

Pour une autonomie des recherches en traduction automatique

«The starting point is the realization that translation is impossible in theory, if not in practice.»

(Punya Sloka RAY, Yale Un., «A Philosophy of Translation», in BABEL, VIII, 1962-4, p. 182)

Il serait superflu, dans les colonnes de cette revue, de s'étendre sur l'actualité du problème de la traduction. L'économie, la politique, la science ne se conçoivent plus qu'à l'échelle internationale. Cette évolution requiert des échanges d'information toujours croissants et il est patent que la disparité des langues ne constitue pas le moindre des handicaps au développement des relations humaines. Les traducteurs le savent bien: le moment viendra bientôt (s'il n'est déjà venu) où l'énorme production¹⁾ de livres et d'articles dépassera la mesure d'un travail artisanal et exigera la mise en œuvre de techniques nouvelles.

Or, depuis une quinzaine d'années, les procédés de traitement automatique ont connu un développement considérable. D'une manière générale, ils se caractérisent, comme l'a montré l'article précédent, par leur capacité de traiter à grande vitesse des masses importantes d'information.

Historiquement, l'apparition de la traduction automatique s'explique par la conjonction de ces deux facteurs: besoins accrus en traduction et avènement de l'ordinateur électronique.²⁾

Ainsi, l'idée de la traduction automatique est née d'une nécessité pratique. Et c'est dans un esprit pratique, utilitaire, qu'ont été menées les premières investigations en ce domaine. On n'a cependant pas tardé à constater que la langue était un objet d'analyse beaucoup trop souple, trop nuancé, en un mot trop humanisé, pour qu'on le traitât impunément avec

si peu d'égards, à la manière dont on aborde l'étude des phénomènes physiques.

L'idée, à peine éclosée, allait-elle faire faillite, faute de fondements théoriques? Les théoriciens ne pouvaient manquer d'apporter tous leurs soins à ce nouveau-né, qu'ils n'avaient pas enfanté, sans doute, mais qui leur offrirait certainement l'occasion d'élaborer de nouvelles théories. Chacun éleva le bébé, selon ses us et moyens, à l'image et ressemblance de sa propre discipline: le linguiste y alla d'une nouvelle systématique à laquelle il donna parfois un petit air de structuralisme (fort à la mode); le mathématicien convertit les nuances de l'expression linguistique en formules (en algorithmes, pardon), ne faisant ainsi ni de la linguistique ni sans doute en définitive des mathématiques³⁾; le logicien, enfin, argumenta... Quant au bébé, il semble devoir rester longtemps encore au stade infantile.

Nous voudrions, dans les lignes qui suivent, développer quelques considérations (ou rappeler quelques évidences!) au sujet des moyens que sont les ordinateurs et des problèmes théoriques fondamentaux de la traduction; nous énoncerons enfin quelques principes méthodologiques qui en découlent.

1. L'ordinateur, un moyen limité.

Un ordinateur se conforme à un programme de travail, ensemble d'instructions codées⁴⁾ dont la séquence a été préalablement déterminée. L'ordinateur ne prend donc au-

cune initiative (au sens humain du terme); il n'apprend pas davantage⁹). L'ordinateur étant un exécutant, La Palice l'eût dit, il convient de n'attendre de lui qu'un travail d'exécution.

2. Problèmes théoriques.

Etant donné les possibilités actuelles des ordinateurs, et sans préjuger des pouvoirs nouveaux que pourrait leur conférer, dans l'avenir, l'ingéniosité humaine, quels sont, sur le plan théorique, les problèmes que soulève à présent l'automatisation de la traduction?

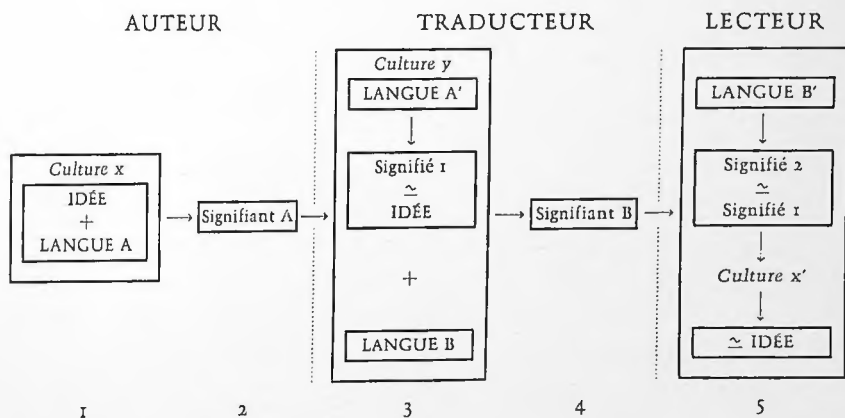
La traduction automatique est souvent présentée comme une simulation de la traduction humaine. Assertion un peu hâtive car, si ces deux opérations visent, dans le principe, le même but, à savoir la production d'un texte traduit, elles se déroulent selon des processus bien différents. C'est qu'en effet leurs moyens diffèrent puisque la totalité de l'expérience humaine n'est pas transmissible à un ordinateur sous forme de programme.

