

LA TRADUCTION AUTOMATIQUE BREF HISTORIQUE

Note préliminaire

Il ne sera pas question ici de traduction assistée par ordinateur (TAO) mais uniquement de traduction automatique (TA) au sens le plus strict du terme. Alors qu'en TAO la machine ne fournit pour chaque segment linguistique que des propositions de traductions parmi lesquelles le traducteur, posté devant son écran, doit choisir, ou du moins aux-quelles il doit réagir (systèmes interactifs), en TA c'est un texte entier qui est traité sans intervention humaine. Une confusion entre ces deux types de traduction (ou d'aides à la traduction) règne dans l'esprit d'un public mal informé, confusion qui, nous l'espérons, se dissipera à mesure que l'information sera mieux diffusée.

DANS LES PAYS OCCIDENTAUX industrialisés le besoin en information technique et scienti-fique s'est accru de façon sensible pendant la dernière décennie – voire les deux dernières décennies. Cette information est produite dans diverses langues que les ingénieurs, techniciens, chercheurs ne connaissent pas obligatoirement. Les traducteurs spécialisés sont relativement rares sur le marché, et lorsqu'ils sont bons ils sont surchargés de travail. Par ailleurs les grandes compagnies industrielles ou commerciales, les grandes administrations nationales et internationales ont parfois leurs propres services de traduction, mais là aussi les « traducteurs maison » ont vu leur charge de travail augmenter de telle sorte qu'ils ont dû s'intéresser aux aides automatisées que l'information pouvait leur fournir.

L'idée même de la traduction automatique remonte au début des années 1950, époque où l'on attendait un maximum des « calculateurs électroniques ». En fait, dès 1946 sous la plume de W. Weaver et A.D. Booth (cités par M. Nagao dans *La Recherche*), l'idée d'une traduction automatique était émise.

On pensait alors qu'il suffisait de remplacer les mots d'une langue par ceux d'une autre langue pour obtenir un texte traduit. Actuellement nous ne pouvons qu'être étonnés de cette naïveté. Les traducteurs humains savent que le processus de traduction est complexe : identification des unités de signification, repérage du sens, réexpression du sens (si possible!) dans d'autres structures syntaxiques et avec d'autres modèles lexicaux et morphologiques, avec recherche de compromis lorsque le couple de langues traitées ne permet pas un transfert simple. Les États-Unis ont d'abord confié à l'Université

Georgetown à Washington un programme de recherche en vue de la traduction automatique du russe vers l'anglais. Cet essor rapide est lié intimement à un événement technico-politique précis : le lancement réussi du premier satellite Spoutnik le 4 octobre 1957. Puisque les Russes avaient réussi la mise sur orbite du premier satellite, les Américains décident d'être les premiers à marcher sur la Lune. Connaître l'état de la technologie et les orientations de la recherche soviétique était alors vital pour que la politique spatiale américaine puisse être appliquée. Cela explique les sommes attribuées aux recherches pour la traduction automatique. Malheureusement les délais escomptés pour réussir se révèlent vite beaucoup trop brefs. Les recherches n'aboutissent pas à des résultats satisfaisants assez rapidement et le Sénat ordonne l'ouverture d'une enquête. La conclusion en est le fameux « rapport ALPAC » (ALPAC = Automatic Language Processing Advisory Committee) qui semble clore définitivement le chapitre des travaux vers la traduction automatique en 1967.

