des comportements interpersonnels et de rendre des services aux usagers. On pourrait par exemple imaginer l'existence d'un petit système complice adapté à chaque propriétaire. Le système aurait une interface multicapteur et multimodale robuste rendant possible la saisie de sons et d'images dans un environnement bruyé et la possibilité de stocker et d'organiser l'information. Un utilisateur dialoguant avec ce système pourrait, par exemple en réunion, retrouver l'information, identifier les personnes dans la salle, etc.

En conclusion, des progrès considérables en reconnaissance de la parole ont été faits au fil des ans, mais nous sommes encore loin des performances de l'être humain. La difficulté vient des nombreux niveaux de variabilité du signal de parole: conditions acoustiques, divers accents, parole superposée, parole spontanée, etc. La production d'un système universel de reconnaissance de la parole n'est pas encore pour demain. Nous devons également faire face à de nombreux défis quand vient le temps de créer des applications allant au-delà de la reconnaissance de la parole, comme celles de la traduction parole-parole ou celles de la compréhension de conversation entre deux personnes. Ces domaines innovants sont des enjeux scientifiques pour l'avenir et nous sommes encore loin de réaliser ce que les films de science-fiction nous promettaient voilà plusieurs années.

*Laurence Devillers,
maître de conférences à l'Université de Paris-Sud et
membre du LIMSI, France.*

Présentations
effectuées lors de la
vitrine des produits
et des prototypes

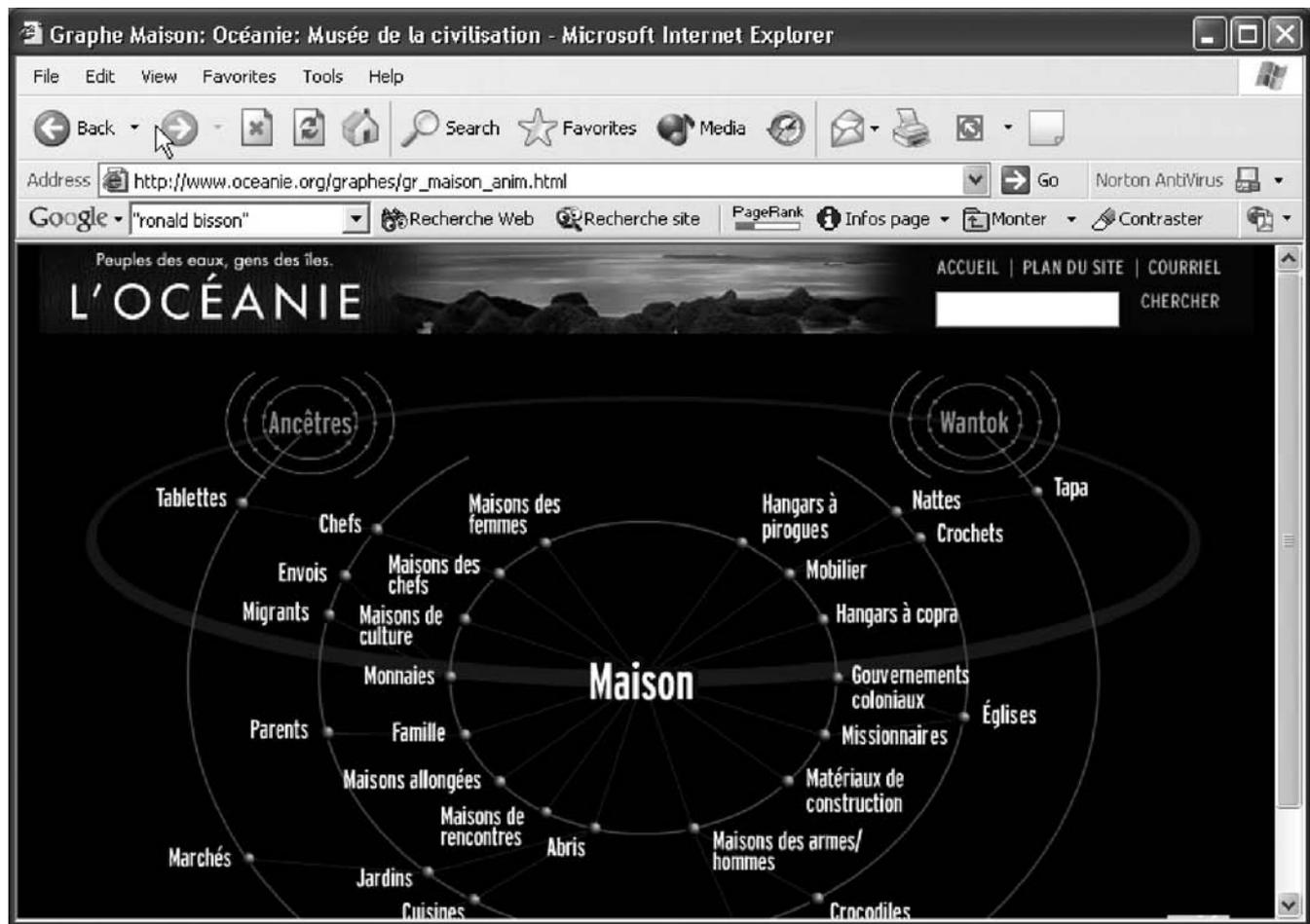
Présentations effectuées lors de la vitrine des produits et des prototypes

Oceanie.org

Oceanie.org, site Web créé dans le cadre d'un projet Québec-France, vise à faire connaître l'Océanie et les peuples qui l'habitent. La section *Excursion* du site propose des résumés sur l'histoire du continent, ses cultures, son environnement et les enjeux qu'il doit relever. La section *Graphes* présente, pour sa part, une vision hypertexte de certains aspects culturels de l'Océanie.

Le site fonctionne à l'aide d'entrées appelées *attracteurs*. Créés à partir d'un modèle dérivé des neurosciences, les attracteurs permettent de créer des grappes de sens (ex. : une grappe constituée en son centre de l'attracteur *maison* et en périphérie de concepts comme *famille* et *mobilier*).

Oceanie.org a été conçu dans le cadre du projet Encyclopédie culturelle hypermédia de l'Océanie (ECHO). Pierre Maranda, professeur émérite à l'Université Laval, dirige l'équipe scientifique du projet.



Terminotix

Fondée par des professionnels de la traduction, Terminotix (www.terminotix.com) est une entreprise spécialisée dans le domaine des aides informatiques à la traduction. Par les logiciels qu'elle commercialise et par les services qu'elle offre, Terminotix souhaite aider les

professions langagières à se donner des outils et des pratiques de travail modernes et efficaces.

Logiterm est le produit phare de Terminotix. Fruit de la collaboration de Claude Bédard, d'Agir Inc., du Laboratoire RALI (Université de Montréal) et de Terminotix, Logiterm se veut un outil de traduction assistée par ordinateur d'utilité universelle.

Terminotix Inc. / LogiTerm / Visite guidée - Microsoft Internet Explorer

Recherche

Base Terminologie
Recherche restreinte

Voici le résultat d'une recherche lancée au moyen du bouton **R. floue**. LogiTerm présente le nom des documents de fiches en cause et le contenu des principaux champs.

