

John Chandioux

## LES PROMESSES DE LA TRADUCTION AUTOMATIQUE

### La tour de Babel

La majeure partie de l'information qui circule aujourd'hui sur Internet est encore en anglais, mais le besoin de communication en d'autres langues se fait sentir de manière croissante. Les problèmes posés par la communication entre deux internautes de même langue sont quasiment résolus. On a d'abord offert des solutions ad hoc pour la transmission et l'affichage des caractères en différentes langues. Ces techniques deviennent caduques avec la généralisation de la norme Unicode, qui fait passer le support multilingue dans le domaine public.

Le problème de la communication entre deux internautes qui ne comprennent pas la même langue demeure entier. Les progrès prodigieux de la micro-informatique au cours des vingt dernières années nous ont habitués aux miracles et certains pensent sans hésitation à des solutions comme le traducteur universel de Star Trek. N'en déplaise aux investisseurs qui gravitent autour d'Internet, la seule chose dont on est sûr après plus de quarante ans de recherche et développement en traduction automatique, c'est que ce n'est pas demain la veille.

L'histoire de la traduction automatique est en partie celle d'utilisateurs potentiels qui n'ont pas su faire la différence entre leurs besoins et l'état de développement de la technologie.

### Les vestiges de la guerre froide

La plupart des grands projets de traduction automatique sont nés entre 1958 et 1966 des besoins de traduction à partir du russe engendrés par la guerre froide.

Le meilleur exemple est celui de l'utilisation du système Systran par la Foreign Technology Division de l'armée de l'air américaine. Ce système a été utilisé pendant plus de vingt ans pour traiter la totalité des textes en langue russe sur

lesquels les services de renseignements pouvaient mettre la main. Il était doté d'énormes dictionnaires russe-anglais couvrant un grand nombre de domaines et le charabia produit par le système était examiné par des spécialistes qui étaient chargés d'isoler les textes ayant une valeur stratégique ou scientifique.

Cette application n'est pas sans rappeler Internet, mais s'agit-il de traduction automatique ou de filtrage en langue étrangère ? La réponse devient claire quand on sait que la plupart des textes retenus étaient ensuite traduits manuellement.

La recherche en traduction automatique a connu un coup d'arrêt en 1966 avec la sortie du rapport ALPAC de la National Science Foundation qui concluait à l'impossibilité d'une traduction automatique de qualité. La plupart des chercheurs américains se sont alors tournés vers une science émergente : l'intelligence artificielle.

La technologie n'en a pas moins évolué au cours des ans. Les développeurs de la fin des années cinquante concevaient la traduction automatique comme un simple problème de décryptage. À chaque mot du texte incompréhensible correspondait un mot de la langue connue. Ils ont vite découvert que la traduction n'était pas biunivoque, qu'un même mot pouvait avoir plusieurs catégories grammaticales et qu'à chaque catégorie grammaticale pouvaient correspondre plusieurs sens.

Des tests de compréhension ont alors révélé que si un locuteur humain lisait un texte en déplaçant une feuille de carton avec un trou qui laissait paraître quelques mots contigus, beaucoup d'ambiguïtés pouvaient être levées. Ceci a donné naissance à l'analyse par micro-contexte, qui est le principe de fonctionnement des systèmes de traduction automatique dits de première génération.

### [La deuxième génération à la rescousse](#)

Les chercheurs du GETA (Groupe d'étude pour la traduction automatique) à Grenoble ont été les premiers à jeter les bases d'une nouvelle génération. La traduction humaine peut se décomposer en trois étapes : la compréhension du message en langue source, la transposition de la teneur du message en langue cible et la formulation du message selon les règles de la langue cible. Parallèlement, un

système de traduction automatique de deuxième génération opère au niveau de la phrase et la traite en trois étapes : l'analyse, le transfert et la génération.

Le premier système de traduction automatique de deuxième génération est entré en exploitation le 24 mai 1977 à Montréal; il s'agit du système METEO<sup>®</sup>.

