

LA TRADUCTION AUTOMATIQUE, UNE IMPOSSIBLE SCIENCE DU LANGAGE APPLIQUÉE

Jacqueline Léon*

RESUMO: O artigo aborda características e tendências das diferentes fases pelas quais passou a tradução automática, assim como as tendências do momento atual.

Palavras-chave: tradução automática, ciência da linguagem, linguística computacional.

Comme on le sait, la TA est la première expérience de mécanisation du langage à l'aide d'une machine électronique. Apparue dès 1948 en Grande-Bretagne et aux Etats-Unis, elle s'est d'emblée située comme une application ayant des ambitions industrielles, à savoir la production de traductions en série de bonne qualité correspondant à une demande sociale, industrielle ou commerciale.

On sait également qu'avec la publication du rapport de l'ALPAC en 1966, la traduction automatique (TA *infra*), jugée non rentable, n'a plus eu droit de cité aux USA, même sous la modeste appellation de traduction automatique assistée par un pré- ou un post-éditeur humain. Si l'on se re-

^(*) CNRS/Laboratoire des Théories Linguistiques, Université Paris 7, ENS Lettres et Sciences Humaines.

porte trente-cinq ans après cette liquidation quasiment totale, nous sommes face à deux phénomènes apparemment contradictoires:

- i) La TA a survécu et semble être à l'heure actuelle un secteur important du traitement automatique des langues.
- ii) Cinquante ans après le début des recherches, ses acteurs sont nombreux à établir un constat d'échec.

On citera ici la position de Newton, éditeur d'un recueil d'articles sur la TA publié en 1992, qui déclare dans la préface de son ouvrage:

“Les chercheurs ont tenté de développer des systèmes d'émulation de la traduction humaine depuis une quarantaine d'années. Qu'est-ce qui fait que le langage naturel transforme cette entreprise en un défi insurmontable? Alors que d'autres applications informatiques se sont soldées par des succès rapides et durables, la TA n'a pas réussi à pénétrer le marché mondial de façon significative. Est-ce que cela reflète une simple répugnance à adopter les systèmes de TA ou cela signale-t-il un problème plus fondamental et insoluble?”¹

Pour rendre compte de cette survivance, mais aussi de ce piétinement de la TA, nous nous proposons, dans cet article, de retracer l'histoire de ce domaine qui s'est considérablement modifié au fil des décennies, tant par le développement technologique des ordinateurs, que par la transformation de la demande en traductions. Le rapport aux sciences du

⁽¹⁾ *“Researchers have been attempting to develop systems to emulate the human translation process for some forty years. What is about natural language that makes this such a daunting challenge? While other software applications have achieved rapid and lasting success, MT has failed to penetrate the worldwide market to any appreciable extent. Does this merely reflect a reluctance to adopt it, or does it signal a more fundamental and intractable problem”.*

langage s'est aussi considérablement modifié et on peut se demander si la TA est parvenue à être la science du langage appliquée qu'elle aspirait devenir. On se posera également la question de savoir ce qui fait que le langage résiste ou du moins à quel moment les acteurs de la TA rencontrent-t-ils de la résistance? On se demandera aussi pourquoi c'est précisément dans les années 90 qu'apparaissent ces énoncés de constats d'échec de la part des acteurs? Enfin pour parler d'échec ou de réussite de la TA, il faut savoir ce qu'on entend par bonne traduction concernant un produit de la machine, et là encore les critères d'évaluation des systèmes de TA et de leurs résultats se sont profondément modifiés le long des années.

Pour examiner ces points, je propose d'envisager cinq périodes² qui nous amènent à la période contemporaine:

- 1^{ère} période (1948-1960): idées et expérimentations;
- 2^{ème} période (1960-1966): l'analyse syntaxique;
- 3^{ème} période (1966-1980): la survie et la "force brute" des grands systèmes;
- 4^{ème} période (1980-1990): le tournant japonais et l'automatisation de la communication;
- 5^{ème} période (depuis 1990): le retour des méthodes empiriques (grands corpus et lexique).

⁽²⁾ On notera que cette périodisation ne se conforme pas à celle généralement alléguée par les acteurs et qui repose sur les différents types de méthodes: 1^{ère} génération (1950-66): méthode directe mot à mot; 2^{ème} génération (années 70): méthode par transfert; 3^{ème} génération (à partir des années 80): méthode interlangue incorporant des techniques de l'intelligence artificielle. Cette périodisation qui ne tient compte que des méthodes n'est pas exacte puisque, dès les années 50, les trois méthodes existent déjà, élaborées à partir des différentes conceptions de l'automatisation de la traduction. De plus, comme cela vient d'être évoqué, l'évolution de la TA ne se résume à un changement de méthodes.

1. PREMIÈRE PÉRIODE (1948-1960): IDÉES ET EXPÉRIMENTATIONS

1.1. WARREN WEAVER ET LES MÉTHODES PROBABILISTES

Concernant la première période, il est intéressant de reprendre les premières propositions de Weaver (1955) dans son Memorandum de 1949 intitulé “*Translation*”. Dans ce texte, Weaver cite la lettre qu’il envoie à Wiener pour le convaincre des avantages de la TA.³ Il en expose les avantages politiques et humanistes: paix dans le monde, échanges scientifiques et techniques qui font écho à la demande en traductions des USA à l’époque (traduction du russe vers l’anglais pour des raisons stratégiques et militaires).

Il aborde un certain nombre de points qui resteront au coeur du problème du traitement automatique du langage et de sa formalisation: le problème de la signification multiple et des ambiguïtés qui reste, pour le TAL (traitement automatique des langues), un problème posé de façon indépendante; les bases logiques du langage et la possibilité de langues ou de grammaires universelles; l’application des méthodes probabilistes au langage.

Pour Weaver, l’idée qu’il existe des invariants, des éléments de base communs à toutes les langues est issue des techniques de déchiffrement et de cryptographie: la fréquence des lettres; la combinaison des lettres, les intervalles entre les lettres sont en grande partie indépendants de la langue. Il conçoit la traduction comme un problème de cryptographie et propose de s’attaquer aux problèmes d’ambiguïtés par l’examen du contexte linguistique à l’aide des méthodes statistiques héritées de la théorie de l’information et du chiffre.⁴ Il

⁽³⁾ Warren Weaver et Norbert Wiener sont deux mathématiciens, proches du groupe de la première cybernétique organisateur des Macy Conferences (1946-1953). Pour plus de détails, voir Léon (1997).

⁽⁴⁾ Cf. Shannon et Weaver: *The Mathematical Theory of Communication*, 1948.

suggère qu'un contexte de $2N$ mots adjacents à gauche ou à droite (les N -grammes, N variant selon le type de texte traduit) devrait suffire.⁵

Il faut noter que les méthodes probabilistes sont très utilisées pendant les années 50. Ainsi, en 1955, Harris définit une procédure de segmentation d'une chaîne de phonèmes de façon à définir la notion de morphème de façon probabiliste. Toutefois la méthode de Weaver n'est pas possible à mettre en oeuvre à l'époque. Il n'existe pas de corpus suffisamment représentatif d'une langue donnée. Les ordinateurs ne sont pas assez puissants pour stocker toutes les chaînes de di-grammes et de tri-grammes nécessaires à établir des statistiques pour cette langue. Weaver le reconnaît lui-même dans son texte (p. 21). Ce n'est que dans les années 90, après une éclipse d'une trentaine d'années, que ces techniques seront reprises.

