

Pourquoi réunifier la traductologie?

Monique Slodzian

Professeur émérite à l'INALCO. Centre de recherches ERTIM

Introduction

La concomitance entre le règne de la sémantique positiviste et la promotion euphorique de la traduction automatique, au tournant de la deuxième guerre mondiale, constitue le point de départ de la réflexion qui suit. Le paradigme dominant en était la tripartition syntaxe/sémantique/pragmatique - peu contesté jusqu'ici par le TAL- qui, à la suite des travaux de Chomsky ou de Šaumjan, abritait sous son aile les recherches sur les langages pivots, reprenant la vieille utopie d'une *interlingua* universelle, indépendante des langues particulières. Le présupposé d'un niveau conceptuel extérieur aux productions langagières sur lequel était fondé le réductionnisme prôné par l'empirisme logique (Slodzian, Hermès 49, 2007) devenait ainsi pour longtemps la pierre angulaire de nombreux projets proposés par le TAL, qu'il s'agisse des ontologies multilingues (WordNet et EuroWordNet) ou de la théorie des sous-langages techniques (Picht et Draskau, 1985; Kittredge et Lehberger, 1982; Grishman et Kittredge, 1986).

Il instituait *de facto* une coupure théorique infranchissable entre traduction humaine et traduction machine. Ce divorce était peu questionné en raison du statut de scientificité que l'épistémologie néo-logiciste (de Frege à Carnap et à Wüster) accordait à « la langue pour la science », dégradant du même coup la traduction humaine au rang d'artisanat aussi douteux que le langage ordinaire lui-même. Cette dichotomie qui mettait hors-jeu la problématique de la traduction humaine avait en outre l'avantage de retirer aux traducteurs professionnels toute légitimité pour juger des résultats de la traduction machine. Pour ce nouveau pôle technologique qu'était la traduction automatique, dans lesquels les pays développés avaient massivement investi, il n'était pas négligeable de mettre en sourdine les critiques, justifiées ou non, qui pointaient les piètres performances des systèmes en lice. C'est ainsi que, sous prétexte d'une normalisation culturelle commandée par la mondialisation des échanges et l'essor des technologies de la communication, la traduction automatique s'est, dès ses débuts, soustraite au champ de la traductologie et à toute déontologie réaliste.

Notre thèse est qu'en retirant aux études de traductologie toute légitimité pour juger à la fois de la qualité des textes traduits automatiquement et de la pertinence des procédures humaines (pré- et post-édition) impliquées dans le processus, on a figé pour longtemps le champ des études. A l'évidence, l'option théorique qui fondait la traduction automatique devait demeurer un socle intangible, quels que soient les résultats obtenus, étant implicitement admis que les problèmes classiques de la traduction humaine seraient un jour dépassés par la traduction automatique.

De fait, les premières brèches sont venues de l'informatique computationnelle elle-même qui, soumise à la contrainte des résultats en traduction machine, a dû proposer des solutions plus performantes et surtout moins coûteuses. Ainsi, la traduction humaine a-telle été subrepticement réintroduite par le biais de processus semi-automatisés appelés traduction « assistée » ou encore « mémoire de traduction ». Or l'évolution des systèmes à base de règles vers ces systèmes hybrides (T.A.O, modèles statistiques,...) ne requérait plus les postulats théoriques de l'empirisme logique.

Nous nous demanderons pourquoi ce tournant empirique a eu si peu de conséquences sur la théorie et la pratique de la traduction automatique. Les nouvelles techniques dites « hybrides » s'accommodent en quelque sorte d'une version faible de la sémantique vériconditionnelle, comme si les échecs constatés à divers moments par d'éminents théoriciens de la T.A. (Bar Hillel, Hutchins, par exemple) ne pouvaient remettre en cause la conception sémiotique qui présidait aux principes de la solution automatique.

Ce défaut d'attitude réflexive devient encore plus flagrant lorsque, à la fin des années quatre-vingt, la disponibilité de grandes collections de textes, unilingues et bilingues, permettent à la traduction automatique de s'émanciper du modèle à base de règles. La réalité physique des textes, centrale pour la linguistique de corpus, ouvre une problématisation différente du couple langue/texte. Force est de constater, que le travail de reconception d'une linguistique au palier du texte n'a été mené qu'à la marge. Le changement radical que constitue l'approche statistique par rapport à l'approche par règles n'a entraîné qu'une révision superficielle du paradigme de l'empirisme logique.

On se demandera alors si la traductologie contemporaine, confinée malgré elle dans la problématique de la traduction humaine pourra, ou saura, tirer les conséquences théoriques des avatars de la traductique, la traduction humaine ayant elle-même stagné dans des théorisations faibles, dérivées du paradigme communicationnel théorisé par la pragmatique positiviste (Morris).