Le gouvernement américain coupe tous les crédits relatifs aux recherches pour la traduction automatique. Les gouvernements qui avaient également commencé à lancer des projets, tel que la France ou le Japon, suppriment aussi les subventions. À ce moment les recherches deviennent le fait soit d'entreprises privées, soit d'Universités suffisamment importantes pour ne pas dépendre des crédits gouvernementaux, soit d'organisations internationales pour lesquelles la question de la traduction est vitale. Le problème de la traduction automatique semble à certains essentiellement financier (*cf.* Loll Rolling dans une interview à *Science et Avenir*, Mai 1980) bien que pour d'autres (*cf.* T. Wonograd cité dans *La Recherche*, déc. 1983) l'absence de connaissance précise sur le fonctionnement de l'intelligence et de la compréhension soit en fait la pierre d'achoppement de la traduction automatique. On a pu rêver un certain temps d'un système automatique reproduisant, imitant, les démarches du cerveau humain. Mais à mesure que l'informatique devient réalité, que les applications sont mises au point, on s'aperçoit que tout est beaucoup plus complexe qu'on ne l'a imaginé. Certains espoirs disparaissent par exemple celui d'obtenir une traduction parfaite de la *Pravda* en anglais. Mais en même temps les vues sur l'avenir deviennent plus réalistes et les projets plus modestes vont pouvoir donner des résultats.

Pendant ce temps le Dr Toma continue à travailler en Californie et le système SYSTRAN est mis au point pour le couple russe-anglais, du russe vers l'anglais. Ce premier SYSTRAN semble donner satisfaction puisque l'Armée de l'Air américaine (USAF) à Dayton, Ohio, le met en applications dès 1970.

En 1976, la Commission des Communautés Européennes acquiert le système SYSTRAN et entreprend de le développer pour le couple anglais-français. Un an plus tard le couple français-anglais commence à être élaboré, et l'année suivante, 1978, il peut déjà produire des résultats. En 1979 la synthèse de l'italien permet d'avoir aussi le couple anglais-italien. Trois couples de langues peuvent ainsi être traités par SYSTRAN.

Il était, semble-t-il, assez naturel que ce soit un organisme tel que la Commission des Communautés Européennes qui se fasse le promoteur de la traduction automatique. En effet le volume de textes officiels, rapports techniques et politiques et autres documents

d'information spécialisés qui y sont utilisés et qui y sont traduits dans les sept langues officielles – allemand, anglais, danois, français, grec, italien, néerlandais – qui sont maintenant neuf : espagnol et portugais en plus – est tel que le Service de Traduction de la Commission des Communautés Européennes peut bien être déclaré le plus grand d'Europe. Traduire est une obligation légale à laquelle la Commission ne peut échapper, différant en cela d'un organisme industriel où la direction peut espérer réduire le volume de traduction en obligeant ses techniciens ou ingénieurs à connaître plusieurs langues étrangères de façon à utiliser les documents primaires. Près de la moitié du budget communautaire était utilisé pour la traduction...

La Commission des Communautés Européennes avait un besoin pressant d'aide informatique à la traduction. Pour autant la traduction automatique a-t-elle été introduite sans heurt? Les difficultés techniques peuvent être résolues avec du temps et de l'argent (ou de l'argent et du temps!) mais les difficultés politiques et psychologiques sont d'une autre nature. Loll Rolling, chef du Service de l'Innovation et du Marketing à Luxembourg et promoteur des « plans d'action pour l'amélioration du transfert de l'information entre les langues européennes » n'a pas eu la tâche facile. Le dictionnaire automatique (EURODICAUTOM) était aisément reconnu par les traducteurs comme une aide appréciable. En effet les dictionnaires imprimés sont toujours en retard, surtout dans les domaines techniques ou scientifiques de pointe. Un dictionnaire informatisé peut très aisément être mis à jour. De plus il est très facilement consultable (chaque traducteur peut avoir dans son bureau un terminal de consultation). Par contre la traduction automatique – tout comme les robots en général – soulèvent un vent de crainte profonde. La machine va-t-elle remplacer l'homme? Les traducteurs vont-ils perdre leur emploi?

La traduction automatique n'a pas tout de suite été introduite au Service de Traduction mais au Service de l'Innovation. Seuls les traducteurs volontaires durent ou purent s'en servir. Ni à cette époque, ni maintenant un traducteur n'est obligé de travailler avec la traduction automatique. Mais, demandera-t-on, puisque la traduction est automatique, quel rôle ont encore à jouer les traducteurs?

