

## LA TRADUCTION AUTOMATIQUE ET LES ALEAS DE LA LANGUE

*Kadi Allah Fayçal  
(Univesité d'Oran)*

La traduction dite automatique c'est-à-dire sans l'intervention directe ou indirecte de l'homme, qui est remplacé par la machine, dans le 'processus' de traduction, n'est qu'à ses premiers balbutiements. Pour s'en convaincre, il suffit de jeter un coup d'œil sur les différents logiciels commerciaux lorsqu'on leur soumet des textes ou simplement des phrases à traduire. C'est généralement la catastrophe !!!

Ceci est dû à la complexité de la langue ( d'une seule langue prise individuellement). Et lorsqu'on fait la 'juxtaposition' de deux langues, la complexité est double eu égard à l'extrême variété des moyens morphologiques qui expriment une même notion grammaticale et la différence dans la valeur sémantique associée. Ainsi les fonctions que le français ou l'anglais caractérisent par l'ordre des mots ou des propositions, d'autres langues les caractérisent au moyen de désinences casuelles. On rencontre même un cumul de prépositions et des désinences casuelles, cas de l'allemand et du russe par exemple. La morphologie des noms en français, par contre, n'est pas trop modifiée (sauf pour le pluriel par exemple). Mais elle est touchée dans les pronoms, distinguant ' il Le voit' de 'il Lui donne un livre'. Quant au vocabulaire, il offre un grand nombre d'homonymies et de polysémies, qui posent autant de problèmes.

La traduction automatique se résout donc, comme toute traduction, dans la transformation d'un texte écrit dans une langue dite 'langue de départ' en un texte écrit dans une langue dite 'langue d'arrivée'. Cette transformation met en jeu les ressources grammaticales et lexicales des



deux langues. Cette attitude d'affrontement et d'interaction des deux langues nécessite une gestion minutieuse des différences des deux systèmes linguistiques. Ceci n'empêche pas l'existence de spécialistes dans le domaine qui pense le contraire en essayant de simplifier les choses. C'est le cas, par exemple du Pr Bengt Sigurd, directeur du département de linguistique de l'université de Lund, en Suède, qui estime que mettre au point des programmes de traduction automatique par ordinateur véritablement efficaces et rapides n'est pas si compliqué et qu'il suffirait que les textes introduits dans la machine soient adaptés à cette fin, c'est-à-dire que l'on supprime, notamment, les prépositions aux sens ambigus, les expressions propres à une langue et qui n'ont aucun sens dans une autre et que l'on évite d'utiliser des mots polysémiques, c'est-à-dire qui présentent plusieurs sens. Le Pr Sigurd vide, ici, la langue de son âme en la réduisant et en délimitant sa substance grammaticale et lexicale pour se faciliter la tâche et ainsi contourner les écueils lors du processus de traduction automatique. Mais la langue en cela ne peut être fragmenter sans 'ankyloser' l'expression de la pensée humaine. Car l'objectif de la T.A est de pouvoir assumer la langue en tant que telle, c'est à dire à tous les niveaux que ce soit lexical, structural ou sémantique sans restriction aucune.

Plus réaliste, par contre, est le diagnostic de Martin Kay professeur de linguistique à l'université de Stanford lorsqu'il annonce que « la traduction automatique est en piteux état. ». En effet les experts de la traduction automatique n'ont pu offrir après plus de cinquante ans de recherche que des systèmes d'aide à la traduction très imparfaits. La raison, selon lui, est que l'ordinateur n'a pas accès au sens des mots qui varie selon le contexte.

La traduction automatique doit donc avoir comme support une étude scientifique des principes sur lesquels de telles traductions doivent être fondées. Ce qui veut dire qu'un travail théorique est nécessaire globalisant et l'aspect linguistique et l'aspect technique. Allant dans ce sens des tentatives bien que fructueuses à petite échelle et assez



simplexiste, comme la distinction des acceptions d'un même mot, des idiotismes, des flexions nominales (féminin, pluriel) et verbales (personnes, temps et modes), les modifications dans l'ordre des mots d'une langue à l'autre, se voyaient confronter aux méandres et à la complexité de la langue. Notre tâche, dira Ingue, « sera longue et difficile parce qu'on ne sait pas encore assez de choses aujourd'hui sur les langues, et sur la façon dont nous traduisons ». La difficulté est dès lors, non pas double mais triple, elle concerne la langue elle-même (ou les deux langues), la traduction et enfin l'aptitude à automatiser la traduction. Par souci de pragmatisme, l'expérience anglaise limite le champ d'investigation à la seule langue. Pour elle, la traduction automatique ne peut être effective que si on arrive à répertorier tous les mots d'une langue grammaticalement par catégories, et si toutes ces catégories, grammaticales et syntaxiques, peuvent être recensées complètement.

