

NRC Publications Archive Archives des publications du CNRC

Traduction automatique: avantages et retombées de la traduction automatique statistique et le logiciel Portage du CNRC

Conseil national de recherches du Canada. Technologies de l'information et des communications

For the publisher's version, please access the DOI link below./ Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

<https://doi.org/10.4224/21274927>

NRC Publications Archive Record / Notice des Archives des publications du CNRC :

<https://nrc-publications.canada.ca/eng/view/object/?id=91f8b1a9-5052-47e7-9c94-033efe904396>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=91f8b1a9-5052-47e7-9c94-033efe904396>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

Traduction automatique

Avantages et retombées de la
traduction automatique statistique et le
logiciel *Portage* du CNRC

avril 2015

Introduction

La traduction automatique (TA) est une façon de transposer au moyen d'un ordinateur un texte écrit dans une langue naturelle en un texte d'une seconde langue naturelle qui sera idéalement son équivalent sémantique et grammatical. Pour le commun des mortels, le système de TA se résume à un encadré noir à l'intérieur duquel on dactylographie le texte dans sa langue d'origine ou indique à l'ordinateur quel fichier il doit traiter. Quelques secondes plus tard, la machine en propose une version dans la langue souhaitée.

Toutefois, il suffit de creuser un peu plus pour constater que le système de traduction automatique peut, selon sa nature – par apprentissage statistique ou à base de règles – influencer sur les résultats obtenus. Depuis 2003, une équipe de spécialistes de calibre international du Conseil national de recherches du Canada perfectionne un système breveté de traduction automatique statistique (TAS) baptisé *Portage*¹. L'article que voici remet la TAS en contexte et compare cette technologie à l'autre grande méthode de traduction automatique en usage : la traduction automatique à base de règles (TABR). Durant la majeure partie de sa courte vie, la traduction automatique avait pour paradigme principal la TABR, dont la genèse peut être retracée à l'année 1948 quand, dans une note envoyée à quelques collègues, le mathématicien Warren Weaver suggéra que l'on pourrait utiliser l'ordinateur pour réaliser des traductions, un peu comme on y avait recouru durant la Deuxième Guerre mondiale pour déchiffrer des codes. Subséquemment, quand la linguistique chomskyenne s'imposa, vers le milieu des années 1960, la traduction automatique en vint à épouser des règles de formalisation déclarative analogues.

Durant cette période, les systèmes de TA étaient mis au point par des linguistes et des informaticiens qui s'efforçaient de programmer l'ordinateur afin qu'il imite le fonctionnement d'un traducteur tel qu'ils se l'imaginaient, c'est-à-dire analyser le

¹ L'État canadien s'intéresse activement à la traduction automatique depuis 1963. Au fil des ans, il a financé la recherche dans ce domaine dans divers laboratoires publics et universités. *Portage* est le produit le plus récent de cette longue démarche.

texte de départ pour s'en faire une quelconque représentation utile, puis transposer cette représentation abstraite en un texte équivalent de la langue d'arrivée. Pour y parvenir, ces chercheurs devaient produire d'impressionnants dictionnaires et précis grammaticaux pour l'ordinateur, tâche manuelle titanesque, aussi laborieuse qu'onéreuse. Les systèmes qui en émergèrent n'étaient pas particulièrement robustes : ils échouaient souvent à produire un résultat quelconque parce qu'il s'avère extrêmement difficile de rédiger une grammaire exhaustive et un dictionnaire complet pour quelque chose d'aussi complexe, en perpétuelle évolution, qu'est une langue naturelle. À ces difficultés, il faut encore ajouter celles tout aussi redoutables de compiler des dictionnaires bilingues et des grammaires contrastives capables d'établir un pont adéquat entre les deux langues.

