

Enjeux qualitatifs et éthiques de la traduction automatique

État des lieux des recherches menées au sein de l'axe « nouvelles technologies » du Centre Interdisciplinaire de Recherches en Traduction et en Interprétation

Perrine Schumacher

Université de Liège, CIRTI/ Université de Genève, TIM
p.schumacher@uliege.be

Damien Hansen

Université de Liège, CIRTI/ Université Grenoble Alpes, LIG-GETALP
damien.hansen@uliege.be

Resumen

Cet article présente les recherches récemment menées au CIRTI sur les nouvelles technologies, à travers la synthèse de deux thèses de doctorat sur le thème de la post-édition en contexte académique et de l'utilisation des outils de traduction automatique dans le domaine littéraire. Pour finir, cette contribution revient sur quelques conclusions et perspectives tirées de ces différents travaux, mais aussi de ces quelques années de réflexion et de collaboration.

Palabras clave: traducción automática neuronal – posesición – didáctica de la traducción – traducción literaria – ética

1. Introduction

De tout temps, le secteur de la traduction a été confronté au changement et contraint de s'adapter : la profession de traducteur·ice ne cesse ainsi de se redéfinir, favorisant l'émergence de nouvelles pratiques et de nouvelles tâches. Récemment, la montée en puissance de la traduction automatique (TA) neuronale, portée par les récents progrès en matière d'intelligence artificielle, a suscité de nouveaux questionnements d'ordre tant professionnel et éthique que pédagogique. En conséquence, le Centre Interdisciplinaire de Recherches en Traduction et en Interprétation (CIRTI)¹ de l'Université de Liège (ULiège) a développé un axe de recherche spécifique aux nouvelles technologies, s'intéressant notamment à la post-édition (PE) de traductions automatiques (Schumacher 2023), à la personnalisation des outils d'aide à la traduction dans le domaine littéraire (Hansen 2024b), à la traduction

de jeux vidéo (Houlmont 2022) ou à une combinaison de ces recherches (Hansen y Houlmont 2022). Dans ce contexte, cet article entend rendre compte des travaux menés au CIRTI en présentant ici une synthèse des deux premières thèses de doctorat soutenues dans ce domaine. La prochaine section est ainsi consacrée à une première thèse, intitulée "La post-édition de traduction automatique en contexte d'apprentissage : effets sur la qualité et défis pour l'enseignement de la traduction" (Schumacher 2023). La deuxième partie de cette contribution portera quant à elle sur un travail intitulé "Évaluation experte d'un prototype d'aide à la traduction créative : la traduction littéraire automatique individualisée au regard de ses enjeux traductologiques, éthiques et sociétaux" (Hansen 2024a). En conclusion, nous montrons que ces deux travaux, bien que divergeant dans leur ancrage, leurs pistes de recherche, leur public et leurs méthodes, se rejoignent finalement dans leur positionnement final

1 <https://www.cirti.uliege.be/>.

et leurs recommandations, faisant ressortir l'approche fondamentalement humaine des recherches sur les technologies menées au CIRTI, mais aussi dans les institutions partenaires à Genève et à Grenoble qui ont rendu possible l'exploration de ces nouveaux horizons.

2. La traduction automatique en contexte d'apprentissage

Depuis plusieurs années, le secteur de la traduction connaît un véritable tournant technologique induit en grande partie par l'arrivée de nouvelles technologies en traduction, telles que la traduction automatique neuronale. En conséquence, les pratiques professionnelles évoluent et l'actualisation des programmes de formation en traduction est devenue un enjeu crucial à la fois pour l'enseignement de la traduction et pour l'avenir de la profession. Dans ce contexte, cette thèse s'est consacrée à l'étude d'une pratique en plein essor dans le secteur des services langagiers, dont le développement est directement lié aux progrès en TA (Poirier 2022) : la post-édition. Par *post-édition*, nous entendons l'intervention humaine visant à modifier et à améliorer un texte traduit automatiquement par machine.

2.1. Objectifs et méthodologie

Premièrement, l'objectif majeur de ces travaux a été d'explorer les effets de la post-édition sur la qualité finale de productions, en contexte pédagogique, en comparant ce processus de traduction à la traduction dite « humaine » (TH)². Deuxièmement, nous avons

cherché à évaluer les connaissances et les perceptions des étudiant·es en traduction à l'égard de la TA et de la PE, étant donné que celles-ci influenceraient directement leur usage de la technologie, ainsi que leurs compétences en PE (O'Brien 2022). Troisièmement, notre volonté a été de contribuer à la réflexion sur l'intégration d'une formation en TA et en PE dans les cursus des apprenant·es en traduction, ainsi que sur les défis majeurs que pose la PE de TA aux traducteur·ices, en l'état actuel de la technologie.

Nos principales questions de recherche sont les suivantes :

- Question A – Quels sont les effets de la post-édition (anglais-français) en contexte d'apprentissage sur la qualité d'un texte cible ?
- Question B – Quelles sont les connaissances et les perceptions des étudiant·es en traduction sur la traduction automatique et sur la post-édition ?

Pour répondre à la question A, nous avons adopté une démarche expérimentale en menant deux expériences contrôlées (ci-après « prétest » et « test ») en contexte d'apprentissage. Le prétest s'est déroulé en 2018 avec 28 étudiant·es de 3^e bachelier en traduction (ULiège), tandis que le test a été effectué en 2021 avec 24 étudiant·es de 2^e master en traduction (ULiège). Envisagé comme une étape méthodologique, le prétest a servi à affiner les paramètres expérimentaux, à reformuler nos hypothèses et à adapter les critères d'évaluation de la qualité en vue du test.

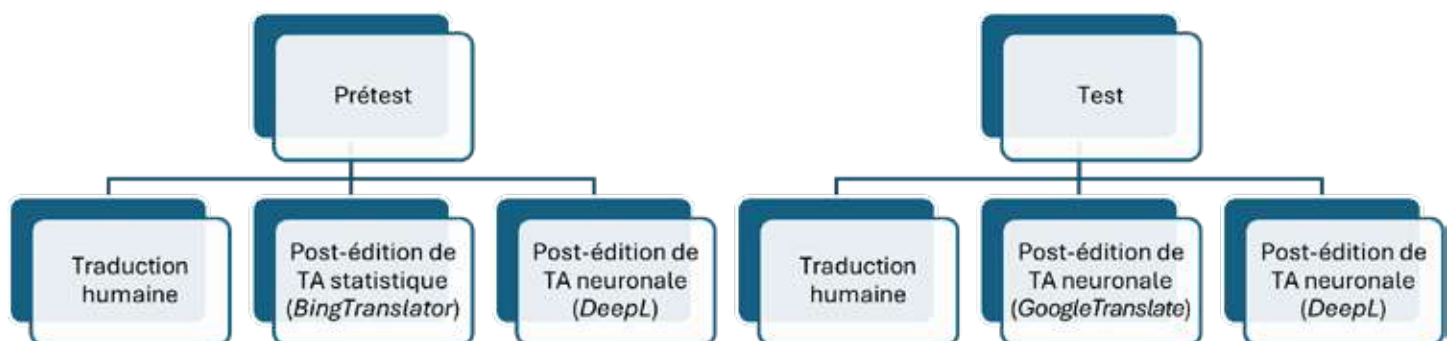


Fig. 1: Configuration du prétest et du test

2 Traduction sans recours à un outil de traduction automatique.

Pour répondre à la question B, deux sondages en ligne combinant questions ouvertes et questions fermées ont été soumis à deux reprises (2018 et 2021) auprès d'étudiant-es de 3^e bachelier, et de 1^{re} et 2^e master en traduction (ULiège). Les étudiant-es ont traduit trois textes sources de l'anglais au français (article de presse, article économique, article de vulgarisation

scientifique) selon trois modes de traduction. Pour plus de précisions, consulter Schumacher (2023).

