

Séminaire Sogesci — Sogesci Seminarie

(suite - vervolg)

ASPECTS LINGUISTIQUES DE LA TRADUCTION AUTOMATIQUE DES LANGUES

par M.-O. HOUZIAUX

Université de Liège

Position du problème.

Parmi les faits de civilisation, la langue occupe une place privilégiée : véritable véhicule de la pensée — au sens le plus large du terme — elle constitue le principal moyen de communication entre les hommes. Sa raison d'être est de faciliter le rapprochement et la collaboration des individus, de contribuer à la consolidation des liens de toutes espèces qui les unissent : religieux, moraux, intellectuels, affectifs, politiques et économiques.

Mais si, de tous temps, le fait linguistique a joué, au sein des groupes unilingues, un rôle éminemment social, de tous temps aussi, la diversité, la multiplicité des langues s'est révélée un sérieux handicap à l'entente et à la compréhension des groupes ethniques dans leurs rapports mutuels.¹

Depuis la seconde guerre mondiale, la primauté de l'anglais continue en s'affirmant à faciliter les relations internationales, tant sur les plans scientifique et culturel que dans le domaine de la vie politique et économique. Mais l'Histoire nous a appris la précarité relative de semblables privilèges et nul n'ignore que la langue française a perdu une part non négligeable de

¹ Il n'est sans doute pas inutile de rappeler ici que, depuis l'Antiquité, de nombreux efforts ont été accomplis en vue de faciliter la communication entre les peuples de parlers différents, depuis l'usage des langues mortes jusqu'aux récents travaux sur la traduction automatique.

Ce sont les lettrés qui, les premiers, ont eu l'idée de se servir d'un idiome commun qui leur permit d'échanger leurs idées et de confronter leurs opinions dans le but d'accroître le savoir humain. Faut-il s'étonner, dès lors, du caractère essentiellement savant des langues qu'ils employèrent : langues mortes ou langues artificielles ? Issu d'un dialecte populaire, le sanscrit védique, puis classique, était déjà, par sa régularité et sa logique, une langue artificielle. En Europe occidentale, l'usage généralisé du latin comme langue de culture a joué un rôle considérable dans l'épanouissement de la civilisation gréco-romaine et dans l'expansion du christianisme. On sait qu'en France, par exemple, le latin resta jusqu'au XVI^e siècle la langue judiciaire et celle des actes publics. Il fallut d'autre part attendre le *Discours de la Méthode* (1637) pour que la langue française trouvât progressivement droit de cité dans les ouvrages philosophiques, scientifiques ou scolaires. De nos jours encore, l'usage du latin peut être considéré comme un facteur important de l'unité doctrinale de l'Eglise catholique.

A la fin du XIX^e siècle, on vit apparaître une langue nouvelle, entièrement artificielle : le volapük. Cette invention, due à l'Allemand Johann-Martin Schleyer, se voulait de portée universelle; elle devait d'ailleurs connaître un certain succès puisque, vers 1889, vingt-trois journaux étaient édités en volapük et quelque trois cents « volapüksklubs » se vouaient à la propagation de ce parler original. Peu de temps après, ce dernier était supplanté par un rival *ejusdem farinae*, le fameux espéranto du médecin russe Zamenhof. L'espéranto a suscité un intérêt appréciable mais on ne le connaît plus guère aujourd'hui que de nom.

la très large audience dont elle jouissait naguère encore, même dans les pays slaves et germaniques. D'autre part, l'adoption de l'anglais comme langue commune pour les échanges internationaux, les congrès, etc., si elle recueille tous les suffrages des anglophones, est loin de bénéficier de l'unanimité parmi les autres groupes linguistiques, naturellement attachés à leur parler national, part importante de leur patrimoine culturel, et qu'ils voudraient inviolable.²

Ces dernières années enfin, de nombreuses initiatives ont été prises dans le but de favoriser les relations entre les pays de langues différentes : intensification de l'enseignement des langues vivantes au niveau scolaire, développement de méthodes nouvelles d'apprentissage des langues, publication de répertoires multilingues. Marquée par la disparition progressive du concept politique de nationalité, notre époque, plus que toute autre, ressent ainsi le besoin impérieux de trouver une solution à ce problème de la diversité des langues.