Le débat ou la polémique sur l'efficacité de la T.A s'est intensifié par la suite, sur la possibilité ou non de traduire automatiquement par exemple la poésie et la littérature. Il y a ceux qui furent pessimiste quant à cette possibilité. Ainsi l'un d'eux disait : « dans la mesure où la traduction est un art, exigeant du traducteur l'exercice de ses plus hautes capacités créatrices, les engins mécaniques seront probablement d'un faible secours ... ce type de traduction doit toujours être laissé à l'être humain ».

Pourtant on n'en est pas encore là, et une phrase aussi simple peut poser d'énormes problèmes à un ordinateur. Il suffit, pour s'en convaincre, d'examiner quelques exemples d'ambiguïté.

La plus simple étant l'ambiguïté lexicale. Ainsi le mot 'avocat' peut avoir le sens de fruit ou d'homme de loi selon que l'on écrit :

- J'ai mangé un avocat.

ou

- J'ai rencontré un avocat.

Si ce type d'ambiguïtés lexicales peut, dans une certaine mesure, être facilement distingué en procédant par des règles catégoriques de



levées d'ambiguïtés. Il n'en est pas de même, par contre, pour d'autres comme par exemple du mot 'pêche' dans :

- La pêche était bonne.

Ici c'est une seule phrase et le sens du mot 'pêche' n'est pas évident : s'agit-il du fruit ou de l'acte de pêcher ? Le recours aux règles catégoriques n'est d'aucun secours sachant, qu'à l'inverse des deux premiers exemples ( pour 'avocat'), il n'y a pas de choix préférentiel de catégories pour le mot 'pêche'. Le contexte, dans ce cas, est primordial pour opter, de la part de la machine, pour l'une ou l'autre des traductions du mot 'pêche' dans n'importe quelle langue. Cependant le contexte reste inaccessible à la machine et pour éviter une traduction erronée du mot, on procède de deux manières : soit en donnant tous les sens possibles du mot dans le texte traduit, soit en procédant par une analyse statistique du texte de départ c'est à dire en cherchant par exemple dans le texte de départ des mots se rapportant au domaine de 'l'acte de pêcher' comme 'poisson', 'bateau'...ce qui, toutefois, n'est toujours pas vrai....

L'ambiguïté structurale, quant à elle, apparaît lorsque la catégorie grammaticale du mot n'est pas évidente. Dans ' il exige le savoir', par exemple, la traduction sera différente selon qu'on considère 'le' comme un article et donc 'savoir' comme un nom, ou bien 'le' comme pronom et 'savoir' comme verbe. La traduction automatique ou le sens donné automatiquement, ici, dépend de l'analyse grammaticale qui y est faite.

L'ambiguïté structurale profonde est encore d'une grande subtilité dès lors que la structure grammaticale est claire mais l'interprétation n'est pas univoque. Par exemple pour ' il les pousse à la limite', bien qu'on sait que 'pousse' est un verbe, on ne saurait trancher sur le sens à donner à la phrase. Est-ce Que la personne est entrain de les encourager ou bien les traîne-elles vers une limite donnée ? C'est-à-dire est-ce une action physique concrète ou une action statique abstraite ?

Il y a aussi l'ambiguïté dite sémantique lorsque le syntagme peut jouer plusieurs rôles dans le sens de la phrase comme par exemple



'l'enfant est innocent'. Ici le syntagme 'l'enfant' est soit référentiel à un enfant bien précis, soit général référant à tout enfant quel qu'il soit.

La dernière ambiguïté est celle dite pragmatique et qui provient de l'emploi de pronoms ou de reprises pour lesquels la référence est mal cernée. L'exemple suivant illustre bien cela : 'la balle touche la fille, elle se blesse'. La machine ne pourra savoir à qui est fait référence par le pronom 'elle', à la 'balle' ou bien à la 'fille', pour pouvoir opter pour une traduction adéquate de celui-ci. Elle a donc besoin de connaissance sur le monde pour savoir que c'est la fille qui se blesse et non la balle c'est à dire que l'ordinateur a besoin d'informations de type 'les êtres animés se blessent par contre les objets non animés ne se blessent pas'.