En 1988, un groupe du T.J. Watson Research Centre d'IBM avança l'idée que la traduction automatique pourrait s'effectuer par apprentissage statistique. Quelques membres du groupe avaient participé antérieurement à la création de systèmes statistiques de reconnaissance de la parole – autre problème complexe tenant depuis longtemps en échec l'approche basée sur des règles.

Contrairement à la TABR, la TAS ne repose pas sur des règles grammaticales de formulation ni sur les entrées d'un dictionnaire. Au lieu de se fier aux intuitions du linguiste concernant un langage, ces systèmes puisent leurs connaissances linguistiques directement dans un vaste corpus de traductions. L'application d'algorithmes d'apprentissage machine aux textes permet d'estimer automatiquement la probabilité qu'un mot ou une phrase X dans la langue d'origine se traduise par le mot ou la phrase Y (ou W , ou $Z...$) dans la langue d'arrivée. Le système de TAS comporte deux grands éléments linguistiques, tous deux de nature probabiliste : un module traductionnel qui correspond grosso modo au dictionnaire bilingue du système TABR et un module linguistique qui garantit la fluidité du texte résultant en maximisant la probabilité de chaque terme W de la traduction, selon les choix retenus pour les n termes précédents.

De nombreux autres modules statistiques et un algorithme de base appelé « décodeur » effectuent les recherches afin de restituer le meilleur texte possible dans la langue cible, par application de ces ensembles probabilistes de connaissances linguistiques au texte d'origine. Pour revenir à la définition initiale donnée plus haut, on pourrait dire en quelque sorte que le module traductionnel veille à la fidélité sémantique du texte final, et le module linguistique, à sa rectitude grammaticale.

Avantages de la TAS

Comparativement à la TABR, la TAS présente plusieurs avantages majeurs, le premier étant sa **polyvalence**. En effet, avec la traduction automatique basée sur des règles, chaque fois que l'on souhaite ajouter une nouvelle langue T , il est nécessaire de mobiliser des linguistes informaticiens spécialisés dans cette langue et des lexicologues bilingues qui connaissent bien la paire $S > T$, pour autant qu'on puisse en trouver.

La TAS, en revanche, applique les mêmes algorithmes d'apprentissage machine à n'importe quelle paire de langues, ou presque. Une personne possédant un système de TAS anglais > français qui décide de s'attaquer à une traduction de l'anglais à l'espagnol (ou du portugais au bulgare, et ainsi de suite) peut le faire sur-le-champ, à condition – évidemment – qu'elle dispose du matériel voulu pour réaliser l'apprentissage machine. Par là, on entend simplement un nombre suffisant de documents correctement traduits entre les deux langues concernées². Par conséquent, la TAS diminue le temps, les efforts et l'argent investis dans le développement de nouveaux systèmes de traduction automatique de plusieurs ordres de grandeur. Ce qui exigerait des mois et de nombreuses centaines

² Combien de textes bilingues faut-il pour former un système de TAS? Les spécialistes ont tendance à dire que, plus le corpus est grand, meilleurs seront les résultats. Néanmoins, un corpus de taille minimale s'avère nécessaire au départ pour que les probabilités qui en émanent soient fiables. On cite souvent comme chiffre une centaine de milliers de phrases dans les deux langues, soit approximativement un million de mots.

d'heures-personnes de travail avec la traduction automatique à base de règles peut désormais être réalisé (littéralement) du jour au lendemain.

Un autre avantage appréciable de la TAS, par rapport à la TAR, est sa **robustesse**. Avec la TAS, chaque phrase dans la langue d'origine peut être restituée de multiples façons dans la langue d'arrivée, bien que certaines propositions soient habituellement beaucoup plus probables que d'autres. En conséquence, les systèmes de TAS produisent toujours un résultat. Sur ce plan, ils sont donc naturellement plus robustes que les systèmes de TAR plus anciens qui, souvent, dépendent d'une analyse grammaticale complète de la phrase d'origine.

