LES VOIES ACTUELLES DE LA TAO AU CANADA

par **Elliott Macklovitch**
et **Pierre Isabelle**

Chercheurs
Centre canadien de recherche
sur l'informatisation du travail
1575 boul. Chomedey
Laval, Québec
Canada H7V 2X2

1. Bref historique: les débuts

L'histoire de la traduction assistée par ordinateur (TAO) au Canada remonte à 1965, date à laquelle le Conseil national de recherche décide de subventionner des recherches universitaires dans ce domaine à Saskatoon et à Montréal. Ces projets naissants allaient heureusement être épargnés lorsque, l'année suivante, la parution du tristement célèbre rapport ALPAC dévasta le monde de la TAO. C'est pour des raisons différentes qu'en 1968 le projet de l'Université de Saskatoon fut abandonné. Cette même année, Alain Colmerauer assuma la direction du groupe TAUM (Traduction automatique de l'Université de Montréal), et mis bientôt au point son logiciel SYSTÈMES-Q, qui rendit possible l'assemblage d'un premier prototype complet de traduction automatique fondé sur les principes des systèmes de deuxième génération.

En 1973, le Secrétariat d'État, dont le mandat est d'assurer les services de traduction pour tout le gouvernement fédéral, devient le principal bailleur de fonds du groupe TAUM, et en 1975, le groupe réalise pour son compte un prototype de traduction automatique conçu spécialement pour traduire les prévisions météorologiques. Livré dès l'année suivante, TAUM-MÉTÉO allait devenir sans contredit la réussite la plus spectaculaire de toute l'histoire de la TA.

Fort de ce succès, TAUM accepte en 1977 un deuxième contrat du Secrétariat d'État, beaucoup plus ambitieux celui-là, pour le développement d'un système de traduction des manuels d'entretien de l'avion CP-140. Malheureusement, les difficultés techniques et organisationnelles de la tâche avaient été sous-estimées. Après trois ans d'efforts, TAUM-AVIATION commencait à peine à décoller lorsque le Conseil du trésor commanda une évaluation indépendante qui allait conclure que le système, considéré par certains comme le plus évolué de son époque, était encore trop limité pour permettre une production rentable à court terme. Peu après, le bailleur de fonds retira son soutien et, en 1981, le groupe TAUM fut formellement démembré.

2. . . . après TAUM

La fin du groupe TAUM a porté un dur coup à la TAO au Canada, surtout en ce qui concerne la recherche et le développement. Au même moment où on assistait dans plusieurs pays à un renouveau d'intérêt pour la TAO, le Canada entreprenait une période de réflexion. Toute activité dans le domaine n'en cessait pas pour autant. Seulement, de chef de file qu'il était dans *le développement* de ses propres systèmes de TA, le Canada devenait plutôt un expert dans le domaine de l'*essai* et de l'*exploitation* de systèmes commerciaux développés par d'autres.

Le Bureau de la traduction du Secrétariat d'État constitue le plus grand service de traduction au pays. À la recherche de moyens de mieux faire face à une charge de travail qui approche les 300 millions de mots par année, il fut amené à faire l'essai d'à peu près tous les systèmes de TA sur le marché. Jusqu'à ce jour, aucun système n'a pu satisfaire ses exigences élevées de qualité. Depuis 1987, le Bureau fait l'essai du système LOGOS (dans sa version anglais-français), qui est maintenant installé dans quatre sites. Aucune évaluation officielle n'a encore été rendue publique.

Le Bureau de la traduction n'est pas le seul grand service à essayer LOGOS. Récemment, au Nouveau-Brunswick, la firme qui a obtenu le contrat de la Défense pour la construction de nouvelles frégates a créé une filiale nommée Lexi-tech dont le premier mandat est d'assurer la traduction de l'énorme volume de documentation engendré par ce programme de construction. Persuadée que les moyens traditionnels ne suffiraient pas à la tâche, la direction de cette société décida de miser sur l'informatisation: lecture optique pour la saisie des textes sources, système d'éditique pour la mise en forme des documents, et LOGOS pour la traduction automatique. La rumeur veut toutefois que Lexi-tech soit maintenant à la recherche d'alternatives à LOGOS.

