

John Chandioux

Histoire de la traduction automatique au Canada

Introduction

À **une** époque où le législateur canadien entend permettre aux employés fédéraux de travailler indifféremment en anglais ou en français, il n'est pas possible de faire abstraction des possibilités de la traduction automatique. Si l'on pense que la simple traduction de la documentation technique de l'Aurora, dont les Forces armées canadiennes viennent de faire l'acquisition, représente le travail annuel de plusieurs dizaines de traducteurs techniques spécialisés en aéronautique, on peut même dire que c'est parfois la seule solution.

L'année 1976 restera dans les annales de la traduction comme celle où la traduction automatique, souvent redoutée et presque toujours mal comprise, a acquis droit de cité. Les faits suivants parlent d'eux mêmes :

- Mars 1976 : attribution au projet TAUM de l'université de Montréal d'un contrat pour le développement d'un système de traduction automatique de manuels techniques par le Bureau des traductions ;
- Mai 1976 : livraison par le projet TAUM du système METEO pour la traduction automatique des prévisions météorologiques destinées au grand public ;
- Septembre 1976 : création à l'Université d'Ottawa d'une option traduction automatique en maîtrise de traduction ;
- Octobre 1976 : tenue d'une table ronde sur la traduction automatique dans le cadre du congrès de l'ATIO.

Mais voyons plutôt comment on en est arrivé là.

Les débuts de la traduction automatique

Bien que les soviétiques revendiquent la paternité de la traduction automatique pour un système mécanique qui semble n'avoir jamais été réalisé, on s'accorde généralement à reconnaître le rôle de pionnier joué par A.D. Booth dès 1946. Il faut bien avouer que dans ses débuts la traduction automatique est

un produit de la guerre froide, ce qui explique l'intérêt attaché au couple de langues russe-anglais. Le premier système opérationnel devait être complété en 1963 : le système de Georgetown, encore utilisé aujourd'hui au centre de recherches nucléaires d'Oakridge (USA) et abandonné depuis peu par le centre EURATOM d'Ispra (Italie). Quelques années plus tard, un système analogue était commercialisé par un ancien programmeur de Georgetown et a été introduit l'année dernière sur le marché canadien.

On peut schématiser le fonctionnement de ce genre de système en se représentant une fenêtre qui laisse apparaître 3 mots environ ; celle-ci se déplace le long de la phrase et chaque mot est désambigué sur la base de son microcontexte en consultant des tables de compatibilités syntaxiques et sémantiques. Ensuite, une série de routines ad hoc déplacent certains éléments de la phrase pour répondre aux particularités de la langue cible.

Les insuffisances de ce genre d'approche sont apparues très tôt et, encouragés par les progrès de la linguistique transformationnelle et de l'informatique, plusieurs centres de recherches universitaires se sont lancés dès 1960 dans la réalisation de ce qui devait devenir la deuxième génération de traduction automatique. En 1966, le rapport ALPAC, qui n'aurait jamais été pris au sérieux s'il n'avait été publié sous l'égide d'institutions aussi vénérables que la National Academy of Sciences, mettait une fin abrupte à vingt années de recherches. Quelques mois plus tard, des quelque 20 centres recensés par Yehoshua Bar-Hillel en 1958, il n'en restait plus qu'une poignée.

La traduction automatique au Canada de 1965 à 1971

Dès 1964, l'Imprimeur de la Reine pense à la traduction automatique pour obtenir rapidement une traduction approximative révisable de documents officiels. Le Conseil national de recherches commence alors à subventionner le projet CETADOL (Centre d'études pour le traitement automatique de données linguistiques) de l'Université de Montréal et le projet de Kathleen H. V. Booth (épouse de A. D. Booth) de l'Université de Saskatchewan. Des subventions seront également accordées pendant trois ans au Cambridge Language Research Unit de Margaret Masterman ; ce groupe britannique a une approche originale mais ne semble pas le plus avancé dans ses travaux.

Le projet de Saskatoon prend immédiatement une orientation statistique et base le fonctionnement de son système sur des tables de fréquence. En ce sens il est l'héritier direct de la première génération et rencontre les mêmes difficultés ainsi qu'en témoignent les derniers rapports publiés en 1971.

Le projet de Montréal au contraire, dirigé par Guy Rondeau et conseillé par le CETA de Grenoble avec lequel il entretiendra des relations privilégiées, entreprend la réalisation d'un système reposant sur un modèle linguistique. Les débuts sont lents (1965, 1966) : étude de la morphologie de l'anglais, catégorisation syntaxique, classification des mots grammaticaux. Mais en 1967 l'implantation de l'algorithme de Kuno-Woods permet d'aborder sérieusement les problèmes d'ana-

lyse syntaxique et justifie plusieurs études théoriques. Un nouveau pas en avant sera fait en 1968 avec les W-grammaires, formalisme de description syntaxique introduit par Van Wijngaarden et utilisé pour la première fois dans la définition du langage ALGOL. Les linguistes peuvent enfin effectuer des transformations syntaxiques à leur gré mais aimeraient avoir un langage qui se prête mieux aux manipulations d'arborescences. C'est alors que Alain Colmerauer a l'idée d'un langage dans lequel on peut écrire toutes les phases d'un système de traduction automatique : les Systèmes-Q, encore utilisés aujourd'hui. En 1971 le projet CETADOL devient le projet TAUM (Traduction automatique à l'Université de Montréal) et sous la direction d'Alain Colmerauer présente le premier système complet de traduction automatique.

Parallèlement à ces travaux, Brian Harris et Thomas R. Hofmann font leurs expériences de transcodage (Pidgin translation). Il s'agit d'un mot à mot ou à chaque terme de la langue source on fait soit correspondre une traduction valise soit un terme nouveau mais facile à comprendre (light -> lumiléger). Les auteurs ne prétendent pas offrir une traduction de qualité mais un texte compréhensible à des personnes ne connaissant pas la langue source.

Le projet TAUM depuis 1972

En 1972 le Conseil national de recherches ne subventionne plus que le projet TAUM pour enfin passer la main un an plus tard au Bureau des traductions du Secrétariat d'État. Les linguistes, sous la direction de Richard Kittredge, améliorent le système expérimental TAUM 71 et en présentent une version augmentée en 1973. Entretemps les informaticiens commencent à concevoir de nouveaux logiciels. L'année 1974 est marquée par la programmation d'une morphologie générative complète du français et par les premières expériences en REZO, adaptation réalisée par Gilles Stewart des réseaux de transition de William Woods. En 1975, le projet TAUM se voit attribué parallèlement à ses recherches son premier contrat de développement : le système METEO. En 1976, le projet s'oriente résolument vers les applications en obtenant le développement d'un système beaucoup plus ambitieux encore. Les deux prochaines années seront donc consacrées à la réalisation d'un système de traduction automatique de manuels techniques qui sera vraisemblablement utilisé dans un premier temps pour traduire la documentation technique de l'Aurora.

Conclusion

Certes l'histoire de la traduction automatique au Canada ne se résume pas à celle du seul centre qui ait survécu ; un grand nombre de personnes ont contribué de près ou de loin au développement de cette spécialité par leur appui, leurs critiques et leurs encouragements. Mais on est en droit de se demander où en serait le Canada aujourd'hui s'il avait attendu le rapport ALPAC pour s'intéresser à la question.