

## L'algorithme sert-il les traducteurs ? Conditions et contexte de travail avec les outils de traduction neuronale

Claire Larssonneur

Université Paris 8, France

---

### **Are algorithms translator-friendly? Some considerations on the impact of neural machine translation on translator working conditions – *Abstract***

Literary translation is widely seen as incompatible with machine translation, which is most effective on technical content. Yet machine translation tools can in fact be harnessed for editorial content and may improve the ergonomics of the translation process: by offering cognitive support and customization, they can potentially reduce working hours and physical strain. Like all technological innovations, machine translation has the potential to impact the socio-economic conditions of the translation profession. NMT is efficient and less costly than CAT software; however, it may disempower translators, and data exploitation is rife. The translation market appears to be shifting into the hands of a few powerful Internet companies with the financial means to develop language models. As a result, due to changes in online practices, the translation market is breaking down into distinct markets for multilingual content management, APIs and editorial publishing. This raises ethical issues and calls for new forms of regulation.

### **Keywords**

machine translation, ergonomics, economics, ethics, regulation

## 1. Introduction

Évaluer une technologie nouvelle au seul prisme de son ancrage scientifique, plus ou moins performante, plus ou moins novatrice, est tentant mais réducteur. Les moteurs de traduction neuronale, nés des progrès de l'intelligence artificielle, ne font pas exception. L'avènement d'interfaces gratuites de traduction comme Google Translate ou DeepL permet à tout un chacun de traduire rapidement des échanges informels (courriels) ou des contenus de loisir (textes, séries, etc.). L'intégration de modules de traduction automatique (API ou *application programming interface*) au sein même des sites Web ou des applications de notre quotidien (par exemple le logiciel de traitement de texte MS Word) accroît considérablement leur accessibilité, et par là même leur compétitivité internationale. Le nombre de solutions de traduction automatique est en plein essor : dans son enquête 2022 sur les pratiques professionnelles des traducteurs, la Société Française des Traducteurs en recense 26 moteurs et la liste n'est pas exhaustive. Parallèlement, le plurilinguisme des outils s'accroît en continu. Ainsi le moteur de traduction automatique intégré dans la suite MS Office s'enrichit de nouvelles langues à un rythme quasi mensuel : le somali et le zoulou sont ajoutés en mars 2022, puis le basque et le galicien à partir du 12 avril 2022, le féroïen depuis le 25 avril 2022 (Microsoft Translator, 29 mars 2022, 12 avril 2022, 25 avril 2022). En novembre 2022, Microsoft Translator couvre 111 langues, Google Translate 133 langues, DeepL 29 langues ; Naver Papago, un moteur coréen, en couvre 13, Baidu translate (le moteur chinois) 200.

Ces outils, dont on voit combien ils sont appelés à investir notre quotidien partout dans le monde, infléchissent notre conception de ce qu'est une machine, et par ricochet, de l'activité qu'ils appareillent, ici la traduction. Simondon, philosophe de la technique, réfutait

l'idée très répandue que « *l'utilité* » suffit à définir la technique. Plutôt que d'être un simple outil à utiliser, chaque machine concrétise des pensées et des gestes, concentre des savoirs et des savoir-faire, et constitue ainsi une singularité intelligemment agencée, pensée dans un environnement social donné. (Hérisson, 2016, p. 147).

Les algorithmes de traduction ont bien une efficacité sociale, au sens où ils produisent des effets psychologiques, économiques et sociaux sur les pratiques humaines de traduction, qu'elles soient collectives ou individuelles, professionnelles ou grand public, visibles ou invisibles.

Bien sûr, un tel développement suscite la crainte d'un remplacement de l'humain par la machine. Toutefois, le marché de la traduction est complexe et le récit du grand remplacement encore largement une fiction. En effet, la demande mondiale de traduction explose car des personnes ou des institutions qui ne l'envisageaient pas y recourent plus volontiers<sup>1</sup>, mais cette demande accrue est de plus en plus prise en charge par les moteurs, sans ou avec moins d'intermédiation humaine. L'activité se déporte vers d'autres types de prestation : post-édition, transcréation, gestion de projet, computation linguistique. Dans un contexte de mutation rapide du secteur, l'effet des technologies de traduction neuronale sur le travail des traducteurs mérite d'être évalué en détail. Afin de dresser un portrait plus complet et plus nuancé de la situation que le seul affrontement de l'homme et de la machine, on examinera trois dimensions du travail : la question souvent occultée de l'ergonomie de travail, la dimension économique de ces outils, leur coût de production et les gains qu'ils permettent, et enfin les enjeux éthiques associés. Notre champ d'étude portant ici spécifiquement sur la traduction d'édition, nous nous concentrerons sur les moteurs grand public (DeepL, Google Translate, Microsoft Word), qui sont à l'heure actuelle l'option technique la plus accessible aux traducteurs littéraires.

---

<sup>1</sup> En 2022 Statistica estime le poids du marché mondial de la traduction à 56,18 milliards de dollars. La demande de traduction a augmenté de 40% durant l'épidémie de Covid. La part de marché des solutions automatiques s'élèverait à 3 milliards de dollars en 2027 (voir Lim (n.d.)).

## 2. Ergonomie des moteurs neuronaux : un gain indéniable

L'ergonomie est « la discipline scientifique qui vise la compréhension fondamentale des interactions entre les humains et les autres composantes d'un système, et la profession qui applique principes théoriques, données et méthodes en vue d'optimiser le bien-être des personnes et la performance globale des systèmes » dans la définition proposée en 2000 par l'IEA. Avant de revenir sur la convergence supposée des deux objectifs (bien-être et performance), il faut rappeler que l'ergonomie travaille toujours sur trois dimensions : l'engagement du corps ou ergonomie physique, celui de l'esprit ou ergonomie cognitive, et l'engagement institutionnel ou ergonomie de l'organisation.

### 2.1. Un gain de temps, moins de fatigue

Les technologies d'IA ne changent pas fondamentalement la dimension organisationnelle du travail du traducteur, déjà généralement en télétravail et le plus souvent maître de ses horaires et de son rythme de travail. Il en va autrement sur le plan de l'ergonomie physique.

