Revue de Traduction et Langues Volume21 Numéro1/2022
Journal of Translation Languages مجلة الترجمة واللغات
ISSN (Print) : 1112-3974 EISSN (Online) : 2600-6235

Post-édition de TA neuronale à la DGT et qualité des textes finaux : étude de cas

Neural Machine Translation Post-Editing in DGT and Final Text Quality: A Case Study

Loïc De Faria Pires
University of Mons-Belgium
Loic.DEFARIAPIRES@umons.ac.be
English studies: Literature, Language, Interpretation and Translation- ELLIT
0000-0002-9980-7175

Comment citer cet article :
De Faria Pires, L. (2022). Post-édition de TA neuronale à la DGT et qualité des textes finaux : étude de cas. *Revue Traduction et Langues21* (1), 77-98.

Reçu: 07/06/2022; **Accepté**: 21/08/2022, **Publié:** 31/08/2022

L'auteur correspondant : De Faria Pires Loïc

Keywords

Institutional Translation; Neural Machine Translation; Post-Editing; Product; Quality

Abstract

This article aims at presenting the results of a case study carried out in collaboration with the European Commission's Directorate General for Translation. This study analyses the quality of contents post-edited from Neural Machine Translation (NMT) proposals (eTranslation NMT engine) by translators with varied translation experience levels. Two types of participants were recruited: "Blue Book" interns (i.e. recently graduated translators taking part in a 5-month paid internship in DGT) and in-house translators. In order to proceed with this analysis, we used an evaluation grid created by French researchers Toudic et al. (2014), and containing nine error categories, as well as four types of effects which guide raters when they attribute severity penalties to errors. The reliability of this tool was verified by an interrater agreement score: 583 revision marks were compared in terms of 1) severity penalty, 2) category and 3) raw MT responsibility by two investigators. As far as methodology is concerned, for each source text, a NMT proposal from the eTranslation engine was post-edited by a DGT translator (10 participants; 7 in-house translators and 3 "Blue Book" interns) and revised by a DGT colleague. This procedure follows the typical DGT workflow: texts are usually first translated by a translator, then systematically revised by a colleague from the same (or sometimes, a different) translation unit. The evaluation of PE text quality was thus carried out through the revision marks introduced in the PE texts. Each of these revision marks was categorised and was attributed a penalty score ranging from 1 (minor) to 5 (critical), according to the perceived distortion of the original message and intention that the source text is supposed to convey. Severity penalties were then normalised using a 100-word basis, in order for the results to be comparable between participants and texts: a total penalty score was computed for each text, and then accordingly divided to reach a 100-word penalty score. These normalised scores enabled us to compare the perceived quality of the texts provided by our participants. Though our results cannot be generalised, since the study presented here is a case study for which no significance score could be computed (not enough data), several conclusions were reached: the overall PE text quality is higher in participants with high experience levels (senior translators) than in junior translators; participants with lower experience levels produce PE texts containing more fidelity and terminology problems than their more experienced counterparts, and professional experience does not seem to have an influence on the proportion of errors directly caused by NMT proposals. Several organisational constraints limited the scope of our study. First, the modest number of participants did not provide for significant results. Hence, a deeper study could be carried on with more volunteers, in order to reach more generalisable results. Secondly, each participant provided us with an uneven number of texts and PE words. This is due to the very nature of our study, in the framework of which translators provided us with texts coming from their daily translation tasks, which limits the quantity of collected data but increases natural validity. Furthermore, the authentic context in which this study was implemented did not enable us to collect process data: further studies could include said data, which would provide for more representative results and provide us with an insight in translators'

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

cognitive processes when post-editing. In this context, eye-tracking data could be collected, and methods such as questionnaires and think-aloud protocols could be implemented in order to link process data to the quality scores obtained in our study. Finally, studying additional language pairs would be relevant, since NMT quality tends to vary according to these.

Mots clés

Post-édition ; Produit ; Qualité ; Traduction automatique neuronale ; Traduction institutionnelle

Résumé

Dans cet article, nous nous proposons de présenter les résultats d'une étude de cas menée conjointement avec la DGT de la Commission européenne. Cette étude a pour objet l'analyse de la qualité des contenus post-édités à partir d'une traduction automatique neuronale (TAN) chez des traducteurs aux différents niveaux d'expérience. Aux fins de cette analyse, nous nous sommes appuyé sur une grille d'évaluation vérifiée au moyen d'une mesure d'accord interévaluateurs. Pour chaque texte source, une proposition de TAN issue du moteur eTranslation a été post-éditée par un traducteur de l'institution (10 participants, dont 7 fonctionnaires et 3 stagiaires « Blue Book ») et révisée par un collègue. Pour mener à bien notre analyse, nous avons évalué la qualité des textes post-édités à l'aune des marques de révision insérées dans les versions post-éditées, marques que nous avons classées en catégories et auxquelles nous avons attribué une pénalité de sévérité allant de 1 (mineur) à 5 (critique). Les pénalités ont ensuite été normalisées sur une base de 100 mots, afin de rendre les scores comparables entre les participants. Malgré le caractère non généralisable des résultats, nous avons pu conclure que la qualité des contenus post-édités est supérieure chez les traducteurs expérimentés de notre cohorte, que les participants les moins expérimentés produisent des textes post-édités contenant davantage de problèmes de sens ou de terminologie par rapport à leurs homologues plus expérimentés, et que l'expérience professionnelle ne détermine pas la proportion de phénomènes erronés directement provoqués par la TAN.

1. Introduction

Au cours de la dernière décennie, les progrès de la traduction automatique neuronale (TAN) se sont révélés fulgurants (Valdez et Lomeña Galiano, 2021, p. 86). Succédant aux approches statistiques, la TAN s'est imposée sur un large fragment du marché professionnel de la traduction, ce qui a débouché sur une réorganisation des pratiques professionnelles (Loock, 2018, p. 787) allant de pair avec de nouvelles compétences à développer (Schumacher, 2020, p. 82). Dans ce contexte, la pratique de la post-édition (PE) s'est rapidement répandue, que ce soit dans sa forme dite « rapide » ou « complète », pour des raisons de coût et de productivité (Robert, 2010, p. 138).

Des débats ont également vu le jour autour de la notion de qualité : si la question d'une qualité au rabais ne se posait pas en traduction dite « humaine » (TH), l'apparition de la notion de « *good enough* » liée à la pratique de la post-édition rapide (Koponen,

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155

2016, p. 12), a provoqué une levée de boucliers dans les rangs des professionnels de la traduction. Les institutions européennes, y compris la Commission européenne et sa direction générale de la traduction (DGT) n'ont pas échappé à cette évolution des pratiques, même si la qualité visée dans le cas des documents post-édités, qui sont sujets à une post-édition complète, est similaire à la qualité attendue d'une TH (Valero Garcés, 2018, p. 116). Si, pour maintenir un rythme de travail soutenu, les institutions européennes doivent avoir recours à la post-édition de TA, notamment lorsque les récupérations dans les mémoires de traduction sont inexistantes ou insatisfaisantes, les questions de qualité restent primordiales, au regard de la valeur juridique de certains textes traités par la DGT (Biel, 2019, p. 32) et, plus largement, de l'image d'excellence que tente de renvoyer l'institution en ce qui concerne la traduction.

