Les lettres et la machine : un état de l'art en traduction littéraire automatique

Damien Hansen

▶ **To cite this version:**

Damien Hansen. Les lettres et la machine : un état de l'art en traduction littéraire automatique. Traitement Automatique des Langues Naturelles, 2021, Lille, France. pp.61-78. hal-03265904

HAL Id: hal-03265904
https://hal.science/hal-03265904

Submitted on 23 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles
Lille, France, 28 juin au 2 juillet 2021

Les lettres et la machine : un état de l'art en traduction littéraire automatique

Damien Hansen
Université de Liège, CIRTI, 4000 Liège, Belgique
Univ. Grenoble Alpes, CNRS, Grenoble INP*, LIG, 38000 Grenoble, France
damien.hansen@uliege.be

Résumé

Étant donné la récente vague d'intérêt pour la traduction littéraire automatique, cet article vise à recenser les travaux déjà parus sur le sujet, tout en partageant quelques prises de position sur ce thème. Nous commencerons par présenter les travaux précurseurs qui ont motivé ces différentes recherches, ainsi que les résultats obtenus plus récemment dans divers scénarios et pour diverses paires de langues. Pour terminer ce tour d'horizon, nous exposerons les débuts de nos travaux pour la paire anglais-français, avant d'évoquer les préoccupations et les avantages à prendre en compte dans les discussions autour de cette technologie.

Abstract

Machines in the humanities: current state of the art in literary machine translation

Given the recent surge of interest for literary machine translation, this article aims to give an overview of the works already published on the topic, and to share a few points of view on the same subject. First, the pioneering research that motivated these various studies will be presented, as well as the results recently reported in different scenarios and language combinations. To conclude this survey, we will present preliminary findings of our work on the English-French pair, before mentioning some of the concerns and advantages we should include in discussions involving this technology.

Mots-clés : Traduction automatique, post-édition, TAO, littérature, prose, poésie.
Keywords: Machine translation, post-editing, CAT tools, literature, prose, poetry.

1 Introduction

Depuis les premiers travaux en traductologie, la traduction d'ouvrages littéraires a toujours reçu une attention privilégiée et occupé une place un peu à part, en particulier si on les compare à d'autres domaines techniques ou appliqués (Lauvault-Olléon, 2011). Par moment, certains ont voulu qu'elle ne soit pas même possible pour les humains (Genzel *et al.*, 2010 ; Greene *et al.*, 2010 ; Tavailkoski-Shilov, 2019). L'arrivée de la traduction automatique (TA) ne fait pas exception à la règle et reflète d'ailleurs très bien les prises de position diamétralement opposées sur le sujet. D'un côté, il n'est plus surprenant de voir des personnes clamant la parité des systèmes de TA face à la traduction humaine (Hassan *et al.*, 2018), et la littérature est souvent invoquée dans ce cas pour justifier l'observation, même si de nombreuses voix plus mesurées ont depuis remis ce constat en question (Läubli *et al.*, 2018 ; Toral *et al.*, 2018). À l'inverse, ses détracteurs et détractrices ont fréquemment recours aux systèmes en ligne pour montrer que quelques lignes de prose mal traduites suffisent à démontrer sa parfaite inutilité (Toral & Way, 2015b).

* Institute of Engineering Univ. Grenoble Alpes

Cette œuvre est mise à disposition sous licence Attribution 4.0 International.

Pourtant, très peu d'attention a été portée à la traduction littéraire automatique (TLA) jusqu'à présent (Voigt & Jurafsky, 2012 ; Besacier & Schwartz, 2015 ; Moorkens *et al.*, 2018 ; Matusov, 2019). Malgré l'intérêt croissant pour la TA au cours de ces dernières années et les diatribes passionnées qu'elle suscite parfois lorsqu'on la rapproche du secteur littéraire, force est de constater que peu d'études avaient abordé la chose d'un point de vue empirique jusqu'à récemment (Toral & Way, 2015a). Au départ, cet intérêt était plutôt lié à la rencontre de la linguistique computationnelle et de la littérature, comme en atteste le *workshop on computational linguistics for literature*, organisé depuis 2012 par l'ACL. Avant encore, les approches précédant la TA statistique étaient détournées à des fins créatives et appréciées pour la consonance avant-gardiste qu'elles donnaient aux textes produits (Kenny & Winters, 2020).

Aujourd'hui, on constate néanmoins une avancée progressive du modèle de production combinant la TA et post-édition (PE), et ce, depuis les systèmes de TA probabilistes (Besacier & Schwartz, 2015 ; Toral & Way, 2015a). Au vu de ces avancées, il est dès lors devenu légitime de se demander quelle pouvait être la performance de ces outils sur les textes littéraires, d'autant plus que le changement de paradigme apporté par les systèmes neuronaux est réputé donner de meilleurs résultats sur les textes d'une plus grande richesse lexicale et d'une plus grande complexité syntaxique (Bentivogli *et al.*, 2016). Cette question est rendue plus pertinente encore par le fait que la traduction littéraire est et a toujours été indispensable aux échanges culturels à travers le monde, mais aussi par le fait qu'il s'agit d'une tâche particulièrement longue et coûteuse (Voigt & Jurafsky, 2012). Selon Besacier & Schwartz (2015), la TLA pourrait même avoir un intérêt pour chacun des maillons de la chaîne de traduction : des éditeur·trice·s aux lecteur·trice·s, sans oublier les auteur·e·s et les traducteur·trice·s. Dans tous les cas, une revue des travaux déjà accomplis est nécessaire si l'on veut aborder la chose de manière raisonnée.

2 Les précurseurs

Comme nous le remarquions, la traduction a historiquement porté un grand intérêt pour la littérature, et notamment pour la poésie. Ceci est dû en partie à l'idée tenace d'une activité qui serait plus « noble » que les autres et qui lui vaut aujourd'hui d'être décrite par certains, selon Toral & Way (2015a), comme « le dernier bastion de la traduction humaine », renforçant davantage l'idée d'une opposition nette entre l'homme et la machine.

Cette idée n'avait pourtant pas empêché quelques auteur·e·s de s'essayer à la traduction automatique de poésie. Dans l'un des premiers travaux sur la question, Genzel *et al.* (2010) avaient par exemple mis au point un système de TA statistique obéissant à des contraintes de longueur, de métrique accentuelle et de rime. Entraîné sur des données hors domaine, celui-ci parvenait raisonnablement à satisfaire les demandes formelles, au détriment malgré tout de la qualité. Greene *et al.* (2010) avaient pour leur part produit deux systèmes statistiques pareillement contraints par la métrique, qui avaient pour objectif de générer et de traduire des poèmes. Dans le cas du système de traduction, entraîné sur des données à la fois similaires et extérieures au domaine pour augmenter la couverture lexicale, les chercheur·euse·s avaient remarqué qu'il produisait parfois des traductions semblables à un segment de référence ou à des fragments de plusieurs références, parfois de nouvelles traductions valides, et qu'il pouvait être aisément modulé du point de vue de la forme, bien que le résultat n'était pas toujours fluide.

Ces études s'inscrivaient par ailleurs généralement dans une tradition visant à recourir aux capacités de la machine pour mieux comprendre les caractéristiques des textes littéraires. C'est le cas par exemple de Voigt & Jurafsky (2012), dont l'intérêt portait justement sur les difficultés que posent les textes en prose et que doit surmonter la TA (tout comme les humains) pour traduire ce type de textes. Les auteurs relevaient à cet égard l'importance de la cohésion textuelle pour les textes littéraires, en montrant que ceux-ci affichent des chaînes référentielles bien plus denses que d'autres domaines tels que la presse.

Partant, les auteurs suggéraient que les recherches futures sur la TA devraient non seulement pouvoir compter sur des données d'entraînement adaptées au domaine, mais aussi prendre en compte les éléments textuels dépassant le cadre de la phrase, pour que celle-ci soit véritablement efficace en littérature.

Plus récemment, les tentatives de traduction automatique de poèmes ont trouvé des échos grâce aux progrès apportés par la traduction automatique neuronale (TAN). Ghazvininejad *et al.* (2018) ont ainsi présenté un système entraîné sur des données hors domaine et adapté par la suite sur un corpus parallèle de chansons. Les résultats reportés montrent que le modèle parvient à produire des poèmes du français vers l'anglais en respectant les contraintes formelles imposées, et que des humains jugent la production acceptable dans environ 78 % des cas, voire bonne et très bonne pour presque 50 % des exemples. La traduction de poèmes représente toutefois une minorité des recherches sur la traduction littéraire automatique. La plupart d'entre elles se penchent en effet sur la prose, et c'est donc ce versant de la TLA qui nous occupera dans la suite de l'article.