Évidemment, cette robustesse présente peu d'intérêt si la qualité des traductions obtenues laisse trop à désirer. Or, on a prouvé, et de manière concluante, dans les concours internationaux ouverts comme ceux organisés par l'armée des États-Unis et le National Institute of Standards and Technology (NIST) que les systèmes de TAS ont tendance à engendrer des traductions de **qualité supérieure** à celles obtenues avec leurs contreparties TAR. Les meilleurs systèmes de TA actuels s'appuient tous sur l'apprentissage statistique, et la TAS est à la base même du **système *Portage* du CNRC**.

Aujourd'hui, les systèmes statistiques dominent totalement le panorama de la traduction automatique. Quelques-uns parmi les plus connus, comme *Google Translate* et *Microsoft Translator*, sont disponibles gratuitement sur Internet, ou peuvent être exploités pour une somme très modique. D'autres, de source ouverte (la plupart reposent sur *Moses*³), sont offerts par des distributeurs qui les adaptent aux besoins de leurs clients. Face à ces multiples possibilités, celui qui songe à la traduction automatique serait en droit de se demander « Pourquoi me donner la

³ *Moses* est un système de TAS complet, créé initialement sous la direction de Philipp Koehn. Le système devait constituer l'équivalent universitaire des systèmes brevetés par les grandes sociétés comme Google et Microsoft. Soutenu financièrement par la Commission européenne, *Moses* a récemment été élargi pour accepter de nombreuses paires de langues en usage en Europe.

peine d'adopter le système *Portage*? », qui n'est pas un gratuiciel. Pourtant les raisons ne manquent pas.

Portage est l'un des meilleurs systèmes de TA par apprentissage statistique actuellement disponibles dans le monde. Les résultats obtenus par ce système lors des récents concours du NIST l'ont prouvé de manière éclatante. En effet, *Portage* a coiffé tous les autres systèmes au poteau pour les traductions du chinois à l'anglais, et s'est classé bon deuxième pour celles de l'arabe à l'anglais – un véritable exploit! L'équipe *Portage* du CNRC a été invitée à deux reprises à participer à d'ambitieux et prestigieux projets financés par la DARPA, la Defense Advanced Research Projects Agency, du gouvernement des États-Unis, de concert avec des laboratoires renommés tels Raytheon BBN et RWTH (Allemagne). Parallèlement, le groupe a collaboré activement avec d'importants partenaires commerciaux comme *Systran*, entreprise qui brille au sommet, sur le marché de la traduction automatique, depuis des décennies.

Beaucoup d'utilisateurs éventuels de la traduction automatique redoutent – et la crainte est légitime – que la sécurité de leurs documents soit compromise, et hésitent soit à expédier ceux-ci hors du pays pour les faire traduire, soit à les télécharger dans le nuage par Internet. Plus rares encore sont ceux qui accepteraient de voir leurs documents employés pour perfectionner le moteur TA d'une autre entreprise, sur lequel ils n'exerceraient aucun droit⁴. *Portage* offre une solution simple à ces préoccupations. Normalement, le système est installé sur un serveur local, dans les locaux sécurisés de l'utilisateur. Personne d'autre n'y a accès, pas plus qu'aux documents traités. Comme nous le verrons ci-dessous, l'utilisateur pourra facilement intégrer le logiciel à sa plateforme de traitement des documents existante.

⁴ C'est notamment le cas de *Microsoft Translator*, bien que Microsoft insiste sur le fait que les textes téléchargés sur ses serveurs de traduction ne puissent être récupérés ni reconstitués par qui que ce soit.

Possibilités commerciales associées à *Portage*

De ses origines à la fin de la Guerre froide, la traduction automatique était surtout employée par l'armée et les services de renseignement qui finançaient aussi principalement la recherche et le développement dans ce domaine. En effet, la masse de documents traitée par ces services dépassait, et de loin, la capacité des traducteurs disponibles. Le seul espoir de remédier à pareille situation consistait à automatiser la traduction.

