

Entre comparaison et raison : la qualité de la traduction automatique

Une comparaison détaillée des résultats de la traduction automatique statistique avec la traduction humaine met en évidence les points forts de la traduction humaine.

Par Éric Poirier, docteur en traduction et traducteur agréé

Depuis le début des années 2000, les modèles de traduction automatique statistique (TAS) connaissent des progrès importants dont les plus récentes méthodes sont expliquées dans *Statistical Machine Translation*¹, de Philipp Koehn, professeur à l'université Johns-Hopkins et figure de proue de la TAS. On retiendra surtout que ces systèmes utilisent comme point de départ d'immenses corpus de textes parallèles (déjà traduits par des professionnels) sans lesquels l'approche statistique ne produirait rien de valable. Pour les entreprises, les services de traduction automatique de Google Traduction, offerts gratuitement dans plus de 100 langues, se présentent aujourd'hui comme un outil d'optimisation de site Web. Mais qu'en est-il de la qualité de ces traductions d'un point de vue professionnel?

Dans un article publié en 2012 dans *L'Actualité langagière*², André Guyon, expert-conseil en technologies de l'information et traducteur autonome, constate que « la traduction automatique donnera souvent des phrases d'aussi bonne qualité, voire meilleures, que celles de la mémoire de traduction quand on a affaire à des correspondances floues de haute qualité (pourcentage de correspondance élevé) ». Bref, la qualité semble être au rendez-vous, au moins dans le cas de certains types de correspondances. Et si les progrès techniques continuent leur évolution (rapide) des dernières années, on peut envisager une amélioration continue et sensible des résultats produits par la machine. Que peut-on prévoir alors des différences récurrentes qui subsisteront entre les traductions des systèmes de TAS et les traductions professionnelles? Cette question se pose avec d'autant plus d'acuité qu'on trouve de plus en plus d'études et de publications qui portent sur la comparaison entre les deux modes de traduction.

Un constat qui peut surprendre

La plupart des chercheurs et des spécialistes s'entendent sur le fait que la traduction machine de type probabiliste (TAS) est principalement défaillante sur le plan de la formulation en langue d'arrivée. Ce qui semble étonnant, par contre, c'est que le problème récurrent (le plus fréquent et le plus important en nombre) n'est pas celui de la traduction proprement dite, c'est-à-dire le résultat du transfert du sens en langue d'arrivée (le terme *résultat* est ici important, car il n'y a pas à proprement parler de transfert du sens) mais celui des relations grammaticales et textuelles entre les mots, termes et expressions transférés en langue cible. Cette situation concorde avec l'expérience subjective de certains langagiers selon laquelle les textes contiennent peu de notions, de concepts ou d'éléments de sens entièrement inédits, de sorte que leur traduction, assez bien établie, présente elle aussi un caractère prévisible que les statistiques textuelles permettent justement de mettre au jour. Un autre problème de la TAS, qui est étroitement lié à l'absence d'évaluation de la fluidité de la formulation en langue d'arrivée, est le fait que la TAS donne souvent lieu à des formulations (parties d'énoncés ou de segments) qui sont totalement dépourvues de sens. Même si les traductions humaines ne sont pas totalement exemptes de ces erreurs, ce qui caractérise la TAS c'est la récurrence et donc l'importance de ces problèmes dans les textes traduits.

Ces conclusions concordent tout à fait avec les résultats d'une étude réalisée en 2013 par l'auteur de ces lignes sur les ruptures lexicogrammaticales (les propriétés formelles) de la traduction d'un texte produite (et révisée) par des traducteurs professionnels qui sont mises en évidence par comparaison avec la traduction d'un système de TAS (*Google Traduction, en l'occurrence*)³.

L'analyse et la catégorisation des différences textuelles constatées systématiquement dans la comparaison des segments des deux modes de traduction permettent de dégager trois principaux types de ruptures de la traduction humaine avec la traduction machine parmi une dizaine de catégories de ruptures entre les segments traduits automatiquement et les segments traduits par des professionnels (fait important, les ruptures ont été mesurées entre les deux traductions et non pas entre le texte source et le texte cible). Le tableau qui suit présente les données quantitatives sur ces trois types de ruptures ainsi que quelques exemples relevés dans la comparaison des deux modes de traduction étudiés : la traduction effectuée par Google d'un document publié par la Banque du Canada et sa traduction officielle diffusée sur le site Web de l'organisme. Le pourcentage indiqué désigne le rapport entre le nombre de ruptures de la catégorie sur l'ensemble des 1593 ruptures relevées dans le corpus. Les exemples fournissent les éléments essentiels du contexte (même si les segments ne sont pas complets) et les ruptures sont indiquées en caractères gras.

