



Revue internationale
International Web Journal
www.sens-public.org

Le fantasme de la traduction automatique : esquisse d'un imaginaire frelaté

Geneviève Has (Université d'Ottawa)

Résumé : Cette contribution vise à explorer l'imaginaire qui entoure la traduction automatique (TA), notamment en ce qui concerne la recherche fondamentale et ses applications. Il s'agit de retracer les étapes importantes de l'histoire de la traduction automatique pour mettre en lumière certaines des attentes alimentées par l'imaginaire technologique et, en retour, montrer comment ces attentes ont teinté à la fois les directions qu'a emprunté la recherche, les applications qui en sont nées et l'accueil qu'on leur a réservées. Une fois ces tensions esquissées, nous proposons un recadrage de cet imaginaire, lequel pourrait s'incarner dans la figure du cyborg.

Mots-clés : Traduction automatique, Warren Weaver, cyborg, imaginaire technologique.

Abstract : This contribution aims to explore the representations of machine translation (MT), specifically within fundamental research and its applications. By retracing the milestones of machine translation history, we shed light on some of the expectations fueled by technological myths, and, in return, show how these expectations curbed fundamental research's orientations, the applications born from it as well as their reception. Once these tensions are laid out, we propose a reframing of these conceptions, reframing that could be informed by the model of the cyborg.

Keywords : Machine translation, Warren Weaver, cyborg, technological myths.

Le fantasme de la traduction automatique : esquisse d'un imaginaire frelaté

Geneviève Has

There is a broad continuum of ways in which man and machine can share the translation responsibility. It ranges from Bar-Hillel's FAHQT (Fully Automatic, High Quality Translation) to Kay's HTLGI (Human Translation Like God Intended).

(Merle D. Tenney, « Machine Translation, Machine-aided Translation, and Machine-imposed Translation », *Tools for the trade: Translating and the Computer 5*, 1985, p. 105)

En 1997, quand Alta Vista fournit son application gratuite de traduction automatique, elle la baptise *BabelFish*, en référence au poisson du même nom imaginé par Douglas Adams¹. L'étonnant symbiote jaune n'est que l'une des innombrables incarnations du mirage de la traduction automatique, que l'on conçoit à tout coup sans intervention humaine. Tant dans la littérature qu'au cinéma, les exemples pullulent : des puces « microsoft » de l'univers de Gibson à C3PO, l'aimable diplomate cybernétique, en passant par le traducteur universel de *Star Trek*, les auteurs de science-fiction rivalisent d'imagination pour expliquer comment un tel exploit pourrait devenir réalité, mais sans jamais vraiment en livrer le mécanisme.

Ceci dit, l'idée d'éliminer l'élément humain de la traduction, puisque subjectif, lent ou encore encombrant, ne se limite pas au domaine de l'imaginaire, et ne date pas d'hier. On recense, en Allemagne au XVIIe siècle, la première tentative de produire des traductions mécaniquement, à l'aide de tables de calcul. Johannes Becher, un moine de Speyer, avait conçu un métalangage mathématique décrivant le sens de nombreux mots dans plusieurs langues. Il s'agissait, en quelque sorte, d'une tentative de formalisation des langues naturelles et de leurs relations complexes, tentatives qui seront reprises plus tard avec quelque succès. Ainsi, une phrase dans n'importe quelle langue incluse dans le système de Becher pouvait être traduite mécaniquement en fonction de formules préétablies². Il faudra évidemment attendre que s'écoule la première moitié du XXe siècle, avec la naissance des supers calculateurs et autres cerveaux mécaniques,

¹ Pour la référence originale, voir : <http://www.bbc.co.uk/cult/hitchhikers/guide/babelfish.shtml>

² Karl Heinz Freigang, « Automation of Translation: Past, Presence, and Future », *Revista Tradumàtica* [en ligne], n° 0, octobre 2001, p.1, <<http://www.fti.uab.es/tradumatica/revista/num0/articles/khfreigang/imprimir.pdf>> (page consultée le 2 décembre 2011).

Le fantasme de la traduction automatique : esquisse d'un imaginaire frelaté

avant de voir de réelles percées (même si elles restent timides) dans le domaine du traitement automatique des langues, et particulièrement de la traduction automatique (TA).

Je propose donc de revenir sur les grandes étapes de la recherche en TA en vue de souligner quelques-uns des écarts entre réalité et fiction, entre possibilités et attentes, et de montrer ainsi que les maladroites et l'inadéquation que l'on reproche encore souvent aux systèmes de TA les plus sophistiqués n'ont d'autres sources que cet imaginaire grandiose, qui ne peut que décevoir une fois ramené à hauteur du réel.

L'âge d'or

Le milieu du XXe siècle, qui voit la naissance de la TA comme nous la connaissons aujourd'hui, est effervescent à bien des égards. D'une part, pour la première fois de l'histoire, les ordinateurs, ces « cerveaux électroniques », sont suffisamment puissants pour répondre aux rêves d'automatisation des scientifiques. D'autre part, la Guerre Froide engendre un besoin impérieux pour les Américains d'être au fait des activités des Russes et vice-versa. Ainsi, les années 50 et 60 se voient le théâtre d'une course effrénée à qui pourra le premier maîtriser à la fois la machine et le langage. Les prédictions des principaux intéressés font preuve, avec le recul, d'un optimisme bon enfant qui paraît friser la naïveté.

On s'accorde généralement pour attribuer la paternité de la recherche en traduction automatique à Warren Weaver, alors à l'emploi de la Rockefeller Foundation³. Dès 1947, Weaver se demande, dans sa correspondance⁴ avec Norbert Wiener, le premier cybernéticien, s'il est concevable que les ordinateurs puissent servir à la traduction de textes rédigés en langues naturelles (par opposition à des textes codés dans une « langue artificielle »). Wiener lui répond

³ Plusieurs références à l'appui, parmi de nombreuses autres.

Alan Melby, *The Possibility of Languages*, Amsterdam, Philadelphie, John Benjamins Publishing Company, 1995, p.17.

Kristin Demos et Mark Frauenfelder, « Machine Translation's Past and Future », *Wired* [en ligne], no 8.05, mai 2000, p. 1, <<http://www.wired.com/wired/archive/8.05/timeline.html>> (page consultée le 5 décembre 2011).

