

L'informatique au service de la traduction

Bernard Vauquois

Volume 26, numéro 1, mars 1981

L'informatique au service de la traduction
Machine Aids to Translation

URI : <https://id.erudit.org/iderudit/004556ar>
DOI : <https://doi.org/10.7202/004556ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)
1492-1421 (numérique)

[Découvrir la revue](#)

Citer cet article

Vauquois, B. (1981). L'informatique au service de la traduction. *Meta*, 26(1), 8-17. <https://doi.org/10.7202/004556ar>

L'informatique au service de la traduction*

BERNARD VAUQUOIS

INTRODUCTION

Pendant les Journées d'Études du Festival International du Son de 1975, j'avais eu l'occasion de présenter l'évolution des recherches en traduction automatique (c'est-à-dire, traduction par ordinateur) depuis les toutes premières tentatives qui eurent lieu vers le milieu des années 50. Si en 1975, malgré tous les signaux qui montraient l'arrivée prochaine d'une demande considérable de traductions et l'urgence de moyens automatisés pour faire face à cette demande, le sujet pouvait paraître futuriste ou du moins d'avant-garde, en 1980 ce même sujet est en plein dans l'actualité. En effet, en 1976, l'Université de Montréal mettait en place son système de traduction automatique anglais-français des bulletins météorologiques; ce système, actuellement opérationnel, traduit la plupart des bulletins émis sur tout le territoire du Canada (environ 5 000 bulletins par jour). En 1976 également, le gouvernement canadien chargeait la même équipe de la réalisation d'un système plus ambitieux pour traduire (toujours d'anglais en français) les manuels de maintenance d'un avion construit sous licence au Canada. À la même époque, la Communauté Économique Européenne signait un premier contrat pour la réalisation d'une traduction automatique anglais-français des textes documentaires sur l'agriculture par le système SYSTRAN. Depuis lors, ce contrat a été étendu à d'autres couples de langues. En mars 1977 le Groupe d'Études pour la Traduction Automatique à l'Université de Grenoble a pris l'initiative d'organiser un colloque Université-Industrie avec la participation des autorités de tutelle et d'administrations intéressées (y compris la CEE) pour sensibiliser les autorités compétentes et les promoteurs éventuels à cette activité dont le débouché est assuré, pour montrer ce que le laboratoire de Grenoble avait déjà réalisé et pour amorcer la phase de développement industriel. En février 1978, la CEE lançait le projet d'études EUROTRA destiné à faire coopérer les diverses équipes européennes pour aboutir à un système européen de deuxième génération. En 1979, après dix années d'interruption, les Japonais ont repris leurs travaux dans ce domaine avec la rapidité et l'efficacité qu'on leur connaît. Des grandes compagnies constructrices d'ordinateurs aux États-Unis, au Japon, en Allemagne, s'inté-

* Communication présentée aux Journées d'Études 1980 du Festival International du Son (3-8 mars 1980).

ressent à la traduction automatique. La Société LOGOS, aux États-Unis, qui avait réalisé en 1973 un système de traduction anglais-vietnamien reprend ses activités après quelques années de demi-sommeil. Enfin, tout le monde a pu se rendre compte du succès rencontré l'année dernière par les « traductrices de poche » vendues pourtant à un prix exorbitant pour des performances très médiocres !

En nous mettant à l'écart de toute publicité exagérée, en faisant le point de ce qui existe et en prenant conscience de ce qui se développe dans les laboratoires de recherches nous allons essayer de montrer ce que l'informatique peut et doit apporter à la traduction des langues naturelles dans l'avenir immédiat et à moyen terme. Il nous faudra tenir compte non seulement des aspects scientifiques et techniques des méthodes de traduction automatisée, mais aussi accorder une importance au moins égale aux aspects psychologiques et économiques que l'introduction de l'informatique dans un nouveau secteur professionnel ne manquera pas de faire apparaître.

Nous allons donc voir successivement :

- les méthodes de la traduction automatique ;
- les limites qui lui sont imposées dans un avenir prévisible ;
- le rôle de l'informatique dans diverses activités professionnelles ;
- comment harmoniser sous le contrôle du traducteur un processus de traduction qui fait intervenir à la fois l'homme et l'ordinateur.

