

Le syndrome de Sisyphe dans la recherche en technologies langagières au Canada

Geneviève Has

Mémoire
présenté
au
Département d'Études françaises

Comme exigence partielle au grade de
Maîtrise ès Arts (Traductologie)
Université Concordia
Montréal, Québec, Canada

Août 2013

© Geneviève Has, 2013

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Geneviève Has

Entitled: Le syndrome de Sisyphe dans la recherche en technologies langagières au Canada

and submitted in partial fulfillment of the requirements for the degree of

Maîtrise ès Arts (Traductologie)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Chair
Pier-Pascale Boulanger, Université Concordia

Examiner
Lynne Bowker, Université d'Ottawa

Examiner
Christine York, Université Concordia

Supervisor
Debbie Folaron, Université Concordia

Supervisor
Benoit Léger, Université Concordia

Approved by _____

Chair of Department or Graduate Program Director

Dean of Faculty

Date: September 5th, 2013

RÉSUMÉ

Le syndrome de Sisyphe dans la recherche en technologies langagières au Canada

Geneviève Has

L'industrie de la traduction au Canada est caractérisée par des facteurs sociolinguistiques, politiques et historiques qui l'ont longtemps mise hors d'atteinte des fluctuations internationales de la demande en traduction. Or, ici, comme dans le reste du monde, les effectifs de traduction ne peuvent répondre à la demande croissante. Pour tailler sa place sur le marché mondial, le Canada devrait pouvoir compter sur son industrie des technologies langagières. Cependant, considérant sa position de pionnier dans les années 70 (avec des succès comme TAUM-MÉTÉO et Termium), le Canada accuse aujourd'hui un retard considérable. Nous avons voulu retracer l'histoire des technologies langagières au Canada, particulièrement celle des laboratoires de recherche fondamentale, pour mettre en lumière les défis qu'ont dû relever les chercheurs canadiens. Le mémoire comprend donc une présentation du contexte sociohistorique qui a vu naître les technologies de la traduction au Canada ainsi qu'une étude centrée sur quatre laboratoires canadiens : TAUM, CITI, RALI et CNRC/CRTL. Notre analyse révèle donc que les facteurs sociologiques, politiques et linguistiques propres au Canada ont causé des tensions, des incohérences et des dissensions qui ont considérablement ralenti la recherche fondamentale en technologies langagières au pays.

REMERCIEMENTS

Je souhaite tout d'abord exprimer ma plus grande reconnaissance à mes deux directeurs, Benoit Léger et Debbie Folaron, qui m'ont soutenue et encouragée tout au long de ma recherche. Leur confiance m'est précieuse. J'aimerais aussi remercier le département d'Études françaises de m'avoir confié plusieurs charges de cours, puisque sans les échanges animés et inspirants que j'ai pu entretenir avec les étudiants, ce mémoire ne serait pas le même. Je leur dois beaucoup. Il me faut aussi souligner la générosité des professeurs du département, particulièrement Philippe Caignon, Danièle Marcoux, Sophie Marcotte, Sylvain David, Pier-Pascale Boulanger et Christine York, qui n'ont jamais hésité à partager leurs connaissances et leur expérience, ni lésiné sur les encouragements.

Un grand merci à Pierre Isabelle, Elliott Macklovitch et Jean-François Simard, qui ont gracieusement accepté de répondre à mes nombreuses questions et qui m'ont fait découvrir un monde fascinant. Leur apport à ma réflexion est inestimable.

Merci aussi à mes anciens collègues du Bureau de la traduction, avec qui j'ai fait mes premiers pas dans l'univers des technologies langagières, à mes collègues étudiantes, Luce, Mona, Dominique, Alexandra, Carol, particulièrement, pour avoir coloré mes études de si belle façon. Claudia, surtout, pour les séances d'écriture et de fou rire. Merci à tous ceux qui ont discuté avec moi de mes émois technologiques (et philosophiques) : Yves, Jocelyn, Pierre-Alain, Marc, Simon, Pierre et Tommy. Un grand merci à AnneMarie Taravella pour ses conseils, ses bons mots, son enthousiasme. Un merci tout particulier à Sylvie Ditcham pour m'avoir hébergée, corps et âme, au moment crucial, et à Johanne Durocher-Norchet pour sa relecture, ses conseils, son amitié. Enfin, merci du fond du cœur à Catherine, à Evelyn et à Vincent, vous savez pourquoi.

Mon travail de recherche a pu être entamé grâce à la bourse de maîtrise Figura, le Centre de recherche sur le texte et l'imaginaire (antenne Concordia), et à celle du Rhoda and Arthur Rodbell Graduate Scholarship, je leur exprime toute ma gratitude. J'aimerais aussi souligner l'apport du laboratoire NT2, et de ses collaborateurs, et le plaisir que j'ai eu à m'y impliquer.

Pour Diane
1959-2010

Table des matières

Table des figures.....	vii
Introduction	1
Chapitre 1 Les technologies langagières	11
1.1 Terminologie des technologies langagières.....	11
1.2 De la traduction humaine à la traduction automatique.....	13
1.3 Les saveurs de la traduction automatique.....	16
1.3.1 Traduction automatique (TA) et traduction entièrement automatique de haute qualité (TEAHQ)	16
1.3.2 Traduction machine assistée par l’humain (TMAH).....	19
1.3.3 Architectures.....	24
1.3.3.1 Approche directe.....	25
1.3.3.2 Approches à base de règles (ou approches indirectes)	30
1.3.3.2.1 Approche par interlangue	31
1.3.3.2.2 Approche par transfert	34
1.3.3.3 Traduction automatique à base de corpus.....	40
1.3.3.3.1 Traduction automatique à base d’exemple	40
1.3.3.3.2 Traduction automatique statistique.....	42
1.4 Traduction humaine assistée par ordinateur (THAO) et traduction assistée par ordinateur (TAO).....	47
Chapitre 2 Contexte canadien.....	57
Chapitre 3 Les grands laboratoires canadiens, étude de cas.....	72
3.1 Traduction Automatique Université de Montréal.....	72
3.2 Centre d'innovation en technologies de l'information (CITI).....	78
3.3 Recherche appliquée en linguistique informatique (RALI)	91
3.4 Centre National de Recherches Canada (CNRC) et le Centre de recherche en technologies langagières (CRTL).....	94
Conclusion	101
Bibliographie	106
ANNEXE I – Formulaire de consentement	118
Pierre Isabelle	118
Elliott Macklovitch.....	120
Jean-François Richard	122

Table des figures

Figure 1 – Termes courants en traduction automatique	12
Figure 2 – Traduction automatique et traduction humaine (Hutchins et Somers, 1992, p. 148, [nous traduisons])	14
Figure 3 – Classification des technologies langagières	14
Figure 4 – Modèle pour l'utilisation future des technologies de traduction (Quah, 2006, p. 154 [nous traduisons])	18
Figure 5 – Modèle de TMAH (Quah, 2006, p. 12 [nous traduisons])	22
Figure 6 – Architecture des systèmes de TA (Quah, 2006, p. 168 [nous traduisons])	25
Figure 7 – Triangle de Vauquois, approche directe (adapté de Isabelle, 2010, p. 6)	26
Figure 8 – Système de TA directe (Hutchins et Somers, 1992, p. 72 [nous traduisons])	26
Figure 9 – Triangle de Vauquois, approche par interlangue (adapté de Isabelle, 2010, p. 8)	32
Figure 10 – Modèle de TA par interlangue (Hutchins et Somers, 1992, p. 74 [nous traduisons])	32
Figure 11 – Triangle de Vauquois, approche par transfert (adapté de Isabelle, 2010, p. 7)	36
Figure 12 – Modèle de système utilisant l'approche par transfert (Quah, 2006, p. 74 [nous traduisons])	37
Figure 13 – Modèle de TABE (Quah, 2006, p. 81 [nous traduisons])	41
Figure 14 – Modèle de TAS (adapté de Quah, 2006, p. 78 [nous traduisons] et Kuhn, 2010, 23-24)	44
Figure 15 – Tableau comparatif de traductions fournies par Google Translate à trois ans d'intervalle (données 2010 dans Kuhn, 2010, p. 26)	46
Figure 16 – Fonctionnement d'un logiciel de mémoire de traduction seul	52
Figure 17 – Structure et processus d'assemblage de la TransBase (Isabelle et coll., 1993, p. 1138 [nous traduisons])	87

Introduction

L'histoire des technologies langagières canadiennes en est une d'occasions manquées. Ce constat s'est imposé sans équivoque au cours de nos recherches. Occasions manquées, puisque le Canada, son industrie de la traduction, ainsi que celle des technologies qui l'accompagnent inévitablement devraient, pour des raisons que nous détaillerons au cours de cette réflexion, être aujourd'hui en position de chef de file sur le plan international. Force est de constater que nous sommes loin du compte aujourd'hui.

Nos recherches ont montré que la traduction au Canada jouit d'un statut unique. Ici, la demande en traduction n'est pas tributaire d'un marché local fluctuant ou le résultat d'exigences économiques dictées par les gros joueurs internationaux, mais bien d'un cadre législatif stable et solide. La *Loi sur les langues officielles*, adoptée en 1968, a eu pour effet de créer une « bulle » qui rend le Canada unique en ce qui concerne la traduction. En effet, puisque la *Loi* oblige le gouvernement canadien à rendre accessibles tous les documents officiels à la fois en anglais et en français, et que la plupart de ces documents sont produits en anglais, la demande en traduction vers le français est volumineuse et constante. Cette « bulle » est unique parce qu'elle met l'industrie de la traduction à l'abri du marché international, quoique ce dernier, il faut le préciser, est aussi influencé par des exigences législatives qui lui sont propres. À l'abri, puisque les traducteurs canadiens n'ont en général pas à se former pour la traduction multilingue, comme c'est souvent le cas ailleurs dans le monde¹. Par ailleurs, les traducteurs canadiens n'ont jamais été suffisamment nombreux pour répondre à la demande.

¹ Cette différence est particulièrement évidente lorsque l'on compare les programmes de traduction offerts au Canada avec ceux des pays européens. Dans « L'enseignement de la traduction au Canada », Geneviève Mareschal souligne que « Le marché de la traduction au Canada étant axé sur l'anglais et le français, les programmes de traduction ont été conçus autour de ces deux langues, et souvent de ces deux langues seulement. Cette conception de la formation à la traduction professionnelle va à l'encontre de la tradition

Cette insuffisance des effectifs de traduction ne date pas d'hier et est loin d'être restreinte au Canada. Les tentatives d'automatisation de la traduction remontent aussi loin que le XVII^e siècle en Europe. Déjà, à l'époque, on essaie d'encoder la langue de départ et la langue d'arrivée, pour ensuite concevoir des calculs compliqués visant à passer de l'une à l'autre sans que l'opérateur n'ait besoin de connaître la langue d'arrivée. Ces premières ébauches, bien que grossières, ont tout de même donné le ton aux tentatives subséquentes, notamment celles de l'Américain Warren Weaver au tournant de la guerre froide. Cette période, qui voit la naissance de la traduction automatique (TA) comme nous la connaissons, est effervescente 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 médias et des scientifiques font preuve, avec le recul, d'un optimisme bon enfant qui frise la naïveté. Rapidement, cependant, l'élan des chercheurs allait être stoppé net. En 1966, l'Automatic Language Processing Advisory Committee (ALPAC) conclut que la traduction automatique est inefficace, inadéquate et trop coûteuse. Dès lors, le financement se tarit. Les groupes de recherche sont démantelés, les scientifiques se tournent vers d'autres secteurs de recherche, qui ont l'heur de mieux plaire aux bailleurs de fonds.

Ce que les historiens relèvent rarement, c'est que le *Black Book of Machine Translation* (le rapport de l'ALPAC, publié en 1966, dont nous reparlerons) ne condamne pas d'emblée

européenne, par exemple, où les programmes exigent des candidats la connaissance d'au moins trois langues, sinon quatre, et offrent une formation dans de nombreuses combinaisons linguistiques. » (Mareschal, 2005, p. 252)

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. Parallèlement, la motivation derrière la recherche se déplace lentement des visées militaires vers des objectifs plus économiques (même si le nombre insuffisant de traducteurs reste un facteur constant à travers le monde). À preuve, de nombreux groupes de recherche sont financés par des entreprises privées qui cherchent à diminuer leurs coûts de traduction. Par ailleurs, au cours des années 1970, de nombreux chercheurs ne succombent plus au mirage de la TEAHQ, notamment Martin Kay et Yeoshua Bar-Hillel, deux figures de proue de la recherche en TA.

Outre le fait que les militaires ne sont plus les principaux commanditaires de la recherche en TA, il faut souligner que le développement rapide de l'informatique joue un rôle prépondérant dans l'évolution de la recherche. Si l'on considère, en effet que l'ENIAC (Electronic Numerical Integrator and Computer), le premier ordinateur, conçu en 1946 aux États-Unis, pouvait effectuer 5000 additions simples à la seconde, mesurait 2,4 m x 0,9 m x 30,0 m et avait englouti plus de cinq millions en dollars américains d'aujourd'hui, il est évident que les ressources disponibles dans les années 70 paraissaient astronomiques. En comparaison, le Cray I, installé au Los Alamos National Laboratory, aux États-Unis, en 1976 traitait 166 millions d'opérations à la seconde et pouvait stocker un million de mots (l'ENIAC n'avait pas de capacité de stockage), reposait sur une base d'un diamètre de 2,63 m, s'élevait à 1,96 m et se détaillait entre 20 et 33 millions de dollars américains

d'aujourd'hui. Ce bond en avant vertigineux est essentiel au développement de systèmes de plus en plus complexes et subtils. Il va sans dire que sans eux, la traduction automatique n'existerait tout simplement pas. Ces superordinateurs de quatrième génération permettent l'archivage de données sur une échelle jamais vue et ouvrent des voies que la TA n'avait pu emprunter auparavant.

C'est dans ce contexte que la recherche sur les technologies appliquées à la traduction s'amorce au Canada. Ici, comme nous l'avons déjà fait remarquer, c'est la *Loi sur les langues officielles* qui constitue la motivation première pour la recherche en TA. De nombreux projets voient le jour, principalement dans les universités canadiennes et au Bureau de la traduction (BT). Nous avons aussi relevé quelques entreprises qui ont vu le jour au sein de ministères fédéraux, notamment Justice et Agriculture Canada. Au meilleur de notre connaissance, le privé est entré relativement tard dans la course de la R et D. Ceci dit, les scientifiques canadiens sont, au tournant des années 70, à la fine pointe de la recherche. L'un des succès de la TA les plus reconnus mondialement a vu le jour en 1977 à Montréal. TAUM-MÉTÉO, commissionné par le Secrétariat d'État du Canada (le futur Bureau de la Traduction), se veut une tentative de diminuer le nombre de traducteurs affectés à la traduction de bulletins météorologiques, tâche fastidieuse, s'il en est une. Le système est un succès et est déployé au service météorologique, pour la plus grande joie des traducteurs. Cet exploit, l'implantation à grande échelle d'un système de traduction automatique au sein d'un organisme gouvernemental, ne sera pas répété avant les années 2010. Portage, un système de traduction automatique à base statistique, développé par les scientifiques du CNRC (sous la gouverne de Pierre Isabelle, qui a participé à MÉTÉO) a été déployé tout juste l'année dernière au BT. Il est évidemment trop tôt pour déterminer si l'entreprise aura le même succès à long terme que TAUM-MÉTÉO.

Toutefois, le fait qu'en près de 40 ans il soit impossible de retracer une collaboration fructueuse et d'une telle ampleur entre les chercheurs canadiens en TA et le plus gros fournisseur de traduction au pays, alors que les deux groupes relèvent de l'autorité du gouvernement fédéral, qui rappelons-le, a tout intérêt à chercher à diminuer les coûts engendrés par l'obligation de traduire tous ses documents dans les deux langues officielles, a de quoi faire sourciller. C'est cette question épineuse qui a guidé la réflexion que nous exposerons ici. Pourquoi, alors que le gouvernement canadien disposait d'une motivation forte à soutenir la recherche en TA, pouvait se targuer de l'une des plus grandes réussites du domaine et employait toujours les chercheurs qui ont réalisé ce projet, la collaboration ne s'est-elle pas poursuivie?

Il n'existe évidemment pas de réponse toute faite à cette question. Le travail que nous avons entrepris en est donc un d'investigation. Au fil de nos lectures et de nos entretiens, toutefois, une possible réponse s'est esquissée. Malgré une apparente collaboration, les acteurs des différents secteurs (gouvernement, chercheurs, industrie) évoluent en fait en vase clos, les uns ignorants des contraintes des autres, ces derniers sans percevoir les motivations des premiers. Il ne s'agit pas ici de faire porter le blâme à l'un ou l'autre de ces groupes, mais bien de souligner la dynamique qui sous-tend leurs relations, et de comprendre pourquoi ils n'ont pas su tirer profit des occasions de collaboration qui s'offraient à eux.

Pour ce faire, nous avons choisi d'étudier quatre laboratoires canadiens de recherche en technologies langagières, en commençant par TAUM, dans les années 60, pour finir avec le CRNC/CRTL, encore en opération de nos jours. Ainsi, les quatre laboratoires traités dans ce mémoire l'ont été parce qu'ils sont les héritiers directs de TAUM. Nous avons cherché à donner une voix en traductologie aux chercheurs de traduction automatique, qui souvent

évoluent dans des cercles complètement différents (quoique certains fassent exception). En fait, le plus souvent, ce sont les traducteurs et les traductologues qui ne veulent pas s'aventurer dans le monde du traitement automatique des langues, un monde mathématique, ou statistique, en tous les cas, étranger. Cette intention teinte évidemment notre réflexion.

D'autres ont colligé de façon plus technique le récit que nous nous apprêtons à relater, John Hutchins, notamment. Cet historien de la traduction automatique a non seulement publié des dizaines d'ouvrages, des centaines d'articles et présenté d'innombrables communications sur le sujet, il a aussi compilé une énorme base de données comprenant la (presque) totalité des écrits de toutes formes sur la traduction automatique, la traduction assistée par ordinateur, le traitement informatisé des langues et la linguistique computationnelle. La *Machine Translation Archive* (<http://www.mt-archive.info/>) est la référence sur le sujet. Elle comprend près de 20 000 entrées organisées selon différents paramètres de recherche. Il nous a fallu faire un tri dans toutes ces informations. Évidemment, nous renvoyons à Hutchins à de nombreuses reprises puisqu'il s'est l'auteur de nombreux comptes rendus historiques. Bibliothécaire de formation et de carrière, il rédige d'abord des articles sur la linguistique et sur la recherche d'information, puis se spécialise en traduction informatique. Il fait preuve, dans ses comptes rendus, d'une précision technique que les néophytes (dont nous étions) ont parfois du mal à suivre. Il était donc primordial, non seulement de rassembler l'information technique pertinente chez Hutchins, mais aussi l'information historique qu'il était possible de glaner en filigrane dans ses publications. Cependant, il nous a fallu chercher ailleurs pour compléter ces informations qui ont constitué la pierre d'assise de notre démarche.

Ainsi, nous avons aussi consulté des auteurs, tout aussi précis, mais qui écrivent pour un public profane, comme Lynne Bowker (traductologie), Marie-Claude L'Homme

(traductologie), Alan Melby (linguistique, traduction), Frank Austermühl (traductologie), Harold Somers (ingénierie linguistique), Chiew Kin Quah (traduction, technologies langagières), parmi d'autres. Leurs explications, mises en contexte, leurs schémas et leur vulgarisation nous ont été d'une aide incomparable pour arriver à distinguer des concepts aussi obscurs que la traduction automatique statistique à base d'exemple ou la différence entre un concordancier et un bitexte. Nous espérons leur avoir rendu justice en compilant leurs explications. Nous avons aussi fait appel à quelques grands noms de la recherche en traduction automatique, comme Martin Kay, de la Stanford University ou Yeoshua Bar-Hillel, de la Hebrew University, lorsque nécessaire. Ce faisant, nous avons buté sur deux principaux obstacles tout au long de ce travail de débroussaillage technico-historique : la fragmentation de l'information et la multidisciplinarité des auteurs. En fait, les deux problèmes sont résolument liés. En effet, le premier est en quelque sorte la conséquence du second. Il va sans dire que les auteurs que nous mentionnons ici ont tous déjà effectué un travail de compilation substantiel, mais nous avons tout de même dû colliger un nombre important d'articles, de monographies, d'encyclopédies, de ressources de toute nature afin de nous assurer de bien comprendre les nuances propres à chacune des disciplines pour les résumer efficacement. L'extraction et la triangulation des données historiques a aussi été un défi particulier. Chacun de ces auteurs présente sa vision, sa perspective sur l'un ou l'autre des aspects liés aux technologies langagières, et la mise en commun de ces multiples facettes forme une image kaléidoscopique parfois déconcertante. Ce mémoire se veut une tentative de synthèse de ces points de vue.

Puisque nous nous sommes employée à relater l'histoire des laboratoires canadiens, nous nous sommes évidemment concentrée sur les publications des chercheurs canadiens actifs au sein de ces laboratoires et avons dû passer sous silence la perspective d'autres

acteurs tout aussi importants, comme les institutions gouvernementales, les joueurs de l'industrie et les universités. Nous nous sommes surtout intéressée aux circonstances qui entourent le développement de tels outils ou encore la naissance de tel laboratoire. Pour ce faire, il a fallu recouper l'information, distinguer l'aspect sociohistorique des données purement techniques. Dans certains cas, nous avons dû remonter directement à la source, puisqu'il nous manquait certains détails. Elliot Macklovitch (Bureau de la traduction, TAUM, CITI, RALI), Pierre Isabelle (TAUM, CITI, RALI, CRNC) et Jean-François Richard (Terminotix) ont eu la gentillesse de nous faire part de leur point de vue sur les événements que nous relatons ici dans le cadre d'entretiens informels. Ces entretiens nous ont permis de glaner des détails absents des comptes rendus officiels et des publications scientifiques. Nous nous sommes employée à faire ressortir les motivations des chercheurs, à expliquer les mouvements que nous avons identifiés au cours de nos lectures préparatoires.

Au cours de nos recherches, nous avons aussi rencontré des gens mêlés de près ou de loin au secteur des technologies langagières : chercheurs, professeurs, traducteurs, gestionnaires, entrepreneurs, fonctionnaires, étudiants, et nos discussions ont aussi influencé notre analyse. Parce qu'évidemment, nous ne prétendons pas à une objectivité sans failles, nous avons décidé de nous concentrer sur un seul aspect du problème. Nous avons aussi dû laisser de côté certains projets ou initiatives qui sans manquer d'intérêt, soit faisaient double emploi, soit illustraient un point accessoire. Aussi souvent que possible nous les avons mentionnés en note.

Considérant la richesse et l'exhaustivité de nos sources, d'aucuns pourraient se demander pourquoi nous avons décidé de les revisiter. En fait, l'histoire des technologies langagières au Canada, même si elle a déjà été relatée, est morcelée au gré des publications

dans des domaines disparates et nous n'avons pu nous fier qu'à une seule autorité en la matière pour répondre à la première question qui nous a mise sur la piste de ce projet d'études. Nous avons été frappée, lors de notre passage au Bureau de la traduction, de 2008 à 2010, par la diversité des outils mis à la disposition des traducteurs et par le peu de connaissance (et nous devons l'avouer, du peu d'intérêt) que ces derniers avaient à leur sujet. Pourquoi les traducteurs que nous avons rencontrés démontraient-ils tant de méfiance envers des outils qui, somme toute, visaient à leur faciliter la tâche? C'est pourquoi nous avons décidé de plonger dans l'histoire des laboratoires canadiens, en l'inscrivant dans le contexte plus large de l'industrie de la traduction et en empruntant une approche sociohistorique.

Nous désirons donc que ce mémoire, en raison de la visée vulgarisatrice et synthétique qui nous a animée tout au cours de la rédaction, puisse servir de point de départ à l'étude des technologies langagières, particulièrement d'une perspective canadienne. Puisque les ressources déjà disponibles sont en grande partie distribuées à travers plusieurs disciplines et rédigées pour la plupart en anglais et dans un style très technique, nous espérons que notre travail préliminaire de compilation, de terminologie et de traduction puisse faciliter la tâche à ceux qui choisiront de remonter d'autres filières, comme celle de l'industrie, ou celle des institutions, et ainsi compléter le portrait dont nous esquissons ici les premiers traits.

Nous amorcerons notre analyse en définissant plus précisément les termes et concepts liés aux technologies langagières que nous utiliserons. Le chapitre 1 sera donc consacré à une mise en contexte des différentes technologies langagières, de leur fonctionnement et du cadre théorique sur lequel elles s'appuient. Cette partie se veut résolument vulgarisatrice et s'adresse particulièrement à des lecteurs qui n'auraient aucun bagage technique concernant

la question. Son découpage, croyons-nous, facilite la lecture et permet de trouver rapidement la section d'intérêt pour le lecteur un peu plus avancé qui voudrait vérifier un point ou en approfondir un autre. Nous poursuivrons notre discussion en délimitant le cadre sociopolitique particulier du Canada. Ainsi, le chapitre 2 trace le portrait de l'industrie langagière au Canada et montre pourquoi et en quoi les facteurs socio-politico-linguistiques sont le moteur de la recherche et de l'industrie des technologies langagières. Cette charnière nous permet d'articuler comment le Canada aurait pu tirer parti des conjonctures historiques qui lui sont propres pour soutenir la recherche fondamentale et de suggérer une piste quand à expliquer pourquoi il ne l'a pas fait. Nous consacrerons, enfin, le chapitre 3 à l'histoire de quatre grands laboratoires canadiens de recherche en TA : TAUM, le CITI, le RALI et le CNRC/CRTL en vue de faire ressortir les lignes de tension et les incohérences qui ont empêché le Canada de demeurer à la tête du peloton de la recherche en TA. Ces laboratoires, nous l'avons mentionné, ont été choisis parce qu'ils constituent en quelque sorte une seule et même lignée, sont les témoins d'un seul et même héritage scientifique. Nous croyons que relater leur histoire nous permet de mieux comprendre les enjeux de la recherche fondamentale et appliquée dans le contexte des technologies langagières canadiennes. Nous concluons notre réflexion en récapitulant les défis que les instances gouvernementales, les chercheurs en TA, les traducteurs et les universités devront relever ensemble pour consolider l'industrie canadienne de la traduction et des technologies langagières.

Chapitre 1 Les technologies langagières

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-impaired Translation », *Tools for the trade: Translating and the Computer* 5, 1985, p. 105)

It is time to adopt a new paradigm for our profession: Not only is it acceptable to use computers well – it is critical to our success. (Jost Zetzsche, *The Translator's Tool Box: A Computer Primer for Translators*, 2012, p. 1)

1.1 Terminologie des technologies langagières

Nous avons choisi de rassembler sous le terme « technologies langagières » les différents outils informatiques qui soutiennent le traducteur dans sa tâche. Ce terme, plutôt large, rend bien compte de la diversité des approches et des outils conçus spécifiquement pour les traducteurs, mais aussi de celle des outils que ceux-ci peuvent s'approprier sans qu'ils aient nécessairement été conçus à leur intention. En effet, le vocable « technologies langagières » permet d'inclure dans notre propos de la recherche qui n'est pas nécessairement en lien avec la traduction, en linguistique computationnelle ou en traitement de l'information par exemple, puisque les applications développées dans ces champs sont souvent d'une grande utilité pour les technologies appliquées à la traduction. Par ailleurs, selon notre expérience au gouvernement du Canada du moins, c'est le terme qui est en usage. Il s'agit néanmoins d'un terme très large qu'il convient de diviser en catégories plus digestibles, d'autant plus que les spécialistes, en fonction de leur domaine, ne s'entendent pas nécessairement sur la terminologie.

Dans les ouvrages et articles sur le sujet, nombre d'acronymes et de termes semblent parfois se recouper. Voici les principaux (les synonymes sont regroupés) :

Termes français	Termes anglais
<ul style="list-style-type: none"> • Traduction automatique (TA) • Traduction machine (TM) 	<ul style="list-style-type: none"> • Machine translation (MT)
<ul style="list-style-type: none"> • Traduction entièrement automatique de haute qualité (TEAHQ) • Traduction machine entièrement automatique de haute qualité (TMEAHAQ) 	<ul style="list-style-type: none"> • Fully automatic high quality translation (FAHQT) • Fully automatic high quality machine translation (FAHQMT)
<ul style="list-style-type: none"> • Traduction machine assistée par l'humain (TMAH) • Traduction automatique assistée par l'humain (TAAH) 	<ul style="list-style-type: none"> • Human-aided/assisted machine translation (HAMT)
<ul style="list-style-type: none"> • Traduction humaine assistée par ordinateur (THAO) 	<ul style="list-style-type: none"> • Machine-aided/assisted human translation (MAHT)
<ul style="list-style-type: none"> • Traduction assistée par ordinateur (TAO) 	<ul style="list-style-type: none"> • Computer aided/assisted translation (CAT) • Machine aided/assisted translation (MAT)
<ul style="list-style-type: none"> • Traduction humaine (TH) • Traduction humaine classique (TAC) 	<ul style="list-style-type: none"> • Human translation (HT) • Traditional human translation (THT)

Figure 1 – Termes courants en traduction automatique

Comme le remarque Chiew Kin Quah dans *Translation and Technology* (2006), il n'est pas toujours aisé de bien faire la distinction entre les différentes catégories, principalement parce que la terminologie change selon la discipline (p. 6), mais nous nous emploierons tout de même dans les prochaines sections à les définir plus précisément. Qui plus est, nous les définirons en français, ce qui sous entend un travail de traduction et de terminologie d'une relative ampleur, puisque nombre des sources auxquelles nous référons sont de langue anglaise. Règle générale, les traducteurs et traductologues préfèrent parler de traduction automatique, tandis que les informaticiens (particulièrement en linguistique computationnelle) optent plus souvent pour le terme traduction machine. Nous avons toutefois remarqué qu'en français, du moins, le terme « traduction automatique » semble supplanter « traduction machine ». Il se peut aussi que le terme « traduction machine » soit en perte de vitesse en raison de la sensibilité des utilisateurs, notamment les traducteurs, qui ne sont pas chauds à l'idée de voir leur travail assimilé à une série de tâches mécaniques. Par ailleurs, même les informaticiens déplorent de constater qu'ils doivent composer avec

[...] an old-fashioned term [machine translation], redolent of a pre-electronic and pre-computer age. More particularly, however, this term implies only purely

automatic systems not involving human participation at any stage. It excludes, almost by definition, all kinds of computer aids for translation. (Hutchins, 2000, p. 62)

Nous verrons en effet que la participation humaine dans la traduction automatique est, à quelques exceptions près, inévitable.

1.2 De la traduction humaine à la traduction automatique

Si traductologues, traducteurs, linguistes et informaticiens ne s'entendent pas nécessairement sur les attributs qui caractérisent la traduction automatique (TA) ou la traduction assistée par ordinateur (TAO), voire la traduction humaine (TH), il est tout de même possible de dégager un terrain d'entente en ce qui concerne les technologies langagières. Tous les spécialistes qui s'intéressent à ces dernières d'une façon ou d'une autre les positionnent le long d'un continuum. En somme, les différences tiennent à la place qu'occupe chacun des types de technologie sur le continuum, mais surtout à leurs frontières, quoique ces dernières tendent à s'estomper au rythme du développement des nouvelles applications. En outre, les catégories, autrefois nettement séparées, se combinent désormais et donnent naissance à des hybrides que l'on serait bien en peine de classer² (voir Hutchins, 1995, p. 433).

Hutchins et Somers, en 1992, proposent une première schématisation du continuum, que nous reproduisons ici.

² Ceci dit, il est important de souligner que l'hybridité est une caractéristique des technologies en général, d'une part parce que les utilisateurs continuent de se servir des « vieilles » technologies, obsolètes ou non (pensons par exemple à la télécopie qui continue d'exister avec le courriel, alors que les deux technologies ont la même visée); d'autre part, parce que les formats de conservation et d'échange de données ne sont pas uniformes, les concepteurs et développeurs doivent s'assurer de la compatibilité de leur technologie avec ce qui existe déjà. Dans le contexte de la TA, l'hybridité est certainement l'une des pistes de recherche les plus prometteuses (Quah, 2006, p. 85).

