

LA TRADUCTION AUTOMATIQUE

Frédérique LAB

La "Traduction Automatique" s'inscrit dans un ensemble de recherches menées depuis le début des années 1950 - recherches que l'on a coutume de regrouper sous l'appellation de "traitement automatique des langues naturelles".

Les premières recherches dans ce domaine ont été entamées dans l'euphorie des premiers grands succès de l'informatique : les réussites remportées dans le traitement des données numériques avaient alors amené les informaticiens à s'intéresser à d'autres domaines que les domaines purement mathématiques, en particulier celui du langage. On peut avancer que les désillusions qui ont suivi dans le traitement des langues naturelles sont dues à ce que, d'un point de vue théorique, l'idée qui animait les chercheurs de l'époque était que, pour travailler sur le langage, il s'agissait de procéder à une simple substitution de données : les données numériques seraient remplacées par les données alphanumériques que constituaient les mots. La Traduction Automatique des débuts était donc essentiellement une traduction de mots. De fait, la part dévolue aux dictionnaires était énorme, celle dévolue à la syntaxe réduite au minimum. Les problèmes relatifs au sens d'une phrase, d'un texte, n'étaient pas abordés en tant que domaine de recherche spécifique, pas plus que l'activité de traduction n'était envisagée comme une activité ayant sa problématique propre.

Il a fallu un certain nombre d'échecs des systèmes de la première génération pour que les ingénieurs impliqués dans ces travaux accordent aux recherches linguistiques la part qui leur revenait.

QUELQUES DÉFINITIONS

Nous donnerons de la Traduction Automatique la définition suivante : un système de Traduction Automatique est un système informatique qui a :

- pour entrée un texte "t₁", ou texte source écrit dans une langue "L₁" ou langue d'origine, et n'ayant pas subi d'aménagements spéciaux préalables au traitement automatique qu'il va subir,
- pour sortie un texte "t₂" ou texte traduit écrit dans une langue "L₂" ou langue cible, tel qu'il n'ait pas à subir de transformations pour être reconnu par les utilisateurs comme une traduction du texte t₁.

Cette définition assigne à la Traduction Automatique des objectifs qu'elle n'a pas encore atteints : à la date où nous écrivons, il n'existe pas de système de Traduction Automatique qui réponde strictement à cette définition. Il suffit en fait de supprimer la condition exprimée par "tel qu'il n'ait pas à subir de transformations pour être reconnu par les utilisateurs comme un traduction du texte t₁" pour que l'on puisse dresser des recherches menées dans ce domaine un bilan moins négatif. En effet, il existe de nombreux systèmes de Traduction Automatique, dont certains sont d'ores et déjà opérationnels comme SYSTRAN, TAUM-METEO, METAL ou les systèmes des sociétés américaines ALPS Systems ou Weidner ; cependant ils ont tous recours, après la phase de traitement automatique, à des traducteurs humains qui assurent une révision du texte produit.

Ceci nous amène à préciser les termes de "Traduction Automatique" (T.A.) et de "Traduction Assistée par Ordinateur" (T.A.O.). On peut en effet distinguer deux types de traduction informatisée :

- La **Traduction Automatique**, telle qu'elle est née au début des recherches et qui se donnait pour objectif de réunir dans le programme informatique toutes les informations nécessaires à la traduction et qui devait donc, pour réussir, fournir en sortir une traduction du texte de départ et ce sans intervention humaine (cf. notre définition).
- La **Traduction Assistée par Ordinateur**, terme apparu postérieurement à celui de Traduction Automatique, en fait après les échecs des premiers systèmes. Il s'agit d'une conception de la traduction informatisée où le savoir humain est sollicité. Mais cette appellation recouvre elle-même deux choses différentes : en effet l'intervention humaine peut-être sollicitée
 - soit à la fin du traitement automatique pour une révision du texte final,
 - soit au cours du traitement (et ce autant de fois qu'il est nécessaire). Il s'agit donc là d'une approche dite interactive.

En réalité, la plupart des systèmes de traduction informatisée sont, au départ, des systèmes de "Traduction Automatique" au sens premier du terme (c'est-à-dire sans intervention humaine) et c'est au vu des résultats produits que les concepteurs optent pour le terme de "Traduction Assistée par Ordinateur" puisqu'ils se trouvent contraints, en fin d'analyse, de prévoir un recours à un ou des traducteurs humains afin d'améliorer la traduction. Dans ce cas, le terme de "Traduction Assistée par Ordinateur" est choisi parce qu'il y a eu une impossibilité à remplir les conditions de réussite de la Traduction Automatique telle que nous l'avons définie.

Les systèmes interactifs reposent sur une approche différente : le système est conçu, dès le départ, pour faire intervenir le savoir humain. L'utilisateur est sollicité, et le pouvoir de décision lui revient, dès qu'un problème "indécidable" en l'état du système se pose. Au cours du déroulement du programme, et donc de la traduction, le système "demande" à l'utilisateur de procéder à tel ou tel choix, le plus souvent en lui proposant une liste de possibilités ; il peut s'agir de choix d'ordre lexical, syntaxique ou stylistique.

TRADUCTION AUTOMATIQUE ET ÉVALUATION DES RÉSULTATS

Faut-il conclure du fait que l'ensemble des systèmes de T.A. ont recours à un traducteur humain en fin de traitement que ces systèmes sont des échecs ? De là, faut-il renoncer à mener des travaux en Traduction Automatique ? Si l'on pose que les systèmes de T.A. doivent, pour avoir réussi dans leurs objectifs, remplir les conditions exprimées dans notre définition, ceci amène à s'interroger sur l'évaluation de cette réussite. Comment évaluer la qualité du texte produit par un système de Traduction Automatique ? Comment évalue-t-on la qualité d'une traduction, quelle que soit la façon dont elle a été obtenue ? Y-a-t-il accord, lorsqu'il s'agit de juger d'une traduction "humaine", sur la qualité de cette traduction ?