Muhammad Raji Zughoul et Awatef Miz'Il Abu-Alshaar, « English/Arabic/English Machine Translation: A Historical Perspective », *Meta*, vol. 50, n° 3, 2005, p. 1024.

Michael Rosner, « Machine Translation and Human Translators », Malte, *Sunday Times*, 25 avril 2005, p.1.

Karl Heinz Freigang, *loc. cit.*, p.1.

Celia Rico Pérez, « From Novelty to Ubiquity: Computers and Translation at the Close of the Industrial Age », *Translation Journal* [en ligne], vol. 5, n° 1, 2001, p.2, <<http://translationjournal.net/journal/15mt2.htm>> (page consultée le 28 novembre 2011).

John Hutchins, « Retrospect and Prospect in Computer-based Translation », *Proceedings of MT Summit VII "MT in the great translation era"*, Low Hwee BOON (dir.), Singapour, AAMT, 1999, p. 1.

Marie-Claude L'Homme, *Initiation à la traductique*, Montréal, Linguatex, 2008, p.12.

⁴ Cette correspondance est reprise dans le Mémoire de Weaver intitulé *Translation*. Une copie des textes originaux peut être consultée à l'adresse suivante : <http://www.mt-archive.info/50/Weaver-1947-typescript.pdf>.

par la négative⁵, mais sa réflexion se poursuit néanmoins, et en 1949, il rédige un mémo qui passera à l'histoire et jettera les bases de la recherche en TA, pour le meilleur et pour le pire. De cette première correspondance, on cite généralement le passage suivant :

I have a text in front of me which is written in Russian but I am going to pretend that it is really written in English and that it has been coded in some strange symbols. All I need to do is strip off the code in order to retrieve the information contained in the text.⁶

L'analyse acceptée généralement est que le mémorandum influencera la traduction automatique pour les décennies à venir, lui donnant une saveur résolument (et négativement) cryptographique, surtout en regard des approches subséquentes. J'ai aussi fait mienne cette analyse dans mes recherches précédentes⁷, jusqu'à ce qu'une lecture plus attentive me permette de remarquer que Weaver n'était pas aussi ignorant des difficultés intrinsèques au traitement automatique des langues (comme la traduction automatique) que l'on a bien voulu le rapporter. Mathématicien et conseiller scientifique, Weaver souligne d'emblée son ignorance des subtilités du jeune domaine de la linguistique computationnelle et de l'automatisation des langues naturelles. Il fait néanmoins preuve d'une vision remarquable, décrivant avec une justesse étonnante les trois générations de systèmes de TA⁸ qui nous sont aujourd'hui familières. Qui plus est, il esquisse

⁵ « [...] as to the problem of mechanical translation, lui répond-il, I frankly am afraid the boundaries of words in different languages are too vague and the emotional and international connotations are too extensive to make any quasi mechanical translation scheme very hopeful. I will admit that basic English seems to indicate that we can go further than we have generally done in the mechanization of speech. [...] At the present time, the mechanization of language, beyond such a stage as the design of photoelectric reading opportunities for the blind, seems very premature. » (Weaver, 1955, p. 19) Wiener, pourtant le mieux outillé pour attaquer le problème aux dires de Weaver, avait bien identifié l'écueil qui guetterait les chercheurs quelques années plus tard. Ainsi, nombre de spécialistes, incluant le « père de la TA » lui-même ont émis des doutes quant aux chances de succès de l'entreprise, mais leurs doléances n'ont pas eu l'effet modérateur escompté.

⁶ Warren Weaver, « Translation », *Machine Translation of Languages*, William S. Locke et Andrew. D. Booth (dir.), Cambridge (Mass.), The Technology Press of the MIT, 1955, p. 18.

⁷ Voir entre autres Geneviève Has, 2014.

⁸ Première génération : approche directe (il s'agit d'une approche mot-à-mot, utilisant principalement des dictionnaires de correspondance). Weaver la mentionne surtout comme l'approche privilégiée dans l'état « actuel » de la recherche. (Weaver, *op. cit.*, p. 18-20)

Deuxième génération : approches indirectes (les approches à base de règles et celles par interlangue sémantique, s'attaquant au problème de la TA en formalisant le processus à l'aide de représentations des langues de départ et d'arrivée, ou à l'aide d'une interlangue sémantique. Des modules séparés de transfert d'une langue à l'autre entrent alors en jeu pour compléter la traduction). Weaver expose longuement sa conception des universaux du langage et justifie ainsi la possibilité de concevoir un système de TA en formalisant complètement le sens de tout énoncé linguistique (*Ibid.*, p. 14-16). Ce type d'approche a fait l'objet de projets de recherches, mais n'a donné que peu de résultats. Weaver, bien conscient de la complexité des langues et des limitations techniques des ordinateurs, propose aussi une approche se basant sur une logique transformationnelle (*Ibid.*, p. 22-23), que l'on peut assimiler à l'approche par règles qui sera la méthode privilégiée des années 70 à 80.

Troisième génération : approches par corpus (il existe deux approches par corpus, celles à base d'exemples, qui relève du domaine de l'intelligence artificielle et les approches statistiques, surtout celle utilisant les *n-gramme*, qui permettent de résoudre le problème du sens à l'aide du contexte et produisent une traduction en fonction de sa probabilité). Weaver se demande, avec beaucoup de justesse quelle

aussi l'idée, qui sera reprise par de nombreux chercheurs par la suite, que les langues restreintes et les vocabulaires contrôlés pourraient constituer une piste non négligeable⁹. Enfin, il est important de souligner, à notre avis, que Weaver avait vu juste en se demandant si viser un système parfait ne relevait pas de l'utopie et en proposant plutôt de viser une performance acceptable pour un nombre tolérable de cas¹⁰. En somme, alors que la position de Weaver semble nuancée et que ce dernier recommande en toutes lettres une approche statistique¹¹, ses contemporains et les historiens n'ont retenu du mémorandum que les aspects cryptographiques et mathématiques au détriment du reste; les premiers, probablement en raison des limites de l'informatique de leur temps, les seconds, puisque c'est cette approche qui a tenu le devant de la scène jusque dans les années 70.

