

Traduction automatique et biotraduction : le mariage forcé

Christine Breyel-Steiner et Thierry Grass

**Édition électronique**

URL : <https://journals.openedition.org/traduire/2350>

DOI : [10.4000/traduire.2350](https://doi.org/10.4000/traduire.2350)

ISSN : 2272-9992

Éditeur

Société française des traducteurs

Édition imprimée

Date de publication : 15 juin 2021

Pagination : 94-106

ISSN : 0395-773X

Référence électronique

Christine Breyel-Steiner et Thierry Grass, « Traduction automatique et biotraduction : le mariage forcé », *Traduire* [En ligne], 244 | 2021, mis en ligne le 15 juin 2021, consulté le 01 juillet 2021. URL : <http://journals.openedition.org/traduire/2350> ; DOI : <https://doi.org/10.4000/traduire.2350>

Traduction automatique et biotraduction : le mariage forcé



Christine Breyel-Steiner,
Thierry Grass

*Lorsque le coup de tonnerre éclate, il est trop tard
pour se boucher les oreilles.*

Sun Tzu, *L'Art de la guerre* (1078)

Le modèle neuronal et l'apprentissage profond en traduction automatique ont déjà fait couler beaucoup d'encre, alors que la start-up allemande DeepL GmbH, sise à Cologne, n'a que quatre ans d'existence. Vis-à-vis de son service de traduction automatique gratuit et ultra-rapide, les traducteurs professionnels ont trois attitudes contradictoires: il y a tout d'abord les défaitistes pour qui le métier n'a plus d'avenir. À l'opposé, les négationnistes prétendent que la traduction automatique n'égale et n'égalerà jamais la traduction humaine. Entre les deux se situent les traducteurs conscients qu'il faut vivre avec son temps (existe-t-il d'ailleurs une alternative?) et donc travailler avec les nouveaux outils et se plier aux nouvelles méthodes de post-édition définies dans la norme ISO 18587-2017. Dans cet article, nous ne discuterons pas

du bien-fondé de l'une ou l'autre de ces attitudes puisqu'en fin de compte c'est le marché qui décidera. Nous tenterons d'abord d'expliquer en quoi la traduction automatique neuronale est révolutionnaire pour ensuite commenter les réactions des traducteurs à l'aide d'un sondage que nous avons effectué auprès des membres de la SFT en 2019. En effet, pour la plupart d'entre nous, issus de formations en sciences humaines, il est plus difficile d'appréhender des modèles reposant sur des calculs sophistiqués que de se lamenter sur la croissante mathématisation du monde.

La traduction automatique à apprentissage profond : un nouveau paradigme

La traduction automatique (désormais TA) a eu des débuts laborieux et certains se souviennent de la fameuse phrase *The vodka is good but the meat is rotten*, traduction de la version russe du proverbe biblique *The spirit is willing but the flesh is weak* (Hutchins, 1995 : 17-18). Cette phrase, en fait un canular, était censée montrer l'incapacité des systèmes de traduction de première et deuxième générations à traduire le sens. Car c'est bien ce dont il est question quelque 70 ans après les premiers essais : le sens est-il accessible à une machine ? Le Trésor de la langue française informatisé donne de l'intelligence une des définitions suivantes : « Fonction mentale d'organisation du réel en pensées. » Si l'on s'en réfère à cette seule définition, l'intelligence n'est accessible qu'aux êtres humains, alors comment peut-on parler d'intelligence artificielle ? La simulation du cerveau humain à l'aide d'algorithmes ne date pas d'hier : en 1949, le neuropsychologue canadien Donald Hebb posait les fondamentaux des réseaux de neurones modernes (Wang & Raj, 2017 : 7). L'un des apports de Hebb est d'avoir montré que deux neurones en activité simultanément créent ou renforcent leur connexion, de sorte que l'activation de l'un par l'autre sera plus facile à l'avenir. Cela est aussi valable pour des mots : la connexion entre deux unités se renforce à mesure que la fréquence des cooccurrences de ces deux unités augmente. Pour prendre un exemple trivial, sachant que *Bank* en allemand peut se traduire soit par « banc », soit par « banque », on traduira *Paul sitzt auf*

der Bank par «Paul est assis sur le banc» et non «Paul est assis sur la banque»: la traduction neuronale est capable d'inférer que dans un corpus où les occurrences de *Bank* et de *sitzt* sont fréquentes, il faudra traduire *Bank* par «banc» et non par «banque». Cette «mémoire» des cooccurrences est possible à l'aide d'algorithmes d'auto-apprentissage et d'une représentation des unités lexicales reposant sur des vecteurs comportant 500 à 1000 nombres réels, une capacité de calcul inaccessible à l'être humain. Un tel calcul mathématique n'est toutefois pas assimilable à la démarche cognitive de recherche du sens.