Aussi raisonnable que puisse paraître cette approche, personne n'est parvenu à développer un système général de traduction automatique de deuxième génération, même en se limitant aux textes scientifiques et techniques.

La première étape suppose qu'on sache bâtir un analyseur d'une langue qui soit un tant soit peu exhaustif. Or ce n'est pas possible sans faire intervenir le sens et on ne sait toujours pas comment représenter et traiter le sens en dehors d'un domaine restreint.

La seconde étape est plus ambitieuse encore car elle suppose une manipulation du sens dans la mesure où le transfert est la transposition d'idées et non simplement de structures syntaxiques.

Paradoxalement, la troisième étape pose peu de problèmes car elle applique les règles syntaxiques et morphologiques d'une langue de manière déterministe.

On sait déjà que même avec une composante syntaxique, le système de deuxième génération ne suffirait pas à la tâche. En particulier, la phrase n'est pas un contexte suffisant pour résoudre certains problèmes de traduction et d'autres ne peuvent être tranchés qu'avec une compréhension du texte et une connaissance du monde.

### Les logiciels grand public

Les logiciels de traduction automatique disponibles dans les magasins ont tous pour objectif de traduire un texte quelconque d'une langue dans une autre. Les vendeurs de ces logiciels prétendent que les traducteurs mettent la barre trop haut pour protéger leur profession et que les utilisateurs peuvent se contenter d'une mauvaise traduction.

La véritable question est de savoir quel est l'objectif visé.

Si l'objectif est de se faire une idée du sujet d'un texte écrit dans une langue qu'on ne comprend pas, la technologie existe, mais donnons-lui plutôt un nom qui n'induit pas les utilisateurs en erreur : logiciel de filtrage ou de dépistage de l'information.

Si l'objectif est d'obtenir une traduction totalement compréhensible d'un texte en langue étrangère, aucun système sur le marché ne satisfait à cette exigence.

Si l'objectif est de fournir un outil qui facilite le travail de traduction, quelques clarifications s'imposent.

Quand on parle d'industries de la langue, la technologie est tellement embryonnaire par rapport à l'objectif à atteindre que les outils disponibles ne doivent pas être mis entre les mains d'utilisateurs qui n'ont pas les connaissances nécessaires pour faire preuve de sens critique. De même qu'un correcteur grammatical n'est utile qu'à une personne qui sait écrire, un logiciel de traduction automatique est un outil de travail pour celui ou celle qui sait traduire.

Ceci étant dit, les logiciels de traduction automatique sont-ils utiles pour le traducteur de métier ? En règle générale, non !

Quand on donne un texte à traduire à un traducteur qu'on ne connaît pas, on reçoit en général un texte qui est diffusable avec un minimum de révision. Plus rarement, on reçoit un texte mal traduit. Le travail a été bâclé ou l'auteur de la traduction ne comprend rien au domaine et a fait des contresens majeurs ou encore, ce n'est pas un traducteur de métier. Dans ce cas, cela prend souvent moins de temps de retraduire le texte au complet que d'essayer de le réviser.

Le traducteur humain a en général le même sentiment face à la sortie d'un logiciel de traduction automatique, d'autant plus que la correction des fautes faites par ces logiciels nécessite beaucoup de manipulations.

## Un aveu d'échec

L'informatique a fait plus de progrès en vingt ans que la traduction automatique. On peut acheter aujourd'hui pour quelques centaines de dollars une version micro-ordinateur des systèmes développés à coup de millions sur les plus gros ordinateurs disponibles pendant la guerre froide. Certes les dictionnaires sont plus complets, les interfaces sont plus conviviales et il y a plus de langues offertes, mais la qualité de traduction est sensiblement la même.

La plupart des grands organismes multilingues ont envisagé la traduction automatique à tour de rôle. Certains sont allés jusqu'à financer de nouvelles recherches. Mais ils ont tous dû se rendre à l'évidence et se tournent aujourd'hui vers des solutions plus modestes comme les outils d'aide à la traduction.