Des considérations d'ordre géopolitique ont par ailleurs fixé les principales langues étrangères pour lesquelles ont été développés les systèmes de TA. En d'autres termes, ce n'est pas un hasard si la première illustration de traduction automatique donnée à New York, en 1954, portait sur une conversion du russe à l'anglais. Les sommes faramineuses injectées par le gouvernement américain dans l'élaboration d'un système de TA du vietnamien à l'anglais durant les années 1960, en pleine guerre au Vietnam, puis du farsi à l'anglais, dans les années qui ont précédé et suivi la destitution du shah d'Iran ne sont pas non plus accidentelles⁵. Dans ces cas de figure, la traduction automatique devait répondre à un besoin pressant au niveau de la sécurité nationale.

L'armée et les services de renseignement continuent d'exploiter largement la traduction automatique aux États-Unis. Depuis quelques années, leur dépendance envers cette technologie s'est même accrue de façon draconienne, car en plus de ce qui s'écrit, on surveille maintenant ce qui se dit, ce qui se filme et ce qui se bavarde sur les médias sociaux. Le moins que l'on puisse dire est que les attentats du 11 septembre 2001 ont avivé l'intérêt du département de la Défense américaine pour la traduction automatique, l'incitant à injecter de lourdes sommes dans la R-D pertinente et à mettre sur pied les concours internationaux mentionnés précédemment.

⁵ Logos Corporation a été la principale entreprise à profiter des sommes investies par les États-Unis dans du vietnamien et du farsi.

Pour des motifs analogues, l'arabe est subitement devenu le centre d'attraction, suivi de près par le chinois. Les organismes de financement ont insisté sur la rapidité avec laquelle les systèmes devaient absolument être développés, à partir d'une documentation moins volumineuse et plus variée. On peut donc avancer sans grand risque de se tromper qu'une bonne partie des progrès remarquables réalisés récemment dans le domaine découlent des fonds engloutis par l'armée et les services de renseignement dans cette technologie.

Il importe de souligner que, dans un tel contexte, la traduction automatique sert principalement à **glaner de l'information**. Même si les traductions obtenues s'avéraient parfaites (ce qui n'est certes pas le cas), aucun service de renseignement n'aurait le temps de parcourir une documentation aussi volumineuse. La traduction automatique n'est alors qu'un maillon dans une plus vaste chaîne ayant pour but de **passer au crible** l'information, bref de trier une quantité colossale de documents en langue étrangère afin de dénicher ceux qui présentent le plus d'intérêt, pour ensuite faire traduire ceux-ci par des humains. Dans une telle optique, les systèmes de TA existants fonctionnent à merveille.

Fait digne de mention, l'armée et les services de renseignement ne sont plus les seuls à s'intéresser à la collecte des informations présentes dans les documents en langue étrangère. Consécutivement à la mondialisation du commerce, les entreprises souhaitent de plus en plus savoir ce que les gens pensent de leurs produits et de leurs services sur les marchés de la planète. Depuis quelques années, une nouvelle industrie du **renseignement commercial** a donc vu le jour. Cette industrie vit de la collecte et de l'analyse de masses de données volumineuses, principalement récoltées dans les médias sociaux, dont l'importance ne cesse de croître, et qui aident, soutient-on, les entreprises à adapter et à peaufiner leurs stratégies commerciales.

Il y a quelques années, les grandes multinationales étaient pratiquement les seules à faire affaire sur les marchés qu'on dit « étrangers ». À présent, de plus en plus de PME s'y aventurent à leur tour, et elles aussi désirent savoir rapidement ce qu'on

pense d'elles dans d'autres langues. Les sociétés spécialisées dans le renseignement commercial sont contraintes de se tourner vers la traduction automatique pour satisfaire la demande relative au traitement et à l'analyse de textes multilingues sur une grande échelle. De fait, l'équipe *Portage* du CNRC a réalisé quelques travaux exploratoires sur l'analyse translinguistique des sentiments avec l'entreprise canadienne MediaMiser⁶.