Ce sont là quelques types d'ambiguïtés qu'on rencontre souvent en T.A en phase d'analyse de la langue de départ et qui restent dans la plupart des cas insolubles. Il y en a d'autres qu'on rencontre lors de la phase de génération en langue cible. Il y a aussi les ambiguïtés syntaxiques, qui dépasse le cadre du mot isolé, et qui sont nombreuses aussi.

Cette situation dramatique, à laquelle fait face la T.A aujourd'hui, vient du fait que la perception du mot par l'ordinateur diffère de celle de l'homme. Pour la machine le mot n'est qu'une suite de caractères encadrée de séparateurs. Ensuite, il y a le rôle du contexte car la plupart des phrases qui sont ambiguës pour l'ordinateur nous paraissent cependant très claires, grâce à notre connaissance de celui-ci.

Tous ces problèmes restent, toutefois subordonnés et dépendants à la fois aux unités de traduction et à un degré plus poussé aux unités sémantiques c'est à dire la façon de découper la phrase. Ceci est du au fait que pour créer un dictionnaire 'interne' de correspondance, le mot n'est pas toujours l'outil idéal. Et quand c'est le cas, sa perception, comme mentionné précédemment, par l'ordinateur diffère. Par définition, un mot, pour un ordinateur, est une suite de caractères encadrée par des séparateurs (espace blanc ou signe de ponctuation). Ainsi apparaissent les premiers accrocs (au stade du mot isolé) tels que les abréviations, les sigles ... En français, par exemple, le point '.'



n'indique pas toujours la fin d'une phrase et est partie intégrante du mot ( min., O.N.U ...), le caractère blanc est tantôt séparateur, tantôt partie intégrante du mot (parce que, bien que...). En arabe, un même mot peut être découpé de plusieurs manières ou être considéré comme une seule unité de traduction comme le 'mot': بِغَال qu'on peut découper en une conjonction 'ب' et l'adjectif 'غَال' (signifiant 'cher'), ou bien être considérée comme un seul mot signifiant pluriel de 'mule'. L'unité de traduction ou unité sémantique peut dépasser le cadre d'un seul mot ainsi :

Elle boit (son café chaud).

Elle boit (son café) (chaud).

Dans le premier exemple, si on considère 'son café chaud' comme une seule unité de traduction (complément d'objet de 'boire'), le sens serait que la personne est habituée à boire du café chaud. Par contre dans le deuxième exemple, 'son café' (pris comme une seule unité) est complément de 'boire', et le sens serait qu'en ce moment la personne est entraînée de boire du café chaud. C'est ainsi que le sens dépend du découpage en unité de traduction ou en unité sémantique .

Il y a aussi cette prédisposition d'intelligence et d'analyse instantanée qu'à l'être humain, chose qui, jusqu'à présent, reste absente chez la machine. Par quel moyen faut-il, alors, rendre la machine intelligente, capable d'analyser et d'opter ? Comment faut-il représenter cette somme gigantesque d'informations qu'on a sur le monde pour rendre la machine pragmatique ? Et surtout comment trouver une représentation sémantique combinée à un raisonnement logique afin d'éviter les risques d'ambiguïtés ? Car même pour comprendre la phrase la plus simple, nous devons connaître le contexte, la nature de l'interlocuteur et certaines données du monde en général ; l'ensemble de ces connaissances dépasse de loin les capacités actuelles des programmes les plus sophistiqués.

Toutes ces questions nous amènent au problème central de la représentation du langage pour qu'il soit intelligible à la machine. Si pour compter, calculer ou opérer numériquement, l'homme à imaginer

la representation décimale des nombres, il n'en est pas de même lorsqu'il s'agit de représenter le langage. Alors à quant est-ce que une representation linguistique universelle capable de prendre en charge tous ces aléas de la langue?....

### Bibliographie

- 'modularité et construction d'informations linguistiques', F.Debili et C.Fluhr
- 'problèmes de la traduction automatique', T.A.Documents (1968)
- 'la traduction automatique', M.Nagao, 'la recherche'(1983)
- 'les logiciels de traitement des langues naturelles' T.Winograd, 'la recherche'(1984)
- 'les machines de l'intelligence artificielle', J.P.sansonnet, 'la recherche'(1985)