Pour éviter tout biais et garantir la représentativité des productions, les tâches ont été réparties entre les étudiant-es selon un plan croisé (carré latin) en fonction des variables suivantes : texte source, mode de traduction et ordre des tâches (Figure 1).

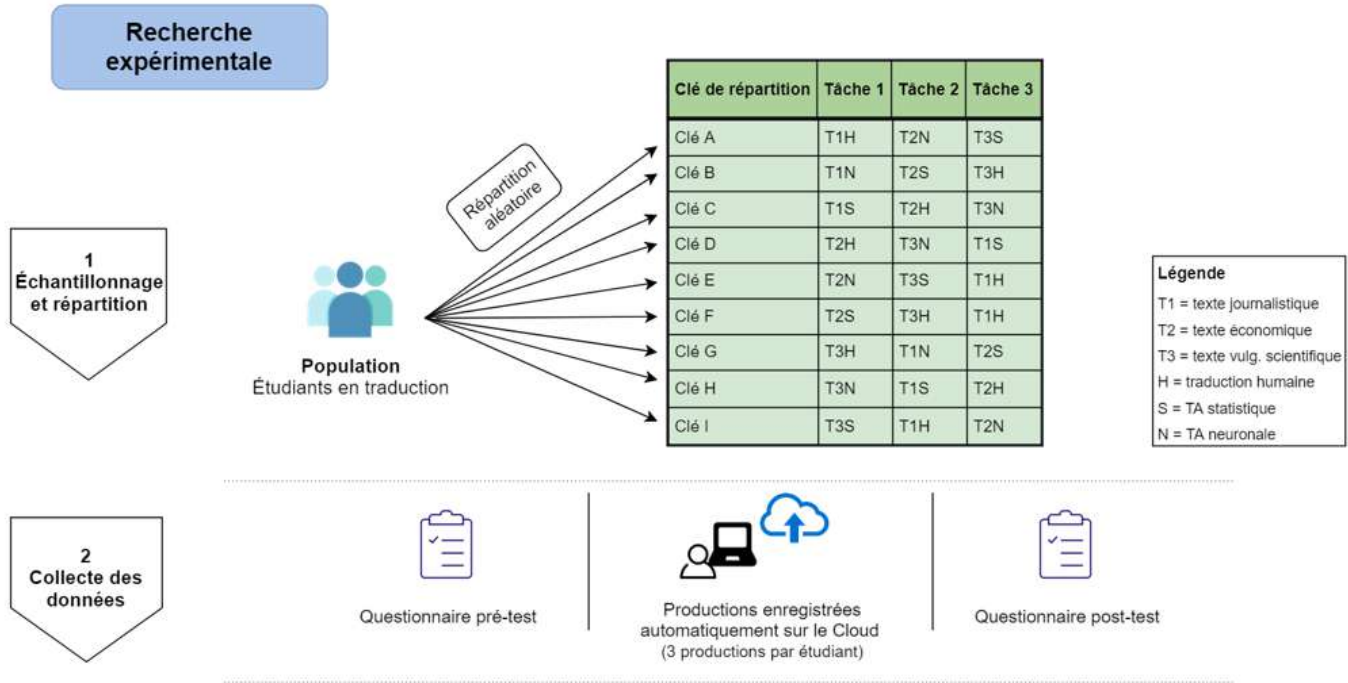


Fig. 2: Configuration de l'expérience

À la suite de ces expériences, les corpus sont ainsi composés des textes sources, des TA brutes, des textes

post-édités et des TH. La composition finale est représentée sur les figures 2 et 3.

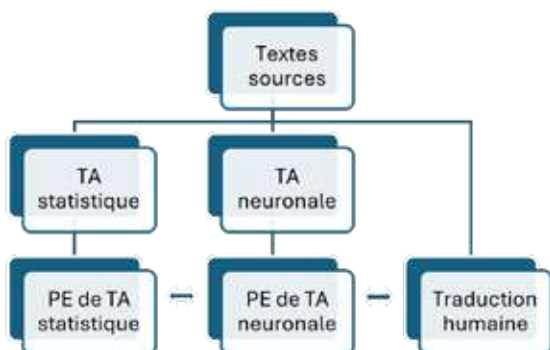


Fig. 3: Corpus Prétest

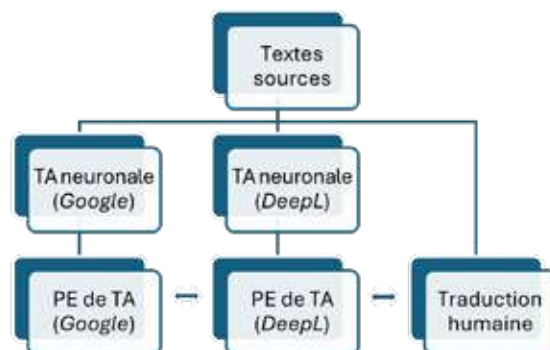


Fig. 4: Corpus Test

2.2. Exploitation des données : approches qualitative et quantitative

Pour mesurer les effets de la PE sur la qualité d'un texte cible, nous avons procédé à l'étude comparative des productions, recueillies anonymement. Les données empiriques ont été analysées en combinant une approche qualitative et une approche quantitative.

L'évaluation humaine de la qualité des textes cibles a été réalisée par cinq enseignantes de notre filière ayant plusieurs années d'expérience dans l'enseignement de la traduction. Les productions ont été évaluées à l'aveugle, sans distinction entre TH et PE. Elles ont eu recours à une typologie d'erreurs spécifique, développée pour cette étude, similaire à celle employée dans la filière traduction de l'ULiège. Cette méthode a permis un repérage précis des erreurs, leur classification par catégorie, le calcul de la moyenne des erreurs entre les évaluateurs, et l'attribution d'une note sur 20 à chaque production selon une pondération prédéfinie. L'évaluation s'est déroulée en deux phases : une portant sur l'« acceptabilité », qui concerne le respect des normes et usages de la langue cible, et une autre sur l'« adéquation », qui consiste à repérer les erreurs de traduction, évaluant ainsi la fidélité au texte source. Pour une explication détaillée de la méthode d'évaluation, voir Schumacher (2023).

À partir des relevés d'erreurs et des notes attribuées à chaque production, nous avons procédé à une analyse quantitative combinant la statistique descriptive, la statistique inférentielle et le calcul de mesures linguistiques automatiques, que nous décrivons avant d'en venir aux résultats.

Statistique descriptive

Nous avons exploré les trois modes de traduction en observant la distribution des erreurs par catégorie, la qualité des productions par mode de traduction, la moyenne des notes, le taux de réussite et les écarts-types selon le mode de traduction et selon le texte source, mais aussi en comparant les performances de différents systèmes de TA et les performances par étudiant-e.