En quoi une traductologie, intégrant les acquis pour solde de tous comptes de la traduction automatique depuis sa période intégriste jusqu'à sa période réformiste (intégrant l'assistance humaine) et, enfin, l'avènement des corpus (alignés et comparables), nous semble-t-elle nécessaire?

Parce qu'un renouvellement de la traductologie -dans la mesure où la traduction requiert une linguistique du texte- est central pour le dépassement d'une linguistique du signe encore prédominante. Au plan pratique, si modestes soient-elles, les tentatives qui ont été menés à l'INALCO pour outiller une linguistique textuelle ne peuvent qu'asseoir cette conviction. Le TAL dans son ensemble est partie prenante du débat.

Le règne de la sémantique positiviste

Cicéron déjà posait les fondements d'une sémiotique de la traduction : *Je n'ai pas cru nécessaire de rendre mot pour mot; c'est le ton et la valeur des expressions dans leur ensemble que j'ai gardés. J'ai cru qu'il me fallait payer le lecteur non pas en comptant pièce par pièce, mais pour ainsi dire en pesant la somme en bloc.*(Du meilleur genre d'orateur, 46 avant J.-C, texte établi par Henri Bornecque, Les Belles Lettres, 1921)¹.

¹ Cité par Ballard M., 1998, Introduction et édition annotée de l'ouvrage de Bachet de Méziriac, *De la traduction*, 1635, Presses de l'Université d'Ottawa)

Un plaidoyer contre le littéralisme qui aurait dû épargner aux générations de traducteurs qui se sont succédé le besoin de justifier les ajouts ou les retranchements qu'ils opéraient sur le texte à traduire. Souvent empêtrés dans une vision idéalisée de la traduction, ils estimaient devoir rendre des comptes en expliquant les raisons pour lesquelles ils avaient fait tel ou tel choix, ayant étoffé ici, retranché là. En même temps, ces excuses témoignent qu'ils avaient conscience de deux faits essentiels. D'une part que les langues n'ont pas la même concentration et, surtout, que la translation de textes d'une langue vers une autre suppose une phase d'interprétation. Pour dire les choses autrement, la pratique de la traduction à un niveau professionnel suppose une connaissance intuitive du statut sémiotique des langues et des textes. Ce qui nous amène à affirmer l'importance théorique de la traductologie pour la linguistique: elle lui impose ni plus ni moins le statut sémiotique des langues et des textes².

Comment expliquer que l'on ait eu à se justifier de fuir le mot-à-mot? C'est qu'il en va de même de la théorie du sens comme de la loi de la gravitation universelle! Jusqu'à Newton, on croyait que certains corps tombaient, tandis que d'autres s'élevaient vers le ciel...En linguistique, on croyait que les mots étaient des étiquettes de concepts ou d'objets distincts jusqu'à ce que Saussure rejette le dogme de la référentialité. Mais les croyances sont tenaces et la vision référentielle du langage persévère sans fléchir.

L'utopie de la machine à traduire lui a redonné une légitimité inespérée qui justifie des croyances et des pratiques relevant d'une théorie du sens inadéquate. Que l'on songe au mythe encore répandu du chinois comme langue d'avenir en raison de la supposée référentialité des idéogrammes et de la simplicité de sa grammaire ou encore du succès mondial des ontologies multilingues à la WordNet. Comme s'il suffisait de construire de gigantesques ressources multilingues au bon format pour résoudre le problème de la traduction. Les convictions théoriques des premiers chercheurs en TAL étaient à l'avenant, laissant croire que, pour travailler sur le langage, il suffisait de procéder à une simple substitution de données: les données numériques seraient remplacées par les données alphanumériques que constituaient les mots.

La Traduction Automatique des débuts était donc essentiellement une traduction de mots. D'où l'importance énorme des dictionnaires, la part réduite au minimum dévolue à la syntaxe et le silence sur le texte. Les problèmes relatifs au sens d'une phrase, d'un texte, n'étaient pas abordés en tant que domaine de recherche spécifique, pas plus que l'activité de traduction n'était envisagée comme une activité ayant sa problématique propre. Comme si les problèmes théoriques de la traduction machine devaient se résoudre par la construction d'ensembles finis et complets de données : en stockant les mots et leurs sens on a cru avoir rempli la plus grande partie de la tâche et on a ainsi nourri l'illusion que l'on parviendrait à un dénombrement satisfaisant des catégories morphologiques, syntaxiques et sémantiques dont sont faites les langues.