La pensée et la langue sont étroitement liées, celle-ci n'est cependant pas l'expression totale, parfaite et univoque de celle-là. Parce qu'elle procède de la personne humaine, la parole, langue actualisée, s'élabore selon des modes très diversifiés. Libre dans les limites de la langue, lieu commun théorique de toutes les paroles, elle s'oppose par son protéisme à la rigueur de la formulation mathématique, essentiellement artificielle et univoque. Cette inadéquation de la pensée et de la parole explique, comme nous allons le voir, toutes les difficultés de la traduction.

Qu'est-ce que traduire? C'est prendre connaissance d'un signifiant de langue A, en dégager le signifié et transposer ce dernier en un signifiant de langue B:

Signifiant A → Signifié → Signifiant B.

On le voit: le nœud de la démarche, c'est le signifié. Examinons la séquence de plus près et intégrons-la dans un schéma qui va de l'auteur du texte écrit en langue A à la personne qui en lit la traduction en langue B.



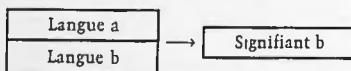
Au départ, un auteur qui dispose d'une culture x et d'une idée qu'il exprime dans sa langue A (d'un niveau déterminé). Sa démarche aboutit à la rédaction d'un texte, le signifiant A⁰. L'auteur n'écrivant pas à l'in-

tention du traducteur, celui-ci ne reçoit du message transmis qu'un signifié amoindri et peut-être déformé. Ne disposant pas nécessairement de la culture requise du lecteur, il ne peut avoir une compréhension globale du

signifiant A. Néanmoins, pour autant que la langue A de l'auteur soit suffisamment correcte, c'est-à-dire représentative de la moyenne des langues des locuteurs A, le traducteur arrivera à saisir les relations syntagmatiques fonctionnelles (groupe du sujet, centre de ce groupe, etc.). Il transposera ensuite, avec une fidélité variable, le signifié 1 en un nouveau texte, le signifiant B. De ce dernier se dégagera pour le lecteur un signifié 2, que sa culture x', voisine de celle de l'auteur, et le macro-contexte aideront à rapprocher de l'idée.⁷⁾

Comparons maintenant à ce processus celui de la traduction automatique. Les différences, on s'en doute, interviendront au niveau 3. Celui-ci se trouvera amputé, faute de moyens appropriés, de son élément central, le signifié. L'automate étant dépourvu de culture, reste la connaissance de la langue. Mais, sur ce plan même, les relations syntagmatiques ne pourront être découvertes que par la seule analyse de la Forme. Or, les langues A et B n'étant décrites que très imparfaitement, le seul outil dont dispose l'automate est lui-même incomparablement moins adapté que son correspondant humain. Enfin, l'automate n'aura pas du texte une approche globale; sa méthode est essentiellement analytique, à la différence de l'appréhension synthétique, de la compréhension (*cum-prehendere*) — même relative — du signifiant par le traducteur.⁸⁾ L'élément central de notre schéma se présentera donc de la manière suivante:

AUTOMATE



3. Principes méthodologiques.

Ces principes ne sont que les corollaires des considérations qui viennent d'être développées.

1° L'ordinateur n'ayant pas accès au signifié il faut avoir constamment à l'esprit que tout est dans le signifiant écrit.

2° Le niveau d'une véritable sémantique, globale et intégrée dans une culture et une

situation (entités extra-linguistiques) est donc, actuellement, inaccessible. On n'atteint pas le signifié, mais une classification notionnelle préalable (humain, animal, végétal; animé, inanimé, etc.), qui semble d'ailleurs donner des résultats encourageants, permet sans doute une représentation sommaire, une formulation partielle du niveau abstrait des faits de conscience sémantiques.