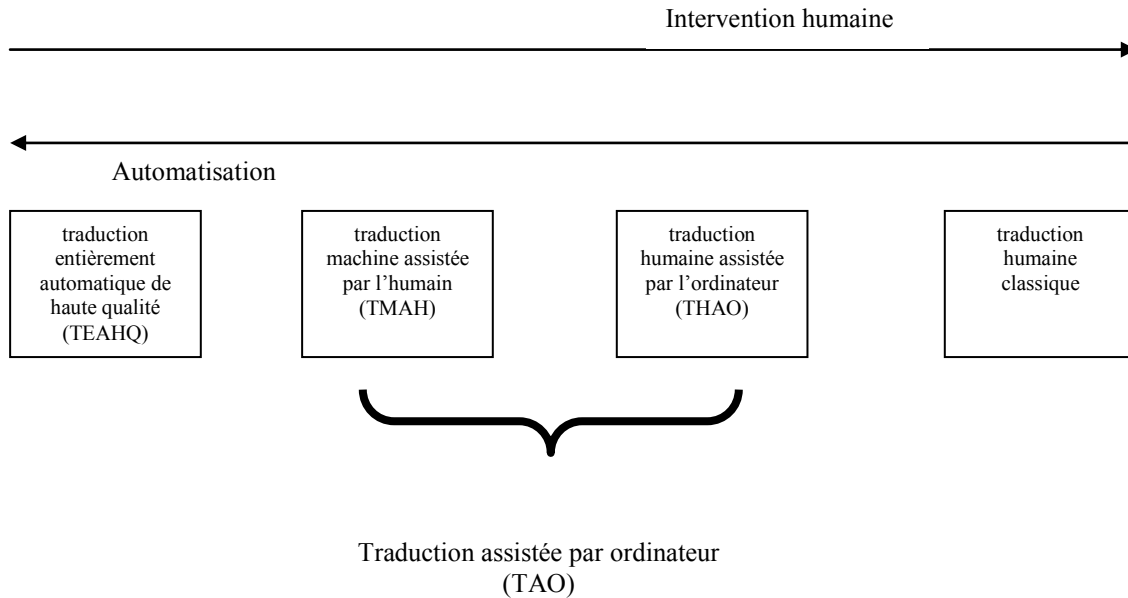


Figure 2 – Traduction automatique et traduction humaine (Hutchins et Somers, 1992, p. 148, [nous traduisons])

Les différents types de technologies langagières sont donc distribués le long d'un axe double, de la traduction complètement humaine à la traduction complètement automatisée. C'est la proportion d'automatisation, ou de participation humaine, selon le point de vue, qui varie selon les auteurs. Lynne Bowker, par exemple, conçoit les technologies langagières en termes de soutien au traducteur. Elle propose donc trois catégories distribuées le long de l'axe TA/TH.

TA	TAO	TH
<ul style="list-style-type: none"> • Systèmes de traduction automatique 	<ul style="list-style-type: none"> • Outils de saisie de données • Outils d'analyse de corpus • Systèmes de gestion de terminologie • Mémoires de traduction • Outils de localisation et de traduction de page Web • Outils de diagnostic 	<ul style="list-style-type: none"> • Outils de traitement de texte • Correcteurs d'orthographe et de grammaire • Ressources électroniques (p. ex., cédéroms) • Internet (p. ex., www, courriel)

Figure 3 – Classification des technologies langagières (Bowker, 2002, p. 7 [nous traduisons et nous avons inversé l'ordre des catégories pour permettre la comparaison avec la figure 2])

Cependant, chacune des technologies présentes sur le continuum de Bowker implique l'humain. Dans sa définition de la TA (MT- machine translation au glossaire), l'auteure indique que « it is the computer, rather than the human, that produces an actual draft translation » (Bowker, 2002, p. 149) et renvoie à l'entrée sur les TAO (CAT). Puisque l'ouvrage de Bowker s'adresse avant tout aux traducteurs désirant utiliser les technologies langagières, il est tout naturel de retrouver l'intervention humaine comme point focal de la TA. Récemment, cette dernière a d'ailleurs fait son entrée dans des applications de TAO comme SDL Trados et Logiterm, comme une option parmi d'autres outils à la disposition du traducteur qui, même s'il ne participe pas à la production automatisée de la traduction, peut tout de même décider de l'utiliser en tout ou en partie.

Comme Lynne Bowker, nous sommes d'avis que la traduction humaine ne peut plus se concevoir sans l'apport d'une technologie, qu'elle soit pensée expressément pour les traducteurs (comme une mémoire de traduction) ou non (comme un logiciel de traitement de texte). D'autres auteurs, c'est le cas de Frank Austermühl, abondent dans le même sens. « The use of information and communication technology (ICT) is a *fait accompli* in the lives of today's language professionals » remarque-t-il en 2001 dans *Electronic Tools for Translators* (p. 7). Il reprend par ailleurs le schéma d'Hutchins et Somers tout en précisant que pour les visées de son ouvrage, il élimine à la fois la traduction entièrement automatisée de haute qualité et la traduction humaine classique des sujets traités. Encore une fois, c'est l'intervention humaine, ou pour reprendre les mots d'Austermühl, le carrefour où l'humain et la machine peuvent coopérer (*ibid.*, p. 11), qui est la caractéristique principale des technologies langagières telles que définies par le traducteur ou le traductologue.

1.3 Les saveurs de la traduction automatique

1.3.1 Traduction automatique (TA) et traduction entièrement automatique de haute qualité (TEAHQ)

La European Association for Machine Translation (EAMT) soutient que la TA « is the application of computers to the task of translating texts from one natural language to another » (EAMT, 2013). Le groupe de certification de l'International Association for Machine Translation³ (IAMT) propose, quant à lui, qu'un système de traduction automatique commercial soit défini comme suit : « A software system is a “machine translation system” if it takes input in the form of full sentences at a time and generates corresponding full sentences (not necessarily of good quality) » (Hutchins, 2000, p. 63). Chez Hutchins et Somers, la TA « is the now traditional and standard name for computerised systems responsible for the production of translations from one natural language into another, with or without human assistance. (1992, p. 3)

Ces définitions s'accordent donc sur le fait que la TA doit, à partir d'un texte source de langue naturelle, produire un texte cible aussi en langue naturelle, excluant d'emblée les systèmes qui produiraient des traductions mot-à-mot et les « dictionnaires de traduction ». Cependant, ces définitions restent plutôt vagues quant au degré de participation de l'humain dans le processus. Ce sujet, nous y avons fait allusion, est loin de faire consensus.

Hutchins et Somers, dans la figure 2, opposent à la TH classique la traduction entièrement automatique de haute qualité (TEAHQ). La TEAHQ est en quelque sorte le Graal de la recherche en automatisation de la traduction. Un tel système devrait arriver à traduire n'importe quel texte, général ou non, de langue naturelle, sans intervention

³ La International Association for Machine Translation (IAMT) regroupe la European Association for Machine Translation (EAMT), la Association for Machine Translation in the Americas (AMTA) et la Asia-Pacific Association for Machine Translation (AAMT). À notre connaissance, ni l'AMTA ni l'AAMT ne proposent de définition officielle pour la TA.

humaine en amont ou en aval (en d'autres mots, aucune prédiction ou postédition ne doit être nécessaire), de telle sorte que l'on ne puisse faire la différence avec une traduction humaine. C'était évidemment l'objectif premier de la TA, cependant, les chercheurs constatent assez rapidement qu'il sera presque impossible de réaliser cette ambition. En conséquence, la TA aujourd'hui cherche à automatiser la traduction, mais sans pour autant viser la perfection (précisons ici que pour les raisons que nous exposerons au chapitre 2, les canadiens sont particulièrement pointilleux quant à la qualité des textes traduits et que leurs exigences ont longtemps eu pour conséquence un manque de tolérance important pour les maladroites des systèmes de TA).

La mondialisation de l'information et l'ouverture des marchés qui en découle rendent la TA (tout comme la TAO, par ailleurs), plus nécessaire que jamais. L'International Data Corporation, une firme d'analyse des marchés, prédisait en 2011 que la quantité d'information créée et partagée dans la sphère numérique dépasserait les 1,8 zettaoctet et que cette croissance serait exponentielle (International Data Corporation, 2011, p. 1). Une grande partie de ces données doivent être traduites, pour une raison ou une autre, et il est évident que les traducteurs professionnels ne suffisent pas à la demande⁴. Si la traduction collaborative, c'est-à-dire la collaboration entre traducteurs de différents horizons pour un projet commun, trop souvent sans rémunération et en général par le truchement de technologies, est une tendance indéniable (Désilets, 2007; Munday, 2008, p. 282; O'Hagan, 2011, notamment), ce mouvement est encore loin de pouvoir livrer la marchandise (sans compter que les experts ne s'entendent pas sur le fait que cette approche soit une pratique souhaitable). La traduction automatique et l'automatisation de la traduction sont donc des

⁴⁴ Pour cette raison, il se peut que le seuil de tolérance auquel nous référerions au paragraphe précédent s'abaisse d'autant, considérant les nouveaux besoins en traduction.

pistes de solution inévitables pour répondre à la demande, dans la mesure où l'on peut définir clairement les degrés de qualité requis en fonction de la nature des textes à traduire. Pour baliser l'utilisation de la TA dans ce contexte, Quah propose le modèle de Schäler, Way et Carl (2003).

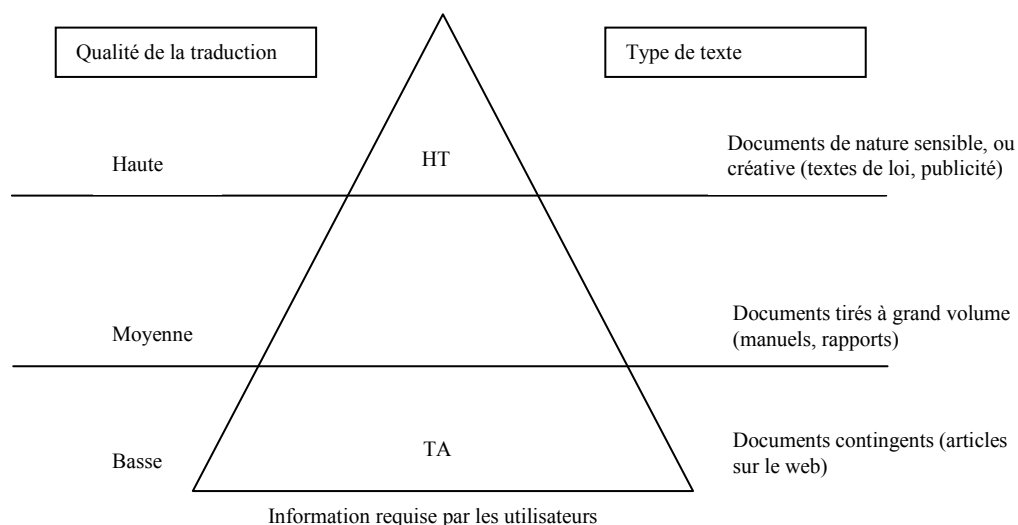


Figure 4 – Modèle pour l'utilisation future des technologies de traduction (Quah, 2006, p. 154 [nous traduisons])

Ainsi, la TA est un outil essentiel au partage de l'information. Cependant, puisque la grande quantité de données (textuelles ou non) à traduire que l'on retrouve au bas de la pyramide ne nécessite pas une traduction « parfaite », les systèmes commerciaux de TA, comme SYSTRAN ou ProMT, ou encore les systèmes gratuits comme GoogleTranslate sont tout à fait adéquats. La TAO (au centre de la pyramide) se prête mieux à la traduction de documents à forte répétition, comme les manuels d'utilisation et les rapports de nature diverse, puisque le traducteur humain s'assure de la qualité des choix de traduction proposés par la machine, tandis que les documents de nature sensible ou créative exigent le savoir-faire humain. En somme, la TEAHQ reste un objet de recherche fondamentale,

particulièrement pour la linguistique computationnelle (à laquelle elle est intimement liée), mais elle n'est plus une finalité.

1.3.2 Traduction machine assistée par l'humain (TMAH)

La traduction machine assistée par l'humain peut à la fois être considérée comme une forme de TA et comme une forme de TAO. Pour les besoins de notre réflexion, nous considérerons, comme Quah (2006, p. 8) et L'Homme (2008, p. 10) notamment, que la TMAH s'apparente plus à la TA qu'à la TAO, puisque c'est la machine qui complète la plus grande part du travail. L'intervention humaine peut advenir en amont (préédition), en aval (postédition) ou au cours du processus (désambiguïsation).

La préédition peut prendre de nombreuses formes. Les systèmes comme TAUM-MÉTÉO (voir chapitre 3), fonctionnant sur des textes de spécialité, à vocabulaire contrôlé, en sont un bon exemple. Ces systèmes sont inutiles sur des textes généraux et exigent que les textes qui leur sont soumis soient traités à l'avance. D'autres systèmes peuvent exiger que l'on modifie à l'avance des phrases non idiomatiques ou que l'on rectifie des erreurs grammaticales ou syntaxiques avant de les passer dans la machine. Ainsi, la préédition visera surtout à assurer la qualité de la sortie machine en vue de minimiser le besoin de postédition, à uniformiser le produit final, en somme, à accélérer l'obtention d'une traduction automatique acceptable pour le lecteur final, quel qu'il soit (Quah, 2006, p. 43-44).

La postédition, à l'opposé du processus, consiste aussi à s'assurer que la traduction fournie par le système de TA atteint un certain standard, lequel peut varier grandement en fonction de la visée du texte (Quah, 2006, p. 11). Horguelin et Pharand, dans *Pratique de la révision* (2009), indiquent que la postédition consiste à réviser la sortie machine d'une traduction automatique, citant ici la définition du Grand Dictionnaire Terminologique

(p. 5). Ce faisant, les auteurs assimilent la postédition à la tâche de réviser une traduction humaine, en vue de s'assurer un produit fini de qualité. Cependant, L'Homme réfute catégoriquement cette analogie, « car les erreurs faites par un logiciel de TA diffèrent radicalement de celles imputables au traducteur » (2008, p. 264). Emma Wagner, ancienne réviseuse à la Commission européenne, précise que les sorties machine contiennent des erreurs qu'aucun humain, fût-il un enfant, ne commettrait et qu'elles ne doivent pas, en conséquence être révisées de la même façon (1985, p. 2). Elle indique que le postéditeur est limité par les choix de la machine, que ce dernier ne peut exercer la même liberté qu'un réviseur : il doit, pour que son travail soit efficace, se limiter à la modification (plutôt que de corriger des erreurs) et oublier la notion de beauté ou d'élégance du texte, car il s'agit ici d'adapter le texte à sa fonction (*ibid.*, p. 2-3). En effet, la postédition est souvent utilisée pour rendre un texte intelligible en vue de seulement informer le lecteur, sans que le texte ne soit destiné à la publication (« gisting »). En conséquence, il est clair à notre avis que la postédition ne peut être pratiquée de la même façon que la révision traditionnelle.

Jeffrey Allen, dans le chapitre sur la postédition de *Computer and Translation : A Translator's Guide* (2003), résume la visée de la postédition et des défis qu'elle pose au traducteur. Il est important, dit-il, de bien mesurer l'impact psychosocial lié aux particularités du métier. L'une de ces difficultés est ce qu'Allen appelle le « red pen syndrome » qui met en jeu la pression de productivité imposée aux réviseurs-traducteurs. En résumé, il s'agit d'une certaine tendance à réviser extensivement les textes (« high quality revising ») qui leur sont soumis, en vue, entre autres, de former les jeunes traducteurs et de, concurremment, fournir des textes de très haute qualité. Cependant, pour que la postédition soit avantageuse, l'objectif à atteindre est d'effectuer le moins de modifications possible pour arriver à un texte « acceptable ». Cette ambition est en

complète opposition avec les pratiques qui ont cours dans le secteur langagier. Il va sans dire que lorsque l'on tente de convaincre ces réviseurs, forts de dizaines d'années d'expérience, de garder au minimum les modifications et de ne pas tenir compte de l'élégance du texte, le processus ne va pas sans accrocs (Allen, 2003, p. 305). Allen offre alors deux scénarios possibles : d'une part une surcorrection, d'autre part une sous-correction (*ibid.*). La méconnaissance de la nature et de la visée de la postédition compte pour beaucoup, à notre avis, dans la méfiance des traducteurs envers la TA et la TAO (particulièrement les mémoires de traduction)⁵.

Enfin, l'intervention humaine au cours du processus de traduction automatique est utilisée dans les systèmes dits « interactifs ». Ces systèmes sont à distinguer des logiciels de TAO puisque que dans le cas de la TA interactive, la machine effectue la traduction automatiquement, et s'arrête sur les segments ambigus pour « demander » à l'humain de clarifier, tandis qu'un logiciel de TAO, incluant ou non une composante TA, laisse totalement les rênes au traducteur⁶ (Quah, 2006, p. 85).

Le modèle de TMAH de Quah résume de façon très concise à la fois la nature de l'intervention humaine qui a lieu, mais aussi le moment où l'interaction humain/machine intervient dans le processus.

⁵ Soulignons que l'on cherche aussi à automatiser, du moins partiellement, la postédition. Si les machines font des erreurs reconnaissables qu'aucun humain ne commettrait, elles sont en outre complètement incapables de les corriger au fil du temps. La réponse naturelle voudrait que l'on corrige l'erreur à la source, mais pour des raisons techniques, ou économiques, il arrive que cela ne soit pas possible. Allen raconte comment le service de traduction de la Commission européenne a, en 1999, déployé un module de postédition automatique pour épargner à ses langagiers une partie de cette tâche fastidieuse, et de ce fait, l'accélérer. Le principe en est simple : on fournit au module un corpus de documents post-édités pour analyse et ce dernier en extrait les constructions fautives les plus fréquentes, qu'il pourra ensuite identifier automatiquement dans les nouveaux documents. La démarche a en plus l'avantage de fournir une base solide sur laquelle concevoir des directives précises pour les postéditeurs (Allen, 2003, p. 313-315).

⁶ Notons au passage que si des systèmes ont été conçus selon ce principe, ils n'ont pas l'heur de plaire aux traducteurs et, au meilleur de notre connaissance, n'ont pas été commercialisés à grande échelle.

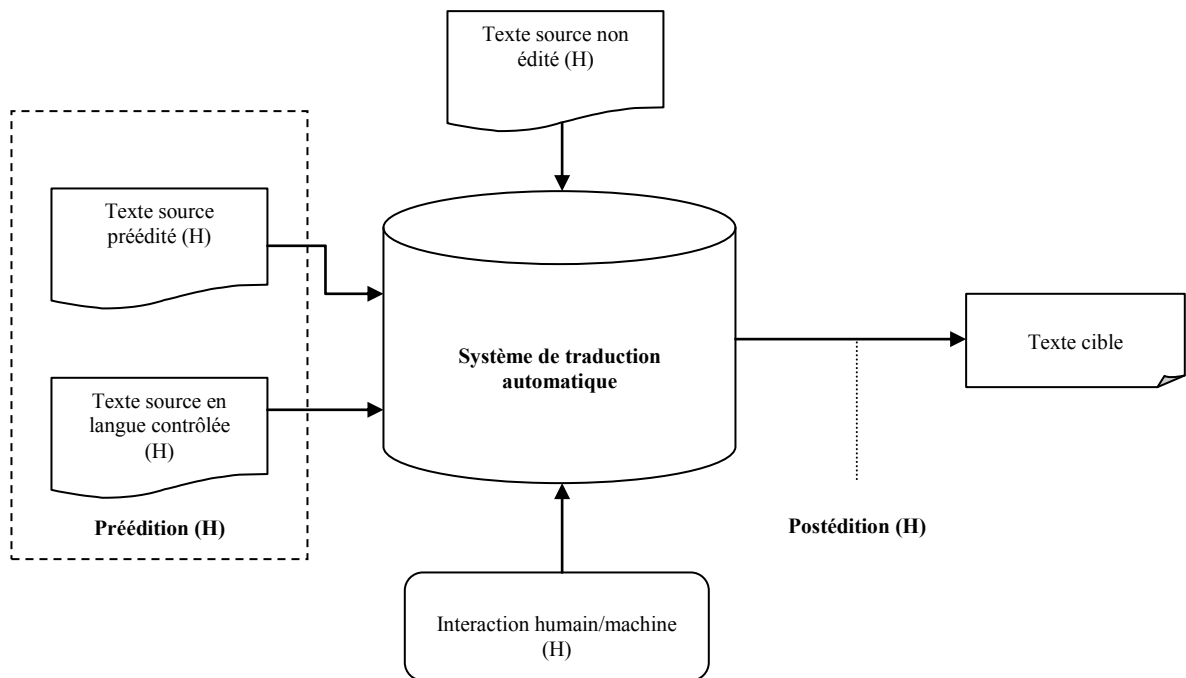


Figure 5 – Modèle de TMAH (Quah, 2006, p. 12 [nous traduisons])

Quah utilise ce modèle pour illustrer l'interaction humain/machine dans le contexte de la TMAH. Cependant, comme nous l'avons remarqué plus haut, très peu de systèmes de TA peuvent se passer du concours de l'humain. C'est pourquoi, dans le cadre de ce mémoire, nous assimilerons la TMAH à la TA. Pour les cas où l'intervention humaine ne serait pas requise, tant en théorie qu'en pratique, nous parlerons de TEA(HQ).

Au milieu du XX^e siècle en Occident, la Deuxième Guerre mondiale est terminée et le climat général est optimiste. La toute-puissance de la science moderne semble inattaquable. Les scientifiques s'emploient à repousser les limites de la connaissance dans des domaines aussi variés que l'aérospatiale, la médecine, l'agriculture, les communications, et bien sûr, l'informatique. Dans ce domaine, les exigences de la guerre ont été le moteur principal du génie humain. À titre d'exemple, les machines servant à décrypter les communications ennemies pullulent : la Bombe, conçue par Alan Turing et Gordon Welchman pour contrer la puissance d'Enigma, la machine à coder allemande, ou le Colossus qui servait à

intercepter les communications allemandes envoyées à l'aide du Lorenz, et dont l'existence n'a été rendue publique qu'au tournant des années 70 (Computer History Museum, 2013).

Puisque les premiers ordinateurs, les cerveaux électroniques comme on les appelait, avaient demandé des investissements faramineux, il n'est pas surprenant de constater qu'après la guerre, on ait voulu continuer à les utiliser. Cependant, les montants en cause sont si imposants que la plupart des premiers ordinateurs sont financés par le secteur militaire. Les tâches sont à l'avenant; l'Electronic Numerical Integrator and Computer (ENIAC), le premier ordinateur non spécialisé, est commissionné par l'armée américaine et sert à calculer des tables d'artillerie. John von Neumann, mathématicien et inventeur de l'architecture des ordinateurs modernes s'en sert pour des calculs essentiels au développement de la bombe à hydrogène. Le Selective Sequence Electronic Calculator d'IBM (IBM SSEC), un hybride capable de modifier lui-même ses instructions et « d'agir » en conséquence, est utilisé pour calculer les tableaux de position de la lune en vue des missions Apollo (Computer History Museum, 2013b).

On retrouve pour la première fois le terme « machine translation » en 1949, dans un mémo de Warren Weaver de la Rockefeller foundation (Melby, 1995, p. 17; Demos et Frauenfelder, 2000, p. 1; Zughoul et Abu-Alshaar, 2005, p. 1024; Rosner, 2005, p. 1; Freigang, 2001, p. 1; Pérez, 2001, p. 2; Hutchins, 1999, p. 1, L'Homme, 2008, p. 12, parmi d'autres). En 1949, donc, Weaver dirigeait la Rockefeller Foundation et était en relation avec de nombreux scientifiques, informaticiens et chercheurs. Dès 1947, il mentionne, dans sa correspondance avec Norbert Wiener, le premier cybernéticien, que les ordinateurs pourraient servir à la traduction de textes rédigés en langues naturelles (par opposition à des codes créés artificiellement). Sa réflexion se poursuit, et en 1949, il rédige le mémo qui jettera les bases de la recherche en TA. De *Translation*, on cite souvent 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. (Weaver, 1955, p. 18)

Cette conception très cryptographique du langage constitue la fondation des systèmes de TA de première génération. Weaver, remarque Melby, considère que les ordinateurs sont très similaires au cerveau humain et qu'en conséquence, il devrait être relativement aisé de les faire traduire (Melby, 1995, p. 18). La solution de Weaver à l'ambiguïté est simplement de donner un peu de contexte au système, c'est-à-dire lui fournir quelques mots avant et après le mot qui pose problème⁷ (*ibid.*). Il n'en reste pas moins que cette conception cryptographique de la TA prévaudra encore pendant de nombreuses années.

1.3.3 Architectures

Qu'ils nécessitent ou non l'intervention d'un humain, les systèmes de TA peuvent aussi être classés selon leur architecture. Les trois principales catégories sont l'approche directe, les approches par règles et enfin les approches par corpus. Ces deux dernières sont, elles aussi, subdivisées en différentes catégories. Hutchins, dans « Machine Translation : A brief History » indique que l'approche directe est caractéristique de la première génération des systèmes de TA, tandis que la seconde génération réfère en général à des systèmes conçus selon l'approche indirecte. La troisième génération est, quant à elle, principalement constituée de systèmes à base de corpus (1995, p. 433).

⁷ Melby donne l'exemple du mot « cut » pour l'ambiguïté : s'agit-il du nom ou du verbe? Selon Weaver, la présence d'un pronom personnel devant indiquerait un verbe, tandis que si l'on trouve un pronom possessif, il s'agirait d'un nom. À cela, Melby répond : « But what about the sentence “her knife cut well but his cut poorly, since it was dull”? [...] Unfortunately, the more one works on such problems, the more complex they appear. » (Melby, 1995, p. 19)

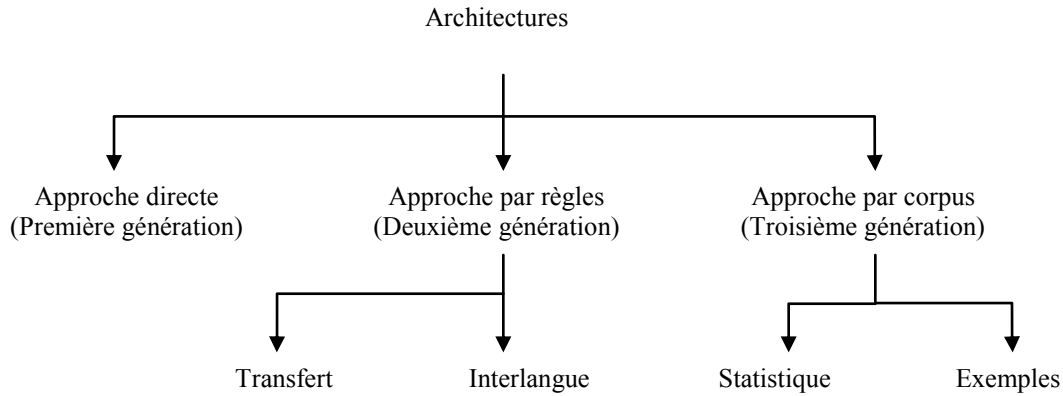


Figure 6 – Architecture des systèmes de TA (Quah, 2006, p. 168 [nous traduisons])

Si les générations se succèdent au fil des ans, on n’abandonne pas obligatoirement l’ancienne approche pour la nouvelle; certains systèmes sont par ailleurs hybrides et conjuguent plusieurs approches. Par souci de clarté, nous les détaillerons séparément.

1.3.3.1 Approche directe

Comme son nom l’indique, l’approche directe, qui se trouve au bas du triangle de Vauquois (voir figure 7), consiste à passer directement d’une langue à l’autre, à l’aide d’un ensemble de règles de correspondance, sans toutefois faire appel à de l’analyse syntaxique ou grammaticale. C’est l’analyse morphologique qui est au cœur du processus, c’est-à-dire que le système analyse les mots, et leurs flexions le cas échéant, dans le texte puis les associe à l’entrée correspondante dans les dictionnaires bilingues (Hutchins et Somers, 1992, p. 72). En effet, les systèmes qui utilisent l’approche directe comprennent, en plus de l’ensemble de règles particulier à la paire de langues traitée, un dictionnaire plus ou moins étoffé. Les systèmes les plus complexes comportent plusieurs dictionnaires juxtaposés, agissant en succession (Wheeler, 1987, p. 114-208). Le résultat est une traduction souvent littérale, qui donne parfois mauvaise presse à la TA, avec ou sans raison⁸.

⁸ 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*

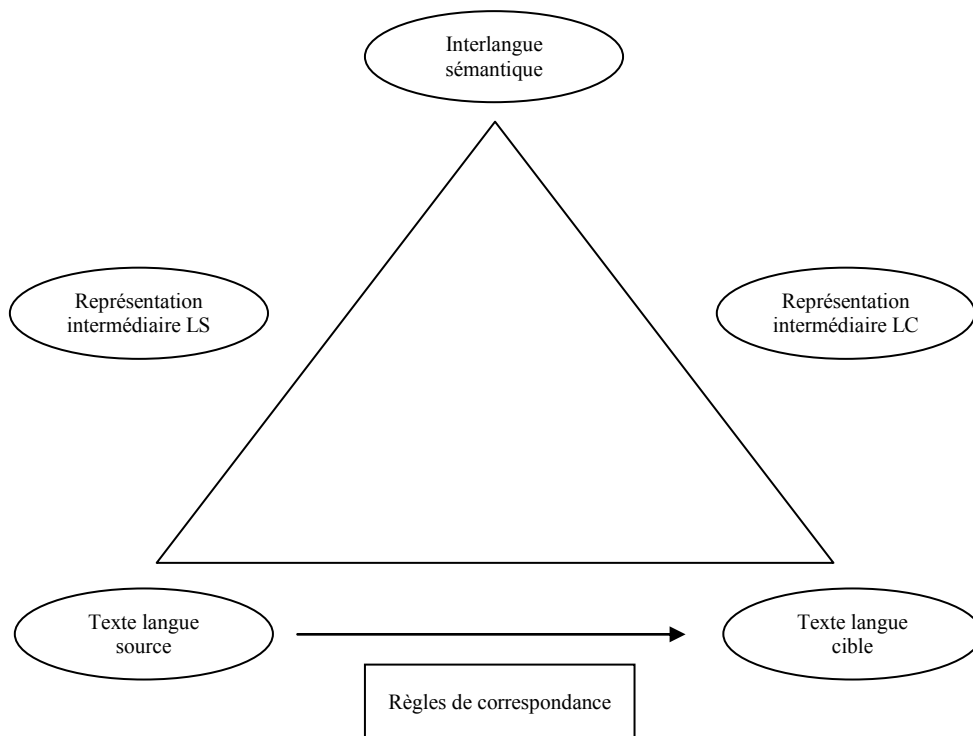


Figure 7 – Triangle de Vauquois, approche directe (adapté de Isabelle, 2010, p. 6)

L’approche directe a une architecture assez linéaire, comme le montre la figure 8. Le texte en langue source (LS) subit une analyse morphologique, dont le résultat est comparé au contenu du dictionnaire bilingue (ou des dictionnaires, selon le cas), pour ensuite passer par un module qui remet les correspondances ainsi trouvées dans l’ordre de la langue d’arrivée (LA).

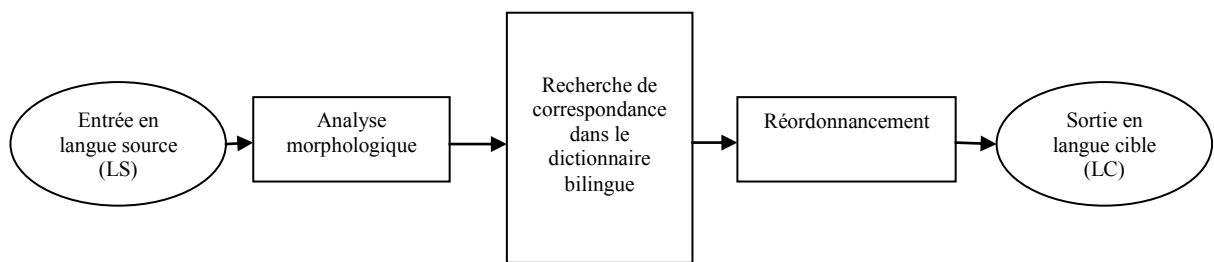


Figure 8 – Système de TA directe (Hutchins et Somers, 1992, p. 72 [nous traduisons])

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” » (Wheeler, 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.

Les systèmes des années 50 présentaient une structure relativement simple. La première tentative réussie de traduction automatique 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) (Hutchins, 1994, p. 15). La réaction est immédiate et euphorique : les scientifiques promettent la TA (c'est-à-dire la TEAHQ) dans un délai de cinq ans⁹. 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 lendemain (*ibid.*). 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. » (*ibid.*, p. 16) Les attentes sont à leur paroxysme. C'est bien là que le bât blesse, et nous le verrons, cet épisode ne cesse de se répéter tout au long de l'histoire de la TA.

Cependant, la démonstration de Georgetown présentait quelques caractéristiques qui continuent encore 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 exemple, dans le cas de Georgetown-IBM, on a rigoureusement sélectionné quarante-neuf 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 personne (Hutchins, 1994, p. 17). Par ailleurs, comme le remarque Melby, les systèmes de traduction automatique doivent toujours être ajustés, ce qu'il compare à

⁹ 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]).

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 traduire (Melby, 1995, p. 20). 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 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 affluent (Melby, 1995, p. 19).

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 cocasses (voir note 8). 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 lorsqu'on s'attaque à un marché comme celui de l'Union européenne, comme le souligne Pierre Isabelle : « pour traduire entre n langues on a besoin de $n*(n-1)$ modules de règles (CE : $23*22 = 506!$) » (Isabelle, 2010, p. 9). 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 il 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 le nombre de ces dernières ne soit démesuré.