1 — MÉTHODES DE LA TRADUCTION AUTOMATIQUE

Nous entendons sous ce titre les activités de recherches qui visent à faire traduire des textes par ordinateur sans autre intervention humaine qu'une pré-édition réduite à des indications typographiques, des indications de « hors textes » (tableaux, figures,...) et de formules. En 1975, nous avons mis en relief l'opposition entre les méthodes de première et de deuxième génération.

La méthode de première génération, fondée initialement sur l'exploitation d'un dictionnaire bilingue s'est perfectionnée progressivement par adjonctions de programmes destinés à traiter soit des ambiguïtés, soit de phénomènes contrastifs entre les deux langues, soit de nombreux cas particuliers d'origine lexicale.

La méthode de deuxième génération repose sur les deux principes suivants :

- a) *Principe de la recherche de descripteurs structuraux pour représenter chaque phrase de texte à traduire*

Ce principe conduit à des systèmes de traductions en trois étapes successives : analyse du texte source pour associer à chaque phrase le descripteur structural qui lui convient, transfert du descripteur lié à la langue source en un descripteur correspondant lié à la langue but, enfin génération à partir de ce dernier descripteur d'un texte en langue but qui donne ainsi la traduction.

Le choix de tels descripteurs peut être plus ou moins ambitieux; si l'on vise très haut, les concepts qui interviennent dans cette représentation peuvent atteindre un degré d'universalité tel que le même descripteur représente toutes les «phrases synonymes» non seulement dans une langue mais aussi dans un grand nombre de langues de familles différentes.

b) *Principe de la séparation entre les données linguistiques et les programmes*

Au lieu de programmer directement pour chaque langue (et même pour chaque typologie de textes dans une langue) les opérations d'analyse, de transfert et de génération, on écrit une fois pour toutes des programmes généraux qui constituent des modèles théoriques pour ces différentes phases.

Ces programmes reçoivent les données linguistiques : grammaires, dictionnaires et stratégies. Ils produisent les programmes qui effectuent alors les traductions. Ce sont des programmes générateurs de programmes. Ils ne dépendent pas des langues en cause et ils doivent se prêter à différentes théories linguistiques ainsi qu'au plus grand nombre possible de modèles algorithmiques. De tels logiciels sont peu nombreux (5 ou 6 dans le monde). Celui du Groupe d'Études pour la Traduction Automatique de Grenoble, appelé ARIANE-78, est utilisé non seulement sur place mais aussi dans d'autres universités à l'étranger.

2 — EXEMPLE DE REPRÉSENTATION STRUCTURALE DANS UNE MÉTHODE DE DEUXIÈME GÉNÉRATION

Considérons les phrases suivantes :

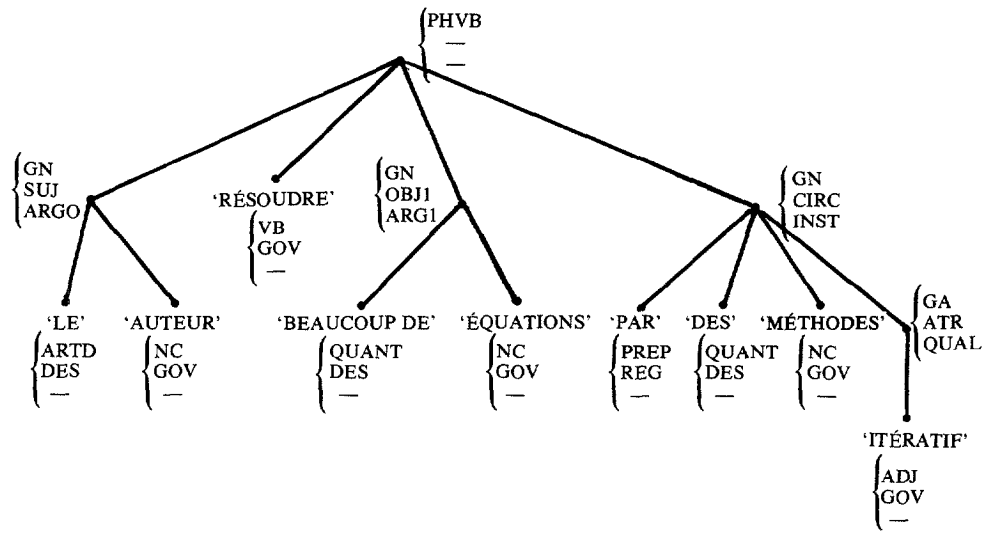
- 1) L'auteur résoud beaucoup d'équations par des méthodes itératives.
- 2) Beaucoup d'équations sont résolues par l'auteur par des méthodes itératives.
- 3) On résoud beaucoup d'équations par des méthodes itératives.
- 4) Beaucoup d'équations se résolvent par des méthodes itératives.
- 5) Il se résoud beaucoup d'équations par des méthodes itératives.