Les résultats sont classés par ordre de pertinence :

- Les occurrences qui comportent tous les mots demandés, contigus et dans l'ordre, sont affichées en priorité.
- Puis celles où tous les mots existent de façon non contiguë ou dans une autre séquence;

1/4

LogiTerm 2.0

Fichier Édition Outils Bases Options Fenêtres Aide

Grille des résultats C:\LOGITERM-1\bases\termino

Nom du fichier	eng	fra	dom	sou
d:\terminotix\comst\files07.doc	managed internet access solution	solution d'accès Internet géré	CCOM	Yvonne lex nf
d:\terminotix\comst\files06.doc	managed access link	liaison d'accès gérée	CCOM	Rapport annuels
d:\terminotix\comst\files06.doc	Internet access provider Internet service provider, commercial access provider, public access provider, commercial Internet provider	fournisseur d'accès Internet, prestataire de services Internet, prestataire d'accès à Internet, fournisseur d'accès à Internet, fournisseur d'accès Internet	INET TELE CCOM	CB, Lexique OLF Internet..
d:\terminotix\comst\files06.doc	Internet service provider Internet access provider, ISP, IAP	fournisseur d'accès Internet	CCOM INET TELE CCOM	CB
d:\terminotix\comst\nouvelles.doc	wireless internet access	accès Internet sans fil	TELE CCOM	CB
d:\terminotix\comst\files06.doc	Internet access provider/ IAP	fournisseur d'accès Internet/ FAI	CCOM TELE	Lexique Téléglobe
d:\terminotix\comst\files06.doc	Internet access	accès Internet	TELE CCOM	CB
d:\terminotix\comst\files06.doc	Internet access	accès à Internet	INFO TELE CCOM	CB
d:\terminotix\comst\files06.doc	high-speed Internet access	accès ultra-rapide à Internet	TELE CCOM	Doc. Epsilon

Champ principal: managed internet access link ENG

Chercher dans: Sous-bases

Terminologie Bitextes Plein texte

Champ sec: FRA

Domaines: CCOM,TELE Modifier... Aide Pluriel

247 occurrence(s) dans 34 document(s), 04 pertinence(s)

97

Oralys

Conçu par la société Oralys (www.oralys.com), le communicateur Oralys est une solution unique et conviviale qui permet à l'utilisateur de communiquer rapidement et facilement par le toucher. Ses utilisations sont multiples:

– téléthèse pour personnes éprouvant des difficultés de communication orale (dysphasie, aphasie, problème d'élocution);

- outil de réadaptation, orthophonie, autisme, déficience intellectuelle;
- outil de communication, téléthèse pour personnes handicapées;
- apprentissage d'une langue parlée et écrite.



MultiCorpora

MultiTrans, produit vedette de la compagnie MultiCorpora (www.multicorpora.com), est un outil de soutien à la traduction et un outil de gestion langagière qui permet de créer rapidement, à partir de vastes recueils de documents préalablement traduits et d'autres contenus, des banques facilement parcourables d'expressions de toutes

longueurs qui peuvent être visualisées dans leur contexte initial. En combinant sa technologie évoluée d'indexation et de recherche se fondant sur les corpus à une infrastructure client-serveur élaborée de gestion terminologique, MultiTrans fournit aux rédacteurs, aux traducteurs, aux terminologues et aux consommateurs de contenus des capacités sans pareil en matière de soutien à la traduction, de compréhension du langage et de gestion terminologique.

The screenshot shows the 'Démonstration Interactive de MultiTrans - Microsoft Internet Explorer' window. The interface is divided into several sections:

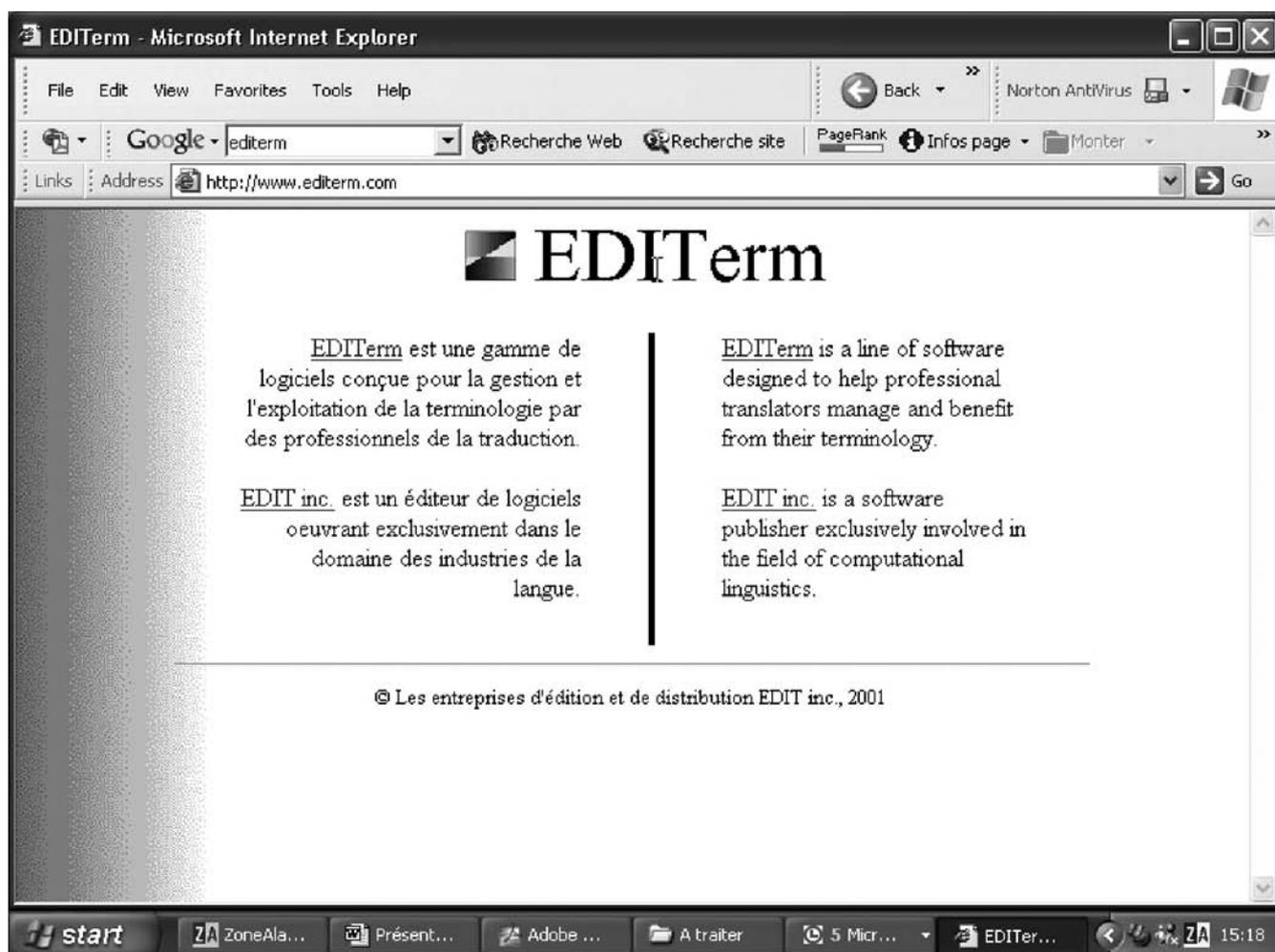
- Left Panel:** Contains instructional text in French explaining how to use the search tool and how to view translations in context.
- Search Panel:** Features a search bar with 'prime minister' entered, options for 'Mots consécutifs' (checked) and 'Résultats partiels' (unchecked), and a results table.
- Results Table:**

#	Source	Phrase	Mot
1	en001	94	5
2	en001	114	2
3	en001	159	14
4	en001	309	5
5	en001	321	5
- Source (English):** Displays the original text snippet: "[Translation] OATHS OF OFFICE Right Hon. Jean Chrétien (Prime Minister, Lib.) moves for leave to introduce Bill C-1, respecting the administration of oaths of office. [Motion agreed to and bill read the first time.] [English] SPEECH FROM THE THRONE".
- Cible (French):** Displays the translated text snippet: "[Français] LES SERMENTS D'OFFICE Le très hon. Jean Chrétien (premier ministre, Lib.) demande à présenter le projet de loi C-1 relatif à la prestation des serments d'office. [La motion est réputée adoptée et le projet de loi est lu pour la première fois.] [Traduction] DISCOURS DU TRÔNE".
- Bottom Panel:** Shows '250 correspondances', 'Source : 94', 'Cible : 92', and 'Delta : -2'.