Malgré cette importance de la qualité, aucune étude ne s'était, à notre connaissance, penchée sur la qualité des contenus post-édités à partir d'une TAN en contexte institutionnel avant la nôtre, qui a été réalisée dès l'adoption de la TAN dans le département de langue française de la DGT, en février 2019. Nous avons choisi d'analyser la qualité des textes PE sous le prisme de l'expérience professionnelle, en raison de plusieurs études antérieures ayant adopté le même paradigme dans des domaines différents. Ainsi, notre question de recherche principale était l'identification des différences causées par l'expérience professionnelle en termes de qualité des contenus post-édités à partir d'une TAN.

Nous avons mené ce travail selon trois hypothèses, qui seront davantage décrites dans la partie méthodologique du présent article : 1) la qualité d'ensemble des contenus post-édités est similaire dans les différents groupes de traducteurs indépendamment de leur expérience professionnelle (Čulo *et al*., 2014, p. 207; Daems, 2016, p. viii) ; 2) la fréquence de certains types de marques de révision apparaissant dans les contenus post-édités diffère selon que ces contenus ont été post-édités par des traducteurs peu expérimentés ou plus expérimentés (Depraetere, 2010, p. 1; De Almeida et O'Brien, 2010, p. 7; Guerberof Arenas, 2014, p. 72; Daems, 2016, p. 102) ; 3) les traducteurs les moins expérimentés se laissent davantage influencer par la traduction automatique brute que leurs homologues plus expérimentés (Depraetere, 2010, p. 6).

2. Cadre théorique

Nous nous attacherons ici à illustrer les fondements théoriques sous-tendant la recherche proposée.

2.1 Traduction automatique neuronale

La TAN, qui succède aux approches statistiques, est le paradigme de TA le plus récent. Même si la qualité des traductions fournies est réputée supérieure aux anciennes approches, ces avantages sont parfois nuancés par quelques limites propres au fonctionnement particulier de ces approches.

2.1.1 Fonctionnement

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Au contraire des approches statistiques (TAS), qui formulaient des propositions de TA en se fondant uniquement sur des équivalences statistiques tirées de corpus parallèles (Specia, 2012), les approches neuronales de TA intègrent une composante d'intelligence artificielle permettant un apprentissage dit « profond » et une conceptualisation des contenus traduits (Cho *et al*. 2014, p. 103).

Bien qu'ils soient également entraînés à partir de corpus parallèles, les moteurs de TAN sont capables de conceptualiser les contenus traduits au moyen d'une forme artificielle de réflexion reposant sur une composante algorithmique dénommée « couche cachée » (Bentivogli *et al.*, 2016). Le désavantage majeur de cette nouvelle approche est l'impossibilité de consulter les raisonnements de la machine (*ibid.*), ce qui rend toute correction de mauvaises solutions proposées par la machine plus ardue que dans le cas des approches précédentes. Néanmoins, des gains de qualité et de productivité peuvent être obtenus par rapport aux approches statistiques, ce qui explique la prépondérance des approches neuronales actuellement.

2.1.2 Comparaison aux approches antérieures

En termes de productivité, les différents auteurs qui se sont penchés sur la question concluent généralement qu'elle est plus élevée lors d'une PE de TAN que lors d'une PE de TAS (quelles que soient les paires de langues). Citons les travaux de Shterionov *et al*. (2018, p. 230-231), qui ont vérifié ce postulat pour toutes les paires de langues étudiées à l'exception de la paire anglais-chinois, ou encore de Jia *et al*. (2019, p. 20), qui ont vérifié ceci pour la paire anglais-chinois. Même les textes littéraires, réputés désavantagés lors d'un traitement par TA, ont fait l'objet d'études en la matière. Ainsi, Toral *et al*. (2018, p. 6) ont observé que, par rapport à une TH, la TAN permettait des gains de productivité de 36 %, alors qu'ils n'étaient que de 18 % dans le cas de la TAS. Néanmoins, des travaux infirmant cette productivité supérieure en PE de TAN ont également été réalisés, comme le soulignent Ragni et Nunes Vieira (2022, p. 144-145) dans leur recensement des études sur la question.

En ce qui concerne la qualité de la TAN en comparaison avec la TAS, la première est souvent jugée supérieure, comme explicité par Daems et Macken (2019, p. 118) à partir d'une revue détaillée de la littérature sur la question tenant compte de divers types de textes et différentes paires de langues. Cette tendance est notamment observée en évaluation humaine de la qualité (Way et Forcada, 2018, p. 192-193 ; Mutal *et al*., 2019, p. 5), modalité qui nous intéresse ici.

Ces meilleures performances de la TAN par rapport aux approches antérieures, ainsi que l'adoption de plus en plus large de la TAN sur le marché de la traduction, ont été les raisons pour lesquelles les institutions européennes, dont la DGT, ont franchi le cap et progressivement adopté ce nouveau paradigme dans le cadre de leurs activités de traduction (Eisele, 2018, p. 5).

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155

2.1.3 Avantages et inconvénients

Et, de fait, ces approches neuronales de TA présentent de nombreux avantages, qu'il convient néanmoins de nuancer par les quelques écueils restants.

Outre la qualité globalement supérieure dont nous avons fait état ci-dessous et le temps de post-édition moindre, citons la réelle conceptualisation des contenus traduits, grâce à la représentation vectorielle du texte source (Toral et Sánchez-Cartagena, 2017, p. 1063) ; un meilleur ordre des mots en langue cible que les approches statistiques (Bentivogli *et al.*, 2016, p. 9) ; ainsi que la nécessité de ne plus disposer que d'un corpus bilingue, alors qu'un corpus bilingue et un corpus monolingue en langue cible étaient requis pour les approches statistiques (Koehn, 2017).

Toutefois, le plus gros désavantage de la TAN réside dans sa tendance à privilégier la fluidité du texte cible plutôt que la fidélité au texte source : en cas de données insuffisantes ou de mauvaise compréhension, la proposition de la TAN pourra être linguistiquement correcte en langue cible, sans toutefois respecter le sens du texte original (Forcada, 201, p. 301 ; Koehn, 2017), ce qui représente un danger de taille en cas d'inattention du post-éditeur. La sensibilité des moteurs de TAN aux données bruitées constitue également un défi (Koehn, 2018), au même titre que les difficultés de prise en charge des phrases très courtes ou très longues (Bentivogli *et al.*, 2016, p. 9 ; Toral & Sánchez-Cartagena, 2017, p. 1069 ; Koehn, 2018).

2.2 Post-édition en tant que produit

Malgré ces quelques écueils, la post-édition de TAN est pratiquée dans les institutions internationales. Nous envisagerons uniquement le produit de post-édition, les études de processus ne faisant pas partie de la recherche présentée.