La TA allait aussi devoir encaisser un dur coup au cours des années 60, avec la publication du rapport de l'ALPAC en 1966. Ce rapport, aussi connu sous le nom de *Black Book on Machine Translation*¹⁰, allègue, en somme, que la TA est inefficace, inadéquate et trop coûteuse (ALPAC, 1966). L'impact 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.

Le problème, comme nous l'avons proposé plus haut, tient en partie au fait que les promesses des premiers chercheurs, promesses qui s'appuyaient sur des résultats spectaculaires, ont gonflé les attentes des organismes finançant la recherche. En effet, de telles démonstrations, qui parsèment l'histoire de la TA, ont une fâcheuse tendance à exagérer les possibilités d'avancées. Il ne faudrait pas pour autant jeter le blâme sur les chercheurs et les accuser de doper leurs réussites. Les laboratoires sont à la merci de leurs commanditaires, lesquels demandent des résultats concrets. Dans les circonstances, on ne peut se formaliser de leur désir de montrer l'avancement de leurs travaux. Cependant, comme le remarquait Pierre Isabelle au cours d'un entretien (Isabelle, 2012), les démonstrations sont des armes à deux tranchants. Trop d'enthousiasme peut mener à un arrêt de financement, soit parce que le commanditaire voit trop grand et est déçu par les fruits réels de la recherche, soit, au contraire, parce que les résultats sont si probants (toujours de l'avis du bailleur de fonds), que le commanditaire exige une mise en marché

¹⁰ Pour une discussion complète sur le rapport de l'ALPAC, incluant aussi des nuances quant à la condamnation de la TA, voir Hutchins, 1996.

immédiate, même si les chercheurs auraient voulu affiner leur produit¹¹. Dans les deux cas, il semble qu'une certaine incompréhension soit au cœur du problème. À cela, il faut ajouter, comme nous l'avons avancé plus haut, le fait que les démonstrations sont souvent menées dans des conditions de laboratoire extrêmement rigoureuses qui ne peuvent être reproduites en situation normale.

En résumé, la première décennie de la recherche en TA est caractérisée par un enthousiasme débordant, enthousiasme qui s'appuie sur des démonstrations optimistes et des progrès techniques rapides. Le financement est surtout assuré par les gouvernements, qui ont des visées militaires renforcées par la guerre froide. De fait, les recherches se concentrent surtout sur la paire anglais-russe. Il ne faut pas pour autant croire que seuls les États-Unis sont concernés par la TA. On trouve aussi des projets en Europe, en URSS et au Japon. Si le rapport de l'ALPAC a eu un impact important aux États-Unis, la recherche reste néanmoins présente à travers le monde, mais les scientifiques doivent se résigner à abandonner l'idée d'atteindre la TEAHQ sur des textes de nature générale.

1.3.3.2 Approches à base de règles (ou approches indirectes)

Nous sommes donc à la fin des années 60, après la publication du rapport de l'ALPAC. On surnomme la période de 1967 à 1976 « The quiet decade » (Hutchins, 2001, p. 16). Si les militaires financent moins la recherche en TA, les gouvernements européens, canadien et japonais, eux, sont toujours de la partie. Hutchins (2001) et Somers (2003) 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

¹¹ Nous verrons plus tard que la pression peut aussi venir du marché, plutôt que des bailleurs de fonds. Les attentes des clients quant à la fréquence des mises à jour logicielles peuvent elles aussi faire pression sur les développeurs.

SYSTRAN, un système américain et en poursuit le développement (Somers, 2003, p. 5). 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 chercheurs : les succès sont si probants que les universités et le secteur privé décident d'appliquer leurs découvertes à la traduction automatique anglais-japonais (Somers, 2003, p. 5-6). Au début des années 70, à Montréal, le groupe Traduction Automatique Université de Montréal (TAUM) s'emploie à créer un système qui permettra de traduire des bulletins météorologiques (Somers, 2003, p. 5-6; Hutchins, 2001, p. 16). Les États-Unis ne sont pas complètement absents, malgré la douche froide de l'ALPAC : le groupe de recherche de l'Université du Texas planche sur le système METAL, qui traite la paire anglais-allemand (Hutchins, 2001, p. 14).

Ainsi, à la fois pour pallier les lacunes de la première génération et pour faire mentir les détracteurs de la TA, les chercheurs se tournent vers une abstraction plus poussée des langues. C'est ainsi que naissent les systèmes de deuxième génération, qui fonctionnent à base de règles. Il s'agit ici de concevoir des représentations abstraites des langues qui, couplées à des algorithmes de transfert et à des dictionnaires d'équivalence, permettront le passage de la LS à la LC. On peut distinguer deux principales variantes de l'approche à base de règles (ou indirecte) : l'approche par interlangue et l'approche par transfert.

1.3.3.2.1 Approche par interlangue

S'appuyant sur des théories de linguistique structurelle qui émergent à l'époque, les chercheurs s'intéressant à la traduction automatique à base de règles (TABR) s'efforcent d'améliorer leur performance avec une nouvelle approche, par interlangue cette fois. Ils

poussent alors l'abstraction sémantique au maximum et le transfert se fait au niveau d'une interlangue neutre, ainsi que l'illustre la figure suivante.

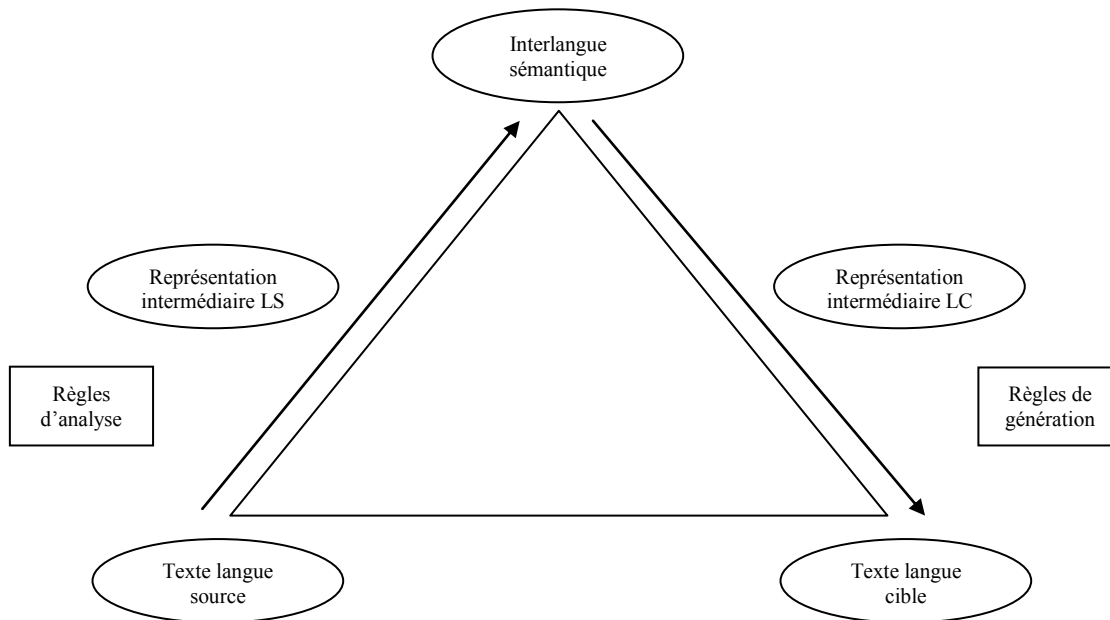


Figure 9 – Triangle de Vauquois, approche par interlangue (adapté de Isabelle, 2010, p. 8)

L'idée consiste à s'affranchir des langues naturelles en concevant une représentation sémantique pure vers laquelle et de laquelle on peut traduire. Ainsi, dans le cas d'un système réversible à deux langues, seulement deux modules par langue seraient nécessaires : un module d'analyse et un de génération :

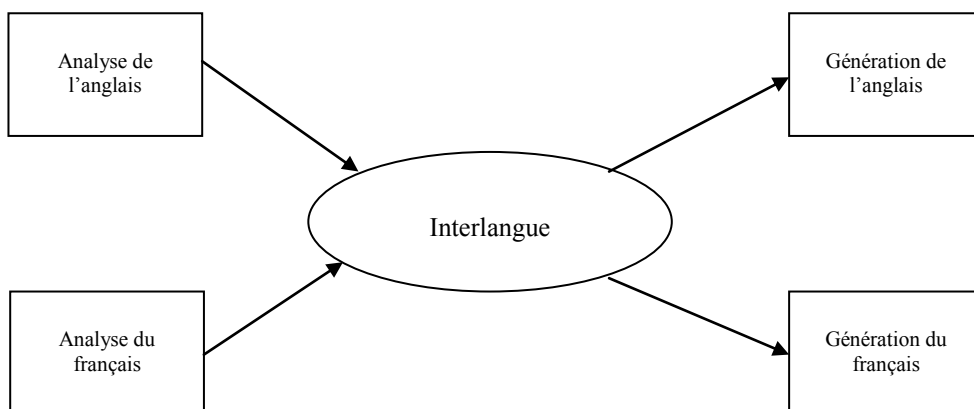


Figure 10 – Modèle de TA par interlangue (Hutchins et Somers, 1992, p. 74 [nous traduisons])

L'approche par interlangue tient plus de la recherche théorique que pratique et intègre des notions et des concepts fortement liés à la linguistique, mais aussi à l'intelligence artificielle. En effet, il s'agit ici de « comprendre » la langue de départ afin de représenter le sens du texte dans une interlangue sémantique qui n'a aucun lien avec les langues naturelles. Si la modularité de ce type de système séduit les chercheurs qui visent la TA multilingue, l'approche par interlangue n'est pas sans désagréments. L'Homme souligne qu'il est difficile de bien définir ce qu'est une interlangue neutre : « quelle représentation sera complètement indépendante des langues traitées tout en retenant toutes leurs propriétés? » (2008, p. 257-258) À ce sujet, Pierre Isabelle remarque à juste titre qu'en pratique, l'interlangue n'est que trop souvent de l'anglais en lettres majuscules et il pose la question cruciale : « Si l'interlangue n'est pas neutre, fait-on mieux qu'une double traduction? » (Isabelle, 2010, p. 18). Hutchins et Somers ajoutent qu'il peut être délicat de représenter en interlangue toutes les nuances de sens nécessaires pour exprimer un (et à plus forte raison, plusieurs) concept dans toutes les langues (1992, p. 123), car, rappelons-le, ce type d'architecture requiert que l'on ne retourne pas vers la langue de départ pour désambiguïser, puisque le module d'analyse doit produire une représentation à partir de laquelle on pourra traduire vers toute autre langue. Ils en donnent pour exemple un système qui inclurait l'espagnol, l'anglais et l'allemand, et pour lequel il faudrait que l'interlangue contienne une distinction entre les pattes d'animaux et les pattes d'une chaise, puisqu'une telle nuance est nécessaire pour l'espagnol, même si elle ne l'est pas pour les deux autres langues (*ibid.*). En d'autres mots, l'approche par interlangue, bien que particulièrement intéressante sur le plan théorique, se bute à des problèmes d'ordre pratique (notamment la quantité d'information qu'il faut fournir au système pour lui faire comprendre le texte) qui se sont vite avérés insurmontables. Ainsi, l'approche par interlangue n'a bien souvent

donné naissance qu'à des systèmes jouets qui n'ont que peu d'applications pratiques et en conséquence n'a jamais vraiment percé le marché des systèmes commerciaux (Quah, 2006, p. 73; L'Homme, 2008, p. 258 et Isabelle, 2010, p. 18). En somme, l'approche par interlangue a souffert du même type de pression que l'approche directe avant elle : les attentes ont été gonflées par des suppositions optimistes, non pas à propos de la nature de la traduction, mais en regard de la linguistique théorique et du concept des universaux du langage (Quah, 2006, p. 73).

Cela dit, la recherche fondamentale en TA par interlangue a tout de même porté ses fruits. Pour remédier aux failles inhérentes de cette approche, les scientifiques ont principalement emprunté deux avenues. Les uns se tournent vers l'intelligence artificielle pour fournir au système les connaissances non linguistiques et les formalismes qui lui permettront de « comprendre » le texte à traduire et de « raisonner » en vue d'atteindre l'interlangue sémantique¹² (Isabelle, 2010, p. 16-17). Les autres se sont tournés vers l'autre variante de la TABR : l'approche par transfert (Hutchins et Somers, 1992, p. 8).

1.3.3.2.2 Approche par transfert

L'approche par transfert requiert un niveau de formalisation des langues naturelles plus poussé que pour l'approche directe, mais moindre que pour l'approche par interlangue. En outre, dans les systèmes de première génération les règles de transfert étaient, la plupart du temps directement programmées dans l'ordinateur, en langage d'assemblage¹³, avec pour conséquence que les algorithmes et les dictionnaires ne pouvaient être facilement modifiés (Isabelle, 1987, p. 249). Avec cette variante de la deuxième génération, les chercheurs ont

¹² Par exemple, le système KANT, et son successeur, KANTOO. Pour une description de son architecture et de son fonctionnement, voir Nyberg III et Mitamura, 1992; et Mitamura, Nyberg et Carbonell, 1994, entre autres.

¹³ Langage de programmation qui sert à communiquer directement avec l'ordinateur : « An assembler is a translator that translates source instructions (in symbolic language) into target instructions (in machine language), on a one to one basis. » (Salomon, 1993, p. 1)

voulu s'affranchir de la machine et ont conçu des langages faisant interface avec le système. Ces métalangages permettent d'écrire des grammaires qui sont beaucoup plus « compréhensibles » pour les linguistes et les traducteurs (*ibid.*), lesquels sont déjà bien implantés dans les équipes de recherche en TA. Les systèmes de deuxième génération sont pour cette raison à la fois plus stables et plus maniables.

En outre, après le choc de 1966, le gouvernement américain, qui finançait une grande partie de la recherche en TA au cours des années 50 et 60, avait retiré son soutien aux chercheurs, comme nous l'avons souligné. Cependant, les besoins de traduction, qui au départ devaient beaucoup aux exigences de la guerre froide (donc du russe vers l'anglais), ne vont pas en diminuant, au contraire. Les marchés commencent à s'ouvrir et le secteur privé, en plus des institutions gouvernementales, exige à son tour des traductions rapides et de bonne qualité pour commercialiser ses produits. Des entreprises comme Siemens, Carterpillar, ou Xerox entrent en scène. Les équipes de recherche doivent donc se repositionner et abandonner l'idée de produire des traductions parfaites¹⁴. L'approche par transfert, les formalismes linguistiques, les systèmes à sous-domaine et les vocabulaires contrôlés sont des réponses à ces nouvelles exigences.

Comme la figure 11 l'illustre, l'approche par transfert, qui a aussi une architecture en partie modulaire puisque les règles d'analyse et de génération sont associées à des langues indépendantes, propose une abstraction sémantique moins profonde que l'approche par

¹⁴ En fait, aussi tôt que 1951 (c'est-à-dire à peine une dizaine d'années après le début de la recherche sérieuse sur la TA), l'Israélien Yehoshua Bar-Hillel, premier à être embauché expressément comme chercheur en TA, met en doute la possibilité d'un jour atteindre ce qu'il appelle la « pure MT » : « It seems obvious that fully automatic MT, i.e. one without human intervention between putting the foreign text into the reading organ of the mechanical translator and reading off its output, is achievable only at the price of inaccuracy, if only for the reason that no method is feasible, for the time being, by which the machine would eliminate semantical ambiguities. » (Bar-Hillel, 1951, p. 229, souligné par l'auteur) Il réitère ses doutes en 1960 dans son article « A Demonstration of the Nonfeasibility of Fully Automatic High Quality Translation », paru dans *Advances in Computers*.

interlangue. Le module de transfert, cependant, est associé à une seule paire de langues et fonctionne dans un sens unique, quoique des tentatives de réversibilité aient été entreprises, avec plus ou moins de succès (nous y reviendrons au chapitre 3). Le système TAUM-MÉTÉO, que nous décrirons plus en détail dans la 3^e partie de notre mémoire, est un bon exemple de TABR de deuxième génération.

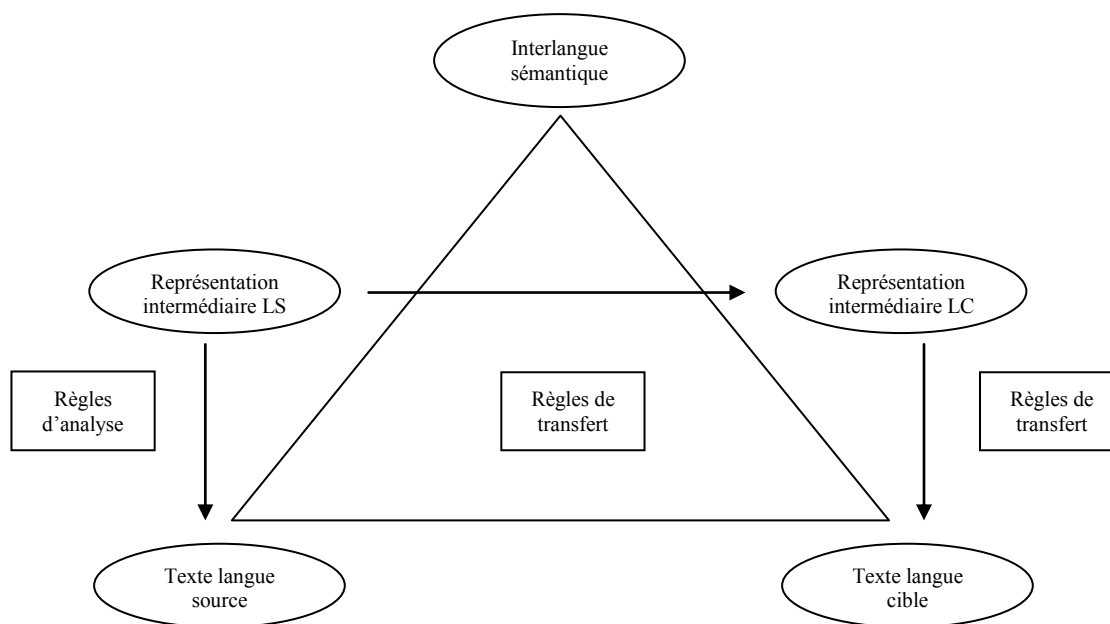


Figure 11 – Triangle de Vauquois, approche par transfert (adapté de Isabelle, 2010, p. 7)

Le système de TABR par transfert commence donc par traiter (à l'aide des règles d'analyse) le texte source en trois étapes (segmentation, analyse morphologique et « parsing¹⁵ ») pour aboutir à une représentation formelle de la phrase, soit un arbre syntaxique à la Chomsky. Cette représentation formelle est exprimée à l'aide d'un métalangage spécialisé comme SYSTEM-Q, SISIF, REZO ou LEXTRA¹⁶. L'arbre syntaxique de la LS passe alors par le module de transfert, qui le transpose en arbre

¹⁵ Ici, nous préférons le néologisme « parsing » à l'équivalent « analyse syntaxique » proposé par le GDT, puisque le processus inclut souvent une analyse sémantique en plus de l'analyse syntaxique. Qui plus est, le GDT accepte « parseur » comme le « Programme de traitement du langage qui, à partir d'un texte en langue naturelle, aboutit à une représentation intermédiaire du sens de ce texte », ce qui rend tout à fait compte de l'opération effectuée ici.

¹⁶ Ces métalangages ont été conçus par TAUM (Isabelle, 1987, p. 247-248).

syntactique équivalent de LC. De là, le module de génération prend la relève et convertit l'arbre syntaxique de LC en un texte de langue naturelle. La figure 12 illustre en détail le processus en fonction des modules du système, plutôt que du niveau d'abstraction comme dans le triangle de Vauquois de la figure 11.

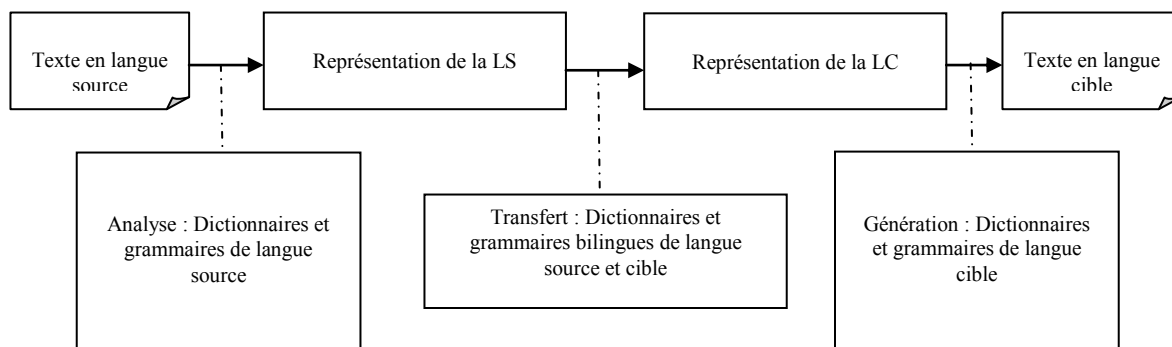


Figure 12 – Modèle de système utilisant l'approche par transfert (Quah, 2006, p. 74 [nous traduisons])

L'approche par transfert a elle aussi ses limites. 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 le garantir. En outre, les modules d'analyse ont du mal à traiter des ambiguïtés grammaticales qui ne ralentiraient pas un traducteur humain. L'exemple par excellence (Isabelle, 2010, p. 14, notamment) pour illustrer les faiblesses de cette approche est « Time flies like an arrow », qui, pour une machine, est très difficile à analyser hors contexte.

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, « The 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 structurelle et de l'informatique. Par ailleurs, les 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. Cependant, il ne faut pas en conclure que seule l'approche par interlangue a la préférence des chercheurs. SYSTRAN, qui utilise l'approche directe, continue d'être perfectionné (Hutchins, 2001, p. 17), et TAUM-MÉTÉO est l'un des premiers à utiliser l'approche par transfert.

Au cours des années 80, c'est l'approche indirecte par transfert qui règne sans conteste sur le paysage de la TA¹⁷ (Hutchins, 1993, p. 11). En effet, les approches par interlangue ne semblent pas remporter les succès escomptés, comme nous l'avons vu à la section 1.3.3.2.1, et l'on préfère se pencher sur des approches moins ambitieuses (Hutchins, 2001, p. 17), 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.

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

¹⁷ ARIANE (Université de Grenoble), METAL (Université du Texas), SUSY (Université de Saarbrücken), Mu (Université de Kyoto) et EUROTRA (Commission européenne) en sont les exemples typiques.

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 TAO chez les traducteurs¹⁸ (Hutchins, 1998, p. 15).

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 » (1993, p. 11). 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 traduction (*ibid.*). 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'implanter¹⁹. Qui plus est, les premiers systèmes commerciaux 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 acceptables (Hutchins, 1990, p. 277-278). Ces conditions sont essentielles au changement de paradigme qui nous mènera à la TA contemporaine.

¹⁸ L'adoption des outils de TAO s'est faite successivement, d'abord chez les traducteurs techniques, pour la plupart employés par les grandes entreprises, puis par les traducteurs indépendants, plus généralistes. Les outils de terminologie et de recherche sont les premiers à être implantés, puis les systèmes plus complexes.

¹⁹ Par exemple, la terminologie informatisée a le vent dans les voiles. Il n'y a qu'à penser à LEXIS en Allemagne, qui ne cesse de croître depuis 1965, ou EURODICOTAUM (CE) et TERMIUM (au Canada) qui sont devenus des références incontournables. Le privé n'est pas en reste, comme le démontre le succès de la base de données terminologique TEAM, mise en place par Siemens (Hutchins, 1990, p. 276).

1.3.3.3 Traduction automatique à base de corpus

La traduction automatique à base de corpus, bien que catégorisée comme la troisième génération de système de TA, a fait ses débuts dans les années 60 (Quah, 2006, p. 76; Hutchins, 1994a, p. 4-4). Cependant, à l'époque, les ressources informatiques ne pouvaient tout simplement pas digérer la quantité de données nécessaire au succès de cette approche. Au tournant des années 90, toutefois, les approches à base de corpus sont résolument à l'avant-scène.

Il existe deux variantes de la TA à base de corpus : la traduction automatique à base d'exemple (TABE) et la traduction automatique statistique (TAS). Ces approches s'appuient toutes deux sur des corpus bilingues (voire multilingues) alignés pour produire une traduction, mais utilisent une méthodologie tout à fait différente pour ce faire. Les deux variantes ont cependant en commun de ne pas avoir recours à des règles linguistiques lors de l'analyse ni lors du choix des équivalents en langue cible (LC) (Quah, 2006, p. 77).

Les approches par corpus sont basées sur des traductions humaines de grande qualité et compilées, le plus souvent, par des organismes gouvernementaux. Le Hansard canadien est le premier à avoir servi pour ce type de recherche (chez IBM, pour Candide), mais le gouvernement japonais produit aussi un corpus parallèle important, tout comme l'Union européenne, et désormais, Internet offre une nouvelle source quasi inépuisable de corpus bilingues ou multilingues pour les chercheurs (Hutchins, 1999, p. 31).

1.3.3.3.1 Traduction automatique à base d'exemple

L'approche à base d'exemple consiste à produire une traduction en se basant sur un corpus de traductions existantes pour en tirer des analogies, qui serviront ensuite à générer le texte cible. Selon Makoto Nagao de l'Université de Kyoto, premier à proposer la TABE, le traducteur n'a pas recours à l'analyse linguistique profonde quand il traduit, il fait plutôt

appel à un principe que Nagao appelle l’analogie (Nagao, 1984, p. 175-176). Ce processus, toujours selon Nagao, consiste à d’abord décomposer la phrase à traduire en segments, ensuite à transférer ces derniers en segments correspondants dans la langue cible, puis à recomposer la phrase complète en langue cible (*ibid.*). La TABE a donc comme visée de reproduire ce processus. La figure 13 illustre la structure de ce type de système.

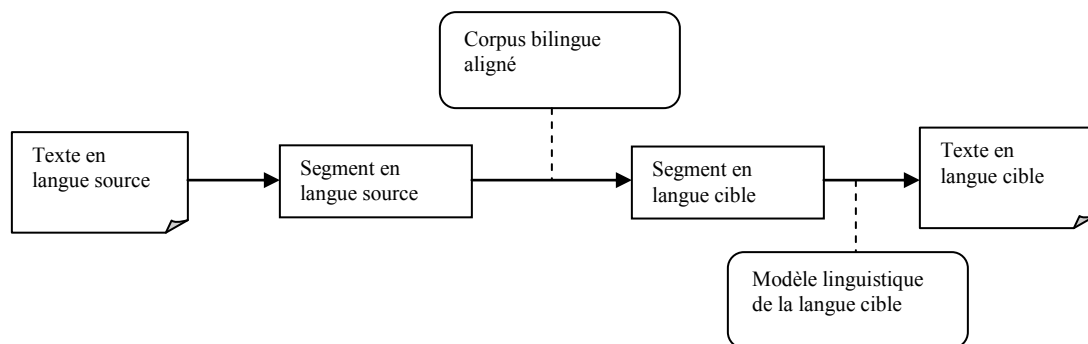


Figure 13 – Modèle de TABE (Quah, 2006, p. 81 [nous traduisons])

Ainsi, une fois le texte en langue source décomposé en plusieurs segments, le système compare ces derniers aux segments présents dans le corpus bilingue, puis les en extrait pour générer le texte cible en les combinant en fonction du modèle linguistique de la langue cible, modèle produit à l’aide de méthodes sémantique ou statistique, selon les groupes de recherche (Hutchins, 2001, p. 23-24). Souvent, plus d’un segment est nécessaire pour générer une traduction et les systèmes de TABE sont conçus de façon à permettre une telle combinaison. Cette dernière caractéristique les différencie des mémoires de traduction (dont nous discuterons dans la section sur la TAO) qui ne peuvent effectuer de telles combinaisons et se limitent à rappeler les segments déjà traduits (Quah, 2006, p. 83, Kit, Pan et Webster, 2002, p. 60).

Somers résume les principaux avantages qu’offre la TABE, après dix ans de recherches. Tout d’abord, les exemples qui constituent la base de données de ces systèmes sont tirés de données linguistiques réelles et permettent en conséquence de produire des constructions

qui apparaissent naturellement dans le discours. Ensuite, il est aisé d'ajouter aux connaissances linguistiques du système : il suffit d'ajouter des exemples. Par ailleurs, la nature empirique de l'approche, par opposition à celle, théorique, de la TABR, évite aux chercheurs d'avoir à créer des grammaires complexes et de se voir confrontés aux casse-têtes qui émergent lorsque deux ou plusieurs règles sont en conflit. Enfin, la TABE pourrait être une façon de s'affranchir du joug de la traduction littérale²⁰, encore trop présent en TA (Somers, 1998, p. 9).

Cependant, Hutchins remarque que la TABE est le plus souvent utilisée en conjonction avec d'autres approches traditionnelles à base de règles (notamment les systèmes ATR et SHALT) et qu'en conséquence, certains observateurs préfèrent s'abstenir de commenter quant à la validité et à l'efficacité réelle de cette méthode, à tout le moins tant que cette dernière ne fait pas cavalier seul (Hutchins, 1994a, p. 4-5).

1.3.3.2 Traduction automatique statistique

La traduction automatique statistique, comme son nom l'indique, se base sur des principes statistiques pour construire un modèle pour la traduction. Cette approche repose sur le théorème de Bayes et s'affranchit complètement des principes de la linguistique computationnelle. On associe souvent à la TAS l'affirmation-choc de Fred Jelinek, chef du groupe de recherche d'IBM sur la reconnaissance vocale, citée par Robert Mercer à l'occasion de la Fourth International Conference on Theoretical and Methodological Issues in

²⁰ Somers indique, concernant ce dernier avantage que « The example-based approach seems to offer some relief from the constraints of "structure-preserving" translation » (1998, p. 9). Pour expliciter ce qu'il entend par « structure-preserving », nous renvoyons le lecteur à un autre article de Somers, où il explique que, traditionnellement, en ce qui concerne la TA, la traduction d'un texte dans son ensemble est considérée comme un assemblage de la traduction de ses parties. Il oppose la vision des chercheurs en TA (« Let's produce translations that are as literal as we can get away with ») avec celles des traducteurs (« structure preserving translation as last resort ») (Somers, 1990, p. 2).

Machine Translation : « Each time I fire a linguist, my performance goes up » (Whitecomb, 1992, p. 1).

Quah explique que le théorème de Bayes permet de combiner des traductions existantes avec de nouvelles données (le texte source) pour prédire la traduction la plus probable (Quah, 2006, p. 77). Comme la TABE, la TAS est une méthode empirique qui s'appuie sur un ensemble de données organisées, le corpus bilingue ou multilingue aligné. Il s'agit, selon Hutchins, d'aligner des phrases, des groupes de mots et des mots individuels de deux textes parallèles (un texte et sa traduction) 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 (Hutchins, 2001, p. 23).

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 » (Quah, 2006, p. 78, Kuhn, 2010, p. 15).

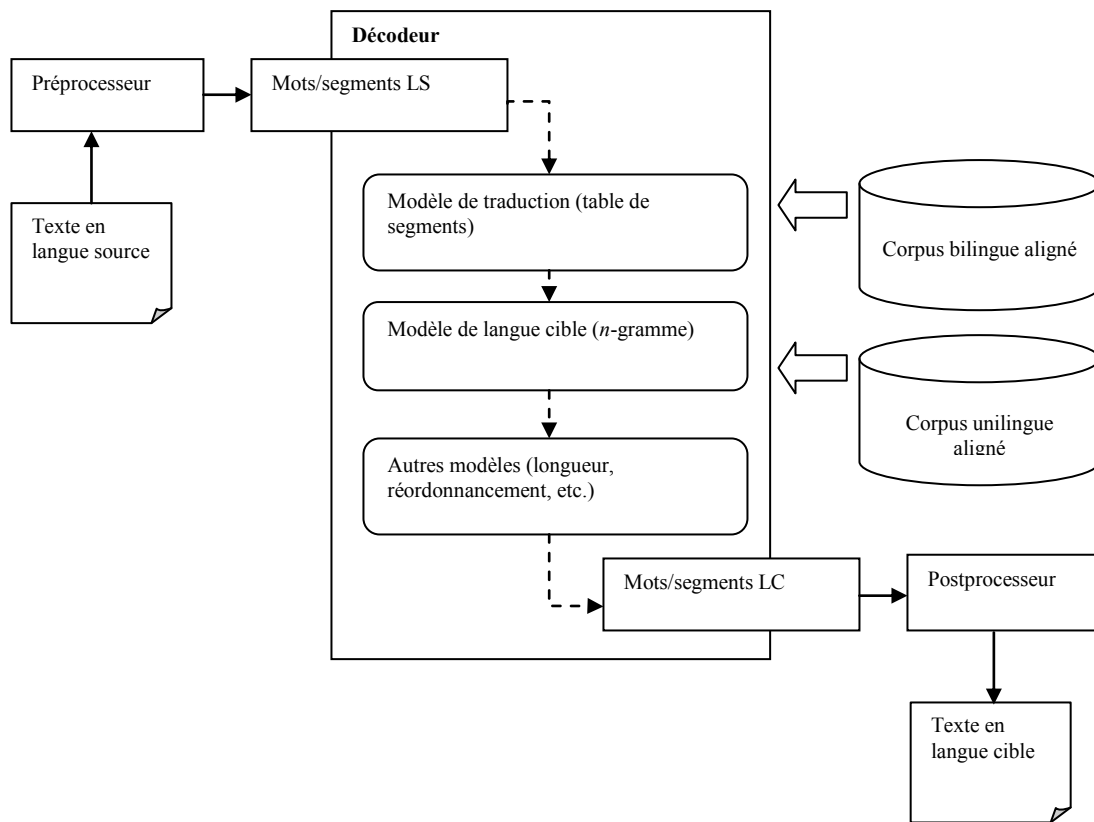


Figure 14 – Modèle de TAS (adapté de Quah, 2006, p. 78 [nous traduisons] et Kuhn, 2010, 23-24)

La figure 14 montre l'architecture d'un système de TAS, qu'il soit à base de mots ou à base de phrases (segments). Le cœur de cette structure est le décodeur, c'est lui qui accomplit toutes les tâches. Le corpus bilingue aligné n'entre pas dans le système en soi, mais il fournit les données en nombre suffisant pour en extraire les différents modèles.