Un changement d'échelle

L'expérience de Georgetown, popularisée en 1954, et durant laquelle soixante phrases transcrites du russe en alphabet latin ont été traduites en anglais, marque les débuts de la TA. Les espoirs sont grands, mais les résultats, décevants, ne sont pas très convaincants. Ils suffisent néanmoins pour que des financements publics et notamment militaires soient octroyés au développement de cette nouvelle technologie. Successivement, les rapports ALPAC (Automatic Language Processing Advisory Committee) aux États-Unis (1966) et Danzin (1992) en Europe sont assez sceptiques quant à la possibilité d'arriver un jour à une TA de haute qualité avec le modèle prévalant à ce stade: à savoir, la TA basée sur des règles et ses avatars que sont la traduction directe, la traduction par transfert et la traduction par interlangue. L'horizon paraît donc bouché pour la TA jusqu'au milieu des années 1990.

Dans le courant des années 2000, le modèle statistique reposant sur l'utilisation de corpus parallèles à grande échelle et traduits par des humains va être à l'origine d'un regain d'intérêt des donneurs d'ordre et du public pour la TA avant que, voilà environ quatre ans, le modèle neuronal ne constitue une nouvelle révolution. L'expression *Big Data* revêt ici toute son importance: alors qu'un moteur de TA statistique est construit à partir d'un corpus d'un à cinq millions de segments bilingues alignés, un moteur hybride statistique/neuronal utilise entre 100 millions et plus d'un milliard de tels segments (Koehn/Wiggins, 2020). Il faut aussi prendre

en considération que ce n'est pas seulement la quantité, mais aussi la qualité des segments traduits qui va influencer sur la qualité finale de la TA, d'où la nécessité de corpus particulièrement bien constitués reposant sur des traductions humaines validées. La qualité du moteur de traduction neuronale DeepL est associée à la qualité des traductions humaines présentes dans son corpus de travail et en partie issues de Linguee.

Les corpus parallèles

Les systèmes de TA statistique, de TA statistique factorisée ou encore neuronaux reposent sur de très grandes mémoires de traduction. La traduction a lieu par assemblage de fragments plus ou moins longs et souvent hétéroclites extraits de textes alignés dans les corpus parallèles. Il va de soi que la reconnaissance floue (*fuzzy matching*) ou probabilité de reconnaissance d'un segment a plus de chances de se rapprocher des 100 % dans une mémoire d'un milliard de phrases que dans une mémoire d'un million de phrases. De plus, la TA neuronale considère la phrase entière, ce qui évite la phase d'assemblage de fragments de traduction piochés à divers endroits dans la mémoire de traduction, typique du modèle statistique.

Mais le principal avantage du modèle neuronal est ailleurs : il présente pour nouveauté de procéder par une analyse contextuelle globale dans laquelle, à tous les niveaux d'analyse, le sens des mots sera représenté en fonction de leur entourage. Les mots sont regroupés en ensembles plus larges sémantiquement homogènes baptisés « plongements de mots » (*word embeddings* en anglais), représentations vectorielles de chaque mot du dictionnaire sous forme de nombres réels. L'opération prend non seulement en compte le contexte du mot considéré, mais aussi celui des mots les plus proches sémantiquement (Poibeau/Reboul 2018).

Du fait que la TA neuronale tout comme la TA statistique « se servent » de traductions humaines, il est essentiel que les corpus parallèles soient abondants (plus d'un milliard de mots comme nous l'avons dit). Cependant, il est tout aussi important que ces corpus reposent sur des données de bonne qualité, c'est-à-dire sur des traductions validées et sans erreurs.