2.2.1 Post-édition complète

Nous entendrons la PE dans la présente contribution au sens de PE dite « complète », seule post-édition pratiquée à la DGT : « In the case of the DGT and its extremely high-quality standards, only full post-editing is applied and even the post-edits are revised by a second person » (Vardaro *et al.*, 2019, p. 2).

La PE complète peut être définie comme PE permettant « d'aboutir à un texte post-édité dont la qualité est comparable à celle d'une TH » (Schumacher et Sutera, 2022, p. 5) : nous étudierons donc la PE menée au sein de la DGT comme un produit devant égaler le niveau de qualité qui serait attendu d'une TH, à savoir un niveau excellent.

2.2.2 Évaluation de la qualité en post-édition

Même si la qualité en TA et en PE peut être étudiée au moyen de métriques automatiques ou semi-automatiques, telles que le score BLEU (pour la TA brute) ou des mesures comme le *Human Translation Edit Rate* (HTER) permettant de calculer un score de qualité en fonction du nombre de changements devant être apportés à la TA brute pour obtenir un résultat fidèle et fluide (Snover *et al.*, 2006, p. 223-224), nous nous centrerons

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

ici sur une évaluation humaine qui, bien que plus coûteuse et bruitée (Snover *et al.*, 2006, p. 223), permet une étude plus fine et libre des phénomènes relevés.

2.3 État de l'art – expérience professionnelle et qualité

Cette modalité d'évaluation humaine de la qualité nous permettra d'étudier les textes PE de la DGT selon trois aspects précis : la qualité globale des contenus, les catégories de phénomènes relevés par les réviseurs institutionnels dans les textes PE, et la responsabilité de la TA brute dans les problèmes relevés dans la post-édition. Nous présenterons donc la littérature liée à ces modalités, ainsi que les hypothèses que nous en avons dégagées.

2.3.1 Qualité globale

En ce qui concerne la qualité globale des contenus post-édités, notons que la littérature ne permet pas de dégager un consensus sur la question. Deux grandes tendances s'affrontaient lors de l'élaboration de notre méthode d'analyse : une qualité équivalente indépendamment de l'expérience professionnelle des personnes en charge de la post-édition ou, au contraire un effet de l'expérience sur la qualité des contenus PE produits. Ces tendances sont parfois tirées de travaux ayant étudié des PE produites à partir d'une TAS, et non d'une TAN : ceci constitue une limite à garder à l'esprit.

Čulo *et al.* (2014, p. 207) et Daems (2016, p. viii) s'accordent sur le fait que la PE serait de qualité généralement équivalente, quel que soit le niveau d'expérience des opérateurs la pratiquant. Cette dernière, qui a travaillé sur des textes de nature journalistique post-édités de l'anglais vers le néerlandais par ses participants, remarque ce qui suit : « Overall quality was comparable for students and professionals » (*ibid.*).

Au contraire, des auteurs tels que Guerberof Arenas (2014, p. 71), qui a travaillé avec un texte relatif à une interface utilisateur post-édité de l'anglais vers l'espagnol, ont conclu qu'une plus grande expérience en traduction allait de pair avec une qualité supérieure des contenus PE.

Nous avons, dans notre cas, retenu le premier cas de figure lors de la formulation de notre première hypothèse de travail : « La qualité globale des produits post-édités par les traducteurs peu expérimentés et les traducteurs plus expérimentés ne présente pas de différences ».

2.3.2 Catégories de problèmes

Notre deuxième hypothèse de travail, liée aux catégories de problèmes de PE relevés dans les textes institutionnels traités, est la suivante : « La fréquence de certains types de marques de révision apparaissant dans les contenus post-édités diffère selon que ces contenus aient été post-édités par des traducteurs peu expérimentés ou par des traducteurs plus expérimentés ».

Cette hypothèse découle de conclusions atteintes par plusieurs auteurs dans leurs travaux respectifs.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

En premier lieu, plusieurs auteurs ont ainsi conclu que les personnes présentant une vaste expérience de traduction avaient tendance à produire des textes PE plus fidèles au texte source que leurs homologues moins expérimentés. Ceci a notamment été le cas pour Depraetere (2010, p. 1), qui a dit des étudiants ayant participé à son étude : « most errors occur in the field of calque or translation loss » et de Daems (2016, p.viii), qui a conclu : « […] professionals made fewer adequacy errors [than students] ».

Une tendance similaire a été remarquée en lien avec la terminologie: « […] translators with less experience were less thorough with terminology and with instructions than were the more experienced ones » (Guerberof Arenas, 2014, p. 72).

Dès lors, nous étudierons ici cette hypothèse sous le prisme, d'une part, des phénomènes liés à la fidélité et, d'autre part, des phénomènes liés à la terminologie.

2.3.3 Influence de la TA brute

Finalement, une idée largement répandue lorsque l'on s'attache à étudier la post-édition est que les traducteurs les moins expérimentés auraient davantage tendance à subir une influence négative de la TA lorsqu'ils post-éditent. Ainsi, ces derniers produiraient davantage d'erreurs de PE directement imputables à la TA brute que leurs homologues plus expérimentés.

Ce postulat est notamment soutenu par Depraetere : « Students do not seem to experience problems keeping in translations that are less than perfect, and this is no doubt indicative of a striking difference in mindset between translation trainees and experienced translators » (2010, p. 6).

Dès lors, la troisième hypothèse envisagée ici est la suivante : « Les traducteurs les moins expérimentés se laissent davantage influencer par la traduction automatique brute que leurs homologues plus expérimentés ».

3. Méthodologie

Si nous avons conscience que les hypothèses formulées ci-dessus ne sont pas directement en lien avec le contexte institutionnel de la DGT et ne se penchent pas sur l'étude d'une PE réalisée à partir d'une TAN, nous devons expliquer ceci par la temporalité de notre étude : lors du début de la recherche présentée ici, la TAN venait d'être mise en œuvre à la DGT, et le recul n'était pas suffisant pour être en mesure de trouver des travaux similaires menés avec une TAN.

3.1 Contexte

C'est précisément en raison du faible nombre d'études sur la qualité en post-édition institutionnelle que nous avons décidé de mener la présente recherche.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

3.1.1 DGT, traduction automatique et post-édition

L'utilisation de la TA à la DGT ne date pourtant pas d'hier : la Commission européenne a, en effet, conclu un contrat avec le fournisseur de TA SYSTRAN en 1976, avant de passer, en 2010, au moteur de TAS interne MT@EC et d'adopter, de façon progressive et à partir de 2017, le moteur neuronal eTranslation (Stefaniak, 2020, p. 263-264). Cela fait donc près de 50 ans que la TA est exploitée dans l'institution.

Malgré cette utilisation assez ancienne, les traducteurs de la DGT ne sont pas contraints d'utiliser la TA. Ils ont le choix, lorsqu'ils acceptent un nouveau projet de traduction, de l'activer ou non. Lorsqu'elle est activée, elle apparaît directement dans l'interface de traduction pour les segments pour lesquels les récupérations dans les mémoires de traduction sont insuffisantes ou inexistantes.