Plus particulièrement dans le domaine de la science, les chercheurs ne peuvent souvent s'informer en temps utile des progrès réalisés dans leur discipline à l'étranger, faute de disposer de traductions. Selon la revue *International Electronics*,³ on estime à 250.000.000 de mots l'importance des textes publiés mensuellement en Union soviétique, dont il conviendrait que les agences gouvernementales et les hommes de science américains possédassent la traduction pour se tenir au courant de l'activité scientifique en U.R.S.S. Les équipes de traducteurs russe-anglais constituées aux Etats-Unis ne peuvent travailler à un rythme suffisant pour répondre à des besoins de cet ordre et par ailleurs toujours croissants.

On le voit, il s'agit d'abord, comme le fait remarquer Delavenay⁴, d'un problème de vitesse et de quantité. Aussi eût-il été étonnant qu'on ne songeât point à le résoudre en exploitant les possibilités nouvelles offertes par l'automatisme et les machines à traiter l'information.⁵

² Sous l'influence de l'anglais, la langue française ne cesse de s'appauvrir... d'emprunts et de se transformer, selon le mot d'Etiemble, en un horrible « franglais », que le Professeur en Sorbonne appelle aussi « sabir atlantique ».

³ « Computers Speed Language Translation », juin 1962, pp. 16 et sv. Voir aussi : Paul W. HOWERTON, « The Parameters of an Operational Machine Translation System », in MT, vol. 6 (1961), p. 108 et sv.

⁴ E. DELAVENAY, *La machine à traduire*, Coll. « Que sais-je ? », n° 834, Presses universitaires de France, 1959, p. 9.

⁵ A côté des études sur la traduction automatique des langues, des recherches ont été récemment entreprises dans le domaine de la documentation automatique. Les deux orientations sont semblables en apparence mais, quoiqu'elles puissent avoir des méthodes communes, il faut se garder de les confondre :

— la traduction automatique (*automatic translation*), dont l'objet final ne se différencie pas de celui de la traduction du type traditionnel, recourt à des méthodes qu'on pourrait appeler de « linguistique appliquée »;

— la documentation automatique (*information retrieval*), qui a pour but de faciliter les recherches bibliographiques ou analogues, fait principalement appel aux procédés de la statistique.

Il ne sera pas question ici de cette dernière orientation.

Bref historique de la traduction automatique des langues (T.A.L.)

Afin de mettre en lumière les développements considérables et rapides pris par les études sur la T.A.L., nous rappellerons ici quelles en furent les étapes principales.

Si l'on exclut l'invention du Russe Smirnov-Trojansky, déposée à Moscou en 1933 et qui n'eut pas de suite, l'idée de recourir à l'automatisme pour effectuer des travaux de traduction remonte à 1946 et est due au Dr Sc. Andrew Booth, de l'Université de Londres. Il suggéra à Warren Weaver, vice-président de la Fondation Rockefeller, d'utiliser les calculatrices électroniques à des fins de traduction. Tandis que Booth entrevoyait la solution d'une traduction mot à mot, rendue possible, croyait-il, par la très grande capacité des mémoires des calculatrices, Weaver tentait d'appliquer les méthodes de décodage employées par les services secrets de l'armée pendant la dernière guerre. Avec la collaboration de Britten, Booth élaborait les « instructions » permettant à une calculatrice électronique de consulter et d'exploiter un dictionnaire automatique en vue d'une traduction mot à mot.

En 1948, l'Anglais Richens apporte à la méthode de Booth une première amélioration : celle de l'analyse automatique des désinences, qui permet de réduire considérablement le volume du lexique et, partant, d'en accroître la vitesse de consultation. L'année suivante, Weaver propose la résolution des problèmes d'ambiguïté sémantique par l'exploration du contexte. En 1950, Reifler lance l'idée de la préparation des textes en vue de leur traduction automatique (pré- et post-édition); Oswald et Fletcher étudient la syntaxe allemande au point de vue de la T.A.L., tandis qu'au M.I.T. le logicien Bar-Hillel entreprend des recherches sur la structure des langues. C'est en 1952 que, sous le patronage de la Fondation Rockefeller et convoquée par le M.I.T., se tient une première conférence sur le sujet, à laquelle participent des linguistes et des ingénieurs. Deux ans plus tard, Dostert et Garvin, de l'Université de Georgetown, réalisent une première expérience de traduction automatique russe-français, utilisant six règles de syntaxe et un lexique de deux fois 250 mots.