Ainsi, stimulés en grande partie par l'impulsion de Weaver, de nombreux chercheurs des années 50 s'emploient à décoder le langage, en vue de dégager une structure universelle qui s'appliquerait à toutes les langues. Rapidement, la première tentative réussie de traduction automatique¹² a été réalisée, le 7 janvier 1954, du russe vers l'anglais, à l'aide d'un dictionnaire de 250 mots et de six règles de transfert syntaxique, dans les laboratoires de l'Université Georgetown (en collaboration avec des chercheurs d'IBM)¹³.

La réaction est immédiate et euphorique : les titres de journaux laissent entendre que le processus est sans effort et que les ordinateurs sont d'ores et déjà prêts à prendre la relève : « Russian is turned into English by a fast electronic translator », peut-on lire en première page du New York Times le lendemain¹⁴. Le Christian Science Monitor, rapporte le 11 janvier que « The brain didn't even strain its superlative versatility and flickered out its interpretation with a nonchalant attitude of assumed intellectual achievement. »¹⁵

Cependant, la démonstration de Georgetown présentait quelques caractéristiques qui continuent jusqu'à aujourd'hui d'accabler les chercheurs. En effet, le grand public a rarement conscience que les conditions de laboratoire sont très strictes lors de ces démonstrations. Par

devrait être la valeur minimum de n en vue d'obtenir une traduction acceptable dans la majorité des cas. (*Ibid.*) Les systèmes de TA actuels s'appuient sur une méthodologie très semblable.

⁹ Warren Weaver, *Op. cit.*, p. 20.

¹⁰ *Ibid.*, page 24.

¹¹ « And it is one of the chief purposes of this memorandum to emphasize that statistical semantic studies should be undertaken, as a necessary preliminary step. » (*Ibid.*, p. 24)

¹² À ce sujet, voir la vidéo d'un entretien avec les chercheurs de l'équipe Georgetown-IBM, qui présentent leurs prédictions (http://www.youtube.com/watch?feature=player_embedded&v=K-HfpsHPmvw [TAUS, 2010]).

¹³ John Hutchins, « The Georgetown-IBM demonstration, 7th January 1954 », *MT News International*, n° 8, 1994, p. 15.

¹⁴ *Ibid.*, page 15.

¹⁵ *Ibid.*, page 16.

exemple, dans le cas de Georgetown-IBM, on a rigoureusement sélectionné 49 phrases à traduire, qui devaient respecter certains critères précis. En effet, toutes les phrases étaient déclaratives et simples et tous les verbes étaient à la troisième personne¹⁶. Par ailleurs, comme le remarque Melby, les systèmes de traduction automatique doivent toujours être ajustés, ce qu'il compare à l'auditeur qui manipulerait les boutons de sa radio pour améliorer la réception. De la même façon, explique-t-il, les systèmes de TA sont ajustés en fonction des phrases qu'on leur demande de traduire¹⁷. Il va sans dire que les scientifiques de Georgetown-IBM avaient consciencieusement entraîné leur système pour le grand jour, ce qui a certainement contribué à impressionner à la fois le public et les bailleurs de fonds, alors que la nature même du système rendait difficile la reproduction de ces exploits à grande échelle. Il n'en reste pas moins que, dès lors, les équipes de recherche en TA ont le vent dans les voiles et les subventions affluent¹⁸.

Rapidement, toutefois, l'approche directe laisse entrevoir quelques-unes de ses faiblesses. Tout d'abord, le nombre de règles nécessaires au traitement de textes généraux dépasse rapidement les capacités des linguistes et des ordinateurs qu'ils utilisent. L'ambiguïté, notamment, est un problème très criant pour l'approche directe. La traduction d'expressions imagées ou de métaphores donne souvent des résultats pour le moins cocasses¹⁹. Par ailleurs, cette approche produit des systèmes qui ne fonctionnent qu'avec une seule paire de langues, souvent dans une seule direction. On peut tout de suite constater le casse-tête potentiel lorsque l'on s'attaquera à un marché du calibre de celui de l'Union européenne puisque, comme le souligne plus tard Pierre Isabelle : « pour traduire entre n langues on a besoin de $n*(n-1)$ modules de règles (CE : $23*22 = 506!$) »²⁰ ²¹. Tous ces désagréments, cependant, n'ont été pleinement constatés et reconnus que bien des années plus tard, et les systèmes à approche directe ont tout de même régné sans compétition pendant près de trois décennies.

¹⁶ *Ibid.*, page 17.

¹⁷ Alan Melby, *op. cit.*, page 20.

¹⁸ *Ibid.*, page 19.

¹⁹ Peter J. Wheeler, dans *Machine Translation Today*, relate l'anecdote suivante : « if it [Systran] did not produce the apocryphal sentence about steak and vodka, [it] certainly did translate *La Cour de justice envisage la création d'un cinquième poste d'avocat général* as "the yard of justice is considering the creation of a fifth general avocado station" » (Peter Wheeler, « Systran », *Machine Translation Today: The State of the Art*, Margaret King (dir.), Édimbourg, Edinburgh University Press, 1987, p. 192.) Pour les détails concernant « the apocryphal sentence about steak and vodka », ou les mythes tenaces concernant la TA, voir Hutchins, 1995, p. 17-18.

²⁰ Pierre Isabelle, *Introduction à la traduction automatique par règles* [en ligne], 2010, p. 9, <<http://linguistech.ca/display657>> (page consultée le 20 juin 2013).

²¹ Il n'en reste pas moins que l'un des systèmes commerciaux les plus anciens, SYSTRAN, a été conçu en approche directe (quoique depuis ses débuts, des modifications y ont été apportées et que le SYSTRAN se trouve maintenant sous la catégorie des systèmes hybrides). Les systèmes directs sont maintenant utilisés pour des paires de langues qui ont une structure et un vocabulaire grandement similaires, de façon à ce que les concepteurs puissent tirer profit de cette similarité et concentrer leurs efforts sur les différences entre les langues, sans que ces dernières ne soient démesurées.

Un coup dur

Dans la foulée du succès de Georgetown-IBM, on croyait bien avoir trouvé la clé permettant de réduire en poussière la tour de Babel et de voir l'avènement d'un monde où la traduction n'aurait plus besoin d'intervention humaine. Les réjouissances allaient être de courte durée. En effet, les efforts de toute une génération de chercheurs allaient se voir presque anéantis par le rapport de l'ALPAC (Automatic Language Processing Advisory Committee) en 1966. Ce rapport, aussi connu sous le nom de « Black Book on Machine Translation », allègue, en somme, que la TA est inefficace, inadéquate et trop coûteuse²². L'impact, que l'on peut expliquer par la grande différence entre les attentes des bailleurs de fonds et la réalité de la recherche, fait trembler le monde de la traduction automatique et sonne le glas de nombreuses initiatives dans le domaine. Dès lors, les subventions de recherche et de développement pour la TA se voient réduites presque à néant, particulièrement en Amérique du Nord.