« Traduction Automatique » signifie bien que l'on entre dans l'ordinateur un texte sans préparation (sans pré-édition) et qu'il en ressort après quelques secondes ou quelques minutes de traitement un autre texte. On appelle ce dernier « traduction brute ». Le texte produit automatiquement en langue cible présente généralement (sauf certains domaines et certaines structures morphosyntaxiques) des erreurs, et non des fautes (*cf.* A.M.L.L., 1983b). Les erreurs ne sont pas terminologiques, au contraire nous dirions que la machine est plus fiable que l'homme pour ce qui est de la terminologie spécialisée. Elles sont morphologiques ou syntaxiques, et dues à des lacunes dans la programmation ou à des ambiguïtés. Les linguistes de la Commission des Communautés Européennes, avec l'aide des traducteurs participant au développement de la Traduction Automatique, travaillent en permanence à l'amélioration du système, et de la traduction de chacun des couples de langues traitées, sous la direction de Ian M. Pigott à Luxembourg. Parfois le texte est compréhensible et utilisable malgré ses erreurs. Mais le plus souvent il demande à être post-édité.

Le Centre de Recherches Nucléaires de Karlsruhe (RFA) utilise le système SYSTRAN sans post-édition et estime que les résultats sont satisfaisants pour les besoins des spécialistes du Centre. Par contre la Commission des Communautés Européennes ne laisse aucun document sans post-édition. Pourquoi ces différences? Tiennent-elles à la qualité de la sortie machine? Les attentes et les demandes des divers « clients » de traduction automatique ne sont pas les mêmes parce que l'utilisation qui est faite des textes traduits n'est pas la même. Lorsqu'on cherche les données brutes, l'information scientifique fondamentale pour ses propres besoins de technicien ou d'ingénieur, et lorsque l'on doit diffuser un texte parmi les représentants officiels de plusieurs groupes de spécialistes de plusieurs pays, l'attention donnée à la formulation ne sera pas la même. À plus forte raison si ces textes doivent servir de base à des discussions d'ordre politique. C'est pourquoi dans le premier cas, la post-édition peut être évitée et dans l'autre cas elle est obligatoire.

Quels types d'erreurs sont les plus répandus? Une étude d'erreurs a été menée sur un échantillon de textes communautaires en 1982 (*cf.* Rapport d'étude réalisée par le CEDEL, et A.M.L.L., 1983a). Les erreurs ont été classées et évaluées quant à leur fréquence. Parallèlement une autre étude était menée par Veronica Lawson sur l'évaluation du temps nécessaire à la correction des erreurs. Pour résumer, il est apparu que les erreurs portant sur le lexique sont rares et, si elles existent, elles sont corrigées aisément. Les erreurs morphologiques sur l'opposition masculin/féminin par exemple sont beaucoup plus rares qu'on ne l'aurait pensé. Les erreurs de syntaxe sont les plus difficiles à corriger. Il ne suffit pas toujours d'indiquer un déplacement par une flèche, il faut parfois reconstruire les phrases. Ainsi par exemple, lorsqu'un énoncé comporte en français une cascade de prépositions, en particulier *de*, le système ne reconnaît pas toujours l'incidence des adjectivations (complément de nom). Inversement, dans un groupe anglais « substantif + plusieurs adjectivations » le système ne reconnaît pas toujours où se situe le noyau du groupe.

On peut classer les erreurs subsistant après traduction automatique par SYSTRAN en trois types :

1. erreurs portant sur le mot isolé,
2. erreurs portant sur les relations,
3. erreurs portant sur les structures de l'information.

Chacun de ces types se subdivise en catégories que nous allons ici préciser.