Fin 1955, des expériences sont effectuées à l'aide de la calculatrice B.E.S.M. de l'Académie des Sciences d'U.R.S.S.; elles portent sur la traduction en russe de textes techniques et scientifiques rédigés en anglais. La partie linguistique des travaux est confiée à I.K. Bel'skaya, tandis que les problèmes de programmation sont étudiés par I.S. Mukhin, L.N. Korolev, S.N. Razumovsky, N.P. Trifonov et G.P. Zelenkevich.

Depuis lors, les recherches n'ont cessé de se multiplier dans des sens divers, allant des études théoriques relatives à certains systèmes linguistiques, à la tentative de confier aux machines la découverte, du moins partielle, des mécanismes de la langue.

Problèmes linguistiques.

Nous nous attacherons maintenant à décrire les principaux problèmes que pose, sur le plan linguistique, la réalisation de la T.A.L. S'il nous arrive de mettre assez souvent l'accent sur la difficulté de ces problèmes, c'est qu'il appartient au linguiste de les relever et de les définir afin que des solutions valables puissent leur être apportées.

1. — *La langue, phénomène social et intellectuel.*

Fondamentalement, la langue est un phénomène intellectuel puisqu'elle vise à l'expression de la pensée. Sa formation résulte non seulement d'une convention sociale plus ou moins stricte, mais en premier lieu d'un processus psychologique d'objectivation. C'est cette faculté d'objectivation de l'esprit qui fait du soliloque une démarche raisonnable et proprement humaine. Parole et activité intellectuelle sont deux phénomènes distincts mais étroitement solidaires, certains disent même indissociables. Si une idée trouve difficilement une expression entièrement satisfaisante dans la parole, c'est-à-dire dans la langue actualisée, elle ne prend cependant corps que par l'expression de la langue, principale ressource du langage. C'est ce qu'exprime fort bien M. Dessaintes⁶ lorsqu'il écrit : « Sans doute y a-t-il toujours un élément logique dans la phrase; celle-ci prend sa source dans la pensée, que la pensée soit la perception de la réalité concrète ou la représentation abstraite. Les schèmes conceptuels se rangent en catégories de pensée, et celles-ci doivent s'intégrer tant bien que mal dans des catégories et des mécanismes grammaticaux. Mais la pensée n'aboutit qu'une fois informée par la langue, et la langue n'a de raison d'être que comme expression et véhicule de la pensée.

» A notre avis, poursuit le même auteur, une des raisons essentielles pour infirmer le parallélisme logico-psycho-grammatical réside dans le fait suivant, et que l'on n'a pas suffisamment mis en lumière : la pensée (états de conscience, états d'âme...) est non seulement intuitive et globale, mais intemporelle. Or, pour s'exprimer, elle doit se soumettre à la linéarité du discours, c'est-à-dire à des contingences temporelles et matérielles. La pensée se déforme toujours quelque peu en entrant dans les structures linguistiques. »

Comme l'usage d'une langue, sa compréhension implique donc plus qu'une simple convention; elle suppose l'intervention, le concours actif de l'intelligence. Or, jusqu'à nouvel ordre, l'intelligence doit être considérée comme une faculté proprement humaine. Cette distinction entre la pensée et ce que, par une trop facile analogie, on appelle parfois « cerveau électronique », il convient de l'avoir constamment à l'esprit lorsqu'on aborde les

⁶ *Eléments de linguistique descriptive*, La Procure, Namur-Bruxelles 1960, pp. 27 et 28.

problèmes de traduction automatique, si l'on veut éviter de verser dans la fiction.