Ce que les historiens relèvent rarement, c'est que le *Black Book* ne condamne pas d'emblée toute recherche sur l'automatisation de la traduction, mais bien les efforts qui visent à atteindre la traduction entièrement automatique de haute qualité (TAEHQ) sur texte général sans intervention humaine. Ainsi, si peu d'équipes visent cette dernière (il se trouve tout de même quelques irréductibles, particulièrement hors des États-Unis), on commence à explorer partout dans le monde de nouvelles voies comme les systèmes à langue restreinte, ou alors des systèmes ne fonctionnant qu'à l'intérieur de domaines de spécialité.

The Quiet Decade

En fait, si les militaires américains abandonnent pour le moment la recherche en TA, les gouvernements européens, canadien et japonais, eux, sont toujours de la partie. Hutchins²³ et Somers²⁴ s'accordent pour dire qu'en raison de facteurs sociopolitiques et culturels, la demande en traduction dans ces pays n'a jamais diminué. En Europe, les échanges multilingues entre les différents pays motivent des entreprises adaptées à cette réalité. La Commission européenne rachète SYSTRAN, un système américain de deuxième génération et en poursuit le développement²⁵. D'autres projets voient le jour en France, en Allemagne et en Italie. Au Canada, c'est le bilinguisme législatif qui motive la recherche. Au Japon, ce sont les avancées informatiques réalisées par les chercheurs en vue de manipuler le système d'écriture qui aiguillonne les

²² ALPAC, *Language and Machines: Computers in Translation and Linguistic*, Washington, D.C., National Academy of Science and National Research Council, 1966.

²³ John Hutchins, « Machine Translation over Fifty Years », *Histoire Épistémologie Langage*, tome 23, fascicule 1, 2001, p. 7-31.

²⁴ Harold Somers, *Computers and Translation: A Translator's Guide*, John Benjamin, Amsterdam, 2003.

²⁵ *Ibid.*, p. 5.

chercheurs : les succès sont si probants que les universités et le secteur privé décident d'appliquer leurs découvertes à la traduction automatique Anglais-Japonais²⁶. De l'approche cryptographique mot-à-mot des premières années, on passe à une approche linguistique qui est caractérisée par la formalisation des langues naturelles, soit via une interlangue, soit à l'aide de grammaires formelles et de métalangages informatiques. Une passation des pouvoirs, en somme, des cryptographes et mathématiciens vers les spécialistes de la linguistique computationnelle.

Cependant, ces approches de deuxième génération ont aussi leurs faiblesses. Tout d'abord, il faut supposer que le texte source est grammaticalement correct. Pour des essais contrôlés en laboratoire, la question est peu pertinente, mais si le système vise la commercialisation et sera utilisé pour des textes généraux, il est impossible de garantir un résultat à tout coup. En outre, les modules d'analyse ont du mal à traiter des ambiguïtés grammaticales qui ne ralentiraient pas un traducteur humain. L'exemple par excellence²⁷ pour illustrer les faiblesses de cette génération est « Time flies like an arrow », dont la structure pour une machine est très difficile à analyser hors contexte et pourrait ainsi être traduit par quelque chose comme « Les mouches du temps aiment la flèche ».

Il nous faut rappeler que les équipes de recherche en TA se sont tournées vers les spécialistes de la linguistique computationnelle pour résoudre les problèmes de l'automatisation de la traduction et ont délaissé au même rythme les cryptographes qui avaient fait leur succès. De plus, à la fin des années 70 et au début des années 80, les progrès en informatique permettent aux chercheurs de concevoir des systèmes de plus en plus complexes et de traiter des volumes de données sans commune mesure avec les précédents.

Ainsi, la « Quiet Decade » est dominée par les approches par interlangue et par les formalismes linguistiques. On expérimente aussi avec les langues restreintes, les vocabulaires contrôlés et les sous-domaines. Les systèmes gagnent en complexité et deviennent modulaires. La recherche en TA est principalement influencée par les domaines de la linguistique structurale et de l'informatique. Par ailleurs, de plus en plus de chercheurs ont abandonné l'idée de la traduction complètement automatique : l'interactivité, avec en bonne place la postédition, fait son apparition comme partie intégrante des processus de traduction automatique.

Au cours des années 80, c'est l'approche indirecte par transfert qui règne sans conteste sur le paysage de la TA²⁸. En effet, les approches par interlangue ne semblent pas remporter les succès

²⁶ *Ibid.*, p. 5-6.

²⁷ Pierre Isabelle, *op. cit.*, p. 14.

²⁸ John Hutchins « Latest Developments in Machine Translation Technology: Beginning a New Era in MT Research », *MT Summit IV: International Cooperation for Global Communication. Proceedings, July 20-22, 1993*, Kobe, Japon, 1993, p. 11.

attendus, et l'on préfère en conséquence plancher sur des approches moins ambitieuses²⁹, mais plus réalistes. Les universités et les organismes gouvernementaux sont toujours au premier plan de la R et D à travers le monde, mais le rôle du secteur privé commence à prendre de l'ampleur que l'on n'avait pas vue jusqu'alors.

En d'autres mots, force est de constater que la « Quiet Decade », même si elle laissera sa marque dans l'imaginaire de la recherche en TA et dans celle des principaux commanditaires, notamment en rendant presque taboue l'appellation « traduction automatique »³⁰, n'aura pas été stérile sur le plan de la recherche, mais aura plutôt été un passage obligé pour en arriver aux systèmes dont nous profitons aujourd'hui.

Une révolution statistique

Le domaine de l'informatique subit aussi des transformations importantes qui auront des répercussions sur la TA et la TAO. En effet, la fin des années 70 voit la naissance des premiers ordinateurs personnels et au cours des années 80, ces derniers se répandent à grande échelle dans les milieux professionnels. Les avancées en micro-informatique transforment en profondeur le monde de la traduction : la standardisation des plateformes, le développement des interfaces-utilisateurs et l'apparition des périphériques de stockage de données plus performants sont autant de facteurs qui ont facilité l'adoption des outils de traduction assistée par ordinateur (TAO) chez les traducteurs³¹, outils qui, comme leur nom l'indique, ne permettent pas de traduire de façon automatique, mais qui assistent les traducteurs dans leur travail.