Statistique inférentielle

Pour tester si la qualité des productions a été influencée par le mode de traduction ou par le texte source, nous avons recherché des différences statistiquement significatives à deux niveaux : la qualité du texte cible a été mesurée tantôt par le nombre d'erreurs relevées dans les productions, tantôt par les notes attribuées aux productions selon une pondération prédéfinie. Nos expériences étant réalisées selon un plan factoriel à mesures répétées, avec deux facteurs de variabilité (le mode de traduction et le texte source) et plus de deux échantillons appariés, nous avons employé un test ANOVA à deux facteurs de type III, en mesures répétées.

Mesures linguistiques automatiques

Enfin, nous avons mené une analyse linguistique automatique afin de comparer la langue post-éditée (PE), la langue traduite (TH) et la langue traduite automatiquement (TA), et de repérer la présence de traits caractéristiques à une langue de PE (*post-editeuse*) qui se différencierait de la langue traduite (*translationese*) et de langue traduite automatiquement (*machine translationese*). Pour ce faire, nos données ont été soumises à trois mesures linguistiques automatiques : la richesse lexicale, la longueur moyenne des phrases et l'équivalence syntaxique.

2.3. Résultats

Question A : Étude du produit

L'évaluation humaine des productions et l'analyse quantitative montrent que, comparée à la TH, la PE de TA, qu'elle soit statistique ou neuronale, n'a pas eu d'influence négative sur la qualité finale des productions. Cependant, les résultats des deux expériences révèlent un effet nivelant de la PE de TA neuronale sur la qualité finale. Nous avons effectivement observé une relation inverse entre le niveau d'un-e étudiant-e en TH et la qualité de sa production en PE de TA neuronale : plus l'étudiant-e a de faibles compétences en TH, plus il-elle tire avantage de la PE et, à l'inverse, plus il-elle est compétent-e en TH, plus la PE entraîne une baisse de qualité de la production finale.

Analyse globale du prétest

Dans le prétest, la PE de TA neuronale a permis d'obtenir une qualité globale supérieure à celle de la PE de TA statistique et de la TH. L'analyse statistique a indiqué que les étudiant-es ont commis significativement moins d'erreurs d'acceptabilité en PE de TA neuronale et ont obtenu de meilleures notes par rapport à la PE de TA statistique et à la TH, quel que soit le texte source. Toutefois, aucun lien n'a pu être inféré entre le mode de traduction et la distribution des erreurs d'adéquation. Ainsi, la qualité des textes post-édités dépendrait donc, notamment, du paradigme de TA (statistique ou neuronale).

Analyse détaillée du prétest

L'analyse détaillée des catégories d'erreurs a indiqué que la distribution d'erreurs en PE n'est pas comparable à celle observée en TH. Les PE (TA statistique et TA neuronale) contiennent davantage de calques et moins d'erreurs d'orthographe qu'en TH. Les PE de TA neuronale contiennent moins de glissements de sens par rapport à la TH et moins d'erreurs de grammaire et de syntaxe par rapport à la PE de TA statistique et à la TH, avec davantage de problèmes de cohérence en PE de TA statistique.

Analyse globale du test

Dans le test, il apparaît que les textes post-édités à partir de *Google Traduction* et *DeepL* sont de qualité globale comparable, voire supérieure à ceux traduits humainement, ce qui est particulièrement le cas en PE de *DeepL*. Les résultats statistiques indiquent que les étudiant-es ont commis significativement moins d'erreurs d'acceptabilité en post-éditant à partir de *DeepL* par rapport à la PE de *Google Traduction* et à la TH. Cependant, aucune relation de dépendance n'a été observée entre le mode de traduction et les erreurs d'adéquation ou les notes attribuées aux productions.

Analyse détaillée du test

L'analyse détaillée des catégories d'erreurs révèle que les productions humaines contiennent plus d'erreurs de grammaire et de syntaxe et moins de calques que

les textes post-édités (*DeepL* et *Google Traduction*). De plus, les PE de *DeepL* présentent moins de problèmes de typographie, de ponctuation et de vocabulaire. Ces résultats suggèrent, entre autres, que la qualité d'un texte post-édité en contexte d'apprentissage dépend notamment du moteur de TA neuronale employé (*Google Traduction* ou *DeepL*).

Vous avez dit *post-edite*?

Afin de déceler la présence de caractéristiques spécifiques à la langue post-éditée (*post-edite*), nous avons approfondi notre étude en calculant des mesures linguistiques automatiques et en procédant à une analyse qualitative des erreurs repérées dans les productions. Voici un récapitulatif des résultats pour les deux expériences.

Richesse lexicale

Malgré des valeurs très proches, les résultats montrent que les TH sont très légèrement plus denses lexicalement que les TA brutes et que les PE. Nous avons constaté également que les productions humaines présentaient le degré de diversité lexicale le plus élevé, ce qui nous a permis de conclure à une plus grande richesse lexicale en TH par rapport à la TA brute et à la PE de TA (statistique et neuronale). En outre, les PE de *DeepL* se révèlent les moins riches lexicalement comparativement à la TH, à la PE de *Google Traduction*, ainsi qu'à la PE de TA statistique.

Longueur moyenne des phrases

Les résultats rendent compte d'une longueur moyenne des phrases plus élevée en langue traduite par rapport aux textes sources. Les valeurs calculées en PE sont très proches de celles obtenues en TH, ce qui laisse entendre que les versions post-éditées se rapprochent davantage des TH que des textes sources. Concernant le foisonnement, les PE de TA neuronale présentent des taux de foisonnement comparables à ceux des TH, tandis que les PE de TA statistique foisonnent moins que les TH.

Équivalence syntaxique

Calculés avec le module Python *ASTrED* (Vanroy *et al.* 2021), les résultats des trois indices d'équivalence syntaxique révèlent la même tendance : les PE sont plus proches syntaxiquement des textes sources que les TH. Parmi elles, les PE de *DeepL* sont celles qui s'éloignent le plus des textes sources. En résumé, les résultats obtenus en matière de richesse lexicale et d'équivalence syntaxique ont permis de confirmer l'existence d'un *post-editeuse*, caractérisé par une moindre richesse lexicale et une plus grande proximité syntaxique avec la langue source par rapport aux TH, particulièrement en PE de *DeepL*.

Analyse qualitative des erreurs

L'analyse qualitative des erreurs contenues dans les productions nous a permis d'illustrer, à l'aide d'exemples tirés des corpus, certains des effets tant positifs que négatifs du recours à la (PE de) TA neuronale par rapport à la TH. Il s'avère que la majorité des erreurs en PE étaient déjà présentes dans la TA brute et n'ont pas été corrigées par les étudiant-es. L'analyse a également mis en évidence des caractéristiques propres au *post-editeuse* : maîtrise des règles d'usage du français (grammaire et orthographe), marques de littéralité, effacements d'informations, ruptures de cohésion textuelle et de cohérence, irrégularité terminologique et perte de richesse linguistique.