EDITerm

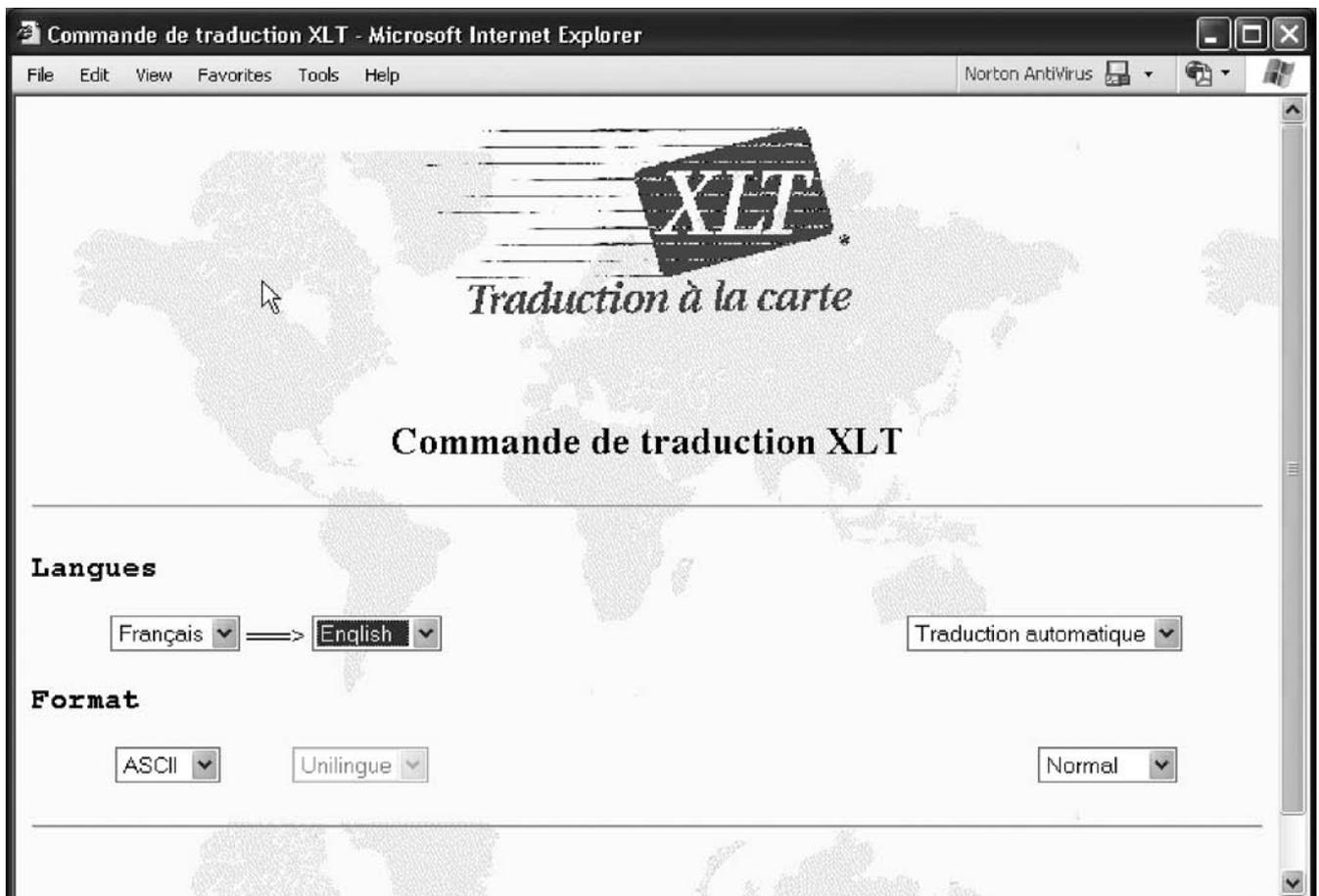
EDITerm (www.editerm.com) est une solution de gestion terminologique conçue pour les langagiers par la société John Chandioux experts-conseils. EDITerm vise essentiellement à assurer l'utilisation d'une terminologie uniforme et précise par trois types d'utilisateurs: les terminologues, qui sont responsables du contenu linguistique; les rédacteurs unilingues, qui doivent

maintenir une productivité élevée tout en disposant d'une terminologie exacte et à jour dans leur langue; les traducteurs, qui doivent produire aussi rapidement que possible une version fidèle d'un texte dans une autre langue, tout en utilisant une terminologie appropriée dans la langue cible. Ce qui distingue EDITerm des autres bases de données terminologiques, c'est la disponibilité de modules de dépouillement, d'aide à la traduction et de diffusion de la terminologie.



Socatra

La société Socatra (www.socatra.ca) décrit le logiciel XLT comme le plus puissant des systèmes de traduction automatique. Lorsqu'il faut traduire rapidement une note de service, un discours, un contrat ou tout autre document, la traduction automatique XLT peut fournir un premier jet en quelques secondes.



OTIL

L'Observatoire du traitement informatique des langues et de l'inforoute (OTIL) réalise des travaux de veille technologique et documentaire, et diffuse cette information; il promeut l'utilisation d'outils TIL et la production de contenus en français; il travaille, en collaboration avec les observatoires francophones, à des projets en normalisation, formation, veille technologique et documentaire, développement terminologique, évaluation

et labellisation; enfin, il dresse et tient à jour un répertoire électronique des intervenants francophones en TIL et de leurs productions.

L'*Inventaire francophone du traitement informatique des langues* (<http://cfwb.otil.org>) permet aux utilisateurs de trouver un produit ou service francophone répondant à leurs besoins; aux professionnels de promouvoir leurs activités TIL et de localiser des partenaires potentiels; aux autorités de la Francophonie d'appuyer leur politique sectorielle sur des données de qualité mises à jour régulièrement.

Inventaire francophone du traitement informatique des Langues - Microsoft Internet Explorer

File Edit View Favorites Tools Help Norton AntiVirus

Rifal Inventaire Francophone du Traitement Informatique des Langues

Centres de recherche Sociétés Projets Produits Services Ressources Distributeurs

- Domaines d'activité
- Adresse, contacts
- Liens vers projets, produits, services

Centres de recherche ?
Centre universitaire / non universitaire, centre d'excellence, académie, ...

Signalez vos activités
Saisie de données
Pour en savoir plus :
Qui peut s'inscrire ?
Comment s'inscrire ?

A consulter
Agenda TIL
Infos TIL

TIL : Traitement Informatique des Langues

Entrez un ou plusieurs critères de recherche

Nom (début)

Domaines

Tous les domaines

Certains domaines:

Analyse et génération de texte Outils inforoutiers

Education, formation Rédaction et édition de texte

Evaluation Systèmes experts

Informatique documentaire Traduction, lexicographie, terminotique

Localisation Traitement de la parole

Normalisation, internationalisation Traitement de l'écrit

Localité

Pays

Tous

Afrique de l'Ouest Haïti République Démocratique du Congo

Belgique (CFWB) Madagascar Roumanie

Canada Maroc Suisse

Centrafrique Québec

France

Termdat

TERMDAT contient les données d'EURODI-CAUTOM, banque de données terminologiques de la Commission de l'Union européenne à Luxembourg, et celles des collections terminologiques suisses (par exemple, de l'administration fédérale, de l'administration du canton de Berne, etc.).

TERMDAT est l'une des plus grandes banques de données terminologiques du monde. Elle est accessible gratuitement à tous les collaborateurs et collaboratrices des collectivités suisses de droit public, y compris les universités et les instituts de recherche (plus de 2 000 utilisateurs et utilisatrices actuellement en Suisse).