La langue peut se définir comme un moyen d'expression constitué de signes — parlés ou écrits — groupés selon des règles particulières d'assemblage. En d'autres termes, vocabulaire, morphologie et syntaxe sont les composants matériels de la langue. La traduction réside essentiellement dans l'analyse d'un texte de langue A pour y découvrir un signifiant, support d'un signifié, et la restitution de ce signifié par la voie du signifiant correspondant dans une langue B. Cette opération n'est simple qu'en apparence. Tout discours, en effet, comprend, outre ce que nous avons appelé les composants matériels, des éléments implicites, sous-entendus ou signifiés globalement. Le discours n'a de cohérence et de raison d'être que par les synthèses successives et la plupart du temps inexprimées dont ses parties font l'objet. Il importe de remarquer en outre que les éléments de ces synthèses sont souvent extra-linguistiques : mémoire, connaissance du sujet ou de circonstances diverses, vie affective, activité sensorielle, etc.

Ainsi, lorsqu'un individu parle ou écrit, il ne fait pas uniquement appel à la connaissance qu'a de la langue utilisée son interlocuteur ou son lecteur; il doit aussi pouvoir compter sur les facultés mentales de ce dernier, et notamment sur son imagination créatrice, capable de suppléer aux carences de l'expression et de résoudre les ambiguïtés sémantiques. Illustrons ces considérations de quelques exemples :

L'agitateur est (fut) arrêté.

Nous aperçûmes aussi les *locataires* de nos *maisons*; ces dernières venaient à notre rencontre.

Nous aperçûmes aussi les maisons de nos *locataires*; ces dernières venaient à notre rencontre.

Je vois *bien* qu'on ne m'approuve pas (... ne m'approuvera pas...; je n'aie pas mes lunettes).

Un homme *lâche la frappe* (la bride, la bascule dans le vide).

Fiez-vous aux femmes !

De telles « difficultés » sont plus fréquentes qu'on ne serait tenté de le croire à première vue. Il suffit de s'imposer la lecture attentive de quelques pages d'un texte courant pour se rendre compte que la parole (langue actualisée) recèle bien des ambiguïtés d'ordre grammatical, que nous résolvons

grâce à notre faculté de comprendre, au sens général et étymologique du terme (*cum-prebendere*, prendre ensemble, saisir globalement).⁷

La grammaire ne suffit donc pas à lever certaines ambiguïtés sémantiques. *A fortiori*, la solution d'une traduction mot à mot ne peut-elle être considérée que comme tout à fait provisoire, même si, d'un point de vue pratique, elle n'est peut-être pas dénuée d'intérêt. Quant à l'idée d'assimiler la traduction à une opération de décodage, elle est aujourd'hui abandonnée : « Dans le codage et le décodage, écrit à ce sujet Panov ⁸, nous ne changeons que la forme extérieure des mots, sans toucher au langage lui-même, et il est normal qu'un travail de ce genre s'accommode fort bien des méthodes formelles. Mais, dans la traduction, c'est le langage lui-même que nous modifions, c'est-à-dire le système entier de la pensée, système hautement complexe et nuancé, élaboré au cours des âges par chaque nation et qui est étroitement solidaire de la manière de penser et de vivre, et de l'histoire du peuple qui l'emploie. »

2. — *Le processus général de la traduction automatique.*

Ce processus comprend trois grandes étapes :

- analyse du texte à traduire, écrit en langue A (dite langue d'entrée ou langue source);
- conversion éventuelle des éléments d'analyse en éléments de synthèse;
- synthèse des éléments dans la langue B (dite langue de sortie ou langue cible).

La réalisation des première et troisième étapes exige une étude approfondie de la langue sur les plans du lexique, de la morphologie et de la syntaxe, le domaine de la sémantique s'étendant à tout le système linguistique. La deuxième étape, celle de la conversion, est requise dans tous les cas où il y a divergence entre les langues A et B quant aux moyens d'expression (subjonctif à rendre par un conditionnel, ou un impératif; élément à omettre ou à ajouter, dans le cas de l'article, par exemple, etc.).

⁷ Sans songer aux problèmes de traduction automatique, Ch. Bally écrivait déjà (*Linguistique générale et linguistique française*, p. 27) : « Pourquoi ne ferait-on pas une enquête complète sur les déficits de la langue normale ? Il vaudrait la peine de noter, au moment où elles surgissent, les fautes que l'on commet en parlant et en écrivant *correctement*. Ces notations, centralisées et commentées, constitueraient un riche matériel d'observation. On verrait que la pathologie de la langue correcte s'étend au lexique, à la syntaxe, à la morphologie, au système phonologique. On ne s'étonnera pas que, malgré l'indigence de la documentation, nous fassions çà et là quelques incursions dans ce domaine. »

⁸ P. Yu. PANOVA, *Automatic Translation*, Pergamon Press, 1960, pp. 51-52.