Wiener, quant à lui ne croit pas à la TA. Dans sa réponse à Weaver, il laisse entendre que le problème des ambiguïtés et du sens des mots n'est pas soluble. Il prend l'exemple du Basic English, sous-ensemble de l'anglais comprenant 2000 mots dans lequel à chaque mot correspond un seul sens. Même un sous-ensemble de ce type, dit-il, ne peut éliminer complètement les ambiguïtés (exemple de *to get*). Ce qui est mis en avant par Wiener, c'est le problème de l'opacité du langage naturel. L'utilisation des 'langages contrôlés', syntaxes limitées et autres, sera toujours mise en avant par les auteurs de systèmes de TA comme une façon de contourner le problème.

1.2. L'IDÉE DE "BONNE TRADUCTION"

Il est intéressant de voir que Wiener et Weaver n'envisagent pas la TA de la même façon: Derrière la position de Wiener,

⁽⁵⁾ Kaplan (1950) a montré que la résolution des mots polysémiques se trouve dans l'examen des mots proches plutôt que dans celui des mots lointains et qu'un ou deux mots avant ou après suffisent.

il y a l'idée que la machine est incapable de fournir de bonnes traductions, au sens de traductions équivalentes aux traductions humaines, à cause de l'ambiguïté constitutive du langage que même les langages restreints ne peuvent contourner. Pour Weaver au contraire, les scientifiques, principaux destinataires de la TA, peuvent se contenter de traductions grossières comportant un certain pourcentage d'erreurs. Toutefois le statut de ces erreurs n'est pas clair et on peut s'interroger, en reprenant les termes de Loffler-Laurian (1996), sur la signification d'une 'traduction bonne à 80%'. Est-ce que cela veut dire que 80% de phrases ressemble à des phrases humaines, ou que 80% de mots a reçu un équivalent correct, ou encore que la traduction a été jugée acceptable par 80% des personnes consultées. C'est une question fondamentale pour la TA.

1.3. LES TENANTS DE LA TRADUCTION DE GRANDE QUALITÉ ENTIÈREMENT AUTOMATISÉE

Chez d'autres expérimentateurs de TA dans les années 50, l'objectif est beaucoup moins modeste. En se donnant une définition de la traduction comme celle que Léon Dostert propose en 1955, les expérimentateurs se donnent comme objectif de fournir par la machine une traduction équivalente à une traduction humaine, une traduction de grande qualité entièrement automatisée:

*“La traduction est cette branche des sciences du langage appliquées consacrée au problème du transfert de sens d'un ensemble structuré de symboles apparaissant dans une culture donnée (la langue source) dans un autre ensemble structuré de symboles apparaissant dans une autre culture (la langue cible)”.*⁶

⁽⁶⁾ *“Translation is that branch of the applied science of language which is specifically concerned with the problem – or the fact – of transference of meaning from one*

Ce type de définition, met en jeu l'idée, issue de l'anthropologie culturelle, que les langues sont ancrées dans des cultures et aussi le fait que le transfert du sens est au centre du processus de traduction. En se donnant ce type de définition, ils espèrent fournir par la machine des traductions semblables à celles fournies par les traducteurs humains. C'est dans ce sens qu'ils revendiquent pour la TA un statut de Science du Langage appliquée.

Or le groupe de Dostert est le plus gros groupe américain, le Centre de TA de Georgetown University, qui a présenté la 1ère démonstration de TA sur ordinateur dite 'Georgetown-IBM' à New York en janvier 1954. C'est principalement à ce groupe que s'attaque Bar-Hillel dans son rapport de 1960. Il les accuse, en prétendant faire de la traduction de grande qualité entièrement automatisée, d'être à la limite de l'escroquerie -il est vrai qu'il n'existe pas d'évaluation des travaux du groupe et peu de publications. Bar-Hillel dit, et il défendait déjà cette position dans sa première évaluation de 1951 qu'il faut clarifier les objectifs. Si on se donne, dit-il, comme objectif la traduction automatique, c'est-à-dire la production de traductions en séries, il faut renoncer soit à la qualité, soit à l'autonomie de l'output, c'est-à-dire accepter la post-édition et le fait que la machine soit assistée par l'humain.

Rappelons que l'argument de Bar-Hillel contre la possibilité de la traduction de bonne qualité entièrement automatisée est bien connu. Il tient aussi aux ambiguïtés. Il s'agit des ambiguïtés sémantiques que la machine ne peut résoudre faute de connaissances de sens commun ou de savoir encyclopédique. Alors que *the pen was in the box* est traduit de façon correcte par la machine (*pen*= crayon), *the box was in the pen* (*pen*=parc à bébé) ne peut être traduit car la ma-

patterned set of symbols occurring in a given culture (the original language) into another set of patterned symbols occurring in another culture (the "final" language)" (cf. Locke et Booth, 1955, p. 124).

chine ne peut comprendre le contexte dans lequel cette phrase peut être produite *Little John was looking for his toy box. Finally he found it. The box was in the pen. John was very happy.* L'argumentation de Bar-Hillel tient à la nature du langage, en l'occurrence à son ancrage corporel dans un espace à trois dimensions que ne connaît pas la machine.⁷

Par ailleurs Bar-Hillel est totalement hostile au compromis qui consiste à fournir une sortie qui ne comporterait que 5% d'erreurs, revendiquée, à la suite de Weaver, comme un bon résultat par un certain nombre de groupes, comme à toute sortie qui réduirait la fiabilité de la machine.

2. DEUXIÈME PÉRIODE (1960-1966): L'ANALYSE SYNTAXIQUE

2.1. DE LA TA À L'ANALYSE SYNTAXIQUE

On peut dire que la première période est foisonnante d'idées qui ont pu être plus ou moins mises en oeuvre dans des expérimentations selon la disponibilité des machines: les méthodes probabilistes, les langues intermédiaires sémantiques⁸ élaborées de façon distincte par le groupe britannique, ou par les Russes comme Mel'cuk, les méthodes empiriques

⁽⁷⁾ Sur cette question, une discussion s'engage avec le groupe britannique de Cambridge qui propose une solution par thésaurus et champ sémantique, à l'origine des premières bases de connaissances en intelligence artificielle. Bar-Hillel objecte qu'un thésaurus donnerait la même traduction pour *pen* dans les deux exemples suivants: *the pen was in the inkstand* et *the inkstand was in the pen*. Voir sur ce point Léon (2000).

⁽⁸⁾ Les britanniques et les russes, dès les années 50, proposent une méthode de TA par langue intermédiaire; les premiers à partir de collocations utilisant la notion de signification contextuelle de Firth (méthode par thésaurus), les seconds à partir de l'idée d'une protolangue construites à partir des significations lexicales et syntaxiques de toutes les langues (à l'exclusion des éléments morphologiques par définition singuliers pour chaque langue). Sur ce point voir Archaimbault, Léon (1997), Léon (2000).

sur corpus mises au point par certains groupes Américains. Il est intéressant de noter que la plupart des idées à la fois théoriques et méthodologiques apparaissent pendant cette première période, même si la plupart sont évincées lors de la seconde période pour réapparaître une ou deux décennies plus tard. Lors de cette période toutefois, un certain nombre de commencent à être posés. En effet contrairement à ce qu'avance Josselson (1971), dans un article qui prend le relais de l'évaluation de Bar-Hillel de 1960, les problèmes théoriques de la traduction sont posés dès les années 50-60: le fait que certains mots dans une langue n'aient pas leur contre-partie dans une autre; le fait qu'un mot peut être exprimé par un groupe de mots;⁹ le fait que deux langues puissent avoir des structures complètement différentes ou qu'elles n'aient pas de correspondances univoques au niveau des parties du discours;¹⁰ la grammaire constrative constituant vraisemblablement un des apports principaux de la TA à la linguistique.