Ainsi, l'automatisation grandissante du monde du travail (phénomène qui ne se limite pas à la traduction, bien sûr) fait naître de nouveaux besoins. En conséquence, observe Hutchins, « MT was coming out of the laboratory onto the marketplace and into the office »³². Au cours de cette décennie, explique-t-il, les faiblesses des systèmes de TA sont obligamment reconnues et l'on propose ces derniers, non plus comme un moyen de remplacer les traducteurs, mais bien comme des aides à la traduction³³. Il n'est donc pas surprenant de constater que nombre de chercheurs s'emploient à développer la TAO et que nombre d'institutions et de grandes organisations qui œuvrent en traduction cherchent à l'implanter. Qui plus est, les premiers systèmes commerciaux

²⁹ *Id.*, « Machine Translation over Fifty Years », *Histoire Épistémologie Langage*, tome 23, fascicule 1, 2001, p. 17.

³⁰ Marie-Claude L'Homme, *op. cit.*, p. 14.

³¹ John Hutchins, « The Origins of the Translator's Workstation », *Machine Translation*, vol. 13, n° 4, 1998, p. 15.

³² *Id.*, « Latest Developments in Machine Translation Technology: Beginning a New Era in MT Research », *MT Summit IV: International Cooperation for Global Communication. Proceedings, July 20-22, 1993, Kobe, Japon, 1993*, p. 11.

³³ *Ibid.*

font leur apparition sur le marché. ALPS et Weidner sont les fers de lance du mouvement, mais plutôt que de les promouvoir comme de la TA, on les commercialise comme des aides à la traduction puisqu'ils ont besoin de beaucoup d'intervention humaine pour produire des traductions acceptables³⁴. Ces conditions sont essentielles au changement de paradigme qui nous mènera à la TA contemporaine.

Le prochain grand virage de la traduction automatique est amorcé à la fin des années 80, sur les chapeaux de roues, par Fred Jelinek, chef du groupe de recherche d'IBM sur la reconnaissance vocale, à l'occasion de la Fourth International Conference on Theoretical and Methodological Issues in Machine Translation : « Each time I fire a linguist, my performance goes up »³⁵. La traduction automatique statistique (TAS) est une méthode empirique qui s'affranchit complètement des modèles linguistiques rigides et s'appuie sur un ensemble de données organisées, le corpus bilingue ou multilingue aligné. Il s'agit d'aligner des phrases, des groupes de mots et des mots individuels de deux textes parallèles pour ensuite calculer la probabilité que n'importe quel des mots dans une phrase donnée en langue source corresponde avec un ou des mots qui font partie de la phrase qui lui est liée dans le texte en langue cible. L'hypothèse qui sous-tend la TAS est la suivante : une phrase (S) de langue source peut avoir un grand nombre de traductions (T) et chacune de ces traductions a une probabilité plus ou moins grande d'être adéquate; en théorie, il n'y a pas de traduction « correcte » ou « incorrecte »³⁶. C'est l'approche qu'utilisent aujourd'hui *Google* et *Microsoft*, notamment, avec modifications à l'avenant, bien sûr, et avec le succès que l'on connaît.

En d'autres mots, le concept de *n-grammes* est la solution de Jelinek au problème du modèle de langue, solution qui lui permet en fait d'éviter la formalisation linguistique³⁷. Les chercheurs en TAS l'ont plus tard adapté à leurs besoins, car Jelinek s'était plutôt concentré sur les trigrammes (séquence de trois mots)³⁸. Néanmoins, le principe reste inchangé. Il s'agit simplement de segmenter automatiquement le corpus unilingue de langue cible aligné en séquences de *n-mots*. Ces séquences sont ensuite compilées et après analyse statistique, le système attribue à chacune d'elles une probabilité basée sur sa fréquence d'occurrence. Les séquences, ainsi analysées et associées à leur « score » de probabilité forment le modèle de langue, qui fournira la sortie en

³⁴ *Id.*, « Out of the Shadows: a Retrospect of Machine Translation in the Eighties », *Terminologie et Traduction*, n° 3, 1990, p. 277-278.

³⁵ Tony Whitecomb, « Statistical methods gaining ground », *Language Industry Monitor*, n° 11, Sept-Oct 1992, p. 1.

³⁶ Chiew Kin Quah, *Translation and Technology*, New York, Palgrave Macmillan, 2006, p.78.

Roland Kuhn, *Introduction à la traduction automatique statistique (TAS)* [en ligne], 2010, p.15 <<http://linguistech.ca/display658>> (page consultée le 25 juillet 2013).

³⁷ *Ibid.*, p.11.

³⁸ *Ibid.*

langue cible³⁹. Le corpus bilingue aligné subira le même sort, mais les séquences seront enregistrées dans la table des segments du système de TAS en fonction de la probabilité que le segment source soit traduit par un ou des segments cibles donnés. Cette table est gage de la fidélité de la traduction⁴⁰.

La TAS offre de nombreux avantages, mais a aussi certains inconvénients. D'une part, il est facile d'entraîner les systèmes : il suffit d'ajouter des textes à leur corpus pour les renforcer. Il en découle aussi qu'ils sont facilement adaptables à un domaine en particulier : plus on fournit des textes précis et spécialisés au corpus (qui présentent donc moins d'ambiguïtés), mieux le système apprendra le nouveau domaine et pourra fournir de bonnes traductions. Ce type d'apprentissage a aussi pour conséquence qu'il est relativement peu coûteux (notamment en ce qui concerne les ressources humaines) d'entraîner et de modifier ces systèmes. Par ailleurs, la TAS traite admirablement les expressions idiomatiques et offre des traductions beaucoup plus agréables au lecteur que ce que pouvaient offrir certains systèmes à base de règles. Cependant, les systèmes de TAS ont des inconvénients à l'avenant. Par exemple, ils n'ont aucune connaissance sémantique ni pragmatique et les registres de langue restent un mystère pour eux. Plus grave, puisqu'ils n'ont recours à aucune grammaire, ils commettent parfois des fautes d'accord qu'un système à base de règles n'aurait jamais perpétrées⁴¹.