Nous allons voir comment peuvent se représenter ces phrases par des schémas arborescents à trois niveaux d'interprétation.

- a) niveau d'association de classes morpho-syntaxiques (telles que NOM COMMUN, ADJECTIF, PRÉPOSITION ou encore GROUPE NOMINAL, PHRASE INFINITIVE, PHRASES VERBALE,...)
- b) niveau de construction des fonctions syntaxiques (SUJET, COMPLÉMENT, OBJETS,...)
- c) niveau de la construction logique (prédicats, places d'arguments, relations circonstancielles telles que INSTRUMENT, MANIÈRE, LIEU, CAUSE,...).

La phrase 1) va fournir le schéma suivant :

FIGURE I



a) Les classes morpho-syntaxiques qui figurent dans ce schéma sont :

PHVB	pour	phrase construite autour d'un verbe conjugué
GN	"	groupe nominal
GA	"	groupe adjectival
VB	"	verbe conjugué
ARTD	"	article défini
QUANT	"	quantificateur
NC	"	nom commun
ADJ	"	adjectif
PREP	"	préposition

b) Les fonctions syntaxiques :

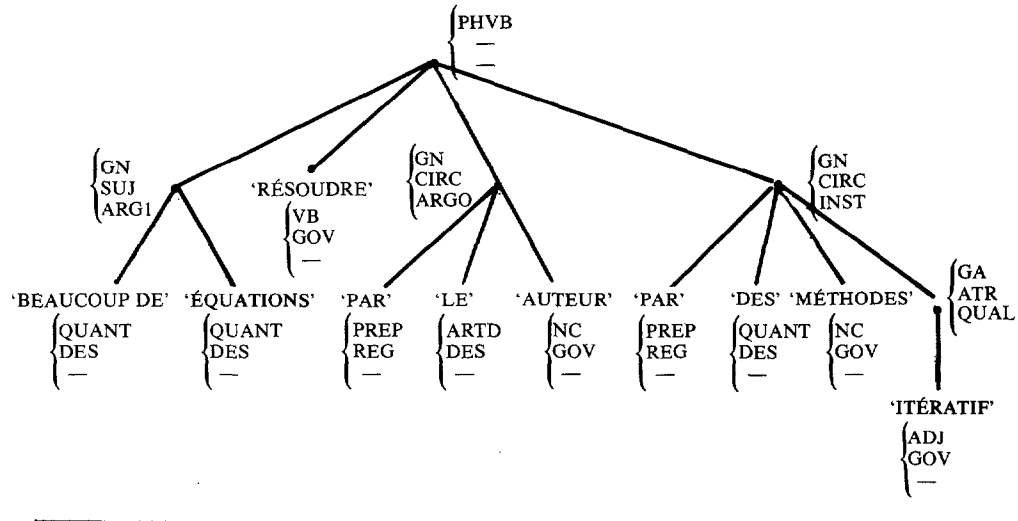
SUJ	pour	sujet
OBJ1	"	premier objet
CIRC	"	complément circonstanciel
DES	"	désignateur
GOV	"	gouverneur du groupe morpho-syntaxique
REG	"	régisseur du groupe morpho-syntaxique.

c) Les relations logiques :

ARGO	pour	la première place d'argument d'un prédicat
ARG1	"	la deuxième place d'argument d'un prédicat
INST	"	la relation « instrumentale » entre un groupe morpho-syntaxique et un prédicat.