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 linguistique (Kuhn, 2010, p. 11). Les chercheurs en TAS l'ont plus tard adapté à leurs besoins (nous y reviendrons en détail dans la section sur l'historique), car Jelinek s'était plutôt concentré sur les trigrammes (séquence de trois mots) (*ibid.*). 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 langue cible (*ibid.*). 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 traduction (*ibid.*). Les autres modèles ne sont pas toujours présents, mais si besoin est, et en fonction des paires de langues à traduire, ils sont eux aussi constitués de façon statistique.

Ainsi, le texte à traduire passe d'abord dans le préprocesseur qui en régularise la casse et la ponctuation. De là, le texte est envoyé au décodeur, qui successivement, découpe les phrases en segments, retrouve ces derniers dans la table des segments, collige les segments correspondants en langue cible pour former des hypothèses et les compare aux différents modèles de façon à pouvoir leur donner un score de probabilité. Le décodeur sélectionne ensuite les hypothèses qui présentent le plus haut score pour former le texte en langue cible. Enfin, le postprocesseur restaure la casse et la ponctuation et recrache la sortie machine (Kuhn, 2010, 23-24).

La TAS offre de nombreux avantages, mais a aussi certains inconvénients. D'une part, comme pour la TABE, il est facile d'entraîner les systèmes : il suffit d'ajouter 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. Les registres de langue restent un mystère pour les systèmes de TAS. Plus grave, puisqu’ils n’ont recours à aucune grammaire, ils commettent parfois des fautes d’accord qu’un système à base de règle n’aurait jamais perpétrées (Kuhn, 2010, p. 25).

La figure suivante, 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’offre aujourd’hui 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.

Figure 15 – Tableau comparatif de traductions fournies par Google Translate à trois ans d’intervalle (données 2010 dans Kuhn, 2010, 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] » (Kuhn, 2010, p. 26). 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 revirement 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.

1.4 Traduction humaine assistée par ordinateur (THAO) et traduction assistée par ordinateur (TAO)

La figure 3 (à la section 1.2 de ce mémoire) donne une idée de la diversité des outils informatiques à la disposition du traducteur pour l'assister dans sa tâche et nous ne prétendons pas nous attarder sur chacun d'entre eux. Bowker (2002), L'Homme (2008), Quah (2006), Austermühl (2001) auxquels nous avons fait référence décrivent avec nombre de détails ces outils, leurs caractéristiques et leurs modalités d'utilisation. Bert Esselink, dans *A Practical Guide to Localization* (2000) propose aussi une revue complète des outils qui peuvent non seulement servir à la localisation, mais aussi à la traduction plus traditionnelle. Enfin, Jost Zetzsche offre, avec *The Translator's Tool Box : A Computer Primer for Translator v. 10.0* (2012), un manuel pratique pour les traducteurs qui voudraient affiner leur maîtrise de l'informatique appliquée à la traduction. Ce dernier va jusqu'à consacrer un

chapitre aux systèmes d'exploitation (Windows, particulièrement), un autre aux suites bureautiques (comme Microsoft Office et Libre Office), ce que l'on trouve rarement dans les ouvrages sur la TAO.

Par ailleurs, le monde de la TAO est en quelque sorte plus complexe que celui de la TA, moins bien délimité. Il n'existe pas de filiation droite, où une architecture ou un paradigme particulier aurait directement évolué du précédent. Qui plus est, le monde des TAO est en constante mutation, ce qui a pour conséquence, notamment, que les ouvrages consacrés à ce sujet sont parfois dépassés quelques mois à peine après leur publication. À preuve, *The Translator's Tool Box* propose une nouvelle version presque à chaque année (un peu à la façon des logiciels, d'ailleurs). Vouloir en traiter de façon exhaustive dans ce mémoire serait présomptueux et requerrait des ressources que nous ne pouvons y consacrer dans le cadre de ce travail. C'est pourquoi nous nous contenterons de résumer certains concepts qui nous semblent importants en regard de la recherche en TA pour la visée de ce mémoire.

Sur le continuum des technologies langagières, nous l'avons montré, se trouve à une extrémité la TA, qui, dans sa forme idéale, ne fait pas intervenir l'humain, et de l'autre côté, la traduction humaine traditionnelle. Marie-Claude L'Homme indique qu'en ce qui concerne la THAO, « L'initiative appartient à l'humain, qui recourt à la machine pour accélérer des tâches annexes. » (L'Homme, 2008, p. 10) Bowker, quant à elle, offre la définition suivante de la TAO :

In CAT, human translators are responsible for doing the translation, but they may make use of a variety of computerized tools to help them complete this task and increase their productivity. Therefore, whereas MT systems try to replace translators, CAT tools support translators by helping them to work more efficiently. (Bowker, 2002, p. 4)

Par ailleurs, dans son *Dictionary of Translation Technology*, Chan Sin-wai distingue la TAO et la THAO (même s'il indique dans la foulée que les deux termes peuvent être

synonymes). Selon lui, la THAO « refers to a type of human translation with limited assistance from the machine. It does not remove from the translator the burden of actually performing the translation » (Sin-wai, 2004, p. 38) tandis que la TAO « refers to the use of a number of computerized tools by human translators to increase their productivity. » (*ibid.*) Ainsi, peut-on penser, à la lumière de ces définitions, que la THAO désigne plutôt le concept général relatif à l'utilisation de la technologie en traduction tandis que la TAO désigne la pratique de la traduction à l'aide d'outils informatiques de toute nature. Cependant, l'usage est encore loin d'être fixé. Jost Zetzsche remarque à ce sujet que « we [the translators] often use "CAT" as a synonym for so-called "translation memory tools", when the latter is really only a sub-category of the former » (2012, p. 187). Ce flou terminologique peut en partie être expliqué par le fait que, comme l'observe Debbie Folaron dans l'entrée sur les outils de traduction (Translation Tools) du *Handbook of Translation Studies* (2011),

CAT tools initially permeating the market in the mid-1990s often did not coincide with the educational or experiential points of reference and profile of many translators. Rather, they reflected the research domains that had conceptualized them: computational linguistics and machine translation. (p. 4)

En effet, banques de données terminologiques, mémoires de traduction, bitextes, concordanciers, et aligneurs, ces outils ont d'abord été conçus dans les laboratoires de TA²¹ (grâce à des concepts comme la segmentation, l'alignement et la création de corpus bilingues) en vue d'automatiser entièrement la traduction, puis adoptés, plus ou moins rapidement par les traducteurs professionnels²². Ce sont donc sur ces derniers que nous nous concentrerons dans cette section.

²¹ Ceci dit, si les laboratoires de TA ont bel et bien conçu ces outils, ce sont souvent les entreprises qui doivent produire beaucoup de traduction à l'interne qui développeront des applications intégrées pour faciliter la tâche à leurs employés. Ces logiciels seront ensuite mis en marché, dans certains cas.

²² Surtout du domaine technique. Ailleurs, l'adoption s'est faite plutôt sur le tard.

Parmi les premières ressources technologiques à être utilisées par les traducteurs, nous retrouvons en bonne place les banques de données terminologiques, les glossaires et les dictionnaires (Hutchins, 1998, p. 290). Bien que, d'emblée, les dictionnaires bilingues conçus pour les systèmes de TA ne conviennent pas du tout aux traducteurs²³, le German Army Translation Service, par exemple, alors sous la direction de Friedrich Krollman, s'est penché sur la conception de bases de données linguistiques qui pourraient non seulement servir aux traducteurs, mais aussi aux lexicographes responsables de produire des glossaires pour la TA, l'accès en ligne et l'archivage de traductions (*ibid.*). Selon L'Homme (2008, p. 137-138), les principaux avantages d'informatiser ainsi les données terminologiques et lexicographiques sont :

- Pouvoir tirer parti des recherches précédentes et de celles effectuées par ses collègues;
- Saisie accélérée des réponses trouvées (si l'outil terminologique est compatible avec l'éditeur de texte, un simple copier-coller suffit);
- Uniformiser facilement la terminologie pour un type de texte, un client, selon les préférences personnelles, etc.

Les bases de données terminologiques créées à l'époque de Krollman, comme EURODICAUTOM (Union européenne), TEAM (chez Siemens) ou TERMIUM (Bureau de la traduction) sont toujours en usage et parfois mises à la disposition du public. Il est aussi possible de créer ses propres bases de données en fonction de ses besoins et de programmer une composante de recherche dans celles-ci. Les systèmes de gestion de terminologie tout faits comprennent un outil pour structurer les données selon les besoins de l'utilisateur, une base de données (communément appelée une base de termes) et une

²³ En effet, les dictionnaires qu'utilisent les systèmes de TA présentent des informations grammaticales, sémantiques et syntaxiques détaillées qui ne sont d'aucune utilité aux traducteurs. Qui plus est, ces dictionnaires sont composés de termes de la langue courante, tandis que les traducteurs doivent plutôt avoir accès à du vocabulaire technique et spécialisé (Hutchins, 1998, p. 290).

fonction de recherche plus ou moins avancée, selon le cas (Quah, 2006, p. 105). De plus, suite à la diversification des produits offerts aux traducteurs, un format d'échange de données²⁴ a été mis en place pour faciliter le partage entre collègues ou d'un donneur d'ouvrage à un client, par exemple.

La mémoire de traduction est un outil qui permet de rappeler des traductions déjà enregistrées dans une base de données. L'Expert Advisory Group on Language Engineering Standards (EAGLES) définit la mémoire de traduction comme « a system which scans a source text and tries to match strings (a sentence or part of thereof) against a database of paired source and target language strings with the aim of reusing previously translated materials. » (EAGLES, 1996, p. 139) Les mémoires de traduction sont particulièrement utiles dans le cas de textes à très haute répétition, comme les rapports annuels ou les manuels techniques. Contrairement aux systèmes de traduction automatique (particulièrement de type TABE, avec lesquels elles partagent certaines caractéristiques comme nous l'avons mentionné précédemment), les mémoires de traduction n'incluent aucune analyse linguistique, la recherche dans la base de donnée se faisant en comparant les caractères (lettres, ponctuation, espaces). Le système peut rappeler deux types de correspondances : exactes et floues (Austermühl, 2001, p. 136). Les correspondances exactes sont remplacées telles quelles dans le texte cible, mais doivent tout de même être acceptées (donc vérifiées) par le traducteur. Les correspondances floues (« fuzzy matches ») sont des segments présents dans la base de données, mais qui ne correspondent pas totalement au segment à traduire. En général, le traducteur règle lui-même le pourcentage de correspondance qu'il souhaite se voir proposer (rarement en deçà de 50 %).

²⁴ Le format TBX, ou TermBase eXchange (ISO 30042), mis en place par la Localization Industry Standard Association (LISA), démantelée depuis. C'est l'Industry Specification Group (ISG) qui a repris le flambeau.

Il doit ensuite modifier lui-même le segment cible. Les correspondances sont proposées en ordre décroissant (*ibid.*, p. 137). Certaines applications de mémoire de traduction offrent aussi des filtres qui leur permettent de traiter différents formats de textes (pdf, html, txt, doc, docx, etc.), ainsi, le traducteur n'a pas à se préoccuper de la mise en page et peut se concentrer sur la traduction (Quah, 2006, p. 98).

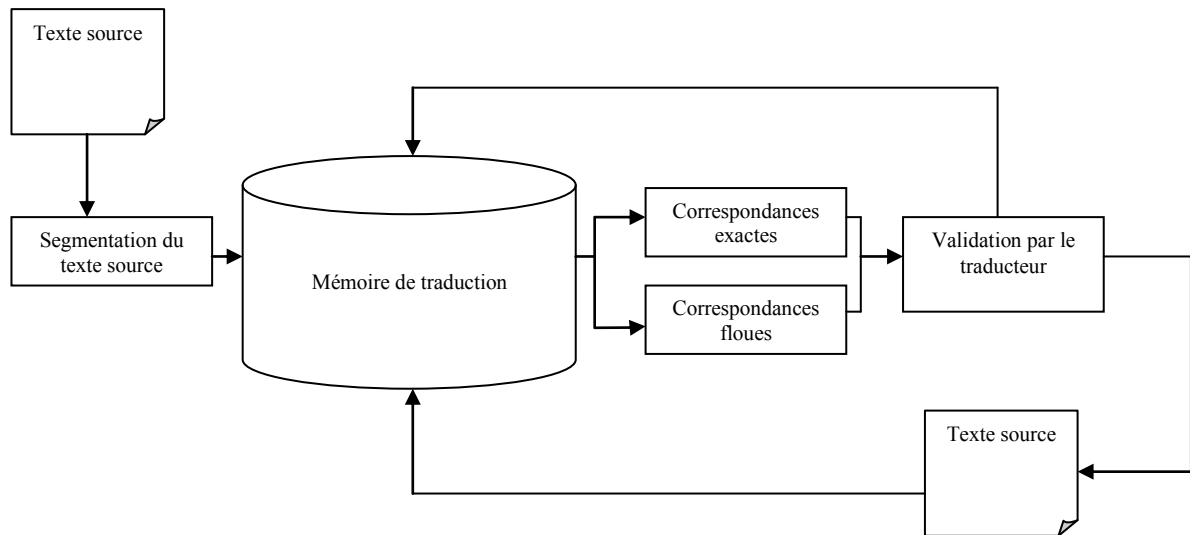


Figure 16 – Fonctionnement d'un logiciel de mémoire de traduction seul

La figure 16 montre en détail toutes les étapes du processus et illustre aussi les modes d'alimentation de la mémoire de traduction. Le texte source est séparé en segments, qui sont comparés un par un à la mémoire de traduction. Le système renvoie les correspondances exactes et floues, qui sont alors validées par le traducteur. Après validation de chaque segment, ces derniers sont enregistrés dans la mémoire de traduction, avec le segment en langue source, pour réutilisation dans le même texte, le cas échéant. Ainsi, la base de données est alimentée en temps réel. Une fois tous les segments validés, le traducteur reconstruit le texte cible, au format originel. Il est aussi possible de verser le texte source et le texte cible dans la mémoire de traduction après coup. Il s'agira alors d'utiliser un aligneur, c'est-à-dire une application qui divise les textes au format numérique

en segments pour ensuite les apparier. Nous y reviendrons. Enfin, comme pour les bases de données terminologiques, il existe aussi un format d'échange standard pour les mémoires de traduction, le Translation Memory eXchange (TMX)²⁵. Ce type d'outil est particulièrement apprécié des traducteurs, puisqu'il leur laisse la décision finale quant à la traduction. Notre propre expérience au Bureau de la traduction nous a d'ailleurs confirmé que les traducteurs apprécient particulièrement utiliser la mémoire de traduction lorsqu'ils ne sont pas forcés à accepter ses propositions²⁶.

Le bitexte est une composante importante des TAO. Le concept a été introduit pour la première fois en 1988 par Brian Harris, professeur à l'Université d'Ottawa. Les spécialistes s'entendent pour dire que cette contribution a été essentielle à l'avènement des mémoires de traduction (Hutchins, 1998, p. 299; L'Homme, 2008, p. 163). En somme, le bitexte de Harris est une structure constituée d'unités de traduction (un segment source et un segment cible) et qui forment un hypertexte dans lequel on peut faire des recherches²⁷. Couplé à l'idée d'Alan Melby, de la Brigham Young University, d'appliquer le concept de concordancier bilingue²⁸ à la traduction, la table est mise pour la TAO telle que nous la connaissons (Hutchins, 1998, p. 300). Cependant, réaliser un bitexte, ou un concordancier,

²⁵ Le TMX a aussi été mis en place par LISA. Suite au démantèlement de l'organisation, le format est maintenant sous licence Creative Commons et peut être consulté ici : <http://www.gala-global.org/oscarStandards/tmx/tmx14b.html>.

²⁶ À ce sujet, voir les travaux de Mathieu LeBlanc, notamment « Translators on Translation Memory (TM). Results of an Ethnographic Study in Three Translation Services and Agencies » (2013), qui traite de la mémoire de traduction dans un contexte canadien.

²⁷ « [...] a bi-text is not two texts, but a single text in two dimensions, each of which is a language. [...] The image of the text in two dimensions refers to the text as a whole. Yet translators do not translate whole texts at one fell swoop. They proceed a little at a time, and as they proceed each spurt, each segment forms a fragment of bi-text in their minds. Bi-text retains this structure when it is recorded on paper or in a computer: that is to say, not only is the whole text a bi-text but each segment combines ST and TT. » (Harris, 1988, p. 1)

²⁸ « A concordancer is an electronic tool which has been used in language learning, literary analysis, corpus linguistics, terminography and lexicography. It allows the user to select a particular word or phrase and display the uses of that word or phrase in the selected corpus in order to show where and how often it occurs, and in what linguistic context. » (Quah, 2006, p. 111; voir aussi L'Homme, 2008, p. 148-162 pour le détail de l'utilisation des concordanciers)

n'est pas aussi simple que l'on pourrait le croire. Les aligneurs sont des outils spécialisés qui permettent la segmentation et l'assemblage des segments sources et cibles. Certains aligneurs se basent sur des repères graphiques (signes de ponctuation, marques de paragraphe) pour la segmentation, mais il arrive souvent que l'ordre des phrases soit changé en cours de traduction, ce qui rend cette méthode peu fiable (L'Homme, 2008, p. 164). D'autres paramètres, comme les noms propres ou les chaînes de caractères numériques peuvent aussi servir de repère, ou encore la longueur relative d'un segment source typique par rapport à son segment cible (*ibid.*, p. 165). Ce sont les chercheurs en TAS, finalement, qui ont permis le développement plus poussé des aligneurs automatiques en utilisant des algorithmes statistiques, et surtout, en ne cessant jamais de les perfectionner (Hutchins, 1998, p. 300).

L'idée d'un poste de travail de traducteur rassemblant une panoplie d'outils pour faciliter la traduction remonte assez loin dans l'histoire des technologies langagières. Martin Kay, chercheur en linguistique informatique, propose en 1980 ce qu'il appelle « The Translator's Amanuensis », un terminal informatique, équipé d'un écran, d'un clavier et d'une souris²⁹ à l'aide duquel le traducteur saisit sa traduction en « écrasant » le texte original. Il peut aussi, à l'aide de son éditeur de texte, remplacer automatiquement certains mots choisis, en rechercher d'autres à l'aide d'un dictionnaire intégré et modifiable, s'assurer automatiquement de l'uniformité de sa terminologie et insérer d'un clic des

²⁹ Il faut préciser qu'au moment où Kay rédige son article, le micro-ordinateur tel qu'on le connaît aujourd'hui n'est pas encore commercialisé. Néanmoins, Kay étant employé chez Xerox, il se peut qu'il ait eu accès au Xerox Alto (<http://www.xtimeline.com/evt/view.aspx?id=292762>, consulté le 29 juillet 2013), l'un des premiers ordinateurs personnels. Il est intéressant de rapporter la description complète qu'il fait du terminal à la disposition du traducteur : « The device I am about to describe, which I call The Translator's Amanuensis does not exist and probably never will. [...] Suppose that the translators are provided with a terminal consisting of a keyboard, a screen, and some way of pointing at individual words and letters. The display on the screen is divided into two windows. » (Kay, 1980, p. 12) C'est, à notre avis, une preuve remarquable de la vision à long terme de Kay, puisque rien ne pouvait garantir que les traducteurs adopteraient cet outil de travail.

traductions que son logiciel pourrait lui suggérer. Une telle application ferait aussi bon usage de sa mémoire et pourrait proposer au traducteur des solutions qu'il a déjà trouvées à des passages similaires ou qui se répéteraient. (Kay, 1980, p. 12-20).

Cependant, Somers (2003) avance que l'idée d'utiliser l'ordinateur (qu'il soit personnel ou central) remonte à avant les années 70 (p. 13). Il remarque avec justesse que le rapport de l'ALPAC, en décourageant la recherche en TA, suggère de tourner les efforts vers le développement d'outils d'aide à la traduction (p. 14). Alan Melby est aussi considéré comme l'un des « pères » de la TAO, comme nous l'avons exposé plus haut, non seulement pour avoir pensé à appliquer le concept de concordancier à la traduction, mais aussi pour le concept de « Translator Workstation ». Au cours des années 80, il a publié nombre d'articles à ce sujet. Le premier que nous avons relevé, « Computer-Assisted Translation Systems : The Standard Design and A Multi-level Design » (1983) décrit un système intégré comprenant un traitement de texte, des applications de soutien pour la terminologie, des aides au prétraitement ainsi qu'un système local de TA. Son système présente trois niveaux d'intervention :

- Le premier « the lowest level of aid to the translator » comprend le traitement de texte, les aides terminologiques, des dictionnaires bilingues et une application qui permet la saisie rapide de termes fréquemment utilisés. Le traducteur saisit lui-même sa traduction dans le système, à partir du texte source en format papier.
- Le deuxième, « an intermediate level at which the source text must be available in machine readable format », offre les mêmes outils que le premier niveau, mais permet aussi de vérifier automatiquement le texte source pour repérer des erreurs d'orthographe ou des termes inhabituels. En outre, le traducteur traduit directement à l'écran, segment par segment, profitant des propositions tirées des dictionnaires bilingues, offertes en temps réel par le système.
- Le troisième niveau « requires a separate machine translation system and interface to it ». Le traducteur reçoit une traduction automatique du texte source, mais qui n'aura peut-être pas été complètement traduite. Le système de TA est paramétré de telle façon que s'il n'arrive pas à analyser, à transférer ou à générer un segment donné, il en garde trace et fournit l'information tant au

traducteur qu'à l'équipe de développement. Le traducteur peut ainsi faire un choix éclairé pour corriger la traduction machine, ou choisir de traduire lui-même les segments à problème et l'équipe de développement peut profiter de données réelles pour améliorer le système de TA. (Melby, 1983, p. 175)

Le poste de travail du traducteur (Translator's Workstation) a subi de nombreuses transformations depuis et a gagné en complexité. Désormais, en plus des outils bureautiques et des aides à la traduction traditionnels, il faut y ajouter les outils nécessaires à la localisation (web, logicielle et de jeux vidéo), à la gestion de projet, à la traduction audiovisuelle ou encore à la publication assistée par ordinateur (« desktop publishing »). Les wikis, les blogues, les communautés virtuelles et les réseaux sociaux sont aussi devenus des ressources importantes pour les traducteurs. En fait, l'histoire de la TAO jusqu'à aujourd'hui mériterait qu'on lui consacre une étude complète, particulièrement en ce qui concerne le contexte canadien.

Chapitre 2 Contexte canadien

Loin d'être autonome et neutre, la technologie est tissée de part en part de liens sociaux, psychologiques, esthétiques, politiques, économiques et avant tout humains, profondément humains.

(Michel Puech, *Homo sapiens technologicus*, 2008, p. 53)

Comme nous l'avons démontré dans le chapitre précédent, de nombreux facteurs influencent la recherche et le développement des technologies langagières. Les politiques gouvernementales, les demandes de l'industrie, les avancées technologiques, les autres domaines de recherche en sont quelques-uns auxquels nous avons fait allusion. Dans le cas de la traduction, les associations professionnelles jouent aussi un rôle important. En outre, les universités, à la fois comme creuset de formation pour les traducteurs (qui utiliseront ou non les technologies langagières) et comme institution de recherche sont aussi des acteurs importants à considérer.

La première impulsion qui pousse la recherche en TA en Occident est, comme nous l'avons énoncé à de nombreuses reprises, celle des objectifs militaires. Ce n'est qu'avec l'ouverture des marchés internationaux, qui débute dans les années 70, pour s'accroître au cours des années 80 et, sans conteste, exploser au tournant des années 90 (au rythme de l'expansion d'Internet) que la recherche en TA reprend son ardeur des débuts. En effet, jusqu'à tout récemment, ce sont les besoins en traduction qui motivaient le financement de la recherche. Par exemple, le comité mandaté par l'ALPAC pour faire rapport sur les technologies langagières dans les années 60 propose deux principales raisons d'octroyer des fonds à la recherche en TA : d'une part, si cette recherche dans un domaine de pointe répond à la mission de l'organisation, d'autre part, si la R et D peut résolument promettre des réductions de coût, une augmentation notable de la performance ou de combler des besoins opérationnels (Hutchins, 1996, p. 2). L'ALPAC ne retiendra que la deuxième

raison, et conclura ses travaux en affirmant que la TA ne répond à pas à ce critère de financement, avec les conséquences que l'on sait (*ibid.*).

Au Canada, pourtant, les vagues causées par le rapport de l'ALPAC se font moins sentir. Le bel enthousiasme et l'optimisme invétéré des premiers chercheurs s'intéressant à la traduction automatique se sont inévitablement refroidis, particulièrement aux États-Unis, après les critiques acerbes du rapport de l'ALPAC. « Quelques mois plus tard [soit après la publication du rapport], des quelque 20 centres recensés par Yehoshua Bar-Hillel en 1958, il n'en restait plus qu'une poignée », remarque John Chandioux (1977, p. 55).

L'expression « traduction automatique » restera elle-même quasiment taboue pour au moins une décennie (L'Homme, 2008, p. 14). Le gouvernement canadien, à l'opposé de la tendance mondiale, lance néanmoins deux projets de traduction automatique, dont l'un, TAUM-MÉTÉO, un système qui utilise l'approche directe, basée sur un sous-langage, aura des répercussions importantes (Macklovitch, 1997; Chandioux, 1977), comme nous l'exposerons plus loin.

Si le projet de Kathleen Booth à l'University of Saskatchewan (un autre système de TA utilisant l'approche directe) ne semble pas avoir eu de suite³⁰, son existence même témoigne de la volonté du gouvernement canadien de soutenir la recherche dans le domaine. Dès 1964, soit deux ans avant la publication du *Black Book*, c'est l'Imprimeur de la Reine (responsable de tous les besoins d'impression du gouvernement) qui examine les promesses de la traduction automatique pour accélérer la production des documents officiels dans les deux langues. En effet, parmi les facteurs qui expliquent l'essor de la

³⁰ Hutchins (1986, section 12.5) y consacre une sous-section dans son ouvrage, tandis que Chandioux rapporte que le projet est « l'héritier direct de la première génération [de systèmes de TA] et rencontre les mêmes difficultés [...] » (1977, p. 55). Le projet sera abandonné après le départ de Booth pour Lakehead University en 1972 (Hutchins, 1986, section 12.5).

recherche en traduction automatique au Canada se trouve en bonne place, comme le remarque Macklovitch, l'appareil législatif qui garantit les droits linguistiques des citoyens. Par exemple, l'adoption en 1969 de la *Loi sur les langues officielles* a pour conséquence directe, selon lui, de décupler les besoins publics de traduction (1997, p. 201). De ce fait, poursuit-il, le gouvernement fédéral se trouvait à l'époque avidement disposé à entendre toute proposition pouvant l'aider à réduire ses dépenses en traduction, « including what must then have appeared as far-fetched computerized systems. » (Macklovitch, 1997, p. 201) Soulignons que c'est vraisemblablement pour cette raison que les contrecoups du rapport de l'ALPAC se sont vus atténués au Canada.

Sans le moindre doute, et la suite des événements l'illustrera, on peut affirmer, à l'instar de Macklovitch, que le volume accru de traduction qui découle directement du bilinguisme officiel au Canada a été un facteur clé, sinon, le facteur clé, expliquant la vivacité de la recherche publique en technologies langagières au pays et que l'intérêt du gouvernement pour l'automatisation de cette tâche est né de l'impératif de réduire les coûts engendrés par la traduction. C'est aussi pour cette raison que les scientifiques canadiens ont dominé le monde des technologies langagières au cours des années 1970 (et par la suite, quoiqu'avec moins de moyens), s'imposant à l'avant-garde de la recherche et développant des prototypes innovateurs et audacieux. Qui plus est, à l'époque, l'industrie canadienne des technologies langagières (comme on la désignera plus tard) n'a ni la vigueur ni l'ampleur de ses homologues américaines ou européennes et les entreprises privées peinent à soutenir la recherche et le développement comme le font Xerox, IBM ou Seimens, pour n'en nommer que quelques-uns.

Dans son article « From Culture to Business: Federal Government Translation in Canada », Brian Mossop, ancien traducteur et réviseur du Bureau de la traduction, affirme

que « The Translation Bureau is of interest precisely because it is very old as translating organizations go. » (p. 3) De plus, comme nous le verrons dans les paragraphes suivants, le gouvernement fédéral en fait le centre de sa politique linguistique, ce qui a pour conséquence d'accroître son influence sur toute l'industrie de la traduction canadienne. L'argument principal de Mossop est que la politique du gouvernement fédéral, en ce qui concerne la traduction, est passée d'une orientation socioculturelle à une préoccupation économique (*ibid.*, p. 25). Cependant, il avance que l'arrivée des technologies au BT n'est pas un marqueur de changement au Canada, puisqu'elles y étaient avant ce passage à une politique économique (*ibid.*, p. 14). Nous voudrions prendre son argument à revers et montrer que c'est le changement d'attitude du BT par rapport à sa mission qui a influencé sa politique quant aux technologies langagières.

Nous avons affirmé que c'est la *Loi sur les langues officielles*, passée en 1969, qui soutient la constante demande en traduction au Canada. Cependant, le gouvernement fédéral doit tout de même composer avec le bilinguisme de sa population depuis longtemps déjà³¹. Jusqu'à 1934, ce sont des fonctionnaires fédéraux disséminés à travers les différents ministères qui traduisent les documents officiels (débat, lois, communications avec le public, etc.). Au nombre de 91 à ce moment, les traducteurs employés par le gouvernement sont alors regroupés par le Secrétariat d'État en un bureau central. Mossop affirme que cette centralisation a eu pour effet de standardiser les pratiques de traduction (*ibid.*, p. 3). Cependant, la décision ne fait pas l'unanimité, particulièrement chez les francophones : Fikri Méléka rapporte en 1977 dans *Meta* qu'il y a trois principaux pôles d'opposition :

³¹ La traduction gouvernementale au Canada remonte à 1867, au Pacte de la Confédération, qui garantit la place du français au Parlement. Les premières traductions, produites vers 1872, le sont par des pigistes. Ce sont les Débats qui sont les premiers objets de la traduction au gouvernement et c'est pour assurer la traduction de ces derniers que l'on embauche les premiers traducteurs titulaires au début du XX^e siècle. (voir Méléka, 1977).

d'une part, les hommes politiques libéraux, qui voient cette initiative des conservateurs comme d'un mauvais œil, d'autre part, les traducteurs³² eux-mêmes, qui craignent de perdre leur autorité et leurs avantages sociaux en quittant les ministères, et enfin, la presse francophone, qui exige des assurances et qui suspecte le gouvernement de vouloir « “parquer” les Canadiens français dans les services de traduction » (Méléka, 1977, p. 58). Ce dissentiment préfigure, à notre avis, les tensions qui surviendront plus tard au cours de l'histoire des technologies langagières. En effet, cet épisode met en scène les pouvoirs politiques, l'industrie (les traducteurs et les associations professionnelles) et la presse (qui sera remplacée par les universités, donc un certain milieu intellectuel). Chacun de ces protagonistes a évidemment à cœur ses propres intérêts et peine à voir objectivement les avantages potentiels³³ d'une telle centralisation.

Les esprits se calment néanmoins et le Bureau des traductions voit le jour en 1934. Le Bureau grandit au fil des ans, la demande pour ses services allant toujours en augmentant. Méléka observe une demande accrue de traduction au cours de la guerre comme en témoigne la création de nouveaux services de traduction notamment chez les Forces armées et aux Approvisionnement et services (*ibid.*, p. 61). La fin de la guerre a aussi une influence et en 1948, le Bureau compte désormais 175 traducteurs. Cependant, c'est réellement de 1968 à 1978 que le Bureau connaît l'une de ses plus fortes expansions : Mossop rapporte qu'en 1968, l'effectif combiné du Bureau se chiffre à 525, pour culminer à 1908 en 1978 (Mossop, 2006, p. 4). Avant l'adoption de la *Loi sur les langues officielles*, le gouvernement était seulement tenu de traduire (vers le français surtout) les débats

³² L'Association technologique de langue française (ATLF), qui regroupe les traducteurs de la région Hull-Ottawa, s'y oppose vivement : « il y a péril en la demeure », déclare le président (cité par Méléka, 1977, p. 58). L'ATLF deviendra l'Association des traducteurs et interprètes de l'Ontario (ATIO).