Doc. Ref 1	ACH	SYL95	2100004	AG8 JUA	4	990710
	<u>si</u>	<i>Wytweide</i> (1); bestockte Weide(2); bewaldete Weide(3)				
	<u>si</u>	Fläche, auf der Waldbestockungen und offene Weideplätze mosaikartig abwechseln und die sowohl der Vieh- als auch der Forstwirtschaft dient.				
	<u>si</u>	(1)(2)Waldgesetz, Art. 2 Abs. 2 Bst. a (AS 1992 2521); (3)Fillbrandt, Wörterbuch für Forsteinrichtung, 1993; (DF)Waldverordnung, Art. 2 (AS 1992 2538)				
	<u>si</u>	REG:CH; USG:(1)(2)CH,(3)				
	<u>si</u>	pascolo alberato				
	<u>si</u>	Superficie che serve all'allevamento di bestiame nonché alla selvicoltura sulla quale si alternano, in guisa di mosaico, estensioni boschive e pascoli aperti.				
	<u>si</u>	(VE)LF Foreste, art. 2 cpv. 2 lett. a (RU 1992 2521); (DF)O Foreste, art. 2 (RU 1992 2538)				
	<u>si</u>	REG:CH				
	<u>si</u>	pâturage boisé				
	<u>si</u>	Surface sur laquelle alternent, en forme de mosaïque, des peuplements boisés et des pâturages sans couvert et qui sert aussi bien à la production animale qu'à l'économie forestière.				
	<u>si</u>	(VE)LF Forêts, art. 2 al. 2 let. a (RO 1992 2521); (DF)O Forêts, art. 2 (RO 1992 2538)				
	<u>si</u>	REG:CH				

Sylex

Le projet Sylex (<http://issun17.unige.ch/sylex/>) a pour but de créer un outil qui permette l'affichage multilingue de mots, de phrases ou de sous-phrases, avec leur contexte dans les textes déjà traduits du *Recueil systématique du droit fédéral*. Cet outil doit permettre aux traducteurs de rechercher des exemples de traductions ainsi qu'aux réviseurs de vérifier la traduction d'un mot, d'une phrase ou

d'une tournure de phrase dans son contexte. Il se distingue des concordanciers multilingues intégrés aux logiciels du type « poste de travail de traducteur » par ses possibilités de recherche plus sophistiquées, en termes de requêtes possibles, et plus pointues, en termes de résultats obtenus, puisque basées sur les techniques linguistiques plutôt que sur les techniques statistiques de « correspondance approximative ».

SYLEX (Version 2.0) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Mail Address

Google ://issun17.unige.ch/sylex/ Recherche Web Recherche site PageRank Infos page Monter Contraster

Projet SYLEX

TIM/ISSCO 2000/2002

Langue	Expressions	Mode	Connecteur	Catégorie syntaxique	Affichage
Français	bilinguisme et cantons	Préfixe	Et	Contraintes sur les mots. p.e. suisse.ADJ ou suisse.NOUN	Par phrase
Allemand		Préfixe	Et	liste des catégories	Sequentiel
Italien		Préfixe	Et	ADJ AUX CONJ DET	
				NOUN NUM PR-NOUN PRON	
				PREP SEPART VERB	
				1 Envoyer	
				Mode d'emploi	

Outil de consultation multilingue des textes fédéraux

TIM/ETI (ex-ISSCO), Gilbert ROBERT, Maghi KING, Pierrette Bouillon et Sabine Lehmann
V4.0 - Janvier 1999

Mode d'emploi

Le projet Sylex a pour but de créer un outil qui permette l'affichage multilingue de mots, de phrases ou de sous-phrases, avec leur contexte dans les textes déjà traduits du *Recueil Systématique du Droit Fédéral*. Cet outil doit permettre aux traducteurs de rechercher des exemples de traductions ainsi qu'aux réviseurs de vérifier la traduction d'un mot, d'une phrase ou d'une tournure de phrase dans son contexte. Il se distingue des concordanciers multilingues intégrés aux logiciels du type

start ZoneAla... Présent... Adobe ... A traiter 5 Micr... 3 Inte... 14:59

Le *Grand dictionnaire terminologique*

Fruit de 30 années de travail de la part d'une équipe de terminologues chevronnés de l'Office québécois de la langue française, le *Grand dictionnaire terminologique*

(www.granddictionnaire.com), maintenant doté d'une nouvelle interface multifonctionnelle, donne accès à près de trois millions de termes français et anglais du vocabulaire industriel, scientifique et commercial, dans 200 domaines d'activité.

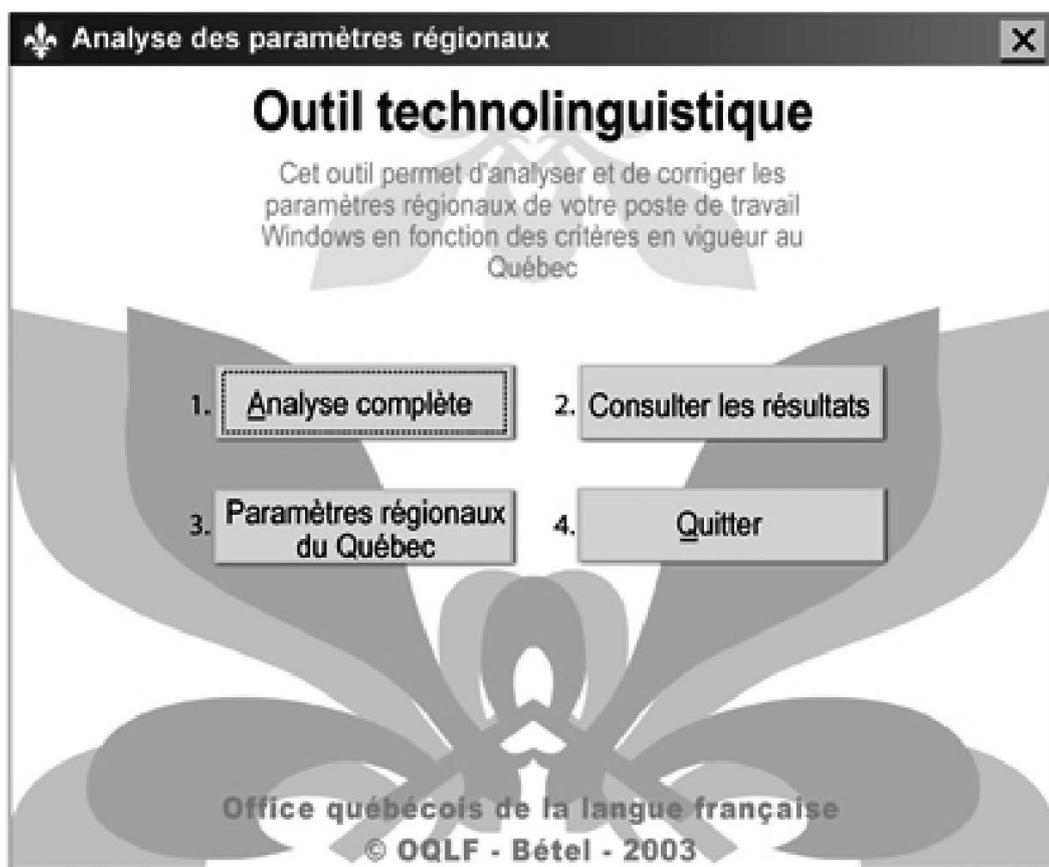
The screenshot displays the web interface of the Grand dictionnaire terminologique. At the top, there is a navigation bar with the logo of the Office québécois de la langue française and the title 'Le grand dictionnaire terminologique'. Below this is a search bar with the text 'internet' and a search icon. The interface is divided into several sections:

- 1. Recherche:** Includes dropdown menus for 'Langue d'interrogation' (français) and 'Langue de l'équivalent' (aucune), and a search box containing 'internet'.
- 2. Index:** A list of search results including 'Internet informatique (1)', 'Internet 2', 'Internet à haut débit', 'Internet à haute vitesse informatique (1)', 'télécommunication (1)', 'internetais', 'Internet Assistant', 'Internet commercial', 'Internet complet', 'Internetel', 'Internet enfoui', 'Internet Explorer', 'Internet gratuit', 'Internet haut débit', 'Internet haute vitesse', and 'Internet II'.
- Domaine(s):** A list of domains: '- informatique', 'Internet', and '- télécommunication'.
- français:** The main entry: 'accès à haut débit à Internet n. m.'
- Équivalent(s):** The English equivalent: 'English high-speed Internet access'.
- Définition:** 'Ensemble des techniques qui peuvent être utilisées pour augmenter la vitesse de transmission des données, dans les réseaux donnant accès à Internet.'
- Sous-entrée(s):** A list of synonyms: 'synonyme(s)', 'accès Internet à haut débit n. m.', 'Internet à haut débit n. m.', 'accès à haute vitesse à Internet n. m.', 'accès Internet à haute vitesse n. m.', 'Internet à haute vitesse n. m.', 'accès rapide à Internet n. m.', 'Internet rapide n. m.', 'accès Internet large bande n. m.', and 'Internet large bande n. m.'