De même, le descripteur associé à la phrase (2) sera le suivant :

FIGURE II



On remarquera que le groupe morpho-syntaxique « BEAUCOUP D'ÉQUATIONS » qui avait la fonction syntaxique de « premier objet » dans la phrase (1) possède celle de « sujet » dans la phrase (2) mais conserve dans tous les cas la relation logique (indépendante de la construction syntaxique) de ARG1 dans le prédicat « RÉSoudre ».

De même le groupe morpho-syntaxique « L'AUTEUR » dans la phrase (1) ou « PAR L'AUTEUR » dans la phrase (2) change de fonction syntaxique mais garde toujours la place de premier argument (ARGO) dans sa relation logique avec le prédicat « RÉSoudre ».

Dans les 3 phrases suivantes, on fait apparaître 3 constructions syntaxiques différentes mais qui ont la même organisation logique et qui diffèrent des deux premières par le fait que la place de premier argument est vide ou indéterminée.

On a ainsi les schémas ci-contre pour les phrases (3), (4), (5) :

À l'aide d'un nombre suffisant de fonctions syntaxiques et de relations logiques on est toujours capable, à la main, d'associer à chaque phrase, aussi compliquée soit-elle, une description structurale à trois niveaux ; comme nous l'avons fait pour les exemples précédents.

Par contre, obtenir le bon résultat, et seulement le bon, par un processus d'analyse automatique n'est pas une opération garantie à 100%. De même la solution des ambiguïtés de sens sur les mots ne peut être trouvée dans tous les cas.

FIGURE III

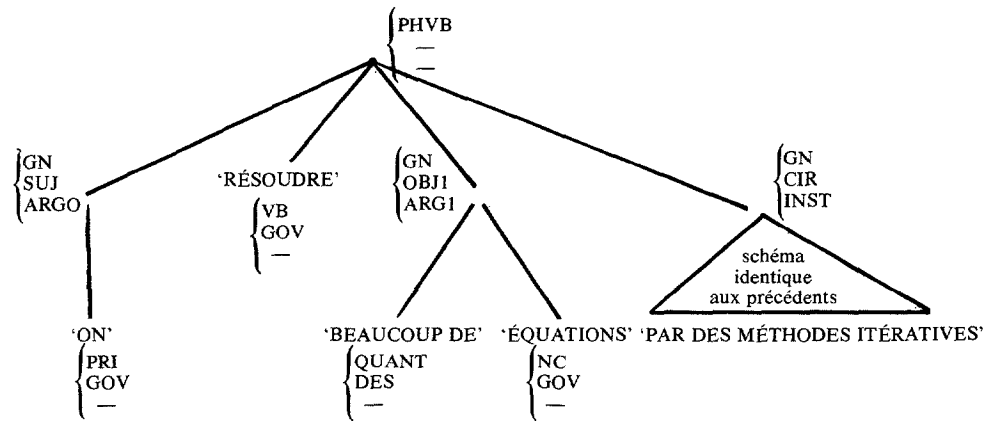
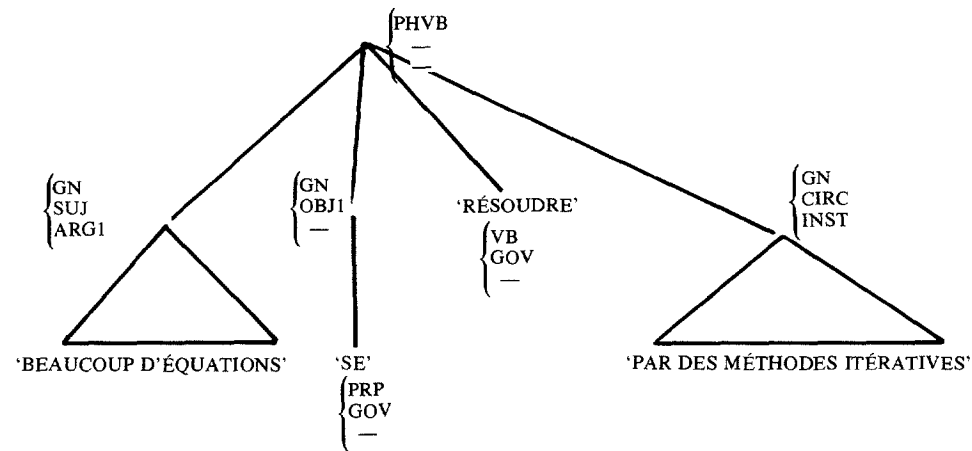