3 Les premières approches statistiques

Comparant justement la traduction automatique de prose et de poésie depuis le français vers l'anglais, Jones & Irvine (2013) offrent une des premières études sur le sujet. Bien que les chercheuses insistent sur l'importance d'évaluer la production d'un point de vue qualitatif, elles offrent tout de même un score BLEU[1] pour la poésie (16.62) et pour la prose (30.05). Dans l'analyse, il ressort que la TA effectue des choix intéressants, mais que les erreurs les plus importantes (choix des temps et choix lexicaux) soulèvent la nécessité d'entraîner ces systèmes sur des données littéraires. La littérature étant en effet un domaine de spécialité, les auteures précisent que cet impératif d'adaptation au domaine vaut donc tout autant que dans d'autres cas d'utilisation de la TA, même si l'objectif des textes créatifs est fondamentalement différent comme nous le verrons plus bas.

Or, on peut trouver des essais d'adaptation pour la prose chez Besacier & Schwartz (2015) et Toral & Way (2015a). Dans le premier cas, une nouvelle est traduite de l'anglais vers le français à l'aide du système Moses (Koehn *et al.*, 2007). Celui-ci est entraîné sur 25 M de phrases hors domaine issues de la campagne IWSLT (Federico *et al.*, 2012) et post-édité ensuite par tranches d'un tiers. Chaque tranche résultante est alors utilisée pour adapter progressivement le système, qui obtient finalement un score BLEU situé aux alentours de 34.79 (par comparaison avec le texte final post-édité). S'il semble que les données ne sont pas suffisantes dans cette étude pour mener à une amélioration du résultat, les auteurs notent en revanche une diminution du temps de PE au fil de l'exercice, bien qu'il soit difficile de savoir si elle est due à l'adaptation. La durée totale de la tâche est tout de même fortement réduite par rapport à une traduction libre (divisée par deux avec une PE non experte), ce qui aurait l'avantage de réduire le coût de la traduction. Les chercheurs évoquent néanmoins la nécessité de s'assurer que la qualité n'y est pas sacrifiée, car la TA a tendance à copier le texte source et ne tient pas compte des références culturelles.

Le second cas est plus proche de Voigt & Jurafsky (2012), dans le sens où les auteurs partagent un intérêt pour les caractéristiques textuelles des textes originaux et de ceux traduits par la machine. Les mesures choisies ici sont deux propriétés qui jouent typiquement en défaveur de la TA lors de la traduction d'ouvrages littéraires ou d'articles de presse, par comparaison avec d'autres domaines (technique et institutionnel par exemple). Premièrement, le degré de spécialisation du domaine (*domain narrowness*) est évalué en fonction du caractère prédictif des modèles de langues entraînés en espagnol pour chaque domaine. Étonnamment, le corpus de presse termine loin devant le corpus littéraire, qui arrive au centre, plus proche des documents techniques et institutionnels. Dans le deuxième cas, une comparaison du degré de liberté calculé pour les traductions anglais-espagnol et espagnol-catalan montre que le domaine importe peu sur la prise de liberté, tandis que la proximité entre les langues paraît plus que déterminante.

[1] BLEU (Papineni *et al.*, 2002) est la métrique la plus utilisée pour comparer les systèmes de TA.

Toujours dans Toral & Way (2015a), les auteurs évaluent la traduction d'un roman de l'espagnol vers le catalan, qu'ils comparent à celle d'un quotidien pour le même couple de langues. Le système utilisé est entraîné principalement à partir d'un corpus d'articles de presse (et de données issues du Web pour le modèle de langue), auxquels sont ajoutés deux romans d'un même auteur pour créer un second système adapté au domaine littéraire. Dans ce cas, il apparaît que l'adaptation au domaine permet d'améliorer sensiblement les résultats (jusqu'à 53.76 points BLEU, contre 49.15 pour le système maison entraîné uniquement sur le corpus de presse et 46.52 pour Google Traduction). Les auteurs indiquent également que 20 % des phrases produites par le meilleur système sont en tous points identiques à la référence, bien qu'il s'agisse des phrases les plus courtes (7,15 % du total de mots), et que 10 % de phrases supplémentaires ne diffèrent de la référence que par cinq caractères ou moins. Par ailleurs, deux locuteur·trice·s bilingues chargé·e·s d'évaluer une sélection de 101 segments originaux et leur traduction rapportent avoir trouvé environ 40 % des phrases de la TA équivalentes à la référence, et presque 20 % de meilleure qualité (en réalité à cause des prises de liberté du traducteur), pour un total de 60 % des phrases affichant une qualité au moins égale à la traduction humaine.

4 L'approche neuronale

Ces trois études pionnières pointent ainsi toutes les trois le besoin de spécialiser les systèmes sur des données littéraires. Les deux essais présentés pour la traduction automatique probabiliste, cependant, n'ont que peu de données à leur disposition, ce qui s'explique en outre par la faible quantité de corpus disponibles dans ce domaine. L'une des conclusions tirées par Besacier & Schwartz (2015) est qu'il faudrait précisément pouvoir compter sur une plus grande représentativité des textes littéraires. Or, nous pouvons constater aujourd'hui un changement en la matière. Celui-ci a été accompagné par la venue de la traduction neuronale, et en particulier des réseaux de neurones récurrents avec mécanisme d'attention (Bahdanau *et al.*, 2015), pour lesquels l'adaptation au domaine (*domain adaptation*) est justement devenue un sujet de choix, y compris en littérature. Dans ce contexte, nous partageons ici quelques publications parues depuis 2018, en particulier autour du projet démarré dans Toral & Way (2018), qui est probablement le plus conséquent à ce jour. Remarquons aussi que l'année 2019 aura été marquée par l'apparition simultanée d'études pour plusieurs autres paires de langue, notamment grâce à l'organisation du premier atelier consacré aux « Qualités de la traduction littéraire automatique ».

4.1 La paire anglais-slovène

L'une de ces publications concerne la paire anglais-slovène, pour laquelle Kuzman *et al.* (2019) mettent au point plusieurs systèmes de traduction personnalisés dont les performances sont comparées à celles de Google Traduction. Ces systèmes sont entraînés via le module OpenNMT (Klein *et al.*, 2017) sur des données génériques (OPUS[2]), auxquelles sont ajoutés soit un corpus de textes littéraires variés (9 romans du corpus SPOOK[3]) soit un roman de la même auteure et de la même traductrice que le texte à évaluer. Dans le premier cas, le score BLEU obtenu est de 18.50, contre 19.01 et 20.75 pour le second. Les chercheur·euse·s reportent par ailleurs des scores pour d'autres systèmes entraînés uniquement sur le roman (1.78), uniquement sur le corpus littéraire varié (6.61) ainsi que sur un mélange de toutes les données (16.02). Étonnamment, l'ajout de données littéraires supplémentaires semble faire baisser le score, par comparaison au système entraîné sur le roman et sur le corpus hors domaine. Google Traduction en revanche affiche des résultats supérieurs dans tous les cas (21.97), ce que les auteur·e·s attribuent à la taille fortement limitée de leur corpus d'entraînement, et ce, bien que l'ajout de données adaptées à une auteure et à sa traductrice permette de se rapprocher de ce score.

[2] Cf. Tiedemann (2012).
[3] Cf. http://nl.ijs.si/spook/.

4.2 La paire gaélique écossais-irlandais

Dans la foulée, Ó Murchú (2019) s'est essayé lui aussi à la post-édition d'un texte littéraire, à traduire du gaélique écossais vers l'irlandais. Intergaelic, l'outil utilisé pour cette tâche, est un système hybride disponible en ligne et entraîné sur un corpus relativement restreint. Sans surprise, le texte ainsi produit comporte à la fois des phrases parfaites en l'état et des passages qui devaient entièrement être retraduits, bien que le résultat soit largement correct d'un point de vue grammatical. Par comparaison avec la traduction humaine produite par le même auteur, la TA atteint un score BLEU de 35, mais paraît moins naturelle par endroit et plus proche de la structure du texte source. De plus, le traducteur note que certains éléments ne sont pas traduits et qu'une intervention humaine est presque systématiquement requise dans le cas des noms propres, des régionalismes, des néologismes et des interjections. Il ajoute cependant que la PE est 31 % plus rapide et qu'elle lui a permis d'éviter des erreurs. Les réactions des maisons d'édition contactées et de l'auteur original vis-à-vis de ce projet ont par ailleurs été très positives et le roman sera quant à lui publié.

4.3 La paire anglais-russe

Matusov (2019) s'est lui aussi intéressé à la TLA, en tentant pour sa part d'évaluer un système pour la traduction de l'anglais vers de russe et un second de l'allemand vers l'anglais. Les systèmes en question sont entraînés grâce au module RETURNN (Zeyer *et al.*, 2018). Dans le premier scénario, les corpus utilisés sont un mélange de données hors domaine et de données littéraires (270 K phrases provenant d'OPUS Books[4]), augmentées avec un corpus synthétique de romans (2.3 M de phrases[5]) à la manière de Sennrich *et al.* (2016). Pour ce couple de langues, l'auteur reporte des points BLEU de 13.9 pour Google Traduction, de 14.2 pour le système non adapté et de 15.2 pour le système adapté. Étant donné la faible fiabilité de cette métrique (comme nous le verrons ci-dessous), une évaluation humaine a également été conduite, sous la forme d'une classification d'erreur personnalisée. Dans l'ensemble, l'évaluation indique que 20 % des phrases ne contiennent pas d'erreur et que le système maison donne lieu à moins d'erreurs graves (de sens et de syntaxe) que Google Traduction, bien qu'il produise plus d'erreurs mineures. Les erreurs de sens demeurent toutefois les plus fréquentes, tandis que le nombre d'omissions, les erreurs liées à la cohérence et les problèmes de registre sont jugés trop nombreux pour un extrait de cette taille.