³³ Nous ne nous prononçons pas, toutefois, sur le bien-fondé de cette décision, car nous ne prétendons pas avoir fait les recherches qui nous permettraient d'émettre une opinion éclairée.

parlementaires et quelques autres documents sur demande. Or, au lendemain de l'entrée en vigueur de la *Loi*, il fallait désormais fournir un environnement de travail dans les deux langues à tous les fonctionnaires en plus de procurer à tous les citoyens des services dans la langue de leur choix. Mossop explique le tour de force que cela représente : à l'époque, seulement 40 % des postes requéraient l'usage des deux langues, en conséquence, il fallait traduire les documents destinés au public préparés par les fonctionnaires unilingues anglophones et par les fonctionnaires unilingues francophones (*ibid.*). Ainsi l'entrée en vigueur de la *Loi sur les langues officielles* a deux principales retombées, qui influenceront la recherche en TA. D'une part, et nous l'avons répété à maintes reprises, la demande en traduction, en plus d'avoir augmenté de façon significative, se maintient à l'abri des fluctuations extérieures des marchés. Cette stabilité aurait pu avoir pour résultat de garantir le financement de la recherche en TA, à notre avis. Cependant, certaines tensions entre les acteurs du monde de la traduction, comme les visées centralisatrices du gouvernement, ont empêché la collaboration nécessaire à un investissement constant en TA. D'autre part, la *Loi* a eu pour résultat d'affermir la raison d'être socioculturelle de la traduction au Canada. Comme l'énonce très bien Jean Delisle à l'occasion du 50^e anniversaire de la *Loi* : « more than an intellectual exercise or an industry, translation is part of Canada's political and sociolinguistic reality » (cité par Mossop, *ibid.*, p. 5). En effet, c'est par la traduction effectuée par le gouvernement fédéral, donc sous la responsabilité du Bureau, que l'identité linguistique canadienne, francophone particulièrement, se consolide. Mossop avance que la traduction est au cœur du processus de bilinguisation du Canada puisqu'elle « bring[s] about a cultural change, that is, a change in how Canadians conceived of the country and an enhanced role from the French language in Canadian public life. » (*ibid.*, p. 6) L'exercice

est périlleux et les enjeux sont importants. Jacques Poisson, traducteur conseil, met en garde contre l'un des écueils de la *Loi*, dans une table ronde publiée chez *Meta* :

On connaît les risques d'interférences qui ont été abondamment étudiés, mais on s'est moins arrêté à l'appauvrissement du français par l'absence d'utilisation des ressources qui ne sont pas suggérées par l'anglais. De même, faute d'un ressourcement original, les ressources culturelles qui sous-tendent la vie d'une langue finissent par s'estomper, ce qui aboutit à une langue sans racines, artificielle [...] (Coty et coll., 1975, p. 64)

Ce rôle de francisation est très présent à l'esprit des traducteurs du Bureau à l'époque et perdure aujourd'hui. Nous croyons que cette attitude explique, du moins en partie, la méfiance de certains traducteurs et gestionnaires du Bureau quant à la qualité du français que l'on peut obtenir à l'aide de la TA. En plus du problème de la qualité des sorties machine, il se peut que les traducteurs aient craint que le recours systématique à une traduction rapide n'élimine les efforts de création terminologique essentiels au maintien du français dans un monde de plus en plus technique. Ainsi, sauf exception (MÉTÉO), les projets et les bancs d'essais de TA au Canada n'ont jamais eu l'aval des décideurs et c'est plutôt les solutions de TAO qui ont été implémentées.

En dépit des réserves de certains, le Bureau s'engage dans la voie des technologies langagières. Mossop remarque à juste titre que la recherche en TA n'est pas la première entreprise technologique du Bureau. La centralisation des effectifs de traduction a permis aux administrateurs de rationaliser certaines opérations, comme en témoigne l'organisation du travail à la fin des années 60. Le Bureau employait « 1,200 salaried translators working in large office-based groups and producing over 200 million words a year using either electric typewriters or dictaphones, with clean copies being retyped by a small army of secretaries », raconte-t-il (*ibid.*, p. 14). L'auteur fait le lien entre les objectifs socioculturels du Bureau et l'introduction de la TA dans son processus de travail : « The idea of using

Machine Translation to increase the quantity of translations produced goes all the way back to the beginning of the socio-cultural era initiated by the Official Languages Act » (*ibid.*, p. 13). Il ajoute qu'à la fois MÉTÉO et TERMIUM sont des projets qui visent à créer et rendre accessible de la terminologie bilingue pour les fonctionnaires fédéraux (*ibid.*). Cette affirmation est tout à fait cohérente avec notre argument précédant quant au caractère essentiel du soutien au français pour le gouvernement fédéral et renforce l'idée que la *Loi sur les langues officielle* est le facteur déterminant pour la recherche en TA au Canada, particulièrement en ce qui concerne le gouvernement fédéral et son Bureau de la traduction.

Cependant, rappelons-le, Mossop avance que le Bureau de la traduction a vu la nature de ses objectifs passer de socioculturelle à économique au cours des années 90. Ce revirement est survenu alors que la demande en traduction continuait d'augmenter et que le nombre de traducteurs au Bureau diminuait. Le gouvernement conservateur de Brian Mulroney décide alors d'augmenter la part des traductions envoyées à des pigistes, plutôt que d'autoriser une vague d'embauches, toujours selon Mossop (2006, p. 5), qui rapporte deux décisions cruciales de ce gouvernement quant à l'avenir du Bureau : d'une part, on a jugé que l'industrie de la traduction au Canada, en majeure partie des micro-entreprises, ne pouvait prendre le relais et en conséquence, la privatisation du Bureau était hors de question. D'autre part, les organismes et ministères ont obtenu leur propre budget de traduction et ont été autorisés, dès lors, à choisir leurs prestataires de services de traduction (interne, privé ou le Bureau), choix qu'ils ne détenaient pas auparavant (*ibid.*, p. 8). Ces décisions prennent tout leur sens lorsqu'on observe le prochain mouvement de l'administration fédérale. En 1993, le Secrétariat d'État est démantelé, et les questions qui relèvent de sa responsabilité, comme la promotion des langues officielles, passent à Patrimoine Canada. La traduction, sous la responsabilité du Secrétariat d'État, se retrouve

quant à elle sous l'égide de Travaux publics et services gouvernementaux. Le gouvernement enfonce le dernier clou en 1995, alors qu'il transforme le Bureau de la traduction en organisme de service spécial (OSS). Les OSS ont en somme une plus grande liberté administrative, mais doivent rendre des comptes en ce qui concerne leur équilibre budgétaire et assurer leur rentabilité. En résumé, comme le dit Mossop, « The Bureau had in effect become a sort of publicly owned translation company » (*ibid.*, p. 9). Il cite l'objectif du Bureau, tel que stipulé dans le rapport que ce dernier présente au Parlement en 1996 : « to provide translation, interpretation and terminology services that are acknowledged as representing the best value, in a commercially-oriented, business-like manner, as outlined in the business plan » (*ibid.*). On ne pourrait être plus clair. Ainsi, si la traduction est désormais considérée comme un service, le modèle de gestion du Bureau est à l'avenant et s'aligne sur les pratiques de l'industrie.

Dès lors, le Bureau préfère acheter des technologies clé en mains plutôt que de financer la R et D, à tout le moins pour la TA. Au cours des années 80 et 90, le Bureau met à l'essai trois logiciels différents : Microcat (de Weidner), Alps et Logos. Les deux premiers essais se soldent par un échec. Nous n'avons pas trouvé de compte rendu pour Alps, mais Macklovitch a supervisé les essais pour Microcat et pour Logos. Il s'avère que Microcat ralentit considérablement les utilisateurs, produit des traductions de piètre qualité nécessitant beaucoup de postédition et ne peut être ajusté pour les besoins du Bureau (Macklovitch, 1991, p. 39-42). En fait, les traducteurs refusent catégoriquement de travailler avec le logiciel lorsqu'on le leur propose après les essais (*ibid.*, p. 39).

Le cas de Logos est intéressant, puisque le Bureau avait déjà mis en place des versions précédentes du logiciel, pour ensuite demander à Macklovitch d'évaluer les performances linguistiques de la dernière version. Les responsables, plutôt que de se concentrer sur des

questions d'efficacité ou d'ergonomie, veulent savoir « what sorts of problems had the developer managed to rectify over the three versions thus far delivered, and what sort of errors persisted? » (*ibid.*, p. 42) Il est aussi important pour les gestionnaires d'évaluer la nature des erreurs les plus fréquentes et leur impact sur la quantité de postédition requise ainsi que, toujours selon l'avis de Macklovitch de déterminer ce qui pourrait être corrigé dans les versions subséquentes (*ibid.*, p. 42). Il devient donc évident, à notre avis, que la recherche fondamentale en TA n'est plus l'objectif premier du Bureau, si elle l'a déjà été, mais qu'il s'agit plutôt d'accélérer la production de ses effectifs. Les efforts déployés par le Bureau pour développer un poste de travail du traducteur sont une autre indication de ce changement de cap. Nous verrons dans la partie suivante comment les travaux des laboratoires comme le CITI ou le RALI ont été intégrés dans les opérations régulières du Bureau, mais il nous faut aussi souligner les initiatives prises à l'interne. Outre le Poste de Travail du Traducteur (PPT 1 et 2) conçu par le CITI, le Bureau lance en 1990 le projet LATTER (L'Atelier du TERminologue), un poste de travail informatisé répondant aux besoins de ses terminologues (Tremblay, 1992, p. 768). Par ailleurs, les postes de travail réguliers des traducteurs sont constamment améliorés avec de nouvelles ressources matérielles et logicielles.

Au tournant des années 2000, le virage économique du Bureau de la traduction est bien amorcé. Le Bureau reste le plus gros employeur en traduction au Canada et l'industrie langagière, tant en traduction qu'en technologies langagières, ne peut rivaliser. Selon le *Rapport final du comité sectoriel de l'industrie canadienne de la traduction*, commandé par Industrie Canada et publié en 1999, 35 % de la demande canadienne en traduction provient du secteur public et de cette proportion, le Bureau en rafle 60 % (p. 18-20). L'industrie de la traduction est morcelée, avec plus de 80 % des cabinets qui ont un chiffre d'affaires de

moins de 500 000 \$ par année³⁴ (p. 20). Qui plus est la rentabilité est loin d'être au rendez-vous. Le rapport indique que l'industrie de la traduction a du mal à se relever de la récession de 1990 : le bénéfice brut se chiffre entre 20 et 25 %, tandis que le bénéfice net dépasse à peine les 5 % (*ibid.*, p. 21). Dans ces conditions, il est évident que les cabinets n'ont pas les reins suffisamment solides pour soutenir la recherche en TA ou en TAO. Les technologies langagières font tout de même partie du paysage, mais il semble que « le taux de pénétration des outils électroniques ou informatiques plus sophistiqués d'aide à la traduction est relativement bas. » (*ibid.*, 29).

L'industrie des technologies langagières, quant à elle, semble aussi avoir besoin de soutien. En 1999, toujours selon le rapport, il existe une quinzaine d'entreprises de recherche en technologies langagières (*ibid.*, 35). À l'image de l'industrie de la traduction en général, l'industrie des technologies langagières est elle aussi morcelée, même si elle connaît une croissance qui avoisine les 20 % (*ibid.*, p 37). Cependant, très peu d'entreprises exportent leurs produits, la plupart se concentrant soit sur le développement soit sur la commercialisation dans le marché canadien. Le Comité sectoriel juge que le manque de financement est le principal problème qui freine le développement de l'industrie. Malgré des partenariats fructueux avec les universités, « la recherche institutionnelle effectuée dans ce domaine d'activité au Canada n'est pas à la hauteur de ce que font d'autres pays ou régions du monde » (*ibid.*, p. 39). En réponse à cette insuffisance, le gouvernement crée en 2003 l'Association de l'industrie de la langue/Language Industry Association (AILIA), qui a pour mandat de promouvoir et soutenir l'industrie de la langue. L'AILIA trace une *Carte routière technologique*, qui brosse un portrait de l'industrie, énonce ses objectifs et propose une stratégie de renforcement du secteur des technologies langagières. Si la mission de

³⁴ Le rapport indique que la proportion est à peu de choses près la même aux États-Unis.

l'association est de « Réunir les forces vives et devenir la voix de l'industrie de la langue au Canada » (AILIA, 2003, p. 6), elle ne dispose toutefois à notre connaissance d'aucun budget substantiel.

Les départements de traduction dans les universités canadiennes à travers le pays ont des préoccupations grandement similaires, puisqu'ils répondent à une demande relativement uniforme. Les premiers programmes universitaires en traduction voient le jour à Montréal, Québec et Ottawa dans les années 60 (Mareschal, 2005, p. 250), en écho aux besoins du marché, qui comme nous l'avons exposé est dominé par le Bureau de la traduction. À l'instar de Mossop, Mareschal affirme que la traduction « comme profession, comme industrie et comme programme d'études [...] est issu[e] de facteurs principalement socio-politico-linguistiques et non pas, comme on pourrait s'y attendre, de facteurs purement économiques et commerciaux » (*ibid.*, p. 251). Les programmes, relativement similaires d'une université à l'autre, visent évidemment à former des traducteurs, mais tendent depuis une dizaine d'années à offrir des cours d'informatique spécialisés. Cependant, Mareschal constate que le principal obstacle que les universités doivent surmonter pour se tenir à jour est bien évidemment le manque de financement (*ibid.*, p. 260). Les laboratoires informatiques sont relativement nouveaux dans le monde de la traduction. En outre, « le recrutement de professeurs spécialisés dans les outils informatique d'aide à la traduction – des oiseaux rares – s'avère difficile » (*ibid.*). En effet, puisque à ce jour aucun programme spécialisé en technologies langagières n'est offert, il n'y a que peu de diplômés qui peuvent enseigner à pied levé. Ce problème devrait cependant se résorber au cours des prochaines années, à notre avis, principalement parce que les technologies langagières sont appelées à prendre une place de plus en plus importante dans la formation des jeunes traducteurs, qui pourront ensuite être embauchés comme spécialistes.

La recherche en TA, toutefois, est loin d'être aussi répandue au Canada. Si Mareschal relève quatorze programmes de traduction en 2005, la *Carte routière technologique* ne fait état que de deux centres de recherche universitaires en technologies langagières en 2003 : le RALI (Recherche appliquée en linguistique informatique) et l'Université du Québec en Outaouais. Le rapport final du Comité sectoriel de l'industrie de la traduction relève que « Dans le milieu, on déplore l'isolement dans lequel sont plongées les différentes équipes de recherche canadiennes » (1999, p. 83). Isolées les unes des autres, certes, mais aussi des chercheurs en traduction et en traductologie. En effet, la recherche fondamentale en TA au Canada, particulièrement dans les laboratoires indépendants, se fait plutôt sur le traitement des langues naturelles et ces laboratoires, même si leurs recherches peuvent être appliquées à la traduction, n'ont que peu de contacts avec les départements de traduction. L'une des raisons qui expliquent ce morcellement tient à une constatation du Comité sectoriel :

Le Canada n'a pas vraiment de politique de recherche précise en matière de langue sur laquelle pourrait s'appuyer l'industrie pour se développer. Il s'agit d'un domaine de recherche parmi d'autres. Pourtant, le Canada est l'un des rares pays à avoir une loi sur les langues officielles et à bénéficier d'une industrie structurée. (1999, p. 82)

En effet, non seulement la recherche en TA a historiquement été laissée au bon vouloir des différents acteurs de l'industrie, mais le gouvernement lui-même a fait preuve d'un manque de vision cohérente que nous nous expliquons mal considérant sa position de pionnier des technologies langagières, son intérêt à les développer et la grande qualité de ses équipes de recherche³⁵. Nous croyons que le changement d'orientation que le Bureau de la traduction a subi au cours des années 90 a pu jouer un rôle dans l'abandon de la recherche

³⁵ Par ailleurs, nous n'avons pas trouvé dans les domaines prioritaires de recherche du Conseil de recherche en science humaine de mention quant à la traduction (http://www.sshrc-crsh.gc.ca/funding-financement/programs-programmes/priority_areas-domaines_prioritaires/index-fra.aspx), encore moins du côté du Conseil de recherches en sciences naturelles et en génie du Canada (http://www.nserc-crsng.gc.ca/NSERC-CRSNG/Index_fra.asp).

fondamentale. Il va sans dire que les préoccupations du Bureau de la traduction, puisque ce dernier est le joueur principal dans l'industrie, doivent avoir un poids considérable dans les décisions gouvernementales quant à l'orientation de la recherche en TA. Ainsi, puisque le Bureau tente, au cours des années 90, de rationaliser ses opérations et de fonctionner dans une logique de profit et pertes, il abandonne la recherche fondamentale et cherche plutôt à intégrer des solutions clé en mains. À cela il nous faut ajouter que les relations entre les écoles de traduction et le Bureau ne sont pas toujours sans accroc, comme en témoigne la décision unilatérale de ce dernier de refuser d'embaucher des finissants de deuxième cycle pour favoriser les bacheliers sans consulter les universités³⁶. Si la querelle a trait à la formation des traducteurs et aux critères d'embauche du Bureau, elle est néanmoins symptomatique d'un dialogue de sourds entre le gouvernement et les institutions d'enseignement.

En somme, le contexte canadien est unique en raison de l'importance socio-historique de la traduction. Alors que la force de l'infrastructure publique, législative et institutionnelle aurait dû réunir les différents acteurs pour soutenir l'industrie de la langue et des technologies langagières, certaines décisions essentiellement économiques ont plutôt mené à son morcellement. De pionnier de la recherche en TA dans les années 70, le Canada accuse aujourd'hui un grand retard en recherche fondamentale sur le plan international et sous-utilise ses chercheurs. Ce retard, il nous faut le préciser, ne signifie pas que les chercheurs canadiens ne sont plus à la fine pointe de la recherche, mais plutôt que le Canada aurait pu, et surtout aurait dû, occuper une place encore plus importante dans le monde des technologies langagières (recherche et commercialisation). Si les chercheurs

³⁶ À ce sujet, voir l'article d'Annie Brisset : « Formation des traducteurs : les critères du Bureau canadien de la traduction sont-ils judicieux? » paru en 2008 dans *TTR*.

canadiens avaient pu profiter du soutien institutionnel auquel ils auraient été en droit de s'attendre, ils auraient sans aucun doute poussé encore plus loin leurs recherches. S'il y a eu des partenariats fructueux entre le gouvernement, les groupes de recherche, l'industrie et les universités, ces partenariats n'ont jamais été soutenus par une vision et une politique cohérentes de soutien et de développement de la recherche.

Chapitre 3 Les grands laboratoires canadiens, étude de cas

Car les objets sont aussi associés à des institutions qui sont des lieux de production, de transmission et de préservation du savoir.
(Milad Doueïhi, *Pour un humanisme numérique*, 2011, p. 10)

3.1 Traduction Automatique Université de Montréal

Le Centre d'études pour le traitement automatisé des données linguistiques (CETADOL), basé à l'Université de Montréal voit le jour en 1965. Ses débuts sont lents (Chandioux 1977, p. 55), mais le groupe évite dès le départ les écueils qu'ont pu rencontrer ses prédécesseurs ailleurs dans le monde en adoptant une approche modulaire³⁷ et axée sur des langues restreintes (Macklovitch, 1997, p. 200). Isabelle, dans un chapitre de *Machine Translation Today: The State of the Art* (1987, p. 247), attire l'attention sur le moment où le groupe redéfinit ses objectifs de recherche, changement de cap qu'exprime sa nouvelle dénomination : Traduction Automatique, Université de Montréal (TAUM). Ce changement de nom donnera le ton à l'orientation du groupe et de ses successeurs : plutôt que de se pencher sur le traitement automatique des langues, les chercheurs basés à l'Université de Montréal s'emploieront dès lors à concevoir des systèmes de traduction automatique à base de règles.

C'est donc sous la gouverne d'Alain Colmerauer, qui prend la relève de Guy Rondeau (premier directeur du CETADOL), que le groupe amorce ce virage vers la TA en 1968. Isabelle et Macklovitch, tous deux collaborateurs chez TAUM, s'accordent pour dire que le SYSTEMS-Q (langage de programmation servant à décrire des règles linguistiques) de

³⁷ L'approche modulaire, comme nous l'avons vu au chapitre 1 consiste à diviser le système en modules (analyse, transfert, synthèse) ce qui permet une manipulation plus aisée. Plus précisément, le composant qui traite les modèles linguistiques est séparé de ceux qui produisent la traduction. Parfois, on sépare aussi les faits lexicaux ou syntaxiques selon qu'ils appartiennent à la LS ou à la LC. Il se peut aussi qu'un autre module consiste en des règles qui décrivent les différences structurales et lexicales des deux langues. Les avantages sont nombreux : les modèles linguistiques peuvent servir indifféremment à l'analyse et à la synthèse, les modules d'analyse et de synthèse peuvent être réutilisés pour d'autres couples de langues et les algorithmes de transfert sont aussi réutilisables (Isabelle et Macklovitch, 1986, p. 115).

Colmerauer était au cœur des efforts du groupe. En 1971, un autre éminent chercheur, Richard Kittredge, prend la direction TAUM. Macklovitch estime que ce dernier « was instrumental in introducing what came to be known as the sublanguage approach. » (1997, p. 200) Cette approche a pour principe fondamental que les langues spécialisées utilisées par les experts de certains domaines (comme la météorologie) laissent relativement peu de place à l'ambiguïté des langues naturelles. En somme, la relative simplicité de ces sous-langues permet une description de la structure des textes (tant en LS qu'en LC) qui y ont recours convenant particulièrement bien à la conception de systèmes de traduction automatique³⁸. Kittredge a vu juste en concentrant les efforts de son équipe sur un système capable de traduire des textes rédigés en langage restreint³⁹, permettant ainsi à TAUM de réaliser une percée unique dans le monde de la TA.

Bénéficiant depuis 1973 du soutien financier du Secrétariat d'État du Canada (lequel lui commissionne en 1975 le développement d'un système de TA pour les bulletins météorologiques canadiens), l'équipe livre en 1977 TAUM-MÉTÉO, reconnu mondialement comme l'un des plus grands succès en TA à ce jour⁴⁰ (voir, entre autres, L'Homme, 2008; Vauquois, 1981; Tremblay, 1992; Hutchins, 1986). D'autres ont rapporté de façon exhaustive les détails tant historiques que techniques qui entourent TAUM-

³⁸ Nous résumons ici très brièvement le propos de Kittredge exposé dans « Variation and Homogeneity of Sublanguages » (1982, p. 107-108). Le chapitre décrit avec force détails la façon dont Kittredge a appliqué ce principe dans les cas de MÉTÉO et d'AVIATION par la suite.

³⁹ Nous considérons que, même si le système requière une intervention humaine en amont (les textes qui y sont soumis doivent respecter certains critères spécifiques et ne sont pas naturels), cette intervention ne fait pas de MÉTÉO un système interactif au sens où nous l'entendons à la section 1.3.2).

⁴⁰ Hutchins rapporte en 1986 que « Although limited in scope, TAUM-MÉTÉO is the first, and so far only, MT system regularly producing translations which are not edited before being made available for public consumption. » (Chap. 13, p. 6) En 2005, Langlais et coll. précisent que des « Spin-offs of that research system, namely MÉTÉO-1 and MÉTÉO-2 (Grimaila and Chandiooux, 1992), have been in continuous use since 1984 translating up to 45,000 words a day. » (p. 167) À notre connaissance, la situation n'a pas changé depuis 2005.

MÉTÉO⁴¹. Il nous semble cependant important d'insister sur le fait que l'interface de TAUM-MÉTÉO est conçue en fonction des besoins des traducteurs, ce qui tranche sur les pratiques des autres laboratoires, même s'il faut tout de même aux utilisateurs une période d'adaptation avant de se familiariser avec cette nouvelle méthode de travail. Une fois cette période écoulée, Chandieux et Guéraud résument la situation au service météorologique du Bureau : « Un équilibre a maintenant été atteint entre les deux notions : machine-menace et machine-panacée. » (1981, p. 21) C'est d'ailleurs grâce au concours des traducteurs du service où le système a été déployé que le logiciel a pu voir son taux d'efficacité passer de 50 % à 80 % (Hutchins, 1986, chap. 13, p. 6-7).

Trois ans après sa mise en service, les traducteurs ne s'en passeraient plus : « [...] on s'habitue très bien à laisser le système faire 80 % du travail en un temps record et, les jours de panne, il devient fort difficile de reprendre le collier de la traduction entièrement manuelle. » (Chandieux et Guéraud, 1981, p. 21) En effet, les traducteurs n'avaient pas à revoir les traductions du système, mais si ce dernier devait faillir, il fallait remplacer la machine au pied levé, avec la surcharge de travail que l'on imagine.

Fort du succès de MÉTÉO, le groupe poursuit sur la même lancée, dirigé cette fois par Marcel Paré, avec le concours de Pierre Isabelle à la direction scientifique, et s'emploie à concevoir un autre système, TAUM-AVIATION, visant à traduire l'ensemble des manuels techniques concernant la maintenance d'aéronefs (CP-140). Les prouesses de TAUM-MÉTÉO avaient placé la barre très haute pour TAUM-AVIATION et le défi était de taille. Aux sept chercheurs initialement affectés au groupe, s'ajoutent rapidement treize nouveaux collaborateurs, soit vingt personnes affectées à temps plein (Isabelle et Bourbeau, 1985,

⁴¹ Voir notamment Hutchins, 1986, section 13.1; Isabelle, 1987, p. 247-277; Tremblay, 1992; L'Homme, 2008.

p. 19). Le Bureau, en plus de financer l'aventure, avait délégué une équipe de traducteurs pour assister les linguistes et programmeurs de TAUM. Le groupe pouvait par ailleurs compter sur le soutien d'André Petit, traducteur spécialisé en aéronautique pour raffiner la compréhension que le système de TA avait de la langue restreinte utilisée (Macklovitch, 2013).

Une telle collaboration étroite avec des traducteurs constituait une première dans la discipline et elle a donné sa couleur à la recherche et au développement au Canada⁴². Pour le groupe, les traducteurs sont d'une aide essentielle en aval : Chevalier et coll. (1981) décrivent en détail le rôle des traductologues⁴³ affectés à leur équipe. Les tâches des spécialistes de la traduction consistent à conseiller les programmeurs quant aux facteurs qui gouvernent la traduction d'une langue à l'autre, écrire les entrées du dictionnaire de transfert, concevoir des tests qui départageront les bons des mauvais équivalents, etc. En somme, les auteurs assimilent le rôle du traductologue à celui du linguiste, c'est-à-dire qu'ils conçoivent la traductologie comme essentiellement descriptive⁴⁴ (Chevalier et coll., 1981, p. 45).

Il ne faut pas en déduire que le traducteur professionnel n'a pas son mot à dire dans le processus. En effet, à la lumière de leur collaboration avec les traducteurs associés à TAUM, Chevalier et coll. définissent le rôle du traducteur : « c'est à lui qu'il appartient

⁴² Il existait tout de même de la recherche en TA dans les autres parties du monde. Quelques groupes américains ont continué à développer des systèmes (par exemple, METAL, Texas University), tout comme des groupes européens (particulièrement le groupe de l'Université de Grenoble). L'approche par interlangue (voir section 1.3.3.2.1) adoptée par ces équipes de recherche s'est vite avérée trop rigide (voir Hutchins 2001, p. 6-8).

⁴³ Selon les auteurs, « la traductologie ne doit pas être confondue avec les techniques et les normes apprises par ceux qui deviennent traducteurs professionnels » (p. 35). Une note à la fin de l'article remercie les traducteurs qui ont participé à TAUM-AVIATION et indique que c'est de leur expérience que l'article est tiré (p. 45), nous en concluons que les traductologues auxquels l'article fait référence sont ces traducteurs.

⁴⁴ Au sens où le traductologue décrirait les opérations nécessaires au transfert vers la LC, sans un système formel et exhaustif. Il n'est donc pas question ici de « Descriptive Translation Studies ».

surtout de découvrir et de suggérer des domaines d'application pour la traduction automatique. » (*ibid.*) L'article de Chevalier et les préoccupations qu'il soulève sont, à notre connaissance, uniques dans les publications de l'époque. Il faudra attendre encore de nombreuses années pour retrouver chez d'autres groupes de recherche un tel intérêt pour l'apport des traducteurs et traductologues au développement de la traduction automatique.

Avec le recul, les chercheurs reconnaissent qu'ils se sont laissé emporter par leur enthousiasme : « the domain is too complex for results comparable to the of TAUM-METEO », écrivent ainsi Isabelle et Bourbeau (1985, p. 19). Le groupe présente malgré tout un système opérationnel vers la fin des années 1970, après trois ans d'efforts intenses. Le système est d'une impressionnante complexité, mais « it came nowhere close to meeting the client's expectations for a fully operational, user-friendly document production chain capable of handling all of the aircraft's many sub-domains. » (Macklovitch, 1997, p. 201) Un rapport (Gervais, 1980) commandé par le Secrétariat d'État juge que l'entreprise est un échec (Tremblay, 1992, p. 762) et l'on abandonne le projet. La débandade est telle que le groupe de recherche est démantelé en 1981.

Il y a un parallèle éclairant à établir entre les événements des années 1960 et la débâcle du groupe TAUM à la fin des années 1970. En effet, comme nous l'avons signalé au Chapitre 1, l'impressionnante efficacité du système Georgetown-IBM, présentée à un public ignorant des contraintes inhérentes à la traduction automatique est en partie responsable des attentes irréalistes par rapport à la traduction machine. Les chercheurs eux-mêmes ne sont pas étrangers à ce malentendu. Bar-Hillel, dans son célèbre article « A Demonstration of the Nonfeasability of Fully Automatic High Quality Translation », se désole de la

[...] reluctance of many MT workers to recognize that the idea of inventing a method for fully automatic high quality translation (FAHQT) is just a dream which will not come true in the foreseeable future. By not realizing the practical futility of this aim, [...] they have misled themselves and the agencies which sponsored their research into not being satisfied with a partly automated translation system whose principles are well understood today, and instead to wait for the real thing which was believed, and made to believe, to be just around the corner. (1960, p. 174)

Ses judicieux conseils restent lettre morte, car peu des groupes de recherche en TA visent moins que l'excellence, et pour ces derniers, cela passe nécessairement par la traduction automatique de haute qualité. Leurs subventionnaires n'en demandent d'ailleurs pas moins, le *Black Book* tout comme l'hécatombe qui frappe les laboratoires américains qui s'ensuit en faisant foi. La même dynamique s'observe au Canada comme les cas de TAUM-MÉTÉO et de TAUM-AVIATION le démontrent. À n'en pas douter, le succès de MÉTÉO a laissé croire tant aux chercheurs de TAUM qu'à leurs bailleurs de fonds que le même exploit était réalisable pour un système d'une plus grande complexité reposant sur une sous-langue autrement plus vaste et hétéroclite que lors du premier essai. Macklovitch formule la chose de la façon suivante : les efforts requis pour reproduire le succès de MÉTÉO avec AVIATION étaient proportionnels à ceux qu'auraient dû déployer les frères Wright pour envoyer un vaisseau spatial sur la lune (Macklovitch, 2013). Il n'en reste pas moins qu'AVIATION obtient, malgré ce qu'on a pu lui reprocher, de bons résultats (entre 60 et 80 % des phrases entrées dans le système sont traduites adéquatement).

La décision du groupe de ne pas programmer de sortie de type « analyse partielle⁴⁵ » a pour conséquence que les phrases non comprises par la machine ne sont simplement pas traduites, ce qui ajoute à l'impression d'inefficacité du système (Isabelle et Bourbeau,

⁴⁵ L'équipe aurait pu choisir de programmer la sortie pour obtenir une traduction mot-à-mot en cas d'analyse partielle, c'est-à-dire quand le système ne comprend pas la phrase qui lui est soumise, traduction imparfaite, certes, mais néanmoins utile. Isabelle et Bourbeau indiquent que cette décision a été prise parce que « these strategies make the quality of the output totally unpredictable. » (1985, p. 19)

1985, p. 19). En somme, AVIATION aurait gagné à être développé plus complètement avant de subir l'évaluation menée par Antoni Gervais. Comme pour les groupes américains avant lui, TAUM est victime à la fois de son enthousiasme et des attentes des subventionnaires. Il en résulte que le Canada, en démantelant son plus important groupe de recherche et de développement en TA délaisse le domaine qu'il dominait jusqu'alors, au moment précis où un renouveau se fait sentir dans le reste du monde.