1. Le mot isolé : l'erreur n'affecte qu'un mot, elle peut être corrigée simplement par remplacement par un autre mot, le reste de l'énoncé demeurant égal à lui-même. Les mots qui sont erronés et qu'il faut post-éditer font partie des cinq catégories :

- a) vocabulaire général (ignorances car domaine non couvert par les dictionnaires, ambiguïtés);
- b) terminologie (dans les domaines couverts par les dictionnaires de la CCE, ignorance du système car néologie, ex. mot d'emprunt ou création d'auteur – des guillemets parfois signalent ce type de néologie, mais pas toujours);

- c) noms propres et sigles (traitement similaire au vocabulaire général par traduction au lieu de conservation);
 - d) relateurs, en particulier prépositions (les prépositions à emplois multiples telles que *à, de, en, pour, par*, présentent de particulières difficultés à la traduction; elles se manifestent dans les déterminants nominaux et dans les complémentations verbales);
 - e) déterminants de liste finie du nom : articles et démonstratifs.
2. Erreurs concernant les relations :
- a) formes des verbes (temps);
 - b) voix verbales (passif/ actif) et degré de personnalisation de l'énoncé (passif – impersonnel par *on* – impersonnel par *il*);
 - c) expression ou non des modalités;
 - d) négations.
3. Erreurs concernant les structures et l'information :
- a) introducteurs de phrases et relations logiques entre énoncés;
 - b) ordre des mots,
 - c) problèmes généraux d'incidence;
 - d) informations ajoutées ou retranchées.

Ce classement des erreurs a été réalisé en vue d'un tri des documents. En effet, il est peu efficace de soumettre à la traduction automatique des textes qui, sans doute, sortiront avec une proportion d'erreurs trop élevée, c'est-à-dire qui demanderont une post-édition longue et coûteuse. Il est possible d'évaluer la probabilité de traduction automatique satisfaisante ou rentable. Tous les documents communautaires ne sont donc pas traduits par SYSTRAN.

Lorsque la traduction automatique était encore expérimentale les posts-éditeurs n'avaient pas de consigne précise. Ils devaient simplement « corriger » le texte. À mesure que les recherches entourant la post-édition se sont développées, deux pratiques différentes de post-édition ont été mises au point : la post-édition rapide (PER) et la post-édition conventionnelle (PEC). La PEC vise à fournir un texte aussi proche que possible de ce qu'aurait été une traduction humaine. La PER donne un texte clair, correct, compréhensible, conforme, mais sans « recherche de style ». Pour les documents techniques ou tous ceux que l'on lit rapidement, en diagonale (et que l'on jette presque aussitôt) une telle post-édition est amplement suffisante : le contenu du document n'est pas lié au style. Cette question a pu être discutée (*cf.* A.M.L.L. « Recherches lexicales et syntaxiques sur les discours scientifiques et techniques », thèse pour le Doctorat d'État, Université de la Sorbonne Nouvelle, 1986). Par contre pour les documents qui sont publiés ou largement diffusés, et *a fortiori* ceux qui ont une valeur officielle, une post-édition conventionnelle est nécessaire (voir A.M.L.L. 1986a et b).

Actuellement la Commission des Communautés Européennes à Luxembourg et à Bruxelles utilise largement la traduction automatique avec post-édition rapide.

On ne peut terminer ce bref aperçu sur la traduction automatique sans mentionner un système automatique mis au point par l'Institut Textile de France : TITUS. Celui-ci a été

développé pour quatre langues : français, anglais, allemand, espagnol. À la différence de SYSTRAN qui accepte tout texte, TITUS fonctionne en « syntaxe contrôlée ». Cela signifie que certaines structures morpho-syntaxiques sont interdites. Cependant, petit à petit TITUS acquiert plus de souplesse. Il est utilisé par le Centre de Documentation Scientifique et Technique du C.N.R.S., pour la traduction de résumés de documents scientifiques.