A partir des années cinquante, le mythe de la traduction automatique a rendu inaudible la vieille réponse que, à l'instar de Cicéron, les bons traducteurs ont donnée à la question de la qualité première d'une traduction - à savoir la fidélité à tout le texte et non à des signes isolés.

2 Cette réflexion renvoie à l'article magistral de François Rastier « Linguistique interprétative et fondements sémiotiques de la traduction », dans *De la linguistique à la traductologie*, Tatiana Milliaressi ed. Presses universitaires du Septentrion, Lille.

La conception traditionnelle du signe et de la signification³ qui a longtemps contredit l'intuition des traducteurs s'est imposée sans droit de réplique, sous une forme souvent outrageusement simplifiée. On trouvera dans l'oeuvre de François Rastier une analyse exceptionnellement forte d'un phénomène qui entrave le développement des sciences du langage dans leur ensemble. Voici ce qu'il dit de la sémantique positiviste, fondatrice de la technologie du traitement automatique des langues:

*« La tripartition sémiotique syntaxe/sémantique/pragmatique marque l'incidence majeure du positivisme logique sur les sciences du langage. Elle est réputée les diviser sans reste. Elle constitue pour leur développement, depuis quarante ans, l'obstacle épistémologique principal. Notamment, elle est fondée sur un paradigme du signe (selon la sémiotique de Morris et Carnap) et non du texte. Encore ce signe est-il réduit à une simple expression ».*⁴

A quelles conditions le coup de force de la sémantique positiviste a-t-il été possible?

Pour cela, il faut aller voir de plus près du côté des promoteurs de la traduction automatique. Soucieux de rassurer la communauté des traducteurs, Warren Weaver et Victor Ingve du MIT, aussi bien que Booth et Leonard Brandwood de l'équipe anglaise, de même que leurs collègues soviétiques, répètent à l'unisson qu'ils comptent s'en tenir à la traduction des textes scientifiques et remettent à plus tard les problèmes touchant à la traduction des textes littéraires. L'argument est bien sûr que les difficultés sémantiques des premiers sont bien moindres. Au VIII^e Congrès international des Linguistes, en 1957, le rapporteur de la section « Machines à traduire » pose que « la traduction qui sera exécutée par la machine est, certainement pour ce qui est du présent,⁵ considérée comme devant être celle de textes scientifiques et techniques seulement, eu égard aux difficultés supplémentaires posées, par exemple par les textes littéraires »⁶.

C'est l'impératif technologique de progressivité qui commandait cette précaution, plus que l'argument théorique. Le renoncement n'est donc que provisoire, si l'on en croit Brandwood: « La question de faire passer un chef d'oeuvre de littérature écrit en langue étrangère, dans une traduction digne de respect, est une question d'une grande difficulté. On a soutenu l'opinion extrême qu'une telle opération n'est généralement pas possible même pour un spécialiste humain, et, par conséquent, encore moins pour la machine. Cette vue nous semble ultra-pessimiste ». Raisonement qui, accessoirement, disculpera les traductions médiocres livrées par la machine et qui traite avec la plus grande désinvolture le trésor séculaire constitué par les millions d'ouvrages traduits dans le monde. En bref, la machine échoue, mais l'homme aussi.

Selon les promoteurs de la traduction automatique, à quoi tient « la grande difficulté » des textes littéraires? Les uns invoquent les « valeurs sentimentales » du langage opposées aux « valeurs logiques » qui caractérisent les textes techniques ou scientifiques, lesquels incluent les « informations économiques, législatives ou politiques ». Et les grands linguistiques qui

3 Rastier, François, 2010-2011, *Texto!* Vol XV n°4, vol XVI n°1

4 Rastier, François, 1994, « Sur l'immanence en sémantique », paru dans les *Cahiers de Linguistique Française*, pp325-335

5 Souligné par l'auteur de ce texte

6 *Proceedings of the VIII International Congress of Linguists*, Oslo University Press, 1958

ont accompagné l'essor de la traduction automatique, comme Bloomfield, Sapir ou Morris ne sont pas en reste puisque, pour le premier, la langue littéraire n'est, semble-t-il, qu'un usage particulier des moyens expressifs de chaque langue; Sapir, pour sa part, se dérobe prudemment dans une note: « Je ne puis m'attarder à définir ce qu'est « une forme plus délicatement expressive », qui mérite d'être appelée littérature, ou art; et, qui plus est, je ne sais pas exactement, ce qu'il en est. Il nous faut admettre *a priori* ce mot de littérature »; la position de Morris n'est pas moins évasive: « l'expressivité des signes est (...) une propriété additionnelle de ces signes, en plus et au-delà de leur signification ».