3° Les travaux de morpho-syntaxe et de sémantique, quoique fondés sur une approche linguistique des problèmes, gagneraient sans doute à ne pas s'embarasser de questions non pertinentes. En recherche pure, la curiosité peut être licite, voire souhaitable. En recherche appliquée, une fois un projet arrêté, toute recherche qui ne tend pas à la réalisation du projet encourt le reproche d'un écart, d'une digression. N'est-ce pas poursuivre une chimère de poser, comme on le fait parfois, des questions du genre: «Ce texte a-t-il un sens?»? Tout texte à traduire est en effet censé significatif mais il est une question qui se pose réellement dans l'analyse d'un texte à traduire automatiquement: «Telle analyse du texte est à rejeter; pourquoi la machine l'a-t-elle retenue? Y a-t-il dans le contexte un élément formel qui permettrait de lever l'ambiguïté?»

Un exemple encore. Savoir quel sens peut avoir, dans un contexte déterminé, une phrase (écrite) comme: «Bel exploit! Je vous félicite» n'est pas affaire de traduction automatique — ni peut-être de traduction. A quoi bon s'attarder à ce problème puisque, l'antiphrase étant un procédé d'expression généralisé, le sens sera transféré, à l'intention d'un lecteur supposé intelligent, par le simple passage direct du signifiant A au signifiant B!

Ce principe d'action et de rendement qui préside aux recherches de sciences appliquées est, à notre sens, un facteur important de leur progression.

4° C'est une vue de l'esprit que de différer les réalisations jusqu'à ce qu'une théorie satisfaisante soit complètement élaborée.⁹⁾ L'histoire de la linguistique, d'ailleurs, comme celle de la philosophie, n'évoque-t-elle pas un mouvement de flux et de reflux, de grands esprits rejetant ce que d'autres grands esprits avaient conçu?¹⁰⁾

La méthode expérimentale, en linguistique appliquée, celle qui consiste à voir le rendement d'un procédé, peut donner des résultats appréciables.¹¹⁾ D'autre part, à l'inverse de la mentalité exigeante qui caractérise souvent la linguistique pure, la linguistique appliquée considère modestement comme encourageants les résultats positifs d'une hypothèse, même si celle-ci ne rend compte que partiellement des éléments du discours (excluant, par exemple, les cas d'ambiguïtés insolubles sans le recours à la situation ou au macro-contexte¹²⁾, le cas d'ambiguïtés voulues, les jeux de mots, etc. ...) et ne fournit pas une traduction en tous points conforme aux impératifs syntaxiques propres à la langue de sortie.

5° La question de savoir si, conçues de telle manière, les recherches en traduction automatique relèveraient ou non de la linguistique, nous semble, dans ces conditions, parfaitement académique. La linguistique serait

plutôt une science auxiliaire. Et il faudrait se garder tout autant de confondre linguistique appliquée et linguistique structurale, ces disciplines offrent, il est vrai, plus d'une analogie puisque toutes deux considèrent la langue comme un système formel. Mais la première se différencie nettement de la seconde par son caractère pragmatique (étude exclusivement synchronique des langues; critère d'efficacité incompatible avec un esprit de système qui modèlerait le réel et tenterait de le plier à ses structures, voire à ses a priori).

Ceci dit, étant donné la manière nouvelle — strictement objective — dont procéderait la linguistique appliquée à la traduction automatique, il est vraisemblable qu'elle jetterait quelque lumière sur le phénomène linguistique. Cet apport serait, somme toute, accidentel et comme un sous-produit de la recherche appliquée.