D'autre part, comme ces systèmes ont besoin de données très abondantes pour inférer des connaissances, ce peut être un problème pour les langues dites MoDiMEs (les Moins Diffusées et Moins Enseignées) moins bien dotées en matière de corpus et moins représentées sur internet.

L'apprentissage profond

Le *deep learning* ou apprentissage profond constitue un type d'apprentissage par la machine fondé sur des réseaux de neurones artificiels dans lequel de multiples couches de traitement (*layers*) sont utilisées pour extraire des informations de niveau progressivement plus élevé à partir des données. Ou, pour citer Poibeau/Reboul (2018): «Le système, en découvrant progressivement des régularités de façon incrémentale, identifie des groupes de mots linguistiquement liés. Autrement dit, sans que la syntaxe – c'est-à-dire les relations entre les mots – soit directement et explicitement formalisée, le système la reconstitue de lui-même en partie.»

Si le principe est relativement simple, son application demande de très grandes quantités de calculs. Lors d'une traduction, il s'agit d'abord de récupérer des segments de traduction (sous la forme de séquences de mots) dans des corpus parallèles étiquetés. Le réseau neuronal intervient alors pour prévoir la probabilité d'une séquence de mots dans un contexte donné, et l'apprentissage profond utilise plusieurs couches qui calculent simultanément la probabilité de diverses séquences possibles à partir d'informations vectorielles sur les mots. La masse d'informations qui résulte de ces représentations est traitée de façon dynamique par des processeurs graphiques (GPU) de façon analogue à celle d'une image.

Les implications de ce nouveau modèle

L'évolution de la TA neuronale est fascinante, et les systèmes progressent considérablement d'une année sur l'autre. Il suffit pour s'en convaincre de faire traduire le même texte par DeepL à un an d'intervalle. Il y a peu de temps se posait encore le problème de la gestion de certaines terminologies absentes, mais là encore les systèmes évoluent, et

les terminologies médicale, juridique, économique et autres, issues de domaines plus marginaux, sont implémentées par les «petites mains» des industries de la langue, à savoir encore nous les traducteurs, nous qui avons fourni les corpus, c'est-à-dire la matière qui fait fonctionner ces logiciels, sans en avoir reçu de rétribution, cela dit entre parenthèses.

La norme ISO 18587 : post-édition et révision

Cette norme, qui date déjà de 2017 (<https://www.iso.org/standard/62970.html>) et dont le titre exact est «Services de traduction – Post-édition d'un texte résultant d'une TA – Exigences», est passée presque inaperçue à sa publication. Or elle équivaut à la consécration d'une méthode de travail censée désormais s'imposer à tous les acteurs; et si l'ISO 18587 n'est pas destinée à remplacer la norme ISO 17100-2015 «Services de traduction – Exigences relatives aux services de traduction», elle la revisite sur un certain nombre de points. Tout d'abord, une nouvelle terminologie est mise en place: ainsi, la post-édition n'est pas une correction (la correction étant réservée aux traductions humaines), mais un remaniement de traductions automatiques (dites traductions brutes ou ébauches). La post-édition n'est pas non plus une révision, puisque la révision intervient en bout de chaîne avec la relecture et consiste, en vertu de la norme ISO 17100-2015, en une «comparaison du texte source et du texte cible visant à vérifier la cohérence terminologique, le style et le niveau de langue». Selon l'ISO 18587, la révision est désormais l'«examen bilingue du contenu dans la langue cible par rapport au contenu dans la langue source afin d'évaluer son adéquation avec l'objectif convenu»; il n'est donc plus ici question de style.

La norme effectue aussi une distinction entre post-édition complète, «permettant d'obtenir un produit comparable à un produit obtenu par une traduction humaine», et post-édition superficielle, «permettant d'obtenir un produit qui soit simplement compréhensible». La qualification de la traduction comme «produit» est aussi éclairante!

De fait, le but avoué de la post-édition est triple et se place exclusivement sur le plan quantitatif:

- «améliorer la productivité de la traduction;

- améliorer les délais d'exécution ;
- rester compétitif dans un environnement où les clients font preuve d'une demande croissante de TA».