Application de la traduction automatique à la diffusion

Tel qu'il a été indiqué précédemment, lorsqu'ils auparavant l'information, les services de renseignement militaires et commerciaux se servent essentiellement de la traduction automatique aux fins d'**assimilation**, soit pour aider les gens à lire et à saisir en partie le contenu d'un texte rédigé dans une langue étrangère. Mais qu'en est-il de l'utilisation de la traduction automatique dans un but de **diffusion**, c'est-à-dire se servir du système pour produire et éventuellement publier un document dans une autre langue? Après tout, cela n'est-il pas le but premier et ultime de la traduction automatique?

Depuis sa naissance, on a à maintes reprises tenté d'employer la traduction automatique à cette fin, au public comme au privé. Aux États-Unis, des entreprises comme Systran, Weidner et Logos (pour ne mentionner qu'elles) ont expressément été fondées dans ce but durant la seconde moitié du siècle dernier. Outre Atlantique, où les multiples langues officielles de l'Union européenne ont engendré une demande énorme du côté de la traduction, la Commission européenne a investi considérablement d'argent et d'efforts dans l'adaptation d'un système commercial de traduction automatique susceptible de l'aider à subvenir à ses besoins avant de se lancer dans un ambitieux programme de développement bien à elle⁷.

⁶ Par « analyse des sentiments », on entend la catégorisation des opinions exprimées dans un document au moyen d'un ordinateur.

⁷ Le programme, baptisé Eurotra, a fonctionné de 1978 à 1992. La CE finance toujours de nombreux projets de R-D en traduction automatique et sur des technologies connexes.

Au Canada, le Bureau de la traduction a abondamment testé plusieurs systèmes de TA commerciaux pour déterminer si cette technologie l'aiderait à mieux composer avec l'escalade de la demande pour ses services, tout en réduisant parallèlement ses coûts. La plupart de ces tentatives n'ont rien donné. Jusqu'à tout récemment, la traduction automatique n'avait servi que très peu et de loin en loin à la diffusion.

Bien qu'il soit difficile d'obtenir des chiffres fiables, il est indéniable que la part du marché mondial de la traduction détenue par la traduction automatique était jusqu'à ce jour très modeste, pour ne pas dire infime. La raison pour cela est fort simple : de manière générale, la qualité des textes issus de la traduction automatique n'est pas assez bonne pour qu'on en autorise la publication ou pour qu'on les exploite de façon rentable. Même lorsqu'on combine la traduction automatique à une personne qui révisera, corrigera et améliorera méthodiquement le travail après coup, la rentabilité n'est pas au rendez-vous, à moins que le premier jet produit par la machine n'atteigne un certain degré de qualité. Sinon, corriger ce que propose l'ordinateur exige plus de temps qu'il en faudrait pour traduire le document d'emblée.

Le système *Météo* du Canada, spécifiquement conçu pour traduire les [bulletins météorologiques](#) quotidiens d'[Environnement Canada](#), est l'un des rares cas où la traduction brute réalisée par ordinateur parvient à une qualité suffisante pour être diffusée. Néanmoins, le succès du système est attribuable au fait qu'il a précisément été créé pour les bulletins météo, qui emploient une terminologie très précise et particulière. Or, les langagiers doivent le plus souvent composer avec des textes fort variés, beaucoup plus ardues à traduire que des bulletins météorologiques pour la plupart, et rien, jusqu'à tout récemment, ne les incitait à lorgner du côté de la traduction automatique pour les aider à produire les documents de qualité réclamés par leurs clients. Cette vérité s'applique aussi à la majorité des services de traduction internes des grandes sociétés.

