

*Aperçu sur le traitement automatique de la
Langue et Traduction*

(Recherche en cours)

GHENIMI, Abdelouafi

(Université d'Oran Es-Sénia)

La langue est le moyen, peut être, le plus important de communication entre les êtres humains dans toutes situations sociales. C'est une structure d'habitudes et un système de comportements. Avec quelques mots, dans un ordre donné et un ton donné, un individu exprime non seulement une idée donnée mais une situation qui lui est propre et caractéristique.

La traduction est une opération réalisée sur les langues. C'est un processus de substitution d'un texte dans une langue en un texte dans une autre. C'est une opération qui a toujours servi pour faciliter la communication entre les habitants de la planète. En fait, les premières traces de la traduction datent depuis 3000 B.C., durant le Vieux Royaume Egyptien, dans la région du premier Cataracte, Eléphantine, où des inscriptions en deux langues ont été relevées. Elle devint facteur significatif en 300 B.C. quand les Romains s'emparèrent en masses de plusieurs éléments de la culture grecque, notamment la totalité de l'appareil religieux. Au douzième siècle, l'occident prit contact avec l'Islam en Espagne moresque. La situation favorisa les deux conditions essentielles de la traduction

à grande échelle : une différence qualitative dans la culture et le contact continuels entre deux langues. Quand la suprématie musulmane en Espagne s'effondra, l'Ecole de Traduction de Tolède traduisit les versions arabes des ouvrages scientifiques et philosophiques grecques¹.

La traduction en tant que processus est toujours unidirectionnelle. D'une langue d'origine ou source (LS) vers une langue cible (LC). Or, il est possible qu'un texte (LS) dont tous les mots sont traduits ne donne pas, dans l'autre langue (LC) le sens original des idées, alors que la mission de la traduction en fait une priorité. Ceci peut trouver l'explication dans les constituants même de la langue; principalement, la grammaire et le vocabulaire. La valeur sémantique des mots n'est pas la même dans toutes les langues «*le mot est le point de rencontre entre langue, culture et histoire...* ». Comme la structure de la phrase, la syntaxe et la grammaire ne sont pas identiques dans toutes les langues.

Pour y remédier, la traduction se trouve attribuée une nouvelle définition; qui est '*le remplacement du matériel textuel dans une langue (LS) par le matériel textuel équivalent de l'autre (LC)*' [CATFORD65].

Avec le développement exponentiel de la technologie et la tentative de la porter aux pays en voie de développement, la publication simultanée du même livre dans différentes langues, le devenir du transport et de la

¹. Voir Newmark, P. (1982), *Approaches to Translation*, Pergamon Press, Grande Bretagne.

télécommunication, l'élimination des frontières économiques entre différents pays et communautés linguistiques suscitent des besoins croissant de la traduction et lui octroient une importance suprême sur le plan scientifique, académique, politique et stratégique. Les deux dernières décennies de ce siècle ont connu l'avènement de l'ordinateur et des autoroutes de l'information, chose qui élève la quantité de messages échangés à travers le globe à l'exponentiel.

La traduction, par conséquent, ne peut que suivre et prendre des dimensions proportionnelles avec les masses de gens et d'informations qu'elle concerne. Ainsi, la nécessité d'investir dans le sujet et de développer des procédures en mesure de répondre aux nouveaux besoins à toutes les justifications requises se fait de plus en plus sentir.

En effet, il n'est pas étonnant que l'ordinateur qui était principalement à l'origine de la marée informationnelle et communicationnelle se voit sollicité pour apporter des solutions aux problèmes de Traduction. Comment va-t-il agir ? Quelles sont ses capacités et ses performances dans le domaine ? Et quelles sont ses limites ?

Beaucoup d'interrogations auxquelles il serait difficile de répondre mais que nous tenterons de cerner dans cet article.

En outre, l'on ne peut investir dans le sens (meaning), qui est, après tout, la finalité de la traduction, sans avoir réglé définitivement le problème de la forme. Nous sommes

suffisamment convaincu que l'étude de la structure de la langue permet de mieux comprendre le sens dont elle est porteuse, dans le cas où ce sens est accessible. Mais si la compréhension de la structure est fondée sur le sens et la compréhension du sens elle-même étant fondée sur la structure, nous voilà ainsi tombé dans un cercle vicieux, appelé dans la langue de la machine 'dead lock'². D'ailleurs, il y a bien longtemps que des linguistes-anthropologues tels que E. Sapir et L. Bloomfield ont recommandé une séparation entre la description de la forme et celle du sens. Z. Harris et N. Chomsky ont renforcé cette position quelques décennies plus tard par le développement d'une sorte de linguistique formalisée dont les descriptions morpho-syntaxiques ont généré des approches mécaniques de la langue.