Au bilan, aucun critère sérieux, linguistiquement et culturellement établi, qui permette de distinguer la langue littéraire d'avec la langue commune. Comment ne pas être frappé par la faiblesse de ces analyses? Et les linguistiques de la génération suivante qui s'attaqueront au problème, en particulier en France avec le fonctionnalisme de Martinet⁷ et les travaux de stylistique de Bailly⁸ n'iront guère au-delà des généralités. On en reste à l'opposition fonction expressive versus fonction esthétique, qui commande celle qui est censée distinguer la langue scientifique ou technique de la langue littéraire ou poétique.

Circularité du raisonnement. Georges Mounin dresse ainsi le bilan des années structuralistes: « Traduire les textes littéraires c'est donc déceler les connotations, puis déceler les traits esthétiquement pertinents (formels ou sémantiques) par lesquels elles sont présentes dans le texte source, puis trouver les moyens correspondants qui seront poétiquement ou littérairement pertinents dans le texte-cible (...) »⁹. Autant dire que l'on ne quitte pas le cadre de l'opposition dénotation versus connotation, conforme à la tripartition « langue pour la science », « langage ordinaire », « langue littéraire », institutionnalisée par l'influente école de terminologie de Vienne¹⁰. Et, quoiqu'on en pense, le fonctionnalisme plus sophistiqué d'un Jakobson ne fera que déplacer l'accent sur le troisième volet, à savoir la pragmatique. C'est l'essor de la pragmatique qui explique le succès des différentes écoles d'analyse du discours se réclamant de la socio-linguistique au cours des années 80-90.

Par ailleurs, les oeuvres de linguistes se réclamant du logicisme du Cercle de Vienne comme Ogden et Richards¹¹ ou Catford¹² ont eu un retentissement considérable qui a encore renforcé les positions de la traduction automatique. Sans chercher ici à approfondir ces connexions, rappelons qu'Ogden et Richards sont les promoteurs du Basic English, appelé à devenir la *lingua franca* de la communication universelle. Les auteurs distinguent dans le langage deux fonctions, la fonction référentielle dont le but est de parler du monde de façon adéquate et la fonction émotive, source d'erreurs et de confusion (d'où l'invitation de Richards dans sa théorie des fictions à « expulser la rhétorique mystificatrice »). Leur sémiotique consiste en

7 Martinet, A., 1960, *Éléments de linguistique générale*, Armand Colin

8 Bailly, Charles, 1909, *Traité de stylistique française*

9 Mounin, id., 1976, *Linguistique et traduction*, Bruxelles, Dessart et Mardaga

10 Slodzian, M., 1994-1995, « La doctrine terminologique, nouvelle théorie du signe au carrefour de l'universalisme et du logicisme », *ALFA*, Halifax, Nova Scotia, Canada.

11 Ogden, C.K., Richards, I.A., *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*, New York, Harcourt, 1946.

12 Catford, J.-C., *A Linguistic Theory of Translation: An Essay in Applied Linguistics*, Oxford University Press, 2000 (1ère éd. 1965).

une « science de la symbolisation » (*Science of Symbolism*) déterminant l'emploi correct des mots dans le raisonnement. Quant à l'ouvrage magistral de Catford, *A Linguistic Theory of Translation*, tout en s'affichant comme une théorie linguistique, il relève clairement des langages formels et des propositions de Carnap dans *Logical Syntax of Language*. En concevant l'équivalence traductive comme un transfert de règles grammaticales instanciant des entités fixes susceptibles d'acculturation (en chinois, russe, etc;), il se place hors du champ de pertinence de la traduction concrète et, sous couvert de théorie linguistique, justifie les fondements théoriques de la traduction automatique¹³.

De la traduction automatique à la traduction assistée: une révolution silencieuse

Trois options se dégagent des travaux des linguistes anglo-saxons des années 50 à 70 consacrés à la traduction automatique:

- soit le modèle de langue à langue, appelé traduction directe, fondé sur des règles linguistiques écrites par des linguistes qui analysent la structure grammaticale de la phrase en langue source (arbre syntaxique), construisent un arbre syntaxique équivalent en langue cible et remplacent les « feuilles » (mots) de la langue source par des « feuilles » de la langue cible;
- soit l'approche par langue pivot ou *interlingua* qui se fait en deux étapes: de la langue source vers l'interlangue, puis de l'interlangue vers la langue cible. Elle présente théoriquement l'avantage d'inclure à moindre coût des langues supplémentaires (ainsi le Basic English était-il pour certains le candidat tout désigné pour remplir la fonction d'*interlingua*);
- soit enfin une approche qui repose elle aussi sur une étape de transfert. Le passage du texte source au texte cible présuppose une représentation désambiguïsée des deux. Cette procédure appelée triangle de Vauquois, en référence aux travaux de Vauquois et Boitet (1985) s'apparente au modèle d'analyse/synthèse développé en URSS (Piotrovski à Léninegrad; Apresjan et Melčuk à Moscou).