Le tableau suivant, dont les données ont été tirées de Kuhn (2010), illustre les forces et les faiblesses de la TAS, avec des traductions tirées de Google Translate. Nous y avons ajouté la traduction qu'offrait récemment Google Translate en comparaison.

Anglais	Traduction GT (09/2010)	Traduction GT (07/2013)
That's another kettle of fish	C'est une autre paire de manches	C'est une autre paire de manches
The girl is attractive	La jeune fille est jolie	La jeune fille est attrayant.
The girl who arrived last Wednesday is attractive.	La jeune fille qui est arrivé mercredi dernier est attrayant.	La jeune fille qui est arrivé mercredi dernier est attrayant.

Tableau comparatif de traductions fournies par Google Translate à trois ans d'intervalle (données 2010⁴²)

³⁹ *Ibid.*

⁴⁰ *Ibid.*

⁴¹ *Ibid.*, p. 25.

⁴² *Ibid.*, p. 26.

On remarque tout de suite le changement qui a eu lieu dans les résultats obtenus. Le premier exemple reste inchangé et confirme la remarque de Kuhn sur la maîtrise des expressions idiomatiques par les systèmes de TAS. Le dernier exemple reste lui aussi inchangé et illustre bien les catégories d'erreur d'accord typiques de la TAS qui surviennent lorsque l'adjectif qui qualifie un nom est placé trop loin dans la phrase pour que système puisse « comprendre » qu'il doit être accordé. La même remarque explique la traduction boiteuse d'« attractive » dans le troisième exemple. Cependant, la construction du deuxième exemple ne devrait pas poser problème à *Google Translate*, puisque le nom et l'adjectif qu'il qualifie sont très proches l'un de l'autre. Kuhn indique qu'« un système TABR aurait probablement omis jeune et mis attrayante », mais que puisque la traduction de *Google Translate* est moins littérale, « La TAS a bien marché ici [sic] »⁴³. Que s'est-il passé en trois ans? La jeune fille est restée, mais elle devenue « attrayant », le système commettant une erreur qu'il n'avait pas faite au départ. Évidemment, il y a peu de chance qu'une telle chose se produise avec un système à base de règles, puisque les ressources que requiert un changement dans ce type de système sont très élevées. Ce désagrément illustre bien le prix à payer si l'on veut profiter des avantages qu'offre un système malléable : les modifications peuvent aller trop loin. Bien sûr l'exemple présenté ici est extrême, puisque *Google Translate* s'appuie en partie sur les modifications proposées par ses utilisateurs, mais la situation peut facilement se reproduire dans un système implanté en entreprise si des mesures ne sont pas prises pour assurer une certaine qualité au corpus d'entraînement. Il n'en reste pas moins, que malgré ces accroc, la TAS reste une approche fiable dont l'efficacité n'est plus à prouver.

Un malentendu qui perdure

Qu'en est-il du traducteur, dans tout cela? Depuis les années 50, nous avons vu défiler cryptographes, mathématiciens, linguistes, informaticiens, statisticiens, mais pas de traducteurs. Puisqu'il s'agissait d'ailleurs au départ de les éliminer, il ne faut donc pas se surprendre de constater que la majorité se soit tenue loin de la recherche en traduction automatique. Dans « *The Proper Place of Men and Machines in Language Translation* », Martin Kay montre bien le calvaire que les traducteurs, et leurs textes, doivent subir aux mains des machines à traduire :

There was a long period—for all I know, it is not yet over—in which the following comedy was acted out nightly in the bowels of an American government office with the aim of rendering foreign texts into English. Passages of innocent prose on which it was desired effect this delicate and complex operation were subjected to a process of vivisection at the hands of an uncomprehending electronic monster that transformed them into stammering streams of verbal wreckage. These were then placed into only slightly more gentle hands for repair. But the damage had been done. Simple tools that would have done so much to make the repair work easier and more effective were not to be had

⁴³ *Ibid.*, p. 26.

presumably because of the voracious appetite of the monster, which left no resources for anything else. In fact, such remedies as could be brought to the tortured remains of these texts were administered with colored pencils on paper and the final copy was produced by the action of human fingers on the keys of a typewriter. In short, one step was singled out of a fairly long and complex process at which to perpetrate automation. The step chosen was by far the least well understood and quite obviously the least apt for this kind of treatment.⁴⁴

Il faut malheureusement avouer que la situation a perduré jusqu'à tout récemment. Pour illustrer ce que j'entends par là, je vous propose deux exemples relativement récents dont j'ai été personnellement témoin.

Au cours du congrès de l'OTTIAQ en 2008, où Pierre Isabelle, pionnier de la recherche en traduction automatique au Canada, venait présenter les résultats d'un banc d'essai du tout dernier système de traduction automatique qui s'est tenu au Bureau de la traduction (BT). Portage, qui depuis a gagné nombre de compétitions internationales, avait très bien performé chez les traducteurs du Bureau de la traduction, lesquels ont une réputation de puristes à peu près inégalée. Ces derniers se voyaient agréablement surpris de l'idiomaticité des traductions produites par le système, vantant son efficacité et sa facilité d'utilisation et certains ont même demandé à continuer à utiliser la bête. En dépit de tout cela, le pauvre conférencier a dû essayer les attaques qui fusaient de toutes parts lors de la période de questions. Pourtant, les traducteurs présents dans la salle, en 2008, je vous le rappelle, pour la plus grande partie des pigistes, ne pouvaient évidemment prétendre dédaigner l'ordinateur, Internet et autres outils informatiques. Notre erreur à tous, traducteurs, informaticiens, donneurs d'ouvrage, confondus, est d'avoir espéré (ou craint) ne serait-ce qu'un instant, que la machine remplacerait l'humain dans la délicate (et mystérieuse) opération qu'est la traduction. Des représentations fictionnelles, comme le *BabelFish*, que j'ai mentionné en introduction, ainsi que des promesses telles que celles rapportées par les journalistes à l'occasion de la démonstration de Georgetown-IBM contribuent à perpétuer cette méprise.