Question B : Étude des connaissances et perceptions

Les avis de 179 étudiant-es (105 en 2018 et 74 en 2021) ont révélé les tendances suivantes :

- le recours à la TA et PE chez les étudiant-es s'est intensifié entre 2018 et 2021. Leur outil de prédilection pour la TA en ligne est *DeepL*, suivi de *Google Traduction* et de *Reverso* ;
- la majorité des étudiant-es a perçu un gain de temps en PE par rapport à la TH. Et quelques étudiant-es attestent l'utilisation de la TA comme moyen d'accéder au sens de certains passages du texte source ;
- la plupart des étudiant-es considèrent la qualité d'une TA comme « satisfaisante » ou « bonne », avec amélioration de leur satisfaction au fil du temps. Cette amélioration des perceptions est certainement due aux progrès réalisés en TA neuronale, ainsi qu'à son intégration progressive dans la formation en traduction à l'ULiège ;
- la qualité en TH est généralement jugée supérieure, mais les étudiant-es anticipent une amélioration continue des performances en TA, bien que peu croient que la qualité d'une TA sans intervention humaine dépassera un jour la qualité d'une TH ;
- la familiarité des apprenant-es avec la PE a augmenté en 2021, grâce à l'introduction d'exercices spécifiques dans les cours de traduction à l'ULiège à partir de la 3^e année d'étude, conjuguée à l'essor de cette pratique dans le monde professionnel ;
- de nombreux-ses étudiant-es sont convaincu-es qu'il est possible d'obtenir en PE une qualité comparable à la TH. En 2021, les étudiant-es sont davantage satisfait-es de la PE qu'en 2018, ce qui porte à croire que l'intégration de cette matière dans leur formation a eu des effets positifs sur leurs perceptions ;
- consciente du bouleversement technologique que traverse le secteur de la traduction et des enjeux qui en découlent pour la profession, la très grande majorité des étudiant-es est en faveur d'une formation spécifique aux outils de TA et PE ;
- les perceptions globales des étudiant-es envers la TA et PE sont positives, malgré leur expérience limitée ;
- en 2018 comme en 2021, les participant-es ont considéré la tâche de TH comme étant plus exigeante que la PE. C'est la PE de *DeepL* qui a été jugée la moins exigeante pour les deux expériences ;
- plus de la moitié des participant-es (54 % en 2018 et 50 % en 2021) estiment que la TH leur a permis d'obtenir une meilleure qualité qu'en PE ;

- la PE de TA neuronale (*DeepL* et *Google*) est jugée plus satisfaisante que la PE de TA statistique, avec un taux de satisfaction global comparable à la TH ;
- en 2018, 43 % des étudiant·es de 3^e bachelier ont exprimé une attitude plus favorable envers la TA et PE après avoir participé au prétest. En revanche, en 2021, seuls 17 % des participant·es de 2^e master ont montré une évolution similaire.

Enjeux, atouts et défis de l'apprentissage de la PE

Dans la dernière partie de cette thèse, nous avons souligné la pertinence de repenser les cursus de formation en traduction en fonction des évolutions technologiques de la profession. Nous avons proposé à cet égard une réflexion sur le contenu et sur les avantages d'une formation à la TA et à la PE, ainsi que sur le moment idéal pour intégrer cette formation dans les programmes d'études en traduction. Dans la lignée de Bowker et Ciro (2019), de Loock (2019) et de Rossi (2019), nous plaidons pour une approche raisonnée des outils de TA en vue de former des traducteur·ices averti·es, autonomes et responsables.

Enfin, en nous appuyant sur notre expérience pédagogique et sur les résultats de notre étude, nous avons dressé une liste de sept grands défis en PE auxquels semblent être confronté·es les traducteur·ices, aussi bien novices que chevronné·es, et sur lesquels il convient de porter un point d'attention particulier dans l'enseignement de la PE. Il s'agit de pouvoir développer la capacité de discernement, de trouver le juste équilibre entre sur-édition et sous-édition, d'assurer la cohésion et la cohérence textuelles, d'adapter la qualité aux modalités de PE, d'atténuer l'effet fantôme de la TA, de déjouer l'illusion de fluidité en TA et de revaloriser le rôle de l'humain dans le processus de traduction.

3. La traduction littéraire automatique individualisée

Dans ce même contexte d'avancées progressives des logiciels d'aide à la traduction, particulièrement notables dans les cursus de formation avec le tournant neuronal de 2016–2017, cette seconde thèse s'est intéressée aux usages technologiques d'une communauté de traducteur·ices professionnel·les qui a

historiquement été laissée tout à fait en marge de ces discussions, et ce, bien que leur pratique a entre-temps largement été influencée par le numérique. De fait, le rapprochement entre traduction automatique ou traduction assistée par ordinateur (TAO) et traduction littéraire a longtemps semblé relever de l'antithèse. Cette recherche part cependant de l'idée que l'on n'a pas même cherché à développer des outils adaptés à cette pratique ni à penser différemment ces outils ou leur usage (Hansen 2024b), et ce, bien que l'on recense des cas d'utilisation existants et potentiellement croissants dans ce secteur, en particulier pour la TAO (Daems 2022). Étrangement, c'est surtout la TA et peut-être même plus l'IA qui ont amené ce sujet sur la table, notamment par l'engouement public et médiatique suscité par les outils reposant sur les réseaux de neurones, mais aussi parce que ces outils amènent de nouveaux usages et surtout des risques socio-économiques forts pour cette profession, qui se vérifient déjà en partie aujourd'hui.

Dès lors, si cette étude témoigne d'une volonté de proposer une vue d'ensemble sur la technologie, tout en soulignant les problématiques transversales de leur inclusion en traduction littéraire, ce travail se centre principalement sur le cas particulier de la TA, prenant comme prémisse le besoin d'outils plus appropriés et plus personnalisés dans ce domaine. Dans cet article, nous présenterons brièvement la méthodologie de départ et les dispositifs techniques mobilisés, pour répondre dans les deux sections suivantes à deux questions de recherche :

- Est-il possible d'adapter un système de traduction automatique au style individuel d'un·e traducteur·ice ?
- Quelles seraient les retombées d'un tel outil sur la pratique et la profession de la traduction littéraire ?

Par ces questions, nous espérons ainsi promouvoir une vision plus nuancée et raisonnée sur ces sujets, afin de tirer le meilleur parti des outils et d'en prévoir au mieux les répercussions.

3.1. Repenser la technologie : des systèmes adaptés et personnalisés

Si l'IA a permis d'amener le sujet sur le devant de la scène, la TA est certainement le principal moteur des recherches sur la traduction littéraire et les nouvelles technologies. Dans ce contexte, l'attention a très tôt été portée sur la possibilité d'adapter ces systèmes à ce domaine spécifique (Kenny y Winters 2024), une entreprise qui passe notamment par la création d'outils de TA spécialement prévus pour traduire de la littérature. Cette étude est cependant la première à proposer une telle adaptation non pas uniquement au domaine mais à une personne en particulier, en suggérant une adaptation au style individuel d'une traductrice littéraire.

Pour ce faire, nous avons suivi une méthodologie classique de l'entraînement des systèmes de TA, rendue possible par le module *OpenNMT* (Klein *et al.* 2017). Celui-ci nous a en effet permis de mettre au point un modèle de TA basé sur le *Transformer* (Vaswani *et al.* 2017) –celui-là même qui est aujourd'hui à la base d'outils comme *ChatGPT*, mais qui avait déjà rendu possible l'arrivée de *DeepL*– à partir de corpus traditionnellement utilisés et librement disponibles pour la paire anglais-français (comptabilisant environ quatre millions de segments et ci-après dénommés « corpus général »). Là où notre méthode s'écarte du schéma habituel est que nous avons ensuite mis un terme à l'entraînement pour relancer celui-ci uniquement à partir de traductions littéraires effectuées par la traductrice, de manière à ce que le système apprenne –en quelque sorte– de ce corpus sur-spécifique (ci-dessous, « corpus personnel »). Il s'agit donc fondamentalement d'un changement de paradigme important, mais finalement très simple à réaliser d'un point de vue technique³.