L'idée de la Traduction Automatique n'est pas récente. En fait, presque simultanément aux USA et en URSS, les premières réflexions sur le sujet eurent lieu dès la première moitié des années 1950, c'est-à-dire avec l'apparition de la première génération d'ordinateurs d'où elle obtint son nom d'ailleurs, 'première génération de la TA'. A cette époque - la fin de la deuxième guerre mondiale - la progression du nombre de traductions avait pris une allure vertigineuse. Il était donc naturel que l'idée de la TA fasse son chemin. Elle fut l'une des

². 'dead lock' ou 'interblocage' surgit dans la machine quand deux processus interdépendants se trouvent dans une situation où chacun attend le résultat de l'autre pour continuer l'exécution.

premières applications non numériques de l'informatique.

Les efforts furent concentrés, dans un premier temps, sur la confrontation des lexiques. Et les pionniers du domaine ont assimilé le processus de la traduction au décryptage des messages ennemis durant la guerre. Warren Weaver, de la Fondation Rockefeller, conçut l'idée que les langues étrangères pourraient être considérées comme des messages codés dont la traduction pourrait être effectuée selon les mêmes principes que le décryptage. Ainsi, la TA exploitait des dictionnaires bilingues et connut le développement des méthodes de recherche syntaxique. L'opération consistait en deux étapes et semblait devoir livrer les résultats escomptés: le remplacement des mots d'un texte par des mots correspondants, puis la restructuration des phrases produites dans la langue d'arrivée selon les règles syntaxiques de cette dernière. L'espoir attendu fut inestimable. On pouvait espérer qu'une fois les lexiques pertinents établis, les règles syntaxiques et morphologiques respectives mises en mémoire, la conversion d'une langue sera réalisable avec des ordinateurs suffisamment puissants. Cependant, en 1962, Y. Bar Hillel, un de ses pionniers, dressa un diagnostic à ce propos aussi véridique que décevant. Il exposa les limites de la machine qui est «incapable de raisonner comme le cerveau humain car elle ne possède pas son intuition ». En cas de polysémie, elle ne peut trier en

fonction de la situation. Elle ne dispose pas de pré-acquis cognitifs et sémantiques³.

Nous citerons ci-dessous quelques types de traduction pour illustrer l'ampleur du problème auquel doit faire face non seulement le linguiste mais également l'informaticien.

- Traduction entière

Le texte entier est soumis au processus de traduction; c'est à dire, chaque partie du texte (LS) est remplacée par le matériel textuel du (LC).

- Traduction partielle

Des parties du texte (LS) sont laissées non traduites; elles sont tout simplement remplacées par le matériel textuel équivalent du (LC).

- Traduction totale

C'est la traduction au sens habituel du terme dans laquelle tous les niveaux du texte (LS) sont remplacés par le matériel du texte (LC).

Il y a également lieu de mentionner le rang de la traduction. Ce dernier constitue

³ « ... there exist extremely simple sentences in English – and .. for any other natural language – which would be uniquely .. and unambiguously translated into any other language by anyone with a sufficient knowledge of the two languages involved, though I know of no program that would enable a machine to come up with this unique rendering . ».

un autre type de différenciation en traduction qui réfère au rang dans la hiérarchie grammaticale auquel l'équivalence est établie. Dans la traduction normale (totale), les équivalences entre unités grammaticales s'effectuent à tous les rangs. Dans un texte long, la traduction change constamment de rang (mot, groupe, clause, ...). La traduction du mot-à-mot est une traduction à rang limité au mot, il est des cas où elle aboutit à des résultats acceptables, mais ceci arrive rarement:

What time is it ?	→ Quelle heure est-il ?
It is raining cats and dogs.	→ Il pleut chats et chiens.
If I were you,	→ Si *j'étais vous,
My lady in love.	→ Ma dame dans l'amour.

Quant à la traduction littérale, elle aboutit très rarement à des résultats, surtout lorsque (LS) et (LC) sont des langues d'apparemment génétique différent et qu'elles reflètent des cultures différentes :

Never mind about them	→ Peu importe près eux
It is raining cats and dogs.	→ Il pleut des chats et des chiens.
If I were you,	→ Si j'étais toi,

Nous terminerons cet article préliminaire sur l'état de notre recherche en Magister par quelques propos sur les problèmes de la traduction automatique et la langue Arabe. L'arabe écrit est généralement l'Arabe

Moderne Standard (AMS). Il est utilisé dans le Monde Arabe comme langue de communication dans les discours officiels, l'enseignement, la littérature entre autres. C'est une continuité linguistique de l'Arabe Classique.