FIGURE IV

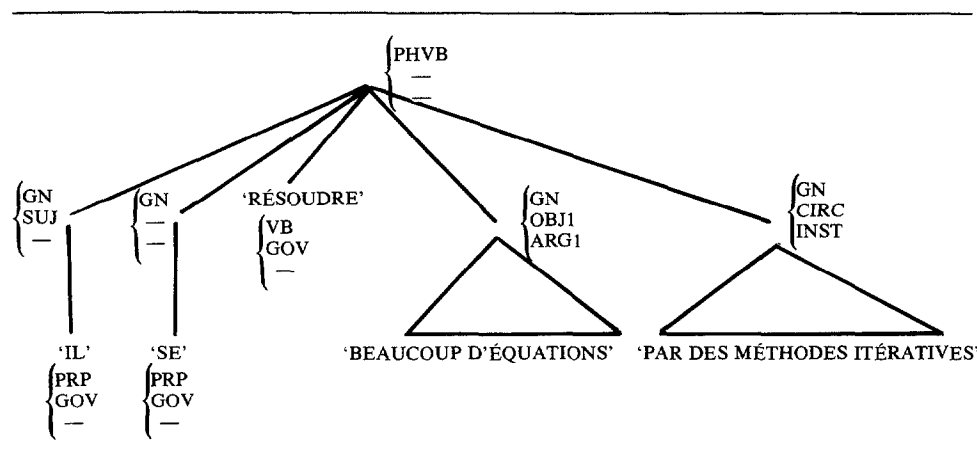


Cependant, on voit bien que la structure considérée au niveau des relations logiques constitue un « invariant » dans le processus de traduction. Ainsi la phrase 4) ne fournira pas la traduction anglaise mot à mot

« Many equations solve themselves by iteration methods » mais pourra être traduite par

« Many equations are solved by iteration methods » comme les phrases 3) et 5).

FIGURE V



3 — LES LIMITES DE LA TRADUCTION AUTOMATIQUE

Imaginer que l'ordinateur pourra remplacer le traducteur professionnel dans toutes les situations et pour tous les besoins est un rêve de science-fiction. Si l'on veut classer les différentes traductions, il faut tenir compte des paramètres suivants :

a – *Typologie du texte* : cela concerne surtout le style utilisé par les auteurs. Les phrases telles qu'elles sont écrites dans les manuels scientifiques, dans les revues, dans les résumés documentaires, dans les articles de journaux, dans les textes administratifs, etc. obéissent à des règles différentes selon cette typologie. Bien qu'il s'agisse de la même langue, on peut dire que l'on est en présence de langages différents.

b – *Le domaine* : ce paramètre intervient évidemment pour le choix du vocabulaire. La terminologie est extrêmement précise dans toutes les langues lorsqu'il s'agit de domaines scientifiques.

c – *La langue source et la langue but* : les aspects contrastifs entre ces deux langues.

d – *Le but de la traduction* : s'agit-il de traduire un document pour acquérir une information intéressante, ou bien pour se contenter d'avoir une idée de son contenu ? S'agit-il de traduire un document pour diffuser l'information soit à titre de documents de travail, soit en vue de son impression, soit pour faire des notices commerciales,...

Pour donner une idée de la multiplicité des types de traductions qu'on peut rencontrer, en choisissant 4 types de textes, 5 domaines, 6 langues (cas des institutions des Communautés Européennes), 3 buts différents, on arrive à 1 800 « profils » de traductions.