Quelques années plus tard, en 2011, j'ai recueilli les commentaires de quelques traducteurs du BT sur la question des technologies en traduction dans le cadre de ma scolarité de maîtrise. Ces traducteurs se sont tous montrés très conscients de la différence entre « traduction automatique » (traduction faite entièrement par un ordinateur) et « traduction automatisée » (traduction faite par un agent humain aidé d'un ordinateur). Le premier traducteur remarque à ce sujet que « [l]a traduction automatique donne des textes de très mauvaise qualité » tout en mentionnant que « [l]a traduction automatisée est un outil très précieux pour les traducteurs. »⁴⁵ La peur de se

⁴⁴ Martin Kay, *The Proper Place of Men and Machines in Language Translation*, Palo Alto, Xerox Palo Alto Research Center [en ligne], 1980, p. 3. <<http://www.mt-archive.info/Kay-1980.pdf>> (page consultée le 10 janvier 2013).

⁴⁵ Traducteur 1, 2011.

faire remplacer par une machine reste néanmoins très préoccupante, comme le remarque un deuxième intervenant, qui nuance cependant ses propos pour conclure que « la traduction automatique pourrait nous permettre de nous "débarrasser" de textes répétitifs en [sic] ennuyeux et de nous concentrer sur de réels défis. »⁴⁶ L'avis d'un troisième traducteur est encore plus tranché. Bien qu'il ne s'oppose pas à l'utilisation de la traduction automatisée, la traduction automatique le rebute. Il explique : « Tant que ce procédé en sera encore à ses balbutiements, cela ne m'intéresse pas et, au contraire, je dois le combattre et m'opposer à son utilisation, voire son développement, si je veux sauvegarder mon gagne-pain pour quelque temps encore. »⁴⁷

Si ce type de prise de position sans équivoque tend lentement à disparaître, c'est surtout parce que les professionnels de la traduction n'ont plus le choix et doivent s'adapter. En effet, plus que l'informatisation du poste de travail, qui a débuté dans les années 70 pour les grands cabinets, puis qui s'est répandue chez les pigistes avec la montée de l'informatique personnelle, c'est l'apparition d'une culture numérique à laquelle ils ne peuvent plus se soustraire qui met en branle un changement profond dans l'univers de la traduction. Car la localisation, soit « the linguistic and cultural adaptation of digital content to the requirements and locale of a foreign market, and the provision of services and technologies for the management of multilingualism across the digital global information flow »⁴⁸, change la donne. Michael Cronin explique bien le changement que cela implique :

As a result of the digital revolution of the late twentieth century, text has become part of digital content [...] Underlying the informatics revolution is the convertibility, the ultimate translatability, of all content to the binary code of machine language. Computers, which initially only received text, now receive sound and images (both static and animated). At one level, the problem for the translator schooled in written and printed textual traditions is how to deal with these multi-modal textual objects.

En somme, il s'agit de considérer la traduction à l'ère numérique, non plus comme une étape d'un processus à automatiser, mais plutôt comme un processus à intégrer dans un contexte numérique. Il y a un peu plus de dix ans, Doug Robinson proposait un virage en ce sens : « it may be more fruitful to shift the discussion of human-machine interfaces in the translation field from the MT/CAT/HT triad into the realm of cyborg translation. The motto of this new approach would be: all translators are cyborgs. »⁴⁹

⁴⁶ Traducteur 2, 2011.

⁴⁷ Traducteur 3, 2011.

⁴⁸ Schäler, 2007, p.157.

⁴⁹ Douglas Robinson, « Cyborg Translation » dans Susan PETRILLI, (dir.), *La traduzione*. Édition spéciale d'*Athanos: Semiotica, Filosofia, Arte, Letteratura* [en ligne], 10-2, 1999-2000, p. 219-233, <<http://home.olemiss.edu/~djr/pages/writer/articles/print/index.html>> (page consultée le 2 novembre 2014).

Le traducteur et la machine, dit-il, forment une entité traduisante complexe et organisée. Il ne s'agit pas de simple collaboration, mais de symbiose, l'un tirant profit des enseignements de l'autre. « The machine is one of the human's limbs or organs. Together they are a cyborg. »⁵⁰

(Re)Construire l'imaginaire du traducteur-cyborg

Si l'imaginaire du cyborg paraît trop connoté aux yeux de certains, je propose néanmoins de chercher du côté des récepteurs de la traduction et des technologies, particulièrement en ce qui concerne la production culturelle « atypique » pour recadrer la traduction dans un contexte numérique.

Le cas de deux fansubbers d'animés japonais, que j'ai examiné dans des travaux précédents⁵¹, peut mettre en lumière de nouvelles dynamiques d'appropriation et de distribution culturelle qui sont soutenues par une relation à la technique exempte des préjugés de leurs contreparties professionnelles. La traduction d'un animé peu connu, *Berserk*, est passée dans ce cas précis dans les mains de deux enthousiastes français, qui ont fait preuve d'une ingéniosité et d'une créativité surprenante quant à l'utilisation des ressources techniques à leur disposition. Passant d'un « raw » déjà sous-titré en portugais, ils utilisent *Google Translate* et leur connaissance de l'univers berserkien (qui compte un manga de plusieurs volumes et une série animée) pour produire la version française, et *Subtitle Edit* pour intégrer les sous-titres au film. Constatant que la version anglaise se faisait elle aussi attendre, ils utilisent le même principe pour la langue de Shakespeare.

Ce que l'exemple de *Berserk* démontre par ailleurs c'est que les fans membres de ces communautés agissent à la fois comme commanditaires (même si l'échange d'argent est plutôt l'exception), producteurs, traducteurs, distributeurs, critiques et consommateurs de « nouveaux » animés, en s'appropriant non seulement toutes les étapes de production, mais aussi le produit en soi, l'adaptant selon une éthique qui leur est propre.

En somme, si l'imaginaire traditionnel lié à TA, tant sur le plan de la recherche fondamentale, de la recherche appliquée et de l'utilisation professionnelle est « frelaté » comme j'ai tenté ici de le démontrer, il est urgent, pour favoriser la poursuite des avancées technologiques et pour assurer la survie de la profession de traducteur, de recadrer l'utilisation des technologies de traduction automatique et automatisée à l'aide de nouvelles représentations plus positives. La figure du cyborg, comme manifestation de l'appropriation de composantes technologiques par l'humain, ou encore, celle des fansubbers, comme figures rebelles, voire romantiques,

⁵⁰ *Ibid.*

⁵¹ Voir Geneviève Has 2013 et 2014b.

transgressant les frontières établies de la profession, pourraient toutes deux servir de hérault/héros pour une nouvelle ère de la traduction à l'ère de la culture numérique.