La périodisation généralement invoquée par les acteurs (cf. note 2) a tendance à masquer le fait que les idées principales ont été mises au jour dès les années 50 et contribue à leur oubli systématique, et a pour effet d'entretenir l'illusion d'un progrès de la TA.

Dans la période 1960-66, c'est l'analyse syntaxique qui est mise en avant comme la seule voie de recherche possible pour faire avancer la TA. En conséquence, la production en séries de traductions est reléguée à des temps futurs et l'évaluation des sorties de la machine n'est plus à l'ordre du jour. Ce qui va conduire à terme à la suppression des recherches en TA au profit de la seule linguistique computationnelle.

⁽⁹⁾ Cette question avait cependant été posée par Bar-Hillel dans son article 'Idioms' de 1955.

⁽¹⁰⁾ Ce problème avait été abordé dès les années 50 par Mel'cuk dans son élaboration d'une langue intermédiaire (cf. Mel'cuk, 1960).

Sur le plan du TAL, toutefois, la description de la grammaire aboutit à la construction et à la compilation de dictionnaires électroniques dont tous les linguistes pensent qu'ils sont une étape indispensable pour l'analyse syntaxique et la TA. On assiste au développement du 'grammar coding', à savoir du stockage de l'information linguistique, morpho, syntaxique et sémantique, dans chaque entrée de dictionnaire électronique en vue d'un traitement automatique ultérieur. On assiste ainsi à l'apparition des premiers outils du TAL (cf. Léon 1998a, 1999).

Il faut se demander pourquoi l'analyse syntaxique devient ainsi dominante comme méthode exclusive de TA. La première explication tient bien sûr au développement des grammaires formelles, notamment de la grammaire générative,¹¹ apparemment plus facilement programmable dans des parseurs. Et on voit se développer des modèles de grammaires destinées à l'analyse syntaxique pour la TA comme les grammaires de dépendance inspirées de Tesnière, grammaires stratificationnelles par Sydney Lamb. Sont associées à ces modèles des stratégies de parsage syntaxique, dont les plus connues sont la *predictive analysis* et la *fulcrum approach*. On notera que le parsage syntaxique favorise une approche par transfert, préconisée par Yngve (1964), qui suppose la description et la formalisation d'une seule langue à la fois. C'est d'ailleurs pourquoi on a appelé cette période: la 2ème génération par transfert (cf. note 2).

Le second élément est la disparition des méthodes empiriques et probabilistes qui connurent plusieurs adversaires. Bar-Hillel (1960) se déclare contre l'approche empirique

⁽¹¹⁾ On notera en particulier le ralliement de Bar-Hillel à la GGT. Dans son rapport de 1960, il critique sa grammaire catégorielle présentée dans *Language* 29 (1953: 47-58) pour adhérer à la théorie transformationnelle de Chomsky, qui, dit-il, complète le modèle CI et le modèle en phrases noyaux de Harris. Le fait de devoir compléter la grammaire CI par des transformations rend la tâche de la TA plus grande mais devrait, à terme, la faciliter.

qui dénigre les grammaires et les dictionnaires a priori, jugés normatifs et donc de peu d'utilité dans l'analyse de l'écrit. Cette approche se donne comme objectif la construction "manuelle" de règles de grammaire à partir de corpus en langue-source et cumulative en fonction de l'accroissement du corpus. De même les dictionnaires sont issus de la compilation de traductions "à la main" au fur et à mesure. L'objection de Bar-Hillel repose sur le fait que les grammaires traditionnelles sont fondées sur des corpus autrement plus vastes que ceux utilisés dans la TA et il exprime des doutes sur le fait de pouvoir dériver des règles de grammaires de textes réels.

Un autre opposant, c'est Chomsky dans *Structures syntaxiques* qui s'est opposé radicalement aux méthodes statistiques et à l'approche empirique (fin du chap 2) dans son argumentation destinée à voir sur quelles bases distinguer suites grammaticales / agrammaticales. Le premier argument avancé par Chomsky concerne l'impossibilité d'assimiler l'ensemble des phrases grammaticales à un corpus particulier d'énoncés recueilli par le linguiste dans son travail d'enquête. Une grammaire reflète le comportement du locuteur, qui, à partir d'une expérience finie et accidentelle de la langue, peut produire et comprendre un nombre infini de phrases nouvelles. Le second argument réfute qu'une définition de la grammaticalité puisse être fondée sur le sens. Ainsi la phrase *colorless green ideas sleep furiously* n'a pas de sens mais est grammaticale. Le troisième argument postule que la notion de 'grammatical en anglais' ne peut être assimilée à celle d'ordre élevé d'approximation statistique. L'aptitude de quelqu'un à produire et à reconnaître des énoncés grammaticaux n'est pas fondée sur des notions d'approximation statistique. Chomsky donne l'exemple suivant. Dans le contexte *'I saw a fragile -'*, les mots *'whale'* et *'of'* peuvent avoir une fréquence identique (= 0) dans l'expérience linguistique antérieure d'un locuteur. Mais ce locuteur reconnaîtra immédiatement que l'une de ces substitutions, mais non l'autre, donne une phrase grammaticale.

On notera qu'en mettant en avant la notion de compétence du locuteur, se définissant par son aptitude à distinguer une phrase grammaticale d'une phrase agrammaticale, Chomsky se débarrasse de la question du sens, mais il pose aussi une vision anhistorique du langage.

Le développement des grammaires formelles, la mise à l'écart des méthodes empiriques statistiques et sur corpus, et des études sémantiques issues de groupes non américains, non dominants dans le champ de la TA comme application pratique, comme celles des britanniques CLRU ou des russes comme Mel'cuk, ont conforté les Américains dans l'idée que, dans un premier temps, la poursuite des recherches en TA reposait entièrement sur l'analyse syntaxique automatique, voire y était équivalente, et dans un second temps que la TA étant impossible et inutile, pour aboutir à l'idée que seule la linguistique computationnelle, assimilée à l'étude des parseurs syntaxiques, était digne d'intérêt.

2.2. DE LA TA À LA LINGUISTIQUE COMPUTATIONNELLE

Comment s'est effectué cet abandon progressif de la TA au profit de la seule analyse syntaxique? Les problèmes de traduction qui n'étaient pas réglés en appliquant des règles dérivées de la grammaire étaient rassemblés sous le terme de "sémantique" et traités, au mieux, comme des problèmes d'ambiguïtés. La sémantique, on le sait n'avait pas bonne presse dans la linguistique américaine, néo-bloomfieldienne puis chomskienne et les méthodes par langue intermédiaire, britanniques et russes qui commençaient à développer des langages de représentation sémantique n'étaient pas repris. La sémantique devenait l'impasse. C'est ce que stigmatise Yngve (1964) lorsqu'il parle de 'barrière sémantique' à laquelle se heurtent les parseurs syntaxiques. Il en vient à l'idée qu'on ne pourra obtenir une traduction automatique adéquate que

lorsque la machine pourra comprendre, tâche impossible à penser comme objet d'étude à l'époque, malgré le développement des premiers systèmes de compréhension du langage naturel mis au point dans le cadre de l'IA. Ne plus s'occuper des problèmes de sens, implique qu'on abandonne la TA au profit des seuls parseurs syntaxiques.