Le talon d'Achille de ces systèmes était évidemment la qualité finale de la traduction obtenue. Les utilisateurs, et plus encore les financiers, se sont lassés de l'insuffisance des résultats. L'automatisation totale censée assurer une qualité inégalée et un coût très avantageux se transformait en cauchemar puisqu'il était impossible d'améliorer la traduction sans revenir au texte source. Non seulement toute révision contraignait à une comparaison avec le texte source, mais la pré-édition du texte à traduire (sa simplification) se révélait être une tâche lourde, coûteuse et aléatoire. L'intervention humaine que le processus imposait s'avérait en outre complètement antinomique avec la démarche intellectuelle des traducteurs.

Dès 1959, le linguiste israélien Yoshua Bar-Hillel concluait qu'un haut niveau de qualité resterait toujours hors d'atteinte et qu'on devait s'en tenir à construire des outils fondés sur « une symbiose homme-machine ». Aux Etats-Unis, à partir de 1966, le rapport ALPAC (Automatic Language Processing Advisory Committee) mettra un coup de frein brutal au rêve de la machine à traduire en bloquant les financements¹⁴ et en suspendant les recherches.

On aurait pu s'attendre à ce que ce revers entraîne un retour critique sur les fondements théoriques. Il n'en a rien été. Le cas de la traduction automatique est en ce sens très instructif

13 Slodzian, M., 2007, « Rationalisation des langues et terminologie: d'Odgen à Catford », *Hermès*, 49.

14 Entre 1978 et 1993, les USA auront dépensé 20 millions de dollars dans la recherche en T.A., l'Europe 70 millions et le Japon 200 millions.

sur les rapports entre science et technologie. Le paradigme logiciste a survécu aux déboires des tentatives industrielles de la traduction automatique et s'est réincarné dans des aventures parallèles comme en témoignent les gigantesques programmes d'ontologies multilingues. Les échecs avérés semblent tirer bien peu à conséquence au plan théorique.

Il n'en va pas de même au plan pratique où la technologie opère des mutations sans le moindre souci de cohérence théorique. L'émergence de la traduction assistée (TAO) illustre cette capacité à rebondir.

Le scepticisme de Bar-Hillel a-t-il été entendu? Sans doute, même si l'idée de « symbiose homme-machine » a été souvent interprétée unilatéralement, en donnant la priorité à la machine. Les opérations de pré-édition et de post-édition des textes confiées aux soins des traducteurs ont fait de leur part l'objet de vives critiques et parfois de rejets: dans la « traduction assistée », en dépit de gains de temps, le traducteur est en fait requis pour assister la machine. Les performances des outils à base de « mémoire » dépendent en effet des opérations suivantes:

- textes contrôlés à l'entrée (pré-édition requise);
- terminologies standardisées;
- révision à la sortie (post-édition).

Comme on voit, l'amélioration du résultat continue de dépendre de la qualité de la révision humaine et de la complétude toujours imparfaite des dictionnaires associés.

L'histoire de l'industrie de la traduction automatique depuis les années 60 met en évidence une mutation silencieuse des systèmes à base de règles vers des systèmes hybrides, associant techniques d'apprentissage (à base d'exemples) et méthodes statistiques. Or les présupposés n'en sont pas seulement techniques: « mémoire de traduction » signifie l'abandon de l'utopie de la supériorité de la machine à traduire sur le traducteur humain au profit de la réintroduction de la traduction humaine dans le processus (chaque nouveau segment traduit vient enrichir la mémoire, augmentant ainsi le taux de pré-traduction disponible pour le prochain document similaire).

La technologie consistant à enregistrer dans une base de données des traductions manuelles définies par des caractéristiques pragmatiques (destinataire, domaine, date, etc.), afin de réutiliser partie ou tout de ces textes ultérieurement, constitue bel et bien une rupture avec les illusions de la traduction automatique des débuts. Mais, pour autant, la traduction « assistée » est loin de convaincre tout le monde.

Au terme de ces dix dernières années, les évaluations font davantage état d'améliorations techniques qui s'expriment en termes d'adaptabilité des systèmes, de simplicité d'utilisation, d'accroissement de ressources lexicales etc., que de progrès qualitatifs.