Les implications pour les traducteurs sont multiples et pas toujours bien comprises :

- d'abord, la norme institue la post-édition comme méthode standard de travail ;
- ensuite, elle suppose que le traducteur professionnel connaisse les «défauts» de la TA, ce qui nécessite une vigilance accrue vis-à-vis de textes grammaticalement corrects et apparemment fiables ;
- la pression financière exercée sur les traducteurs s'intensifie du fait de nombreux outils de TAO permettant d'augmenter la productivité : l'attente des donneurs d'ordre est ainsi passée de 2500 à 4000 mots par jour (300 à 500 mots par heure) en l'espace de dix ans ;
- enfin, elle fait peser la responsabilité du texte final sur le traducteur : à lui de détecter les fautes du système... et ce pour une rémunération moindre.

La perception des traducteurs professionnels

Au vu de cette évolution inéluctable de la TA et de la post-édition, il nous a semblé intéressant de connaître et d'analyser leur impact sur les professionnels de la traduction. Le sondage que nous avons réalisé auprès des traducteurs professionnels a été mis en ligne en mars 2019 et adressé aux membres de la SFT. Il s'agissait d'un questionnaire électronique anonyme de dix items, conçu avec LimeSurvey, outil de gestion de questionnaires électroniques.

Nous avons obtenu au total 154 réponses exploitables.

Question 1: Dans quelle tranche d'âge vous situez-vous ?

- Moins de 30 ans : 5 %
- De 30 à 50 ans : 44 %
- Plus de 50 ans : 51 %

Question 2: Depuis combien d'années exercez-vous la profession de traducteur ?

- De 1 à 5 ans : 9 %
- De 5 à 20 ans : 53 %
- Plus de 20 ans : 38 %

Les questions 1 et 2 étaient destinées à mieux cerner les répondants et leur aptitude à adapter leur méthode de travail à la nouvelle donne. Notre hypothèse de départ, selon laquelle les plus anciens (que ce soit par l'âge ou l'ancienneté dans la profession) seraient plus réticents face à la post-édition, reposait sur le constat que les mêmes réticences avaient été exprimées lors de l'arrivée sur le marché des premiers logiciels d'aide à la traduction il y a une vingtaine d'années.

En recoupant les réponses à ces deux premières questions avec les réponses aux questions de fond, nous avons pu confirmer notre hypothèse puisque la grande majorité des sondés entrant dans la catégorie des « anciens » restent sceptiques et très inquiets et ne sont pas prêts à utiliser la TA ou la post-édition ni à s'y former.

Question 3: Utilisez-vous des services de traduction automatique lors de votre travail? Si oui, lesquels?

Les réponses étaient majoritairement négatives puisque seules 41 personnes (soit à peine plus d'un quart des sondés) ont déclaré s'en servir. DeepL est de loin le service le plus prisé de tous (2 personnes ont aussi recours à DeepL dans sa version payante), suivi de Google Translate (7 personnes).

Cette question a également permis de pointer la confusion qui règne encore dans l'esprit de quelques traducteurs entre TA et TAO.

Question 4: Que pensez-vous de la qualité des ébauches de traductions proposées par les services de traduction automatique?

Voici les commentaires les plus fréquemment recueillis auprès des 41 traducteurs qui ont répondu oui à la question 3:

- [Qualité] très variable, du très bon à l'exécration, voire inutilisable (avis partagé par 18 sondés);
- [Qualité] insatisfaisante, mauvaise, non créative (une dizaine de participants au sondage);
- Dépend des services et des combinaisons linguistiques. Certains sont performants comme DeepL, même au niveau du style (6 sondés);
- Tout dépend du texte source. Plus le texte est rédactionnel et « humain », plus le résultat est médiocre (7 sondés);
- Manque de cohérence terminologique (2 sondés);
- Les ébauches de traduction générées « s'améliorent à vitesse grand V ».

Certains vont plus loin encore en voyant dans ces outils des sources d'erreurs potentielles, voire d'aliénation :

«Les systèmes les plus récents donnent parfois une impression superficielle de qualité, ce qui est encore plus dangereux et trompeur.»

«Cela me gêne, on n'a plus la même liberté de penser et de réfléchir, c'est quelque chose comme un conditionnement.»

Question 5 : Connaissez-vous la norme ISO 18587-2017 (Services de traduction - Post-édition d'un texte résultant d'une traduction automatique - Exigences) ?