Ainsi lorsque le rapport de l'ALPAC met fin au financement des recherches en TA aux USA en promouvant à la place la linguistique computationnelle, le terrain était tout à fait préparé. Et on comprend mieux pourquoi le comité comprenait un certain nombre d'anciens acteurs de la TA déjà acquis à la linguistique computationnelle.

Ce soutien à la linguistique computationnelle passe par un soutien explicite à Chomsky et un rejet des néo-bloomfieldiens. Dans l'annexe 19 du rapport, consacrée aux relations entre TA et linguistique, il est mentionné la révolution apportée par Chomsky, dont il est dit qu'il n'a pas à voir directement avec les ordinateurs mais qu'il a apporté des changements fondamentaux grâce au point de vue différent porté par la linguistique sur la nature de la science, sur ce qu'est une théorie scientifique, et sur la relation de l'empirisme à la science.

Le rapport cite à ce propos les actes du 9ème congrès de linguistique qui eut lieu à Cambridge en 1962. Or on sait d'après Murray (1994) que c'est lors de ce congrès que les chomskiens opérèrent leur coup de force pour écarter les néo-bloomfieldiens, notamment Hockett et la phonologie au profit de la syntaxe. Les chomskiens, organisateurs, réservèrent une séance plénière pour Chomsky bien qu'il soit d'une génération plus jeune que les quatre autres conférenciers invités. De même, dans les actes, on lui a réservé plus de pages qu'aux autres.

L'avant dernier paragraphe du rapport témoigne du rejet des néo-bloomfieldiens:

*“Si jamais une simulation par la machine d’une analyse et d’une synthèse linguistique totale devenait possible (ou une traduction orale simultanée), cela ne serait pas à cause de l’adhésion au type de linguistique courante dans les années 50”.*¹²

On notera que, dans ses recommandations, les relations entre machine et humain sont inversées par rapport à ce que propose Bar-Hillel. Alors que celui-ci préconisait une TA humainement assistée, l’ALPAC ne parle plus que de traduction humaine assistée par ordinateur, considérant que même lorsque les sorties de machine sont postéditées, elles sont moins lisibles, même par des scientifiques, que les textes traduits manuellement. Cette assistance consiste dans l’utilisation par les traducteurs de glossaires spécialisés qui le décharge de la tâche fastidieuse de rechercher les termes techniques dans les dictionnaires spécialisés.

Paradoxalement, c’est peut être à ce moment là que la TA a pu être pensée le plus comme une linguistique appliquée: application de la linguistique comme science. Ainsi, l’ALPAC considère que l’apport principal de la TA c’est le passage syntaxique: apport à la linguistique et apport à la programmation. Dans ce rapport est cité un article de Hockett¹³ qui qualifie l’ordinateur de troisième révolution humaine. C’est le premier manipulateur de symboles externe au cerveau humain, susceptible de modifier l’analyse des langues comme le microscope a changé la biologie. Grâce à l’ordinateur, la linguistique peut prétendre, comme la physique, à une mathématisation conséquente. De plus celui-ci a permis de faire le lien entre théorie, études empiriques et applications pratiques.

⁽¹²⁾ *“If ever a machine-aided simulation of total linguistic analysis-synthesis (or voice-to-ear-to-voice translation) becomes possible it will not be because of adherence to the type of linguistic theory widely current around 1950.”*

⁽¹³⁾ Ascher et Hockett, 1964, “The Human Revolution” *Current Anthropology* 5, 135.

Enfin signalons que les débats internes à la linguistique computationnelle sont déjà en cours dans les années 60 et ne feront que se poursuivre après le rapport de l'ALPAC, au moment où la TA devient tabou. Alors que, jusque dans les années 60, toute programmation en TA constituait un tour de force, à cause de l'intrication entre programmation et grammaire qui empêchait toute évaluation linguistique des systèmes et tout progrès, le développement des langages formels va permettre de penser les problèmes de façon déclarative en distinguant la grammaire (la description linguistique), les langages formels (qui rendent les informations linguistiques traitables par la machine) et les stratégies de parsing.¹⁴

3. 3ÈME PÉRIODE (1966-1980): LA SURVIE ET LA “FORCE BRUTE” DES GRANDS SYSTÈMES

Les systèmes qui ont survécu ou qui sont créés, bien sûr ailleurs qu'aux USA, pendant la quinzaine d'années ayant succédé à l'ALPAC, obéissent à des critères très différents de ceux de la période précédente. Il faut distinguer les systèmes industrialisés et les systèmes de recherche. Il reste en effet quelques systèmes de recherche (comme Ariane, développé par le GETA à partir de 1971, ou comme Susy, élaboré à l'Université de Sarrebrück,) qui ont surmonté la crise parce qu'ils venaient seulement de commencer les recherches ou qu'ils étaient soutenus par de grands organismes de recherche, tel le CNRS, préoccupé seulement de façon secondaire par les aspects économiques de la TA.¹⁵ C'est le cas également des systèmes qui répondent à une volonté politique de construire des systèmes multilingues comme Eurotra (1977-1994) pour la communauté européenne, ou le couple Spanam (espagnol-

⁽¹⁴⁾ Cf. sur ce point Josselson (1971) et Kay (1982).

⁽¹⁵⁾ Sur l'histoire des débuts de la TA en France, voir Léon (1998b).

anglais) / Engspan (anglais-espagnol) pour la PAHO (Pan American Health Organisation), lancé en 1976. Dans tous les cas ces systèmes n'ont jamais abouti à une industrialisation.

Le deuxième groupe concerne les systèmes industrialisés qui 'marchent'. Ce sont des systèmes qui répondent à un besoin interne de traductions, limitées à l'information interne d'une entreprise donnée et qui, non destinées à être publiées, peuvent se contenter d'être de qualité médiocre.

3.1. SYSTÈMES DE TA ET LANGAGES RESTREINTS

Le cas de Taum-Météo, qui fait partie de ce second groupe, est intéressant, dans la mesure où il est unanimement reconnu comme un succès et qu'il pose la question des langages restreints. Taum-Météo, mis en oeuvre dès 1965 à l'Université de Montréal, commence à traduire des bulletins météorologiques dès 1977. Ce système a deux caractéristiques qui en font un système en utilisation réelle: il est fondé sur un sous-langage correspondant au style télégraphique utilisé par les rédacteurs des bulletins météo (ne comportant par exemple ni articles ni verbes tensés) ce qui simplifie la traduction. Deuxièmement, c'est le système lui-même qui détermine s'il est capable ou non de traduire une phrase. Sinon il fait appel à un traducteur humain. Ainsi la traduction, quand elle est faite, est entièrement automatique sans intervention humaine. On a pu parler de réussite dans le cas de Météo dans la mesure où le système exécute une tâche de traduction trop fastidieuse pour un traducteur humain et suffisamment contrainte pour que le système ne fasse d'erreurs que dans le cas où l'entrée est mal formée.