3.2 Centre d'innovation en technologies de l'information (CITI)

Au tournant des années 1980, la recherche en TA reprend du service un peu partout à travers le monde. La recherche fondamentale n'avait jamais vraiment été abandonnée, mais, exception faite du groupe TAUM, peu d'équipes avaient travaillé à concevoir des systèmes opérationnels (Isabelle, 2013, Macklovitch, 1990, p. 3). Toutefois, l'ouverture des marchés internationaux, qui engendre une demande grandissante pour la traduction de documents techniques et commerciaux, pousse les bailleurs de fonds à financer des projets qui visent à produire des systèmes de TA fonctionnels. Par exemple, aux États-Unis, de nombreux systèmes commerciaux sont mis sur le marché au cours des années 1980 : LOGOS, ALPS ou SYSTRAN, entre autres (Isabelle et Macklovitch 1990, p. 4). En Europe, ce sont à la fois les gouvernements et les entreprises privées qui soutiennent la recherche en TA, avec des résultats comme EUROTRA ou METAL (qui est maintenant chez Siemens)⁴⁶.

Au Canada, les besoins en traduction, engendrés en grande partie par les exigences de la *Loi sur les langues officielles*, excèdent toujours la capacité de production des traducteurs. Isabelle et Macklovitch résument la situation en ces mots : « Nonobstant les aléas du financement de la recherche au pays, les pressions de la demande en traduction constituent

⁴⁶ Pour un bref survol de l'histoire de ces systèmes, voir Hutchins, 1981. Pour un compte rendu détaillé, Hutchins, 1986.

une donnée inéluctable. » (1990, p. 4) Dans ces conditions, il convient de se demander quelle a été la réponse du gouvernement à ces pressions constantes, puisque, rappelons-le, ce dernier a démantelé son équipe de recherche en TA en 1981.

Ainsi, en 1982 le Secrétariat d'État et le ministère des Communications entreprennent une série de consultations qui aboutiront à la conclusion qu'il est essentiel de réinvestir dans la recherche en TA (Isabelle et Macklovitch, 1990, p. 4). Le résultat de cette consultation est publié par Cognos en 1984 dans un rapport intitulé *La traduction automatique et le traitement des langues naturelles : créneaux pour l'intelligence artificielle au Canada*. Les principales recommandations du rapport préconisent la conception, à très court terme, de nouveaux outils de traduction automatique, la promotion des efforts d'automatisation des langues naturelles et de faire connaître, au public et à « tous les secteurs intéressés du gouvernement, de l'industrie et du monde universitaire », les possibilités qu'offrent ces technologies (résumé des conclusions publié dans les *Actes du colloque international sur l'aménagement linguistique*, Wells, « Les services de soutien linguistique », 1986, p. 363). Le Centre canadien de recherche sur l'informatisation du travail⁴⁷ (CCRIT), fondé en 1985 sous la direction de Pierre Isabelle est la réponse à ce rapport (Wells, 1986, p. 363-364; Isabelle et Macklovitch, 1990, p. 7). De ses bureaux à Laval, l'équipe d'Isabelle s'affaire à concevoir tout un éventail d'outils, pour la plupart destinés à faciliter la vie des traducteurs.

Dès la formation du groupe, et pour les années suivantes, les chercheurs du CITI gardent à l'esprit les enseignements de TAUM et concentrent leurs efforts sur la traduction assistée par ordinateur (TAO), plutôt que la TA : « MAHT [Machine aided human

⁴⁷ Le CCRIT deviendra, au début des années 90 le Centre d'innovation en technologies de l'information (CITI) Puisque ce changement ne concernait que la dénomination du groupe (Isabelle 2013, Macklovitch 2013) nous considérerons qu'il s'agit d'un seul et même laboratoire et y référerons comme le CITI.

translation] is the philosophical cornerstone of the CITI's machine-aided translation program. » (Macklovitch, 1995, p. 138) Étonnamment, le premier outil à sortir des murs du CITI est tout de même un système de traduction automatique, le CRITTER (Isabelle et coll., 1988; Dymetman et Isabelle, 1988; Hutchins, 1994a, p. 9, notamment). Rappelons que l'une des trois recommandations du rapport de Cognos portait sur le développement de systèmes de traduction automatique.

Avec CRITTER, l'équipe met de l'avant un système de deuxième génération (voir section 1.3.3.2.2) dont la structure modulaire rappelle celle de TAUM-MÉTÉO et de TAUM-AVIATION. CRITTER partage aussi avec ses prédécesseurs la caractéristique de se limiter à un corpus de langue restreinte (Isabelle et coll., 1988, p. 261), portant sur les rapports hebdomadaires publiés par Agriculture Canada concernant l'état du marché de la viande et du bétail au pays. Qui plus est, CRITTER est programmé en PROLOG, héritier direct de SYSTEMS-Q, le langage de programmation conçu par Colmerauer, autrefois directeur de TAUM (Hutchins et Somers, 1992, p. 68). Ainsi, l'équipe du CITI inscrit-elle ses recherches dans la foulée de TAUM, en mettant à profit l'expertise chèrement acquise du groupe démantelé quelques années plus tôt. Il ne faut pas en conclure que cette équipe ne fait que tabler sur des acquis. En effet, le groupe se penche aussi sur de nouvelles pistes et parmi celles-ci se trouve en bonne place le concept de réversibilité⁴⁸. La traduction automatique est une entreprise complexe qui nécessite d'énormes ressources humaines et matérielles. Les premiers systèmes, conçus tout d'un bloc, laissaient peu de place pour des modifications. La modularité, nous l'avons vu, a réglé en partie le problème puisqu'il était désormais possible d'apporter des modifications à une partie du système sans créer de

⁴⁸ Le groupe qui conçoit le système Rosetta (Phillips Research Laboratories) est le premier à explorer l'avenue de la réversibilité au début des années 1980. Les projets LMT, EUROTRA, UNITRAN et celui d'ISSCO auront aussi pour but explicite d'y parvenir (Hutchins, 1993, p. 14).

problème (de logique, de programmation, etc.) dans les autres modules. Cette compartimentation a aussi comme avantage, nous y avons fait référence (voir note 3) de permettre l'utilisation des modules pour d'autres couples de langues. Il est donc aussi possible de tirer partie de cette modularité pour concevoir des systèmes qui peuvent traduire d'une LS donnée à une LC donnée et vice-versa. L'idée n'est pas aussi facile à appliquer qu'il le semble au premier coup d'œil, comme nous en avons fait mention.

Isabelle donne un exemple parlant des difficultés que pose la réversibilité, qu'il souhaite atteindre avec CRITTER : si l'on souhaite que le système traduise « tu » ou « vous » vers l'anglais, il suffit de faire savoir au système que dans les deux cas, la traduction devrait être « you », cependant, pour faire le chemin inverse, il ne suffit pas de simplement inverser les modules, il faut programmer une nouvelle règle puisque l'équivalent français de « you » peut être, selon le contexte, « vous » ou « tu » (Isabelle, 1989, p. 67).

Le groupe semble sur la bonne voie. En effet, en 1988, Isabelle, Dymetman et Macklovitch rapportent que « CRITTER'S implementation is still incomplete, but the results obtained so far are promising. » (p. 261) La même année, Dymetman et Isabelle rapportent que « the CRITTER translation system is currently capable of translating moderately complex sentences from English to French and vice-versa, using only one grammar for each language » (1988, p. 15). Le CITI est parmi les premiers à atteindre la réversibilité, mais nous n'avons réussi à trouver aucune information sur le destin du système⁴⁹.

⁴⁹ Pierre Isabelle dans un entretien, indique que le système ne pouvait pas être rentable, puisque le volume que peut produire annuellement CRITTER ne peut rivaliser avec le volume produit par les traducteurs (Isabelle, 2012). Tout de même, le relatif silence des auteurs à ce sujet est quelque peu déconcertant, particulièrement si l'on considère leurs nombreuses publications au sujet des projets qui suivront.

Une fois CRITTER mené à terme, le groupe concentre son attention sur la question des outils de TAO. Nous l'avons déjà fait remarquer, les chercheurs du CITI ont à l'esprit les échecs passés de la TA, qu'ils attribuent en grande partie aux attentes trop élevées des bailleurs de fonds et commissionnaires. En 1990, Macklovitch et Isabelle affirment que « Le programme de TAO du CCRIT se veut une réponse cohérente et réaliste à la situation actuelle en TAO, une réponse qui fasse écho aux leçons que l'on doit tirer de l'histoire tumultueuse du domaine » (p. 4). Macklovitch précise que : « in the vast majority of translation situations, currently available MT systems are simply not able to satisfy user's need and/or expectations. » (1995, p. 141). Il explicite plus tard la position prise par le groupe :

The program elaborated by the CITI's TAO team was a conscientious attempt to respond to what was seen as the past failures of machine translation, in particular, the persistent tendency to confuse the aims and methods of research with efforts to provide short-term solutions to practical problems facing working translators. (Macklovitch, 1997, p. 202)

À la lumière de cette affirmation, il est possible d'affirmer que CRITTER (tout comme MÉTÉO et AVIATION, par ailleurs) serait l'une de ces « solutions à court terme », et que la décision de se tourner vers les outils de TAO permettrait d'éviter que la « confusion quant aux buts et méthodes » des chercheurs ne mène au démantèlement du groupe, comme en 1981. Si cette hypothèse est exacte, nous pouvons affirmer que l'équipe d'Isabelle a vu juste. Le groupe TAUM aura produit deux systèmes et survécu dix ans. Le CITI aura quant à lui survécu plus de douze ans et produit une demi-douzaine d'outils dont plusieurs sont toujours disponibles gratuitement ou alors distribués en partenariat avec des entreprises privées⁵⁰.

⁵⁰ Le CCRIT naît en 1985, change de nom pour CITI au tournant des années 90, et est absorbé (équipe, matériel et autres ressources) par le RALI (Recherche appliquée en linguistique informatique) en 1997. Le

Nous l'avons dit, l'équipe du CITI n'est pas avare de publications. Toutefois, l'un de ces articles se démarque et résume à lui seul l'ampleur de la contribution du CITI aux technologies langagières canadiennes. *Translation Analysis and Translation Automation* (Isabelle et coll., 1993) établit à la fois la philosophie de recherche du CITI, sa filiation scientifique, ses postulats et ses principaux projets.

À la question « Which parts of the [machine translation] process should be given to a human partner? », posée par Yehoshua Bar-Hillel en 1951 (cité dans *ibid.*, p. 1133), l'équipe du CITI répond par la voix de Martin Kay : « [...] machines are gradually, almost imperceptibly, allowed to take over certain functions in the overall translation process. First they will take over functions not essentially related to translation. Then, little by little, they will approach translation itself. » (cité dans *ibid.*, p. 1134) S'appuyant sur les épaules de ces deux géants de la traduction automatique, Isabelle et ses collaborateurs formulent leur propre approche, approche qui, à leurs dires, tient plus de l'automatisation du poste de travail du traducteur que de celle de la traduction elle-même (*ibid.*). Les chercheurs proposent donc, dans cet article, l'analyse de la traduction comme pierre d'assise de leurs efforts de conception et de développement de technologies langagières, lesquelles seront décrites dans la deuxième partie de l'article.

Les auteurs commencent par formaliser le processus de la traduction comme la relation suivante : $tr_{L_1L_2}(S, T)$, où la traduction (*tr*) d'une langue (L_1) à une autre (L_2) met en relation un texte source (S) et un texte cible (T). Les auteurs proposent que, puisqu'il existe en théorie un nombre infini de textes dans chacune des langues, la relation entre S et T peut

groupe original reste au RALI pendant encore quelques années, puis se sépare lentement. Il est donc difficile de rapporter exactement la durée de vie du laboratoire, puisque les chercheurs eux-mêmes ne distinguent que peu ces derniers lorsqu'ils en parlent (Macklovitch, 2013 et Isabelle 2012). Nous relaterons dans les paragraphes suivants le fil des événements.

être décomposée en les éléments constitutifs des textes source et cible (s_n et t_n), lesquels sont des fragments de phrase qui se correspondent. Ainsi, expliquent-ils, un système de traduction automatique est conçu de sorte que pour toute valeur de S, ce dernier fournit une ou plusieurs valeurs T qui font partie de l'ensemble tr_{L1L2} (*ibid.*, p. 1134).

Ils poursuivent leur démonstration en abordant la traduction non pas d'une perspective de production (*machine translation*), mais d'analyse (*translation analysis*). Autrement dit, selon les auteurs, un système d'analyse de traduction s'apparente beaucoup à un système de traduction automatique. En somme, plutôt que de contenir des éléments constitutifs (s_n et t_n) et de produire une traduction qui respecte la relation $tr_{L1L2}(S, T)$, comme un système de traduction automatique, le système d'analyse décide si la relation $tr_{L1L2}(S, T)$ est valide pour une paire de textes dans une combinaison de langues donnée et produit, à partir de ces postulats, un arbre d'analyse qui consiste en des équivalences sous le niveau de la phrase (*ibid.*, p 1135).

En parallèle, les chercheurs soulignent que le bitexte, tel que défini la première fois en 1988 par Harris (voir Harris, 1988 pour le raisonnement complet), est une structure qui permet de ramener une paire de textes qui respecte la relation $tr_{L1L2}(S, T)$ à ses éléments constitutifs (voir aussi section 1.4 de ce mémoire), un peu à la manière de l'arbre d'analyse (*ibid.*, p. 1135). Dans les deux cas, en effet, l'idée consiste à formaliser la relation entre les deux textes en fonction de paramètres que nous qualifierons de « traductifs », c'est-à-dire, qui ont trait à des caractéristiques de textes traduits⁵¹. Le résultat, pour les deux opérations,

⁵¹ Cette idée, utiliser les corpus pour dégager certains traits caractéristiques des textes traduits, pourrait être mise à profit pour explorer la théorie des universaux de la traduction, comme l'a suggéré Mona Baker en 1993. Une autre occasion où le Canada aurait eu un net avantage, grâce à des corpus comme les Hansards. Baker travaillait sur des corpus unilingues et nous ne suggérons pas une parenté entre ses travaux et ceux du CITI, cependant, il serait possible de tirer parti des travaux réalisés par Baker et d'autres (voir Zanettin, 2012 pour un état de la recherche), et de les appliquer aux Hansards.

est une base de données spécialisée, qui organise les informations de façon à permettre à l'utilisateur d'y accéder facilement, rapidement et en contexte.

Isabelle et ses collaborateurs explicitent plus avant leur démarche. Ils avancent que, comme nous l'avons rapporté plus haut, l'analyse de la traduction et la traduction automatique présentent de nombreuses similarités sur le plan formel. Les deux processus diffèrent principalement par le degré de précision requis pour la formalisation (les auteurs parlent de « formal characterisation ») de la relation $tr_{L1L2}(S, T)$ (*ibid.*) En d'autres mots, le degré d'exactitude et d'exhaustivité requis pour obtenir de la traduction automatique de haute qualité est si élevé que cet objectif est presque inatteignable. De la même façon, cette exigence d'exactitude et d'exhaustivité est beaucoup moins élevée si l'on souhaite simplement analyser une traduction⁵². Les auteurs donnent pour exemple les efforts d'IBM pour aligner le Hansard canadien sur une simple base statistique (*ibid.* p. 1135-1136 pour un résumé, voir Brown et coll., 1991 pour le rapport complet). Si les gens d'IBM ont très bien réussi à aligner les textes des Débats canadiens en théorisant la longueur des segments source et cible, Isabelle et ses collègues soulignent que s'ils devaient utiliser le même modèle pour produire une traduction automatique : « an English sentence e would be translated more or less as a random sequence of characters f , whose only notable property is to have a length $l(f)$ that is typical for a translation of an English sentence of length $l(e)$. » (*ibid.*, p. 1136) Nous ne pouvons qu'être d'accord avec les auteurs lorsqu'ils soutiennent qu'un tel système de traduction automatique serait tout à fait inutile en pratique.

Toutefois, et c'est là que tout le raisonnement exposé précédemment prend son sens, les chercheurs du CITI n'en restent pas à ce constat somme toute anecdotique. En effet, en

⁵² Il s'agit ici de travailler avec le concept de résultat sous-optimal, le résultat optimal étant la TEAHQ, le résultat sous-optimal correspond à la TA et à la TAO.

définissant clairement les principes de base communs à la traduction automatique et à l'analyse de traduction, Isabelle et son équipe établissent leur philosophie de recherche pour les années à venir. Ainsi qu'ils le déclarent, « Our basic claim here is only that translation analysis, even based on weak translation models, provides the right foundation for a new generation of translation support tools. » (*ibid.*, p. 1137) Cette nouvelle génération d'outils TAO (la série des « Trans », qu'ils s'emploieront à décrire dans les pages suivantes) n'est pas la première du genre, tant s'en faut⁵³, mais elle a ceci de particulier d'à la fois provenir d'un laboratoire de recherche fondamentale et d'être en partie commercialisée⁵⁴.

« In fact, it is clear that **existing translations contain more solutions to more translation problems than any other available resource.** », écrivent les chercheurs du CITI en 1993 (*ibid.*, p. 1137, les auteurs soulignent). Cet axiome est à la base de toute la démarche des chercheurs du CITI pour les 30 prochaines années. Le premier de la série, TransBase est l'héritier direct du bitexte de Harris. Il s'agit en fait d'une structure qui permet d'aligner ensemble un texte source et un texte cible à l'aide de l'outil d'analyse présenté dans la première partie de l'article, appliqué à la fois sur le texte source et le texte cible. Le bitexte ainsi créé est complètement indexé au niveau du syntagme.

⁵³ Martin Kay en avait déjà esquissé les caractéristiques en 1980, Alan Melby en avait conçu la première mouture en 1978, avec l'*Interactive Translator Station* à la Brigham Young University (Melby, 1980 et Melby et coll., 1981), et Brian Harris en avait donné les moyens en 1988 avec le concept de bitexte, comme nous l'avons exposé au Chapitre 1.

⁵⁴ Nous spécifions TAO, ici, par opposition à TA. Les systèmes de TA sont souvent conçus en laboratoire pour ensuite être mis en marché grâce à des partenariats. SYSTRAN en est le meilleur exemple. Les systèmes de TAO sont plus souvent conçus par des entreprises de technologies langagières ou par des entreprises qui sont dotés d'une division de traduction.

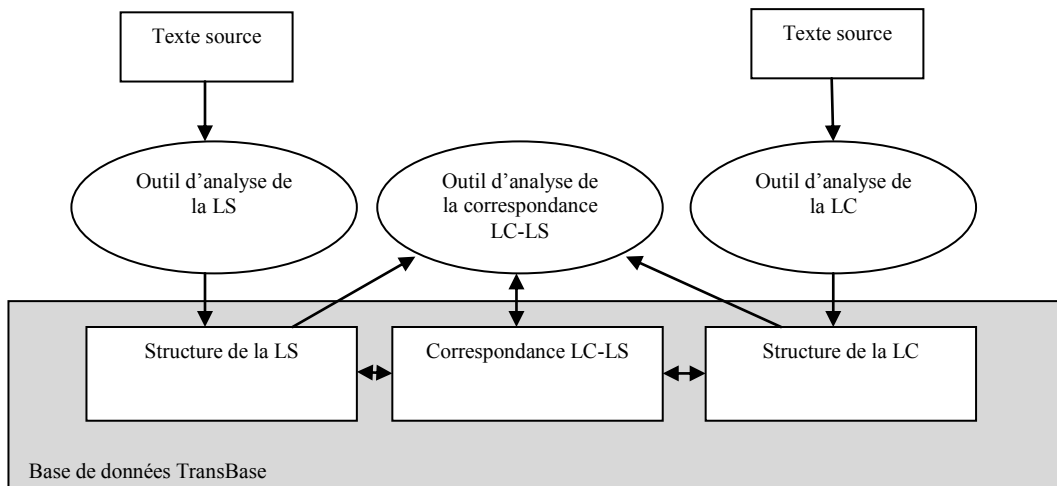


Figure 17 – Structure et processus d’assemblage de la TransBase (Isabelle et coll., 1993, p. 1138 [nous traduisons])

La figure 17 indique la façon dont les outils d’analyse segmentent les textes source et cible, puis les apparie. La figure montre aussi comment les données ainsi colligées sont manipulées de façon symétrique, bien que les concepteurs aient prévu indiquer la source et la cible pour future référence. En outre, à la différence d’une mémoire de traduction, comme nous l’avons vu à la section 1.4, les segments sont conservés en contexte.

TransSearch (<http://www.tsrali.com/>), le deuxième outil conçu par le CITI, et aussi le plus connu, sert à interroger une TransBase à la façon d’un concordancier bilingue. Encore une fois, le principe qui sous-tend cet outil est de faciliter la vie aux traducteurs : « Because the software is primarily aimed at translators, who are likely to use it as just another reference source, it is designed to be used interactively and to provide answers in real-time » (*ibid.*, p. 1138). Le corpus utilisé pour constituer la TransBase que l’on interrogera à l’aide de TransSearch est évidemment le Hansard canadien (pour la première version [il y en aura trois, à ce jour], il s’agit du Hansard de 1986).

La section suivante présente des applications possibles des outils d’analyse décrits plus tôt, mais dont les essais, au moment de la rédaction de l’article, n’ont pas encore permis de développer de solution pratique (*ibid.*, p. 1141-1143). Néanmoins, ces travaux sur les

interférences entre le français et l'anglais mèneront plus tard à TransCheck. En somme, il s'agit de répertorier les faux amis dans le corpus de travail (le Hansard, dans le cas qui nous occupe), pour ensuite les intégrer dans un outil qui vérifierait une traduction humaine et qui identifierait ces faux amis un peu à la façon d'un correcteur orthographique. L'application, qui sera développée plus tard⁵⁵ offre aussi la possibilité de vérifier le transfert correct de valeurs numériques, entre autres (par exemple, le logiciel peut vérifier que tous les points qui séparent les décimales en anglais soient remplacés par des virgules en français).

Enfin, le dernier outil décrit dans l'article « Translation Analysis and Translation Automation » se nomme TransTalk. Tirant parti à la fois de leurs travaux sur TransBase, TransSearch et TransCheck (son prototype), les chercheurs du CITI s'allient au Centre de Recherche Informatique de Montréal (CRIM), un autre centre de recherche financé par Industrie Canada, pour intégrer la reconnaissance vocale à la TAO. Il s'agit ici de permettre au traducteur de dicter sa traduction en langue cible et de bénéficier d'une analyse du texte source pour désambiguïser les mots qui donneraient du fil à retordre à la reconnaissance vocale (*ibid.*, p. 1144). Dans un entretien, Isabelle nous donne l'exemple suivant : pour le cas où la phrase à traduire est « You have beautiful horses », la traduction pourrait être « Vous avez des chevaux magnifiques ». Un logiciel de reconnaissance vocale traditionnel pourrait, en entendant « chevaux », hésiter entre « chevaux » et « cheveux », qui sont très proches phonétiquement. TransTalk, à l'aide de ses modèles de traduction constitués à

⁵⁵ Voir Macklovitch, 1994, « Using bi-textual Alignment for Translation Validation » ou encore Macklovitch, 1995b, « TransCheck – or the Automatic Validation of Human Translation » pour la description détaillée de TransCheck. L'idée de TransCheck vient de l'épouse de Macklovitch. Traductrice au Bureau de la traduction, elle s'échinait sur de longs rapports de Statistique Canada. Elle aurait alors demandé à son mari si le CITI pourrait concevoir un outil qui lui rendrait la révision de ces montagnes de chiffres moins fastidieuses (Macklovitch, 2013). L'épisode peut sembler anecdotique, mais il reflète bien, à notre avis, la proximité des chercheurs du CITI avec les traducteurs (Macklovitch lui-même a toujours été employé par le Bureau et a pratiqué la traduction durant de longues années), proximité qui explique les efforts des scientifiques du CITI à concevoir des outils ergonomiques et pratiques.

partir des outils d'analyse de traduction du CITI, peut reconnaître que la phrase source contient le mot « horses » et qu'en conséquence le mot le plus probable dans le contexte est « chevaux » (Isabelle, 2012).

Nous l'avons relevé en début de section, le CITI est une initiative conjointe du Secrétariat d'État (donc le Bureau de la traduction) et du ministère des Communications. L'un des mandats du CITI était de concevoir un poste de travail du traducteur (PTT) comprenant la fine pointe de la technologie langagière sur un poste de travail informatique. Ainsi, Klaire Tremblay (1992), indique que les logiciels constituant le PTT-1 sont WordPerfect, CompareRite, TextSearch, SM VERBES, Software Bridge, Harrap's, Robert & Collins et DESQView (p. 766). Après évaluation, dit-elle, le CITI propose une deuxième version, le PTT-2, qui exclut TextSearch et Software Bridge, jugés inutiles (*ibid.*, p. 767-768). Dans « La technologie au service du traducteur » (1992b), elle ne fait aucune mention de la composition de l'un ou l'autre des PPT. La description de Macklovitch, dans « PTT-2 : la deuxième version du poste de travail du traducteur du CCRIT » (1992) ne fait pas non plus état des outils de TAO développés par le même laboratoire⁵⁶ (p. 181). Nous en déduisons que la série des Trans n'était pas prête pour le déploiement puisque le poste de travail du traducteur a connu deux vagues, le PTT-1 en 1989 et le PPT-2 en 1991. Le PTT-3 n'offre que deux changements par rapport à son prédécesseur : l'addition d'une application de courrier électronique et le transfert vers une plateforme Windows (Macklovitch, 1993, p. 5). Aucune trace des outils de TAO développés par le CITI. Nous savons néanmoins que la série des Trans se retrouve quelques années plus tard au RALI (nous y reviendrons plus en détails dans la section suivante), sans qu'aucune mention ne

⁵⁶ Rappelons que le CCRIT et le CITI sont deux dénominations pour un seul laboratoire.

soit faite d'essais sur le terrain. Cette absence nous laisse penser que si essais il y a eu, le Bureau a décidé de ne pas conserver les différents logiciels⁵⁷. Cette absence soulève des questions pressantes quant aux motivations du Bureau de la traduction de refuser d'utiliser des technologies qu'il a pourtant commandées, et surtout pour lesquelles il a payé. À tout le moins pouvons-nous affirmer que si le Bureau a bel et bien refusé d'intégrer TransSearch à l'époque du CITI pour ensuite l'offrir à ses traducteurs après sa commercialisation par un tiers, il s'agit sans aucun doute de l'une des occasions manquées auxquelles nous avons fait allusion en introduction⁵⁸.

En 1993, le ministère des Communications, qui chapeaute le CITI, est démantelé (Hillmer, 2013) et le CITI passe sous la tutelle d'Industrie Canada⁵⁹. Quelques années plus tard, à la suite de coupes budgétaires, Industrie Canada décide en 1996 de retirer son financement au CITI, comme à tous les autres laboratoires sous sa tutelle (Macklovitch, 1997, p. 207). Encore une fois, l'expertise accumulée par les chercheurs canadiens est en danger. Ces décisions d'Industrie Canada coïncident avec le changement d'orientation du Bureau de la traduction vers des objectifs plus économiques que sociopolitiques et ne font que confirmer le manque de cohérence dont les institutions gouvernementales n'ont cessé de faire preuve en ce qui concerne la recherche en technologies langagières au Canada. La survie du groupe de chercheurs du CITI, tout comme la pérennité de leurs acquis, semble donc bien incertaine.

⁵⁷ Par ailleurs, lors de notre propre passage au Bureau de la traduction comme agente d'appui à la mémoire de traduction (donc comme spécialiste des TAO), de 2008 à 2010, la série des Trans ne faisait pas partie des applications disponibles, à l'exception de TransSearch, accessible par Internet. Nous relaterons dans la section suivante le fil des événements qui ont mené à la commercialisation de TransSearch.

⁵⁸ À l'époque du RALI, Macklovitch a aussi à deux reprises proposé des projets de recherche interdisciplinaire, rassemblant des chercheurs de nombreuses universités et laboratoires provenant du Canada et de l'Europe, mais qui se sont vu essuyer une fin de non-recevoir (Macklovitch, 2013).

⁵⁹ La plupart des autres organismes sous sa responsabilité passent quant à eux sous l'autorité de Patrimoine Canada.

3.3 Recherche appliquée en linguistique informatique (RALI)

En 1996, Guy Lapalme, alors professeur au Département d'informatique et de recherche opérationnelle (DIRO) de l'Université de Montréal, négocie avec Industrie Canada pour rapatrier les scientifiques du CITI et leur mandat de recherche (Macklovitch, 1997, p. 207; Macklovitch, 2013). Quelques mois plus tard, en 1997, c'est chose faite et l'équipe de Pierre Isabelle s'installe dans ses nouveaux quartiers à l'Université de Montréal. L'équipe du CITI et celle du DIRO unissent leurs forces et créent le laboratoire de Recherche appliquée en linguistique informatique (RALI). Toutefois, la survie du laboratoire n'est toujours pas garantie. En effet, le gouvernement consent à financer le RALI jusqu'en 1999, après quoi, il devra voler de ses propres ailes (Macklovitch, 1997, p. 207). Les montants accordés sont loin d'être faramineux : 500 000 \$ par année, pour douze chercheurs à temps plein et le matériel dont ils ont besoin (Macklovitch, 2013). Le laboratoire doit absolument trouver des partenaires commerciaux pour assurer sa stabilité.

Pour ce faire, les chercheurs peuvent tout de même compter sur des outils dont le développement est suffisamment avancé pour intéresser l'industrie. La série des Trans sera évidemment mise à contribution, mais aussi d'autres outils qui ont été conçus au RALI. Le cas de TransSearch est tout à fait représentatif. En 1996, TransSearch est mis à la disposition du public gratuitement sur Internet. Macklovitch (2000, p. 4) indique que le RALI ne fait aucun effort de promotion de l'outil. Très rapidement, les chercheurs s'inquiètent du nombre grandissant de requêtes et décident d'ajouter une page d'inscription à l'outil pour rassembler des données d'utilisation (*ibid.*). En 1999, le journal d'accès indique quelques 20 000 requêtes par mois (Macklovitch et coll., 2000, p. 4), c'est dire la popularité de l'outil auprès des traducteurs. En 2003, Terminotix conclut un partenariat avec le RALI et reprend TransSearch (maintenant, TSrali) à son compte

(<http://rali.iro.umontreal.ca/rali/?q=fr/TransSearch> [RALI, 2013]). Les chercheurs continuent d'améliorer le produit, mais Terminotix en contrôle l'accès (désormais payant), gère la mise à jour de la base de données et l'hébergement de l'outil. Il nous faut ici souligner que le Bureau de la traduction s'est désormais abonné à TransSearch, avec quelques années de retard. Le partenariat entre le RALI et Terminotix est exemplaire et profite aux deux organisations : d'une part, le RALI doit obtenir du financement extérieur pour continuer ses recherches et n'a pas les ressources pour mettre lui-même en marché ses produits, encore moins pour en assurer le service après-vente; d'autre part, Terminotix n'a ni les ressources humaines, ni les ressources financières pour soutenir des activités de R et D qui lui permettraient de concevoir ses outils à l'interne. Qui plus est, comme le mentionnait Jean-François Richard (président de Terminotix) dans un entretien en 2013, les chercheurs du RALI (et aujourd'hui du CNRC) sont parmi les meilleurs au monde dans le domaine et il serait inefficace de vouloir entrer en compétition avec eux sur leur propre terrain.

Si Terminotix a su profiter de l'expertise du RALI pour TransSearch, d'autres organisations canadiennes n'ont pas su saisir les occasions qui s'offraient. Par exemple, TransType, le dernier né de la série des « Trans » est basé sur un concept novateur que George Foster décrit dans sa thèse de doctorat (Macklovitch, 2013). Il s'agit en somme d'un outil de traduction automatique interactif qui, plutôt que de faire intervenir l'humain pour clarifier le texte source, se base sur la saisie de caractères en temps réel du traducteur pour lui proposer la traduction la plus probable⁶⁰. Macklovitch prend la direction du RALI

⁶⁰ Pour une description sommaire de TransType, voir, RALI, 2013b, pour une description détaillée des concepts sous-jacents et du fonctionnement de l'outil, voir Langlais et coll., 2000; Langlais et coll., 2000a et Foster, Langlais et Lapalme, 2002. Pour l'évaluation du système après les essais, voir Langlais et coll., 2000b et Macklovitch 2006.

en 1999, suite au départ d'Isabelle pour Xerox en France. Il constate immédiatement le potentiel⁶¹ de l'application développée par Foster et se met en quête de financement. C'est le Conseil de recherche en sciences naturelles et en génie du Canada (CRSNGC) qui finance la première phase du projet (RALI, 2013b), mais l'industrie canadienne n'est pas au rendez-vous. Il faut attendre 2002 pour voir le projet renaître, sous la forme de l'initiative internationale TransType 2 (TransType 2, 2005). Le projet est financé à hauteur de 750 000 € par la Communauté européenne (Macklovitch, 2013) et rassemble des groupes de recherche⁶² de France, d'Espagne et d'Allemagne, en plus du RALI, et de la Société Gamma⁶³. Le Canada fournit un financement morcelé : le Programme de recherche-développement coopérative (RDC) du CRSNGC, le ministère du Développement économique et régional et Recherche du Québec, la Société Gamma et l'Université de Montréal fournissent les fonds. Malgré tout, le projet est abandonné en 2005, faute de financement. Macklovitch avance que les défauts de l'application, tels que rapportés dans « TransType2 : The Last Word » (2006), notamment le fait que le logiciel n'apprend pas des erreurs corrigées par l'utilisateur en temps réel (p. 171), auraient pu être atténuées avec la poursuite des travaux, mais que les bailleurs de fonds n'ont considéré que les critiques dans leur décision (Macklovitch, 2013). Cet épisode n'est pas sans rappeler l'affirmation d'Isabelle quant aux dangers de telles démonstrations à un public plus ou moins au fait des exigences de la recherche. Comme dans le cas d'AVIATION, TransType 2 a été jugé et abandonné avant qu'il ne puisse atteindre son plein potentiel. En effet, il est difficile

⁶¹ Il nous faut rappeler que Macklovitch, en plus de ses différentes fonctions au sein des laboratoires depuis 1976 (TAUM, CITI, RALI) n'a jamais cessé d'être employé par le Bureau de la traduction comme traducteur et comme spécialiste des technologies langagières. Son expérience pratique de la traduction, couplée à son expertise en technologies langagières lui ont gagné le respect de tous les acteurs de l'industrie.