Les recherches relatives à l'analyse et celles qui concernent la synthèse comportent évidemment bien des points communs puisqu'elles tentent toutes deux à « programmer » des faits linguistiques, c'est-à-dire à découvrir ce qui, dans la langue, est susceptible de systématisation. Ces études sont même solidaires, les besoins des unes ne pouvant se définir qu'à la lumière des exigences des autres. Ainsi, dans le cas d'une traduction du type $A \rightarrow B$, l'analyse devra réunir, en vue de la synthèse, tous les éléments utiles contenus dans le texte d'entrée.

Analyse et synthèse ont cependant leurs problèmes propres. Nous indiquerons à ce sujet deux différences importantes.

a — Les études d'analyse, pour être d'une application assez étendue, doivent tenir compte de la diversité des moyens d'expression de la langue *A*, dans les limites — souvent imprécises d'ailleurs — de la correction grammaticale. Les études de synthèse, par contre, peuvent se contenter d'opérer un choix parmi les possibilités de la langue *B*; elles peuvent, par exemple, ne retenir qu'une orthographe là où la langue en fournit davantage, que quelques-unes des constructions disponibles pour marquer la condition, la causalité, le but, l'hypothèse.

b — La polysémie et l'homographie, qui posent certainement un des problèmes les plus ardues, n'ont d'importance qu'au stade de l'analyse.

3. — *Lexique, morphologie, syntaxe.*

a — Le lexique.

On sait à quelles grossières erreurs peut conduire l'usage incorrect d'un dictionnaire bilingue. Le français *temps* connaît en néerlandais deux principaux correspondants : *weer* et *tijd*. Il incombe au traducteur d'examiner le sens du contexte pour choisir le terme convenable. Placée devant le même problème, la machine ne pourra le résoudre que si l'exploration de la FORME du contexte fournit des éléments de choix objectifs et matériels, ensemble de signes préalablement répertoriés. Si la forme du contexte n'est pas suffisamment explicite, elle pourrait se contenter de noter *weer, tijd*, laissant au lecteur le soin de retenir le terme conforme au sens de la phrase.

Cette difficulté de polysémie proprement lexicale peut cependant être surmontée, du moins dans certains cas, indépendamment de toute recherche contextuelle. En effet, certains termes polysémiques ne sont généralement employés que dans un seul sens lorsqu'ils figurent dans un exposé à caractère scientifique. Le français *cœur*, par exemple, peut, selon le contexte, signifier : organe de la vie, siège des sentiments, centre (nous ne donnons ici que les sens principaux). Mais, dans une étude médicale, ce terme ne pourra fort probablement signifier que « organe de la vie ». On voit ici comment, par une intervention préalable au niveau du lexique, il serait

possible de trancher, ou plutôt de contourner la difficulté. Le problème de la polysémie lexicale se simplifie donc considérablement quand se réduit le champ d'investigation sémantique, en d'autres termes, lorsque le texte à traduire relève d'un domaine bien particulier de la pensée et du savoir. C'est la raison pour laquelle la constitution de lexiques spécialisés ou idioglossaires permet de simplifier les études linguistiques préparatoires aux travaux de programmation, et est de nature à accroître la qualité des textes obtenus par des procédés automatiques de traduction. L'expérience a toutefois montré que la multiplication des idioglossaires présentait de sérieux inconvénients. En effet, de nombreuses branches de la science ont en commun un important vocabulaire spécial, monovalent sur le plan sémantique. Une « bibliothèque » qui contiendrait tous les idioglossaires constituerait une source d'informations très redondante. Pour éviter un gaspillage de temps et d'argent, on a tendance actuellement à regrouper ces lexiques particuliers en un dictionnaire unique dont les rubriques fournissent, s'il y a lieu, les acceptions propres à telle ou telle discipline. Cette méthode exige évidemment que le programme de traduction indique à quelle branche de la science se rattache le texte à traduire.