Exemples de ruptures entre la TAS et la traduction professionnelle

Types de rupture	Texte source	Traduction Google	Traduction humaine
Variantes codées - 344 sur 1593 (21,6 %)	a) The expected pace of price increases has fallen (3); b) about the outlook; c) Firms affected by (2); d) For the third consecutive quarter; e) sales growth ; f) indicating that; g) to expand their workforces; h) improvements in demand i) however	a) Le rythme attendu de la hausse des prix a baissé (3); b) quant aux perspectives; c) Les entreprises affectées (2); d) Pour le troisième trimestre consécutif ; e) une croissance des ventes; f) ce qui indique que; g) élargir leurs effectifs; h) amélioration prévue de la demande; i) cependant	a) Le rythme prévu d' augmentation des prix a fléchi (3); b) à l' égard des perspectives; c) Les firmer touchées (2); d) Pour le troisième trimestre de suite ; e) une progression de leurs ventes; f) ce qui signifie que; g) augmenter leurs effectifs; h) redressement prévu de la demande; i) toutefois
Variantes fonctionnelles - 301 sur 1593 (18,9 %)	a) Responses to the autumn survey indicate that (2); b) Western Canada; c) Firms reported ; d) suggesting that; e) firms reported an improvement	a) Les réponses de l'enquête de l'automne indiquent que (2); b) l'Ouest du Canada ; c) Les entreprises ont déclaré ; d) ce qui suggère que; e) Les entreprises ont rapporté une amélioration	(a) Les résultats de l'enquête de l'automne révèlent que (2); b) l'Ouest du pays ; c) Les entreprises ont indiqué ; d) ce qui donne à penser que; e) Les entreprises ont observé une amélioration
Postéditions obligatoires - 292 sur 1593 (18,3 %)	a) The balance of opinion on employment edged up in the spring survey (Chart 4). (3) b) Indicators of pressures on production capacity are (2)	a) Le solde d'opinion sur l'emploi a légèrement up dans l'enquête du printemps (Chart 4). (3) b) Indicateurs de tensions sur les capacités de production sont (2)	a) Le solde des opinions concernant le niveau de l'emploi s'inscrit en légère hausse dans l'enquête du printemps (Graphique 4). (3) b) Les indicateurs des pressions s'exerçant sur la capacité de production sont (2)

Dans le tableau précédent, le nombre entre parenthèses renvoie au nombre de ruptures de même type constatées dans le segment cité. Les variantes codées désignent des ruptures d'origine lexicale ou syntagmatique qui impliquent une substitution simple (pas de chassé-croisé, de réduction ou d'expansion dans les segments en cause) dont la relation synonymique est attestée par le dictionnaire en ligne du **Centre de recherche inter-langues sur la signification en contexte (CRISCO)** de l'Université Caen Normandie. Ainsi, les variantes fonctionnelles constituent une classe résiduelle qui réunit des mots de même sens en contexte qui ne sont pas des synonymes attestés dans le dictionnaire consulté. En pratique, les postéditions obligatoires désignent des ruptures qui représentent une faute ou une erreur de type grammatical ou autre dans le texte d'arrivée et qui nécessitent par conséquent une intervention corrective par un locuteur natif de la langue d'arrivée.

Quelques statistiques à l'appui

Une proportion de 21,6 % des ruptures lexicogrammaticales provient de l'emploi par les humains de synonymes codés des mots et expressions plutôt que du correspondant direct « statistiquement significatif » (abstraction faite bien sûr des autres mots déjà présents dans le texte traduit ou qui s'ajouteront à mesure que l'on traduit), tandis qu'un autre groupe de 18,9 % de ces différences provient des variantes fonctionnelles, ce qu'on pourrait appeler des raccourcis pragmatiques et textuels qui découlent de l'interprétation du texte source et du contexte d'énonciation (voir les exemples dans le tableau). Enfin, 18,3 % des différences proviennent des postéditions obligatoires, c'est-à-dire des erreurs de formulation et de grammaire (accords, régime incorrect, mauvaise préposition, etc.) que l'on trouve dans le texte traduit par la TAS mais qui sont absentes de la traduction humaine. À eux seuls, ces trois types de différences entre les deux modes de traduction représentent 58,8 % de toutes les différences constatées entre la traduction machine et la traduction humaine.