Bétel

Bétel, Banc d'évaluation technolinguistique de l'Office québécois de la langue française, permet de vérifier les paramètres régionaux de l'utilisateur courant d'un poste de travail Windows en fonction des normes en vigueur au Québec.

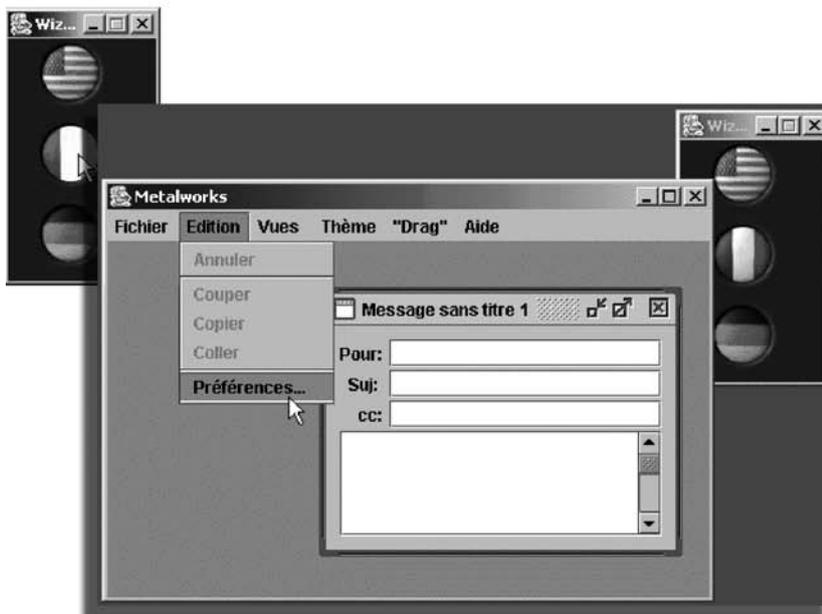
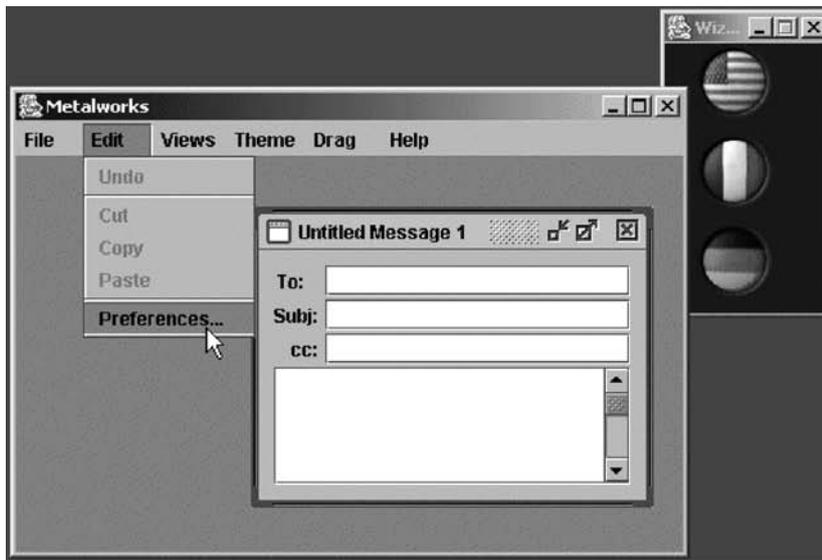
Les paramètres régionaux sont un ensemble de valeurs du système d'exploitation qui influencent la façon dont les programmes trient et affichent les dates, les heures, les monnaies et les nombres. Plus de soixante critères configurés par défaut lors de l'installation d'un système d'exploitation Windows sont examinés.



WizArt

WizArt (www.wizart.com) est un groupe spécialisé dans le développement de technologies de localisation. Les outils de WizArt aident à rendre multilingue n'importe

quelle application Windows, Web ou Java, sans égard à sa langue de départ et sans qu'il soit nécessaire d'en connaître le code source. Plus de 110 compagnies utilisent les technologies de localisation de WizArt, dont Air Liquide, Belgacom, le Groupe Jean Coutru et McDonald's Europe.





Déclaration finale et proposition d'un plan d'action

Déclaration finale

Montréal, le 14 juin 2003

Les participants de la Vitrine-forum sur le traitement informatique des langues (VF), qui s'est tenue les 13 et 14 juin 2003 à Montréal, constatant la nécessité d'intensifier

l'informatisation des langues dans l'espace francophone, déclarent ce qui suit :

1. Pour gérer l'information et favoriser la diversité culturelle – deux défis majeurs pour le XXI^e siècle –, il est indispensable de développer le traitement des langues à l'aide de logiciels de plus en plus évolués. Ces logiciels incluent notamment : les banques de terminologie, les outils de gestion de données textuelles et audio, les aides à la traduction, les outils de reconnaissance et de synthèse de la voix, les correcteurs, les générateurs de documents et les moteurs de recherche.
2. Les besoins restent encore très grands, malgré l'importance stratégique du traitement informatique des langues et les actions déjà entreprises dans ce domaine.
3. L'épanouissement des capacités de traitement informatique du français et des langues partenaires est susceptible d'avoir des retombées à tous les points de vue :
 - a) sur le plan économique, il pourra aider les travailleurs de la Francophonie à être plus productifs;
 - b) sur le plan social, il pourra rendre plus conviviale l'utilisation des ordinateurs par l'ensemble de la population des pays francophones;
 - c) sur les plans linguistique et culturel, les technologies linguistiques, particulièrement les outils d'aide à la traduction, les dictionnaires électroniques bilingues et les moteurs de recherche translingues, rendront possible l'établissement de passerelles entre le français et les autres langues, dont les langues partenaires de la Francophonie, contribuant ainsi à la diversité linguistique et culturelle;
 - d) sur le plan politique, il pourra augmenter la visibilité d'une Francophonie plurielle et renforcer le statut du français et des langues partenaires.
4. Les organismes et les gouvernements de la Francophonie devraient intensifier leurs efforts pour soutenir les chercheurs, particulièrement ceux du Sud, et les industriels actifs dans le domaine du traitement informatique du français et des langues partenaires.

Par conséquent, nous proposons la prise en compte, par les organismes et les gouvernements de la Francophonie, du plan d'action présenté ci-après.

Proposition d'un plan d'action

Vitrine-forum sur l'informatisation des langues

Attendu l'importance stratégique du traitement informatique des langues (TIL) pour l'épanouissement économique, social, culturel, linguistique et politique des pays de la Francophonie, les participants de la Vitrine-forum sur l'informatisation des langues, tenue les 13 et 14 juin 2003 à Montréal, proposent la réalisation des actions suivantes :

Les gouvernements de la Francophonie et certains des opérateurs du Sommet, l'Agence intergouvernementale de la Francophonie (AIF) et l'Agence universitaire de la Francophonie (AUF) sont conviés à appuyer financièrement la production de matériaux langagiers, tels les corpus écrits et oraux, pour le français et les langues partenaires. Ils sont aussi invités à assurer que les ressources langagières développées avec des fonds publics sont mises à la disposition de l'ensemble des acteurs. Dans cette veine, ils devraient particulièrement soutenir le développement et l'élargissement de l'*Inventaire du traitement informatique des langues* déjà mis en place par l'AIF.

Les chercheurs et les industriels actifs dans le champ du TIL n'ont pas toujours à leur disposition les ressources langagières (ex. : les corpus écrits et oraux, les lexiques informatisés, les bases de données terminologiques) dont ils auraient besoin pour leurs travaux. Et lorsque ces ressources existent, ils ont parfois de la difficulté à les repérer ou à y accéder. Le développement et l'élargissement de l'*Inventaire du traitement informatique des langues*, base de données accessible sur Internet qui recense déjà, entre autres, les produits et les entreprises du secteur du TIL, faciliteraient notamment le repérage des matériaux langagiers requis par les acteurs francophones.

Les participants de la Vitrine-forum s'entendent sur le fait que le prochain défi en matière d'informatisation des langues est leur traitement sémantique. Par conséquent, les gouvernements de la Francophonie, l'AIF et l'AUF devraient appuyer

l'établissement de lexiques ou de bases de connaissances sémantiques.