Seuls 6 % des traducteurs interrogés déclarent connaître cette norme, ce qui semble indiquer à quel point la post-édition n'est pas encore entrée dans le quotidien de la grande majorité des traducteurs en France.

Question 6 : Avez-vous l'intention de vous former à la post-édition ?

16 % des sondés disent être prêts à suivre une formation dans ce domaine ou en avoir déjà suivi une. Cette dernière catégorie regroupe majoritairement des traducteurs qui l'ont fait par nécessité (formation imposée par les agences) ou par «manque de travail classique», ou encore de jeunes traducteurs qui ont suivi une initiation à la post-édition durant leurs études.

«[J'ai l'intention de me former à la post-édition] si la qualité de la traduction automatique est suffisante, si la post-édition me procure la même satisfaction dans mon travail que la traduction aujourd'hui et que le fruit de ce travail est suffisamment rémunéré et apprécié.»

Les raisons pour lesquelles la très grande majorité des traducteurs ne souhaitent pas se former à la post-édition sont multiples :

- perte de créativité ;
- travail peu intéressant, voire ennuyeux ;
- perte de temps, prétexte pour tirer les prix vers le bas.

Question 7 : Quels sont les domaines de traduction pour lesquels les services de traduction automatique vous semblent les plus performants ?

Nombreux sont ceux qui estiment que les résultats de TA ne dépendent pas du domaine, mais de la structure syntaxique du texte source.

«Lorsque les textes sont très codifiés et utilisent un langage simplifié (structure de phrase simple, phrases courtes, terminologie sans ambiguïté), ils sont plus faciles à traduire automatiquement. C'est le cas notamment de certains documents techniques.»

«Les textes "créatifs" ne se prêtent pas à la TA.»

Question 8 : Qu'est-ce qui, à votre avis, justifie encore pour un traducteur professionnel de NE PAS recourir à la traduction automatique ?

Les raisons invoquées par les traducteurs sont multiples : la qualité rédactionnelle, la créativité, l'originalité, garder son propre style, le plaisir de traduire, l'intérêt pour son travail, la conscience professionnelle, l'honnêteté intellectuelle, le respect de la confidentialité, les subtilités terminologiques, etc.

«Les subtilités terminologiques dans des domaines requérant une grande précision et des connaissances pointues de la matière (traduction juridique) ; les considérations stylistiques (textes littéraires ou rédactionnels).»

«L'appauvrissement de ses compétences linguistiques à force de pratiquer la post-édition.»

Les langues MoDiMEs ne sont pas non plus les mieux implémentées, faute de corpus parallèles conséquents : «Cela n'existe pas encore en norvégien à ma connaissance.»

Quelques personnes estiment, pour leur part, que rien ne justifie de ne pas utiliser la TA :

«L'égo de nombre de traducteurs qui ont une vision prétentieuse du métier freine l'adoption de ces outils et la redéfinition de l'avenir avec cet outil.»

«Une sorte de fierté professionnelle... ? Un refus de principe ?»

Trois répondants ont exprimé une critique quant à la formulation même de la question :

«La question me semble posée à l'envers, comme si l'on estimait que la situation normale est le recours à la TA, et qu'il serait anormal de s'en passer.»

«D'où vient ce "encore" ? ! C'est curieux, on dirait que vous évoquez une tendance "unstoppable" !»

«Encore, pourquoi encore ? Ça n'est ni inéluctable, ni une fatalité.»

Question 9 : Pensez-vous que la post-édition change fondamentalement les conditions d'exercice de notre métier ?

Pourquoi?

La grande majorité des sondés ont répondu affirmativement : les traducteurs sont conscients des implications de la post-édition, mais expriment aussi une grande méfiance à son égard :

« Oui, sur le fond et la forme. Notre profession sera (encore) moins bien considérée et donc aussi (encore) moins rémunérée. »

« [La post-édition] contraint notre esprit à suivre un chemin prédéfini. »

« La post-édition dépouille le traducteur de sa condition d'auteur. »

Nous n'avons enregistré que quelques réponses négatives à cette question de la part de traducteurs prêts à accompagner le changement :

« Elle les [les conditions de travail] améliore, car elle [la post-édition] nous permet de nous concentrer davantage sur le texte à traduire, d'approfondir nos recherches de manière plus ciblée, de nous donner des idées de traductions pour des cooccurrences données auxquelles nous n'aurions pas spontanément pensé. »

« Qu'on le veuille ou non, ce sera la réalité de la profession d'ici peu et au lieu de se braquer, il faut anticiper, pour bien prendre le virage, être plutôt force de proposition, et profiter des opportunités offertes. »

Question 10 : Comment le traducteur professionnel peut-il se préparer à l'évolution des conditions du marché pour faire face à la concurrence des prestataires de services de traduction automatique en ligne ?