4.4 La paire allemand-anglais

Dans le même article, Matusov (2019) met au point un second système pour la paire allemand-anglais. Contrairement aux études précédentes, celui-ci est entraîné sur l'architecture Transformer (Vaswani *et al.*, 2017), qui a plus récemment permis d'atteindre de meilleurs résultats dans de nombreux domaines. Les corpus utilisés pour ce faire sont les mêmes (50 K phrases d'OPUS Books pour les textes littéraires), à l'exception cependant de celui utilisé pour la rétrotraduction du corpus synthétique (10 M de phrases issues du projet Gutenberg[6]). Cette fois, les résultats sont à l'inverse du cas précédent, avec un score BLEU de 20.2 pour Google Traduction, de 18.5 pour le système généraliste et de 16.2 pour le système adapté. Cette baisse pourrait néanmoins être expliquée par la taille réduite des données d'entraînement dans ce second scénario, ou tout simplement par la rigidité de la mesure BLEU. L'auteur indique d'ailleurs que ces résultats sont à prendre avec un peu de recul, car l'utilisation de la métrique TER montre des résultats tout à fait inverses, ce que constataient déjà Kuzman *et al.* (2019) dans certains cas.

---

4 Cf. Tiedemann (2012).

5 Issues du site Web Lib.ru.

6 Cf. http://www.gutenberg.org/about/.

Ceci se note également dans l'évaluation humaine, puisque les juges estiment de meilleure qualité les textes produits par le système ayant pourtant obtenu le score le plus faible. En effet, 30 % des phrases sont jugées être d'une qualité acceptable. Il apparaît d'autre part que le système de TA adapté utilise un vocabulaire beaucoup plus riche, bien que les erreurs de sens soient plus fréquentes par comparaison avec Google Traduction. Malgré la difficulté liée à la fiabilité des métriques, l'auteur conclut donc que l'adaptation au domaine permet d'obtenir de meilleurs résultats en littérature, en particulier pour le couple allemand-anglais. Cette pratique pourrait en outre faciliter le processus de post-édition, puisque les erreurs les plus graves et les problèmes de syntaxes laissent place à des erreurs plus simples et rapides à corriger. Enfin, il est noté que des systèmes de traduction automatique opérant au niveau du texte et non plus de la phrase pourraient grandement bénéficier à la prise en charge des textes littéraires.

4.5 La paire anglais-néerlandais

Cet intérêt pour les types d'erreurs produits par la TA se note également chez d'autres chercheur·euse·s, intéressé·e·s pour leur part à la traduction littéraire automatique depuis l'anglais vers le néerlandais (Tezcan *et al.*, 2019 ; Fonteyne *et al.*, 2020). Si aucun système sur-mesure n'est proposé dans ce cas, c'est surtout parce que les auteur·e·s cherchent ici à relever les caractéristiques stylistiques et à classifier les erreurs les plus fréquentes des textes produits par les outils en ligne tels que Google Traduction. Dans ce sens, des mesures de richesse lexicale, d'entropie, de couverture lexicale et sémantique et d'équivalence syntaxique sont utilisées. Ces différents éléments sont ainsi censés refléter des traits esthétiques essentiels des textes littéraires, dans la mesure où l'expérience de lecture serait en partie tributaire de celles-ci et où la machine devrait absolument les préserver pour être considérée véritablement efficace dans ce domaine (Tezcan *et al.*, 2019). Concernant la richesse lexicale, il apparaît que le texte produit par la TA dépasse non seulement le texte source, mais aussi la référence humaine. D'autres mesures confirment cette tendance, en montrant cependant un écart plus réduit. Cette richesse apparente pourrait aussi être due en partie aux traductions erronées, ce que semble confirmer la mesure d'entropie en présentant la traduction humaine comme étant plus imprévisible. L'analyse de recouvrement lexical indique quant à elle une faible différence entre les deux traductions, ce qui pourrait signifier que la TA possède un degré de cohésion moins élevé, mais cet écart est en revanche beaucoup plus marqué pour l'équivalence syntaxique, confirmant à nouveau que la TA reste plus proche de la structure du texte source.

Pour ce qui est de l'annotation d'erreurs, les chercheur·euse·s ont employé la taxonomie SCATE (Tezcan *et al.*, 2017), évaluant les deux aspects typiques de fluidité (grammaire, vocabulaire, orthographe…) et de fidélité (ajout, omission, glissement de sens…), et ont étendu celle-ci avec deux critères supplémentaires de fluidité concernant le style et le registre. Pour ces mesures, les résultats indiquent que 44 % des phrases ne contiendraient aucune erreur et que la TAN, connue pour ses difficultés à traiter de longues phrases, parviendrait à ne pas commettre d'erreur dans des phrases allant jusqu'à un maximum de 37 mots. Comme dans d'autres études pour cette paire de langues, les erreurs de fluidité sont plus nombreuses que les erreurs de fidélité, les erreurs liées à la cohérence représentant plus de 50 % des occurrences dans le premier cas et les traductions erronées plus de 80 % dans le deuxième. Les problèmes de style et de registre arrivent quant à eux en troisième position.

Dans une seconde étude, Fonteyne *et al.* (2020) proposent de répéter cette même expérience sur un extrait plus long du texte source évalué par un second juge, afin de vérifier l'accord inter-annotateur. Après avoir pointé les difficultés inhérentes aux annotations d'erreur et nuancé les scores qui semblaient faibles à première vue, les auteur·e·s confirment ainsi les résultats obtenus précédemment concernant le nombre de phrases sans erreur (44 %), la prévalence des erreurs de fluidité sur les erreurs de fidélité, ainsi que les types d'erreurs les plus fréquents. Ils ajoutent en fin de compte qu'il faudrait davantage

d'études sur le sujet afin de voir notamment quels effets ces résultats peuvent avoir sur l'expérience de lecture, car c'est évidemment sur cet objectif que le domaine littéraire se démarque de tous les autres.

4.6 La paire anglais-catalan

D'autres auteur·e·s se sont récemment penché·e·s sur cette question, au cours d'un projet qui réunit déjà quelques publications (Toral & Way, 2018 ; Toral *et al.*, 2018 ; Moorkens *et al.*, 2018 ; Toral *et al.*, 2020 ; Guerberof-Arenas & Toral, 2020). Celui-ci vise à évaluer les performances de la TA sur la paire anglais-catalan en confrontant différents paradigmes. Il est par ailleurs intéressant de noter que ce projet est le premier à utiliser des données d'entraînement uniquement littéraires, là où les études qui précèdent avaient systématiquement recours à des corpus hors domaine afin d'obtenir des données suffisamment larges pour la mise au point du système de TA. Dans certains cas, les données littéraires sont d'ailleurs presque négligeables, ce qui tient une fois encore à la difficulté de trouver des corpus adaptés.

Selon Toral & Way (2018), néanmoins, l'émergence continue des livres électroniques pourrait changer cette réalité, en facilitant la construction de corpus parallèles. Si l'on y ajoute l'arrivée de la TA neuronale, et en particulier sa capacité à produire des traductions moins littérales et lexicalement plus riches que les systèmes statistiques, ces changements pourraient, à en croire les auteurs, changer notre perception vis-à-vis de l'utilisation de cette technologie en littérature. De fait, le corpus d'entraînement utilisé ici a été constitué de manière automatique à partir de 133 romans (1 M de phrases) et de leur traduction. Plus encore, les auteurs ont eux aussi constitué un corpus synthétique, composé d'environ 1 000 romans (5 M de phrases), pour renforcer la représentation de la langue cible. Le système est entraîné avec le module Nematus (Sennrich *et al.*, 2017), et le modèle résultant est un assemblage des quatre meilleures sorties.

Le système neuronal est ensuite comparé à un système statistique, entraîné lui aussi sur ces données. C'est le module Moses qui a servi ici à mettre au point le modèle, pour lequel un corpus hors domaine de 400 K phrases (OpenSubtitles2018[7]) et un autre corpus synthétique de 16 M de phrases (caWaC[8]) ont également été utilisés. Les deux paradigmes sont alors évalués sur 12 romans différents, parus entre 1920 et aujourd'hui. Les scores BLEU obtenus varient entre 17.94 et 38.92 pour la TAN, avec une moyenne de 32.12, contre 16.11 et 37.35 pour la variante statistique, avec une moyenne de 29.09. La variante neuronale montre donc de meilleurs résultats dans tous les cas, avec une augmentation du score BLEU pouvant varier de 3 à 14 % (10 % en moyenne)[9].