La traduction exclusivement automatique, sans intervention humaine, répond certainement au but « acquisition d'information globale sur le contenu d'un document ». D'ailleurs, les résultats de cette sorte sont acquis depuis longtemps

(une quinzaine d'années, au moins) et les systèmes de première génération sont certainement les plus économiques pour accomplir la tâche. Cet emploi de la traduction automatique peut se révéler fructueux un peu dans tous les domaines pour effectuer la sélection d'informations intéressantes que l'on donne à traduire correctement. On évite ainsi des traductions humaines longues et coûteuses pour des textes sans intérêt.

La traduction automatique est aussi gagnante lorsqu'il s'agit de traductions sur des textes de typologie simple. Le meilleur exemple est fourni par le système MÉTÉO réalisé au Canada. Les bulletins météorologiques sont en effet rédigés de façon assez uniforme et sont construits sur un canevas bien établi. Le système reçoit environ 5 000 bulletins par jour; au moins 80% de ces bulletins sont traduits correctement; pour les autres, le système lui-même se rend compte qu'il ne peut les traduire et les fait parvenir immédiatement, grâce au réseau, à un bureau de traducteurs qui lui renvoient les traductions; le système s'occupe de la réinsertion de ces bulletins traduits à l'extérieur dans le lot.

Enfin, on peut compter sur une extension des applications de la traduction automatique à des types de texte un peu plus compliqués (résumés documentaires, par exemple) dans un proche avenir. Mais c'est sans doute sous une forme différente, que les travaux sur le traitement des textes par ordinateur et l'informatique d'une manière plus générale vont intervenir efficacement dans les opérations de traduction.

4 — L'INFORMATIQUE AU SERVICE DE LA TRADUCTION

C'est une évidence flagrante de constater que l'informatique a envahi en quelques années l'activité de nombreux secteurs professionnels. Considérés au moment de leur apparition comme des machines à calculer, les ordinateurs ont permis de réaliser des calculs numériques qui auraient exigé des délais considérables, même inimaginables, au moyen de machines mécaniques. Très vite, l'ordinateur est devenu la machine à traiter l'information, le calcul numérique n'étant plus qu'un cas particulier d'un tel traitement. En outre, il faut se rendre compte que l'ordinateur est un outil qui multiplie les facultés humaines de traitement de l'information par un facteur sans commune mesure avec le facteur que donnent les engins mécaniques pour multiplier les facultés physiques de l'homme. Si un avion supersonique permet d'aller 100 ou 200 fois plus vite qu'un voilier ou si un bulldozer permet d'effectuer le travail de 1 000 individus munis de pioches et de pelles, un ordinateur permet d'effectuer une multiplication ou un tri sur un grand fichier plusieurs millions de fois plus vite que l'individu le plus rapide. Aussi la gestion des grandes bases de données (comptes des clients d'une banque, stocks des grandes entreprises, réservations des places d'avion à l'échelle planétaire,...) est-elle maintenant effectuée par ordinateur.

Il est certain que la profession de traducteur sera aussi touchée par le phénomène informatique. Il appartient aux informaticiens et surtout aux traducteurs que les transformations qui en découleront aient un effet de progrès à la

fois pour les professionnels et pour les usagers de la traduction. Dans ce but il importe de savoir que deux voies sont en train de se développer.

La première pourrait s'appeler «traduction humaine assistée par ordinateur». Il existe déjà des dictionnaires terminologiques multilingues accessibles par ordinateur; il existe aussi de nombreuses banques de données qui communiquent par réseau d'ordinateurs. On imagine facilement un traducteur installé à une console reliée à une machine ou à un réseau, consulter de tels dictionnaires ou de telles banques de données. Au moyen de consoles reliées à la même machine ou au même réseau, on imagine aussi bien des échanges d'informations entre traducteurs. Enfin, si on regarde en détails tout ce que les programmes d'édition et de manipulation de textes peuvent faire actuellement, toutes les tâches de rédaction, de correction, de mise en page, etc., deviennent extrêmement faciles et rapides pour un traducteur qui commande et contrôle toute cette activité à partir de l'écran et du clavier de sa console. Tous ces dispositifs, disponibles dès maintenant, n'effectuent par eux-mêmes aucune traduction; ce sont seulement des outils informatiques mis à la disposition du traducteur pour accélérer et faciliter sa tâche. Ce serait sans doute la phase par laquelle le traducteur professionnel pourrait commencer, pour se familiariser avec l'informatique et pour se rendre compte qu'il garde le contrôle complet sur toutes les opérations.