La question d'utiliser un sous-langage en TA, langage contrôlé ou sous-syntaxe, pose des questions intéressantes. Dans les années 80, au Japon, l'idée d'élaborer un langage contrôlé analysable par n'importe quel système de TA est con-

sidérée comme bonne non seulement pour les machines mais aussi pour les humains. En effet un sous-langage est susceptible de diminuer les ambiguïtés et d'aider les humains à comprendre les contenus clairement. Ainsi, le vieux mythe de la transparence du langage, véhiculé notamment par le Basic English n'est pas mort.

Le cas du système TITUS est aussi intéressant de ce point de vue. Mis au point par l'Institut Textile de France, le système Titus partage avec Taum-Météo le fait de fonctionner en syntaxe contrôlée et d'être utilisé avec succès de façon industrielle et très spécifique. Il est destiné à traduire des résumés d'articles sur les divers aspects des textiles pour les ingénieurs spécialistes en quête de documentation.

Loffler-Laurian (1996) raconte que, dans les années 80, le CNRS a expérimenté l'utilisation de TITUS pour la traduction des résumés publiés régulièrement dans ses bulletins documentaires. Il est alors apparu que les rédacteurs de résumés ne respectaient pas réellement la syntaxe restreinte autorisée pour le traitement automatique. La question de savoir s'il était possible d'imposer aux rédacteurs une syntaxe contrôlée s'est ainsi posée. Dans sa note 69, Loffler-Laurian conclue: *“le sens de la liberté est extrêmement fort lorsque la langue est concernée”*.

3.2. SYSTRAN ET “LA FORCE BRUTE”

Parmi les systèmes de cette période, le cas de Systran mérite aussi d'être relaté de façon spécifique comme exemplaire. A la fois le plus industrialisé et le plus diffusé, encore actuellement, il n'est fondé sur aucun modèle linguistique cohérent, de telle sorte que ses détracteurs le qualifient de 'bricolage linguistique'.

C'est un descendant du GAT (General Analysis Technique) mis au point par Zarechnak dans les années 50 dans le

cadre du groupe de la Georgetown University. Le GAT était fondé sur une méthode empirique, ce qui lui a valu des critiques virulentes de la part Bar-Hillel (cf.§.1.3). Bien qu'ayant cessé d'être développé faute de financements dès 1964, le GAT a été utilisé entre 1963 et 1973 pour traduire des documents du russe en anglais par deux centres: l'Euratom à Ispra en Italie, et l'Atomic Energy Commission's National Laboratory à Oak Ridge, aux USA. Cette version du GAT (v. Hutchins 1982), appelée par un contemporain "la force brute" avait fini par devenir un programme monolithique d'une très grande complexité, sans séparation entre la partie analyse et la partie synthèse, et donc inutilisable.

Réécrit au début des années 70 par Peter Toma, il prend le nom de Systran. Il comporte surtout des améliorations informatiques, telles la modularité de sa programmation, la séparation entre données linguistiques et procédures informatiques, ce qui permet des modifications locales. Tout en restant linguistiquement "bricolé", il acquiert peu à peu les qualités principales d'un véritable système informatique opérationnel: aisance d'accès, possibilité de mise à jour et couverture.

Dès 1974, il est utilisé par la NASA dans le cadre de la collaboration Apollo-Soyouz. En 1976 il remplace le GAT à l'Euratom, et est acheté par General Motors puis par Rank Xerox. La Communauté Européenne à Luxembourg, son principal utilisateur, l'acquiert également en 1976 et emploie une vingtaine de traducteurs utilisant Systran; actuellement, il travaille sur 12 paires de langues à raison de 1000 pages par heure. Depuis le début des années 90, Systran est commercialisé par la compagnie française Gachot.

Systran est exemplaire dans la mesure où, depuis les années 50, il continue à être utilisé malgré son caractère linguistiquement 'bricolé' et non homogène théoriquement. Il répond toujours au même type de demande, à savoir des traductions brutes, dans un domaine très spécialisé, à usage interne d'une entreprise. Au début, il est soutenu par les mi-

litaires, notamment le Foreign Technology Division de l'US Air Force, qui, malgré les tentatives de dissuasion l'ALPAC, s'est obstiné à utiliser les systèmes de TA. Selon ces militaires, la TA est le seul moyen de traduire de grandes quantités de matériel et de fournir rapidement aux clients des traductions d'informations vite caduques.¹⁶

3.3. EVALUATION DES SYSTÈMES DE TA

Pendant les années 70, la demande en traductions se diversifie. On distingue les traductions qui doivent être publiées et donc faire l'objet d'une post-édition, et celles ne servant qu'à fournir une information à une catégorie spécialisée d'utilisateurs, chercheurs ou ingénieurs devant se mettre à jour dans leur discipline, ou bien à des entreprises devant surveiller leurs concurrents. Enfin un troisième type de traduction concerne des textes très routinisés dans un domaine restreint comme les bulletins météos.

Dans ces années, la question de l'évaluation est posée, non plus dans l'absolu et en termes de pourcentage d'erreur ou par comparaison avec la traduction humaine comme dans la première période, mais par rapport à l'utilisateur final qui devient le critère définitif. Aussi voit-on apparaître les premières grandes évaluations d'envergure comme celle de Systran commanditée par l'US Air Force en 1979-80 (voir Wilks 1992). Celle-ci conclut que la TA ne se justifie économiquement que dans le cas de documents en grand nombre devant être traduits toujours dans la même langue et dans la même direction. La TA est alors considérée comme supérieure à la traduction humaine sur le plan de la rapidité et sur celui de l'exactitude et de la cohérence de la terminologie.

⁽¹⁶⁾ C'est également le cas du système Logos, mis au point en 1971 pour la traduction de l'anglais vers le vietnamien des manuels de maintenance d'équipement militaire.

Dès lors, ce type d'évaluation va être adopté comme le modèle absolu d'évaluation pour la TA. La question de la validité des traductions est un des thèmes principaux d'un congrès sur la TA ayant eu lieu en 1971 à l'Université du Texas. Lors du même congrès, Bar-Hillel, qui a changé de point de vue par rapport à son texte de 1960, déclare:

“Tout programme de TA doit être immédiatement testé quant à ses effets sur l'utilisateur humain. C'est lui le premier juge et le juge en dernière instance et c'est lui qui doit dire s'il est prêt à troquer la qualité pour la vitesse et dans quelles proportions”.¹⁷

C'est également la position d'un autre acteur de la période comme Garvin: “Il est clair que la qualité de la traduction doit être mesurée à l'aune des besoins de l'utilisateur: plus les besoins sont grands, plus il est possible de céder sur la qualité”.¹⁸

Cette position centrale de l'utilisateur final devient cruciale et marquera la TA dans la décennie suivante.

4. LE TOURNANT JAPONAIS ET L'AUTOMATISATION DE LA COMMUNICATION (1980-1990)

Le début des années 80 marque un tournant important. Le développement des micro-ordinateurs et des traitements de texte, leur usage démocratisé et la mondialisation de la consommation favorise l'apparition d'une nouvelle étape pour la TA, celle de la commercialisation. Alors que dans les pre-

⁽¹⁷⁾ *“Every program for MT should be immediately tested as to its effects on the human user. He is first and final judge and it is he who will have to tell whether he is ready to trade quality for speed, and to what degree”*.