3.2. Évaluation de la qualité : approches humaines et automatiques

Pour mettre cette idée à l'épreuve, notre choix s'est porté sur un cas d'étude, à savoir les six premiers tomes de la série d'*heroic fantasy Septimus Heap* (Angie Sage, HarperCollins, 2005–2013) et leur traduction française, *Magyk* (Nathalie Serval, Albin Michel, 2005–2012). Ces volumes comptabilisent environ 45 000 segments, représentant notre corpus personnel, et ont servi à l'entraînement du système personnalisé à partir duquel nous avons produit une traduction pour les sixième et septième tomes de la série, que nous avons ensuite comparée à la traduction humaine. Pour faciliter ces comparaisons, nous avons appliqué diverses méthodes automatiques et humaines.

Pour commencer, nous avons eu recours à diverses métriques automatiques employées pour la TA, dont nous donnons ici un exemple représentatif de ces différents scores qui consiste à évaluer la proximité entre une sortie de TA et une traduction humaine de référence. Autrement dit, plus le score donné par la métrique augmente, plus la sortie s'approche de la traduction humaine, ce qui est évidemment souhaitable pour notre système personnalisé. Aux fins de la comparaison, nous donnons ici le score de l'entraînement avant adaptation du système (sur le corpus général), puis sur le corpus personnel, mais aussi sur un corpus littéraire varié de 125 romans, soit 600 000 segments (ci-après dénommé « corpus littéraire ») ou encore sur combinaison de tous ces corpus. Dans la figure ci-dessous, on remarque clairement que ce sont les traductions personnelles qui contribuent le plus à l'amélioration du système, qui s'avère d'ailleurs particulièrement prononcée⁴.

3 Pour une bibliographie et une méthodologie beaucoup plus détaillée, voir Hansen (2024a).

4 Pour cette mesure, on trouve typiquement dans la littérature scientifique des comparaisons l'ordre de quelques points ou dixièmes de points.

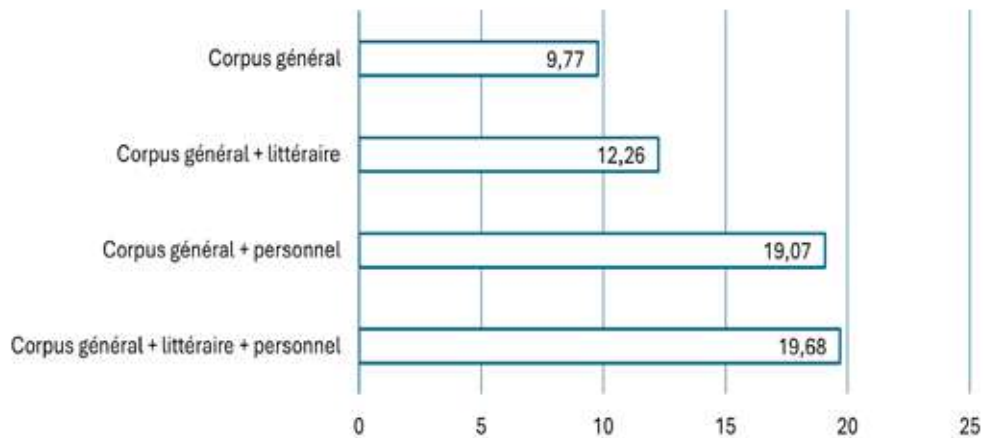


Fig. 5: Comparaison de l'entraînement avec différents corpus (score BLEU)

Comme dans le travail de la section précédente, nous avons également relevé diverses caractéristiques textuelles pouvant donner une appréciation de la qualité des sorties de traduction, mais pouvant similairement être indicatives d'un style donné, et donc de la proximité avec la traduction humaine de référence. Une de ces caractéristiques est la longueur des phrases, pour laquelle la TH présente un coefficient de foisonnement

très faible, indicatif d'une traduction très libre et synthétique, là où une comparaison avec les sorties produites par *Google Traduction* et *DeepL* donne un coefficient standard pour la paire anglais-français. Étonnement, il apparaît ici que l'outil personnalisé s'adapte à cette stratégie, omettant fréquemment certains types d'information ayant tendance à disparaître dans la TH.

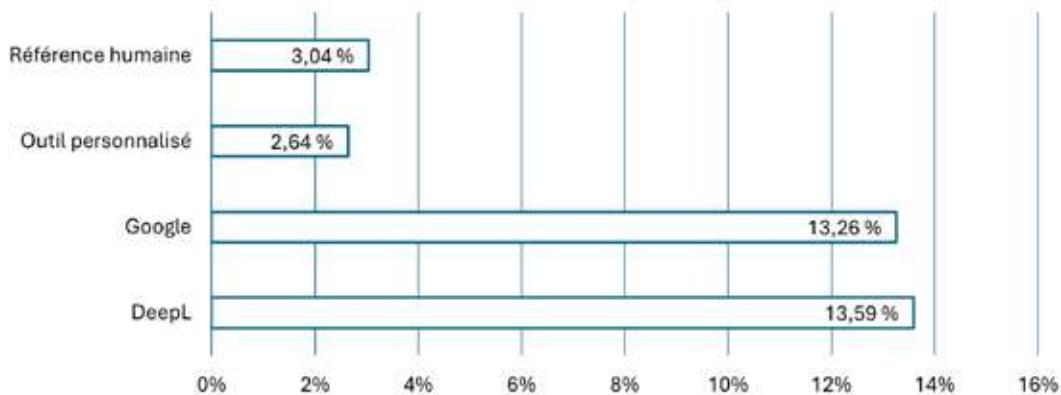


Fig. 6: Comparaison du coefficient de foisonnement entre les différentes traductions

On pourrait s'attendre à ce que ceci conduise à des traductions lexicalement moins riches. Pourtant, les trois mesures correspondantes calculées (*type-token ratio*, *mean-segmental TTR* et *moving-average TTR*)

indiquent une nette amélioration du système de TA personnalisé en ce sens, par rapport aux deux systèmes génériques, et ce, bien que la TH fasse montre d'une plus grande richesse lexicale encore.

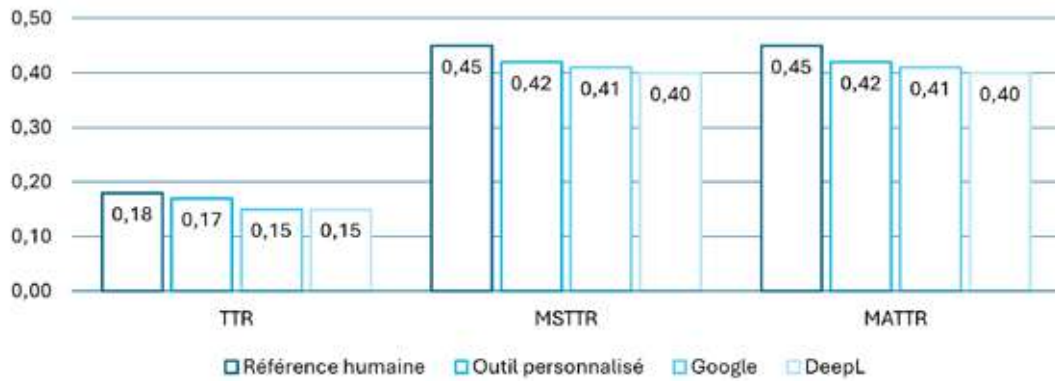


Fig. 7: Comparaison de la richesse lexicale entre les différentes traductions

Il en va de même pour l'« équivalence syntaxique », pour laquelle nous avons relevé trois mesures estimant le degré de proximité entre les structures syntaxiques de la traduction et de la source. Celles-ci ont

été calculées avec *ASTrED* (Vanroy *et al.* 2021) de façon entièrement automatique –avec une marge d'erreur possible, donc.