3.1.2 Études préalables menées à la DGT

Pour ce qui est, précisément, de l'adoption de la TA à la DGT, une étude a été menée par Cadwell *et al.* (2016). Dans cette étude, menée auprès de l'intégralité des départements linguistiques de la DGT, les auteurs ont observé que la majorité des traducteurs activaient la TA, qu'ils la trouvaient très majoritairement utile dans le cadre de leurs fonctions, et qu'ils étaient conscients des compétences nécessaires pour pratiquer une PE de TA (Cadwell *et al.*, 2016, p. 234-235). Les raisons souvent avancées par les traducteurs choisissant d'utiliser la TA sont les suivantes : un gain de productivité, la bonne qualité de la TA brute, la source d'inspiration que fournit la proposition de TA, la réduction des frappes clavier et des clics de souris, et un intérêt personnel à l'égard des technologies (Cadwell *et al.*, 2016, p. 236).

La source d'inspiration fournie par la TA et la réduction des frappes clavier ont d'ailleurs été relevées comme arguments en faveur de l'adoption de la TA dans une étude plus récente menée par Macken *et al.* à la DGT, et plus précisément dans les départements de langue française et finnoise (2020, p. 4). Ainsi, nous constatons que la transition de la TAS vers la TAN n'a pas changé la situation entre les deux études mentionnées. Les résultats de cette deuxième étude ont par ailleurs fait état de gains généraux de productivité en PE de TA par rapport à la TH au sein de la DGT, en dépit de variations individuelles entre les participants (2020, p. 1).

3.2 Participants et corpus

Nous présenterons ici les participants et le corpus sur lesquels repose notre étude. Ce corpus n'est en rien significatif, la participation ayant été de nature volontaire tout en tenant compte de la charge de travail et des textes disponibles. En effet, la participation au projet représentait une surcharge de travail pour les participants, raison pour laquelle le corpus a été collecté auprès de volontaires (au nombre de 10). Étant donné que celui-ci est constitué de textes authentiques issus des tâches quotidiennes de traduction des traducteurs de la DGT qui ont participé, le nombre de mots et la nature des textes fournis par chaque participant diffèrent, raison pour laquelle nous nous

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155

inscrivons ici dans une étude de cas, et non dans une étude empirique qui serait statistiquement significative.

3.2.1 Recrutement

Le recrutement des participants a débuté en 2017, dans le cadre de notre recherche doctorale, qui envisageait alors les textes post-édités à partir de la TAS MT@EC. Un appel interne avait été lancé dans le département de langue française. Dans le cadre de l'étude présentée ici, à savoir la partie concernant les textes post-édités à partir de la TAN eTranslation, un nouvel appel à participation a été lancé en 2019, lors de la transition de MT@EC vers eTranslation dans le département de langue française.

Nous avons pu recruter 10 participants : 7 traducteurs fonctionnaires aux niveaux d'expérience variés (participants « EXP ») et 3 stagiaires « *Blue Book* » (stagiaires venant de terminer leur cursus universitaire et recrutés pour 5 mois à la DGT – participants « BB »). Leurs années d'expérience en traduction à la DGT ont été relevées, ainsi que le nombre de mots PE fournis (en vue d'illustrer la quantité différente de données fournies par chaque candidat). La numérotation non suivie est due à l'organisation interne du projet (certains traducteurs ayant participé aux deux volets de l'étude, tandis que d'autres n'ont fourni des textes que pour la modalité TAS et ont, par conséquent, été supprimés de la liste).

Tableau 1. Participants et mots PE (sources) fournis

Participants	Années d'expérience en traduction à la DGT	Nombre de mots PE fournis
EXP1	25	20 604
EXP3	7,75	13 823
EXP4	10	4 541
EXP5	18	5 224
EXP6	8	11 663
EXP7	10	11 078
EXP8	25	14 429
BB2	0,3	5 596
BB3	0,4	24 588
BB4	0,4	5 636

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155

Si notre intention, lors de l'élaboration du projet, était de comparer des groupes représentatifs de traducteurs, la réalité du terrain ne nous a pas permis de procéder de la sorte.

3.2.2 Corpus : volume et nature des textes

Les textes collectés font partie des tâches quotidiennes à traiter par les traducteurs. Ainsi, les textes traités sont différents pour chaque participant. Le critère principal retenu pour leur sélection a été une grande proportion de segments TA proposés par la machine (généralement plus de 50 % des segments), ainsi qu'une taille ni trop longue, ni trop courte pour les textes en question.

De ce contexte authentique de travail naît une contrainte de taille : la non-équivalence de la quantité de données collectées chez chaque participant, qui ne nous a pas permis de calculer de scores statistiques significatifs. Néanmoins, une base d'analyse par tranche de 100 mots PE a été calculée pour chaque participant, rendant les résultats numériquement comparables.

Une autre contrainte concerne les catégories de documents. En effet, une typologie classant les textes de A à D existe au sein de la DGT. Notre corpus contient des documents de nature A, B et C (A-Documents juridiques, B-Documents administratifs / documents liés aux politiques, C-Documents d'information à destination du public) (DGT, 2015, p. 4). Chaque type de texte présente des objectifs différents, comme le souligne Biel (2017, p. 38) : les textes de type A doivent être traduits de manière extrêmement précise, tandis que les textes de type C permettent une plus grande liberté de formulation, par exemple. Néanmoins, nous n'avons pas collecté assez de textes de chaque type dans la présente étude pour séparer les analyses en fonction des types de textes, bien que nous reconnaissions qu'il serait intéressant de le faire à l'avenir.

3.3 Instruments

Il est communément admis que l'évaluation de la qualité en traduction est ardue à objectiver (O'Brien, 2012, p. 55). La PE ne déroge pas à cette règle.

3.3.1 Outils envisagés et grille retenue

Ainsi, après avoir envisagé plusieurs outils, dont la grille LISA, ou encore le système DQF proposé par TAUS dans le domaine de l'évaluation de la TA/PE, nous avons retenu la grille de Toudic *et al.* (2014).

Cette grille, en plus de permettre l'octroi d'une pénalité de 1 à 5 points à chaque erreur, propose 9 catégories d'analyse : sens, addition/omission, terminologie, phraséologie, grammaire/syntaxe, orthographe/typographie, style, localisation et PAO (formatage) (Toudic *et al.*, 2014, p. 9), ce qui englobe les éléments repris dans nos hypothèses. Une précision est nécessaire entre les notions de « sens » et de « terminologie ». En effet, la terminologie englobe une dimension sémantique non prise en compte dans les catégories proposées (puisque « sens » et « terminologie » font l'objet

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155

de deux catégories distinctes). Néanmoins, il convient de préciser que les quatre effets (non décrits ici) proposés par Toudic *et al*. (2014) pour l'attribution des scores de pénalité contiennent une dimension relative à la précision du texte cible : les problèmes terminologiques altérant lourdement la fidélité au TS ont donc fait l'objet de pénalités plus élevées que les problèmes terminologiques mineurs : la dimension sémantique des soucis terminologique a donc été comptabilisée dans l'analyse. En ce qui concerne les problèmes de « sens », il s'agit des altérations de fidélité au TS ne relevant pas de problèmes liés à la terminologie.