Les participants reconnaissent que l'établissement de lexiques ou de bases de connaissances sémantiques, sur le modèle de *WordNet* – malgré les critiques qu'il est possible de faire de cet instrument anglophone sur le plan technique –, donne une longueur d'avance au monde anglo-saxon. La création d'outils de ce genre pour le français aiderait la Francophonie à faire face à l'émergence de l'Internet sémantique.

Les gouvernements de la Francophonie, l'AIF et l'AUF sont conviés à soutenir financièrement, au sein de leurs programmes respectifs, l'informatisation des langues partenaires.

Par exemple, les gouvernements de la Francophonie, l'AIF et l'AUF pourraient faire les arrimages qui s'imposent, dans le cadre de Teltic⁴⁸, avec les responsables du projet d'Observatoire d'indicateurs stratégiques, pour aider l'Observatoire à adopter des pratiques exemplaires en matière de traitement informatique des langues. Ils pourraient aussi appuyer la réalisation du projet « Déverrouillage » du groupe Telal⁴⁹ sur l'informatisation des langues africaines. Ils pourraient enfin assurer que l'AIF et l'AUF font les arrimages qui s'imposent avec les responsables du projet Teltic sur la traduction automatique dans les trois espaces linguistiques, de manière à en connaître les résultats et, éventuellement, à y prendre part.

Les gouvernements de la Francophonie, l'AIF et l'AUF devraient faciliter, selon leurs moyens, la participation des francophones du Nord et du Sud aux tables de normalisation internationale pour que les nouvelles normes tiennent compte du français et des langues partenaires.

Des organisations internationales comme l'Organisation internationale de normalisation (ISO), l'*Internet Engineering Task Force* (IETF) ou le *World Wide Web Consortium* (W3C) travaillent à l'élaboration des normes et des standards qui

48. Groupe de travail sur les nouvelles technologies de l'information et de la communication des trois espaces linguistiques (Francophonie, Hispanophonie et Lusophonie).

49. Groupe de travail sur l'aménagement linguistique des trois espaces linguistiques (Francophonie, Hispanophonie et Lusophonie).

gouvernent le fonctionnement d'Internet et des ordinateurs qui permettent d'y accéder. Plusieurs de ces normes et de ces standards ont des effets importants sur la place du français et des langues partenaires dans le secteur des technologies de l'information. Il convient donc d'infléchir leur développement et de surveiller leur évolution.

Les gouvernements de la Francophonie, l'AIF et l'AUF devraient veiller à ce que le français soit pris en compte lors des activités internationales menées pour évaluer les prototypes et les produits mis au point dans le secteur du TIL. Ils devraient aussi appuyer la participation des chercheurs et des industriels francophones à ces activités.

Comme le relève L'Agence pour l'évaluation et la distribution des ressources linguistiques (ELDA), « l'évaluation est un facteur de premier plan dans le domaine de l'ingénierie linguistique, [...] et ce,] à différents niveaux. L'évaluation permet aux équipes de [...] recherche et développement] de valider leurs recherches et d'assurer ainsi le développement des systèmes, en constante progression. L'évaluation permet aussi [...] de déterminer] les créneaux vers lesquels il est judicieux de s'orienter pour développer de nouvelles technologies, aboutissant à leur mise sur le marché. L'évaluation permet également aux agences ayant financé des projets d'en mesurer, d'une certaine façon, le [...] rendement de l'investissement grâce à la mise en œuvre de ces projets ». Malheureusement, il arrive que le français ne soit pas pris en compte lors des activités d'évaluation menées en Europe ou ailleurs. Il arrive aussi que, faute de ressources, les acteurs du domaine ne puissent présenter leurs systèmes aux tables d'évaluation.

Les gouvernements de la Francophonie, l'AIF et l'AUF devraient assurer la croissance du réseau d'experts dans le secteur de l'informatisation des langues en :

- **instaurant un programme francophone de stages en entreprise et d'octroi de bourses d'études spécialisées aux étudiants qui mènent des travaux de recherche dans le secteur de l'informatisation des langues ;**
- **accordant aux universités et aux laboratoires de recherche des contrats de recherche dans le secteur du TIL, notamment à cause des effets de tels**

contrats sur la formation et les choix de carrière futurs des étudiants.

Il ne fait pas de doute que la principale façon pour la Francophonie de conquérir un avantage concurrentiel dans le domaine du TIL sera le développement d'une main-d'œuvre de qualité. Nous avons vu que les entreprises et les centres de recherche francophones ont souvent de la difficulté à recruter des ressources humaines compétentes. Plusieurs raisons expliquent que le bassin de chercheurs sur lequel le développement du secteur peut s'appuyer soit trop petit. Parmi celles-ci, on retrouve notamment : la faible connaissance que les étudiants ont du secteur et des possibilités d'y faire carrière ; le manque d'intérêt des professeurs-chercheurs pour le domaine ; la rareté des programmes de cycle en TIL ou, tout au moins, la difficulté de recevoir une formation cohérente et complète dans le champ du TIL.

Les gouvernements de la Francophonie sont conviés à soutenir financièrement les centres de recherche de leurs pays désireux de participer à des projets de recherche internationaux dans le secteur du TIL, en particulier dans le domaine des aides à la traduction et dans celui de la diffusion électronique multilingue d'information scientifique et technique, en français et dans les langues partenaires.

Certains acteurs, comme l'Union européenne, affichent de l'ouverture devant la participation de centres de recherche étrangers à des projets de recherche communautaires dans le secteur du TIL, mais les centres doivent investir des fonds dans les projets.

Les gouvernements de la Francophonie, l'AIF et l'AUF devraient sensibiliser les partenaires de la Francophonie à l'importance d'accorder une attention particulière au français lors de la mise sur pied de grands projets multilingues en TIL.

Plusieurs grands projets nationaux et internationaux existent actuellement dans le champ du TIL. Les gouvernements et les organisations de la Francophonie devraient chercher à nouer des liens avec les organisateurs de ces projets, de façon à y favoriser, le cas échéant, la participation de chercheurs francophones et la réalisation de travaux sur la langue française. Les participants francophones à ces projets veilleront à ce que des ponts

soient construits entre le français et l'ensemble des autres grandes langues.

Les gouvernements de la Francophonie, l'AIF et l'AUF sont conviés à coordonner les activités de veille virtuelle sur les questions d'informatisation du français et des langues partenaires.

Pour avoir du succès, les intervenants du secteur du TIL doivent disposer de renseignements sur les tendances susceptibles d'affecter leur positionnement. Par exemple, ils doivent connaître les percées réalisées dans les centres de recherche francophones et étrangers, la demande, les réseaux de commercialisation, etc. Ces renseignements circulent souvent mal d'un pays à l'autre. Il conviendrait d'aider les acteurs francophones qui font de la veille dans le domaine à collaborer et à échanger de l'information les uns avec les autres.

L'AIF et l'AUF devraient rééditer tous les deux ans, dans un pays différent de la Francophonie, la Vitrine-forum sur l'informatisation des langues pour favoriser la cartographie du secteur et le partage de renseignements entre participants. Ils devraient également évaluer les effets de ces manifestations sur les acteurs du domaine.

Les intervenants du secteur émettent le souhait de pouvoir participer régulièrement à des rencontres réunissant un éventail d'acteurs du milieu aussi large que possible. Ils devraient pouvoir compter sur les gouvernements et les organisations de la Francophonie pour soutenir la tenue de telles rencontres, qui auront notamment pour conséquence de favoriser la circulation d'information entre chercheurs et industriels, et l'établissement de partenariats.

En bref – Nouvelles du Rifal

1. Suivi de l'Assemblée générale du Rifal

La troisième assemblée générale du Rifal, tenue en juin 2004 au siège de l'Agence intergouvernementale de la Francophonie, a été particulièrement fertile en nouveautés pour le réseau et pour la coopération internationale en matière d'aménagement linguistique.