La posture adoptée devant l'écran et les mouvements répétitifs des doigts et du poignet sont en effet susceptibles d'affecter la santé du traducteur, en provoquant par exemple des tendinites ou un syndrome du canal carpien. Passer un texte source dans un moteur de traduction réduit considérablement (parfois entièrement) le travail de saisie du premier jet de traduction, et contribue, avec les technologies de saisie vocale, à soulager le traducteur. Ceci va de pair avec un gain de temps dont l'étendue reste à déterminer. En effet, selon le type de texte et la visée finale de la traduction, il reste plus ou moins de travail de post-édition à effectuer sur le texte pré-traduit automatiquement. Pour des essais, des articles universitaires, ou de manière générale des textes dont le contenu informatif prime sur le style, le gain de temps peut être considérable : jusqu'à près de 50% dans mon expérience. Systran estime le gain de temps pour ses clients entre 30 et 60 % (Communiqué de Systran, 9 novembre 2018). Pour des textes plus créatifs ou qui demandent un travail d'adaptation culturelle, le gain de temps serait plus limité : entre 10 et 30% (Macken *et al.*, 2020). Un autre facteur clef est l'entraînement du traducteur aux techniques de post-édition et sa relation à l'outil : une étude récente montre en effet que les traducteurs qui estiment que les outils neuronaux augmentent leur productivité travaillent effectivement plus vite que ceux qui pensent que ces mêmes outils les ralentissent (Sanchez-Gijon *et al.*, 2019, p. 31). Dans le domaine de la traduction d'édition, le gain de temps n'est pas aussi crucial qu'en traduction technique : une rapidité accrue de 20 à 30 %, c'est important, mais sans doute pas décisif. Enfin et dans la mesure où les moteurs neuronaux sont sujets à des erreurs importantes (contre-sens, hallucinations) que seul l'humain peut corriger, le fantasme d'une traduction quasi instantanée des best-sellers reste ce qu'il est : un fantasme d'éditeur !

### 2.2. Une aide cognitive

Il en va autrement pour l'ergonomie cognitive de la traduction, c'est-à-dire les choix cognitifs opérés par les traducteurs comme la recherche de variantes, les modifications syntaxiques ou encore l'identification des erreurs. La plupart des moteurs proposent également des fonctions de vocalisation des extraits (*text-to-speech*), des dictionnaires, ou même des fonctions plus poussées comme l'extraction du texte au sein d'images (Apple Live Text, Lingvanex). L'étude de Jia *et al.* (2019, p. 71) a précisément montré que la post-édition de textes pré-traduits automatiquement allait de pair avec une réduction significative de la charge cognitive, pour des textes techniques comme pour les textes d'édition, alors que le gain de temps n'était significatif que dans le domaine technique. Une interface comme celle de DeepL permet d'identifier des variantes lexicales pour chaque partie du texte, tout simplement en faisant glisser le curseur. On peut aussi s'inspirer des variantes proposées par plusieurs moteurs. Dans

le cadre d'une expérimentation réalisée le 15 novembre 2021, j'ai ainsi passé des extraits de *The Night Watchman* (Erdrich, 2020) dans DeepL, Google Translate, Microsoft et Lingvanex. Le tableau 1 récapitule les propositions faites pour la séquence suivante dans le texte source : « The government attributed their focus to Indian blood and training in Indian beadwork. Thomas thought it was their sharp eyes—the women of his tribe could spear you with a glance. »

<b>Microsoft Word</b>	Thomas pensait que c'était leurs yeux acérés – les femmes de sa tribu pouvaient vous lancer d'un coup d'œil.
<b>Lingvanex</b>	Thomas pensait que c'était leurs yeux perçants – les femmes de sa tribu pouvaient vous lancer d'un coup d'œil.
<b>Google Translate</b>	Thomas pensait que c'était leurs yeux perçants – les femmes de sa tribu pouvaient vous harponner d'un coup d'œil
<b>DeepL</b>	Thomas pensait que c'était leurs yeux aiguisés – les femmes de sa tribu pouvaient vous transpercer d'un regard.

**Tableau 1.** Traductions via différents outils

Ces quatre propositions restent très proches du texte source, jusqu'à le calquer, mais elles identifient des champs lexicaux qu'on peut réemployer, notamment la figure du harpon qui est très cohérente avec l'univers fictionnel (une réserve indienne où la chasse et la pêche sont des activités importantes). La traduction que je propose réemploie ces suggestions et les retravaille : « Pour Thomas, c'était l'acuité de leur regard. Les femmes de sa tribu vous harponnaient d'un coup d'œil. ». Mobiliser les moteurs ainsi, pour étayer le travail de la variante et parfois dégripper un blocage, peut être très fructueux.

### 2.3. L'enjeu de la personnalisation de l'outil

Qu'il s'agisse des auteurs ou des traducteurs, bien souvent, on identifie l'œuvre de l'esprit à une voix singulière. Or, par définition, les algorithmes proposent les solutions de traduction statistiquement les plus probables : l'opposé d'une voix singulière ! La personnalisation des outils est donc un enjeu important pour les traducteurs d'édition. Au sein des interfaces grand public, on trouve déjà la possibilité de se constituer des glossaires ou des historiques de traduction personnalisés. Toutefois, l'algorithmique reste conçu pour identifier des récurrences. Quelques travaux très récents visent à intégrer à la conception même de l'outil une forme de personnalisation. Une manière de faire consiste à entraîner le moteur sur un corpus de textes d'un traducteur donné ou dans un genre précis : Damien Hansen a tenté cette expérience sur un corpus de fantasy, ce qui demande toutefois des compétences informatiques certaines et un ordinateur robuste (Hansen, 2021). Dans « Towards Modelling the Style of Translators in Neural Machine Translation » (2021), Yue Wang, Cuong Hoang et Marcello Federico, de chez Amazon AI, ont tenté d'infléchir les rouages de l'algorithmique de manière à ce qu'il imite le style d'un traducteur donné. L'objectif est d'identifier les idiosyncrasies stylistiques au moyen d'annotations et de modifier les paramètres pour leur donner priorité. Les équipes d'Alibaba travaillent également sur ce sujet (Lin *et al.*, 2021). Enfin, certains moteurs de traduction, comme LILT, proposent des versions auto-adaptatives, c'est-à-dire qui intègrent les caractéristiques des textes sur lesquels vous travaillez (Martikainen, 2022).

La traduction neuronale n'est pas seulement limitée par sa nature statistique ; elle est aussi décalée des productions authentiques des locuteurs natifs. En effet, comme la plupart des segments alignés sur lesquels sont entraînés les moteurs viennent eux-mêmes de traductions, la traduction neuronale renforce et démultiplie le biais inhérent au texte traduit, en anglais

le *translationese*. Plus on utilise de textes pré-traduits puis révisés, plus le *post-editesse* va se superposer au *translationese* (Daems *et al.*, 2018), ce qui conduit à une forme d'artificialisation et de standardisation de la langue, sur la base des corpus traduits et non plus des productions de locuteurs natifs. C'est ce que souligne l'un des chercheurs vedettes de Google lui-même :

Our empirical findings also raise concerns regarding the effect of synthetic data on model scaling and evaluation, and how proliferation of machine generated text might hamper the quality of future models trained on web-text. (Ghorbani *et al.*, 2021, p. 12)

Enfin, la machine n'est pas le seul élément en jeu et la personnalisation de l'outil passe aussi par une bonne formation des utilisateurs. Le principal écueil en post-édition de moteurs neuronaux est ce que j'appelle « le halo de la machine », à savoir que le fait de travailler sur un texte déjà rédigé dans la langue cible, qui semble fluide et cohérent, court-circuite un certain nombre de réflexes de traduction et empêche une vision critique. Ce phénomène est particulièrement marqué chez les étudiants, qui font trop confiance au texte pré-traduit en omettant de vérifier dans le texte source (Schumacher, 2019 ; Daems, 2016). Or la machine reproduit en effet généralement l'organisation syntaxique des phrases du texte source, ce qui conduit à de très nombreux calques de structure (Martikainen, 2022). Enfin, le dernier point concerne la capacité créatrice des traducteurs eux-mêmes. Les études montrent en effet que la PEMT profite surtout aux étudiants les moins aguerris, leur permettant de faire passer comme acceptables des textes cibles qu'ils peineraient à produire eux-mêmes (Schumacher, 2020, p. 259) : de moins bons traducteurs, qui ne pourraient sans doute pas vivre professionnellement de leurs compétences, parviennent à sortir des textes convenables.