De même, le Canada ayant une obligation légale de traduction, a beaucoup investi dans la traduction automatique. Mais là, deux langues seulement sont traitées. Depuis 1977 TAUM METEO, construit pour le Gouvernement canadien par l'Université de Montréal, est un système automatique effectivement utilisé pour traduire toutes les informations météorologiques. Les recherches se poursuivent pour utiliser l'ordinateur dans d'autres domaines. La Suisse utilise un « mini-système » SEPPLI à Berne (cf. P. Shann et S. Warwick, « Études contrastives dans un système de traduction automatique (SEPPLI) », *Contrastes*, A3, 1983) pour la traduction des offres d'emplois fédérales. L'Allemagne, en particulier l'Université de Saarebrück, a également entrepris des recherches en vue de la traduction automatique. Et le Japon, qui en a un besoin intense pour traduire toutes les notices techniques des produits qu'il exporte, a tôt fait de lancer plusieurs équipes dans ce domaine à la fois à l'Université et dans les Industries de façon à mener de front l'approche théorique et l'approche pratique. Il est évident que celles-ci doivent toujours être combinées, mais hélas dans le domaine de la Traduction Automatique cela a souvent été oublié. Les ingénieurs informaticiens ont souvent cru que leur savoir technique leur permettrait de traiter de la langue de la même façon qu'ils traitent de phénomènes naturels, d'électricité ou de corps chimiques. À l'inverse, les linguistes ont parfois cru qu'ils pourraient seuls résoudre tous les problèmes liés aux langues, y compris ceux de l'automatisme. Nous savons maintenant qu'on n'échappe pas à la collaboration. Surtout dans les domaines de pointe tel que celui-ci.

Enfin il convient de parler du système européen de demain : EUROTRA. Tandis que SYSTRAN fonctionne par couples de langues, EUROTRA est un système fondé sur des modules de transfert permettant la jonction entre l'analyse d'une langue et la synthèse d'une autre. Chaque pays de la Communauté Européenne a mis en place une équipe de chercheurs, de linguistes et d'informaticiens, pour élaborer ce système d'un genre tout à fait nouveau. Pour plus d'informations voir le numéro spécial de *Multilingua* consacré à EUROTRA : *Multilingua*, 5-3, 1986. Un modèle expérimental devrait être présenté à la fin de cette décennie.

Mentionnons également que diverses institutions japonaises travaillent aussi en ce moment à la mise au point de systèmes de traduction automatique. Certaines tentent d'implémenter SYSTRAN. D'autres élaborent leur propre produit, en misant sur les possibilités de la cinquième génération d'ordinateurs. Les recherches pour la traduction automatique sont ainsi de plus en plus liées aux recherches dans le domaine de l'Intelligence Artificielle.

D'autres systèmes en cours d'étude ont été omis ici : que leurs promoteurs n'en prennent pas ombrage : nous avons seulement voulu donner quelques indications d'information générale.

Indications bibliographiques

Articles

- (1979) Ian PIGOTT et Georges VAN SLYPE, Description du système de traduction automatique de la Commission des Communautés Européennes, *Le Documentaliste*, vol. 16, n° 4.
- (1983) Makoto NAGAO, La Traduction Automatique, *La Recherche*, n° 150.
- (1983) Loll ROLLING, Machine Translation's big ten years: 1980-1990, *Actes du congrès EURIM 5*, Aslib ed.
- (1983a) Anne-Marie LOFFLER-LAURIAN, Pour une typologie des erreurs dans la traduction automatique, *Multilingua*, 2-2.
- (1983b) *Id.*, Notion de faute et linguistique contrastive aujourd'hui, *Bulletin de Terminologie de la C.C.E.*, 43.
- (1984) *Id.*, Normes de communication et traduction de textes techniques, *Meta*, vol. 29, n° 2.
- (1984) *Id.*, Machine translation: what type of post-editing on what type of documents for what type of users, *Actes du congrès COLING 84*, Stanford, Californie.
- (1985) *Id.*, Traduction automatique et style, *Babel*, vol. XXXI, n° 2.
- (1986a) *Id.*, Post-édition rapide et post-édition conventionnelle : Deux modalités d'une activité spécifique (I), *Multilingua*, 5-2.
- (1986b) *Id.*, Post-édition rapide et post-édition conventionnelle : critères pour la post-édition rapide (II), *Multilingua*, 5-4.

Volumes

- (1984) *Contrastes*, Hors Série A4, Traduction Automatique – aspects européens .
 - (1986) *Multilingua*, 5-3 (volume consacré à EUROTRA).
-

Source : Ballard, Michel (dir.), *La traduction plurielle*, Lille, Presses Universitaires de Lille, 1990, p. 143-152.