Le Bureau de la traduction du gouvernement canadien a lui aussi dépensé des millions de dollars au cours des ans à évaluer des systèmes comme Systran, Weidner et Logos. Un essai de cinq ans du système Logos a démontré que la traduction automatique avait coûté trois fois plus cher que la traduction manuelle. Le Bureau en est venu à la conclusion qu'aucun logiciel commercial ne pouvait répondre à ses besoins.

On sait désormais que la traduction automatique générale n'est pas viable, mais on sait également quand la traduction automatique peut rendre de réels services. L'objectif initial était trop ambitieux dans l'état actuel de la technique. Pourquoi ne pas restreindre l'objectif en se limitant à certains types de textes, à certains domaines ou encore en réduisant l'ambiguïté des textes lors de leur rédaction ?

## La mémoire de traduction

À Montréal en 1990, la compagnie d'assurances Confédération-Vie prenait livraison d'un système dédié d'aide à la traduction qui lui permettait d'assembler la version française d'une police d'assurance-vie à partir de la version anglaise en vingt minutes au lieu de quatre heures. Le système Général TAO, utilisé aujourd'hui par ManuVie, remplace les clauses types d'un document par leur

traduction tout en générant la mise en page désirée. Il ne s'agit pas à proprement parler de traduction automatique, mais d'une mémorisation des phrases répétitives.

Depuis, des systèmes commerciaux de mémoire de traduction ont vu le jour comme Translation Manager 2 d'IBM ou Translator's Workbench de Trados qui mémorisent toutes les phrases au fur et à mesure de leur traduction.

L'ordinateur distingue très bien ce qui est semblable de ce qui est identique et reproduit avec exactitude les sommes et les dates là où l'humain peut laisser passer des erreurs coûteuses en travaillant par couper-coller.

Mais si les systèmes de mémoire de traduction sont réellement utiles sur du texte répétitif comme les contrats d'assurances, les conventions collectives ou les descriptions de tâches, ils sont complètement inutiles ailleurs.

De plus, la notion de répétitivité est subjective à un point tel que même le traducteur peut s'y tromper. Quoi de plus répétitif que des bulletins météorologiques me direz-vous ? C'est ce qu'affirmait dernièrement l'équipe de traducteurs qui travaille au Centre météorologique du Québec à Montréal. Une étude d'une classe de bulletins pendant six mois démontre cependant que sur plus de 40 000 phrases traitées, seulement 700 d'entre elles se répètent.

### Les sous-langages

On appelle sous-langage un sous-ensemble de la langue ayant ses propres règles syntaxiques et un vocabulaire limité dont les relations sémantiques sont cernables.

L'exemple le plus connu de l'application de la traduction automatique à un sous-langage est le système METEO qui sert à traduire les prévisions publiques, agricoles et maritimes émises par Environnement Canada. Le système est entré en exploitation sur un super-ordinateur du Centre météorologique canadien il y a plus de vingt ans et tourne aujourd'hui sur un réseau de quatre micro-ordinateurs.

La production du système METEO dépassera 30 millions de mots cette année avec un taux d'intervention humaine inférieur à 5 % sur les textes traduits automatiquement.

Certains textes comme les avis, les veilles et les avertissements échappent malgré tout encore à la traduction automatique et sont traduits manuellement en partant des émissions précédentes, ce qui allège quand même le travail de traduction de près de 60 %.

La solution est viable parce que le volume traité permet de justifier un développement continu des composantes linguistiques et informatiques. Le système METEO assure à lui seul près de 15 % du volume du Bureau de la traduction du gouvernement fédéral à un prix de revient de 0,5 cent du mot machine brut avant révision.

La solution est d'autant plus justifiée que c'est le seul moyen d'assurer la disponibilité quasi simultanée de l'information dans les deux langues. La longueur moyenne d'un bulletin est de 250 mots, soit une petite heure de travail pour un traducteur humain alors que le système METEO le renvoie à destination en moyenne quatre minutes après réception.