Comme on le voit après ce rapide tour d'horizon technique, la traduction neuronale peut accroître aussi bien la performance que le bien-être physiologique, et il serait dommage que les traducteurs d'édition s'en privent. En revanche, il est important de comprendre que ces moteurs ne proposent pas de traduction proprement dite, mais qu'ils comptent des énoncés probables à partir de corpus gigantesques : le texte cible ainsi pré-traduit est une proposition qui reste à affiner. L'anthropomorphisme de l'intelligence artificielle suggère que ces outils relèvent d'une cognition supérieure, alors qu'il conviendrait plutôt de les ranger du côté du traitement de texte.

### 3. Économie : plusieurs marchés en un

Si on part du postulat que les moteurs neuronaux ne sont pas des traducteurs, mais des outils professionnels, alors on peut s'interroger sur les conditions économiques de production et d'utilisation de ces outils. Sont-ils rentables ? Si oui, pour qui, dans quel marché ?

#### 3.1. L'intérêt pour les traducteurs

Les principaux outils grand public proposent des interfaces en ligne gratuites, limitées généralement à 5 000 caractères. C'est une solution peu pratique pour des textes longs et qui pose des problèmes de confidentialité sur lesquels on reviendra. DeepL, à l'heure actuelle la meilleure solution pour les langues européennes, propose une version professionnelle payante. DeepL Pro offre de pré-traduire intégralement un quota mensuel de documents déjà formatés (Word, Excel, Powerpoint, pdf) et de les restituer dans le même format. À un manuscrit source correspond un manuscrit cible. Un abonnement annuel pour un utilisateur unique, en mai 2022, monte à 79 € pour 5 documents par mois, ou 239 € pour 20 documents par mois. Pour une entreprise qui compte 3 collaborateurs, l'abonnement annuel de DeepL monte à 719 € pour 20 documents par mois par utilisateur. Soit 720 documents par an, au prix donc d'environ 1€ le document : cette offre économique est évidemment très tentante pour des maisons d'édition et de manière générale tout secteur d'activité qui doit produire

de nombreuses traductions (communication, événementiel, culture etc.). Les autres solutions payantes, comme Lingvanex, pratiquent des tarifs similaires d'environ 80 € par an pour un particulier. Ces prix sont très intéressants par rapport à ceux pratiqués par les logiciels d'aide à la traduction, comme Trados dont l'achat monte à plusieurs centaines d'euros<sup>2</sup>, d'autant plus qu'il est très facile de télécharger et d'utiliser ces outils. Tout se fait en quelques clics, sans avoir à apprendre le maniement de nombreuses fonctions.

### 3.2. L'intérêt pour les développeurs

L'offre est donc particulièrement attractive du côté des clients : traducteurs et diffuseurs. Toutefois, le véritable marché de la traduction pour les développeurs des moteurs neuronaux n'est pas le secteur de l'édition, loin de là. Les grandes plates-formes de l'économie numérique (Google Translate, Amazon, Facebook ou encore Apple) intègrent en effet ces outils à titre gratuit sur leurs interfaces ou dans la liste de leurs applications afin de fidéliser l'internaute, de générer du trafic et de créer du contenu. Pour les mêmes raisons, des entreprises comme DeepL ou Lingvanex développent des modules de traduction (API) à intégrer dans des sites ou plateformes commerciaux. La traduction automatique est vendue comme un outil permettant de conquérir de nouveaux marchés, de faciliter la communication au sein d'une équipe internationale, de protéger les données de l'entreprise, et évidemment de collecter et d'exploiter les données client multilingues. Dans le cas de DeepL API, le logiciel est quasiment gratuit pour l'entreprise et la tarification se fait à l'usage, avec un abonnement d'environ 60 € par an puis une facturation de 20 € par tranche d'un million de caractères traduits<sup>3</sup>. Ici, l'accent n'est mis ni sur le style ni sur la fidélité au texte source, mais sur les mots-clés et la fluidité des interactions : l'outil n'est donc pas vraiment calibré pour la traduction d'édition. Enfin pour les plus gros clients, les prestataires de services linguistiques vendent l'accès à un moteur personnalisé, entraîné spécifiquement sur les documents du client, et à un serveur sécurisé : les contrats se négocient en dizaines ou centaines de milliers d'euros pour des périodes de plusieurs années. Juste avant la pandémie, Systran a ainsi signé un contrat de 200 000 € avec l'Autorité française de régulation bancaire, pour cinq ans (Fournier-Outters, 2018). De telles prestations restent hors de portée d'un petit éditeur, mais sont possibles pour les grands groupes d'édition comme Editis.

### 3.3. L'enjeu des coûts de production

En quelques années, le marché de la traduction a été bouleversé par l'apparition des technologies neuronales et de nouveaux acteurs qui investissent des sommes considérables. Sur l'année 2019, 66 % du panel de 50 articles de recherche les plus pertinents sur Google Scholar en matière de traduction neuronale émanaient de laboratoires d'entreprise : Google AI (9), Microsoft (9), Tencent (7), Facebook (2), Samsung (2). En 2021, sur un panel comparable, les grandes entreprises du Net sont toujours à l'origine de 54 % des articles : un peu moins, mais cela s'explique par l'importance des avancées techniques réalisées en 2019, sur lesquelles les chercheurs ont ensuite capitalisé sans nécessairement faire de nouvelles propositions.

Dire que les investissements sont massifs relève de la litote. En 2019, Amazon a consacré 200 millions de dollars à la recherche en traitement des langues naturelles et Facebook a fondé un consortium dédié à cette thématique, AI Language Research (Cimino, 2019) ; en 2022, Meta

<sup>2</sup> En mars 2023, Trados propose aux freelances l'achat du logiciel à 449 euros (remise de 40% sur le prix initial de 750 euros) ou bien un abonnement annuel à 215 euros. La mise à niveau des dernières fonctionnalités pour ceux ayant acheté le logiciel revient à 171 euros. Pour une entreprise, le coût d'achat est de 2 695 euros et la mise à niveau pour un seul poste revient à 860 euros. (Faut-il donner une référence pour ce genre d'information ?)

