

Exploration des répercussions de la TA neuronale sur la langue cible après post-édition en contexte d'apprentissage : qu'en est-il du *post-editese* ?

Exploring the impact of neural MT on the target language after post-editing in a pedagogical context: What about post-editese?

Perrine Schumacher
CIRTI, Université de Liège (Belgique)

Résumé

Le secteur de la traduction vit un tournant technologique majeur, marqué par une demande croissante en post-édition (PE) et l'arrivée des outils de traduction automatique (TA) dans le quotidien des professionnel·les. Cette étude explore les effets de la TA neuronale sur la qualité de la langue cible après PE (anglais-français) en contexte d'apprentissage. Nos deux corpus de productions d'étudiant·es ont été analysés au moyen de trois mesures linguistiques automatiques : richesse lexicale, longueur moyenne des phrases et équivalence syntaxique. Les résultats montrent que, bien que la densité lexicale et la longueur moyenne des phrases ne diffèrent pas notablement entre la PE et la traduction humaine (TH), les textes post-édités, surtout à partir de *DeepL*, sont moins variés lexicalement et plus proches syntaxiquement de la langue source qu'en TH, révélant l'existence de traits propres à une langue de PE (*post-editese*).

Mots clés

post-édition, *post-editese*, TA neuronale, richesse lexicale, équivalence syntaxique

Abstract

The translation industry is undergoing a major technological shift, marked by a growing demand for post-editing (PE) and the emergence of machine translation (MT) tools as an essential part of the translators' workflow. This study aims to explore the effects of neural machine translation (NMT) on the quality of the target language after PE (English-French) in a pedagogical context. Three linguistic metrics – lexical richness, average sentence length and syntactic equivalence – were used to analyse our two corpora of student's productions. The results show that, although

lexical density and average sentence length do not differ significantly between PE and human translation (HT), post-edited texts – especially when using DeepL –, were found to be less lexically diverse and syntactically closer to the source language than HT, revealing the evidence of typical features in MTPE (post-edited).

Keywords

post-editing, post-edited, NMT, lexical richness, syntactic equivalence

1. INTRODUCTION

Avec l'essor de l'IA, notamment dans le domaine du *deep learning*, l'industrie des langues et, en particulier, le monde de la traduction, connaît un profond bouleversement technologique. La demande en post-édition¹ (PE) croît de manière constante sur l'ensemble du secteur, tandis que les outils de traduction automatique (TA) neuronale sont de plus en plus souvent intégrés à l'environnement de travail des traducteur-ices (Martikainen, 2022), ainsi qu'aux programmes de formation en traduction.

Plusieurs chercheur-ses ont constaté que l'usage de ces outils de TA neuronale avait pour effets d'entraîner une diminution de la richesse lexicale comparativement à la traduction humaine (TH) et ce serait le cas, non seulement, en TA brute, mais également après l'intervention humaine en PE (Farrell 2018 ; Martikainen 2023 ; Toral 2019). Certain-es affirment même que cette tendance peut conduire à un appauvrissement de la langue cible (Deneufbourg 2021 ; Toral 2019 ; Vanmassenhove, Shterionov & Gwilliam 2021). Cette perte de richesse lexicale se manifesterait, entre autres, par des phénomènes de simplification et de standardisation en langue cible. En d'autres termes, ces études montrent que les textes traduits automatiquement, même après PE, tendent à être plus simples sur le plan lexical et plus standardisés qu'en TH. On y observe une uniformité plus marquée tant dans les tournures usitées que dans le choix des mots comparativement à la TH (Farrell 2018).

Dans ce contexte de mutation technologique, cette contribution propose d'étudier les répercussions de la TA neuronale sur la qualité linguistique de la langue cible après PE (anglais-français) en contexte d'apprentissage au moyen de différentes mesures linguistiques automatiques. Ces méthodes sont couramment utilisées en traitement automatique des langues (TAL) et dans les études qui s'attachent à démontrer l'existence de caractéristiques propres à la langue traduite (*translationese*) ou à la langue de post-édition (*post-editedese*). Cette étude est issue de nos travaux doctoraux qui intègrent une combinaison d'approches qualitative et quantitative et qui ont déjà fait l'objet de publications (voir notamment Schumacher 2019 et 2023b).

Après avoir présenté les objectifs poursuivis, le corpus d'étude et l'hypothèse générale de recherche, nous détaillerons séparément les trois mesures linguistiques utilisées. Pour chaque mesure, nous fournirons un aperçu des principales recherches

¹ Par « post-édition », nous entendons l'intervention humaine visant à modifier et à améliorer un texte traduit automatiquement par machine.

traductologiques et linguistiques pertinentes, formulerons une hypothèse spécifique de recherche, détaillerons la méthode de calcul, et analyserons les résultats obtenus. Nous commencerons par la mesure de richesse lexicale à travers deux dimensions : la densité lexicale et la diversité lexicale. Ensuite, nous nous pencherons sur la longueur moyenne des phrases, corrélée au coefficient de foisonnement, avant d'examiner la mesure d'équivalence syntaxique, qui permet d'étudier la proximité syntaxique d'un texte cible par rapport à un texte source. Nous terminerons par une synthèse et conclusion.