Toujours dans Toral & Way (2018), une évaluation humaine conduite sur trois des romans traduits montre que la TA neuronale est perçue d'une qualité équivalente à la traduction humaine pour 17, 32 et 34 % des cas, contre 7, 18 et 20 % pour la TA statistique. Les juges n'étant pas des traducteurs experts, une seconde évaluation est conduite dans Toral *et al.* (2020) et donne alors 15, 32 et 30 % de phrases produites par la TA équivalentes à la traduction humaine pour les mêmes romans. Une évaluation de l'effort de post-édition indique par ailleurs que l'entraînement des systèmes sur des données littéraires permet de réduire considérablement le nombre d'erreurs. Par comparaison avec un texte issu de Google Traduction, celles-ci se limitent en outre à des cas moins sévères, comme le montrait déjà Matusov (2019). Les comparaisons effectuées dans ce travail montrent en dernier lieu un net saut de performance de l'architecture Transformer par rapport aux modèles récurrents avec attention.

---

7 Cf. Lison & Tiedemann (2016) ; http://www.opensubtitles.org/.

8 Cf. Ljubešić & Toral (2014).

9 Toral *et al.* (2020) proposent des comparaisons avec d'autres systèmes sur ces mêmes données et obtiennent un score BLEU dépassant les 40, bien qu'ils ne donnent pas les chiffres exacts.

Ceci se note également lorsque l'on demande à des professionnel·le·s de post-éditer un texte produit par ces deux systèmes. Dans Toral *et al.* (2018), il a ainsi été demandé à six juges de comparer la traduction libre d'un roman et sa post-édition par les systèmes statistiques et neuronaux mentionnés plus haut[10]. Comme dans d'autres études sur la post-édition, les conclusions indiquent que cette tâche diminue l'effort cognitif fourni durant la traduction (42 %), réduit le nombre de frappes (23 %) et augmente la productivité (36 %), y compris lorsqu'elle mène à des pauses plus longues que la traduction libre (25 %). Toutes ces mesures sont les plus marquées lorsque l'on passe du système statistique au système neuronal. Selon Moorkens *et al.* (2018), les traducteur·trice·s n'ont cependant pas toujours conscience de ces gains, et tou·te·s expriment une préférence pour la traduction sans post-édition, car le sentiment de liberté et de créativité y est plus grand. Les chercheur·euse·s remarquent par ailleurs que la plus jeune génération de traducteur·trice·s montre une attitude plus positive envers la TA et accepteraient plus facilement d'y avoir recours que leurs collègues plus expérimenté·e·s. Enfin, l'unique juge à avoir plus d'un an d'expérience avec la post-édition est également celui qui trouve l'aide offerte par la TA la plus utile.

Dans une autre étude, Guerberof-Arenas & Toral (2020) reportent qu'un traducteur et une traductrice littéraires trouvent la TA plutôt utile et qu'elle produit des textes plutôt fidèles que fluides. À choisir entre la post-édition et la traduction libre, les deux ne s'accordent toutefois pas sur la méthode la plus rapide et la moins laborieuse, bien que la deuxième soit toujours préférée. Le retour d'un juge expert pointe d'autre part le fait que le texte littéraire post-édité contiendrait moins d'erreurs, mais aussi moins de tournures créatives. Enfin, une enquête sur la réception de la traduction humaine et post-éditée chez les lecteur·trice·s montre que les deux paradigmes sont évalués presque au même niveau en ce qui concerne leur réception, leur appréciation et l'intérêt suscité, malgré la difficulté à dégager des tendances. Pour cette dernière raison les auteur·e·s concluent à la nécessité de vérifier ces observations avec plusieurs paires de langues, avec d'autres ouvrages, et avec différents genres littéraires.

5 Réévaluer la paire anglais-français avec l'approche neuronale

Dans l'ensemble, donc, nous voyons que les systèmes neuronaux produisent de meilleurs résultats que les systèmes statistiques, en particulier ceux basés sur Transformer, de même que ceux entraînés sur de la littérature par comparaison aux généralistes, pour autant qu'ils soient suffisamment fournis en données. Chacun de ces changements offre ainsi des gains de performances pour le domaine littéraire, de même qu'un aperçu plus objectif des résultats que l'on pourrait attendre dans ce secteur. Selon Kuzman *et al.* (2019) et Toral *et al.* (2020), il serait également utile de voir quelle serait la performance d'un système de TA entraîné spécifiquement sur la production des mêmes auteur·e·s et traducteur·trice·s. Or, c'est précisément sur cette question que se centre notre projet de recherche, qui offrira par la même occasion de nouveaux résultats pour la paire anglais-français grâce à l'approche neuronale. Nos premiers essais, effectués sur un corpus restreint de 6 tomes (45 K phrases) issus d'une saga d'*heroic fantasy*, donnent un score BLEU de 9.24, tandis que Google Traduction et DeepL obtiennent respectivement 10.79 et 10.04. De cela, nous notons tout d'abord le fait qu'un corpus aussi petit obtienne un score comparable aux systèmes en ligne, mais aussi et surtout le faible score de ces systèmes qui obtiennent généralement des points BLEU se situant entre 20 et 30 pour de grands classiques de la littérature. Ceci, nous l'attribuons aux particularités de notre roman, qui affiche un registre relativement soutenu, des régionalismes, des tournures volontairement vieillies, ainsi que de nombreux concepts et néologismes propres à la série, mieux définis comme des irréalias (Loponen, 2009). Sans surprise, l'ajout de données génériques permet de dépasser ces scores et de se rapprocher de ceux obtenus dans les autres langues. Nos prochains travaux s'attarderont donc sur ce processus d'adaptation au domaine et sur les évaluations, automatiques et humaines, de ce texte visiblement plus complexe que d'autres pour la TA.

[10] Il est important de noter que les participant·e·s n'avaient pas tou·te·s une maîtrise parfaite de la langue source et que deux seulement avaient de l'expérience en post-édition.

6 Inconvénients et préoccupations

Évidemment, le développement de la traduction automatique laisse envisager divers changements et pose de nombreuses questions, certaines plus positives que d'autres. S'il est peut-être encore tôt pour se prononcer sur l'évolution de la TLA, celle-ci soulève tout de même des enjeux éthiques et sociétaux qui concernent la profession au sens large. Certains d'entre eux ont déjà été soulevés dans d'autres domaines, mais pourraient déjà ou prochainement concerner la littérature. Pour cette raison, nous pensons qu'il convient de considérer ceux-ci au plus tôt, pour assurer un avenir éthique et durable de la traduction littéraire (Taivalkoski-Shilov, 2019), et de le faire du point de vue de l'humain (Kenny, 2017).

Tout au long de cet article, nous avons par exemple abordé des problématiques de qualité et de créativité. S'il est difficile de tirer des conclusions sur ce sujet, en raison des résultats souvent contradictoires, il reste toutefois certain qu'une mauvaise implémentation de la TA conduirait inévitablement à une baisse de la qualité des textes produits. Celle-ci dépend en effet intimement des conditions de production et des facteurs sociaux qui entourent l'acte de traduction (Taivalkoski-Shilov, 2019). Or, une baisse de qualité pourrait à son tour avoir des conséquences négatives sur l'expérience de lecture, sur la reconnaissance du métier de traducteur·trice et sur le travail de l'auteur·e, mais aussi sur le transfert de culture, l'apprentissage des langues et les compétences linguistiques des lecteur·trice·s (*Ibid*.). Heureusement, les résultats sur ce sujet nous semblent montrer que ces effets sont dus avant tout à un manque d'expérience en post-édition. Le changement de paradigme introduit par la technologie nécessite bien sûr d'être formé et familiarisé à une tout autre méthode de travail pour pouvoir traduire dans des conditions normales et confortables. Ceci vaut tout autant pour la traduction assistée par ordinateur (TAO), dont l'utilisation en littérature a initialement suscité les mêmes rejets que la TA, mais que l'on voit apparaître de plus en plus fréquemment à mesure que les gens s'habituent à ces outils.

Les effets négatifs d'une interface peu ergonomique sur le travail et sur la qualité de la traduction résultante sont d'ailleurs d'ores et déjà bien documentés pour la TAO (Teixeira & O'Brien, 2017). Si le processus est mal conçu, la TA pourrait donc pareillement augmenter la charge cognitive des utilisateur·trice·s (Taivalkoski-Shilov, 2019) et faire en sorte qu'ils ou elles puissent difficilement se détacher des suggestions pour trouver des solutions plus créatives (Şahin & Gürses, 2019). Toral & Way (2015b) insistent dès lors sur le fait que l'introduction de la TA en littérature devrait se faire d'une manière qui s'éloignerait de la post-édition traditionnelle et qui correspondrait mieux à ce domaine. Pour cette raison, notamment, Besacier & Schwartz (2015) mentionnent l'intérêt d'une interface de PE interactive capable d'afficher plusieurs propositions, tandis que Toral & Way (2015a) évoquent un système susceptible d'être entraîné au fur et à mesure que le traducteur travaille. Enfin, le manque de contexte en PE peut mener à un manque d'homogénéité (Besacier & Schwartz, 2015), c'est pourquoi une segmentation par paragraphe devrait être possible dans l'idéal (Moorkens *et al*., 2018 ; Nunes Vieira *et al.*, 2020), ce qui vaut également pour l'utilisation des logiciels de TAO en littérature selon nous.