<sup>3</sup> Cf. <https://www.deepl.com/en/pro/change-plan#developer>

a développé un super-ordinateur qui prend en charge des modèles de langues sophistiqués (Filippone, 2022). Toujours en 2019, Microsoft avait prévu d'investir plus d'un milliard de dollars sur 10 ans, en partenariat avec Open AI, pour développer le modèle de langue GPT-3 (Vincent, 2021), ainsi que ChatGPT, ouvert au public à l'automne 2022.

Car les technologies de traduction neuronales, si elles sont gratuites ou quasi gratuites pour les utilisateurs, sont très coûteuses lors de la phase de conception et d'entraînement des logiciels, et requièrent des ressources hors de portée de la plupart des prestataires de services linguistiques. Le traitement des langues naturelles repose notamment sur la transformation des mots en vecteurs auxquels sont appliquées des techniques semblables à celles qui sont utilisées pour les images. Les travaux des équipes de recherche mobilisent donc en général de 8 à 10 GPU (*graphic processing units*), des puces qui valent plusieurs milliers d'euros chacune (entre 7 000 € et 30 000 €). Ces puces consomment beaucoup d'énergie électrique, et il faut y ajouter la nécessité d'accéder à des serveurs puissants (comme Microsoft Azure), et le coût d'utilisation de la bande passante. Le temps d'entraînement des algorithmes sur ces machines, essentiel à la production d'énoncés de qualité, est donc un facteur décisif. Liu *et al.* (2021) décrivent ci-après leur protocole d'expérimentation, où l'on voit qu'il faut compter sur des centaines de milliers d'itérations :

All Transformer models are first trained by teacher forcing with 100k steps, and then trained with different training objects or scheduled sampling approaches for 300k steps. All experiments are conducted on 8 NVIDIA Tesla V100 GPUs, where each is allocated with a batch size of approximately 4096 tokens. We use Adam optimizer (Kingma & Ba, 2014) with 4000 warmup steps.

Le temps d'entraînement se compte donc en jours, plus souvent en semaines. Dernier point et non des moindres, pour que les moteurs neuronaux produisent des traductions de qualité, il est impératif qu'ils soient entraînés sur des corpus de qualité et de taille conséquente. Parmi les corpus les plus utilisés au sein des équipes de recherche, on peut citer celui du Parlement européen (21 langues, avec des ensembles de 11 à 50 millions de mots) ou les TED Talks (59 langues, de 0,3 à 4 millions de mots). Il y a encore les corpus créés spécialement pour les conférences annuelles de l'ACL (Association for Computational Linguistics) et sur lesquels les chercheurs comparent les performances de leurs innovations : le WMT 2014<sup>4</sup> pour le couple de langues anglais/français contient 36 millions de paires de phrases, le WMT 2019 pour le couple de langues anglais/chinois contient 21 millions de paires de phrases. Les dimensions des corpus d'entraînement pour les algorithmes de traduction n'ont donc rien à voir avec celles des mémoires de traduction des traducteurs indépendants, ni même avec l'œuvre d'un traducteur d'édition, fût-il prolifique.

L'effet de seuil se double d'un accélérateur : en effet, les mêmes technologies d'entraînement des algorithmes sur corpus permettent à de grands opérateurs de développer des modèles de langues (comme GPT-3 et ChatGPT) qui proposent un panel varié de services linguistiques de génération de contenu, en plus de la traduction. La version grand public de ChatGPT permet ainsi, entre autres, d'extraire les mots-clefs d'un texte, d'en faire une synthèse, de classer des informations, de rédiger des micro-histoires d'horreur, ou de suggérer des recettes de cuisines à partir d'une liste d'ingrédients. Si le panel de possibilités est impressionnant, les résultats restent encore limités et entachés de nombreuses erreurs : une requête sur mon identité personnelle formulée en février 2023 renvoie que je serai actuellement le gardien de but du

<sup>4</sup> Chaque année, pour le Congrès dédié à la traduction automatique, un ensemble de données est préparé par couples de langues et mis à disposition des équipes plusieurs mois à l'avance. Ils servent à comparer les performances des outils (voir Statistical and Neural Machine Translation (n.d.)).

PSG, et la liste des 10 principales enseignes de vêtement françaises ne fait apparaître que les maisons de haute couture et les multinationales comme Zara ou H&M. Ces outils ne sont donc pas encore utilisables tels quels professionnellement, mais au vu du rythme auquel la traduction neuronale s'est améliorée, on peut penser qu'ils vont faire partie des outils quotidiens des traducteurs et prestataires de services linguistiques d'ici peu. On peut s'attendre à ce que les offres de services linguistiques proposés par les traducteurs et linguistes évoluent, notamment vers la vérification des informations et l'adaptation fine des contenus.

Pour ces raisons de technicité dans la conception et de coût de mise en œuvre, et même s'il existe des algorithmes open source, la traduction neuronale et les modèles de langues restent des technologies hors de portée des traducteurs individuels, voire de la plupart des agences. Et cela pose un certain nombre de problèmes éthiques.

#### 4. Algorithmes et éthique du travail

On entend ici le terme éthique au sens large, allant au-delà de la dichotomie entre licite et illicite, pour désigner un rapport au travail et entre les différentes parties prenantes qui serait idéalement durable et équilibré (en anglais *sustainable and fair*), en écho aux recommandations formulées par le CEATL (European Council of Literary Translators' Association) en 2018.<sup>5</sup>