La contextualisation de l'opération de traduction introduite par la technologie des « mémoires de traduction », associée ou non à des modules de traduction automatique, n'a donc apporté que des solutions restreintes et fragiles (valables pour des domaines techniques) qui n'ont pas permis à l'industrie de la traduction de consolider suffisamment ses acquis, si l'on en juge par le nombre de faillites retentissantes ou de reventes (Lernout et Hauspie, Logos Corporation, Winger en Suède...). Quelles que soient les avancées et les apports réels de tel ou tel logiciel, la qualité finale des traductions demeure problématique.

Le rapport Hutchins (2003)¹⁵ énumère les obstacles. En premier lieu viennent les nombreux problèmes de traitement du langage naturel non résolus, quels que soient les avancées du domaine et les progrès des systèmes informatiques. « Nombre de ces problèmes sont inhérents au langage lui-même et sont dus aux différences culturelles. Ils existeront toujours et il est peu probable qu'une machine sera un jour capable de les surmonter », reconnaît l'auteur.

Un tel constat -problèmes inhérent au langage lui-même et différences culturelles- est désespérément répétitif. Et les améliorations préconisées ne le sont pas moins: nécessité de contrôler le langage technique, de normaliser les lexiques et d'expurger les spécificités culturelles propres aux textes et aux pratiques qu'ils décrivent.

Or, répétons-le, il y a bien eu changement de paradigme en ce que la notion de « mémoire de traduction » et de techniques d'apprentissage impliquent des textes ou des fragments de textes répondant à des normes culturelles précises et non plus à des mots-signes désincarnés. Il est beaucoup question de « contextualisation des données », de prise en compte du récepteur... Ces observations ont mis en avant la dimension communicationnelle qui fera les beaux jours du fonctionnalisme, mais dont la portée théorique et pratique demeure superficielle. Le cœur de l'affaire, c'est la textualité. Les textes sont par eux-mêmes des systèmes d'organisation des connaissances et les interrelations entre les mots, le mode de tissage textuel, sont réglés, non de l'extérieur, mais par les normes textuelles intrinsèques.

On saisit bien ici à quelle sorte d'obstacle se heurte le monde de la traduction automatique, version forte et version faible confondues, enfermé dans une sémantique positiviste tributaire d'une linguistique du signe réductrice.

L'avènement des systèmes statistiques

Aux yeux des spécialistes, l'introduction de solutions plus opératoires et surtout moins coûteuses, censées mieux répondre aux besoins concrets des traducteurs, n'actait nullement l'échec de la traduction automatique. La traduction automatique et la TAO ne jouaient pas dans la même cour: l'une était l'avenir, l'autre une sorte d'ersatz. Mais la raison pragmatique l'emportant, les deux approches ont été progressivement intégrées dans la gamme des produits proposés par les firmes. Systran, leader mondial des solutions de traduction automatique, ouvre la voie. Soucieuse de regagner du crédit auprès des institutions clientes - telle la Communauté européenne, l'OTAN ou le gouvernement américain -, la société Systran a conçu des solutions hybrides qui, à côté de son moteur de traduction automatique, offrent la possibilité d'enregistrer les traductions au fur et à mesure et d'alimenter parallèlement sa base terminologique. De son côté SDL et ses outils Trados, initialement spécialisée dans les mémoires de traduction et les logiciels de gestion de terminologie, se dotent d'une fonctionnalité T.A. Le processus d'hybridation s'est imposé sur le marché et s'étend aujourd'hui aux modèles statistiques qui, à leur tour, intègrent des modules de « mémoire de traduction » et modules à règles.

La traduction automatique statistique (TAS) s'est pourtant imposée contre les systèmes à base de règles et a proclamé se passer de connaissances linguistiques dans le traitement des deux langues. Elle suppose des modèles statistiques complexes évaluant le poids des mots ou des

15 Hutchins, W.J., 2003. « Machine translation: General overview », in Mitkov, Ruslan, ed., *The Oxford Handbook of Computational Linguistics*, Oxford University Press.

phrases pour obtenir le meilleur appariement entre segments de textes source et cible. Elle suppose de grandes quantités de corpus alignés et comparables qui permettent l'apprentissage automatique d'alignements bilingues.

Stimulés par la problématique de la traduction sur le Web, les systèmes TAS ont produit à leur arrivée, début des années 2000, une onde de choc qui s'exprime, entre autres, par le litige qui oppose Systran à la Commission européenne depuis 2004. Derrière la question juridique portant sur le droit à la propriété d'un code source, s'exprime une crise relative à l'efficacité et à la maintenance de systèmes de traduction dont Systran a longtemps eu l'exclusivité. Jusqu'en 2007, en effet, Google, AOL, Yahoo, AltaVista et Babelfish, par exemple, utilisaient le système Systran. C'est à cette date que Google introduit son propre système à base statistique. Le principe du système repose sur la construction d'un corpus de textes parallèles de plus d'un million de mots et de deux corpus unilingues de plus d'un milliard de mots chacun. Les modèles statistiques construits à partir de ces données servent ensuite à traduire d'autres langues.