⁽¹⁸⁾ *“Clearly, the question of the quality of translation has to be related to user need: the greater the need, the more it is possible to compromise with quality”*.

mières expériences ce qui était visé c'était la traduction scientifique et technique (du russe en particulier), l'orientation est maintenant plus commerciale: il s'agit de traduire des modes d'emplois, des descriptifs de produits ou de machines, conditions indispensables pour leur commercialisation à l'échelon mondial.

Le tournant est impulsé par les Japonais. La TA leur permet de savoir ce que font leurs concurrents, de diffuser leurs produits sur les marchés étrangers et de traduire des manuels d'utilisation, face à une traduction humaine trop chère. En 1982, l'annonce du projet '5ème génération' à l'ICOT (Institute for New Generation Computer Technology) et soutenu par le très puissant MITI (Japan's Ministry of International Trade and Industry) est le point de départ d'un énorme développement. La TA devient une activité prioritaire. Toute entreprise japonaise en informatique investit dans la TA, à la fois sur le plan de la recherche comme du développement. Il faut noter que l'attitude du Japon est radicalement opposée à celle des USA. L'enjeu ne concerne pas simplement un intérêt économique à court terme mais il s'agit d'un pari à long terme sur une société future, fondée sur l'information. A long terme, il s'agit d'intégrer un système de TA sur chaque micro comme faisant partie d'un ensemble de services en ingénierie linguistique.

Par ailleurs, les chercheurs japonais sont attentifs aux caractéristiques de la langue japonaise dans le processus de traduction. Ainsi Nagao (1989), l'un des initiateurs de la TA au Japon, et partisan de la méthode par langue intermédiaire (le langage pivot) du groupe français, le GETA, soutient que les nuances des expressions linguistiques japonaises reflètent des différences culturelles, ce qui exclut d'utiliser le même langage pivot pour traduire à la fois les langues européennes entre elles et le japonais.

Parmi les très nombreux projets développés, citons le projet d'interprétation automatique des appels téléphoniques

(Interpreting Telephony) entre l'anglais et le japonais utilisant la reconnaissance et la synthèse vocales réunissant des Japonais, des Allemands et des Américains. L'intérêt d'un tel projet, qui paraît totalement utopique, c'est l'étude de la simulation de la traduction humaine par la machine, notamment les questions de planification par l'interprète en traduction simultanée. Ce type de recherches, qui n'avaient jamais été tentées jusqu'alors, rejoignent bien sûr l'IA mais sont loin d'être suffisamment avancés pour être intégrés à un système de TA.¹⁹

Parallèlement à ces systèmes extrêmement ambitieux, se développent des systèmes interactifs de traduction automatique assistée par des humains, et des stations de travail de traduction assistée par ordinateur (TAO), offrant aux traducteurs des ensembles d'outils rendus utilisables par la micro-informatique et les logiciels de bureautique.

Par exemple dans le système Cult, lancé en 1968 par la Chinese University de Hong-Kong et destiné à la traduction de textes mathématiques chinois en anglais, la machine interroge un opérateur humain pour résoudre, pendant l'analyse, les ambiguïtés syntaxiques et les homographes, et pendant la synthèse, insérer les articles, les temps et les modes en anglais. Dans le système japonais PECOF, c'est la post-edition qui est interactive: le système cherche à localiser la règle linguistique responsable de l'erreur et propose une révision.

L'apparition de ces dispositifs va modifier de façon considérable le métier de traducteur et être à l'origine de nouveaux métiers. La post-edition, ou révision, exige un savoir-faire particulier plus apparenté à la rédaction technique qu'à

⁽¹⁹⁾ Les recherches commencées en 1984 à l'ATR (Advanced Telephone Research) en collaboration avec le Center for MT de Carnegie Mellon, et l'Université de Karlsruhe, ont connu des développements divers, dont le projet allemand *Verbmobil* (cf. Hauenschild et Heizmann 1997).

la traduction et ne demande aucune capacité de création (littéraire ou non). La post-edition suppose de garder le maximum de la sortie de la TA et de se rendre directement aux points stratégiques du texte plutôt que de le refaire entièrement. Le post-éditeur apprend à développer un ensemble de techniques dépendantes du contexte pour traiter les configurations produites par la machine. Il doit dominer toute la terminologie technique, capturer toutes les nuances, s'assurer de l'équivalence idiomatique, maintenir un registre approprié de façon uniforme, mettre de la cohérence et de la cohésion là où il n'y en a pas, préserver la structure d'information. Les traducteurs et réviseurs maison garantissent la bonne utilisation de formulations propres aux différents corps de l'organisation et permettent d'éviter les hypercorrections abusives. Parfois même, la fonction du réviseur se rapproche de celle de programmeur lorsqu'il doit se former au fonctionnement du système de façon à intervenir sur son amélioration (v. Pigott 1982).

La "customization" du système de TA, qui consiste à mettre au point des procédures d'ajustement des dictionnaires et des grammaires à un domaine spécifique ou à un document particulier demande aussi un savoir-faire particulier. On voit apparaître de nouveaux métiers comme celui d'ingénieur linguiste.

Quant aux stations de travail pour traducteurs, elles offrent un ensemble d'outils qui modifient considérablement le métier de traducteur: affichage des textes source et cible sur l'écran; possibilité de créer, pour un ensemble de documents spécialisés, des dictionnaires terminologiques bilingues spécifiques; consultation de banques de données terminologiques ou de glossaires bilingues en ligne; utilisation de logiciels de fréquences et de concordances; interface avec des systèmes de TA.

Dans les années 90, les mémoires de traduction et la recherche de segments de textes stéréotypés seront utilisées

pour la traduction ou la génération de textes techniques ou commerciaux très formalisés et très répétitifs. Ces stations bénéficieront également des ressources technologiques disponibles sur tous les micro-ordinateurs, tels que les CD-ROM, l'accès au réseau ou les transmissions électroniques.

Avec les années 80, la TA n'est plus confinée à une utilisation maison mais fait partie de l'ingénierie linguistique, au même titre que les dictionnaires électroniques, les bases de données terminologiques et la génération de texte, qui constitue son application privilégiée notamment dans le domaine de la production de textes multilingues. Elle participe d'un mouvement plus général, celui de l'automatisation de la communication, initié dans les années 80 et qui affecte l'ensemble de la société. L'idée de fournir des traductions entièrement automatisées est toutefois révolue. La traduction assistée par ordinateur n'a même rien à voir avec un système de traduction automatique. C'est le traducteur qui est au poste de commandes et on parlera plus volontiers de traduction automatisée que de traduction automatique.

5. DEPUIS 1990, LE RETOUR DES MÉTHODES EMPIRIQUES

Les années 90 voient le renouveau de l'intérêt pour la TA aux USA. Celui-ci est d'abord d'ordre économique: nécessité de traduire la documentation des concurrents japonais, baisse des coûts de l'informatique et hausse des coûts de la traduction surtout pour les langues à alphabets non latins comme le japonais. Ce renouveau est toutefois limité: L'ensemble des centres qui se consacrent à la TA, partiellement financés par la NSF (National Science Foundation), sont bien moins subventionnés qu'un seul centre japonais.