##### 4.1. La maîtrise des outils de travail

Bien sûr, le travail d'un traducteur d'édition relève de l'œuvre de l'esprit, laquelle peut s'incarner dans de multiples formats (papier ou numérique), et les modalités de sa conception restent invisibles. À première vue, le résultat (le texte cible) prime sur le processus. Toutefois, si on ouvre la boîte noire et que l'on étudie le processus de travail lui-même, on observe que l'arrivée des moteurs neuronaux accroît une tendance à la dépossession des outils. Pendant des siècles, muni de papier et de crayons ou de leurs équivalents, le traducteur avait la maîtrise de son outil de travail. L'arrivée du traitement du texte changeait peu cet état de fait car les suites informatiques étaient achetées et installées sur les ordinateurs des traducteurs ; de même pour les logiciels de TAO et les mémoires de traduction que les traducteurs se constituaient. Quand les agences ont commencé à développer leurs plateformes de traduction maison (Heinisch & Iacono, 2019), elles ont acquis la maîtrise de plusieurs dimensions de l'outil, notamment les mémoires de traduction ou bien encore le type d'information affiché à l'écran et bien sûr les statistiques d'activité (temps de connexion, frappes clavier). La situation est un peu différente avec les moteurs de traduction grand public dont l'interface est encore plus simple : deux fenêtres en regard et quelques fonctions périphériques. Une grande partie de ce qui structurait l'information sur le processus de traduction au sein des logiciels de TAO a disparu : les matchs, la segmentation, les codes couleur, les variantes, les pourcentages et les statistiques. On passe instantanément du texte source au texte cible et très peu de personnes sont à même de comprendre les opérations en jeu. Comment savoir quels ont été les corpus utilisés pour l'entraînement ? Comment savoir si certains termes ont été censurés automatiquement parce que Meta par exemple les considérait comme « toxiques » (Stasimioti, 2022) ? Quelles sont les opérations de pondération ou d'étiquetage qui conduisent à privilégier telle expression ou construction plutôt qu'une autre ? Tous ces aspects de la production du texte pré-traduit sont masqués et inaccessibles pour l'utilisateur : l'asymétrie d'information est importante. De plus, la gratuité des interfaces grand public contribue à brouiller la distinction entre ce qui relève d'un usage professionnel et ce qui relève d'un usage personnel, ou de loisirs : travail ou non-travail ? Le travail effectif réalisé par le traducteur sur un texte pré-

<sup>5</sup> « It is essential to create fair terms, a balanced relationship, and good material and moral working conditions for translators. » (CEATL, 2018).

traduit est-il ostensiblement du travail, perçu comme tel par la société (Casilli, 2019) ? Un tel brouillage à terme pourrait mettre en jeu la perception de la traduction comme métier, et la réduire pour le grand public à la simple fonctionnalité d'une application.

#### 4.2. L'exploitation des données

Masquée et inaccessible, l'étendue de l'exploitation des données liées à l'activité de traduction donne le vertige. Les fournisseurs d'outils gratuits peuvent exploiter à leur gré les contenus téléversés (texte et formats) et les données d'activité (date et durée de connexion, manipulations sur l'outil). Ils peuvent conserver ces données longtemps : par exemple Apple conserve jusqu'à deux ans les données des utilisateurs de la dictée vocale sur les Mac ou de Siri, y compris en enregistrant les échanges vocaux (Beky, 2019 ; Hern, 2019). Evidemment, ces formes de surveillance électronique ne sont pas nouvelles, ni limitées au secteur de la traduction : des transactions financières aux interactions sur les réseaux sociaux en passant par les historiques de navigation, toutes les activités en ligne sont monitorées. Toutefois, il ne s'agit pas ici de loisirs mais d'activité professionnelle et, si on lit avec attention les conditions d'usage de Google Translate, Google se réserve le droit d'exploiter à sa guise, à des fins de recherche mais aussi de diffusion ou de modification, tous les contenus qui transitent par ses services.<sup>6</sup> En clair, le texte que vous traduisez sur leur plateforme leur appartient, même si vous avez signé un contrat d'édition par ailleurs. Il y a là une grande différence entre le droit américain et l'univers européen, où les entreprises sont assujetties à la RGPD, c'est-à-dire le règlement général sur la protection des données, en vigueur depuis le 25 mai 2018. DeepL et Lingvanex, qui sont des sociétés européennes, doivent respecter ces lois : la protection des données et la non-conservation des textes téléversés sur leurs outils sont des arguments de vente cruciaux pour leurs abonnements professionnels. Même si, à ce jour, on ne connaît pas d'exemple de traducteurs dépossédés de leur travail par les fournisseurs d'outils, la faille juridique est là et un traducteur d'édition serait avisé d'investir dans une solution professionnelle, plutôt que d'utiliser l'interface grand public. Les traducteurs ne sont pas les seuls concernés par le double mouvement de dépossession et de dessaisissement, en anglais *dispossession and disempowerment*, mais ils sont tout particulièrement concernés par la discussion sur la propriété intellectuelle, qu'il s'agisse des textes cibles, des mémoires de traduction ou des corpus (Moorkens & Lewis, 2019).

#### 4.3. La propriété intellectuelle

La question de la propriété intellectuelle des traductions passées par les moteurs de traduction est particulièrement complexe et ne peut être abordée en détail que par des juristes. Je ne relèverai donc ici que des tendances générales, néanmoins instructives. Dans le droit français, les traductions relèvent du droit d'auteur (article 112-2 du Code de la propriété intellectuelle) et sont donc considérées comme des œuvres de l'esprit, lesquelles doivent avoir un contenu original. Selon Lionel Maurel, interrogé à ce sujet lors du colloque Tralogy 3 le 8 avril 2022, un éditeur ne pourrait publier une œuvre traduite exclusivement par un moteur neuronal et en posséder les droits, car il n'y aurait pas là de contenu original. Il faut *a minima* une intervention humaine de révision et post-édition pour que le contenu devienne une œuvre. Dans le contexte américain toutefois, la propriété intellectuelle repose sur le copyright et non le droit d'auteur. Comme son nom l'indique, le copyright traite du droit de reproduction d'un contenu : ce régime n'est en vigueur que dans les pays de Common Law. Outre l'originalité de la création, le

<sup>6</sup> « This license allows Google to : host, reproduce, distribute, communicate, and use your content — for example, to save your content on our systems and make it accessible from anywhere you go ; publish, publicly perform, or publicly display your content, if you've made it visible to others ; modify your content, such as reformatting or translating it ; sublicense these rights to other users or our contractors. » (Google (5 janvier 2022)).

contenu doit avoir été « fixé », c'est-à-dire inscrit dans un media stable et il est recommandé de le déposer auprès du Copyright Office (s.d.). La question de contenus produits uniquement à partir d'IA (par exemple une traduction purement automatique, *raw MT*) est très épineuse : les droits appartiennent-ils à ceux qui ont créé l'IA ou à celui qui s'est servi de l'IA pour traduire ou créer ? Comme le stipulent les mentions légales de ces sites, les œuvres traduites à partir des plateformes de Google Translate ou Amazon Translate leur appartiennent tant que l'auteur ne s'est pas déclaré comme tel. Dans son jugement du 18 août 2023, le tribunal du district de Columbia a estimé que les œuvres d'art produites uniquement par IA ne peuvent bénéficier du copyright. Mais dans les conclusions du jugement, il est précisé que le tribunal s'est basé sur les pièces administratives dont il disposait, lesquelles ne permettaient pas d'établir une intervention humaine, et ne se prononçait pas sur tous les cas où il y a une part de machine et une part d'intervention humaine (Andrews *et al.*, 2023). Comme les plateformes de traduction automatique, ancrées dans le droit américain, sont accessibles au monde entier, la prudence est de mise, tant pour les traducteurs que pour les éditeurs.