L'avènement du nouveau paradigme statistique en traduction automatique est sur le point de changer la donne. Comme on a pu le lire plus haut, les systèmes de TAS ont prouvé de façon objective et convaincante qu'ils sont capables de fournir de

meilleures traductions que celles des systèmes de TABR. Par ailleurs, la facilité avec laquelle on produit les **systèmes de TAS spécialisés** autorise désormais le développement de systèmes adaptés à des textes d'une nature, d'une terminologie et d'un style particuliers, peu importe le nombre de clients ou de ministères desservis. Là aussi, on a fait la preuve que ces systèmes spécialisés produisent des traductions de meilleure qualité que les systèmes généralistes du genre *Google Translate*.

L'attitude et les attentes de la population à l'endroit de la traduction automatique ont sensiblement changé eux aussi, ces dernières années, aspect aussi important que les progrès réalisés par la technologie eux-mêmes. Il y a peu encore, beaucoup de ceux qui se tournaient vers la traduction automatique étaient persuadés que l'ordinateur les aiderait à se départir de traducteurs lents et onéreux tout en récoltant un produit de qualité, simplement au dé clic d'un bouton. Après tout, si l'ordinateur a permis à l'homme de poser le pied sur la lune et réussit à battre les plus grands maîtres aux échecs, pourquoi ne pourrait-il automatiser cette démarche antique qu'est le passage d'une langue à une autre⁸? Nous ne sommes plus si naïfs aujourd'hui, et on le doit en grande partie à l'expérience que les sites Web, aisément accessibles, comme *Google Translate* et *Bing Translator* ont fait vivre au public.

Non, le traducteur ne disparaîtra pas aussi vite que le tisserand d'autrefois. Cela dit, la qualité des textes obtenus avec la nouvelle génération de systèmes de traduction automatique par apprentissage statistique semble avoir atteint le point où l'on peut désormais envisager un **partenariat productif** entre la machine et un correcteur humain. Cette évolution, et la demande qui ne cesse de grandir, de même que les fortes pressions exercées par le marché en vue d'une réduction des coûts et d'un raccourcissement des délais, incitent de nombreux langagiers et services de traduction internes à voir la traduction automatique d'un œil neuf.

⁸ Oui, effectivement, pourquoi pas? La question mériterait une réponse plus élaborée que l'autorise malheureusement l'espace dont nous disposons.

Le paragraphe précédent mentionnait une exposition de plus en plus grande du public à la traduction automatique, avec le changement d'attitude que cela implique. Pareil changement s'observe également chez les traducteurs. Auparavant, beaucoup de membres de cette profession craignaient de voir disparaître leur gagne-pain à cause de la traduction automatique. Ils sont nettement moins nombreux aujourd'hui, car ils ont vu de leurs propres yeux ce que cette technologie peut et ne peut réaliser, et ils savent que la demande, en poursuivant son escalade, dépasse considérablement la capacité collective de leur profession.

À présent, les traducteurs voient plutôt dans la traduction automatique un instrument inespéré qui leur permettrait d'accroître leur rendement et, dans certains cas, les soulagerait de la fastidieuse corvée que représentent les textes les plus répétitifs et les moins intéressants⁹. Ceux qui gèrent des services de traduction ne regardent plus davantage la traduction automatique du même œil. Dans le passé, beaucoup s'étaient empressés de tester la technologie dans l'espoir qu'elle amènerait des économies en remplaçant les autres outils et ressources employés jusque-là par les traducteurs. Plus maintenant. Un nombre croissant de ces gestionnaires ont compris que la traduction automatique ajoute une corde à l'arc du traducteur, qu'elle augmente son arsenal au lieu de se substituer aux outils sur lesquels il a appris à compter et dont il se sert efficacement.