⁶² Voir RALI, 2005b.

⁶³ La société Gamma est le cabinet qui a donné naissance à Terminotix. Cette dernière est en quelque sorte la branche technologique de Gamma, et a obtenu son indépendance en 1996.

d'exiger d'un prototype qu'il soit complètement opérationnel. C'est pour cette raison que le financement public est essentiel à la recherche fondamentale : il est injuste de demander de soutenir complètement la recherche fondamentale à une industrie aussi morcelée que l'industrie canadienne des technologies langagières, puisque cette dernière n'a pas les reins suffisamment solides pour investir à long terme dans l'amélioration d'outils comme TransType 2. Cette responsabilité envers la recherche fondamentale échoit, à notre avis aux institutions gouvernementales.

Malgré les obstacles, le RALI a tout de même réussi à négocier des ententes avec des partenaires industriels qui sont toujours présents. Parmi eux, notons Terminotix, évidemment, mais aussi KeaText, Druide Informatique, NPL Technologies et Microsoft Research (Chine). Soulignons parmi les partenaires passés Alis Technologies et Machina Sapiens, notamment. Cependant, le RALI n'est plus que l'ombre de lui-même aujourd'hui : composé de douze chercheurs à temps plein en 1996, il n'en reste plus que trois. S'il reste actif au sein de l'Université de Montréal et du DIRO, le RALI a perdu l'avantage en ce qui concerne la recherche en technologies langagières.

3.4 Centre National de Recherches Canada (CNRC) et le Centre de recherche en technologies langagières (CRTL)

En 2003, le gouvernement canadien lance la *Feuille de route pour la dualité linguistique canadienne 2008-2013 : Agir pour l'avenir*, qui prévoit des investissements de plus d'un milliard de dollars pour renforcer la dualité linguistique canadienne (Patrimoine canadien, 2008). La *Feuille de route* du gouvernement indique que 34 millions sont octroyés à TPSGC pour renforcer le secteur langagier, qui sont divisés comme suit : 16 millions pour le Portail linguistique du gouvernement du Canada (mise en ligne gratuite et accessible à tous de TERMIUM), 8 millions pour un Programme de bourses universitaires en traduction

et les 10 millions restants pour l'Initiative de l'industrie de la langue (Patrimoine canadien, 2008b). L'AILIA est le résultat direct de cette initiative, mais un autre 10 millions a en outre été alloué au CNRC pour la création du Centre de Recherche en Technologies Langagières (Industrie Canada, 2011). Le CRTL a donc été fondé en partenariat entre le CNRC, le Bureau de la traduction et l'Université du Québec en Outaouais (UQO) en 2004 (CRTL, 2012). Le Centre se veut un creuset pour l'industrie des technologies langagières canadiennes et regroupe sous son toit une équipe de recherche du CNRC, des employés du Bureau de la traduction et sert de salle de classe pour les cours de technologies langagières du programme de traduction de l'UQO. Mme Janet Dewey, anciennement de chez SDL (l'un des principaux fournisseurs internationaux d'application en TAO) en assure la présidence depuis sa fondation. Tout indique que la visée du CRTL est rassembleuse, mais la suite des événements montre que le résultat est tout autre.

Le premier effet notable de cette initiative est de vider peu à peu le RALI de son personnel. En effet, lorsque le CNRC fonde son « groupe de recherche de l'Institut de technologie de l'information du Conseil national de recherches Canada » (ITI-CNRC), groupe qu'il détachera au CRTL, il fait appel aux services de Pierre Isabelle pour rassembler une équipe de scientifiques. Isabelle quitte Xerox pour revenir s'installer au Canada et se met à la recherche des meilleurs éléments du domaine. De nombreux chercheurs du RALI passent alors au CNRC, puisque leur premier employeur n'a pas la possibilité de leur offrir la même stabilité, ni les mêmes ressources, par manque de financement à long terme (le RALI, rappelons-le doit être auto-suffisant pour bénéficier de financement public). Ainsi, des chercheurs comme George Foster et Michel Simard quittent le RALI pour grossir les rangs du CRNC. En somme, le premier effet de l'initiative du

gouvernement pour renforcer la recherche en technologies langagières a été de cannibaliser l'un de ses propres laboratoires⁶⁴.

Le groupe de recherche a néanmoins le vent dans les voiles et explore de nouvelles avenues de recherche. Isabelle, tirant parti de son expérience avec les outils à base d'analyse statistique au CITI (la série des Trans), mène donc son équipe dans cette nouvelle direction. Leur principale réalisation, Portage, est un système de TAS à base de phrases (voir section 1.3.3.3.2 de ce mémoire) qui utilise des algorithmes sophistiqués et qui peut être entraîné pour améliorer ses performances⁶⁵. Portage rafle les honneurs à l'Open MT de 2012, compétition organisée par le National Institute of Standards and Technology (NIST). Comme pour le CITI et le RALI avant lui, le gouvernement exige que ses chercheurs négocient des ententes de partenariat avec l'industrie. Depuis 2006, le CNRC collabore avec le projet Global Autonomous Language Exploitation (GALE), groupe financé par la Defense Advanced Research Project Agency américaine⁶⁶ (DARPA). Le contrat, d'une valeur de plusieurs millions de dollars autorise le GALE à utiliser Portage comme moteur de TA (Secrétariat du Conseil du Trésor Canada, 2007). De plus, le Bureau de la traduction l'adopte en 2011-2012 (PWC, 2012, p. 33). Multicorpora, de son côté, utilise la technologie PORTAlign sous licence commerciale pour aligner ses bases de données (AILIA, 2013). En juillet 2012, le CRNC annonce que CSL-Lexi-Tech, une compagnie de traduction de

⁶⁴ Il est important de mentionner que cet épisode a eu des conséquences difficiles pour les chercheurs qui ont du faire ce choix. Le monde de la recherche en TA est très petit et les acteurs sont peu nombreux. Qui plus est, Macklovitch est toujours à la tête du RALI à ce moment.

⁶⁵ Pour une description du système, voir Sadat et coll., 2005; pour une discussion sur certaines améliorations apportées au système, voir Johnson et coll., 2006; pour un résumé de la préparation que l'équipe a mise en place pour la compétitions du NIST, voir Chen, Foster et Kuhn, 2010; enfin, pour la présentation du classement de Portage au Open MT 2012 du NIST, voir Foster et Kuhn, 2012.

⁶⁶ Notons que ce nouveau partenariat nous ramène à l'époque des premiers balbutiements de la TA, alors que les objectifs poursuivis étaient militaires. Il ne faut donc pas s'étonner que Portage fonctionne particulièrement bien dans les paires anglais-arabe et anglais-chinois, comme en témoignent ses excellents résultats à l'Open-MT 2012.

calibre international basée au Nouveau-Brunswick se porte acquéreur de licences de Portage pour ses opérations régulières (CNRC, 2012). Qui plus est, en avril 2013, Terminotix ajoute Portage à l'éventail d'outils disponibles dans Logiterm en plus de fournir la possibilité de l'intégrer seul dans n'importe quel autre environnement de traduction (LinguisTech, 2013). Par ailleurs, il est à noter que Portage n'est pas la seule réalisation du CNRC, par exemple, Alain Désilet a conçu Webitext en 2007, un concordancier en ligne qui permet désormais de constituer *ad hoc* des corpus bilingues alignés et interrogeables, selon les besoins de l'utilisateur. En 2012, Terminotix obtenait la licence d'exploitation de l'application web (CNRC, 2012b). Tous ces partenariats sont relativement récents et seul l'avenir pourra dire si le CNRC poursuivra ses activités de recherche en TA pour de nombreuses années. Déjà, cependant, un certain mouvement se fait sentir : l'équipe de recherche jusqu'alors détachée au CRTL a été rapatriée en mars 2013 dans les bureaux du CNRC, privant le Centre de l'un de ses partenaires les plus actifs (Isabelle, 2013). Il n'est pas clair, toutefois si le retrait de l'équipe des locaux du CRTL entraîne nécessairement le retrait du CNRC des activités du CRTL. Pour l'instant, rien ne laisse présager un tel revirement.

Le CRTL a aussi perdu les employés du Bureau de la traduction qui y étaient logés. Lors de notre passage en 2012, le Bureau venait de les rapatrier. Nous avons aussi constaté que le troisième étage de l'édifice, qui devait accueillir des entreprises de l'industrie des technologies langagières était complètement vide. L'Infolettre de LinguisTech (Vol. 2, numéro 10), publiée le 31 juillet 2013 comporte l'annonce suivante :

Le Centre de recherche en technologies langagières (CRTL) est actuellement à la recherche de compagnies qui souhaiteraient louer certains de ses locaux. Situé au 283 boulevard Alexandre-Taché, le CRTL est un incubateur en technologies langagières qui offre à ses locataires un environnement de travail synergique. Au

CRTL, vous aurez notamment la chance d'échanger et de faire du réseautage entre divers agents évoluant dans le même domaine que vous. (LinguisTech, 2013b)

Il semble donc que la situation soit restée la même. Il ne reste, dans l'édifice du boulevard Alexandre-Taché, que LinguisTech et son cabinet de traduction. Il ne faut cependant pas croire que si l'édifice n'est finalement pas devenu le creuset physique des technologies langagières canadiennes, où les différents chercheurs, gens d'affaires, professeurs, traducteurs et étudiants auraient pu collaborer activement dans un même environnement synergique, que le CRTL a nécessairement échoué. De nombreux projets collaboratifs ont été menés à bien sous son égide⁶⁷. Cependant, il est important de se demander quelle proportion du budget total du *Plan d'action pour les langues officielles* a été allouée à des biens immobiliers, proportion qui aurait peut-être pu servir à soutenir encore plus avant la recherche en technologie langagière.

Sous l'égide du CRTL, en tant que moyen de promouvoir les technologies langagières et leur utilisation dans le cadre de la *Feuille de route pour la dualité linguistique canadienne*, LinguisTech se veut « un site Web de référence pour les langagiers qui comprend une boîte [sic] à outils de technologies langagières, un centre de documentation et de formation ainsi qu'une communauté virtuelle. » (LinguisTech, 2011). Le site et son contenu sont tout d'abord offerts gratuitement à tous les internautes, y compris le Bureau virtuel, une plateforme qui permet à l'utilisateur d'avoir accès à distance à une panoplie

⁶⁷ Barçah, un outil de terminométrie sous la responsabilité de Jean Quirion, de l'UQO; BiTerm, un outil terminologique aligné sur TransSearch, sous la responsabilité de Michel Simard du CRNC; Édition archivistique, un projet de numérisation et d'indexation de textes d'archives, sous la direction de Marie Grégoire de l'UQO et de Pierre Isabelle du CNRC; L2Corr, un correcteur orthographique pour des usagers de langue seconde, piloté par Alain Désilets, du CNRC; MemRev, un outil de révision automatisé pour traducteurs, sous la direction de Roger Racine du BT et de Pierre Isabelle du CNRC, parmi d'autres (CTRL, 2012b).

d'outils de TAO, sans devoir négocier avec chaque fournisseur individuellement⁶⁸. Toutefois, puisque le CRTL, comme le CITI, le RALI et le CNRC, doit parvenir à être auto-suffisant sur le plan financier, LinguisTech offre désormais des forfaits d'abonnement aux étudiants, professeurs et aux universités qui souhaiteraient y accéder. Pour l'instant, en raison de restrictions posées par les licences d'utilisation, les traducteurs professionnels n'ont pas accès à l'abonnement. Le reste des ressources (tutoriels, fiches d'aides, bibliothèque virtuelle, forums, etc.) demeure accessible gratuitement à tous.

Enfin, le CRTL a lancé un dernier projet en avril 2012 : le cabinet de traduction étudiant de LinguisTech. Dirigé par Pierre LeBel, le Service de traduction LinguisTech se veut une réponse à la pénurie de stages offerts par le gouvernement du Canada. Selon le communiqué de lancement, le cabinet pancanadien comptera cinq stagiaires à son bord pour le mois de mai 2012 (CRTL, 2012c). Chacun des stagiaires sera supervisé par un membre de l'Ordre de traducteurs et terminologues du Québec (OTTIAQ) et pourra comptabiliser les mots traduits dans le cadre de son stage pour son agrément à l'ordre. LeBel, dans *l'Actualité Langagière* d'octobre 2012, rapporte que des cinq stagiaires embauchés pour l'été, il compte passer à quinze en janvier 2013 (p. 7). Cependant, l'article ne dit rien des mentors ou des clients, sinon que l'OTTIAQ s'engage à considérer les mots traduits pour l'agrément et que les agences considèrent le cabinet « comme pigiste et lui confie des textes à traduire » (*ibid.*). Le 30 août dernier, *l'Infolettre* de LinguisTech

⁶⁸ Les outils présentement disponibles sont les suivants : Adobe (bureautique), Antidote HD (outil d'aide à la rédaction), CatsCradle (outil de localisation), Fluency (environnement de traduction, sur demande), LogiTerm Pro (environnement de traduction), MultiTrans (environnement de traduction), MultiTrans Flow (gestion de traduction), OpenOffice (bureautique), SDL Passolo (environnement de traduction), SDL Trados 2007 (environnement de traduction), SDL Trados Studio (environnement de traduction), Suite Fusion (environnement de traduction), Suite Microsoft Office (bureautique), TradooIT (concordancier), Transflow (gestion de traduction), Transiterm (application de gestion terminologique) et WhiteSmoke (outil d'aide à la rédaction, sur demande) (LinguisTech, 2011b).

annonce que le cabinet est à la recherche de nouveaux stagiaires pour l'automne (LinguisTech 2013c).

En somme, les initiatives du CNRC et du CRTL s'inscrivent toutes les deux dans la dynamique qui anime l'industrie de la traduction et celle des technologies langagières au Canada. D'une part, ces projets ont vu le jour pour répondre à la demande en traduction qui s'est toujours maintenue depuis les années 60 en raison de la *Loi sur les langues officielles*, et d'autre part, le financement qui leur permet d'exister est tributaire d'impératifs économiques exigeants, qui mettent en péril la stabilité de leurs activités. Qui plus est, lorsque l'on retourne en arrière pour observer les tendances générales qui caractérisent la recherche en technologies langagières au Canada, force est de constater que le gouvernement fédéral n'a jamais maintenu de politique cohérente et durable en ce qui concerne ses programmes de recherche scientifique. En effet, les initiatives sont lancées et abandonnées au gré des impératifs budgétaires des gouvernements en place, sans vision à long terme et surtout en faisant table rase à chaque fois, obligeant les chercheurs à tout recommencer d'un laboratoire à l'autre. Jusqu'à aujourd'hui, le Canada a pu compter sur un noyau de scientifiques de calibre international, qui a su tirer les leçons du passé pour conserver une place tout de même intéressante dans le marché mondial, en dépit du manque de soutien. Le CRTL se voulait un projet fédérateur et présentait, à notre avis, le potentiel nécessaire pour fournir un carrefour efficace et engageant de collaboration entre chercheurs, industrie et universités, mais les impératifs de rentabilité et de profits qui caractérisent les institutions gouvernementales depuis les années 80, et qui ne semblent pas en voie de céder le pas, menacent le laboratoire du même sort que celui de ces prédécesseurs.

Conclusion

Nous avons amorcé notre réflexion en posant la question suivante : Pourquoi, alors que le gouvernement canadien disposait d'une motivation forte à soutenir la recherche en TA, pouvait se targuer de l'une des plus grandes réussites du domaine et employait toujours les chercheurs qui ont réalisé ce projet, la collaboration ne s'est-elle pas poursuivie? Nous avons démontré que la motivation forte que nous avons d'emblée attribuée au gouvernement canadien pour soutenir la recherche, c'est-à-dire la pression exercée par la demande constante en traduction générée par la *Loi sur les langues officielles* devait être nuancée et remise en contexte selon de nombreux facteurs.

Il est vrai que cette demande ne diminue jamais et a plutôt tendance à augmenter, tant en raison de la *Loi*, que de la pression des nouveaux marchés. Par ailleurs, nous avons démontré que suite à des coupes budgétaires et à des restructurations institutionnelles les traducteurs ont vu leurs rangs se clairsemer. La motivation derrière le système TAUM-MÉTÉO, et après lui TAUM-AVIATION était de traduire plus, plus rapidement, avec moins de main-d'œuvre. Toutefois la perception des possibilités de la traduction automatique, donc les attentes des décideurs, est trop souvent déformée par des promesses irréalisables. En ce sens, la TA au Canada a souffert de comparaison avec d'autres systèmes apparemment plus performants et les résultats jugés décevants d'AVIATION, par exemple, ne l'étaient qu'au regard d'espérances gonflées.

Un autre effet notable de la législation linguistique canadienne est que l'industrie de la traduction n'est pas qu'une simple industrie. Traduire au Canada, particulièrement vers le français minoritaire, c'est aussi participer à la création de l'identité canadienne, une identité biculturelle dont aucune des parties ne veut céder devant l'autre. Les traducteurs de l'anglais sont particulièrement conscients de leur responsabilité en regard du français. Ils

traquent le faux ami, le calque avec fougue et vivacité et ne sont pas enclins au compromis. Puisqu'une partie non négligeable de la traduction au Canada passe par le Bureau de la traduction, l'organisation, du moins à ses débuts, était particulièrement consciente de cette responsabilité. Des résultats moins qu'exemplaires, peu importe la sophistication du système, n'auront que très peu de chance d'obtenir l'aval des décideurs.

Cependant, au cours des années 90, le gouvernement prend une série de décisions qui renversent la tendance : on a refusé de privatiser le Bureau de la traduction en alléguant que l'industrie ne pouvait supporter cette responsabilité, tout en envoyant une plus grande proportion du travail en sous-traitance pour économiser. Dans la même foulée, on a autorisé les ministères à choisir leurs prestataires de services langagiers plutôt que d'imposer le Bureau de la traduction. Enfin, le Bureau de la traduction est devenu un organisme spécial de service, avec des objectifs de rentabilité à remplir. Les pratiques de gestion du Bureau imitent dès lors le monde de l'entreprise, avec des plans d'affaires, des présidents et vice-présidents, des primes de rentabilité, etc. De fait, le Bureau retire son financement au CITI, alors que le ministère des Communications, l'autre organisme subventionnaire, se voit démantelé et le CITI passe à Industrie Canada. Il ne faut que quelques années pour qu'à son tour Industrie Canada décide que l'automatisation de la traduction coûte trop cher pour ce qu'elle rapporte et démantèle le laboratoire.

Déjà, à ce moment, on constate que le gouvernement fédéral ne considère plus la traduction comme un aspect important de son identité culturelle, mais bien comme un service qui demande que l'on délie les cordons de la bourse. En corollaire, la recherche en technologie langagière devient une dépense pour laquelle il faut recevoir un retour sur investissement à court terme. En même temps, le Bureau de la traduction fait ses propres essais sur des technologies clés en main dont les prouesses sont vantées par leurs

représentants commerciaux mais dont les résultats sont décevants. L'idéologie commerciale et la vue à court terme du Bureau l'empêchent de tirer parti de technologies développées à l'interne pour lesquelles il a déjà payé (la série des « Trans »).

Il n'est donc pas étonnant de voir Industrie Canada sabrer à nouveau dans les acquis du CITI (déjà, le Bureau avait commis la même erreur avec TAUM), et la recherche en technologie au Canada ne doit sa survie qu'à la prévoyance du directeur du DIRO. Le RALI reprend donc les outils que ses chercheurs avaient développés au CITI et les perfectionne. La logique marchande du gouvernement est encore à l'œuvre lorsque ce dernier impose comme condition au financement du laboratoire des partenariats avec l'industrie.

Si le RALI n'est pas démantelé, il est sérieusement affaibli avec le *Plan d'action pour les langues officielles*, qui prévoit la création d'un nouveau laboratoire de recherche en technologies langagières en vue de renforcer l'expertise canadienne dans le domaine et de lui redonner sa place de leader mondial de la recherche dans le domaine. De façon plutôt ironique, le gouvernement canadien saborde les efforts d'un laboratoire pour en soutenir un autre. Cela démontre bien le manque de cohérence du gouvernement quant à sa politique de financement de la recherche.

Au fil de tous ces revirements, la demande en traduction n'a cessé d'augmenter, nous l'avons souligné, et l'initiative du CNRC/CRTL vise précisément à faire face à cette augmentation. Malgré les aléas du financement, les chercheurs canadiens ont pu non seulement conserver leurs acquis, mais ils ont aussi su s'adapter, voire devancer les tendances mondiales, parfois à l'insu de leurs bailleurs de fonds. TAUM-MÉTÉO était l'un des premiers systèmes à utiliser un vocabulaire contrôlé, tandis que la série « Trans »

mettait à profit, en TAO, des méthodes statistiques qui allaient plus tard faire le succès de la TAS.

Nous nous demandions si l'on pouvait, en quarante ans de recherches en technologies langagières trouver un seul exemple de collaboration fructueuse entre le Bureau de la traduction et les laboratoires à l'image de TAUM-MÉTÉO. Il s'avère que Portage pourrait bien relancer le Canada à l'avant-scène de la recherche en traduction automatique. Les résultats du système dans les compétitions mondiales l'ont sacré l'un des meilleurs systèmes sur le marché. Les partenariats (imposés par le gouvernement, il est vrai) avec de nombreuses entreprises privées lui donnent une visibilité hors du monde de la TA dont TAUM-MÉTÉO n'aurait jamais pu rêver. Cependant, le morcellement des équipes de recherche, qui n'ont pas de forum central où échanger, l'isolement des chercheurs en TA du monde universitaire et de l'univers de la traduction, et l'épée de Damoclès qu'est l'incertitude du financement à long terme sont autant d'obstacles qui nous empêchent d'affirmer que le Canada peut finalement reprendre sa place parmi les fers de lance de l'innovation en TA. Lorsque nous avons demandé à Elliott Macklovitch si le Canada pouvait reprendre le terrain perdu en recherche fondamentale sur les technologies langagières, sa réponse a été catégorique : aussi longtemps que le gouvernement ne se dotera pas d'une politique scientifique digne de ce nom, la réponse est négative.

En somme, l'histoire des technologies langagières au Canada est non seulement une histoire d'occasions manquées, mais aussi d'éternel recommencement. Tels le Sisyphe de la mythologie grecque, les scientifiques canadiens doivent à chaque fois remonter la pente et porter seuls le fardeau de la recherche en technologies langagière. À trois reprises, le gouvernement a démantelé des groupes de recherche qui fonctionnaient bien et qui produisaient des résultats tangibles. Résultats dont le plus gros fournisseur de traduction au

pays, qui pourtant finançait ces recherches, n'a jamais vraiment pu profiter. La situation a de quoi inquiéter. D'autant plus que la quantité d'information à traiter et à traduire ne cesse d'augmenter, la collaboration entre les différents acteurs du monde de la traduction et des technologies langagières sera essentielle pour faire face à ces nouveaux défis.

Notre réflexion portait sur un aspect seulement de l'histoire des technologies langagières et le format de ce mémoire ne nous permettait pas de creuser en profondeur les motifs plus subtils qui ont pu avoir une influence. Nous nous sommes de plus concentrée sur les chercheurs en TA et avons dû passer rapidement sur d'autres acteurs du secteur. Une enquête de terrain, à la fois diachronique et synchronique sur les technologies langagières permettrait sans aucun doute de tirer des conclusions beaucoup plus précises sur les raisons qui expliquent l'arrivée tardive de l'industrie dans la recherche, ou le morcellement qui semble caractériser à la fois l'industrie de la traduction et des technologies langagières. En d'autres mots, nous souhaitons soit poursuivre le travail entamé ici en interrogeant cette fois les acteurs de l'industrie et ceux du gouvernement soit en inspirant d'autres chercheurs à poursuivre notre travail d'investigation sur ces questions. Est-ce que seule la domination historique du Bureau de la traduction est à pointer du doigt? Les universités (particulièrement les écoles de traduction) pourraient-elles prendre un rôle plus actif dans l'intégration des technologies langagières? Comment le marché de la traduction canadien s'ajustera-t-il à la nouvelle demande multilingue qui arrive sur sa chasse gardée en raison d'Internet et de l'internationalisation des marchés? L'industrie des technologies langagières peut-elle s'y adapter? Les écoles de traduction ont-elles suffisamment intégré les technologies langagières dans leur formation? Quelles sont les nouvelles pratiques qui sont générées par l'utilisation des technologies, langagières ou non? Il nous faudra regarder en arrière pour y trouver des pistes de réponse.

Bibliographie

- AILIA (2003), *Industrie canadienne de la traduction. Carte routière technologique (2003-2007)*, [En ligne] www.crtl.ca/dl34&display (page consultée le 14 avril 2013).
- AILIA (2013). *In the words of... Michel Mellinger* [En ligne] <http://ailia.ca/In+the+words+of+Michel+Mellinger> (page consultée le 25 juin 2013).
- Allen, Jeffrey (2003). « Post-editing », *Computers and Translation*, Harold Somers (dir.), Amsterdam et Philadelphie, John Benjamins Publishing Company, p. 297-317.
- ALPAC (1966). *Language and Machines: Computers in Translation and Linguistic*, Washington, D.C., National Academy of Science et National Research Council.
- Austermühl, Frank (2001). *Electronic Tools for Translators*, Manchester, Royaume-Uni; Northampton, MA, St. Jerome Publishing.
- Baker, Mona (1993). « Corpus Linguistics and Translation Studies: Implications and Applications », *Text and Technology: in Honour of John Sinclair*, Mona Baker, Gill Francis et Elena Tognini-Bonelli (dir.), Amsterdam et Philadelphie, John Benjamins Publishing Company, p. 233-250.
- Bar-Hillel, Yehoshua (1951) « The Present State of Research on Mechanical Translation », *American Documentation*, vol. 2, n° 4, p. 229-237.
- Bar-Hillel, Yehoshua (1960). « A Demonstration of the Nonfeasibility of Fully Automatic High Quality Translation. Appendix III of “The Present Status of Automatic Translation of Languages” », *Advances in Computers*, vol. 1, p.158-163.
- Bowker, Lynne (2002). *Computer-Aided Translation Technology: A Practical Introduction*, Ottawa, University of Ottawa Press.
- Brisset, Annie (2008). « Formation des traducteurs : les critères du Bureau canadien de la traduction sont-ils judicieux? », *Meta*, vol. 21, n° 2, p. 131-162.
- Brown, Peter F., Jennifer C. Lai et Robert L. Mercer (1991). « Aligning Sentences in Parallel Corpora », *29th Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference: 18-21 June 1991, University of California, Berkeley, California, USA*, Morristown, New Jersey, ACL, p. 169-176.
- Chandioux, John (1977). « Histoire de la traduction automatique au Canada », *Meta*, vol. 22, n° 1, p. 54-56.
- Chandioux, John et Marie-France Guéraud (1981). « MÉTÉO : un système à l'épreuve du temps », *Meta*, vol. 26, n° 1, p. 18-22.
- Chen, Boxing, George Foster et Roland Kuhn (2010). « Fast Consensus Hypothesis Regeneration for Machine Translation », *Proceedings of the Joint Fifth Workshop on*

Statistical Machine Translation and Metrics MATR, Stroudsburg, PA, Association for Computational Linguistics, p 11-16.

Chevalier, Monique, Pierre Isabelle, François Labelle et Claude Lainé (1981). « La traductologie appliquée à la traduction automatique ». *Meta*, vol. 26, n° 1, p. 35-47.

CNRC (2012). *Un chef de file de l'industrie langagière adopte un outil de productivité du CNRC* [En ligne] http://www.nrc-cnrc.gc.ca/fra/realisations/saillants/2012/portage_traduction.html (page consultée le 2 août 2013).

CNRC (2012b). *Un outil en ligne gonfle la productivité des traducteurs* [En ligne] http://www.nrc-cnrc.gc.ca/fra/realisations/saillants/2012/cnrc_webitext.html (page consultée le 2 août 2013).

Computer History Museum (2013). *Timeline of Computer History* [En ligne] <http://www.computerhistory.org/timeline/?category=cmpr> (page consultée le 5 juin 2013).

Computer History Museum (2013b). *Timeline of Computer History — 1948* [En ligne] <http://www.computerhistory.org/timeline/?year=1948> (page consultée le 5 juin 2013).

Coty, Jean-Paul, Robert Dubuc, Mary Plaice, Gisèle Poisson et Jacques Poisson (1975). « Table ronde sur l'évolution de la traduction », *Meta*, vol. 20, n° 1, 1975, p. 58-70.

CRTL (2012). *Le Centre de recherche en technologies langagières en bref* [En ligne] http://www.crtl.ca/Mission_CRTL (page consultée le 6 août 2013).

CRTL (2012b). *Projets Terminés* [En ligne] http://www.crtl.ca/projets_termin%C3%A9s_CRTL (page consultée le 6 août 2013).

CRTL (2012c). *LE PREMIER CABINET DE TRADUCTION ÉTUDIANT PANCANADIEN VOIT LE JOUR!* [En ligne] <http://www.crtl.ca/display345> (page consultée le 6 août 2013).

Debbie Folaron (2011). « Translation Tools », *Handbook of Translation Studies, volume 1*, Yves Gambier et Luc van Doorslaer (dir.), Amsterdam, Philadelphie, John Benjamins Publishing Company, [En ligne] <http://0-www.benjamins.com.mercury.concordia.ca/online/hts/> (page consultée le 9 juillet 2013).

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

Désilets, Alain (2007). « Translation Wikified: How Will Massive Online Collaboration Impact the World of Translation? », [En ligne] <http://nparc.cisti-icist.nrc->

cnrc.gc.ca/npsi/ctrl?action=shwart&index=an&req=8913226&lang=en (page consultée le 12 janvier 2013).

Douheih, Milad (2001). *Pour un humanisme numérique*, Paris, Éditions du Seuil.

Dymetman, Marc et Pierre Isabelle (1988). « Reversible Logic Grammars for Machine Translation », Laval, Québec, Canadian Workplace Automation Research Centre, Department of Communications Canada [En ligne] <http://mt-archive.info/TMI-1988-Dymetman.pdf> (page consultée le 16 mars 2013).

EAGLES (1996). *Evaluation of Natural Language Processing Systems Final Report* [En ligne] <http://www.issco.unige.ch/en/research/projects/ewg96/index.html> (page consultée le 3 juillet 2013)

EAMT (2013). *What is Machine Translation?* [En ligne] <http://www.eamt.org/mt.php> (page consultée le 7 juillet 2013).

Esserlink, Bert (2000). *A Practical Guide to Localization*, Amsterdam, Philadelphie, John Benjamins Publishing Company.

Foster, George et Roland Kuhn (2012). *Conference on Portage and its Evaluation at the NIST "Open MT 2012"* [En ligne] <http://linguistech.ca/article371> (page consultée le 5 janvier 2013).

Foster, George, Philippe Langlais et Guy Lapalme (2002). « Transtype: Text prediction for Translators », *Proceedings of the Second International Conference on Human Language Technology Research*, San Francisco, CA, Morgan Kaufmann Publishers, p. 372-374.

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

Gémard, Jean-Claude (1992). « Traduction et industrie de la langue : nouveau défi pour le traducteur? », *Meta*, vol. 37, n° 2, p. 374-378.

Gervais, Antoni (1980). *Évaluation du système-pilote de traduction automatique TAUM-AVIATION. Rapport final*, Ottawa, Canada, Bureau des traductions, Secrétariat d'État.

Harris, Brian (1988). « Bi-text, a New Concept in Translation Theory », *Language Monthly*, n° 54, p. 8-10.

Hillmer, Norman (2013). « Ministère des Communications », *L'Encyclopédie Canadienne* [En ligne] <http://www.encyclopediecanadienne.ca/articles/fr/ministere-des-communications> (page consultée le 3 août 2013).