Ce qui caractérise ces systèmes c'est qu'ils mettent en oeuvre de plus en plus des stratégies et des techniques mixtes contrairement aux systèmes des années 60 et 70 qui privi-

légiaient les stratégies par transfert et l'analyse syntaxique. Des modules "intelligents" et des modules probabilistes viennent compléter les modules d'analyse linguistique plus traditionnels. Les systèmes américains les plus élaborés, comme le CMU (Center for Machine Translation de l'université Carnegie Mellon) et le KMBT (Knowledge-Based Machine Translation) intègrent des techniques de l'intelligence artificielle comme les systèmes-experts et les bases de connaissances intelligentes. En Europe, aux Pays-Bas, on trouve des systèmes originaux comme Rosetta, conçu chez Philips à Eindhoven développant une interlangue fondée sur la grammaire de Montague; ou le système DLT (Distributed Language Project) utilisant l'espéranto comme interlangue et des corpus alignés pour résoudre les ambiguïtés.

Mais ces systèmes sont loin de pouvoir être commercialisés et le clivage entre stations d'aide à la traduction et systèmes de traduction s'accroît (pour un état de l'art en 1991, v. Nirenburg 1993).

Mais surtout, une des caractéristiques des années 90, c'est le retour des approches empiriques et du traitement statistique de grands corpus (V. Church et Mercer 1993). Ce renouveau, suscité par le succès des méthodes stochastiques dans le traitement du signal et la reconnaissance de la parole dans les années 70, a été renforcé par le regain d'intérêt des linguistes pour le lexique au début des années 80. Par ailleurs, des corpus de données textuelles importants deviennent disponibles grâce aux nouvelles possibilités de traitement des ordinateurs et aux efforts de standardisation des textes (langages TEI et HTML).

Plusieurs méthodes de TA utilisent les mémoires de traduction (v. Somers 1992). Dans le système développé par les Japonais Sato et Nagao, la traduction s'effectue par la comparaison entre la phrase à traduire et des exemples stockés dans une base de données. Enregistrés sous forme d'arbres de dépendance, ces exemples sont comparés aux phrases d'en-

trée par des procédures de pattern-matching. Un module statistique permet ensuite de choisir entre plusieurs candidats.

Les mémoires de traduction peuvent aussi être constituées à partir de corpus bilingues alignés (ou synchronisés), tels le corpus anglais-français des actes parlementaires canadiens, les *Canadian Hansards*, ou les rapports économiques de l'*Union Bank of Switzerland*. L'alignement, c'est-à-dire la mise en correspondance des phrases, est effectué à partir de critères exclusivement statistiques utilisant le fait que les phrases longues dans la langue source ont tendance à être traduites par des phrases longues dans la langue cible et que les phrases courtes ont tendance à être traduites par des phrases courtes.

Ces approches utilisent des dictionnaires d'équivalences eux-mêmes constitués par l'application de méthodes probabilistes. Par exemple, Brown et al. (1990) ont calculé la probabilité pour qu'un mot d'une phrase dans une langue donnée corresponde à zéro, un ou deux mots dans la traduction. Par exemple *the* a une probabilité de 0.610 d'être traduit par *le* et de 0.178 d'être traduit par *la*. Le glossaire ainsi obtenu consiste en des listes de mots avec leurs probabilités de traduction.

Ces méthodes constituent une rupture par rapport aux hypothèses plus ou moins explicites partagées par les approches traditionnelles de la TA. Comme l'indique Somers (1992), elles s'opposent à l'approche compositionnelle selon laquelle la traduction de l'ensemble d'une phrase serait fonction de la traduction de ses parties; approche, qui en s'attachant à préserver en premier la structure de la phrase, est à l'opposé de la pratique du traducteur humain qui ne préserve la structure qu'en dernier ressort. Par ailleurs, jusqu'à présent tous les systèmes de TA ont été construits avec l'idée sous-jacente que le texte source contient assez d'informations pour permettre sa traduction. Autant d'hypothèses que ne partagent pas les mémoires de traduction qui fournissent des traduc-

tions a priori. Alors que la linguistique computationnelle, de même que les approches issues de l'intelligence artificielle, favorisent une approche 'rationnelle' fondée sur des règles (principle-based parsing), les approches 'empiriques' montrent l'importance du lexique, notamment des collocations de mots, pour résoudre certaines contraintes locales (preference-based parsing). Par exemple, l'étiquetage automatique de textes en parties du discours utilisent des probabilités lexicales (considérées comme un dictionnaire) et des probabilités contextuelles (considérées comme une grammaire) pour déterminer des préférences en cas d'ambiguïtés. C'est ainsi que les préférences lexicales permettent de résoudre 90% de l'étiquetage.

Ce recours au lexique a pourtant ses limites quand certains auteurs s'évertuent, dans le but de construire une langue intermédiaire, à extraire des primitives sémantiques à partir des définitions de dictionnaires élémentaires comme le LDOCE (Longman Dictionary of Contemporary English) et de son équivalent espagnol VOX. La consultation des travaux des lexicologues auraient évité à ces auteurs la déconvenue de s'apercevoir que le sous-langage des définitions n'est pas aussi simple et limité qu'ils le pensaient et qu'un "Machine-readable dictionary" peut opposer la même résistance que le langage en général.

CONCLUSION

Kay (1982) conçoit les systèmes de TA et les parseurs syntaxiques comme des méthodes d'évaluation des modèles linguistiques par la technologie.²⁰ Il préconise de continuer à travailler dans le domaine de la traduction entièrement auto-

⁽²⁰⁾ Kay, un pionnier britannique de la TA, a promu, dès 1958 l'idée que la validité d'une description linguistique doit toujours être liée à son utilité et que cette linguistique n'est pas la même selon les objectifs poursuivis.

matique parce que c'est là qu'on apprend sur le langage en général et la traduction en particulier. La TAO, où les enjeux sont davantage la communication entre traducteurs ou le développement de dispositifs spéciaux d'édition que les problèmes linguistiques ne peut constituer un produit dérivé naturel de la TA.

Pour Loffler-Laurian (1996), la TA fournit l'occasion à la linguistique de fonctionner à grande échelle, avec un gros appareillage, comme les grandes sciences physiques.

Mais face à ces conceptions ambitieuses de la TA, on constate une certaine déconvenue qui met fortement en cause son rapport à la linguistique. Des auteurs comme Kay (1992) déplorent le peu de progrès accomplis depuis les années 60, en constatant que les systèmes se sont beaucoup améliorés sur le plan de la vitesse mais peu sur le plan de la qualité, que la machine produit des traductions de niveau bien inférieur à celles produites par des traducteurs humains, demandant l'intervention humaine à tous les niveaux, et qui auraient été jugées inacceptables si elles avaient été rédigées par un traducteur humain. Certains vont même jusqu'à expliquer l'apparent succès de la TA par la seule supériorité de la demande sur l'offre. Rappelons qu'en 1993, les textes commerciaux et techniques constituent plus des deux tiers de la demande en traductions.

Le second constat d'échec concerne l'incapacité de la linguistique à produire des systèmes performants de TA. Cet est d'ailleurs échec parfois difficile à reconnaître par les tenants de l'analyse syntaxique parmi les acteurs de la TA. Ce constat repose sur le fait qu'aucun des systèmes récents les plus robustes linguistiquement, Ariane, Rosetta, DLT et SUSY, n'ont pu être commercialisés, faute de résultats suffisants. Tout au plus les systèmes de TA fondés sur des modèles linguistiques puissants peuvent-ils servir de bancs d'essai aux théories linguistiques computationnelles. C'est le cas notamment de Rosetta.