Un autre bureau de services, SOCATRA de Montréal, dit exploiter un système de TA de sa propre conception, pour offrir à ses clients soit des traductions brutes, soit des traductions révisées. D'après son président, M. Claude Richaud, la commercialisation de son système XLT sera lancé au prochain Salon International de l'Informatique, qui se tiendra à Montréal au mois de juin.

Ailleurs dans le secteur privé, certains consultants continuent à offrir des services en TAO. Laurent Bourbeau (autrefois membre du groupe TAUM) développe pour le compte du Secrétariat d'État une méthodologie d'évaluation des systèmes de TA. John Chandioux (un autre ancien du groupe TAUM) dirige une société qui développe diverses aides à la rédaction, notamment des détecteurs de fautes d'orthographe ou grammaticales, ainsi que des systèmes spécialisées d'aide à la traduction capables par exemple de repérer des passages figés dans des contrats ou des listes de pièces. M. Chandioux est la personne qui a transféré le système MÉTÉO sur micro-ordinateur en 1984 et il continue d'assurer son entretien.

Entretemps, la société CoGenTex (dirigée par le Dr. Richard Kittredge, un ancien directeur du groupe TAUM) a démontré la faisabilité d'une approche susceptible dans certains cas de constituer une alternative à la traduction: la *rédaction bilingue automatique*. Cette approche semble tout indiquée dans le cas de textes confinés à la description de l'évolution d'indices numériques. L'un des premiers prototypes mis au point par CoGenTex effectue la synthèse bilingue de... prévisions météorologiques! On doit semble-t-il se résoudre à voir disparaître d'ici peu l'un des rares systèmes de TA qui fonctionne bien.

Dans les universités canadiennes, les cours sur la TAO suscitent de plus en plus d'intérêt, surtout dans les écoles de traduction où on essaie de préparer les futurs practiciens aux outils de l'avenir. C'est le cas à l'École de traduction de l'Université d'Ottawa, où les professeurs Brian Harris et Ingrid Mayer donnent des cours dans ce domaine, et à l'Université de Montréal où l'on retrouve le professeur Richard Kittredge. Il y a d'autres professeurs qui s'intéressent à la TA au pays, par exemple Chrysianne DiMarco à Waterloo et Veronica Dahl à Simon Fraser, mais leurs efforts demeurent isolés. L'Université de Moncton a récemment annoncé un projet d'envergure: la mise sur pied d'un Centre de recherche et de développement en traductique (CRDT). Pour le moment il s'agit toujours d'un projet, et l'Université vient d'accorder un contrat pour une étude sur les orientations éventuelles de son CRDT.

3. Le CCRIT

La TAO, il faut bien se le dire, n'est pas un domaine où un ou deux individus peuvent réussir des percées, chacun de son côté. La réalisation de grammaires et de dictionnaires informatisés, même pour une seule langue, exige un fort coefficient de main d'oeuvre. À plus forte raison les projets de traduction automatique, qui impliquent par définition au moins deux langues. Depuis le démantèlement du groupe TAUM, le Canada n'avait plus d'*équipe* de R&D en TAO. C'est pourquoi le Ministère des communications a décidé en 1985, lors de la création de son Centre canadien de recherche sur l'informatisation du travail, de mettre sur pied un nouveau groupe de R&D en traduction assistée. Actuellement, ce groupe comprend quatre chercheurs permanents (deux linguistes et deux informaticiens), deux stagiaires universitaires, et quelques étudiants qui travaillent à temps partiel. Considérant l'ampleur du programme visé, ces ressources demeurent plus que modestes.

Le programme de TAO du CCRIT se veut une réponse cohérente et réaliste à la situation actuelle en TAO, une réponse qui fasse écho aux leçons que l'on doit tirer de l'histoire tumultueuse du domaine. La solution globale au problème de l'automatisation de la traduction, à savoir la machine à traduire universelle, n'est guère plus à notre portée aujourd'hui qu'elle ne l'était il y quarante ans. La production d'une traduction de bonne qualité présuppose une intelligence du texte qui aille au-delà de son contenu linguistique explicite, pour faire intervenir pour les capacités cognitives globales du locuteur humain, en particulier sa capacité à raisonner à partir de sens commun et de savoir encyclopédique. Malgré les progrès récents de l'intelligence artificielle, il ne demeure possible de simuler un niveau raisonnable de compréhension qu'à l'intérieur de micro-univers particulièrements étroits. En conséquence, la TA de haute qualité n'est réalisable que dans des circonstances plus ou moins exceptionnelles.