Il peut arriver qu'en cours de travail la machine à traduire rencontre un mot qui n'a pas été inclus dans le lexique enregistré (nom propre, nombre écrit en chiffre, signe conventionnel, abréviation, néologisme, etc.). Dans ce cas, le terme est reproduit tel quel dans le texte de sortie, muni d'un indicatif signalant au lecteur qu'il s'agit d'une simple transcription de l'original. Cependant, lorsque la langue A est l'un de ces systèmes linguistiques qui permettent aux usagers de former des mots nouveaux par composition, il importe de prévoir un programme qui analyse systématiquement tout mot absent du lexique enregistré. On sait qu'en allemand, par exemple, la création de mots composés est un phénomène relativement fréquent.⁹

En ce qui concerne la traduction des expressions idiomatiques, de certains proverbes et locutions, on doit l'envisager comme une opération lexicale portant sur des ensembles de mots à traiter globalement. Ce procédé ne peut cependant, du moins dans tous les cas, faire abstraction des faits grammaticaux. Comparez, par exemple, les locutions suivantes :

Nom d'une pipe ! Sous le manteau.

Mutatis mutandis. De derrière les fagots. To make a long story short.

et Un (ou des, les, ces...) chevalier(s) d'industrie.

Tirer le diable par la queue.

⁹ Erwin REIFLER de l'Université de l'Etat de Washington, a commencé l'étude de ce problème en vue de la traduction automatique de l'allemand. Cf., entre autres travaux, « Mechanical Determination of the Constituents of German Substantive Compounds », *M T*, 2 (1955), n° 1, pp. 3-14.

b — La morphologie.

Pour marquer les rapports entre les parties du discours, la langue use non seulement de constructions déterminées (ordre des termes, prépositions...) mais elle recourt également à des flexions indiquant le genre, le nombre, la personne, le cas, la modalité, l'aspect, le temps, etc. L'abondance, la diversité et la complexité des règles morphologiques varient d'un système linguistique à l'autre.

De nombreux processus morphologiques présentent cependant un automatisme que l'élaboration d'un programme pourra souvent exploiter, et qui permettra une plus grande condensation de l'information au niveau du lexique. C'est notamment le cas pour le verbe, dont les formes — très nombreuses dans certaines langues — peuvent être considérées comme des combinaisons de radicaux particuliers et de terminaisons communes à tout le système de la conjugaison. Pour les besoins de l'enseignement, des grammairiens ont imaginé un moyen pratique de composer les diverses formes verbales françaises. Si ce procédé, dit de l'apparement des temps, n'est guère exploitable tel quel en T.A.L., certains de ses éléments méritent néanmoins d'être retenus. Sur cette base, nous avons élaboré un système qui ramène les 47 formes que comporte la conjugaison aux temps simples d'un verbe comme *blanchir* à une racine (génératrice de plusieurs radicaux) et 29 terminaisons, valables pour tous les verbes du même type. Au total, le système comporte trois types de conjugaison et 36 terminaisons. Une méthode semblable a permis de réaliser l'étude complète de la conjugaison composée et surcomposée de tous les verbes (même défectifs).

c — La syntaxe.

Ce chapitre de la grammaire concerne l'ordre des mots dans la proposition et des propositions dans la phrase, ainsi que les rapports entre les parties du discours (construction des divers types de propositions : tournures affirmatives, négatives, interrogatives, emploi des prépositions, moyens d'expression de la causalité, du but, de la condition, de l'hypothèse, etc.).

En principe, on peut dire que la syntaxe d'une langue est d'autant plus libre que son caractère flexionnel est plus prononcé. C'est parce qu'elles se servaient des cas pour marquer de très nombreux rapports grammaticaux que les langues latine et grecque offraient une syntaxe fort souple. L'ancien français comptait encore deux cas : le cas sujet et le cas régime; la disparition quasi totale du premier engendra un raidissement de la syntaxe, désormais soumise à des règles plus strictes. Sans doute, et à des degrés différents, les langues romanes et germaniques tendent-elles à uniformiser l'ordre des termes de la proposition. Mais les lois qui régissent cet ordre varient consi-

dérablement selon les langues. Il n'est pas douteux qu'en ce domaine la linguistique comparée ne soit d'un grand secours pour relever les différences et les similitudes entre les systèmes syntaxiques des langues.¹⁰

Notons enfin que ce qui a été dit, au chapitre du lexique, sur la traduction des textes dont les sujets sont nettement circonscrits et à caractère scientifique, reste valable pour la résolution des problèmes grammaticaux. Les ouvrages spécialisés, en effet, usent d'une langue qui n'exploite qu'une partie très réduite de l'éventail extrêmement varié des moyens syntaxiques.