2. OBJECTIFS

Cette étude vise à comparer la langue post-éditée (PE), la langue traduite (TH) et la langue traduite automatiquement (TA), et de repérer la présence de traits caractéristiques à une langue de PE (*post-edited*) qui se différencierait de la langue traduite (*translationese*) et de langue traduite automatiquement (*machine translationese*). Plus largement, notre volonté est de contribuer à la réflexion sur l'intégration d'une formation en TA et en PE dans les cursus des apprenant·es en traduction. En effet, étant donné que ces étudiant·es ont déjà accès gratuitement à ces outils, il est crucial de susciter et d'encadrer leur usage informé et raisonné. Et il paraît, dès lors, impérieux de mener une recherche de pointe en la matière et de l'intégrer aux enseignements de façon raisonnée afin de former des futur·es traducteur·ices en phase avec l'évolution de leur pratique et conscient·es des multiples dimensions et enjeux de ces développements.

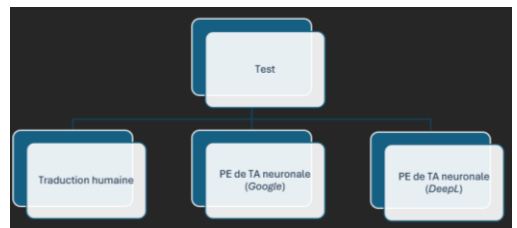
3. CORPUS

Les données sont tirées de deux corpus (2018 et 2021) recueillis dans le cadre de notre travail de thèse de doctorat (Schumacher 2023a). Ces corpus proviennent de deux expériences contrôlées menées avec des étudiant·es en traduction de 3^e bachelier (prétest) et de 2^e master (test). Dans ces expériences, les étudiant·es ont traduit trois textes sources de l'anglais au français selon trois modes de traduction (Figures (1) et (2)).

Figure 1 : Schéma Prétest



Figure 2 : Schéma Test



Nous avons opté pour des moteurs de TA générique, accessibles gratuitement en ligne, car ce sont les plus populaires à ce jour et, sans surprise, ceux que les étudiant·es utilisent le plus fréquemment (Loock et Léchaugnette, 2021). Les textes sources (TS) sont identiques dans les deux expériences. Il s'agit de trois courts

textes (300-340 mots) : un article de presse (*The Guardian*), un article économique (PWC) et un article de vulgarisation scientifique (*Scientific American*). Ainsi, le corpus 2018 (prétest) est composé de 25 TH, de 25 PE de TA statistique (*BingTranslator*), de 25 PE de TA neuronale (*DeepL*) et des six TA brutes. Le corpus 2021 (test) se compose de 24 TH, de 24 PE de *Google*, de 24 PE de *DeepL*, ainsi que des six TA brutes. Pour plus d'informations sur les paramètres expérimentaux, consulter Schumacher (2023a).

4. HYPOTHESE GENERALE

En nous appuyant sur la littérature scientifique, nous estimons que les étudiant·es en traduction sont influencé·es dans une certaine mesure par la TA (*priming effect*) et, dès lors, que nos observations vont révéler l'existence de caractéristiques propres à une langue de PE (*post-editese*), qui la distingueraient de cette sorte d'interlangue propre aux textes traduits appelée *translationese* (Gellerstam 1986). En outre, étant donné que nos corpus sont constitués de productions d'élèves, nous nous attendons à ce que les textes post-édités soient plus proches linguistiquement du TS et de la sortie de la TA que les textes traduits humainement. En effet, nous pensons qu'en PE, les étudiant·es ont notamment tendance à se satisfaire davantage de traductions littérales et à commettre un plus grand nombre de calques² par rapport à la TH. Enfin, nous postulons que cette plus grande proximité linguistique sera observable dans les deux corpus, c'est pourquoi nous ne prévoyons pas d'observer de grandes divergences entre les résultats du prétest et du test.

5. MESURES LINGUISTIQUES AUTOMATIQUES

5.1. Richesse lexicale

En dépit du fait que la richesse lexicale est un concept linguistique qui demeure complexe tant « sa terminologie et ce qu'elle recouvre diffère [*sic*] entre les chercheurs » (Bonvin & Lambelet, 2019 : 66), nous avons souhaité reprendre la définition singulière et subtile qu'en donne E. Brunet (2015) dans son étude statistique portant sur la répétition en écriture, car il est aisé d'établir une analogie avec l'activité de traduction :