La réforme européenne du droit d'auteur (directive du 17 avril 2019) vise à encadrer la création de contenu web afin que les créateurs (articles, vidéos, musique...) soient rémunérés pour leurs productions. L'accent est mis sur les contenus publiés sur les plateformes et pas forcément sur les contenus produits à partir des plateformes : cette loi ne concerne donc que peu les traducteurs. En revanche, elle édicte et formalise le fait que l'auteur d'un contenu en ligne doit être identifié clairement, ce qui renforce sur le principe la position des auteurs (Richir, 6 juillet 2021).

Or les auteurs ne sont évidemment pas toujours identifiés en ligne et absolument jamais quand il s'agit d'exploiter les œuvres des auteurs et traducteurs comme données pour nourrir les algorithmes de traduction. Le corpus d'entraînement de Google Translate est présenté comme composé de 2,2 milliards de segments, ce qui est gigantesque (Ghorbaniet *al.*, 2021). Outre les contenus moissonnés en permanence sur le web par les robots<sup>7</sup>, cette entreprise s'appuie sur les innombrables ouvrages qu'elle a numérisés au travers de Google Books. Restreinte normalement aux ouvrages tombés dans le domaine public au sein des partenariats avec les bibliothèques (Giroud, 8 décembre 2010), cette collecte lancée en 2004 a été étendue dans les faits à toutes sortes d'ouvrages, y compris ceux d'auteurs bien vivants. Et même si l'internaute ne peut lire en ligne que quelques extraits par exemple de *Cloud Atlas*, publié en 2004 par David Mitchell, le fichier complet est bien sûr archivé et exploité sans que l'auteur ne soit dédommagé. Or l'entraînement sur des contenus authentiques, effectivement produits par des humains, est devenu un enjeu industriel.

## 5. L'importance d'une vigilance collective

Si les moteurs de traduction améliorent le quotidien des traducteurs, la concentration de la production aux mains de quelques acteurs et l'opacité du processus invite à la vigilance, selon le principe que la production et la diffusion des énoncés dans toutes les langues et au niveau mondial ne relève pas seulement de la sphère marchande, mais constitue un bien commun (Larsonneur, 2021a). Une première étape consisterait à créer des institutions regroupant les différentes parties prenantes afin d'identifier et d'évaluer les enjeux liés au traitement

<sup>7</sup> Comme de plus en plus de contenus disponibles en ligne ont été traduits, voire générés par des algorithmes, le scénario d'une intelligence artificielle entraînée sur des textes eux-mêmes produits par la machine devient une réalité. Orhan Firat (l'autrice cite un article collaboratif ici), un des chercheurs vedettes de Google Translate, souligne ce point dans la conclusion d'un de ses derniers articles (Ghorbani *et al.*, 2021) : « Our empirical findings also raise concerns regarding the effect of synthetic data on model scaling and evaluation, and how proliferation of machine generated text might hamper the quality of future models trained on web-text. »

automatique des langues. Sur ce point, les conclusions du rapport Villani sur le développement de l'IA en France et les préconisations du Partnership on AI se rejoignent. Car le travail collectif de vigilance ne peut être fructueux que si le fonctionnement des algorithmes et la manière dont les données sont collectées et exploitées deviennent plus transparents. Le rapport Villani (Villani *et al.*, 201, p. 21) précise ainsi que

trois axes en particulier semblent mériter une attention particulière : la production de modèles plus explicables bien sûr, mais aussi la production d'interfaces utilisateurs plus intelligibles et la compréhension des mécanismes cognitifs à l'œuvre pour produire une explication satisfaisante.

Des structures d'audit formalisées, au niveau européen, mais aussi au niveau national, permettraient de mieux identifier les principaux rouages des outils de traduction grand public et de mieux informer les utilisateurs. Ce serait également un message envoyé aux plateformes, pointant *vers* leur responsabilité. En France, ceci pourrait être confié à la DGLFLF, en partenariat avec les principaux syndicats professionnels et les chercheurs, ainsi qu'avec le Défenseur des droits. Toujours pour mieux connaître le fonctionnement et les effets de ces technologies, un répertoire des incidents renseigné par les utilisateurs et les professionnels serait un outil d'évaluation extrêmement utile. Il existe un répertoire des incidents liés à l'IA en général, au niveau international, en anglais : AI Incident Database (2023a) (ou AIID), en place depuis novembre 2020 et financée par le Partnership on IA. Sur un total de près de 2000 signalements en date de décembre 2022, le portail propose 225 résultats à la requête *language*, 61 pour *translation*, et 31 pour *machine translation*, sachant qu'il y a de nombreux doublons. L'incident le plus marquant est l'arrestation en 2017 d'un Palestinien suite à la publication d'un post sur Facebook signifiant bonne journée en arabe, mais traduit automatiquement en anglais par *hurt them* et en hébreu par *attack them*.

Enfin et dans le même ordre d'idées, il pourrait être utile de faire évoluer l'évaluation des traductions. Au lieu de se focaliser sur l'identification des erreurs, il faudrait intégrer la notion de risque encouru (Canfora & Ottman, 2018) et du degré de confiance qu'on attribue à l'outil (Larssonneur, 2021b). La thématique de l'erreur est en effet centrée sur le texte-source et reste dans le domaine de la linguistique : la plupart des fournisseurs de moteurs de traduction affichent d'ailleurs une clause de non-responsabilité pour les défauts des textes cibles. La thématique du risque s'articule plus directement au droit et donc à la responsabilité civile ou pénale : l'AIID définit les incidents d'IA (AI Incident Database, 2023b) à partir des dommages qu'ils produisent dans le monde réel (*real-world harm*), ce qui inclut par exemple une atteinte à la réputation d'une entreprise, l'interférence dans des élections, la censure ou bien encore les poursuites injustes envers des personnes. La diffusion en multilingue et dans le monde entier de certaines informations ou discours produit des effets sur le réel, dont la portée peut être immense. Dans le domaine éditorial, les principaux concernés par le risque sont l'auteur, le traducteur et l'éditeur, pour des mauvaises traductions ou des textes qui heurteraient certaines sensibilités, sur des questions de genre, de discriminations ou d'opinions religieuses, par exemple. Mauvaise réputation sur les réseaux sociaux, interdictions de l'ouvrage ou même attaques contre les auteurs, comme dans le cas de Salman Rushdie, sont des risques réels et sérieux.