La langue arabe est une des langues les plus difficiles à traiter par ordinateur. Par son jeu de caractères et sa richesse morphologique et syntaxique, elle constitue le produit de multiples et complexes problèmes lors de l'implémentation de l'analyseur morphologique. L'un des problèmes étant lié à l'ambiguïté dans l'attribution des informations linguistiques adéquates et la détection de la forme originare. L'écriture de l'arabe est consonantique et elle ne nécessite pas de vocalisation. La plus grande partie des problèmes liés à la reconnaissance de l'écriture de cette langue est essentiellement due à l'absence des voyelles et des signes de gémiation (la chedda ّ) dans les mots.

Citons comme exemples les cas suivants :

1. Ambiguïtés dues à l'absence de vocalisation:

Soit le mot non vocalisé: 'علم'. Hors contexte, ce mot peut vouloir dire :

'عِلْمٌ'	[ʔilmun]	«un savoir»
'عَلَمٌ'	[ʔalamun]	« un monde »
'عَلَّمَ'	[ʔallama]	« il a enseigné »
'عُلِّمَ'	[ʔullima]	"il a été enseigné"

'عَلِمَ'	[ʔulima]	« on a su »
'عَلِّم'	[ʔallim]	« Enseigne ! »(Imperatif)
'عَلِمَ'	[ʔalima]	« il a su »

Certaines ambiguïtés dues à l'absence de la gémination⁴. Celle-ci peut changer la signification des verbes, transformer un nom en un verbe et vice versa, comme dans les cas suivants:

'عَبَرَ'	[ʔabara]	« traverser »
'عَبَّرَ'	[ʔabbara]	« exprimer »
'صَلَّبَ'	[ʃalaba]	« crucifier »
'صَلَّبَ'	[ʃallaba]	« endurcir »

Finalement, il y a lieu également de mentionner les problèmes dus à l'agglutination qui en fait n'est pas caractéristique de la langue arabe comme c'est le cas de la langue turque par exemple. Néanmoins, le problème, réside également dans le phénomène d'agglutination⁵ de préfixes et suffixes qui rendent les frontières de mots

⁴. La gémination entraîne le renforcement de l'articulation et l'accentuation des propriétés de la consonne. En arabe, elle a une fonction différenciative. Elle joue, par ailleurs, un rôle structural dans le développement morphologique du nom et du verbe.

⁵. L'agglutination est la juxtaposition au radical d'affixes distincts pour exprimer des rapports syntaxiques.

invisibles et non marquées par la présence d'un séparateur. Citons comme exemple le cas de:

'يتحاربونهم' [jataħa :rabu :nahum] «ils les combattent»

Un mot arabe peut correspondre à une véritable phrase constituée d'un sujet, d'un verbe et le plus souvent d'un ou de plusieurs compléments. En arabe, un mot est le noyau principal complété par les marques de genre, de nombre, de temps, de mode et de diverses unités minimales significatives de type pronom, article, etc...

Citons comme exemple : أتنذروننا , qui se traduit par 'Est ce que vous vous souvenez de nous ?'

La langue arabe est caractérisée par des traits grammaticaux et des structures syntaxiques propres à elle. La théorie de l'arabe et les structures de base sont bien connues et définies dans les grammaires scolaires et surtout dans le remarquable travail du grammairien du 8^{ème} siècle, SIBAWAYH. Ces structures doivent être respectées lors de la traduction et cela indépendamment de la structure de la phrase dans la langue à partir de laquelle s'effectue la traduction. Le passif n'est traduit en passif que lorsque le sujet est inconnu. Les phrases en 'there is...', largement employées en Anglais, peuvent nécessiter l'emploi de phrases nominales en arabe (مبتدأ و خبر) et cela quand le sens est focalisé sur le nom, par exemple:

Anglais : 'there is a book on the table'



Arabe : 'على المنضدة كتاب' [ʔala l minɖadati kita:bun]

Comme elles peuvent nécessiter l'emploi de phrases verbales quand le sens est focalisé sur le verbe, par exemple:

Anglais : "there was quietness in the city after the police dispersed the protestors"



Arabe : 'ساد المدينة الهدوء بعد قيام الشرطة بتفريق المتظاهرين'
(Voir Annani, 1996)

Références

LYONS, John (1983) *New Horizons in Linguistics*
Penguin books, G.B. , pp. 215-228

GLEASON, H.A.(1988) *Introduction à la linguistique*
Larousse, Paris, 1969.

CATFORD, J.C.(1965)*A Linguistic Theory of Translation*
1st pub, 4th imp 1974, GB.

FROMKIN, Victoria (1988) *An introduction to Language*
& Robert Rodman, », The Dryden Press.

LEDERER M. (1994) "La traduction automatique dans la perspective de la traduction humaine" *La traduction aujourd'hui*, Hachette FLE, Paris.

CARRE, René, DEGREMONT, J.F. GROSS, M. PIERREL, J.M., SABAH, Gérard (1991) Le traitement de l'écrit » *Language humain et machine*, Presse du CNRS, Paris.

Sabah, G. (1988)*L'intelligence artificielle et le langage*
Hermes.