¹⁾ On devrait sans doute écrire: surproduction. C'est un fait bien connu des chercheurs que les publications scientifiques présentent un caractère de redondance très marqué. A l'accroissement de la chose imprimée ne correspond pas — tant s'en fait — un accroissement proportionnel de la connaissance. Notre propos n'est pas ici d'examiner les causes de cette débauche imprimée. Disons seulement que l'un des facteurs responsables en est probablement l'habitude (facile mais nécessaire, nécessaire mais facile ...) qu'ont parfois les organismes de recherche de mesurer la qualité de leurs agents à la quantité de publications de ces derniers. Ajoutons que pareille surechère ne laisse pas de se révéler, en fin de compte, très préjudiciable à l'équilibre même du travail de recherche. Ce serait le moment de rappeler ce qu'écrivait [déjà!] Leibniz en parlant de la recherche scientifique: "... quand je vois ... le peu de concert des desseins, les routes opposées qu'on tient, l'animosité que les uns font paraître contre les autres, j'appréhende que nous ne soyons pour demeurer longtemps dans la confusion et dans l'indigence où nous sommes par notre faute. Je crains même qu'on ne se dégoûte et que par un désespoir fatal les hommes ne retombent dans la barbarie. A quoi cette horrible masse de livres, qui va toujours augmentant, pourrait contribuer beaucoup." Et ce savant ajoutait: "Il ne s'agit pas toujours de faire de grands ouvrages, si chacun ne donnait qu'une grande découverte, nous y gagnerions beaucoup en peu de temps. Une seule remarque ou démonstration de conséquence suffit pour s'immortaliser et pour se faire un mérite auprès de la postérité." REIBNIZ, Fragment tiré des manuscrits conservés à la Bibliothèque de Hanovre, cité par Gaston LAURENT, *Les Grands Ecrivains Scientifiques*, 2^e édition, A. Colin, 1930 (p. 96). Ces remarques, énoncées aujourd'hui, seraient jugées bien naïves ...

²⁾ Il est significatif, à cet égard, de constater que, venue à une époque où l'électronique en était encore à ses débuts, l'invention par Smirnov-Trojanski d'une « machine à traduire » n'ait pas eu de lendemain [1933].

³⁾ "The symbolic nature of language is probably responsible for the widely held but erroneous view that linguistics is a branch of mathematics: a sort of symbols 'looks like' a mathematical formula". David HAYS, « Research procedures in machine translation », in *Natural Language and the Computer*, edited by Paul L. GARVIN, McGraw Hill, 1963 (p. 183). — Les mathématiques, système formel par excellence, procèdent essentiellement d'une démarche déductive; la linguistique, elle, observe d'abord un phénomène, le langage; elle est à la fois inductive et déductive. [Voir, sur

ce dernier point, dans l'ouvrage cité, l'article de Paul L. GARVIN et W. KAROSH, «Linguistics, data processing and mathematics.»]

⁴⁾ L'utilisation de langages symboliques (Fortran, Cobol, Algol, Comit ...) peut faciliter la tâche du programmeur, elle ne modifie en rien le principe d'une totale sujétion de la machine.

⁵⁾ Des travaux sont actuellement en cours en vue de doter les ordinateurs d'une « faculté » d'apprentissage. S'il devait aboutir, il est hautement probable que la physiologie du problème s'en trouverait fort modifiée puisqu'à ce moment la machine serait en mesure de se créer des sortes de réseaux mémoriels, de simuler plus ou moins fidèlement certains processus humains d'analyse et d'intégrer des expériences successives.

⁶⁾ Cette distinction entre culture, idée et langue est évidemment méthodologique et pratique. Tous ces éléments apparaissent bien comme solidaires, voire confondus, dans la même personne, on ne peut les considérer objectivement comme des entités discrètes.

⁷⁾ A propos du travail permanent de synthèse que requiert implicitement toute lecture, cf. notre article: «Aspects linguistiques de la traduction automatique des langues», *Revue des Langues vivantes*, XXX, 1964, 2, pp. 140-153 [pp. 144-145].

⁸⁾ Observons ici que la grande imperfection des descriptions des langues n'empêche pas les locuteurs de pratiquer ces dernières, le plus souvent d'ailleurs sans la moindre référence aux catégories grammaticales. C'est toute la différence avec l'automate.

⁹⁾ Ceci nous rappelle la réflexion d'un professeur de grammaire, émise, lors d'un récent colloque, à l'adresse de collègues théoriciens: «Doit-on attendre de connaître les lois de l'équilibre pour rouler à bicyclette?»

¹⁰⁾ Une question analogue se pose dans l'enseignement de la grammaire, trop souvent soumis aux fluctuations des théories.

¹¹⁾ Qu'on nous permette de mentionner ici une expérience personnelle, au demeurant fort simple et tout à fait limitée. — Un des problèmes de plus épineux que soulève la réalisation de la traduction automatique est certes la levée des polysémies et homographies. Nous plaçant au niveau le moins complexe, celui des ambiguïtés morpho-lexicales, nous avons converti un texte d'une dizaine de pages en une séquence de natures grammaticales. L'ordina-