Au terme de ce tour d'horizon des enjeux économiques et sociétaux afférents aux technologies de traduction neuronale, les progrès en termes d'ergonomie et de coûts pour les traducteurs sont manifestes, même s'il ne faut pas sous-estimer les besoins de formation au maniement de ces outils. Ils induisent un bouleversement majeur du marché de la traduction, où de nouveaux acteurs apparaissent, et qui se scinde en domaines distincts dotés chacun de leur propre

logique. Il existe toutefois de nombreux coûts cachés : d'une part les coûts de production et d'entraînement de ses outils, qui ont un impact environnemental certain, et, d'autres part, les coûts sociétaux en termes de dépossession des outils et de propriété intellectuelle. Il est possible que soit à l'œuvre un infléchissement de notre rapport au langage, qui n'est plus seulement produit par l'humain mais en partie par la machine. Les moteurs neuronaux de traduction, et de manière générale les outils de traitement automatique des langues, sont actuellement clairement parvenus à une autonomie opérationnelle, au sens où ils sont à même de traiter les données et de produire des contenus. La question de leur autonomie décisionnelle, celle de l'interprétation et de l'évaluation des échanges, reste ouverte mais pourrait être résolue si les intelligences artificielles continuent à gagner en sophistication. Si nous voulons conserver la maîtrise des échanges linguistiques et en préserver la diversité, il me semble qu'il serait urgent d'avoir un débat éclairé et collectif sur les enjeux de cette technologie et de poser les jalons d'une régulation du secteur. L'Europe et plus précisément la Commission européenne est un des rares, sans doute le seul, acteurs publics à aborder ces questions : sur le plan technique, leur moteur de traduction automatique eTranslation est accessible gratuitement à toutes les administrations et les petites et moyennes entreprises européennes, sur inscription. Sur le plan juridique, la directive sur le droit d'auteur de 2019 aborde la question des contenus numériques. Mais il n'existe pas à ce jour d'instances nationales ou européennes capables d'effectuer les opérations de veille, d'audit et de préconisation en matière d'économie linguistique sur le modèle de la Haute Autorité de Santé française. Elle serait très utile.

## 6. Bibliographie

- AI Incident Database (2023a). *Welcome to the AI Incident Database*. <https://incidentdatabase.ai/>
- AI Incident Database (2023b). *Defining an "AI Incident"*. <https://incidentdatabase.ai/research/1-criteria/>
- Andrews, C. D., Meredith, M. W., Bloom, C. A., Bieber, C. & Valente, C. J. (30 août 2023). *Federal Court Rules Work Generated by Artificial Intelligence Alone is not Eligible for Copyright Protection*. K&L Gates. <https://www.klgates.com/Federal-Court-Rules-Work-Generated-by-Artificial-Intelligence-Alone-Is-Not-Eligible-for-Copyright-Protection-8-30-2023>
- Beky, A. (22 octobre 2019). *Siri : Apple admet conserver les données jusqu'à deux ans*. Silicon. <https://www.silicon.fr/siri-apple-confidentialite-donnees-85537.html>
- Canfora, C. & Ottmann, A. (2018). *Of ostriches, pyramids and Swiss cheese : Risks in safety-critical translations. Translation Spaces, 7(2), 167-201*. DOI : <https://doi.org/10.1075/ts.18002.can>
- Casilli, A. (2019). *En attendant les robots. Enquête sur le travail du clic*. Le Seuil.
- CEATL (2018). *Guidelines for fair translation contracts*. <https://www.ceatl.eu/tools-of-the-trade/guidelines-for-fair-translations-contracts>
- Cimino, V. (30 août 2019). *Facebook vient de fonder un consortium baptisé AI Language Research*. Siècle Digital. <https://siecledigital.fr/2019/08/30/facebook-vient-de-fonder-un-consortium-baptise-ai-language-research/>
- CNAM (s.d.) *Définition de l'ergonomie*. Consulté le 4 octobre 2023, <https://ergonomie.cnam.fr/ergonomie/index.html>
- Copyright Office (s.d.). *What is copyright ?* Consulté le 4 octobre 2023, <https://www.copyright.gov/what-is-copyright/>
- Daems, J., De Clercq, O., & Macken, L. (2018). *Translationese and Post-editese: How comparable is comparable quality? Linguistica Antverpiensia, New Series – Themes in Translation Studies, 16, 89-103*. <https://doi.org/10.52034/lanstts.v16i0.434>
- Erdrich, L. (2020). *The Night Watchman*. Harper.
- Filippone, D. (25 janvier 2022). *Meta élabore le plus grand supercalculateur dédié à l'IA*. Le Monde informatique. <https://www.lemondeinformatique.fr/actualites/lire-meta-elabore-le-plus-grand-supercalculateur-dedie-a-l-ia-85561.html>
- Fournier-Outters, T. (4 juin 2018). *La traduction automatique à l'ACPR*. Présentation lors de la 27<sup>ème</sup> rencontre du Groupe interministériel sur la traduction, Ministère de l'Economie et des Finances.
- Ghorbani, B., Firat, O., Freitag, M., Bapna, A., Krikun, M., Garcia, X., Chelba, C. & Cherry, C. (2021). *Scaling Laws for Neural Machine Translation*. ArXiv. <https://arxiv.org/pdf/2109.07740.pdf>
- Giroud, V. (8 décembre 2010). *Google Books de A à Z*. Nonfiction. [https://www.nonfiction.fr/article-3996-google\\_books\\_de\\_a\\_a\\_z.htm](https://www.nonfiction.fr/article-3996-google_books_de_a_a_z.htm)