La deuxième voie, dérivée de la traduction automatique mais moins ambitieuse, pourrait s'appeler «traduction automatisée assistée par le traducteur». C'est surtout dans cette perspective qu'il faut aménager le rôle de l'informatique pour que le traducteur en reste le maître. Les systèmes de traduction automatique avaient été prévus pour effectuer un processus sans intervention humaine. Or, la qualité des résultats ne correspond pas toujours à ce qui est attendu et le traitement exclusivement automatique ne suffit plus.

En conséquence, en dehors des cas où la traduction exclusivement automatique s'applique, il faut organiser la combinaison de programmes sur ordinateur et d'interventions humaines. Ces dernières peuvent apparaître soit avant le processus automatique (c'est la préédition), soit au moyen d'interruptions pendant le processus automatique (c'est l'interaction) soit enfin après l'exécution du processus automatique (c'est la révision). Évidemment, chacune de ces actions n'exclut pas les autres possibilités.

— *La préédition.* Ce terme est sans doute ambigu et il faut clarifier ce qu'il signifie. Souvent, il a été entendu par préédition d'un texte, de transformer ce texte pour le mettre sous une forme acceptable par le système automatique. Il s'agit en fait d'un codage ou même d'une traduction d'un texte écrit en langue naturelle en un texte présenté en langue pseudo-naturelle. Cette manière de faire est évidemment inacceptable et nous proposons de comprendre par préédition une annotation facultative et pas nécessairement complète du texte source au moyen d'un certain nombre de marqueurs afin d'aider efficacement le système automatique à résoudre les ambiguïtés les plus difficiles. Il ne faut certainement pas demander davantage à un travail de préédition.

— *La postédition.* C'est le moyen le plus couramment utilisé à l'heure actuelle. Malheureusement, il se borne à fournir aux réviseurs la sortie sur imprimante du texte traduit par le processus automatique. Si les fautes sont nombreuses et répétitives, le réviseur doit continuellement reprendre les mêmes corrections. Assez souvent, il n'est pas possible de corriger le texte de sortie par une simple correction sur des mots, mais il faut reconstruire toute une phrase. C'est le cas général lorsqu'il s'agit de réviser des traductions automatiques obtenues par des systèmes de première génération. On comprend fort bien, qu'après quelques heures d'un tel exercice, le réviseur envoie la sortie de traduction automatique au panier et se met à traduire à partir du texte source.

Un progrès important pourrait être accompli en donnant au réviseur la possibilité de travailler sur console pour bénéficier de tous les programmes qui existent sur l'édition et la manipulation de textes. Mais cette manière d'opérer n'est valable que si la phase automatique a produit des résultats de qualité suffisante.

— *L'interaction.* C'est là, certainement que se situe la combinaison la plus efficace et la plus intéressante entre le traitement automatique et l'intervention humaine.

D'abord, parce que le traducteur garde le contrôle pendant le processus de traduction et n'est plus soumis à réparer « a posteriori » les erreurs du système automatique. C'est dans cette voie qu'il faut penser la traduction automatisée et non changer de nom (traduction automatique → traduction automatisée) le même système, en fonction des possibilités conversationnelles d'accès à l'ordinateur.

Ensuite, parce qu'en considérant le système automatique placé dans un tel contexte, le réviseur (ou le traducteur) pourra lui-même améliorer le système automatique et lui « apprendre » à éviter telle faute ou à acquérir telle habitude.

Cette optique de développement des systèmes de traduction automatisée n'est certes pas encore au point; cependant, elle ne présente pas de difficultés insurmontables et devrait aboutir à des résultats significatifs en quelques années. En tout cas, elle est certainement plus vite accessible que la recherche d'une qualité parfaite (ou presque) par des moyens exclusivement automatiques.