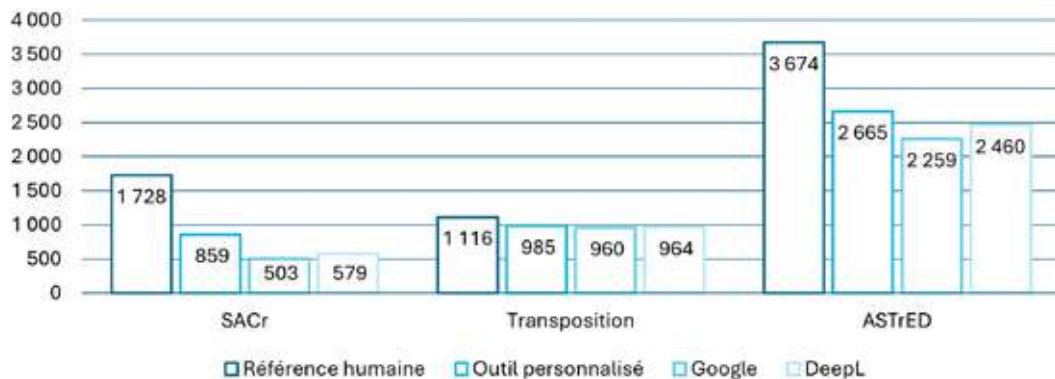


Fig. 8: Comparaison de l'écart syntaxique entre les différentes traductions et la source

De fait, la traduction humaine affiche sans surprise un nombre bien plus élevé de réorganisations syntaxiques (SACr) et de transpositions, mais aussi une distance bien plus élevée avec l'arbre syntaxique du texte source (ASTrED). Les traductions générées par l'outil personnalisé montrent quant à elles une amélioration à ces égards (particulièrement pour la réorganisation

syntactique et la déviation de la structure source), tandis que les sorties des deux autres systèmes restent les plus proches de la source. Étant donné la nature automatique de ces mesures, nous avons toutefois voulu confirmer celles-ci par une lecture attentive (*close reading*) des traductions générées, par une annotation d'erreurs et par des observations générales.

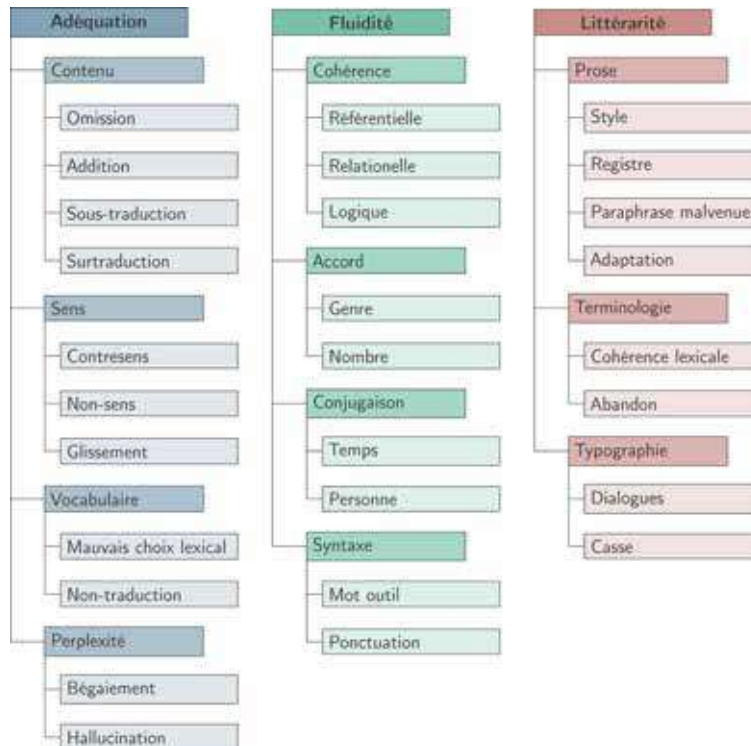


Fig. 9: Typologie d'erreurs employée pour l'annotation

L'annotation d'erreurs, notamment, a révélé que notre système personnalisé produisait finalement peu d'erreurs de fluidité –on sait que c'est une des forces des systèmes neuronaux– mais générait énormément de problèmes d'adéquation, plus problématiques et difficiles à repérer. Représentant 50 % du nombre total d'erreurs, on y trouve principalement des problèmes d'omission et de sous-traduction, de mauvais choix lexicaux, mais aussi des glissements de sens et des non-sens. Pour ce qui est de la littéarité, la moindre proportion d'erreurs démontre selon nous l'efficacité de l'adaptation, excepté pour le style, qui représente 57 % des erreurs de cette catégorie et 17 % du total ⁵.

La lecture plus attentive des traductions produites par ce système personnalisé a permis en outre de confirmer les observations tirées des évaluations automatiques, tout comme la meilleure adéquation du texte cible avec le genre de *l'heroic fantasy*, mais a pareillement mis en lumière d'autres tendances intéressantes, comme la capacité du système à employer des

synonymes, à paraphraser, à produire des traductions plus libres, à fusionner des phrases, à effectuer des modulations, des transpositions, des omissions pertinentes et des contractions, qui reflètent généralement des choix similaires dans la traduction de référence et qui ont traditionnellement été soulevées parmi les limites principales de la TA face à la TH (Ahrenberg 2017: 26). Néanmoins, l'étude ne nous dit pas à ce stade dans quelle mesure cet outil pourrait être utile à la traductrice de ces romans.

3.3. Une recherche technocritique : évaluer les outils en contexte réel

Les évaluations de la TA, qu'elles soient humaines ou automatiques, comportent de nombreuses limites. Le meilleur banc d'essai d'un logiciel d'aide à la traduction reste donc l'évaluation experte, à savoir une évaluation en situation réelle par un-e traducteur-ice professionnel-le. Cette méthode permet en outre d'évaluer non pas uniquement le produit, mais aussi le processus de

5 Pour plus de détails sur les méthodes et les résultats des évaluations, voir Hansen (2024a).

la traduction (Borg 2023: 2–3), d’analyser l’influence de la TA sur les phénomènes cognitifs de la traduction, de même que les aspects ergonomiques de cette interaction humain-machine (Risku 2010).

Nous avons donc proposé à la traductrice française, Nathalie Serval, de tester cet outil personnalisé lors d’un exercice de traduction assistée par ordinateur, au cours duquel nous avons utilisé plusieurs combinaisons d’outils (interfaces de travail et systèmes de TA) ainsi que diverses méthodes d’enquête (observation, protocole de verbalisation, entretiens libres et semi-directifs). S’il est ainsi ressorti de ces expériences que l’outil personnalisé pouvait potentiellement jouer le rôle d’un assistant utile et qu’il montrait certaines améliorations vis-à-vis des systèmes de TA non personnalisés, nos discussions et comparaisons avec la même traduction humaine réalisée 10 ans auparavant ont cependant révélé deux problèmes majeurs, à commencer par l’environnement de TAO, dont il faut reconnaître que les interfaces ont peu changé en quelques décennies, ainsi que les modalités d’interaction avec la TA, pour lesquelles le scénario classique de post-édition s’avère trop contraignant et interfère dans le processus de déverbalisation (Schumacher 2023: 393), et donc avec la voix et le style de la traductrice.