- Google (5 janvier 2022). *Privacy and Terms*. <https://policies.google.com/terms?hl=en-US#toc-permission>
- Hansen, D. (2021). Les lettres et la machine : un état de l'art en traduction littéraire automatique. *Traitement Automatique des Langues Naturelles, Lille, France* (pp. 61-78). <https://hal.science/hal-03265904>
- Heinisch, B. & Iacono, K. (2019). Attitudes of professional translators and translation students towards order management and translator platforms. *The Journal of Specialised Translation* 32, 61-88. [https://jostrans.org/issue32/art\\_heinisch.pdf](https://jostrans.org/issue32/art_heinisch.pdf)
- Hérisson, A. (2016). Les machines humaines de Simondon. De la dialectique utilisateur/concepteur de systèmes d'armes à une émergence de capacités militaires intégrées. *Stratégique, 112*(2), 145-164. DOI : 10.3917/strat.112.0145
- Hern, A. (26 juillet 2019). Apple contractors 'regularly hear confidential details' on Siri recordings. *The Guardian*. <https://www.theguardian.com/technology/2019/jul/26/apple-contractors-regularly-hear-confidential-details-on-siri-recordings>
- IEA. (2000). *What is ergonomics (HFE) ?* <https://iea.cc/about/what-is-ergonomics>
- Jia, Y. & Carl, M. & Wang, X. (2019). How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study. *The Journal of Specialised Translation, 31*, 60-86. [https://www.jostrans.org/issue31/art\\_jia.pdf](https://www.jostrans.org/issue31/art_jia.pdf)
- Larssonneur, C. (2021a). Neural Machine Translation: from Commodity to Commons. Dans R. Desjardins, P. Lacour & C. Larssonneur (dir.), *When Translation Goes Digital* (pp. 257-280). Palgrave Macmillan.
- Larssonneur, C. (2021b). Issues of trust in machine translation. Communication à IATIS 7th Conference *The Cultural Ecology of Translation*. Barcelona.
- Lim, S. N. (s.d.). *2023 Translation Industry Trends and Stats*. Redokun. <https://redokun.com/blog/translation-statistics#translation-industry-statistics>
- Lima, C. (2 mai 2019). *Facebook wades deeper in censorship debate as it bans 'dangerous accounts'*. Politico. <https://www.politico.com/story/2019/05/02/facebook-bans-far-right-alex-jones-1299247>
- Lin, H. et al. (2021). Towards User-Driven Neural Machine Translation. *ArXiv*. <https://arxiv.org/pdf/2106.06200.pdf>
- Liu, Y., Meng, F., Chen, Y., Xu, J. & Zhou, J. (2021). Confidence-Aware Scheduled Sampling for Neural Machine Translation. *ArXiv*. <https://arxiv.org/pdf/2107.10427.pdf>
- Macken, L., Prou, D. & Tezcan, A. (2020). Quantifying the effect of Machine Translation in a High-Quality Human Translation Production Process. *Informatics, 7*(2). DOI : <https://doi.org/10.3390/informatics7020012>
- Martikainen, H. (2022). *Ghosts in the machine: Can adaptive MT help reclaim a place for the human in the loop?*. HAL Open Science. <https://hal.science/hal-03548696/document>
- Microsoft Translator. (29 mars 2022). *Translator welcomes two new languages: Somali and Zulu!*. Microsoft Translator Blog. <https://www.microsoft.com/en-us/translator/blog/2022/03/29/translator-welcomes-two-new-languages-somali-and-zulu/>
- Microsoft Translator. (12 avril 2022). *Break the language barrier with Translator — now with two new languages!*. Microsoft Translator Blog. <https://www.microsoft.com/en-us/translator/blog/2022/04/12/break-the-language-barrier-with-translator-now-with-two-new-languages/>
- Microsoft Translator. (25 avril 2022). *Introducing Faroese translation for Faroese Flag Day!*. Microsoft Translator Blog. <https://www.microsoft.com/en-us/translator/blog/2022/04/25/introducing-faroese-translation-for-faroese-flag-day/>
- Moorkens, J., Toral, A., Castilho, S., & Way, A. (2018). Perceptions of Literary Post-editing using Statistical and Neural Machine Translation. *Translation Spaces, 7*(2), 240-262. DOI : <https://doi.org/10.1075/ts.18014.moo>
- Moorkens, J. & Lewis, D. (2019). Research Questions and a Proposal for the Future Governance of Translation Data. *The Journal of Specialised Translation, 32*, 1-25. [https://www.jostrans.org/issue32/art\\_moorkens.pdf](https://www.jostrans.org/issue32/art_moorkens.pdf)
- Parlement Européen. (2019). *Directive 2019/790 du sur le droit d'auteur et les droits voisins dans le marché unique numérique*. EUR-Lex. <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=celex%3A32019L0790>
- Richir, C. (6 juillet 2021). *Qu'est-ce que la directive européenne sur le droit d'auteur ?*. Toute l'Europe. <https://www.touteurope.eu/economie-et-social/qu-est-ce-que-la-directive-europeenne-sur-le-droit-d-auteur/>
- Sanchez-Gijon, P., Moorkens, J., & Way, A. (2019). Post-Editing Neural Machine Translation vs. Translation Memory Segments. *Machine Translation, 33*(1-2), 31-59. <https://www.jstor.org/stable/45222461>
- Schumacher, P. (2020). Post-édition et traduction humaine en contexte académique : une étude empirique. *Transletters: International Journal of Translation and Interpreting, 4*, 239-274. <https://hdl.handle.net/2268/250040>
- Schumacher, P. (2019). Avantages et limites de la post-édition. *Traduire, 241*, 108-123. DOI : <https://doi.org/10.4000/traduire.1887>
- Stasimioti, M. (14 octobre 2022). *Meta Reveals How Toxic Machine Translation Can Be*. Slator. <https://slator.com/meta-reveals-how-toxic-machine-translation-can-be/>

- Statistical and Neural Machine Translation (s.d.). *EMNLP 2022. Seventh Conference on Machine Translation (WMT22)*. [https://www.statmt.org/wmt22/Communiqué de Systran. \(9 novembre 2018\). De 30 à 60% de gains de productivité grâce à l'utilisation de la traduction neuronale dans les départements traduction des industriels. Decideo. https://www.decideo.fr/De-30-a-60-de-gains-de-productivite-grace-a-l-utilisation-de-la-traduction-neuronale-dans-les-departements-traduction\\_a10961.html](https://www.statmt.org/wmt22/Communiqué de Systran. (9 novembre 2018). De 30 à 60% de gains de productivité grâce à l'utilisation de la traduction neuronale dans les départements traduction des industriels. Decideo. https://www.decideo.fr/De-30-a-60-de-gains-de-productivite-grace-a-l-utilisation-de-la-traduction-neuronale-dans-les-departements-traduction_a10961.html)
- Villani, C. , Schoenauer, M., Bonnet, Y., Berthet, C., Cornut, A.-C., Levin, F. & Rondepierre, B. (2018). *Donner un sens à l'intelligence artificielle : our une stratégie nationale et européenne*. Vie publique. <https://www.vie-publique.fr/rapport/37225-donner-un-sens-lintelligence-artificielle-pour-une-strategie-nation>
- Vincent, J. (2 novembre 2021). *Microsoft is giving businesses access to OpenAI's powerful AI language model GPT-3*. The Verge. <https://www.theverge.com/2021/11/2/22758963/microsoft-openai-gpt-3-azure-cloud-service-ai-language>
- Wang, Y., Hoang, C., & Federico, M. (2021). Towards Modelling the Style of Translators in Neural Machine Translation. In Toutanova, K., Rumshisky, A., Zettlemoyer, L., Hakkani-Tur, D., Beltagy, I., Bethard, S., Cotterell, R., Chakraborty, T. & Zhou, Y., *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1193–1199). Association for Computational Linguistics. <https://aclanthology.org/2021.naacl-main.94/>



 Claire Larssonneur

Université Paris 8  
2 rue de la Liberté  
93526 Saint Denis  
France

[Claire.larssonneur@univ-paris8.fr](mailto:Claire.larssonneur@univ-paris8.fr)

**Biographie :** Claire Larssonneur est maître de conférences en littérature britannique, traduction et humanités numériques. Ses travaux en économie de la traduction portent sur les technologies de traduction et des compétences afférentes, la tarification, les enjeux éthiques et sociétaux. Elle a publié en co-direction *When Translation Goes Digital* (Palgrave-Macmillan, 2020), *Subjectivités numériques et posthumain* (P.U. de Rennes, 2020) et *Le Sujet Digital* (Presses du réel, 2015).



This work is licensed under a Creative Commons Attribution 4.0 International License.