Parmi les technologies les plus populaires figure sans aucun doute la mémoire de traduction, système qui détecte et récupère automatiquement les phrases déjà traduites. Si beaucoup de traducteurs se sont d'abord opposés à l'introduction de tels systèmes, la vaste majorité y recourt maintenant tous les jours et ne s'en priverait plus que de mauvaise grâce. Mais qu'en est-il des nombreuses phrases jamais encore rencontrées, pour lesquelles la mémoire de traduction n'a rien à proposer? Une façon efficace de combiner les deux technologies en pareil cas consiste, pour le système de TA, à incorporer une traduction automatique à l'éditeur

⁹ À dire vrai, il faut y voir le motif à l'origine du système Météo : l'ennui qui assommait ceux contraints de traduire des bulletins répétitifs à mourir.

de la mémoire de traduction, dans une cellule cible qui sans cela resterait vide¹⁰. Inutile d'ajouter que le traducteur peut, à sa guise, accepter, modifier ou rejeter ce que l'ordinateur lui propose, comme il le fait déjà avec les propositions soumises par la mémoire de traduction. Si le système de TA est basé sur un corpus adéquat, correspondant au domaine du texte d'origine, il y a alors de fortes chances que la version suggérée renferme au moins quelques éléments que le traducteur réutilisera à son avantage.

L'amalgame de la traduction automatique et de la mémoire de traduction est d'autant plus naturel que les données d'apprentissage idéales pour « entraîner » le système de TAS sont précisément celles contenues dans la mémoire de traduction. Un autre point important que la mémoire de traduction et la TAS ont en commun est que ces deux systèmes s'améliorent avec le temps. En effet, avec l'addition de nouvelles traductions sanctionnées par l'être humain à la base de données, le système de mémoire de traduction devrait trouver de plus en plus de phrases qui se recoupent; et en même temps, le système de TAS disposera de données plus nombreuses pour mieux estimer les probabilités de traduction et mieux couvrir le domaine pertinent.

Les progrès récents – impressionnants – réalisés en TAS et une association aussi productive des technologies traductionnelles expliquent pourquoi un nombre croissant de services de traduction et de langagiers ne font pas qu'envisager, mais adoptent la traduction automatique. Selon le bulletin de janvier 2015 de la TAUS (Translation Automation Users' Society), quarante pour cent des traducteurs recourent déjà à la traduction automatique d'une manière ou d'une autre. Dans le 243^e numéro du très respecté *Tool Box Journal*, Jost Zetsche cite les données de Memsorce, un système de TA en nuage auquel se sont inscrits au-delà de trente mille utilisateurs. Ces données suggèrent que plus de la moitié des clients de Memsorce consultent régulièrement les systèmes de TA en ligne gratuits offerts par *Google Translate* et *Microsoft Translator*. Au Canada, deux des plus grands

¹⁰ Ou la copie du segment d'origine dans la cellule cible.

cabinets de traduction du pays recourent désormais quotidiennement à *Portage* après avoir constaté que le système augmentait leur productivité, bref leur épargnait temps et argent sans que la qualité ou la confidentialité des textes de leurs clients en souffrent pour autant.

Une pléthore d'applications

Jusqu'à présent, nous n'avons parlé que de deux usages possibles, et diamétralement opposés, de la traduction automatique : à une extrémité du spectre, la traduction automatique brute, employée essentiellement par les services de renseignement militaires et commerciaux pour glaner de l'information, et, à l'autre bout, la traduction automatique révisée, surtout utilisée par les services de traduction et les langagiers aux fins de diffusion. Entre ces deux pôles cependant se situe une vaste gamme d'applications que nous nous bornerons ici à mentionner.