3.3.2 Objectivation de la grille

Notons, par ailleurs, qu'un accord interévaluateurs a été calculé pour vérifier la cohérence de l'attribution des pénalités et du classement en catégories.

L'attribution des pénalités a été vérifiée par un *rho* de Spearman (permettant de prendre en compte l'écart entre les valeurs en cas de divergence) : 583 marques de révision se sont vu attribuer une pénalité située entre 1 et 5 par deux évaluateurs. L'accord obtenu est de 0.483, soit un accord « modéré » (Sim et Wright, 2005, p. 264). Ce score est justifié par quelques écarts de grande ampleur, mais retenons que parmi les 583 phénomènes envisagés, 346 (59,35 %) ont fait l'objet d'un accord parfait, et 172 (29,5 %) d'un écart d'un point.

Pour ce qui est des catégories, un coefficient *k* de Cohen a été calculé, selon les mêmes modalités. Le score obtenu était de 0.6535, soit un accord « satisfaisant » (Santos, 2017).

3.3.3 Procédure d'analyse

L'analyse a été réalisée sur des textes traités dans le cadre des tâches quotidiennes des participants. Notre personne de contact à la DGT nous transmettait des packages issus de Trados Studio, dans lesquels figuraient les textes post-édités par les participants et les marques de révision qui y sont introduites par un collègue. Le statut de chaque segment figurait dans l'interface : nous pouvions donc aisément repérer les segments post-édités. Une fois ces marques de révision consignées dans un document séparé avec le texte source, nous avions la possibilité de récupérer la TA brute pour chaque texte, afin de travailler sur notre troisième hypothèse.

Dans le cadre de la première hypothèse, nous avons attribué une pénalité de 1 à 5 à chaque marque de révision, calculé un score global de pénalité pour les segments PE de chaque texte, et reporté cette pénalité totale sur une tranche de 100 mots PE. Pour chaque participant, nous avons procédé de la sorte, et également calculé un score global pour tous les segments PE fournis par un même participant (en nous fondant sur la pénalité totale et le nombre total de mots fournis). De la sorte, nous avons été en mesure de comparer les participants en partant d'une base commune. Nous avons également comparé le nombre absolu de marques de révision, calculé selon les mêmes modalités.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155

Pour la division en catégories, nous avons classé chaque marque de révision dans l'une des catégories de la grille, et ensuite réalisé les mêmes calculs de nombre de marques et de pénalités que précédemment (sur 100 mots PE), de manière à obtenir le nombre moyen de marques de révision et de points de pénalité pour les marques appartenant à chaque catégorie chez chaque participant.

Enfin, en ce qui concerne la responsabilité directe de la TA dans les problèmes de PE corrigés par les marques de révision, nous avons comparé PE, marque de révision et TA brute pour déterminer, pour chaque phénomène, s'il provenait ou non de la TA brute. Ensuite, pour chaque participant, le pourcentage de marques de révision directement causées par la TA brute a été calculé. Ce classement dichotomique a été vérifié par un *k* de Cohen, pour lequel nous avons obtenu un score de 0.7799 (accord « satisfaisant »). Ci-dessous figure un résumé du dispositif expérimental.

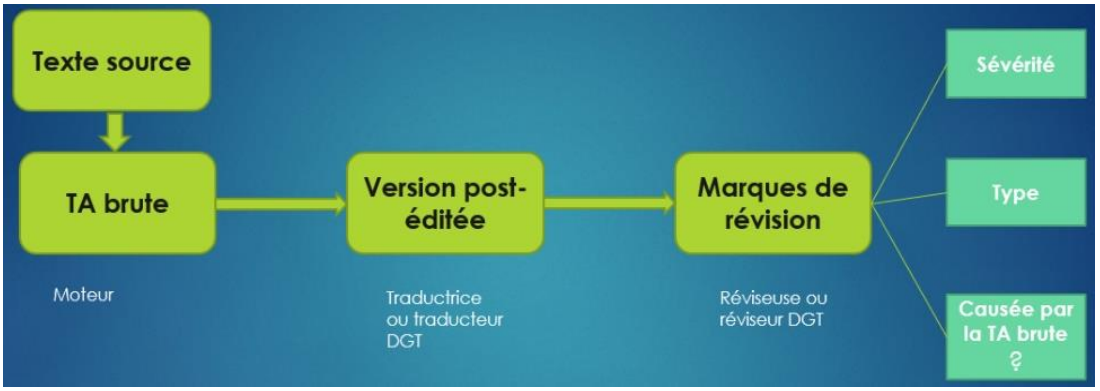


Figure 1. Dispositif expérimental

4. Résultats et discussion

Dans cette section figurent les résultats liés aux trois hypothèses, ainsi que les interprétations que nous pourrons en tirer dans le cas particulier de la DGT.

4.1 Hypothèse 1 : Qualité d'ensemble

Pour rappel, notre première hypothèse faisait état d'une qualité d'ensemble des contenus PE non influencée par l'expérience professionnelle. Voici les résultats obtenus en termes de nombre de marques de révision et de scores de pénalité.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

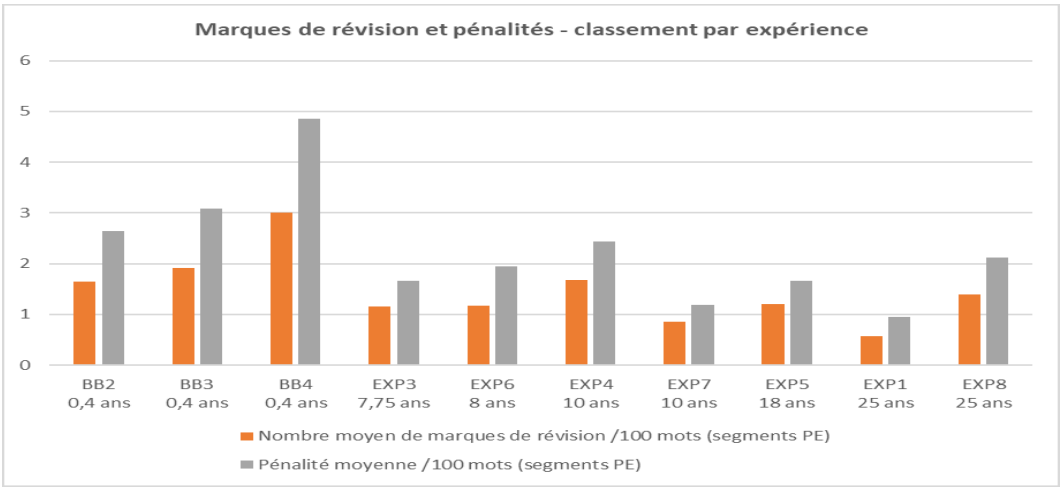


Figure 2. Résultats – qualité PE globale

Les participants sont classés, sur ce graphique, en fonction de leur nombre d'années d'expérience à la DGT (classement croissant). Ainsi, les stagiaires sont les trois premiers à y figurer, suivis par les fonctionnaires.