Convaincant pour une utilisation en ligne, le TAS a des résultats très inégaux et souvent faibles s'agissant de textes entiers. Il nécessite de très grandes quantités de corpus alignés et comparables, ce qui met hors jeu les langues peu dotées. Il n'atteint pas les mêmes performances pour tous les couples de langues. Ainsi, pour la paire anglais -allemand, Babelfish, qui garde son système à base de règles, ferait mieux. Les méthodes d'évaluation (la mesure Bleu ou Senseval) des résultats de la traduction automatique sont si opaques que nous livrons cette information avec précaution.

A nos yeux, la question majeure touche le rapport des concepteurs de TAS à la linguistique de corpus. En effet, pour construire des corpus comparables de qualité, il faut à minima maîtriser les concepts centraux de la textualité que sont les discours et les genres. Donc, les bases mêmes d'une linguistique textuelle. A l'inverse, il semble que pour palier les faiblesses du système, les ingénieurs se tournent vers des modèles de désambiguïsation des mots et on ajoute des modules à règle et des modules « mémoire de traduction ». Encore une fois, l'imagination ne franchit pas les frontières de la linguistique du signe.

De la linguistique de corpus à la textométrie

Ainsi les corpus occupent-ils une place centrale dans des applications technologiques novatrices, sans que dialoguent les deux plans, théorique (à partir des travaux pionniers de Biber aux USA) et technologique avec les promoteurs de la traduction automatique à base statistique.

Ces derniers se sont sans doute inspirés directement des travaux qui ont abouti au Brown corpus, compilation de 1 million de mots au début des années 60. Dans le Brown Corpus, les textes sont caractérisés par des traits hétérogènes: y figurent 500 échantillons, distribués en 15 « genres » qui mélangent domaine, registre, style et d'autres critères censés relever du genre. Les applications visées sont unilingues et essentiellement lexicographiques, d'où sans doute cette définition laxiste du « genre ». Construire la signification des mots pour le dictionnaire en élicitant leurs sens principaux n'est pas traduire un texte.

Une autre étape est franchie lorsque Biber¹⁶ propose une typologie qui caractérise les textes par un ensemble de dimensions organisant des traits linguistiques. L'objectif est de construire

16 Biber, D., 1988, *Variation across Speech and Writing*, Cambridge University Press.

des modèles statistiques pour la description de la variation linguistique en anglais/américain écrit et parlé (conversation, textes académiques, etc.). L'analyse de la variation étant strictement fondée sur les traits linguistiques observés, ces travaux constituent une avancée considérable pour la linguistique descriptive, comme en témoigne le *Longman Grammar of Spoken and Written English*¹⁷.

Ces travaux pionniers qui ouvrent de formidables perspectives en problématisant le couple langue/texte échouent cependant à articuler les études quantitatives tournées vers les traits grammaticaux et lexicaux avec des analyses qualitatives visant les paliers supérieurs de l'organisation textuelle et son interprétation. Biber est le premier à le constater¹⁸. En se référant aux études sur le genre textuel produits par la socio-linguistique anglo-saxonne dont Swales¹⁹ est un éminent représentant, il conclut que l'analyse de patterns discursifs dans des textes isolés ne sont pas reproductibles à large échelle. Il met en doute « la possibilité d'identifier les limites les frontières des segments de discours et de coder le type fonctionnel de chacun des segments par une analyse humaine ».

En d'autres termes, l'enjeu est pour lui de réussir à combiner linguistique de corpus et analyse de discours. Il précise ainsi sa pensée: « dans les études de corpus l'unité d'analyse a été jusqu'ici le texte, -un livre entier, un article de recherche, de presse, etc. -, *alors qu'il y a souvent de grandes variations à l'intérieur des textes, dues aux changements internes de la tâche, de la finalité et du thème communicationnels*²⁰. » Le paradigme sémiotique implicite qui s'exprime ici relève de la pragmatique, troisième volet de la partition élaborée par Morris et Carnap (syntaxe, sémantique, pragmatique) et s'inscrit dans le cadre des théories communicationnelles auxquelles appartiennent les différentes écoles d'analyse du discours.

Les théories communicationnelles adhèrent à un dualisme sémiotique qui sépare expression et contenu. Les différences de registre que Biber veut « capturer » dans les textes sont à ses yeux motivées par des finalités communicationnelles.