### Les langages contrôlés

Si on ne peut pas traduire n'importe quel texte, pourquoi ne pas écrire en fonction de la machine quand on sait que le texte devra être disponible en plusieurs langues.

C'est ce que fait depuis longtemps l'Institut textile de France à Lyon avec un système développé par un étudiant du Massachusetts Institute of Technology. Le système TITUS sert à alimenter une banque multilingue de résumés scientifiques dans le domaine du textile. Chaque texte est entré dans sa langue d'origine en respectant les contraintes lexicales et grammaticales du système et devient automatiquement disponible dans cinq autres langues.

Certains constructeurs ont même utilisé des systèmes généraux de traduction automatique avec cette approche : ce fut le cas de Mitel, il y a quelques années, avec le système Weidner et, plus récemment, de Xerox avec le système Systran.

Mais les systèmes dédiés donnent toujours de meilleurs résultats, comme le démontre le système développé par l'université Carnegie Mellon pour Caterpillar. Ce système permet de traduire plus rapidement et à un coût moindre des manuels rédigés en « Caterpillar English ».

La plus grande surprise nous vient d'Atlanta, où une version spéciale du système METEO a été utilisée pour traduire en français les bulletins émis pendant la durée des derniers Jeux olympiques. Il y avait également deux types de textes à traduire, à cette différence près que les prévisions publiques étaient produites en langage contrôlé par un logiciel de prévision. Malgré une période de mise au point de très courte durée, METEO 96 a traduit 162 745 mots de prévisions publiques avec 99,4 % d'exactitude. Un record toutes catégories !

### Un projet prometteur

Les premières recherches sur les langages contrôlés ont porté sur l'anglais technique, le but étant d'obtenir des documentations non ambiguës tant pour des utilisateurs anglophones que pour des utilisateurs dont l'anglais n'est pas la langue maternelle. Un bon exemple est le « Simplified English » développé par l'AECMA (Association européenne de constructeurs de matériel aéronautique) et dont l'usage est recommandé par l'ATA (American Transport Association).

Le GIFAS (Groupement des Industries Françaises Aéronautiques et Spatiales) mène depuis quelques années des travaux similaires sur le « français rationalisé ».

Il suffit de penser à l'accident de la gare de Lyon, qui aurait été dû à une ambiguïté de lecture de la procédure de remontage des freins de la locomotive, pour réaliser qu'une documentation technique non ambiguë n'a pas de prix.

Un projet de Guide de rédaction, visant la langue administrative, a été déposé par l'Office de la langue française à la Table de concertation sur les industries de la



langue en 1993. Ce projet a été repris par l'Observatoire québécois des industries de la langue qui a jugé préférable, après un survol de l'état de la situation, de l'appliquer à un domaine où il y avait déjà eu des efforts importants, soit l'aéronautique.

L'approche par langage contrôlé consiste à réduire l'ambiguïté des textes en limitant la complexité syntaxique des phrases et en employant les termes dans un sens seulement et avec une seule fonction grammaticale.

Le volet syntaxique du projet consiste à compléter les travaux sur le « français rationalisé » en collaboration avec le GIFAS. Le volet lexical quant à lui consiste en un inventaire de la terminologie en usage chez les constructeurs québécois et européens. Parallèlement, des comités d'harmonisation comme celui du professeur Poitevin de l'École nationale d'aérotechnique à Saint-Hubert font des recommandations d'usage.

Le but du projet est de parvenir à définir un ensemble de règles de rédaction technique permettant de rédiger des manuels d'entretien non ambigus pour l'humain et pour la machine, se prêtant par là même à la traduction automatique.

Si l'ordinateur ne comprend pas ce que nous écrivons, écrivons dans un langage qu'il comprend et laissons la littérature aux humains...

---

Source : <http://www.chandioux.com/document/tradauto.htm>