D'entrée, le graphique témoigne d'une différence non négligeable entre les stagiaires et les fonctionnaires. Il apparaît donc que la qualité d'ensemble des contenus PE est, ici, influencée par le niveau d'expérience professionnelle des participants. Cette tendance ressort des données chiffrées liées à ce graphique, lesquelles figurent ci-dessous.

Tableau 2. Données – qualité PE globale

	Années d'expérience pro (*DGT*)	**Nombre moyen mq. rév /100 mots PE**	**Pénalité moyenne /100 mots PE**
EXP1	*25*	**0,5727**	**0,9464**
EXP3	*7,75*	**1,1575**	**1,6639**
EXP4	*10*	**1,6736**	**2,4444**
EXP5	*18*	**1,2060**	**1,6654**
EXP6	*8*	**1,1661**	**1,9463**
EXP7	*10*	**0,8485**	**1,1825**
EXP8	*25*	**1,3861**	**2,1276**
BB2	*0,4*	**1,6440**	**2,6447**
BB3	*0,4*	**1,9196**	**3,0828**
BB4	*0,4*	**2,9986**	**4,8616**

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Nous observons dans ce tableau que les trois stagiaires ont fourni les textes PE contenant les trois pénalités moyennes les plus élevées, ce qui signifie que la qualité des textes dont ils ont été en charge est, globalement et avec toutes les précautions que nous impose la non-représentativité des données, inférieure à la qualité des textes PE fournis par tous les fonctionnaires. Ainsi, les scores de pénalité s'étalent entre 2,6447 et 4,8616 points/100 mots chez les stagiaires, et entre 0,9464 et 2,4444 chez les fonctionnaires.

En termes de nombre de marques de révision, la tendance est globalement similaire, même si le score d'EXP4 est plus élevé que celui de BB2 (dont la pénalité plus élevée indique une sévérité supérieure des phénomènes relevés).

Ceci ne nous permet pas de vérifier l'hypothèse 1, fondée sur les recherches de Čulo *et al.* (2014, p. 207) et Daems (2016, p. viii). En revanche, nos résultats rejoignent ceux de Guerberof Arenas (2014, p. 71). L'effet nivelant de la TA mentionné par certains auteurs semble donc ne pas exister dans le cas présent.

4.2 Hypothèse 2 : catégories

La deuxième hypothèse de travail faisait état de marques de révision liées au sens et à la terminologie témoignant de problèmes de PE plus sévères chez les traducteurs les moins expérimentés que chez leurs homologues plus expérimentés. Nous avons, pour étudier ceci, repris les chiffres liés à ces deux catégories selon la grille d'évaluation utilisée, dont le détail figure sur les graphiques ci-dessous.

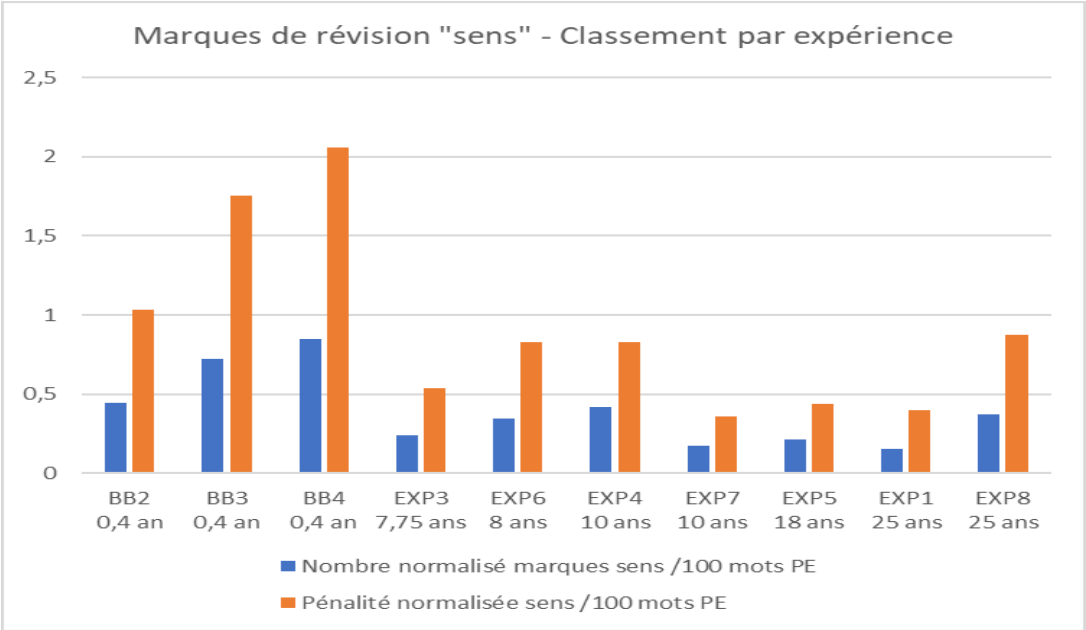


Figure 3. Résultats – qualité PE « sens »

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Nous voyons ici que la tendance observée au niveau global peut également être relevée dans le cas des marques de révision liées au respect du sens. En effet, les trois stagiaires présentent des scores plus élevés que les fonctionnaires, tant en termes de nombre de marques de révision que de scores de pénalité.

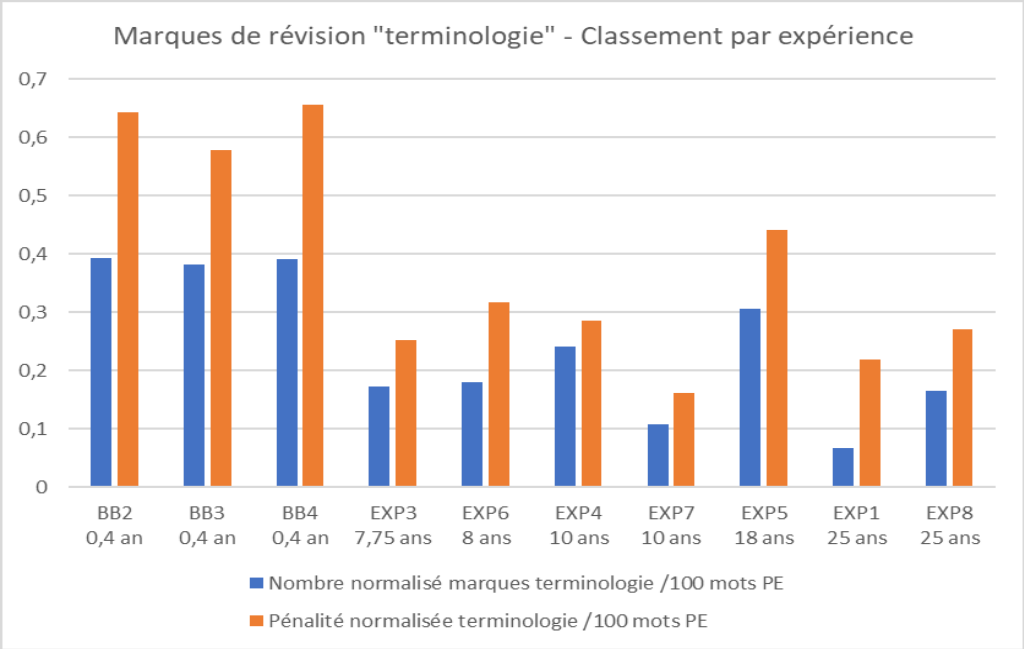


Figure 4. Résultats – qualité PE « terminologie »

La tendance est similaire pour les marques de révision liées à la terminologie : les participants BB présentent systématiquement des scores de nombre et de pénalité plus élevés que les participants EXP.