Toutes ces questions ont en commun qu'en soulevant le problème de l'évaluation de la qualité de la traduction produite, elles abordent le domaine de la Traduction Automatique sous l'angle du produit final. Nous dirons ici que cet aspect du problème - même si, en dernière analyse, c'est bien la qualité du produit final qui mesure l'échec ou la réussite de tel ou tel système de T.A. - ne doit pas occulter le problème de

la qualité de la phase intermédiaire, celle du traitement automatique, celle-là même qui est invisible à l'utilisateur. En effet, on peut tout à fait concevoir un système qui, du point de vue de l'analyse et du traitement automatique, soit relativement pauvre, et éventuellement sans grand intérêt théorique, et qui néanmoins s'approche au plus près de critères de qualité et de réussite tel qu'il est normal de les définir pour une application industrielle par exemple. Ainsi, si un système est conçu pour ne traiter que des textes qui recouvrent un champ lexical relativement restreint et où la syntaxe des phrases est ramené à un certain nombre de structures, alors il est vraisemblable que la qualité de la traduction sera satisfaisante. C'est pourquoi il convient d'être prudent lorsqu'il s'agit d'évaluer la qualité d'un système de T.A. Il faut observer non seulement la qualité du texte produit mais aussi la nature du texte que le système reçoit en entrée, afin de déterminer la qualité du traitement automatique mis en oeuvre. Par ailleurs, dans ce domaine, on est amené à distinguer les projets expérimentaux des projets à destination industrielle ou commerciale, dont les objectifs ne sont pas les mêmes.

La France participe actuellement à un projet expérimental de Traduction Automatique de grande envergure, le projet "Eurotra". Ce projet européen, lancé vers 1975, a pour objectif de mettre au point un système informatique multilingue capable de traduire des textes relatifs au fonctionnement de la Communauté Européenne depuis - et vers - chacune des langues de la Communauté : anglais, allemand, danois, français, espagnol, grec, italien, néerlandais, portugais, soit 72 couples de langues.

Au cours des deux dernières années, l'ensemble des équipes européennes impliquées dans le projet ont travaillé à produire un projet-pilote portant sur un lexique de 2500 entrées par langue. Le premier prototype réel qui doit suivre (au début des années 90) couvrira un lexique de 20 000 entrées.

Les projets commercialisés, eux, ne peuvent se donner des objectifs de traduction aussi ambitieux ; de fait, il s'agit plus de systèmes d'aide à la traduction (en dépit de certaines publicités tapageuses) que de systèmes de traduction automatique. L'accent y est mis essentiellement sur des problèmes de terminologie.

Un bilan que l'on pourrait résumer ainsi : des projets de T.A. expérimentaux ambitieux mais qui n'ont pas encore fait leurs preuves ; des systèmes commercialisés à la disposition du public qui ne peuvent que décevoir les linguistes tant est grand le nombre de problèmes que ces

systèmes ne peuvent résoudre. Ceci dit assez la complexité des problèmes que pose le langage. Reste que les recherches se poursuivent et que les attitudes évoluent : les informaticiens semblent avoir pris la mesure de la complexité de problèmes à traiter ; les linguistes, pour certains, abordent le domaine du traitement automatique des langues naturelles en reconnaissant que s'offre là un champ d'expérimentation où ils se doivent d'aller mettre à l'épreuve leurs théories.

Frédérique LAB

Département de Recherches Linguistiques
Université de Paris VII.

QUELQUES INDICATIONS BIBLIOGRAPHIQUES :

La littérature sur la Traduction Automatique est relativement pauvre et les ouvrages qui lui sont consacré sont en général écrits en langue anglaise. Pour aborder ce domaine, nous conseillons la lecture du numéro 170 de **La Recherche** (octobre 1985), consacré à l'Intelligence Artificielle. Ce numéro regroupe des articles à la fois clairs et documentés sur le traitement automatique de langues naturelles. On lira en particulier l'article de D. Kayser, "Des machines qui comprennent notre langue", qui aborde tous les problèmes liés au traitement automatique de langue (analyseur morphologique, analyseur syntaxique, problèmes d'ambiguïté du langage, etc.), et qui contient lui-même des indications bibliographiques techniques.

- **La machine à traduire**, écrit par G. Mounin, et publié à Paris, par Mouton, en 1964, dont l'intérêt est de résumer la préhistoire de la traduction automatique de 1946 à 1964.
- "La traduction automatique", article de F. Goder, paru dans **Les Cahiers du Département des Langues et des Sciences du Langage, La Traduction**, Université de Lausanne, n°5, 1987 (Faculté des Lettres, Université de Lausanne, 1015, Lausanne).
- **Machine Translation**, ouvrage en langue anglaise regroupant des articles écrits par des auteurs qui sont tous engagés dans des projets de T.A. différents ; cet ouvrage, relativement technique, couvre toutes les recherches actuelles menées en T.A. : **Machine Translation**, D. Nirenburg (éd.), Cambridge, University Press, 1987.

Enfin, un service de Traduction Automatique est disponible sur Minitel - 3615 code **MITRAD** ; il s'agit d'une version commercialisée du système SYSTRAN. La demande de traduction de quelques phrases judicieusement choisies, et dans une langue familière au lecteur, illustrera clairement les types de problèmes qui restent non résolus à ce jour dans les systèmes de T.A. commercialisés.