Ces sources d’interférence posent donc un risque pour la qualité des traductions, mais nos discussions ont aussi soulevé les risques posés par la technologie vis-à-vis des conditions de travail (manque de transparence, rémunération, droits d’auteur·ice, statut, liberté des traducteur·ices sur le processus de traduction, etc.), qui posent en effet des questions sur les aspects sociaux et organisationnels de la profession, sur le contexte d’utilisation plus large des outils d’aide à la traduction et sur les motivations de leur introduction. Or, les enquêtes sur les usages actuels de la TA dans le domaine littéraire montrent que ces questions sont plus que jamais d’actualité aujourd’hui (ATLF 2022). Sur ce, nous proposons en définitive un changement de paradigme dans la manière d’envisager la traduction assistée par ordinateur (incluant la post-édition), mais aussi dans la manière d’interagir avec la TA. Évidemment, cette contribution appelle à davantage d’études sur ces aspects ergonomiques de la technologie, ainsi que des études prenant en compte l’ensemble des

dispositifs technologiques, sans oublier que les outils peuvent et devraient idéalement toujours être adaptés pour mieux soutenir le travail humain.

4. Conclusions et perspectives

Ces deux études complémentaires, en s’intéressant à différents angles traductologiques, linguistiques, techniques et éthiques, convergent dans l’ensemble vers des conclusions similaires. Elles réaffirment la valeur ajoutée de l’humain dans le processus de traduction et témoignent de l’aide que ces outils peuvent apporter aux traducteur·ices. Toutefois, ces mêmes conclusions insistent sur la nécessité d’un développement éthique de la technologie, d’une formation solide, d’une utilisation responsable des outils et d’un environnement de travail décent et durable.

Malgré les progrès indéniables en TA neuronale et les enjeux qu’elle pose notamment pour les formations, il est certain que l’humain continue de jouer un rôle prépondérant et indispensable en traduction. Sur ce constat, nous avons montré comment intégrer ingénieusement la technologie neuronale aux outils de TAO (TA interactive, TA adaptative, etc.) pour recentrer le processus de traduction sur l’humain. Enfin, nous avons exprimé notre conviction selon laquelle l’avenir de la traduction professionnelle va dépendre de la capacité des traducteur·ices à s’approprier les nouvelles technologies et à valoriser leurs spécificités humaines. Nous pensons que les formateur·ices en traduction ont, à cet égard, un rôle déterminant à jouer en développant les capacités d’adaptation des étudiant·es, quelle que soit l’innovation technologique à appréhender, ainsi qu’en mettant l’accent sur la valeur ajoutée de la traduction humaine dans cette « civilisation de l’algorithme » (Rouvroy 2018).

Si cela se vérifie pour la formation des futures générations, il en est tout autant des professionnel·les déjà en activité sur le marché. Une utilisation non raisonnée de la technologie peut en effet avoir des retombées directes sur les conditions de travail et le bien-être des traducteur·ices, ou sur la qualité des textes traduits. Plus spécifiquement, pour la traduction littéraire, on pourra inclure des enjeux plus larges concernant la diffusion de la culture, le plaisir de lecture, la réception

des textes traduits par les lecteur·ices, le travail des auteur·ices, la propriété sur les données, l'écologie, la précarisation des métiers liés aux technologies du langage, etc. Dans l'ensemble, ces défis soulignent la nécessité du dialogue entre les personnes qui développent et qui utilisent les outils (O'Brien y Conlan 2018), mais aussi de mettre les discours sur l'avancée de la technologie en lien avec les conditions de travail (Moorkens 2020), et enfin d'adopter une vision plus transversale de la technologie prenant en compte tous les outils à disposition des traducteur·ices ainsi que leur environnement physique et social. Il est aussi important de ne pas oublier que les outils peuvent toujours être détournés ou réappropriés, et que la valorisation des compétences humaines passe également par le développement de la technologie.

Pour conclure, nous espérons que ces travaux permettront de mieux cerner les atouts et les défis posés

à différents niveaux par les technologies du langage, et que nos résultats et nos réflexions contribueront à faire évoluer les formations –initiales ou continues– de la traduction, en promouvant un usage raisonné de la TA neuronale auprès des apprenant·es, des formateur·ices, des professionnel·les et des autres maillons de la chaîne de traduction. Évidemment, les évolutions que connaît le monde de la traduction ne se limitent pas à l'arrivée de la TA neuronale. Récemment, l'application des grands modèles de langue (LLM) aux technologies d'IA générative ont donné naissance à de nouveaux outils puissants, accessibles à tou·tes. Ces chatbots, tels *ChatGPT* ou *Gemini*, mériteraient également des recherches approfondies, puisqu'ils sont susceptibles de venir enrichir encore la palette d'outils des traducteur·ices, en offrant notamment la possibilité de combiner TA, PE et reformulation, mais parce qu'ils posent par ailleurs des risques exacerbés ou des enjeux nouveaux pour la profession.

Referencias

- Ahrenberg, Lars (2017). "Comparing Machine Translation and Human Translation: A Case Study". En Temnikova, Irina/ Constantin Orasan/ Gloria Corpas Pastor/ Stephan Vogel (eds.): *Proceedings of the First Workshop on Human-Informed Translation and Interpreting Technology*. Varna: ACL, pp. 21–28. DOI: [10.26615/978-954-452-042-7_003](https://doi.org/10.26615/978-954-452-042-7_003).
- ATLF. 2022. "Traduction automatique et post-édition". Sondeo de la ATLF realizado del 20 de noviembre al 13 de diciembre 2022. Paris: ATLF. Disponible en <https://atlf.org/wp-content/uploads/2023/03/ENQUETE-TRADUCTION-AUTOMATIQUE.pdf>.
- Borg, Claudine (2023). *A Literary Translation in the Making: A Process-Oriented Perspective*. Nueva York: Routledge. DOI: [10.4324/9781003150909](https://doi.org/10.4324/9781003150909).
- Bowker, Lynne/ Jairo Buitrago (2019): *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*. Bingley: Emerald Group Publishing. DOI: [10.1108/9781787567214](https://doi.org/10.1108/9781787567214).
- Daems, Joke (2022). "Dutch literary translators' use and perceived usefulness of technology: The role of awareness and attitude". En Hadley, James/ Kristiina Taivalkoski-Shilov/ Carlos S. C. Teixeira/ Antonio Toral (eds.): *Using Technologies for Creative-Text Translation*. Nueva York: Routledge, pp. 40–65. DOI: [10.4324/9781003094159](https://doi.org/10.4324/9781003094159).
- Hansen, Damien (2024a): "Évaluation experte d'un prototype d'aide à la traduction créative : la traduction littéraire automatique individualisée au regard de ses enjeux traductologiques, éthiques et sociétaux". Tesis doctoral, Universidad de Lieja/ Universidad Grenoble Alpes. Disponible en <https://hdl.handle.net/2268/312631>.