Puisqu'il ne saurait pour le moment être question d'une solution optimale aux problèmes de la traduction (la machine à traduire universelle), on n'a d'autres choix que de chercher des *palliatifs* susceptibles de venir à la rescousse des traducteurs. Dans cette perspective, deux des trois volets de notre programme visent le développement de systèmes fondés sur des techniques précises de *sous-optimisation.* Le troisième volet est consacré à des recherches à plus long terme visant l'émergence de systèmes fondés sur des techniques d'avant-garde en intelligence artificielle.

Volet I: Poste de travail du traducteur (PTT)

Les seuls systèmes d'applicabilité générale qui soient présentement réalisables en traduction sont ceux qui laissent au traducteur humain l'entière initiative du processus de traduction, mettant à sa disposition un ensemble intégré d'outils susceptibles de l'aider dans sa tâche.

Cette tâche est complexe, et elle possède de multiples facettes. En conséquence, un poste de travail spécialisé en traduction doit réunir et intégrer de façon cohérente une vaste gamme d'outils. Le CCRIT a déjà réalisé un premier prototype du PTT qui intègre sur un ordinateur personnel (de type PC/AT) un ensemble de logiciels commerciaux. Dans cet environnement, le traducteur dispose d'outils destinés à faciliter:

- la télécommunication de textes;
- la recherche lexicographique et terminologique, y compris
  - un 'cardex' informatisé
  - des dictionnaires naturels informatisés

- l'accès à des banques terminologiques à distance
- des outils d'établissement de concordances;

- diverses aides à la rédaction, y compris
  - un traitement de texte multilingue
  - un vérificateur d'orthographe
  - un 'conjugueur';
- la comparaison de deux versions d'un même texte; et,
- la conversion de fichiers de traitement de texte d'un format à un autre.

Ce prototype modeste mais fonctionnel est actuellement à l'essai au Bureau de la traduction. Les résultats de son évaluation seront connus en octobre 1990. Entretemps, le CCRIT commence à élaborer une version plus évoluée de son PTT.

Volet II: Traduction automatique spécialisée

Le système TAUM-MÉTÉO a clairement démontré que l'approche dite par sous-langue peut obtenir des résultats spectaculaires lorsqu'elle est appliquée à un domaine dont les textes affichent les caractéristiques appropriées. (L'utilisation du système SMART pour la traduction d'offres d'emploi au Ministère de l'emploi et de l'immigration - même s'il traite un volume beaucoup moins important que MÉTÉO - est un autre exemple du bien-fondé de cette approche.) Il nous apparaît extrêmement important de transposer le succès des techniques concernées à un nombre aussi grand que possible de domaines. Alors que le PTT prévu au volet I opte pour la généralité du champ d'application au détriment du degré d'automatisation, la TA spécialisée effectue la démarche inverse. Les deux approches sont parfaitement complémentaires puisque, à moyen terme, le PTT fournira un environnement de travail idéal pour l'exploitation de modules de TA spécialisés.

Le succès d'une entreprise de TA spécialisée est intimement lié à des facteurs comme les suivants:

- simplicité linguistique du domaine choisi
- volume important et constant de textes à traduire
- tolérance des utilisateurs envers un style plutôt rigide

Le premier point est évidemment le plus difficile à mesurer, et pour y parvenir, le CCRIT a participé au développement d'une méthodologie d'évaluation de la complexité relative des sous-langues. Pour chaque sous-langue examinée, cette méthodologie prend en compte des facteurs comme: a) une estimation de la taille du vocabulaire; b) un examen des particularités syntaxiques; et c) un examen du degré de fermeture de l'univers sémantique.

Un certain nombre de sous-langues ont déjà fait l'objet d'études détaillées. Bien que les recherches n'aient pas encore permis d'identifier de cibles idéales pour des applications, nous demeurons optimistes quant aux chances d'y parvenir prochainement. Entretemps, nous étudions, de concert avec CoGenTex, la possibilité d'appliquer l'approche génération bilingue aux rapports publiés par Statistique Canada sur la population active et le commerce de détail.