Ces deux préoccupations pointent de ce fait la nécessité d'intégrer pleinement les traducteurs et les traductrices au centre de ces recherches, de façon à pouvoir adapter les outils à leur travail (Ruffo, 2018). La connaissance de la technologie et l'amélioration des interfaces de travail ne sont cependant pas les seules choses à prendre en considération, puisque les traducteur·trice·s littéraires dépendent aussi des groupes d'édition. L'objectif de ces derniers restant principalement financier, l'arrivée de la TA pose potentiellement un risque pour leurs conditions de travail (Taivalkoski-Shilov, 2019). Comme l'a montré l'utilisation des outils de TAO et de la TA dans d'autres domaines, ces outils conduisent généralement à une baisse de la rémunération et à des délais encore plus courts. Dans des cas plus graves, certains vendent directement des sorties de TA non post-éditées, et l'on pourrait penser que d'autres éditeur·trice·s soient tenté·e·s d'engager des post-éditeur·trice·s non professionnel·le·s à l'avenir (*Ibid*.).

Dans la même ligne d'idée, Şahin & Gürses (2019) ajoutent que les systèmes de TA soulèvent la question du plagiat. Celle-ci serait particulièrement problématique en Turquie, selon les auteurs, où des éditeurs publient des retraductions en partie plagiées, et cette préoccupation est d'autant plus grande si ces outils sont utilisés sans intervention humaine. Néanmoins, la question des droits et de la propriété sur les données se pose quant à elle dans tous les cas, y compris lorsque des humains sont intégrés dans le processus, bien qu'il n'y ait pas encore de régulation à ce sujet (Taivalkoski-Shilov, 2019). De même, les outils automatiques posent un risque pour la visibilité des traducteurs et traductrices (Cronin, 2013). Ceci vaut bien évidemment pour tous les domaines de la traduction, mais peut-être plus encore dans le monde de l'édition, où le manque de reconnaissance représente déjà un défi de taille. Heureusement, aucun des ouvrages considérés ici n'envisage la TA autrement que comme un outil au service des professionnel·le·s, excepté lorsque ces outils servent d'aide aux apprenant·e·s d'une langue étrangère.

Kenny & Winters (2020b), de leur côté, soulignent le fait que la traduction automatique aurait tendance à atténuer la voix des traducteur·trice·s dans le texte cible. Il a aussi souvent été rapporté que les professionnel·le·s se sentent contraint·e·s par la tâche de post-édition et la segmentation par phrases, ce qui pourrait coïncider avec une baisse de la créativité. Toral *et al.* (2018) remarquent en effet que les traducteur·trice·s prennent très peu de liberté par rapport à la structure du texte source. Nunes Vieira *et al.* (2020), en revanche, notent une totale prise de liberté lorsque ces mêmes traducteurs et traductrices possèdent de l'expérience en post-édition. Dans ce même cas, les juges experts ne notent pas de différence concernant la créativité si l'on compare les textes post-édités aux références humaines (*Ibid.*). D'autre part, la TA pourrait au contraire stimuler la démarche créative en offrant d'autres possibilités de traduction ou en facilitant la traduction des passages moins importants, de manière à ce que l'humain puisse se concentrer sur les points les plus difficiles, les passages clefs où l'ingéniosité humaine est nécessaire (Şahin & Gürses, 2019).

7 Avantages

En plus d'une possible augmentation de la créativité, l'utilisation de la TA dans le domaine littéraire laisse se profiler d'autres points positifs. Selon Besacier & Schwartz (2015), on pourrait d'ailleurs y trouver un intérêt à tous les niveaux de la chaîne de traduction. Pour les traducteur·trice·s, ce sont bien entendu les possibles augmentations de la qualité et de la productivité qui ont été largement évoquées jusqu'ici. Selon Taivalkoski-Shilov (2019), les outils tels que la TA et la TAO pourraient en outre réduire la charge cognitive des utilisateur·trice·s et leur permettre de focaliser leur attention sur les tâches les plus complexes. Dans l'ensemble, ces avantages ont donc le potentiel de rendre le travail des traducteur·trice·s plus agréable (*Ibid.*), en particulier s'ils sont susceptibles d'aboutir à une prise d'importance de la dimension créative, comme nous le pensons. L'enjeu est d'autant plus important que le statut des traducteur·trice·s littéraires est déjà extrêmement précaire et les délais de traduction souvent très courts. Évidemment, cela implique de pouvoir apporter des assurances concernant les enjeux éthiques que nous venons de voir.

Pour les éditeur·trice·s, le principal avantage résiderait sans surprise dans la réduction des coûts, dont nous pourrions espérer qu'il profite aux traducteur·trice·s professionnel·e·s. Par ailleurs, la TA pourrait ainsi leur offrir la possibilité d'augmenter les commandes de traductions, faciliter leur accès à de nouveaux ouvrages et diversifier leur catalogue. Pour les auteur·e·s, l'avantage résiderait avant tout dans le fait de voir son œuvre traduite dans un plus grand nombre de langues (Besacier & Schwartz, 2015). Pour les lecteur·trice·s, enfin, ce serait la chance d'avoir accès plus rapidement aux traductions de leurs auteur·e·s favori·te·s (*Ibid.*). La TA serait plus utile encore à celles et ceux qui n'auraient pas accès à leur contenu préféré dans les langues qu'ils pratiquent. Dans ce contexte, la TA se profile comme une aide précieuse à la lecture et à l'apprentissage des langues (Oliver González *et al.*, 2019 ; Matusov, 2019).

Pour terminer, si une utilisation abusive de la TA pose un risque pour la visibilité des traducteur·trice·s, celle-ci pourrait également donner plus de visibilité à d'autres et assurer une plus grande diversité dans le panorama de la traduction littéraire. Les traducteur·trice·s et les auteur·e·s n'ont en effet pas tou·te·s les mêmes chances de voir leurs ouvrages paraître sur le marché de l'édition (Castro, 2020). Les romans écrits dans des langues à plus faible diffusion, par exemple, ont peu de chance d'attirer l'intérêt des maisons d'édition (*Ibid*. ; Toral & Way, 2015b). Or, la traduction automatique pourrait permettre à ces éditeur·trice·s de se familiariser avec d'autres œuvres ou faciliter l'envoi d'un échantillon en vue d'une publication (Matusov, 2019). De la même manière, la TA pourrait en outre favoriser la diversité de la traduction littéraire en introduisant des travaux produits par des personnes issues de minorités (Castro, 2020). Cette dernière observation ne concerne d'ailleurs pas uniquement les personnes victimes de discrimination, mais aussi simplement des traducteur·trice·s émergent·e·s (Tazelaar, 2020) ou des auteur·e·s moins connu·e·s (Matusov, 2019).

8 Discussion

Il n'est pas rare de voir des gens condamner fermement la traduction automatique et son utilisation en littérature, au motif que les résultats obtenus par les systèmes existants seraient bien trop mauvais et que ceux-ci produiraient des traductions trop littérales, là où la traduction des textes littéraires nécessite d'opérer des choix stylistiques signifiants. Or, l'un des avantages reconnus aujourd'hui de la TA neuronale est sa capacité à traduire plus librement, pour autant que les données d'entraînement soient adaptées. Il faut en effet reconnaître qu'aucun des systèmes de traduction auxquels nous avons (librement) accès à ce jour n'a été prévu pour traduire des textes littéraires. Comme le remarquaient déjà Jones & Irvine (2013) :

> « Tout comme les humains, les systèmes de traduction automatique sont capables de produire des traductions qui peuvent être tantôt littérales, tantôt plus libres, et doivent constamment opérer des choix lors du décodage [pour produire une phrase de sortie]. Dans le cas des systèmes SMT [et NMT], ces choix sont dépendants des observations contenues dans les données d'entraînement et de leur fréquence. Lorsqu'ils sont entraînés sur des jeux de données semblables au corpus de test, il est donc probable qu'ils effectuent des choix plus pertinents. »[11]

Pourtant, il aura fallu attendre jusqu'à très récemment pour que des systèmes soient adaptés au domaine littéraire. Et même s'il n'est pas toujours aisé de se faire une idée du niveau de qualité concrètement obtenu en raison du manque d'accès aux résultats, les conclusions reportées semblent prometteuses, bien qu'il reste encore du chemin à parcourir. Les défis particuliers posés par la littérature reflètent en effet les intérêts d'autres pistes de recherche très récentes en traduction automatique, notamment l'augmentation de données (Fadaee *et al*., 2017), l'adaptation au domaine (Chu & Wang, 2018) ou encore la possibilité de traiter les éléments au niveau textuel (Lopes *et al*., 2020).

Dans tous les cas, tou·te·s les auteur·e·s cité·e·s ici se rejoignent sur le fait que l'ingéniosité humaine restera indispensable pour de longues années encore, peu importe le domaine littéraire, technique, etc. Combinée aux capacités de la machine, la créativité humaine pourrait se voir renforcée, tout comme la qualité du texte produit. Cela implique comme nous l'avons dit de pouvoir compter sur une interface ergonomique qui serait spécifiquement prévue à cet effet. Pour cette raison, il nous semble justement que l'environnement de TAO représenterait le point de départ idéal — et qu'il pourrait amener une solution à d'autres problèmes, comme celui de la propriété sur les données — même s'il reste beaucoup de chemin à faire dans ce sens.