4. — *Les différents types de T.A.L.*

Les recherches sur la T.A.L. n'ont pas la prétention de toucher au but du premier coup. Progressant par étapes, elles se corrigent et s'améliorent à la lumière des résultats fournis par l'expérimentation. Nous tenterons maintenant de décrire et de caractériser brièvement ces étapes, par ordre croissant de complexité.

a — *Le mot à mot non morphologique* constitue le stade le moins élaboré de la T.A.L. Toutes les formes de mots enregistrés de langue A figurent dans le lexique accompagnées de toutes les traductions possibles. Ce procédé est celui qui soulève le moins de difficultés d'ordre linguistique, mais il se révèle extrêmement lourd, tant sur le plan de la confection du lexique — très redondant — que pour la lecture des textes de sortie.

b — On pourrait appeler *mot à mot semi-morphologique* un système qui, après décomposition des mots de langue A en racines et désinences, fournirait les racines correspondantes de langue B, suivies des morphèmes de la langue A. Une telle traduction suppose des études théoriques préalables portant sur la morphologie de la langue d'entrée. Le lexique, beaucoup moins chargé que dans le mot à mot non morphologique, est complété par un programme d'analyse des formes. Locke et Booth¹¹ rapportent une expérience qui fut tentée dans ce sens par Yngve dans le but de montrer que la plupart des problèmes de polysémie rencontrés dans les textes étaient de nature grammaticale, morphologique ou syntaxique.

c — *Le mot à mot morphologique* apporte au procédé *b* un complément substantiel. Ici, l'analyse morphologique de la langue A est suivie d'une synthèse des formes en langue B.

d — Avec la *traduction lexico-grammaticale semi-automatique* du type $A \rightarrow B$ sont abordés les problèmes plus complexes de la syntaxe et de la polysémie. Le texte à traduire est adapté par un prééditeur aux possibilités définies par le lexique et les programmes d'analyse et de synthèse gramma-

¹⁰ Comparez par exemple les langues germaniques au point de vue de certaines constructions (inversion du sujet, rejet du verbe, place de l'adjectif épithète...).

¹¹ *Machine Translation of Languages*, London, 1955, p. 82.

ticaux et sémantiques. Un postéditeur, qui peut ignorer la langue B mais est familiarisé avec la langue de synthèse de la machine, remanie éventuellement le texte de sortie de façon à le rendre plus « lisible ».

e — La traduction *lexico-grammaticale automatique* du type $A \rightarrow B$ serait le produit d'une machine disposant de tous les programmes nécessaires pour lever la plupart ou la totalité des ambiguïtés sémantiques. Elle représente le stade de l'automatisme intégral.

f — Il faut enfin signaler le système appelé *G(eneral) M(echanical) T(ranslation)* qui, dans l'état actuel, n'a qu'une existence toute théorique. Ce système, qui devrait permettre des traductions multilatérales ($A \rightleftharpoons B$; $A \rightleftharpoons C$; $B \rightleftharpoons C$, etc.), utiliserait une langue intermédiaire artificielle qui serait, en quelque sorte, le lieu commun de plusieurs langues naturelles. Les recherches en vue d'obtenir la réalisation beaucoup plus modeste d'une traduction unilatérale $A \rightarrow B$ soulèvent de telles difficultés que le projet de systèmes de traductions bilatérales $A \rightleftharpoons B$ ou multilatérales $A \rightleftharpoons C \rightleftharpoons B \rightleftharpoons A...$ semble actuellement, sinon abandonné, du moins ajourné à très longue échéance.