Au moment de la fin du mandat du premier Bureau du Rifal, les participants ont confié le secrétariat général du réseau à l'agence elle-même, alors qu'auparavant, cette fonction était assurée par l'un des membres du réseau désigné parmi les membres du réseau par l'assemblée générale. Il convient de rappeler que le Rifal fait partie du dispositif que l'agence a créé pour mettre en oeuvre sa programmation en matière de langues. Dans ce dispositif, lequel les interventions en matière de langues sont structurées en réseaux, regroupés autour du Conseil international francophone des langues (CIFLA). Ces réseaux sont les suivants :

- le Réseau international du français dans le monde (RIFRAM);
- le Réseau international francophone d'aménagement linguistique (Rifal);
- le Réseau international des langues africaines et créoles (RILAC);
- le Réseau international des littératures francophones (RILIF).

Chacun de ces réseaux trouve sa spécificité dans le partage des actions prévues dans la programmation de l'agence en matière de langues.

Les nouveaux membres du Bureau du Rifal sont les représentants des organismes suivants :

- Office québécois de la langue française;
- Délégation générale à la langue française et aux langues de France;
- Service de la langue française - Communauté française de Belgique;
- Services linguistiques centraux - Confédération helvétique;
- Institut de linguistique appliquée - République centrafricaine
- Centre de linguistique appliquée de Dakar – Module collectif d'Afrique de l'Ouest;
- Institut d'études et de recherches pour l'arabisation - Maroc.

2. Activités du Rifal

En 2004, l'Agence intergouvernementale de la Francophonie a été en mesure de soutenir l'ensemble des actions proposées par les membres de l'assemblée générale du Rifal. Ces actions s'inscrivent dans la continuité des travaux du Rifal en matière d'aménagement linguistique.

Ces travaux sont les suivants :

- Poursuite de la publication de la revue *Les cahiers du Rifal*

Cette revue, tirée à 3000 exemplaires, est diffusée gratuitement et est accessible sur Internet, sur le site du réseau : www.rifal.org. Le présent numéro est constitué des actes de la *Vitrine-forum sur le traitement informatique des langues* qui s'est tenue à Montréal en 2003. Le numéro suivant portera sur le thème « Aménagement linguistique et diversité ». Un appel à contributions est en cours.

- Formation pratique

Un nouveau cycle de stages de formation se tiendra en 2004 sur le double thème de la « terminologie socioculturelle » et celui de « gestion électronique et partage des documents ». Ces stages visent les pays suivants : Burkina Faso, République centrafricaine, Guinée, Haïti, Madagascar, Mali, Maroc, Mauritanie, Niger, Sénégal et Roumanie.

- Poursuite du développement de l'Inventaire francophone du traitement informatique des langues

Cet inventaire répertorie les produits, services, organismes spécialisés en matière de traitement informatique du français et des langues partenaires. Il est accessible sur Internet à l'adresse suivante :

cfwb.otil.org/index.html. Cet inventaire permanent est destiné à être alimenté par les membres du Rifal, de même que par les différents acteurs de ce domaine (chercheurs, producteurs, diffuseurs et utilisateurs des produits et services).

- Transformation de l'*Inventaire des travaux terminologiques en ligne*

Cet inventaire, jadis créé par le Rint (ancien Réseau international de néologie et de terminologie), vise à répertorier les sites Internet qui diffusent de la terminologie en langue française. Il s'agit de transformer ce site afin qu'il donne accès directement aux données terminologiques elles-mêmes, et non pas seulement à l'adresse URL des sites. L'inventaire deviendra ainsi un vaste corpus à

la disposition des francophones donnant accès aux contenus terminologiques de l'Internet.

- Édition et réédition d'ouvrages terminologiques

L'Agence intergouvernementale de la Francophonie soutiendra cette année l'élaboration et la publication des ouvrages suivants :

- *Lexique de l'Internet* (Maroc);
- *Lexique du prêt-à-porter* (Maroc);
- *Lexique de l'urbanisme* (Centrafrique);
- *Vocabulaire de la bureautique* (Centrafrique).

De plus, les lexiques et vocabulaires publiés sous label Rint dans les années 1990 seront réédités. L'ensemble de ces travaux servira d'amorce au développement de banques de terminologie qui seront implantées dans les pays du Sud membres du Rifal en 2005.

- Projet de diffusion sociale de la terminologie à Madagascar

Un projet de diffusion sociale de la terminologie aux 13-25 ans à Madagascar sera réalisé. Il s'agira de produire un vocabulaire bilingue français-malgache / malgache-français comportant deux volets : (1) environnement et (2) développement social, aspects socio-sanitaires, relations avec l'administration.

- Le Rifal au II^e Sommet mondial de la terminologie

Le Rifal sera présent au II^e Sommet mondial de la terminologie qui se tiendra à Barcelone les 26 et 27 novembre 2004.

Ce sommet fait suite à celui qui s'est tenu à Bruxelles en novembre 2001 et au cours duquel une importante déclaration avait été adoptée. Intitulée « Pour une coopération terminologique internationale », cette déclaration, signée conjointement par un grand nombre de réseaux et d'organisations, consiste en fait en un vaste programme destiné à promouvoir la terminologie comme champ d'expérience capital dans la nouvelle économie du savoir et de la communication. Au cours du sommet, l'état d'avancement de la mise en oeuvre de la *Déclaration de Bruxelles* fera l'objet de discussions au cours d'une table ronde présidée par le Rifal.

*Louis-Jean Rousseau,
Secrétaire général du Rifal,
Direction des langues et de l'écrit,
Agence intergouvernementale de la Francophonie,
Paris.*

Les Cahiers du Rifal

Les *Cahiers du Rifal*, revue scientifique du Réseau international francophone d'aménagement linguistique (Rifal), paraissent chaque année à la suite d'un appel à contributions. La thématique abordée est fixée par le Réseau et la direction scientifique du numéro est confiée à une ou plusieurs personnalités actives dans le domaine abordé. La revue publie également des actes de colloque et de séminaires.

Les articles destinés à paraître dans les *Cahiers du Rifal* doivent être soumis au(x) directeurs scientifiques du numéro à paraître.

Direction scientifique de ce numéro:
Réseau international francophone d'aménagement linguistique
(Secrétariat général).

Comité scientifique:
Daniel Blampain (Institut supérieur de traducteurs et interprètes, Bruxelles),
Marcel Diki-Kidiri (Centre national de la recherche scientifique, Paris),
Abdelkader Fassi-Fehri (Institut d'études et de recherches pour l'arabisation, Rabat),
John Humbley (Université de Paris VII),
Chérif Mbodj (Centre de linguistique appliquée de Dakar),
Marie-Claude L'Homme (Université de Montréal),
Michèle Valiquette (Bureau de la traduction, Hull),
Eric Wehrli (Université de Genève).

Comité de rédaction:
Daniel Blampain (Institut supérieur de traducteurs et interprètes),
Martine Garsou (Communauté française de Belgique),
Louis-Jean-Rousseau (Agence intergouvernementale de la Francophonie),
Marc Van Campenhoudt (Institut supérieur de traducteurs et interprètes).

Secrétaire de rédaction:
Marc Van Campenhoudt,
Institut supérieur de traducteurs et interprètes, 34 rue Joseph Hazard,
B-1180 Bruxelles, Belgique,
marc.van.campenhoudt@termisti.be.

Éditeur responsable:
Martine Garsou, Communauté française de Belgique, 44 bd Léopold II,
B-1080 Bruxelles, Belgique.

Conception:
Patrice Junius, *Alternatives Théâtrales*.

Conseillers techniques:
Serge Paulus et Christian Chanard.

Photocomposition et impression:
Édition et imprimerie.

Présentation des tapuscrits:
Les normes de présentation des articles ainsi que la feuille de style associée peuvent être téléchargées sur le site internet du Rifal: www.rifal.org.

Descriptif bibliographique:
Rifal, *Actes de la vitrine-forum sur l'informatisation des langues* (Montréal, 13 et 14 juin 2003), dans *Cahiers du Rifal*, n° 24, décembre 2004, Bruxelles, Agence intergouvernementale de la Francophonie et Communauté française de Belgique, ISSN: 1015-5716.