Volet III: Vers des systèmes de traduction plus intelligents

Dans la mesure où les volets précédents de notre programme atteindront leurs objectifs, les milieux de la traduction au Canada en tireront des bénéfices tangibles. Toutefois, l'ampleur du problème est telle que les produits présentement réalisables (PTT et systèmes de TA spécialisés) ne sauront suffire.

Pour assurer l'avenir, il nous paraît vital d'oeuvrer dès maintenant à l'émergence de systèmes plus puissants, plus 'intelligents'. Si les recherches en traitement des langues naturelles et en intelligence artificielle ont reçu un appui considérable depuis quelques années, il demeure que ces recherches ont rarement été focalisées spécifiquement sur le problème de la traduction, même si ce problème constitue l'un des meilleurs bancs d'essai qui soit.

Le troisième volet de notre programme vise justement à appliquer les plus récentes techniques de l'intelligence artificielle et du traitement des langues naturelles au problème de la traduction. Pour illustrer l'état d'avancement de nos recherches, nous avons construit un système prototype, baptisé CRITTER. Implanté en langage PROLOG sur des stations SUN, CRITTER est un système expérimental qui fait preuve d'un niveau très élevé de maîtrise linguistique. Par exemple, il est capable de reconnaître et de produire des paraphrases lexicales et structurales. En outre, et il s'agit là d'une première, CRITTER est entièrement *réversible*: il peut traduire aussi bien du français vers l'anglais que de l'anglais vers le français, et cela en utilisant exactement les mêmes grammaires et dictionnaires.

La technologie sur laquelle le système CRITTER s'appuie est en fait très largement indépendante du domaine d'application, voire même des langues particulières entre lesquelles la traduction s'effectue. Toutefois, il nous paraît certain que les applications à des sous-langues simples sont celles qui présentent les meilleures chances de succès. En vue de mettre la technologie à l'essai, nous avons incorporé à CRITTER des modules spécialisés pour la traduction de rapports de marchés agricoles. Les résultats obtenus confirment pleinement le potentiel de la technologie

CRITTER.

Le prototype actuel est en outre doté d'une élégante interface de démonstration qui tire parti des possibilités d'affichage des stations SUN. Tout en simulant l'embryon d'un PTT de l'avenir, cette interface permet de mettre en valeur les capacités de CRITTER à effectuer les manipulations complexes de structures linguistiques qui sont requises en vue du calcul d'une traduction adéquate.

Finalement, le CCRIT a également commencé à explorer les possibilités d'appliquer la technologie vocale au processus de la traduction. Dans ce but, on a développé pour le système CRITTER une autre interface qui permet cette fois de dicter le texte de départ et d'effectuer une synthèse vocale de sa traduction machine, et ce, dans les deux directions entre nos langues officielles.

4. Conclusion

La traduction met en jeu les capacités cognitives globales de l'humain. Pour cette raison, la machine à traduire universelle demeure présentement hors d'atteinte. Il est quand même possible de venir en aide aux traducteurs, dont le fardeau ne cesse d'augmenter, pour autant que l'on adopte des stratégies de sous-optimisation adéquates. Face au regain d'intérêt à l'échelle mondiale pour l'automatisation de la traduction, le Canada ne peut se permettre de rester inactif.

Le programme du CCRIT vise à redonner au pays son rôle traditionnel de chef de file en traduction automatisée. À court et à moyen terme, ce programme adopte une approche réaliste, fondée sur des stratégies éprouvées de sous-optimisation: poste de travail spécialisé en traduction, et systèmes orientés vers des sous-langues particulières. À plus long terme, notre programe vise à mettre l'intelligence artificielle au service de la traduction.

MOTS THÈMES

- **traduction assistée par ordinateur (TAO)**
- **machine-aided translation (MAT)**
- **Centre canadien de recherche sur l'informatisation du travail (CCRIT)**
- **Canadian Workplace Automation Research Center (CWARC)**
- **poste de travail du traducteur (PTT)**
- **CRITTER**
- **TAUM**