[11] Nous traduisons.

En tout état de cause, et comme le montrent particulièrement les résultats de Kuzman *et al.* (2019), l'utilisation de données adaptées au domaine ne suffit pas et nécessite de disposer de très larges corpus d'entraînement, car le besoin en données des systèmes de TA reste primordial. Toutefois, le recours aux corpus littéraires pour l'entraînement des modèles neuronaux permet effectivement d'améliorer les résultats et même d'atteindre des performances encourageantes dans le cas d'un système entraîné uniquement à partir de données littéraires suffisamment larges (Toral *et al.*, 2020).

En revanche, l'évaluation des textes produits par la machine présente encore, à l'heure actuelle, un obstacle de taille. Si la métrique BLEU (Papineni *et al.*, 2002) est encore largement utilisée en pratique et peut offrir une estimation de la qualité par rapport à une (et une seule) référence, cette mesure doit être prise avec une précaution toute particulière dans le cas de la littérature (Kit & Wong, 2015 ; Toral & Way, 2015b) et plus encore si elle concerne des systèmes entraînés sur des couples de langue, des genres et des auteur·e·s qui peuvent varier fortement d'une étude à l'autre. Pour cette raison, certain·e·s chercheur·euse·s partagent d'autres métriques[12], comme METEOR (Denkowski & Lavie, 2014), TER (Snover *et al.*, 2006) ou HTER (Snover *et al.*, 2009). Toutefois, ces mesures varient également d'une publication à l'autre, elles offrent parfois des résultats tout à fait contradictoires entre elles et cette pratique reste minoritaire, bien que de nouvelles métriques reposant moins sur les structures de surface, comme YiSi (Lo, 2019) et COMET (Rei *et al.*, 2020), laissent entrevoir de meilleures corrélations avec les jugements humains.

De plus, ces métriques n'aident pas nécessairement les lecteur·trice·s à se représenter concrètement les résultats produits par la TA. Une seconde solution consiste donc à effectuer des classifications d'erreur ou des évaluations humaines, mais la majorité des auteur·e·s rapportent dans ces deux cas des résultats hautement variables entre chaque juge (Kuzman *et al.*, 2019 ; Nunes Vieira et al., 2020 ; Guerberof-Arenas & Toral, 2020). À titre d'exemple, certains trouvent que la TA est moins utile aux traducteur·trice·s littéraires débutant·e·s (Şahin & Gürses, 2019), là d'autres trouvent précisément le contraire (Moorkens *et al.*, 2018). Enfin, les évaluations humaines sont rendues d'autant plus compliquées, et leurs conclusions d'autant plus difficiles à interpréter, par le fait que les évaluateurs sont familiers soit avec la technologie, soit avec la traduction littéraire, mais jamais dans les deux domaines (Nunes Vieira *et al.*, 2020). Parfois même, les évaluateur·trice·s ont très peu d'expérience dans ces deux tâches.

Finalement, ces études constituent tout de même des pistes de recherche intéressantes, qui ont de surcroît l'avantage de fournir de précieuses informations concernant les caractéristiques des textes créatifs ainsi que le travail particulier des traducteur·trice·s littéraires. La TA, sans être — heureusement — parfaite, reste quant à elle un outil potentiellement utile. Il n'est toutefois pas question de sombrer dans le déterminisme technologique pour autant, et c'est pourquoi ces démarches doivent être accompagnées de réflexions concernant les changements négatifs et positifs qu'elle pourrait apporter. Les auteur·e·s savent à quel point il peut être difficile d'être traduit·e. Il arrive même fréquemment que les lecteur·trice·s perdent la possibilité de suivre les suites d'une série ou de terminer une saga qu'ils avaient entamée, pour de simples raisons de coûts, de droits et de revenus. La situation des traducteur·trice·s littéraires n'est d'ailleurs pas toujours plus confortable, et des avancées dans le domaine de la TLA pourraient ainsi apporter des changements positifs à ces égards, tout comme elle pourrait contribuer à détériorer des conditions déjà précaires. Cette piste de recherche reste donc une exploration fondamentalement intéressante, qui pourrait contribuer à mieux équiper les traducteurs littéraires, mais elle doit nécessairement s'accompagner pour cela de considérations sociologiques qui indiqueraient comment l'intégrer au mieux (comme pour toutes les avancées liées au *deep learning*).

[12] Cf. Toral & Way (2015a) ; Kuzman *et al.* (2019) ; Matusov (2019) ; Guerberof-Arenas & Toral (2020).

Remerciements

Je tiens à remercier mes encadrant·e·s et évaluateur·trice·s pour leur relecture, leurs retours éclairants et leurs suggestions judicieuses.

Références

BAHDANAU D., CHO K. & BENGIO Y. (2015). Neural Machine Translation by Jointly Learning to Align and Translate. In Y. BENGIO & Y. LECUN, Éds., *3rd International Conference on Learning Representations: Conference Track Proceedings*, *ICLR 2015*, *San Diego (CA)*, *États-Unis*, *7–9 mai 2015*. ARXIV : 1409.0473.

BENTIVOGLI L., BISAZZA L., CETTOLO M. & FEDERICO M. (2016). Neural versus Phrase-Based Machine Translation Quality: a Case Study. In J. SU, K. DUH & X. CARRERAS, Éds., *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, *EMNLP 2016*, *Austin (TX)*, *États-Unis*, *1–5 novembre 2016*, p. 257–267 : ACL. DOI : 10.18653/v1/D16-1025, ARXIV : 1608.04631.

BESACIER L. & SCHWARTZ L. (2015). Automated Translation of a Literary Work: A Pilot Study. In A. FELDMAN, A. KAZANTSEVA, S. SZPAKOWICZ & C. KOOLEN, Éds., *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, *NAACL-HLT 2015*, *Denver (CO)*, *États-Unis*, *4 juin 2015*, p. 114–122 : ACL. DOI : 10.3115/v1/W15-0713, HAL : hal-01147903.

CASTRO O. (2020). Transnational Feminism, Women Writers in Translation, Stateless Cultures/Literatures in Translation. Présentation, *Creative Translation and Technologies Expert Meeting*, *Université de Surrey, Royaume-Uni*, *29 mai 2020*.

CHU C. & WANG R. (2018). A Survey of Domain Adaptation for Neural Machine Translation. In E. M. BENDER, L. DERCZYNSKI & P. ISABELLE, Éds., *Proceedings of the 27th International Conference on Computational Linguistics*, *COLING 2018*, *Santa Fe (NM)*, *États-Unis*, *20–26 août 2018*, p. 1304–1319 : ACL.

CRONIN M. (2013). *Translation in the Digital Age*. Routledge.

DENKOWSKI M. & LAVIE A. (2014). Meteor universal: Language specific translation evaluation for any target language. In O. BOJAR, C. BUCK, C. FEDERMANN, B. HADDOW, P. KOEHN, C. MONZ, M. POST & L. SPECIA, Éds., *Proceedings of the Ninth Workshop on Statistical Machine Translation*, *EACL 2014*, *Baltimore (MD)*, *États-Unis*, *26–27 juin 2014*, p. 376–380 : ACL. DOI : 10.3115/v1/W14-3348.

FADAEE M., BISAZZA A. & MONZ C. (2017). Data Augmentation for Low-Resource Neural Machine Translation. In R. BARZILAY & M.-Y. KAN, Éds., *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, *ACL 2017*, *Vancouver*, *Canada*, *30 juillet – 4 août 2017*, p. 567–573 : ACL. DOI : 10.18653/v1/P17-2090.

FEDERICO M., STÜKER S., BENTIVOGLI L., PAUL M., CETTOLO M., HERRMANN T., NIEHUES J. & MORETTI G. (2012). The IWSLT 2011 Evaluation Campaign on Automatic Talk Translation. In C. CALZOLARI, K. CHOUKRI, T. DECLERCK, M. U. DOĞAN, B. MÆGAARD, J. MARIANI, A. MORENO, J. ODIJK & S. PIPERIDIS, Éds., *Proceedings of the 8th International Conference on Language Resources and Evaluation*, *LREC'12*, *Istanbul*, *Turquie*, *21–27 mai 2012*, p. 3543–3550 : ELRA.

FONTEYNE M., TEZCAN A. & LIEVEN M. (2020). Literary Machine Translation under the Magnifying Glass: Assessing the Quality of an NMT-Translated Detective Novel on Document Level. In N. CALZOLARI, F. BÉCHET, P. BLACHE, K. CHOUKRI, C. CIERI, T. DELRECK, S. GOGGI, H. ISAHARA, B. MAEGAARD, J. MARIANI, H. MAZO, A. MORENO, J. ODIJK & S. PIPERIDIS, Éds., *Proceedings of the 12th Conference on Language Resources and Evaluation*, *LREC 2020*, *Marseille*, *France*, *11–16 mai 2020*, p. 3790–3798 : ELRA.