- Hansen, Damien (2024b): “The figure of the literary translator amidst new technologies”. En Winters, Marion/ Sharon Deane-Cox/ Ursula Böser (eds.): *Translation, Interpreting and Technological Change: Innovations in Research, Practice and Training*. Londres: Bloomsbury, pp. 80–108. DOI: [10.5040/9781350212978.0012](https://doi.org/10.5040/9781350212978.0012).
- Hansen, Damien/ Pierre-Yves Houlmont (2022). “A Snapshot into the Possibility of Video Game Machine Translation”. En Campbell, Janice/ Stephen Larocca/ Jay Marciano/ Konstantin Savenkov/ Alex Yanishevsky (eds.): *Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas*. Orlando: AMTA, pp. 257–269. Disponible en <https://aclanthology.org/2022.amta-upg.18/>
- Houlmont, Pierre-Yves (2022). “Traduire le jeu vidéo : un équilibre des dynamiques intersémiotiques”. En *Sciences du jeu*, 17. DOI: [10.4000/sdj.4148](https://doi.org/10.4000/sdj.4148).
- Kenny, Dorothy/ Marion Winters (2024): “Customization, Personalization and Style in Literary Machine Translation”. En Winters, Marion/ Sharon Deane-Cox/ Ursula Böser (eds.): *Translation, Interpreting and Technological Change: Innovations in Research, Practice and Training*. Londres: Bloomsbury, pp. 59–79. DOI: [10.5040/9781350212978.0010](https://doi.org/10.5040/9781350212978.0010).
- Klein, Guillaume/ Yoon Kim/ Yuntian Deng/ Jean Senellart/ Alexander Rush (2017). “OpenNMT: Open-Source Toolkit for Neural Machine Translation”. En Bansal, Mohit/ Heng Ji (eds.): *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Vancouver: ACL, pp. 67–72. Disponible en <https://aclanthology.org/P17-4012>.
- Loock, Rudy (2019): “La plus-value de la biotraduction face à la machine”. En *Traduire*, 241, pp. 54–65. DOI: [10.4000/traduire.1848](https://doi.org/10.4000/traduire.1848).
- Moorkens, Joss (2020). “A tiny cog in a large machine: Digital Taylorism in the translation industry”. En *Translation Spaces* 9, 1, pp. 12–34. DOI: [10.1075/ts.00019.moo](https://doi.org/10.1075/ts.00019.moo).
- O’Brien, Sharon (2022): “How to deal with errors in machine translation: Postediting”. En Kenny, Dorothy (ed.): *Machine translation for everyone: Empowering Users in the Age of Artificial Intelligence*. Berlín: Language Science Press, pp. 105–120. DOI: [10.5281/zenodo.6653406](https://doi.org/10.5281/zenodo.6653406).
- O’Brien, Sharon/ Owen Conlan (2018). “Moving Towards Personalising Translation Technology”. En Dam, Helle/ Matilde Brøgger/ Karen Zethsen (eds.): *Moving Boundaries in Translation Studies*. Abingdon: Routledge, pp. 81–97. DOI: [10.4324/9781315121871](https://doi.org/10.4324/9781315121871).
- Poibeau, Thierry (2022): “On ‘Human Parity’ and ‘Super Human Performance’ in Machine Translation Evaluation”. En Calzolari, Nicoletta/ Frédéric Béchet/ Philippe Blache/ Khalid Choukri/ Christopher Cieri/ Thierry Declerck/ Sara Goggi/ Hitoshi Isahara/ Bente Maegaard/ Joseph Mariani/ Hélène Mazo/ Jan Odijk/ Stelios Piperidis (eds.): *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. Marsella: ELRA, pp. 6018–6023. Disponible en <https://aclanthology.org/2022.lrec-1.647/>.
- Risku, Hanna (2010). “A cognitive scientific view on technical communication and translation: Do embodiment and situatedness really make a difference?” En *Target*, 22, 1, pp. 94–111. DOI: [10.1075/target.22.1.06ris](https://doi.org/10.1075/target.22.1.06ris).
- Rossi, Caroline (2019): “Les usages actuels de la traduction automatique”. Conferencia del *Atelier DigiThum 2019 : les humanités numériques en langues*, 17 de octubre de 2019, París. Disponible en https://digiThum.huma-num.fr/atelier/2019/4_rossi.php.
- Rouvroy, Antoinette (2018): “Homo juridicus est-il soluble dans les données?”. En Terwangne, Cécile de/ Elise Degrave/ Séverine Dusollier/ Robert Queck (eds.): *Droit, normes et libertés dans le cybermonde*. Bruselas: Éditions Larcier, pp. 417–444.
- Schumacher, Perrine (2023): “La post-édition de traduction automatique en contexte d’apprentissage : effets sur la qualité et défis pour l’enseignement de la traduction”. Tesis doctoral, Universidad de Lieja/

Universidad de Ginebra. Disponible en <https://hdl.handle.net/2268/302239>.

Schumacher, Perrine (2023): “Traduction humaine et postédition : contrôle qualité en contexte académique”. En *Meta*, 68, 3, pp. 510–536. DOI: [10.7202/1111955ar](https://doi.org/10.7202/1111955ar).

Vanroy, Bram/ Orphée De Clercq/ Arda Tezcan/ Joke Daems/ Lieve Macken (2021): “Metrics of Syntactic Equivalence to Assess Translation Difficulty”. En Carl, Michael (ed.): *Explorations in Empirical Translation Process Research, Machine Translation: Technologies and Applications*. Cham: Springer, pp. 259–294. DOI: [10.1007/978-3-030-69777-8_10](https://doi.org/10.1007/978-3-030-69777-8_10).

Vaswani, Ashish/ Noam Shazeer/ Niki Parmar/ Jakob Uszkoreit/ Llion Jones/ Aidan Gomez/ Lukasz Kaiser/ Illia Polosukhin (2017). “Attention is All you Need”. En von Luxburg, Ulrike/ Isabelle Guyon/ Samy Bengio/ Hanna Wallach/ Rob Fergus (eds.): *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., pp. 6000–6010. Disponible en <https://papers.nips.cc/paper/2017/hash/3f5ee-243547dee91fbd053c1c4a845aa-Abstract.html>

Perrine Schumacher es doctora en Traductología por la Universidad de Lieja y en Tratamiento Informático Multilingüe de la Universidad de Ginebra. En este contexto, trabajó sobre la didáctica y las nuevas tecnologías en traducción, lo que incluye en particular la posesión de traducción automática. A través de sus investigaciones, pretende participar en una reflexión más amplia sobre la integración de las nuevas herramientas tecnológicas en los programas de formación en traducción.

Damien Hansen es doctor en Traductología por la Universidad de Lieja y en Ciencias de la Computación por la Universidad Grenoble Alpes. En este contexto, trabajó sobre la reapropiación de las tecnologías de traducción como herramienta personal de los traductores literarios, centrándose en el caso de la traducción automática adaptada al estilo individual de los traductores, pero otras áreas de investigación incluyen la traducción de videojuegos y la ludología.

El **Centre Interdisciplinaire de Recherches en Traduction** (Centro Interdisciplinario de Investigaciones en traducción) se constituyó como Unidad de Investigación de la Universidad de Lieja en la primavera de 2016 con el objetivo de darle mayor visibilidad a las investigaciones sobre esa temática, que se desarrollaban desde el año 1991. Si bien las investigaciones del CIRTI exploran las distintas facetas de la traducción, ya sea como proceso, producto o práctica, todas contribuyen a una reflexión común sobre las tensiones que atraviesan el acto de traducir, la subjetividad que involucra y su anclaje social. Organizados en torno a cinco ejes temáticos, los estudios se han interesado principalmente por la relación entre la traducción y las prácticas artísticas, los contextos que promueven una transferencia teórica, la diversidad de figuras de traductor/a, las estrategias de no traducción y, más recientemente, las evoluciones ligadas a las nuevas tecnologías.