Ces données permettent ainsi de chiffrer le constat exprimé précédemment selon lequel les lacunes des systèmes de traduction automatique proviennent principalement d'une mauvaise formulation en langue cible des concepts transférés de la langue source. En partant du principe que la traduction humaine est correcte et stylistiquement acceptable, on peut affirmer que la TAS échoue 58 % du temps en traduction spécialisée (le corpus étudié réunissait des extraits de textes économiques et financiers) par des choix lexicaux incorrects ou maladroits (40 %) ou par des erreurs de formulation en langue d'arrivée (18 %). Les variantes fonctionnelles auxquelles recourent fréquemment les traducteurs rappellent hors de tout doute que la traduction machine ne réfléchit pas (encore) et que seuls les humains peuvent faire des analogies et des inférences avec nos univers

culturels ou spatio-temporels. La dernière rupture de l'exemple a) montre bien que la différence entre les deux traductions peut relever à l'occasion du choix du registre; général en TAS et spécialisé en traduction humaine. Et l'interprétation des différences entre les deux modes de traduction confirme aussi que les critères de qualité en traduction ne sont pas toujours discrets et polarisés.

Dans la comparaison entre la traduction machine et la traduction humaine, il ne faut pas opposer la gratuité des systèmes de TAS aux services tarifés de professionnels mais plutôt aux coûts élevés que risquent d'entraîner les (nombreuses) erreurs de traduction des systèmes.

Dans l'équation complexe de la qualité de la prestation des services de traduction, les risques que présente la traduction automatique et que met en évidence l'étude des ruptures lexicogrammaticales doivent être rapprochés avec les points forts de la TAS, à savoir sa gratuité relative (beaucoup de ressources doivent être consacrées à la constitution des corpus et à l'optimisation des systèmes), sa formidable rapidité d'exécution et, vraisemblablement, son indéniable justesse dans le transfert linguistique, laquelle est bien entendu fonction de la qualité des corpus bilingues mis à contribution en amont du processus. En matière de traduction, un service hautement culturel, dont la réussite se mesure de manières très différentes selon le type de texte à traduire (entre les pôles des textes littéraires et pragmatiques), on constate toujours que la perfection n'existe ni chez l'humain ni chez la machine. C'est en faisant preuve de pragmatisme devant l'inaccessible quête de la perfection qu'il faut envisager la cohabitation des produits et services de TAS avec la traduction professionnelle, cohabitation qui est d'ailleurs devenue réalité. Dans ce sens, comme pour ce qui est de tous les produits et services de l'intelligence artificielle (calculs complexes de théorèmes, traitement de la parole, interprétation d'images, robotique, jeux vidéo, systèmes experts et apprentissage automatique, etc.), la machine ne remplacera jamais l'être humain et il y a de l'intérêt de tous de confier à chacun des tâches dans lesquelles il excelle.

Dans la logique marchande, les traducteurs ont tout intérêt à bien connaître les points forts et les points faibles de la TAS pour mieux mettre en évidence leur valeur ajoutée. Malgré tout le cynisme qu'a pu susciter l'arrivée des services de TAS, il faut envisager cette percée comme une avancée technique et ne pas hésiter à opposer à cette technologie des données objectives issues d'une comparaison systématique avec les résultats optimaux obtenus par des professionnels de la traduction.

Éric Poirier, est professeur agrégé au Département de langues modernes et de traduction à l'Université du Québec à Trois-Rivières, où il enseigne depuis 2011.

¹ Koehn, Philipp (2010). *Statistical Machine Translation*, Cambridge University Press : Cambridge.

² Guyon, André (2012). « Mémoires de traduction et traduction automatique », *L'Actualité langagière*, vol. 9, no 3, p. 26.

³ Poirier, Éric (2013). « Procédés de traduction humaine mis en évidence dans les ruptures lexicogrammaticales avec la traduction d'un système de traduction automatique probabiliste », *Texte et corpus 2013, Actes des 7e Journées Internationales de Linguistique de Corpus = Proceedings of 7th Corpus Linguistics in France Conference*.