GENZEL D., USZKOREIT J. & OCH F. (2010). 'Poetic' Statistical Machine Translation: Rhyme and Meter. In H. LI & L. MÀRQUEZ, Éds., *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, *EMNLP '10*, *Cambridge (MA)*, *États-Unis*, *9–11 octobre 2010*, p. 158–166 : ACL.

GHAZVININEJAD M., CHOI Y. & KNIGHT K. (2018). Neural Poetry Translation. In M. WALKER, H. JI & A. STENT, Éds., *Proceedings of NAACL-HLT 2018*, *NAACL-HLT 2018*, *New Orleans (LO)*, *États-Unis*, *1–6 juin 2018*, p. 67–71 : ACL. DOI : 10.18653/v1/N18-2011.

GREENE E., BODRUMLU T. & KNIGHT K. (2010). Automatic Analysis of Rhythmic Poetry with Applications to Generation and Translation. In H. LI & L. MÀRQUEZ, Éds., *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, *EMNLP '10*, *Cambridge (MA)*, *États-Unis*, *9–11 octobre 2010*, p. 524–533 : ACL.

GUERBEROF-ARENAS A. & TORAL A. (2020). The Impact of Post-Editing and Machine Translation on Creativity and Reading Experience. *Translation Spaces*, 9(2), p. 255–282. DOI : 10.1075/ts.20035.gue, ARXIV : 2101.06125.

HASSAN H., AUE A., CHEN C., CHOWDHARY V., CLARK J., FEDERMANN C., HUANG X., JUNCZYS-DOWMUNT M., LEWIS W., LI M., LIU S., LIU T.-Y., LUO R., MENEZES A., QIN T., SEIDE F., TAN X., TIAN F., WU L., WU S., XIA Y., ZHANG D., ZHANG Z. & ZHOU M. (2018). Achieving Human Parity on Automatic Chinese to English News Translation. *ArXiv*. ARXIV : 1803.05567.

JONES R. & IRVINE A. (2013). The (Un)faithful Machine Translator. In P. LENDVAI & K. ZERVANOU, Éds., *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, *LaTeCH 2013*, *Sofia*, *Bulgarie*, *8 août 2013*, p. 96–101 : ACL.

KENNY D., Éd. (2017). *Human Issues in Translation Technology*. Routledge.

KENNY D. & WINTERS M. (2020). Another Way of Looking at Machine Translation and Literary Translation. Présentation, *Creative Translation and Technologies Expert Meeting*, *Université de Surrey, Royaume-Uni*, *29 mai 2020*.

KENNY D. & WINTERS M. (2020b). Machine translation, ethics and the literary translator's voice. *Translation Spaces*, 9(1), p. 123–149. DOI : 10.1075/ts.00024.ken.

KIT C. & WONG T.-M. (2015). Evaluation in Machine Translation and Computer-Aided Translation. In S-W. CHAN, Éd., *The Routledge Encyclopedia of Translation Technology*, p. 213–236. Routeledge.

KOEHN P., HOANG H., BIRCH A., CALLISON-BURCH C., FEDERICO M., BERTOLDI N., COWAN B., SHEN W., MORAN C., ZENS R., DYER C., BOJAR O., CONSTANTIN A. & HERBST E. (2007). Moses: Open Source Toolkit for Statistical Machine Translation. In S. ANANIADOU, Éd., *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Companion Volume, Proceedings of the Demo and Poster Sessions*, *ACL 2007*, *Prague*, *République Tchèque*, *25-27 juin 2007*, p. 177–180 : ACL.

KLEIN G., KIM Y., DENG Y., SENELLART J. & RUSH A. (2017). OpenNMT: Open-Source Toolkit for Neural Machine Translation. In M. BANSAL & H. JI, Éds., *Proceedings of ACL 2017, System Demonstrations*, *Vancouver*, *Canada*, *30 juillet–4 août 2017*, p. 67–72 : ACL. ARXIV : 1701.02810.

KUZMAN T., VINTAR Š. & ARČAN M. (2019). Neural Machine Translation of Literary Texts from English to Slovene. In J. HADLEY, M. POPOVIĆ, H. AFLI & A. WAY, Éds., *Proceedings of the Qualities of Literary Machine Translation*, *Machine Translation Summit XVII*, *Dublin*, *Irelande*, *19 août 2019*, p. 1–9 : EAMT.

LÄUBLI S., SENNRICH R. & VOLK M. (2018). Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation. In E. RILOFF, D. CHIANG, J. HOCKENMAIER & J. TSUJI, Éds., *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, *EMNLP 2018*, *Bruxelles*, *Belgique*, *31 octobre – 4 novembre 2018*, p. 4791–4796 : ACL. DOI : 10.18653/v1/D18-1512, ARXIV : arXiv:1808.07048.

LAVAULT-OLLÉON E. (2011). L'ergonomie, nouveau paradigme pour la traductologie. *ILCEA*, 14, p. 1–17. DOI : 10.4000/ilcea.1078.

LISON P. & TIEDEMANN J. (2016). OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. In C. CALZOLARI, K. CHOUKRI, T. DECLERCK, S. GOGGI, M. GROBELNIK, B. MÆGAARD, J. MARIANI, MAZO H., A. MORENO, J. ODIJK & S. PIPERIDIS, Éds., *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, *LREC'16*, *Portorož*, *Slovénie*, *23–28 mai 2016*, p. 923–929 : ELRA.

LJUBEŠIĆ N. & TORAL A. (2016). caWaC: A web corpus of Catalan and its application to language modeling and machine translation. In C. CALZOLARI, K. CHOUKRI, T. DECLERCK, H. LOFTSSON, B. MÆGAARD, J. MARIANI, A. MORENO, J. ODIJK & S. PIPERIDIS, Éds., *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, *LREC'14*, *Reykjavik*, *Islande*, *26–31 mai 2014*, p. 1728–1732 : ELRA.

LO C. (2019). YiSi - a Unified Semantic MT Quality Evaluation and Estimation Metric for Languages with Different Levels of Available Resources. In O. BOJAR, R. CHATTERJEE, C. FEDERMANN, M. FISHEL, Y. GRAHAM, B. HADDOW, M. HUCK, A. JIMENO YEPES, P. KOEHN, A. MARTINS, C. MONZ, M. NEGRI, A. NÉVÉOL, M. NEVES, M. POST, M. TURCHI & K. VERSPOOR, Éds., *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, *WMT 2019*, *Florence*, *Italie*, *1–2 août 2019*, p. 507–513 : ACL. DOI : 10.18653/v1/W19-5358.

LOPES A., FARAJIAN M. A., BAWDEN R., ZHANG M. & MARTINS A. (2020). Document-level Neural MT: A Systematic Comparison. In A. MARTINS, H. MONIZ, S. FUMEGA, B. MARTINS, F. BATISTA, L. COHEUR, C. PARRA, I. TRANCOSO, M. TURCHI, A. BISAZZA, J. MOORKENS, A. GUERBEROF, M. NURMINEN, L. MARG, M. L. FORCADA, Éds., *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, *EAMT 2020*, *Lisbonne*, *Portugal*, *3–5 novembre 2020*, p. 225–234 : EAMT. HAL : hal-02900686.

LOPONEN M. (2009). Translating Irrealia – Creating a Semiotic Framework for the Translation of Fictional Cultures. *Chinese Semiotic Studies*, 2(1), p. 165–175. DOI : 10.1515/css-2009-0117.

MATUSOV E. (2019). The Challenges of Using Neural Machine Translation for Literature. In J. HADLEY, M. POPOVIĆ, H. AFLI & A. WAY, Éds., *Proceedings of the Qualities of Literary Machine Translation*, *Machine Translation Summit XVII*, *Dublin*, *Irelande*, *19 août 2019*, p. 10–19 : EAMT.

MOORKENS J., TORAL A., CASTILHO S. & WAY A. (2018). Translators' Perceptions of Literary Post-Editing using Statistical and Neural Machine Translation. *Translation Spaces*, 7(2), p. 240–262. DOI : 10.1075/ts.18014.moo.

NUNES VIEIRA L., ZHANG X., YOUDALE R. & CARL M. (2020). Machine Translation and Literary Texts: A Network of Possibilities. Présentation, *Creative Translation and Technologies Expert Meeting*, *Université de Surrey, Royaume-Uni*, *29 mai 2020*.

OLIVER GONZÁLEZ A., TORAL A. & GUERBEROF-ARENAS A. (2019). InLéctor: Neural Machine Translation for the creation of bilingual ebooks. In J. HADLEY, M. POPOVIĆ, H. AFLI & A. WAY, Éds., *Proceedings of the Qualities of Literary Machine Translation*, *Machine Translation Summit XVII*, *Dublin*, *Irelande*, *19 août 2019*, p. VII : EAMT.

Ó MURCHÚ E. P. (2019). Using Intergaelic to pre-translate and subsequently post-edit a sci-fi novel from Scottish Gaelic to Irish. In J. HADLEY, M. POPOVIĆ, H. AFLI & A. WAY, Éds., *Proceedings of the Qualities of Literary Machine Translation*, *Machine Translation Summit XVII*, *Dublin*, *Irelande*, *19 août 2019*, p. 20–25 : EAMT.