Il va de soi que tous ces types de traduction connaissent de très nombreuses variantes, plus ou moins importantes. Faire le point des résultats acquis est donc une tâche très malaisée, d'autant qu'un nombre toujours croissant de centres et de laboratoires poursuivent des recherches dans le domaine de la T.A.L. Les couples de langues les plus étudiées sont les suivantes : russe \rightleftharpoons anglais; allemand \rightarrow anglais; anglais \rightleftharpoons chinois; anglais \rightarrow japonais; français \rightarrow russe. Sur le plan des réalisations, on s'oriente généralement vers un mot à mot morphologique de plus en plus élaboré, et plusieurs résultats montrent que des programmes existent d'ores et déjà qui, dépassant le stade du mot à mot, assurent l'application de certaines règles syntaxiques et la résolution de quelques difficultés de polysémie grammaticale.

CONCLUSIONS

L'idée de confier à une machine des travaux de traduction a suscité et suscité encore ce qu'il est convenu d'appeler des réactions en sens divers.

Un inconscient besoin de merveilleux amène certains à rendre les réalisations de la science et de la technique modernes plus merveilleuses encore qu'elles ne le sont. Une personne de mes connaissances ne m'affirmait-elle pas dernièrement que, dans les réunions internationales importantes, des machines traduisaient simultanément et oralement les paroles des conférenciers ? Je ne parvins pas à convaincre mon interlocuteur enthousiaste que son imagination lui faisait pour le moins brûler les étapes, que, par des associations prématurées, il devait prendre des vessies pour des lanternes puisque enfin ces machines, remarquables en vérité, étaient des interprètes.

Tout aussi dangereuse est l'attitude paralysante du sceptique. Loin de s'en réjouir, certains redoutent la perspective d'une traduction automatique, dont ils considèrent le principe même comme une manière d'insulte, un procédé irrévérencieux à l'égard de la langue et de la pensée. Ceux-là oublient peut-être que la langue est un moyen subordonné à une fin dont la réalisation apparaît de jour en jour plus nécessaire : faciliter la collaboration scientifique internationale et, de façon générale, les relations humaines.

D'autres encore doutent du résultat final parce qu'une connaissance approfondie d'une ou plusieurs langues les amène à constater l'extraordinaire complexité des systèmes linguistiques. Peut-être conviendrait-il néanmoins qu'ils prennent davantage conscience des moyens de l'automatisme et du caractère pratique des buts visés par la T.A.L.

Ces prises de position extrêmes nous permettent de mieux définir quelle est l'attitude, empreinte de réalisme, qu'il convient d'adopter à l'égard du principe de la T.A.L.

1) La réalisation de la T.A.L. demandera encore des études fondamentales importantes avant que l'on puisse considérer le problème comme résolu d'une manière satisfaisante. Mais, comme le dit Figaro : « La difficulté de réussir ne fait qu'ajouter à la nécessité d'entreprendre ».

2) Les recherches sur la T.A.L. ne pourront sans doute aboutir que si elles se développent harmonieusement et par améliorations successives. Aborder la traduction d'œuvres littéraires en prose, et à plus forte raison en vers, apparaît comme une entreprise, sinon hasardeuse, du moins semée d'embûches. En tout cas, elle n'aura de sens que si les problèmes moins ardu que pose la traduction de textes scientifiques ont reçu une solution valable.

3) La nouveauté des moyens mis en œuvre pour traduire automatiquement ne postule pas, à notre sens, le recours à des méthodes d'analyse linguistique révolutionnaires. Rompre avec la tradition en ce domaine, c'est se priver à coup sûr des enseignements d'une science qui a déjà accompli de très importants travaux sur la codification et l'étude comparée des langues.

4) Enfin, les études de traduction automatique appellent la collaboration du linguiste et de l'ingénieur. Si le premier est armé pour étudier le fait linguistique, le second possède la maîtrise de l'outil capable de mettre en œuvre les études théoriques. Dans la réalisation d'un projet aussi vaste et complexe que la T.A.L., les activités de l'un et de l'autre doivent s'exercer solidairement. Les méthodes de travail du linguiste ne peuvent se définir qu'à la lumière des indications fournies par l'ingénieur sur les possibilités et les limites des machines à traiter l'information. De son côté, l'ingénieur ne peut œuvrer avec fruit que si la matière de son expérimentation a été soigneusement préparée par le spécialiste de la langue.