Ainsi, nous pouvons constater que les deux tendances composant notre deuxième hypothèse se vérifient ici : la plus grande sévérité des phénomènes liés au sens relevée par Daems (2016, p. viii) chez les traducteurs les moins expérimentés, et la plus grande sévérité des phénomènes liés à la terminologie relevée par Guerberof Arenas (2014, p. 72). L'hypothèse 2 est donc vérifiée, sans pour autant pouvoir être généralisée à d'autres contextes.

L'expérience en traduction des participants expérimentés semble donc les avantager également lors de la PE de TAN en ce qui concerne ces deux catégories.

4.3 Hypothèse 3 : Responsabilité de la TA brute

Finalement, notre dernière hypothèse postulait que la TA brute était à l'origine de davantage de phénomènes corrigés par des marques de révision chez les traducteurs les moins expérimentés de la cohorte.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155

Tableau 3. Données – responsabilité de la TA brute

Participant	Années d'expérience DGT	Provenance eTranslation – Marques de révision (%)	Provenance eTranslation – Pénalités (%)
EXP1	25	54,24 %	57,87 %
EXP3	7,75	48,75 %	49,56 %
EXP4	10	63,38 %	67,31 %
EXP5	18	65,08 %	65,52 %
EXP6	8	38,23 %	40,53 %
EXP7	10	57,45 %	55,72 %
EXP8	25	50 %	47,56 %
BB2	0,4	39,78 %	38,25 %
BB3	0,4	58,56 %	57,39 %
BB4	0,4	53,25 %	52,19 %

Nous remarquons ici des tendances individuelles plutôt que groupales. Ainsi, dans certains cas, davantage de marques de révision directement engendrées par la TA brute ont été détectées chez les participants les plus expérimentés que chez les stagiaires. En termes d'expérience professionnelle, aucune tendance de groupe claire ne peut être décelée ici concernant l'influence directe du moteur neuronal eTranslation sur les interventions des réviseurs dans les contenus PE fournis par leurs collègues.

Notre troisième hypothèse, fondée sur le travail de Depraetere (2010, p. 6) n'est donc pas vérifiée ici. Nous pourrions donc supposer que l'habitude plus précoce qu'ont prise les jeunes traducteurs de travailler avec une TA les aide à pratiquer une post-édition.

5. Conclusion

Nous nous sommes attaché à étudier les textes post-édités à partir d'une TA neuronale (eTranslation) dans le contexte institutionnel de la DGT, sous le prisme de l'expérience professionnelle. À cette fin, nous avons formulé trois hypothèses fondées sur des travaux antérieurs menés dans des contextes différents, qui postulaient, en substance,

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

que 1) l'expérience professionnelle n'influence pas la qualité globale des contenus PE, 2) certaines catégories de problèmes sont plus fréquentes chez les participants les moins expérimentés (sens et terminologie) et 3) les participants les moins expérimentés se laissent davantage influencer négativement par la TA brute que leurs homologues plus expérimentés. De ces trois hypothèses, seule la deuxième a pu être vérifiée. Par ailleurs, nous avons conclu que, dans le contexte de la DGT, les contenus post-édités à partir de la TA neuronale eTranslation présentent une qualité globalement supérieure chez les traducteurs les plus expérimentés, mais que l'expérience professionnelle ne semble pas ici exercer d'effet sur le degré d'influence de la TA brute sur les problèmes relevés dans ces contenus.

Toutefois, il convient de garder à l'esprit les différentes limites qui rendent impossible la généralisation des résultats. La divergence des types de textes fournis par les différents participants et la quantité différente de données récoltées pour chacun d'entre eux incite à la prudence dans l'interprétation des résultats. Pour fournir des résultats plus solides, il conviendrait de répartir les participants en groupes déterminés selon des tranches d'expérience professionnelle plus restreintes, d'obtenir une quantité équivalente de mots PE pour chaque participant, et de contrôler les types de textes collectés, de sorte à assurer la cohérence du corpus d'étude.

À une telle recherche, il serait pertinent d'ajouter une dimension concernant le processus de PE, en vue d'étudier le lien entre le produit fini fourni par les traducteurs et la manière dont ils pratiquent la post-édition. Il serait intéressant de procéder à des mesures oculométriques, à des mesures des frappes clavier, ou encore à des entretiens individuels et collectifs. Il conviendrait aussi de comparer qualité et processus entre la PE de TAN et la TH à la DGT, de sorte à étudier les éventuels gains de productivité que permet la PE de TA par rapport à la TH, et l'éventuelle influence du changement de modalité sur la qualité des textes.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155

Références

[1]Bentivogli, L., Bisazza, A., Cettolo, M. & Federico, M. (2016). Neural versus Phrase-Based Machine Translation Quality: a Case Study. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

[2]Biel, Ł. (2017). Quality in institutional EU translation: Parameters, policies and practices. In Tomas Svoboda, Łucja Biel, & Krzysztof Łoboda (Eds.), *Quality aspects in institutional translation: Translation and Multilingual Natural Language Processing* vol.8, Berlin: Language Science Press, 31-58.

[3]Biel, Ł. (2019). Theoretical and methodological challenges in researching EU legal translation. Legal Translation. In Ingrid Simonnæs & Marita Kristiansen (Eds.), *Current Issues and Challenges in Research, Methods and Applications*, Berlin: Frank & Timme, 25-39.

[4]Cadwell, P., Castilho, S., O'Brien, S. & Mitchell, L. (2016). Human factors in machine translation and post-editing among institutional translators. *Translation Spaces 5*(2), 222-243.

[5]Cho, K., Van Merriënboer, B., Bahdanau, D. & Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103-111.

[6]Čulo, O., Gutermuth, S., Hansen-Schirra, S. & Nitzke, J. (2014). The Influence of Post-Editing on Translation Strategies. In Sharon O'Brien, Laura Winther Balling, Michael Carl, Michel Simard, & Lucia Specia (Eds.), *Post-editing of Machine Translation: Processes and applications*, Newcastle upon Tyne: Cambridge Scholars Publishing, 200-218.

[7]Daems, J. (2016). *A translation robot for each translator? - A comparative study of manual translation and post-editing of machine translations: process, quality and translator attitude*. Gent: thèse de doctorat soutenue à la Universiteit Gent.

[8]Daems, J. & Macken, L. (2019). Interactive adaptive SMT versus interactive adaptive NMT: a user experience evaluation. Machine Translation 33(1), 117-134.