PAPINENI K., ROUKOS S., WARD T. & ZHU W.-J. (2016). Bleu: a Method for Automatic Evaluation of Machine Translation. In P. ISABELLE, E. CHARNIAK & D. LIN, Éds., *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, *ACL-02*, *Philadelphia (PA)*, *États-Unis*, *7–12 juillet 2016*, p. 311–318 : ACL. DOI : 10.3115/1073083.1073135.

REI R., STEWART C., FARINHA A. C. & LAVIE A. (2020). COMET: A Neural Framework for MT Evaluation. In B. WEBBER, T. COHN, Y. HE & Y. LIU, Éds., *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, *EMNLP 2020*, *16–20 novembre 2020*, p. 2685–2702 : ACL. DOI : 10.18653/v1/2020.emnlp-main.213, ARXIV : 2009.09025.

RUFFO P. (2018). Human-Computer Interaction in Translation: Literary Translators on Technology and Their Roles. In D. CHAMBERS, J. DRUGAN, J. ESTEVES-FERREIRA, J. M. MACAN, R. MITKOV & O.-M. STEFANOV, Éds., *Proceedings of Translating and the Computer 40*, *TC40*, *Londres*, *Royaume-Uni*, *15–16 novembre 2018*, p. 127–131 : Éditions Tradulex.

ŞAHIN M. & GÜRSES S. (2019). Would MT kill creativity in literary retranslation? In J. HADLEY, M. POPOVIĆ, H. AFLI & A. WAY, Éds., *Proceedings of the Qualities of Literary Machine Translation*, *Machine Translation Summit XVII*, *Dublin*, *Irelande*, *19 août 2019*, p. 26–34 : EAMT.

SENNRICH R., HADDOW B. & BIRCH A. (2016). Improving Neural Machine Translation Models with Monolingual Data. In K. ERK & N. SZPAKOWICZ, Éds., *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, *ACL 2016*, *Berlin*, *Allemagne*, *7–12 août 2016*, p. 86–96 : ACL. DOI : 10.18653/v1/P16-1009, ARXIV : 1511.06709.

SENNRICH R., FIRAT O., CHO K., BIRCH A., HADDOW B., HITSCHLER J., JUNCZYS-DOWMUNT M., LÄUBLI S., MICELI-BARONE A. V., MOKRY J. & NĂDEJDE M. (2017). Nematus: a Toolkit for Neural Machine Translation. In A. MARTINS & A. PEÑAS, Éds., *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, *EACL 2017*, *Valence*, *Espagne*, *3–7 avril 2017*, p. 65–68 : ACL.

SNOVER M., DORR B., SCHWARTZ R., MICCIULLA L. & WEISCHEDEL R. (2006). A Study of Translation Error Rate with Targeted Human Annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation of the Americas, AMTA 2006*, *Cambridge (MA)*, *États-Unis*, *8–12 août 2006*, p. 223–231 : ACL.

SNOVER M., MADNANI N., DORR B. & SCHWARTZ R. (2009). Fluency, Adequacy, or HTER? Exploring Different Human Judgments with a Tunable MT Metric. In C. CALLISON-BURCH, P. KOEHN, C. MONZ & J. SCHROEDER, Éds., *Proceedings of the Fourth Workshop on Statistical Machine Translation, EACL 2009*, *Athènes*, *Grèce*, *30–31 mars 2009*, p. 259–268 : ACL.

TAIVALKOSKI-SHILOV K. (2019). Ethical issues regarding machine(-assisted) translation of literary texts. *Perspectives*, 27(5), p. 689–703. DOI : 10.1080/0907676X.2018.1520907.

TAZELAAR F. (2020). CELA: Connecting Emerging Literary Artists. Présentation, *Creative Translation and Technologies Expert Meeting*, *Université de Surrey, Royaume-Uni*, *29 mai 2020*.

TEIXEIRA C. S. C. & O'BRIEN S. (2017). Investigating the cognitive ergonomic aspects of translation tools in a workplace setting. *Translation Spaces*, 6(1), p. 79–103. DOI : 10.1075/ts.6.1.05tei.

TEZCAN A., HOSTE V. & MACKEN L. (2017). SCATE taxonomy and corpus of machine translation errors. In G. CORPAS PASTOR & I. DURÁN-MUÑOZ, Éds., *Trends in e-tools and resources for translators and interpreters*, p. 219–244. Brill - Rodopi.

TEZCAN A., DAEMS J. & MACKEN L. (2019). When a 'Sport' Is a Person and Other Issues for NMT of Novels. In J. HADLEY, M. POPOVIĆ, H. AFLI & A. WAY, Éds., *Proceedings of the Qualities of Literary Machine Translation*, *Machine Translation Summit XVII*, *Dublin*, *Irelande*, *19 août 2019*, p. 40–19 : EAMT.

TIEDEMANN J. (2012). Parallel Data, Tools and Interfaces in OPUS. In C. CALZOLARI, K. CHOUKRI, T. DECLERCK, M. U. DOĞAN, B. MÆGAARD, J. MARIANI, A. MORENO, J. ODIJK & S. PIPERIDIS, Éds., *Proceedings of the 8th International Conference on Language Resources and Evaluation*, *LREC'12*, *Istanbul*, *Turquie*, *21–27 mai 2012*, p. 2214–2218 : ELRA.

TORAL A., CASTILHO S., HU K. & WAY A. (2018). Attaining the Unattainable? Reassessing Claims of Human Parity in Neural Machine Translation. In O. BOJAR, R. CHATTERJEE, C. FEDERMANN, M. FISHEL, Y. GRAHAM, B. HADDOW, M. HUCK, A. JIMENO YEPES, P. KOEHN, C. MONZ, M. NEGRI, A. NÉVÉOL, M. NEVES, M. POST, L. SPECIA, M. TURCHI, K. VERSPOOR, Éds., *Proceedings of the Third Conference on Machine Translation: Research Papers*, *WMT 2018*, *Bruxelles*, *Belgique*, *31 octobre – 1 novembre 2018*, p. 113–123 : ACL. DOI : 10.18653/v1/W18-6312, ARXIV : arXiv:1808.10432.

TORAL A., OLIVER A. & RIBAS-BELLESTÍN P. (2020). Machine Translation of Novels in the Age of Transformer. In J. PORSIEL, Éd., *Maschinelle Übersetzung für Übersetzungsprofis*, p. 276–295. BDÜ Fachverlag. ARXIV : 2011.14979.

TORAL A. & WAY A. (2015a). Translating Literary Text between Related Languages using SMT. In A. FELDMAN, A. KAZANTSEVA, S. SZPAKOWICZ & C. KOOLEN, Éds., *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, *NAACL-HLT 2015*, *Denver (CO)*, *États-Unis*, *4 juin 2015*, p. 123–132 : ACL. DOI : 10.3115/v1/W15-0714.

TORAL A. & WAY A. (2015b). Machine-Assisted Translation of Literary Text: A Case Study. *Translation Spaces*, 4(2), p. 241–268. DOI : 10.1075/ts.4.2.04tor.

TORAL A. & WAY A. (2018). What Level of Quality can Neural Machine Translation Attain on Literary Text? In S. CASTILHO, F. GASPARI & S. DOHERTY, Éds., *Translation Quality Assessment: From Principles to Practice*, p. 263–287. Springer. ARXIV : 1801.04962.

TORAL A., WIELING M. & WAY A. (2018). Post-Editing Effort of a Novel With Statistical and Neural Machine Translation. *Frontiers in Digital Humanities*, 5(9), p. 1–11. DOI : 10.3389/fdigh.2018.00009.

VASWANI A., SHAZEER N., PARMAR N., USZKOREIT J., JONES L., GOMEZ A., KAISER L. & POLOSUKHIN I. (2017). Attention Is All You Need. In I. GUYON, U. V. LUXBURG, S. BENGIO, H. WALLACH, R. FERGUS, S. VISHWANATHAN & R. GARNETT, Éds., *Advances in Neural Information Processing Systems 30*, *NIPS 2017*, *Long Beach (CA)*, *États-Unis*, *4–9 décembre 2017*, p. 5998–6008 : Curran Associates Inc.

VOIGT R. & JURAFSKY D. (2012). Towards a Literary Machine Translation: The Role of Referential Cohesion. In D. ELSON, A. KAZANTSEVA, R. MIHALCEA & S. SZPAKOWICZ, Éds., *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*, *NAACL-HLT 2012*, *Montréal*, *Canada*, *8 juin 2012*, p. 18–25 : ACL.

ZEYER A., ALKHOULI T. & NEY H. (2018). RETURNN as a generic flexible neural toolkit with application to translation and speech recognition. In F. LIU & T. SOLORIO, Éds., *Proceedings of ACL 2018, System Demonstrations*, *Melbourne*, *Australie*, *15–20 juillet 2018*, p. 128–133 : ACL. DOI : 10.18653/v1/P18-4022, ARXIV : 1805.05225.