L'essor phénoménal des médias sociaux et l'engouement pour le cybercommerce ont indirectement engendré une nouvelle demande, colossale, en traduction. En effet, chaque jour, des millions de gens publient commentaires et analyses sur les médias sociaux ou le site Web des entreprises, dans une langue peut-être incompréhensible à ceux qui aimeraient en prendre connaissance. Bien sûr, ces derniers peuvent copier-coller le contenu affiché par les utilisateurs dans un système de TA en ligne gratuit tel *Google Translate*, mais des sociétés ont constaté qu'elles pourraient offrir à leurs visiteurs une traduction plus digne de foi en s'associant à un développeur pour créer un système à moitié spécialisé, adapté à la nature particulière de leurs documents. TripAdvisor a choisi cette solution avec SDL. D'autres sites en ligne populaires, tel eBay, ont pour leur part investi dans le développement d'une technologie de traduction automatique qui leur est propre. Actuellement, les textes affichés n'ont le plus souvent qu'une vie très brève, et il y a peu d'utilité à réviser avec minutie les textes restitués mécaniquement. L'utilisateur n'obtient donc généralement qu'une traduction automatique brute. Celle-ci lui suffit habituellement pour décider s'il fréquentera ou pas tel hôtel ou tel restaurant, ou achètera ou pas un produit quelconque. Avec la solution retenue par eBay

néanmoins, l'utilisateur qui le désire interagit avec le fournisseur en entamant une conversation nourrie en partie par la traduction automatique.

Microsoft a instauré une application légèrement différente de traduction automatique sans vérification en vertu de laquelle l'entreprise actualise en permanence les énormes bases de connaissances en ligne qui appuient la multitude de produits offerts. Recourir à des traducteurs pour cela serait manifestement impensable. C'est pourquoi Microsoft a mis son propre système de TA à la tâche. Bien que le contenu des bases de connaissances soit initialement rédigé en anglais, la clientèle de Microsoft peut le consulter dans sept langues (au dernier recensement) et espérer trouver la réponse à ses questions dans sa langue maternelle. L'entreprise a sondé ses clients pour connaître leur degré de satisfaction. Fait intéressant, ceux qui effectuent une recherche dans la base de connaissances en anglais semblent aussi satisfaits que ceux dont les réponses ont été traduites mécaniquement.

Enfin, on peut évidemment atteindre des niveaux de révision qui se situent entre, d'une part, la traduction automatique brute, sans correction, et, d'autre part, la traduction automatique révisée visant un degré de qualité similaire à celui d'une traduction humaine. Compromis courant, la « postédition légère » suppose que le réviseur corrige uniquement ce qui empêcherait l'utilisateur de saisir le sens du texte. Les traductions de ce genre seront donc émaillées de fautes de grammaire mineures et d'impropriétés stylistiques pourvu que le lecteur comprenne l'essentiel de ce qu'il lit. On recourt souvent à la traduction automatique avec postédition légère pour les communications internes comme le fait, pour ne citer que ce cas, la Commission européenne avec ses notes de service internes et l'ébauche de ses documents.

Inutile de dire que le logiciel *Portage* du CNRC s'adapte aisément à chacune des situations préalablement décrites.

Sommaire

La traduction automatique a connu des progrès remarquables ces dernières années, surtout depuis l'avènement du nouveau paradigme statistique. Le logiciel *Portage* du CNRC figure parmi les meilleurs systèmes de TAS au monde, ainsi qu'on a pu le constater à maintes reprises dans divers concours internationaux. Des importants langagiers du Canada y recourent déjà pour accroître leur rentabilité, alors que des entreprises d'autres secteurs en font présentement l'essai, au même titre que l'armée et les services de renseignement des États-Unis.

Le CNRC souhaite ardemment étendre l'utilisation de *Portage* à un plus grand nombre de partenaires, et le système pourrait engendrer des avantages appréciables dans plusieurs secteurs. *Portage* n'est pas que le ne plus ultra de la traduction automatique, c'est aussi un système moins coûteux que la plupart de ceux offerts par les principaux concurrents du CNRC.

Les gens qui aimeraient en apprendre davantage sur *Portage* – assister à une démonstration pratique ou en savoir plus sur les droits associés aux licences d'exploitation et les possibilités de collaboration, par exemple – sont invitées à communiquer avec le CNRC.

Pierre Charron, chef, Relations avec les clients

Courriel : pierre.charron@nrc-cnrc.gc.ca

Tél. : +1 613-990-0336