Cette approche fonctionnaliste qui cherche à établir une typologie externe des textes (voir les classifications de Swales) passe à côté de la « sémiosis » qui s'accomplit au niveau des textes (corrélation entre plan du signifié et plan du signifiant) et dont le genre est le facteur principal. Le genre qui, à son tour, présuppose un discours s'inscrivant dans des normes linguistiques et culturelles.

Impuissante à concevoir la comparabilité des textes hors de la sémiotique dualiste du positivisme logique, la linguistique de corpus anglo-saxonne échoue à décrire la complexité textuelle.

En revanche, la textométrie, telle qu'elle est conçue et pratiquée par l'équipe autour d'André Salem, ouvre pour la traductologie des perspectives réellement fécondes. La mise en contraste de textes traduits (co-traductions) permet entre autres de poser la question de

17 Biber and al., 1999.

18 Biber, D., Jones, J., K., 2005, *Merging corpus linguistic and discourse analytic research goals: Discourse units in Biology research articles*, in *Corpus Linguistics and Linguistic Theory 1-2*, De Gruyter ed.

19 Swales, J., 1990, *Genre Analysis: English in Academic and Research Settings*, Cambridge University Press

20 Souligné par l'auteur

l'intertextualité dans sa dimension diachronique²¹. Elle échappe surtout à l'écueil du sens comme « ce qui reste d'invariant dans le transcodage »²² en prenant en charge toutes sortes de variables comme la fréquence et la distribution lexicale, la segmentation des paragraphes, la ponctuation, etc., autrement dit l'organisation textuelle. Les comparaisons textométriques de traductions qui ont été menées sur des textes français traduits en russe, chinois et coréen, par exemple, mettent en lumière la comparabilité des textes au delà des langues.

A elle seule, la textométrie ne peut naturellement résoudre les questions herméneutiques posées par les textes dans le parcours traduisant. Une linguistique interprétative est requise pour poser les bonnes hypothèses, sélectionner les variables à privilégier et les interpréter. Elle est garante d'une approche adéquate du principe d'équivalence qui est au coeur du processus de traduction. Rappelons qu'on ne recherche pas l'identité, mais l'équivalence en vertu du principe sémiotique fondamental selon lequel le sens est fait de différences. Tout « change » de l'expression s'accompagne en effet d'une modification du contenu. « Il faut déterminer le modulo de l'équivalence, caractérisé notamment par les genres d'arrivée et de départ et par la pratique en cours »²³

Conclusion

L'intérêt de la linguistique de corpus n'a pas échappé aux auteurs des Translation Studies, dont les travaux ont connu une grande notoriété à partir des années 80²⁴. Très tôt, ils ont eu l'intuition que la linguistique de corpus pouvait beaucoup apporter à la traductologie. C'est le cas de Mona Baker qui a tenté d'exploiter les corpus parallèles et comparables pour explorer « les régularités du comportement traduisant à partir de données empiriques »²⁵. Toutefois, l'intérêt de ce courant inspiré de la théorie descriptive de Toury est limité par sa finalité philosophique, les études sur les corpus comparables étant justifiées par la recherche d'universaux. C'est la porte ouverte à toutes formes de simplification au nom de la loi de la « growing standardization » des cultures.

De fait, l'orientation ultra-fonctionnaliste illustrée en particulier par la théorie du *skopos*²⁶ qui définit la traduction comme « un acte déterminé par l'objectif communicationnel dans l'intérêt de la culture cible » enferme les Translation Studies dans les errements d'une linguistique du discours réduite à sa fonction communicationnelle. La traduction se confond alors dangereusement avec la localisation et les Translation Studies avec les Transfer Studies « qui cherchent à identifier ce qui est inaccessible et d'éliminer ce qui est inaccessible »²⁷. A rebours

21 Salem, A., et al, 2009, Lexico 3, Explorations textométriques, séries textuelles chronologiques.

22 Rastier, F., *Linguistique interprétative et fondements sémiotiques de la traduction*, id.

23 Rastier, F., id.

24 Venuti, L., 2000, *The Translation Studies Reader*, London-New York, Routledge.

25 Pym, A., 2008, « On Toury's laws of how translators translate » in *Beyond Descriptive Translation Studies*, ed Pym and al., Amsterdam-Philadelphia, John Benjamins.

26 Vermeer, H.,J., 1989, « Skopos and Commitment in Translation Theory », in *Redings in Translation Theory*, ed Chesterman, Helsinki, Oy Finn Lectura Ob.

27 Göpferich, S., 2004 « Translation Studies and Transfer Studies » in *Doubts and Directions in Translation Studies*, ed Gambier et al., Amsterdam-Philadelphia, John Benjamins.

de ce que devrait être un programme de recherche pour la traductologie puisque, au final, on aura tué le texte.