[9]De Almeida, G. & O'Brien, S. (2010). Analysing post-editing performance: correlations with years of translation experience. *Proceedings of the 14th Annual Conference of the European Association for Machine Translation*.

[10] Depraetere, I. (2010). What counts as useful advice in a university post-editing training context? Report on a case study. Proceedings of the 14th Annual Conference of the European Association for Machine Translation.

[11] Direction générale de la traduction. (2015). DGT Translation Quality Guidelines. https://ec.europa.eu/translation/maltese/guidelines/documents/dgt_translation_quality_guidelines_en.pdf

[12] Eisele, A. (2018). Rolling out Neural MT within CEF/eTranslation [Communication interne présentée à la DGT].

[13] Forcada, M. (2017). Making sense of neural machine translation. Translation Spaces

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155

6(2), 291-309.
[14] Guerberof Arenas, A. (2014). The Role of Professional Experience in Post-editing from a Quality and Productivity Perspective. In Sharon O'Brien, Laura Winther Balling, Michael Carl, Michel Simard, & Lucia Specia (Eds.), *Post-editing of Machine Translation: Processes and applications*, Newcastle upon Tyne: Cambridge Scholars Publishing, 51-76.
[15] Koehn, P. (2017). What is Deep Neural Machine Translation and how can it be applied? *Webinaire Omniscien Technologies*.
[16] Koehn, P. (2018). Machine Translation Primer - Current Technology and Future Directions. Webinaire Omniscien Technologies.
[17] Koponen, M. (2016). *Machine translation post-editing and effort: Empirical studies on the post-editing process*. Helsinki: these de doctorat soutenue à l'Université d'Helsinki.
[18] Jia, Y., Carl, M. & Wang, X. (2019). Post-editing neural machine translation versus phrase-based machine translation for English–Chinese. *Machine Translation* 33(1), 9-29.
[19] Loock, R. (2018). Traduction automatique et usage linguistique : une analyse de traductions anglais-français réunies en corpus. *Meta* 63(3), 786-806.
[20] Macken, L., Prou, D. & Tezcan, A. (2020). Quantifying the effect of machine translation in a high-quality human translation production process. *Informatics* 7(2).
[21] Mutal, J. D., Volkart, L., Bouillon, P., Girletti, S. & Estrella, P. S. (2019). Differences between SMT and NMT Output-a Translators' Point of View. *Second Workshop on Human-Informed Translation and Interpreting Technology*.
[22] O'Brien, S. (2012). Towards a Dynamic Quality Evaluation Model for Translation. *The Journal of Specialised Translation* 17, 55-77.
[23] Schumacher, P. (2020). La traduction automatique neuronale : technologie révolutionnaire ou poudre de perlimpinpin ? Compte rendu d'une expérience pédagogique. *Al-Kīmiyā*, 18, 67-89.
[24] Shterionov, D., Superbo, R., Nagle, P. et al. (2018). Human versus automatic quality evaluation of NMT and PBSMT. *Machine Translation* 32, 217-235.
[25] Specia, L. (2012). Fundamental and New Approaches to Statistical Machine Translation. *Proceedings of the International Conference on Computational Processing of the Portuguese Language*.
[26] Stefaniak. K. (2020). Evaluating the usefulness of neural machine translation for the Polish translators in the European Commission. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 263–269.
[27] Ragni, V. & Nunes Vieira, L. (2022). What has changed with neural machine translation? A critical review of human factors. *Perspectives*, 30(1), 137-158.
[28] Robert, A. M. (2010). La post-édition : l'avenir incontournable du traducteur ? Traduire 222, 137-144.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Disponible en ligne à https://www.asjp.cerist.dz/en/Articles/155

[29] Schumacher, P., & Sutera, A. (2022). Analyse comparative de post-édition et de traduction humaine en contexte académique. In Carmen Expósito Castro, María del Mar Ogea Pozo, & Francisco Rodríguez Rodríguez (Eds.), *Theory and practice of translation as a vehicle for knowledge transfer*, Séville: Editorial Universidad de Sevilla.

[30] Snover, M., Dorr, B., Schwartz, R., Micciulla, L. & Makhoul, J. (2006). A Study of Translation Edit Rate with Targeted Human Annotation. *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*, 223-231.

[31] Santos, F. (2017). Le kappa de Cohen : un outil de mesure de l'accord inter-juges sur des caractères qualitatifs. https://www.pacea.u-bordeaux.fr/IMG/ pdf/Kappa_ Cohen.pdf

[32] Sim, J. & Wright, C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy* 85(3), 257-268.

[33] Toral, A. & Sánchez-Cartagena, V. (2017). A Multifaceted Evaluation of Neural versus Phrase-Based Machine Translation for 9 Language Directions. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 1063-1073.

[34] 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).

[35] Toudic, D., Hernandez Morin, K., Moreau, F., Barbin, F. & Phuez, G. (2014). Du contexte didactique aux pratiques professionnelles : proposition d'une grille multicritère pour l'évaluation de la qualité en traduction spécialisée. *ILCEA - Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*, 19.

[36] Valdez, C & Lomeña Galiano, M. (2021). Exploration de la traduction automatique neuronale espagnol-français : Pour une traductologie de corpus appliquée à l'analyse des outils de traduction. *Revue Traduction et Langues* 20(1), 85-111.

[37] Valero Garcés, M. C. (2018). Interview with Spanish Language Department. Directorate-General for Translation (DGT) European Commission. José Luis Vega (Head of Department); Alberto Rivas (Quality Officer) and Luis González (Terminologist). *FITISPos International Journal* 5, 114-122.

[38] Vardaro J., Schaeffer M. & Hansen-Schirra S. (2019). Translation Quality and Error Recognition in Professional Neural Machine Translation Post-Editing. *Informatics* 6(3).

[39] Way, A. & Forcada, M. (2018). Editors' foreword to the invited issue on SMT and NMT. *Machine Translation* 32, 191-194.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International

Remerciements
Nous tenons à remercier chaleureusement le réseau LTT pour l'organisation des 12e Journées du Réseau LTT en 2021, ainsi que pour la publication de l'article consécutif à celles-ci. Nous remercions également les sponsors de l'événement, sans qui sa tenue n'aurait pas été possible. Nous tenons également à remercier le CERIST d'héberger la revue TRANSLANG sur ASJP. Finalement, nous remercions les participantes et participants aux 12e Journées du Réseau LTT pour la richesse de nos échanges et les contacts qui ont été créés.

Notice biographique de l'auteur
Loïc de Faria Pires est actuellement chargé de cours à la Faculté de Traduction et d'Interprétation (FTI-EII) de l'Université de Mons (Belgique). Il est titulaire d'un master en traduction multidisciplinaire (2016) et d'un doctorat en langues, lettres et traductologie (2020), obtenus au sein de cette même université. Ses recherches portent sur la post-édition de traduction automatique (neuronale), la traduction institutionnelle, la localisation et la traduction audiovisuelle. Il est en charge des cours d'informatique appliquée à la traduction, d'outils d'aide à la traduction, de localisation, de sous-titrage et de doublage.

Cette œuvre est sous la licence Creative Commons Attribution-NonCommercial 4.0 International