Note finale : Cet article reprend en grande partie mes travaux précédents, mais les revisite à la lumière de nouvelles réflexions. Je remercie les éditeurs de m'avoir permis de partager ces dernières. Voir la bibliographie pour les références complètes.

Bibliographie :

ALPAC, *Language and Machines: Computers in Translation and Linguistic*, Washington, D.C., National Academy of Science and National Research Council, 1966.

DEMOS, Kristin et Mark FRAUENFELDER, « Machine Translation's Past and Future », *Wired* [en ligne], no 8.05, mai 2000. <<http://www.wired.com/wired/archive/8.05/timeline.html>> (page consultée le 5 décembre 2011).

FREIGANG, Karl Heinz, « Automation of Translation: Past, Presence, and Future », *Revista Tradumàtica* [en ligne], n° 0, octobre 2001, <<http://www.fti.uab.es/tradumatica/revista/num0/articles/khfreigang/imprimir.pdf>> (page consultée le 2 décembre 2011).

HAS, Geneviève, « Lorsqu'un Français traduit du japonais sans parler anglais, Guts s'en tire-t-il indemne? Les «fansubs» de Berserk et la notion de l'acceptabilité en traduction », *Journée d'études en littérature et résonances médiatiques*, Montréal, Canada, 2013.

HAS, Geneviève, *Le syndrome de Sisyphe dans la recherche en technologies langagières au Canada*, Mémoire de maîtrise, Département d'études françaises, Université Concordia, Montréal, Canada, 2014.

HAS, Geneviève, « La traduction, au cœur des nouvelles pratiques éditoriales », *Séminaire Écritures numériques et éditorialisation* (Paris-Montréal), Sens-public.org, Montréal, Canada, 2014(b).

HUTCHINS, John, « Out of the Shadows: a Retrospect of Machine Translation in the Eighties », *Terminologie et Traduction*, n° 3, 1990, p. 275-292.

HUTCHINS, John, « Latest Developments in Machine Translation Technology: Beginning a New Era in MT Research », *MT Summit IV: International Cooperation for Global Communication. Proceedings, July 20-22, 1993*, Kobe, Japon, 1993, p. 11-34.

HUTCHINS, John, « The Georgetown-IBM demonstration, 7th January 1954 », *MT News International*, n° 8, 1994, p. 15-18.

HUTCHINS, John, « Machine-Translation: A Brief History », *Concise History of the Language Sciences: from the Sumerians to the Cognitivists*, E. F. K. KOERNER et R. E. ASHER (dir.), Oxford, Pergamon Press, 1995, p. 431-445.

HUTCHINS, John, « The Origins of the Translator's Workstation », *Machine Translation*, vol. 13, n° 4, 1998, p. 287-307.

HUTCHINS, John, « Retrospect and Prospect in Computer-based Translation », *Proceedings of MT Summit VII "MT in the great translation era"*, Low Hwee BOON (dir.), Singapour, AAMT, 1999, p. 30-34.

HUTCHINS, John, « Machine Translation over Fifty Years », *Histoire Épistémologie Langage*, tome 23, fascicule 1, 2001, p. 7-31.

ISABELLE, Pierre, *Introduction à la traduction automatique par règles* [en ligne], 2010, <<http://linguistech.ca/display657>> (page consultée le 20 juin 2013).

KAY, Martin, *The Proper Place of Men and Machines in Language Translation*, Palo Alto, Xerox Palo Alto Research Center [en ligne], 1980, <<http://www.mt-archive.info/Kay-1980.pdf>> (page consultée le 10 janvier 2013).

KUHN, Roland, *Introduction à la traduction automatique statistique (TAS)* [en ligne], 2010, <<http://linguistech.ca/display658>> (page consultée le 25 juillet 2013).

L'HOMME, Marie-Claude, *Initiation à la traductique*, Montréal, Linguatex, 2008.

MELBY, Alan, *The Possibility of Languages*, Amsterdam, Philadelphie, John Benjamins Publishing Company, 1995.

PÉREZ, Celia Rico, « From Novelty to Ubiquity: Computers and Translation at the Close of the Industrial Age », *Translation Journal* [en ligne], vol. 5, n° 1, 2001, <<http://translationjournal.net/journal/15mt2.htm>> (page consultée le 28 novembre 2011).

QUAH, Chiew Kin, *Translation and Technology*, New York, Palgrave Macmillan, 2006.

ROBINSON, Douglas, « Cyborg Translation » dans Susan PETRILLI, (dir.), *La traduzione. Édition spéciale d'Athanos: Semiotica, Filosofia, Arte, Letteratura* [en ligne], 10-2, 1999-2000, p. 219-233, <<http://home.olemiss.edu/~djr/pages/writer/articles/print/index.html>> (page consultée le 2 novembre 2014).

ROSNER, Michael, « Machine Translation and Human Translators », Malte, *Sunday Times*, 25 avril 2005.

SOMERS, Harold, *Computers and Translation: A Translator's Guide*, John Benjamin, Amsterdam, 2003.

TAUS, *1954 Machine Translation Movie* [en ligne], 2010, <http://www.youtube.com/watch?feature=player_embedded&v=K-HfpsHPmvw> (page consultée le 15 mai 2013).

TENNEY, Merle D., « Machine Translation, Machine-aided Translation, and Machine-impeded Translation », *Tools for the Trade: Translating and the Computer 5*, Veronica LAWSON (dir.), Londres, Aslib, 1985, p. 105-113.

Traducteurs 1 à 5, Questionnaire. Non publiés, 2011.

WEAVER, Warren, « Translation », *Machine Translation of Languages*, William S. LOCKE et Andrew. D. BOOTH (dir.), Cambridge (Mass.), The Technology Press of the MIT, 1955, p. 15-24.

WHEELER, Peter, « Systran », *Machine Translation Today: The State of the Art*, Margaret KING (dir.), Édimbourg, Edinburgh University Press, 1987, p. 192-208.

WHITECOMB, Tony, « Statistical methods gaining ground », *Language Industry Monitor*, n° 11, Sept-Oct 1992, p. 1-3.

ZUGHOUL, Muhammad Raji et Awatef Miz'Il ABU-ALSHAAR, « English/Arabic/English Machine Translation: A Historical Perspective », *Meta*, vol. 50, n° 3, 2005, p. 1022-1041.