- Horguelin, Paul A. et Michelle Pharand (2009). *Pratique de la révision (4^e édition)*, Montréal, Linguattech.
- Hutchins, John (1981). « The Evolution of Machine Translation Systems ». *Practical Experience of Machine Translation: Proceedings of a Conference, London, 5-6 November, 1981*, Londres, North-Holland Pub. Co., p. 21-37.
- Hutchins, John (1986). *Machine Translation: Past, Present, Future* [En ligne] <http://www.hutchinsweb.me.uk/PPF-TOC.htm> (page consultée le 26 novembre 2011).
- Hutchins, John (1990). « Out of the Shadows: a Retrospect of Machine Translation in the Eighties », *Terminologie et Traduction*, n° 3, p. 275-292.
- Hutchins, John (1993). « 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, Japan, p. 11-34.
- Hutchins, John (1994). « The Georgetown-IBM demonstration, 7th January 1954 », *MT News International*, n° 8, p. 15-18.
- Hutchins, John (1994a). « Research Methods and System Designs in Machine Translation: a Ten-year Review, 1984-1994 », *Machine Translation: Ten Years On; Proceedings of the Second International Conference Organised by Cranfield University in Conjunction with the Natural Language Translation Specialist Group of the British Computer Society, Held at Cranfield University, England, 12 - 14 November 1994*, [En ligne] <http://www.mt-archive.info/BCS-1994-Hutchins.pdf> (page consultée le 2 juillet 2013).
- Hutchins, John (1995). « 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, p. 431-445.
- Hutchins, John (1996). « ALPAC: The (in)Famous Report », *MT News International*, n° 14, p. 15-21.
- Hutchins, John (1998). « The Origins of the Translator's Workstation », *Machine Translation* vol. 13, n° 4, p. 287-307.
- Hutchins, John (1999). « Retrospect and Prospect in Computer-based Translation », *Proceedings of MT Summit VII "MT in the great translation era"*, Low Hwee Boon (dir.), Singapour, AAMT, pp. 30-34.
- Hutchins, John (2000). « The IAMT Certification Initiative and Defining Translation System Categories », *Fifth EAMT Workshop "Harvesting existing resources", 11-12 May, 2000*, Ljubljana, Slovénie, p. 61-68.
- Hutchins, John (2001). « Machine Translation over Fifty Years », *Histoire Épistémologie Langage*, tome 23, fascicule 1, p. 7-31.

- Hutchins, John et Harold Somers (1992). *An Introduction to Machine Translation*, Londres, Academic Press.
- Industrie Canada (1999). *L'industrie canadienne de la traduction. Stratégie de développement des ressources humaines et d'exportation. Rapport final du Comité sectoriel de l'industrie canadienne de la traduction* [En ligne] www.uottawa.ca/associations/csict/stratf.pdf (page consultée le 2 juin 2013).
- Industrie Canada (2011). *Évaluation finale de l'Initiative de l'industrie de la langue* [En ligne] <http://www.ic.gc.ca/eic/site/ae-ve.nsf/fra/03016.html> (page consultée le 5 août 2013).
- International Data Corporation (2011). *Extracting Value from Chaos* [En ligne] <http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf> (page consultée le 29 juin 2013).
- Isabelle, Pierre (1987). « Machine Translation at the TAUM Group », *Machine Translation Today: The State of the Art*, Margaret King (dir.), Édimbourg, Edinburgh University Press, p. 247-277.
- Isabelle, Pierre (1989). « Towards Reversible MT systems », *MT Summit II August 16-18, 1989, Munich, Germany*, vol. 161, p. 67-68. [En ligne] http://transsearch.iro.umontreal.ca/rali/sites/default/files/publis/towards_reversible.ps (page consultée le 16 mars 2013).
- Isabelle, Pierre (1992). « La bi-textualité : vers une nouvelle génération d'aides à la traduction et la terminologie », *Meta*, vol. 37, n° 4, p. 721-737.
- Isabelle, Pierre (2010). *Introduction à la traduction automatique par règles* [En ligne] <http://linguistech.ca/display657> (page consultée le 20 juin 2013).
- Isabelle, Pierre (2012). Entretien personnel, 12 juillet 2012, Gatineau.
- Isabelle, Pierre (2013). Communication personnelle [courriel] février 2013.
- Isabelle, Pierre et Elliott Macklovitch (1986). « Transfer and MT Modularity », *Proceedings of Coling '86: 11th International Conference on Computational Linguistics, Bonn, August 25th to 29th, 1986, University of Bonn*, Martin Kay (dir.), Morristown, NJ, Association for Computational Linguistics, p. 115-117.
- Isabelle, Pierre et Elliott Macklovitch (1990). « Où en est la traduction automatique? » *Actes du colloque annuel CIPS/CATA, Ottawa*, [En ligne] http://rali.iro.umontreal.ca/rali/sites/default/files/publis/ou_en_est_ps (consulté le 12 février 2013).
- Isabelle, Pierre et Laurent Bourbeau (1985). « TAUM-AVIATION: Its Technical Features and Some Experimental Results », *Computational Linguistics*, vol. 11, n° 1, p. 18-27.

- Isabelle, Pierre, Marc Dymetman et Elliott Macklovitch (1988). « CRITTER: a Translation System for Agricultural Market Reports », *Proceedings of the 12th Conference on Computational linguistics – Volume 1*, Stroudsburg, PA, Association for Computational Linguistics, p. 261-266.
- Isabelle, Pierre, Marc Dymetman, George Foster, Jean-Marc Jutras, Elliott Macklovitch, Francois Perrault, Xiaobo Ren et Michel Simard (1993). « Translation Analysis and Translation Automation ». *Proceedings of the 1993 Conference of the Centre for Advanced Studies on Collaborative research: distributed computing – Volume 2*, Ann Gawnan, Evelyn Kidd et Per-Åke Larson (dir.), (s. lieu), IBM Press, p. 1133-1147.
- Johnson, John Howard, Fatiha Sadat, George Foster, Roland Kuhn, Michel Simard, Eric Joanis et Samuel Larkin (2006). « Portage: With Smoothed Phrase Tables and Segment Choice Models », *Proceedings of the Workshop on Statistical Machine Translation*, Stroudsburg, PA, Association for Computational Linguistics, p. 134-137.
- Kay, Martin (1980). *The Proper Place of Men and Machines in Language Translation*, Palo Alto, Xerox Palo Alto Research Center [En ligne] <http://www.mt-archive.info/Kay-1980.pdf> (page consultée le 10 janvier 2013).
- King, Margaret (1992). « L'évaluation des systèmes de traduction automatique dans le cadre d'un service de traduction », *Meta*, vol. 37, n° 4, p. 817-827.
- Kit, Chunyu, Haihua Pan et Jonathan J. Webster (2002). « Example-Based Machine Translation: A New Paradigm », *Translation and Information Technology*, Chan Sin-wai (dir.), Honk Kong, Chinese University Press, p. 57-78.
- Kittredge, Richard (1982). « Variation and Homogeneity of Sublanguages ». *Sublanguage: Studies of Language in Restricted Semantic Domains*, Richard Kittredge et John Lehrberger (dir.), Berlin, New York, Walter de Gruyter, p. 107-137.
- Kuhn, Roland (2010). *Introduction à la traduction automatique statistique (TAS)* [En ligne] <http://linguistech.ca/display658> (page consultée le 25 juillet 2013).
- L'Homme, Marie-Claude (2008). *Initiation à la traductique*, Montréal, Linguattech.
- Langlais, Philippe, George Foster et Guy Lapalme (2000). « TransType: A Computer-aided Translation Typing System ». *Proceedings of the 2000 NAACL-ANLP Workshop on Embedded machine translation systems – Volume 5*, Stroudsburg, PA, Association for Computational Linguistics. p. 46-51.
- Langlais, Philippe, George Foster, Guy Lapalme et Sébastien Sauvé (2000a). « TransType: a Target-text Mediated Interactive Machine Translation System », *ACL-2000: 38th Annual meeting of the Association for Computational Linguistics, Hong Kong, 1-8 October 2000*, San Francisco, Morgan Kaufmann Publishers, p. 36-37.
- Langlais, Philippe, Sébastien Sauvé, George Foster, Elliott Macklovitch et Guy Lapalme (2000b). « Evaluation of TransType, a Computer-aided Translation Typing System: a

Comparison of a Theoretical and a User-oriented Evaluation Procedures », *LREC-2000: Second International Conference on Language Resources and Evaluation. Proceedings, Athens, Greece, 31 May – 2 June 2000*, Maria Gavrilidou et coll. (dir.), Paris, ELRA, p. 641-648.

Langlais, Philippe, Thomas Leplus, Simona Gandrabur, et Guy Lapalme (2005). « From the Real World to Real Words: The Meteo Case », *Proceedings from the 10th Annual Conference of the European Association for Machine Translation*, Budapest, EAMT, p. 166-175.

LeBel, Pierre (2012). « Services de traduction LinguisTech, le premier cabinet de traduction étudiant pancanadien », *L'Actualité Langagière*, vol. 9, n° 3, p. 4-7.

LeBlanc, Matthieu (2013). « Translators on Translation Memory (TM). Results of an Ethnographic Study in Three Translation Services and Agencies », *Translation and Interpreting*, vol. 5, n° 2, p. 1-13.

LinguisTech (2011). *Qui sommes-nous?* [En ligne] <http://linguistech.ca/About+us+-+fr> (page consultée le 29 juin 2013).

LinguisTech (2011b). *Abonnement* [En ligne] <http://linguistech.ca/Abonnement+bureau+virtuel+LT> (page consultée le 29 juin 2013).

LinguisTech (2013). *Terminotix Adds Statistical Machine Translation to its Products Line* [En ligne] <http://linguistech.ca/article411> (page consultée le 29 juin 2013).

LinguisTech (2013b). « Les technologies langagières au bout des doigts! », *Infolettre de LinguisTech*, vol. 2, n° 10 [En ligne] <http://linguistech.ca/article436> (page consultée le 29 juillet 2013).

LinguisTech (2013c). « Les technologies langagières au bout des doigts! », *Infolettre de LinguisTech*, vol. 2, n° 11 [En ligne] <http://linguistech.ca/article439> (page consultée le 11 septembre 2013).

Macklovitch, Elliott (1991). « Evaluating Commercial MT Systems », *Proceedings of the Evaluators' Forum, April 21st – 24th, 1991, Les Rasses Vaud. Switzerland*, Kristen Falkedal (dir.), Genève, ISSCO, p. 37-49.

Macklovitch, Elliott (1992). « PTT-2 : la deuxième version du poste de travail du traducteur du CCRIT », *L'Environnement traductionnel*, André Clas et Hayssam Safar (dir.), Québec, Les Presses de l'Université du Québec, p. 171-182.

Macklovitch, Elliott (1993). *La troisième version du poste de travail du traducteur*, Centre d'innovation en technologies de l'information, Laval.

Macklovitch, Elliott (1994). « Using bi-textual Alignment for Translation Validation », *Technology Partnerships for Crossing the Language barrier: Proceedings of the First*

Conference of the Association for Machine Translation in the Americas, 5-8 October 1994, Columbia, Maryland, USA, Washington D. C., AMTA, p. 157-168.

Macklovitch, Elliott (1995). « The Future of MT is Now and Bar-Hillel was (Almost Entirely) Right », *Proceedings of the Fourth Bar-Ilan Symposium on the Foundations of Artificial Intelligence*, Moshe Koppel et Eli Shamir (dir.), Menlo Park, CA, AAAI Press, p. 137-148.

Macklovitch, Elliott (1995b), « TransCheck – or the Automatic Validation of Human Translation », *MT Summit V: Luxembourg, July 10-12, 1995: Proceedings*, Genève, EAMT [En ligne] <http://www.mt-archive.info/MTS-1995-Macklovitch.pdf> (page consultée le 20 juillet 2013).

Macklovitch, Elliott (1997). « MT R&D in Canada », *MT Summit VI. Machine Translation: Past, Present, Future. Proceedings, 29 October – 1 November 1997, San Diego, California, USA*, Virginia Teller et Beth Sundheim (dir.), Washington D. C., AMTA, p. 200-210.

Macklovitch, Elliott (2000). « Two types of translation memory », *Proceedings of the ASLIB Conference on Translating and the Computer. Vol. 22*, Londres, Aslib [En ligne] <http://www.mt-archive.info/Aslib-2000-Macklovitch.pdf> (page consultée le 17 mars 2013).

Macklovitch, Elliott (2006). « TransType2: The Last Word », *Proceedings of the 5th International Conference on Languages Resources and Evaluation (LREC 06)*, Gênes, Italie, LREC, p. 167-172.

Macklovitch, Elliott (2013). Entretien personnel, 15 février 2013, Montréal.

Macklovitch, Elliott et Pierre Isabelle (1990). « Les voies actuelles de la TAO au Canada ». *La Tribune des industries de la langue*, vol. 2 [En ligne] <http://rali.iro.umontreal.ca/rali/?q=fr/node/782> (page consultée le 3 mars 2013).

Macklovitch, Elliott, Michel Simard et Philippe Langlais (2000). « TransSearch: A Free Translation Memory on the World Wide Web », *Second International Conference On Language Resources and Evaluation (LREC)*, vol. 3, (s. lieu), LREC, p. 1201-1208.

Mareschal, Geneviève (2005). « L'enseignement de la traduction au Canada », *Meta*, vol. 50, n° 1, p. 250-262.

Melby, Alan (1981). « Translators and Machines – Can they Cooperate? », *Meta*, vol. 26, n° 1, p. 23-34.

Melby, Alan (1983). « Computer-assisted Translation Systems: The Standard Design and a Multi-step Design », *Proceedings of the First Conference on Applied Natural Language Processing: Santa Monica (California), February 1-3, 1983*, Iris M. Kameny et Beatrice T. Oshika (dir.), Morristown, New Jersey, Association for Computational Linguistics, p. 174-177.

- Melby, Alan (1995). *The Possibility of Languages*, Amsterdam, Philadelphie, John Benjamins Publishing Company.
- Melby, Alan, Melvin R. Smith et Jill Peterson (1980). « ITS: Interactive Translation System », *Coling 80: Proceedings of the 8th International Conference on Computational Linguistics, Sept.30 – Oct.4, 1980, Tokyo*, (s. lieu), (s. éd.), p. 424-429.
- Méléka, Fikri (1977). « Le Bureau des traductions (1934-1977) », *Meta*, vol. 22, n° 1, p. 57-65.
- Mitamura, Teruko, Eric Nyberg et Jaime Carbonell (1994). « KANT: Knowledge-based, Accurate Natural Language Translation », *Technology Partnerships for Crossing the Language Barrier, Proceedings of the First Conference of the Association for Machine Translation in the Americas, 5-8 October 1994, Columbia, Maryland, USA*, Washington, DC, AMTA, p. 232-233.
- Mossop, Brian (2006). « From Culture to Business: Federal Government Translation in Canada », *The Translator*, vol. 12, n° 1, p. 1-27.
- Munday, Jeremy (2012). *Introducing Translation Studies: Theories and Applications*, Londres, Routledge.
- Nagao, Makoto (1984). « A Framework of a Mechanical Translation Between Japanese and English by Analogy Principle », *Artificial and Human Intelligence: Edited Review Papers Presented at the International NATO Symposium, October 1981, Lyons, France*, A. Elithorn et R. Banerji (dir.). Amsterdam, Elsevier Science Publishers, p. 173-180.
- Nyberg III, Eric et Teruko Mitamura (1992). *Proceedings of the Fifteenth International Conference on Computational Linguistics: COLING 92, Nantes, 23-28/8/1992*, Christian Boïtet (dir.), Grenoble, IVR Imprimeur, p. 1069-1073.
- O'Hagan, Minako (2011). « Community Translation: Translation as a Social Activity and its Possible Consequences in the Advent of Web 2.0 and Beyond », *Translating as a Social Activity: Community Translation 2.0*, Minako O'Hagan (dir.), Bruxelles, University Press Antwerp.
- Office québécois de la langue française (2001). « Analyseur syntaxique », *Fiche terminologique* [En ligne] http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8366943 (page consultée le 21 juin 2013).
- Patrimoine Canadien (2008). *Feuille de route pour la dualité linguistique canadienne 2008-2013 : Agir pour l'avenir* [En ligne] <http://www.pch.gc.ca/slo-ols/pubs/08-13-LDL/index-fra.cfm> (page consultée le 5 août 2013).

- Patrimoine Canadien (2008b). « Annexe B : les engagements financiers détaillés de la feuille de route 2008-2013 », *Feuille de route pour la dualité linguistique canadienne 2008-2013 : Agir pour l'avenir* [En ligne] <http://www.pch.gc.ca/slo-ols/pubs/08-13-LDL/index-fra.cfm> (page consultée le 5 août 2013).
- Pérez, Celia Rico (2001). « From Novelty to Ubiquity: Computers and Translation at the Close of the Industrial Age », *Translation Journal*, vol. 5, n° 1, [En ligne] <http://translationjournal.net/journal/15mt2.htm> (page consultée le 28 novembre 2011).
- Puech, Michel (2008). *Homo sapiens technologicus*, Paris, Éditions le Pommier.
- PWC (2012). *Translation. Bureau. Benchmarking and Comparative. Analysis. Final Report*, May 15, 2012 [En ligne] <http://www.bt-tb.tpsgc-pwgsc.gc.ca/publications/documents/rapport-report-benchmarking-eng.pdf> (page consultée le 15 juillet 2013).
- Quah, Chiew Kin (2006). *Translation and Technology*, New York, Palgrave Macmillan.
- RALI (2013), *TransSearch : un concordancier bilingue Concordancier bilingue – TransSearch* [En ligne] <http://rali.iro.umontreal.ca/rali/?q=fr/TransSearch> (page consultée le 25 juillet 2013).
- RALI (2013b), *TransType* [En ligne] <http://rali.iro.umontreal.ca/rali/?q=fr/TransType> (page consultée le 25 juillet 2013).
- Richard, Jean-François (2013). Entretien personnel, 22 juillet 2013, Montréal.
- Rosner, Michael (2005). « Machine Translation and Human Translators », Malte, *Sunday Times*, 25 avril 2005, n. p.
- Roudaud, Brigitte (1992). « La traduction automatique : l'ordinateur au service des traducteurs », *Meta*, vol. 37, n° 4, p. 828-846.
- Sadat, Fatiha, John Howard Johnson, Akakpo Agbago, George Foster, Roland Kuhn, Joel Martin et Aaron Tikuisis (2005). « PORTAGE: A Phrase-based Machine Translation System », *Proceedings of the ACL Workshop on building and Using Parallel Texts*, Stroudsburg, PA, Association for Computational Linguistics, p. 129-132.
- Salomon, David (1993). *Assemblers and Loaders*, New York, Ellis Horwood, PTR Prentice Hall.
- Schäler, Reinhard, Andy Way et Michael Carl (2003). « EMBT in a Controlled Environment », *Recent Advances in Example-based Machine Translation*, Michael Carl et Andy Way (dir.), Dordrecht, Kluwer Academic Publishers, p. 83-114.
- Schwab, Wallace (1981). « Traduction et informatique : perspective pour les années 80 », *Meta*, vol. 26, n° 1, p. 48-56.

- Secrétariat du Conseil du Trésor Canada (2007). « Section II – Analysis of Program Activities », *Departmental Performance Reports* [En ligne] <http://www.tbs-sct.gc.ca/dpr-rmr/2006-2007/inst/nrc/nrc02-eng.asp> (page consultée le 6 août 2013).
- Sin-Wai, Chan (2004). *A Dictionary of Translation Technology*, Hong Kong, Chinese University Press.
- Somers, Harold (1990). « Current research in machine translation », *Proceedings of a Workshop on Machine Translation, 2-3 July 1990, UMIST, Manchester*, [En ligne] <http://www.mt-archive.info/UMIST-1990-Somers.pdf> (page consultée le 7 juillet 2013).
- Somers, Harold (1998). « New Paradigms' in MT: The State of Play Now that the Dust has Settled ». *Proceedings of the 10th European Summer School on Logic, Linguistics and Information (ESSLLI), Saarbrücken, 24-28 August 1998: Workshop on Machine Translation* [En ligne] <http://mt-archive.info/ESSLLI-1998-Somers.pdf> (page consultée le 7 juillet 2013).
- TAUS (2010). *1954 Machine Translation Movie* [En ligne] http://www.youtube.com/watch?feature=player_embedded&v=K-HfpsHPmvw page consultée le 15 mai 2013.
- Tenney, Merle D. (1985). « Machine Translation, Machine-aided Translation, and Machine-impaired Translation », *Tools for the Trade: Translating and the Computer 5*, Veronica Lawson (dir.), Londres, Aslib, 1985, p. 105-113.
- TransType 2 (2005). *TransType2 — Computer Assisted Translation* [En ligne] <http://tt2.atosorigin.es/Index.htm> (page consultée le 22 juillet 2013).
- TransType 2 (2005b). *Partners* [En ligne] <http://tt2.atosorigin.es/partners.htm> (page consultée le 22 juillet 2013).
- Tremblay, Klaire (1992). « La technologie langagière au Secrétariat d'État du Canada : une réalité quotidienne », *Meta*, vol. 37, n° 4, p. 761-769.
- Tremblay, Klaire (1992b). « La technologie au service du traducteur », *L'Environnement traductionnel*, André Clas et Hayssam Safar (dir.), Québec, Les Presses de l'Université du Québec, p. 161-170.
- Vauquois, Bernard (1981). « L'informatique au service de la traduction », *Meta*, vol. 26, n° 1, p. 8-17.
- Wagner, Emma (1985). « Post-Editing Systran – A Challenge for Commission Translator », *Terminologie et Traduction*, n° 3, Luxembourg, Commission des communautés européennes, p. 1-7.

- Weaver, Warren (1955). « Translation », *Machine Translation of Languages*, William S. Locke et Andrew. D. Booth (dir.), Cambridge (Mass.), The Technology Press of the MIT, p. 15-24.
- Wells, Winston (1986). « Les services de soutien linguistique », *Actes du Colloque international sur l'aménagement linguistique*, Québec, Les Presses de l'Université Laval, p. 357-364.
- Wheeler, Peter (1987). « Systran », *Machine Translation Today: The State of the Art*, Margaret King (dir.), Édimbourg, Edinburgh University Press, p. 192-208.
- Whitecomb, Tony (1992). « Statistical methods gaining ground », *Language Industry Monitor*, n° 11, Sept-Oct 1992, p. 1-3.
- Zetsche, Jost (2012). *The Translator's Tool Box: A Computer Primer for Translators*, v. 10, Winchester Bay, Oregon, International Writers' Group.
- Zanettin, Frederico (1993). *Translation-Driven Corpora. Corpus Resources for Descriptive and Applied Translation Studies*, Manchester, Royaume-Uni; Northampton, MA, St. Jerome Publishing.
- Zughoul, Muhammad Raji et Awatef Miz'Il Abu-Alshaar (2005). « English/Arabic/English Machine Translation: A Historical Perspective », *Meta*, vol. 50, n° 3, p. 1022-1041.

ANNEXE I – Formulaires de consentement

Pierre Isabelle

CONSENTEMENT À PARTICIPER AU PROJET « DES MACHINES ET DES HOMMES »

Il est entendu que l'on m'a demandé de participer au projet de recherche dirigé par Geneviève Has, du Département d'Études françaises de l'Université Concordia (514-703-5096 — has.genevieve@gmail.com), et supervisé par Benoit Léger, (514-848-2424 poste 7501 — benoit.leger@concordia.ca), et Debbie Folaron, (514-848-2424 poste 7516 — dfolaron@alcor.concordia.ca) tous deux du Département d'Études françaises de l'Université Concordia.

A. BUT

Le but de la recherche a été porté à ma connaissance, à savoir : retracer l'histoire du changement technologique concernant les technologies langagières (traduction assistée par ordinateur, traduction machine, traduction automatisée, etc.) au Canada, des années 70 à aujourd'hui.

B. PROCÉDURE

Il est entendu que le chercheur conduira des entrevues de style journalistique (conversation libre) auprès de personnes ayant participé de près ou de loin à la recherche, au développement, à la commercialisation ou à l'utilisation de technologies langagières des années 70 à aujourd'hui.

Il est entendu que ces entrevues auront lieu là où la personne interrogée le jugera utile, soit en personne, soit par téléphone, soit par voie électronique (courriel, messagerie instantanée, messagerie vidéo, etc.), ou par tout autre moyen de communication.

Il est entendu que ces entrevues seront enregistrées et conservées en vue d'en citer des extraits dans le mémoire.

Il est entendu que le chercheur discutera avec la personne interrogée du temps nécessaire pour chaque entretien.

Il est entendu que la nature des informations ainsi recueillies sera limitée aux faits historiques qui permettront de tracer l'histoire du changement technologique à l'étude.

C. RISQUES ET AVANTAGES

Il est entendu que les risques liés à ma participation au projet de recherche sont minimaux, tout comme les avantages que je pourrais en retirer.

D. CONDITIONS DE PARTICIPATION

- Il est entendu que je suis libre de me désister et d'interrompre ma participation en tout temps sans conséquences défavorables.
- Il est entendu que ma participation à l'étude est :

NON CONFIDENTIELLE (mon identité sera divulguée dans les résultats).

- Il est entendu que les données de l'étude peuvent être publiées.

J'AI LU ATTENTIVEMENT LA PRÉSENTE ENTENTE ET J'EN COMPRENDS LA PORTÉE. JE CONSENS LIBREMENT ET VOLONTAIREMENT À PARTICIPER À L'ÉTUDE EN QUESTION.

NOM (en majuscules)

PIERRE ISABELLE

SIGNATURE



Si vous avez des questions sur la recherche proposée, veuillez communiquer avec le chercheur principal de l'étude :

Geneviève Has, du Département d'Études françaises de l'Université Concordia
(514-703-5096 — has.genevieve@gmail.com)

ou avec :

Benoît Léger, du Département d'Études françaises de l'Université Concordia
(514-848-2424 poste 7501 — benoit.leger@concordia.ca),

ou encore :

Debbie Folaron, du Département d'Études françaises de l'Université Concordia
(514-848-2424 poste 7516 — dfolaron@alcor.concordia.ca)

Si vous avez des questions sur vos droits en tant que sujet de recherche ou participant à une recherche, n'hésitez pas à communiquer avec le conseiller – Éthique de la recherche de l'Université Concordia, par téléphone au 514 848-2424, poste 7481, ou par courriel à ethics@alcor.concordia.ca.

**CONSENTEMENT À PARTICIPER
AU PROJET « DES MACHINES ET DES HOMMES »**

Il est entendu que l'on m'a demandé de participer au projet de recherche dirigé par Geneviève Has, du Département d'Études françaises de l'Université Concordia (514-703-5096 — has.genevieve@gmail.com), et supervisé par Benoit Léger, (514-848-2424 poste 7501 — benoit.leger@concordia.ca), et Debbie Folaron, (514-848-2424 poste 7516 — dfolaron@alcor.concordia.ca) tous deux du Département d'Études françaises de l'Université Concordia.

A. BUT

Le but de la recherche a été porté à ma connaissance, à savoir : retracer l'histoire du changement technologique concernant les technologies langagières (traduction assistée par ordinateur, traduction machine, traduction automatisée, etc.) au Canada, des années 70 à aujourd'hui.

B. PROCÉDURE

Il est entendu que le chercheur conduira des entrevues de style journalistique (conversation libre) auprès de personnes ayant participé de près ou de loin à la recherche, au développement, à la commercialisation ou à l'utilisation de technologies langagières des années 70 à aujourd'hui.

Il est entendu que ces entrevues auront lieu là où la personne interrogée le jugera utile, soit en personne, soit par téléphone, soit par voie électronique (courriel, messagerie instantanée, messagerie vidéo, etc.), ou par tout autre moyen de communication.

Il est entendu que ces entrevues seront enregistrées et conservées en vue d'en citer des extraits dans le mémoire.

Il est entendu que le chercheur discutera avec la personne interrogée du temps nécessaire pour chaque entretien.

Il est entendu que la nature des informations ainsi recueillies sera limitée aux faits historiques qui permettront de tracer l'histoire du changement technologique à l'étude.

C. RISQUES ET AVANTAGES

Il est entendu que les risques liés à ma participation au projet de recherche sont minimaux, tout comme les avantages que je pourrais en retirer.

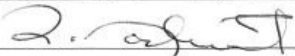
D. CONDITIONS DE PARTICIPATION

- Il est entendu que je suis libre de me désister et d'interrompre ma participation en tout temps sans conséquences défavorables.
- Il est entendu que ma participation à l'étude est :

NON CONFIDENTIELLE (mon identité sera divulguée dans les résultats).

- Il est entendu que les données de l'étude peuvent être publiées.

J'AI LU ATTENTIVEMENT LA PRÉSENTE ENTENTE ET J'EN COMPRENDS LA PORTÉE. JE CONSENS LIBREMENT ET VOLONTAIREMENT À PARTICIPER À L'ÉTUDE EN QUESTION.

NOM (en majuscules) ELLIOTT MAERLOWITZ
SIGNATURE 

Si vous avez des questions sur la recherche proposée, veuillez communiquer avec le chercheur principal de l'étude :

Geneviève Has, du Département d'Études françaises de l'Université Concordia
(514-703-5096 — has.genevieve@gmail.com)

ou avec :

Benoît Léger, du Département d'Études françaises de l'Université Concordia
(514-848-2424 poste 7501 — benoit.leger@concordia.ca).

ou encore :

Debbie Folaron, du Département d'Études françaises de l'Université Concordia
(514-848-2424 poste 7516 — dfolaron@alcor.concordia.ca)

Si vous avez des questions sur vos droits en tant que sujet de recherche ou participant à une recherche, n'hésitez pas à communiquer avec le conseiller – Éthique de la recherche de l'Université Concordia, par téléphone au 514 848-2424, poste 7481, ou par courriel à ethics@alcor.concordia.ca

**CONSENTEMENT À PARTICIPER
AU PROJET « DES MACHINES ET DES HOMMES »**

Il est entendu que l'on m'a demandé de participer au projet de recherche dirigé par Geneviève Has, du Département d'Études françaises de l'Université Concordia (514-703-5096 — has.genevieve@gmail.com), et supervisé par Benoit Léger, (514-848-2424 poste 7501 — benoit.leger@concordia.ca), et Debbie Folaron, (514-848-2424 poste 7516 — dfolaron@alcor.concordia.ca) tous deux du Département d'Études françaises de l'Université Concordia.

A. BUT

Le but de la recherche a été porté à ma connaissance, à savoir : retracer l'histoire du changement technologique concernant les technologies langagières (traduction assistée par ordinateur, traduction machine, traduction automatisée, etc.) au Canada, des années 70 à aujourd'hui.

B. PROCÉDURE

Il est entendu que le chercheur conduira des entrevues de style journalistique (conversation libre) auprès de personnes ayant participé de près ou de loin à la recherche, au développement, à la commercialisation ou à l'utilisation de technologies langagières des années 70 à aujourd'hui.

Il est entendu que ces entrevues auront lieu là où la personne interrogée le jugera utile, soit en personne, soit par téléphone, soit par voie électronique (courriel, messagerie instantanée, messagerie vidéo, etc.), ou par tout autre moyen de communication.

Il est entendu que ces entrevues seront enregistrées et conservées en vue d'en citer des extraits dans le mémoire.

Il est entendu que le chercheur discutera avec la personne interrogée du temps nécessaire pour chaque entretien.

Il est entendu que la nature des informations ainsi recueillies sera limitée aux faits historiques qui permettront de tracer l'histoire du changement technologique à l'étude.

C. RISQUES ET AVANTAGES

Il est entendu que les risques liés à ma participation au projet de recherche sont minimaux, tout comme les avantages que je pourrais en retirer.

D. CONDITIONS DE PARTICIPATION

- Il est entendu que je suis libre de me désister et d'interrompre ma participation en tout temps sans conséquences défavorables.
- Il est entendu que ma participation à l'étude est :
NON CONFIDENTIELLE (mon identité sera divulguée dans les résultats).

- Il est entendu que les données de l'étude peuvent être publiées.

J'AI LU ATTENTIVEMENT LA PRÉSENTE ENTENTE ET J'EN COMPRENDS LA PORTÉE. JE CONSENS LIBREMENT ET VOLONTAIREMENT À PARTICIPER À L'ÉTUDE EN QUESTION.

NOM (en majuscules)

JEAN-FRANÇOIS RICHARD

SIGNATURE



Si vous avez des questions sur la recherche proposée, veuillez communiquer avec le chercheur principal de l'étude :

Geneviève Has, du Département d'Études françaises de l'Université Concordia
(514-703-5096 — has.genevieve@gmail.com)

ou avec :

Benoit Léger, du Département d'Études françaises de l'Université Concordia
(514-848-2424 poste 7501 — benoit.leger@concordia.ca),

ou encore :

Debbie Folaron, du Département d'Études françaises de l'Université Concordia
(514-848-2424 poste 7516 — dfolaron@alcor.concordia.ca)

Si vous avez des questions sur vos droits en tant que sujet de recherche ou participant à une recherche, n'hésitez pas à communiquer avec le conseiller – Éthique de la recherche de l'Université Concordia, par téléphone au 514 848-2424, poste 7481, ou par courriel à ethics@alcor.concordia.ca.