Comme le rappelle Melby (1992), en citant les conseils que lui a donné Hays, un des pionniers de la TA, en 1972, ce n'est pas en trouvant le bon modèle formel en linguistique qu'on va résoudre les problèmes de TA, c'est en fournissant du travail solide fondé sur une linguistique simplifiée. C'était aussi l'intuition de Peter Toma, le concepteur de Systran, qui, contrairement aux scientifiques de l'époque, ne croit pas que la linguistique puisse fournir une solution adaptée au traitement du langage par l'ordinateur. Il est convaincu que le traitement du langage doit être adapté aux possibilités de l'ordinateur plutôt que l'inverse (Loffler-Laurian, 1996).

Que penser également de ces nouveaux systèmes, fondés sur des méthodes probabilistes qui se prévalent de ne pas utiliser de savoir linguistique? Enfin doit-on interpréter le fait que ces auto-critiques soient formulées dans les années 90, comme le signe que le spectre de l'ALPAC est enfin conjuré? La réponse à ces questions passe par le constat d'un double paradoxe: on n'exige plus des systèmes de TA fondés linguistiquement et qui ont reconquis leur légitimité dans la linguistique computationnelle de fournir des traductions; quant aux stations d'aide aux traducteurs, elles n'utilisent les systèmes de TA, quand ils existent, que comme un outil parmi d'autres. Autrement dit, on en arrive à se demander si l'automatisation de la traduction a encore réellement besoin de la traduction automatique.

BIBLIOGRAPHIE

- AKHMANOVA O.S., Mel'cuk I.A., Frumkina R.M., Paducheva E.V., 1963, *Exact Methods in Linguistic Research*, U. Of California Press, Berkeley.
- ARCHAIMBAULT S. et Léon J., 1997, "La langue intermédiaire dans la Traduction Automatique en URSS (1954-1960). Filations et modèles", *Histoire Epistémologie Langage* 19-2: p. 105-132.
- BAR-HILLEL Y., 1955, "Idioms", in *Machine Translation of Languages*, 14 essays, (W.N.Locke and A.D. Booth, eds.), MIT et John Wiley: p. 183-193.

- BAR-HILLEL Y., 1960, "The present Status of Automatic Translation of Languages". *Advances in Computers*, vol.1, (F.C. Alt ed.) Academic Press, N.Y., London: p. 91-141.
- BROWN P., COCKE J., DELLA PIETRA S., DELLA PIETRA R., JELINEK F., LAFFERTY J., MERCER R., and ROSSIN P., 1990, "A statistical approach to machine translation". *Computational Linguistics*, vol. 16, n° 2, p. 79-85.
- CHURCH KENNETH W. and ROBERT L. Mercer, 1993, "Introduction to the Special Issue on Computational Linguistics Using Large Corpora". *Computational Linguistics*, vol. 19, n° 1, p. 1-25.
- DOSTERT L. E., 1955, The Georgetown-IBM experiment, in *Machine Translation of Languages*, 14 essays, (W.N.Locke and A.D. Booth, eds.), MIT et John Wiley, p. 124-135.
- HAUENSCHILD CHRISTA AND SUSANNE HEIZMANN (eds.), 1997, *Machine Translation and Translation Theory*. Berlin: Mouton de Gruyter.
- HARRIS Z.S., 1955, "From phoneme to morpheme". *Language* 31, p. 190-222.
- JOSSELYN H. H., 1971, "Automatic translation of languages since 1960: a linguist's view". *Advances in computers*, 11, 1-58. London: Academic Press.
- KAPLAN ABRAHAM, 1950, "An experimental study of ambiguity in context", *Mechanical Translation*, v. 1, n. 1-3.
- KAY M., 1982, "Machine Translation", *American Journal of Computational Linguistics*, vol. 8, n° 2, p. 74-78.
- KAY M., 1992, Préface de *An Introduction to Machine Translation*, éd. par W. John Hutchins et Harold L. Somers, London: Academic Press Ltd.
- LANGUAGE AND MACHINES. *Computers in translation and linguistics*. 1966. A report by the Automatic Language Processing Advisory Committee (ALPAC), National Academy of Sciences, National Research Council.
- LÉON J., 1997, "Les premières Machines à Traduire (1948-1960) et la filiation cybernétique", in *TAL et sciences cognitives* (H. Madec, ed.), *BULAG*, n° 22, p. 9-34.
- LÉON J., 1998a, "Les premiers outils pour la Traduction Automatique: demande sociale, technologie et linguistique (1948-1960)". *BULAG* n° 23, p. 273-295.
- LÉON J., 1998b, "Les débuts de la traduction automatique en France (1959-1968): à contretemps?". *Modèles Linguistiques*, tome XIX, fascicule 2, p. 55-86.

- LÉON J., 1999, "La mécanisation du dictionnaire dans les premières expériences de traduction automatique (1948-1960)", *History of Linguistics 1996*. vol. II (D. Cram, A. Linn, E. Nowak eds.), p. 331-340 John Benjamins Publishing Company.
- LÉON J., 2000, "Traduction automatique et formalisation du langage. les tentatives du Cambridge Language Research Unit (1955-1960)", in *The History of Linguistics and Grammatical Praxis* (eds. P. Desmet, L. Jookken, P. Schmitter, P. Swiggers) Louvain / Paris: Peeters, p. 369-394.
- LOFFLER-LAURIAN A-M., 1996, *La traduction automatique*. Septentrion, 157 pp.
- MELBY ALAN, 1992, "The translator workstation". *Computers in Translation, A practical Appraisal*, éd. par John Newton, p. 147-165, London: Routledge.
- MEL'CUK, I. A. 1960, "K voprosu o grammaticeskom v jazykeposrednike", *Masinnij Perevod i Prikladnaja Linguistika* 4, 25-451 [Trad. angl. The problem concerning the "grammatical", "in an intermediate language", JPRS/8026, archives ATALA].
- MURRAY STEPHEN O. 1994, *Theory groups and the study of language in North America*. Studies in the History of the Language Sciences 69, John Benjamins Publishing company.
- NAGAO MAKATO, 1989, *Machine Translation: How Far Can It Go?* Oxford University Press.
- NEWTON JOHN ed., 1992, *Computers in Translation. A practical Appraisal*. London: Routledge, 238 pp.
- NIRENBURG SERGUEI (ed.), 1993, *Progress in Machine Translation*. IOS Press, Amsterdam: Netherlands.
- PIGOTT IAN M., 1982, "The importance of feedback from translators in the developement of high-quality machine translation". *Practical experience of machine translation*. ASLIB 1982, éd. par V. Lawson, 61-74, Amsterdam: North-Holland.
- SOMERS HAROLD L., 1992, "Current research in MT". *Computers in Translation. A practical Appraisal*. éd. par John Newton, 189-207, London: Routledge.
- WEAVER WEAVER, [1949] 1955, "Translation" in *Machine Translation of Languages*, 14 essays, (W.N. Locke and A.D. Booth, eds.), MIT et John Wiley, p. 15-23.
- WILKS YORICK, 1992, "Systran: it obviously works but how can it be improved?" *Computers in Translation, A practical Appraisal*, éd. par John Newton, 166-188. London: Routledge.

YNGVE V. H., 1964, "Implications of Mechanical Translation Research".
Proceedings of the American Philosophical Society, vol. 108, n° 5.

RÉSUMÉ: *Cet article aborde les divers périodes de la traduction automatique qui amènent à la période contemporaine.*

Mots-clés: *traduction automatique, science du langage, linguistique computationnelle.*