Les réponses à cette question peuvent être catégorisées comme suit :

1) En misant sur la qualité :

« Se former et proposer des prestations de qualité, meilleures que ce que fait la TA. »

2) En trouvant des niches :

« Proposer des prestations complémentaires qu'aucun système automatique ne peut assurer. »

« Se placer sur des marchés de niche très rédactionnels qui échappent encore à la traduction automatique (presse, édition, etc.). »

3) En changeant de métier à terme :

« J'ai bien peur que le métier de traducteur professionnel soit voué à la disparition... à plus ou moins long terme selon le domaine et la combinaison de langues... »

Conclusion

Le marché mondial de la traduction est estimé à 43 milliards de dollars, d'où l'intérêt de l'industrie. L'enjeu de l'intelligence artificielle, ne nous leurrions pas sur ce point, consiste à substituer des algorithmes non conscients à des humains conscients. Selon les experts de la science et de l'industrie, l'intelligence artificielle sera en mesure d'effectuer toutes les tâches humaines entre 2040 et 2050. Le défi à relever pour pouvoir continuer à exercer notre métier est de savoir anticiper, de favoriser la créativité, la qualité de communication et le travail en équipe. Les formations des traducteurs doivent aussi se diversifier en incluant les techniques d'écriture, la rédaction technique, l'interprétation de liaison, etc.

breyel@unistra.fr
tgrass@unistra.fr

Diplômée de l'ITI-RI de Strasbourg, **Christine Breyel-Steiner** est établie comme traductrice freelance depuis trente ans et occupe parallèlement un poste de professeure associée à temps partiel à la Faculté des langues de l'université de Strasbourg où elle assure des TD de traduction et d'initiation à SDL Trados Studio. Depuis peu, au sein de l'UR 1339 Linguistique, Langues, Parole, elle s'est lancée dans des travaux de recherche portant sur la traduction automatique, son évolution et son utilisation par les traducteurs professionnels.

Professeur en linguistique outillée et terminologie, **Thierry Grass** enseigne l'allemand et la traduction spécialisée à la Faculté des langues de l'université de Strasbourg dont dépend l'Institut de traducteurs, d'interprètes et relations internationales. Ses recherches au sein de l'UR 1339 Linguistique, Langues, Parole sont consacrées à la traduction juridique, à la traduction des noms propres, aux bases de données lexicales, ainsi qu'à la traduction automatique.

SOURCES CITÉES

HUTCHINS John, « *The whisky was invisible, or Persistent myths of MT* », in *MT News International* 11, June 1995, p. 17-18.

POIBEAU Thierry, REBOUL Marianne, « La traduction automatique passe à l'apprentissage profond », *La Recherche*, 28 août 2018, <https://www.larecherche.fr/technologie/la-traduction-automatique-passe-%C3%A0-l%E2%80%99apprentissage-profond>, consulté le 28/04/21.

KOEHN Philipp, WIGGINS Dirk, *Artificial Intelligence and Machine Translation: 2021 and Beyond*, 2020, <https://omniscien.com/webinars/artificial-intelligence-and-machine-translation-2021-and-beyond/>, consulté le 28/04/21.

WANG Haohan, RAJ Bhiksha, *On the Origin of Deep Learning*, 2017, Carnegie Mellon University, https://www.researchgate.net/publication/314093717_On_the-Origin_of_Deep_Learning, consulté le 28/04/21.

Trustpoint.One, *How Artificial Intelligence is Transforming Machine Translation and the Global Language Business*, 2019, <https://trustpoint.one/wp-content/uploads/2019/02/Trustpoint.One-Machine-Translation.pdf>, consulté le 28/04/21.