Les éditeurs

Agence intergouvernementale de la Francophonie

Direction des langues et de l'écrit
13, quai André Citroën
75015 Paris
France
<http://agence.francophonie.org>

Communauté française de Belgique

Service de la langue française
44 bd Léopold II
B-1080 Bruxelles
<http://www.cfwb.be/franca>
et
Commissariat général aux relations internationales
2, Place Saintelette
B-1080 Bruxelles
Belgique
<http://www.wbri.be>

Liste des numéros parus sous le titre *Terminologies nouvelles*

- Consultables à partir du n° 14
à l'adresse www.rifal.org
- 1, mai 1989: *Le Rint: objectifs et perspectives*
 - 2, décembre 1989: *La formation en terminologie*
 - 3, juin 1990: *Harmonisation des méthodes en terminologie* (actes des séminaires de Talence et de Hull)
 - 4, décembre 1990: *Numéro général*
 - 5, juin 1991: *Terminologie et informatique*
 - 6, décembre 1991: *Terminologie et développement I* (actes du séminaire de Rabat)
 - 7, juin 1992: *Numéro général*
 - 8, décembre 1992: *Terminologie et environnement*
 - 9, juin 1993: *Terminologie et développement II* (actes du séminaire de Cotonou)
 - 10, décembre 1993: *Pbraséologie* (actes du séminaire de Hull)
 - 11, juin 1994: *Numéro général*
 - 12, décembre 1994: *Implantation des termes officiels* (actes du séminaire de Rouen)
 - 13, juin 1995: *Terminologie et entreprise*
 - 14, décembre 1995: *Numéro général*
 - 15, décembre 1996: *Banques de terminologie* (actes de la table ronde de Québec)
 - 16, juin 1997: *Enquêtes terminologiques*
 - 17, décembre 1997: *Terminologie et formation*
 - 18, juin 1998: *Terminotique et documentation*
 - 19, décembre 1998 - juin 1999: *Terminologie et intelligence artificielle* (actes du colloque de Nantes)
 - 20, décembre 1999: *De nouveaux outils pour la néologie*
 - 21, décembre 2000: *Terminologie et diversité culturelle*

Liste des numéros parus sous le titre *Cahiers du Rifal*

- Consultables à l'adresse
www.rifal.org
- 22, décembre 2001: *Développement linguistique: enjeux et perspectives*
 - 23, novembre 2003: *Le traitement informatique des langues africaines*
 - 24, décembre 2004: *Actes de la vitrine-forum sur l'informatisation des langues*

Membres du Rifal

Afrique de l'Ouest (Bénin, Guinée, Mali, Mauritanie, Niger, Sénégal)

Coordination:

Centre de linguistique appliquée
Université Cheikh Anta Diop
Dakar-Fann, Sénégal
Téléphone: 221.825.01.26
Télécopie: 221.824.71.47
Courriel: chembodj@ucad.refer.sn

Canada

Bureau de la traduction
Travaux publics et Services
gouvernementaux Canada
Immeuble Richelieu
975, bd Saint-Joseph, 5^e étage
Gatineau (Québec) K1A 0S5, Canada
Téléphone: 1.819.997.33.00
Télécopie: 1.819.997.19.93
Courriel: gabriel.huard@pwgsc.gc.ca

Communauté française de Belgique

Service de la langue française
Ministère de la Communauté française
44, boulevard Léopold II
1080 Bruxelles, Belgique
Téléphone: 32 2.413.22.95
Télécopie: 32 2.413.28.94
Courriel: martine.garsou@cfwb.be

France

Délégation générale à la langue française
6, rue des Pyramides
75001 Paris, France
Téléphone: 33.1.40.15.36.70
Télécopie: 33.1.40.15.36.76
Courriel: bernard.cerquiglioni@culture.fr

Haïti

Faculté de linguistique appliquée
Université d'État d'Haïti
B.P. 668
Rue Dufort n° 38
Port-au-Prince, Haïti
Téléphone: 509 245.12.33
Télécopie: 509 245.91.53
Courriel: Ssylvestre.fla@ht.refer.org

Madagascar

Centre des langues de l'Académie
Académie malgache
B.P. 6217
Antananarivo 101, Madagascar
Téléphone: 261.20.22.624.39
Télécopie: 261.20.22.313.61
Courriel: tenymalagasy@dts.mg

Maroc

Institut d'études et de recherches
pour l'arabisation
B.P. 6216, Instituts
Rabat, Maroc
Téléphone: 212.77.730.09
Télécopie: 212.77.720.65
Courriel: fassi@atlasnet.net.ma

Québec

Office québécois de la langue française
125, rue Sherbrooke Ouest
Montréal (Québec) H2X 1X4, Canada
Téléphone: 1.514.873.07.97
Télécopie: 1.514.873.34.88
Courriel: nicole.rene@oqlf.gouv.qc.ca

République centrafricaine

Institut de linguistique appliquée
Université de Bangui
B.P. 1450
Bangui, République centrafricaine
Téléphone: 236.61.67.59
Télécopie: 236.61.78.90
Courriel: ila@intnet.cf

République démocratique du Congo

Centre de linguistique théorique
et appliquée
B.P. 4956
Kinshasa/Gombé,
République démocratique du Congo
Téléphone: 243.12.62.910
Télécopie: 243.12.46.197
Courriel: anyembwe@yahoo.fr

Roumanie

Institutul de lingvistica Iorgu Iordan
Casa Academiei
Calea 13 Septembrie nr. 13
76100 Bucuresti, Roumanie
Téléphone: 40.1.410.34.09
Télécopie: 40.1.410.34.10
Courriel: ioanar@fx.ro

Suisse

Chancellerie de la Confédération suisse
Services linguistiques centraux
Section française
CH-3003 Berne, Suisse
Téléphone: 41 31. 324. 11. 49
Télécopie: 41 31. 324. 11. 04
Courriel: anne.marie.gendron@bk.admin.ch

Union latine

Union latine
131, rue du Bac
75340 Paris Cédex 07, France
Téléphone: 33.1.45.49.60.60
Télécopie: 33.1.45.44.45.97
Courriel: d.prado@unilat.org

Bulletin d'abonnement

Je soussigné souhaite recevoir gratuitement les *Cahiers du Rifal*.

Nom: _____

Entreprise, organisme: _____

Fonction: _____

Ce bulletin d'abonnement est à adresser à l'organisme représentant votre pays ou, à défaut, au secrétariat général du Rifal.

Adresse: _____



Notes

A series of horizontal dotted lines for writing notes.

Actes de la vitrine-forum sur l'informatisation des langues

(Montréal, 13 et 14 juin 2003)

Réseau international francophone d'aménagement linguistique
(Secrétariat général)

Stéphane Chaudiron :

Le traitement informatique des langues en France

Réjean Roy :

Le traitement informatique des langues au Québec

André Manseau :

Le traitement informatique des langues au Canada

Sylvie Wallez :

*Le traitement informatique des langues dans la
communauté Wallonie-Bruxelles*

Sandra Manzi :

Le traitement informatique des langues en Suisse

Marcel Diki-Kidiri :

Les langues africaines sur la Toile

Jean-Claude Guédon :

Vers une nouvelle écologie mondiale des langues

Marie-Claude L'Homme :

*Formation et recherche dans le domaine du traitement
automatique des langues en contexte universitaire*

Mokhtar Ben Henda :

*Traitement informatisé des langues, langues
nationales et partenariat francophone*

Marcel Diki-Kidiri :

*La recherche sur l'informatisation des langues
africaines*

Robert Chaudenson :

La graphisation des langues africaines

Claude Rigault :

*Les enjeux industriels dans le secteur du TIL :
l'expérience d'un spin-off universitaire*

Jean-Patrice Isaac :

*Le secteur du traitement informatique des langues :
un marché frileux*

Stéphane Chaudiron :

L'avancement des travaux est affaire d'intégration

Velina Slavova :

*Sciences cognitives et traitement informatique des
langues*

Guy Lapalme :

*Les voies d'amélioration des outils de traitement de
l'écrit*

Laurence Devillers :

Les défis en matière de traitement de la parole



Coédité par :
L'Agence intergouvernementale
de la francophonie
et la Communauté française
de Belgique

Secrétariat général du Rifal :
Agence intergouvernementale
de la Francophonie

Direction des langues et
de l'écrit
13, quai André Citroën
75015 Paris
France

Site internet du Rifal :
www.rifal.org