[P]lutôt que de richesse lexicale, on devrait parler de pauvreté lexicale ; quand l'homme puise dans la langue pour actualiser son discours, les mots qui lui viennent à l'esprit sont ceux du voisinage qu'un lien unit à la situation, au sujet, au genre, au registre. [...] Si le hasard présidait seul à la fourniture du matériel lexical, la variété serait plus riche et les redites plus rares. C'est d'ailleurs ce qui rend malaisée l'application au langage des lois classiques de probabilité, le hasard des tirages aléatoires pouvant difficilement servir de modèle à une opération humaine qui manifeste des choix, des refus et des préférences et glisse la liberté dans les mailles de la contrainte. (Brunet, 2015 : 5)

² Selon Delisle (2013), la « traduction-calque » est une « [t]raduction littérale qui consiste à transposer dans le texte d'arrivée les éléments du texte de départ de manière à reproduire leurs aspects sémantiques, étymologiques et temporels. Note : La traduction-calque est le cas extrême de la traduction littérale et tend vers la non-traduction (p. 688).

Ainsi même dans le domaine de la statistique linguistique, la richesse lexicale est un concept relatif et ce n'est pas le père de la discipline, Ch. Muller, qui dira le contraire :

C'est bien évidemment une notion relative, et le vocabulaire d'un texte ne peut être dit "riche" ou "pauvre" que par rapport à d'autres. Il s'agit donc, au-delà des impressions subjectives, de comparer entre elles des valeurs [...] obtenues sur différents textes, et d'arriver ainsi à un classement objectif. (Muller, 1969 : 36)

Comment mesurer cette richesse lexicale ? Dans cette étude, elle a été mesurée à l'aune de deux indices : la densité lexicale et la diversité lexicale.

5.1.1. La densité lexicale

L'indice de densité lexicale (Ure 1971 ; Halliday 1985) est une mesure de complexité linguistique : « a kind of complexity that results from the development of words. In other words, this relates to the notion of *lexico-grammar* in terms of the level of wording in language » (To, Fan & Damon, 2013 : 61). Alors que cet indice est fréquemment employé dans les études sur la PE (Volkart & Bouillon 2022), il ne nous a toutefois pas été possible de dégager une tendance claire puisque les principaux travaux en la matière ont abouti à des résultats contradictoires, parfois au sein d'une même étude (*cf.* Castilho, Resende & Mitkov 2019 ; Castilho & Resende 2022).

Selon Scarpa (2006), plus cette densité lexicale est élevée, plus difficile sera la compréhension du texte par le lectorat. Dans son étude de corpus pour la traduction de l'anglais en italien, l'autrice (*ibid.*) a notamment montré que les textes traduits manuellement présentaient une moindre densité lexicale que les textes sources. Cette même tendance sous-tend l'hypothèse qu'ont formulée Castilho et Resende dans leur étude comparative (TA, PE et TH) pour la traduction de textes littéraires :

As lower lexical density is a way of building redundancy and making a text simpler, the simplification hypothesis claims that HT texts present lower lexical density than comparable non-translated texts. We expected that the MT version would be similar to the source with lower lexical density compared to the HT, and consequently, the PE versions would follow the MT output as the PE versions originated in the MT, meaning PE versions would present lower lexical density than the HT. (Castilho & Resende, 2022 : 9)

Pourtant, pour l'un des deux extraits de roman sélectionnés, l'analyse linguistique infirme leur hypothèse étant donné qu'il ne leur a pas été possible de mettre en évidence des différences de densité lexicale entre ces trois types de productions :

In a closer examination of the translated versions [...], we found that the number of adjectives, adverbs, nouns, and verbs were very similar between HT, MT, and PEs, thus resulting in close lexical density averages (*ibid.*).

En 2019, dans son étude comparative de trois systèmes de TA menée pour cinq combinaisons de langues³, Toral (2019) a conclu à une densité lexicale plus faible en TA et en PE par rapport à la TH. Ses résultats montrent également une différence inattendue entre la TA neuronale et la TA statistique puisqu'il a constaté une plus faible densité lexicale en TA neuronale.

Volkart et Bouillon (2022) ont étudié un corpus composé de projets réels de traduction et de PE de communiqués de presse (anglais-français) émanant de la Banque européenne d'investissement (BEI). Leurs résultats en matière de densité lexicale rejoignent ceux de Toral (2019) et indiquent que les textes cibles sont moins denses que les textes sources et que cette perte de densité est plus grande en PE par rapport à la TH.

5.1.1.1. Hypothèse – densité lexicale

Pour notre part, en nous fondant sur les conclusions de Toral (2019) et de Volkart et Bouillon (2022), nous émettons l'hypothèse que la densité lexicale sera plus faible, non seulement en TA, mais aussi en PE comparativement à la densité lexicale calculée en TH.

5.1.1.2. Méthode de calcul

Pour mesurer la densité lexicale (Figure (3)), il s'agit d'établir le rapport entre le nombre de « vocables lexicaux » ou « lexèmes⁴ » (adjectifs, adverbes, verbes et substantifs) – par opposition aux « vocables fonctionnels » (prépositions, articles etc.) – et le nombre total de mots (occurrences) contenus dans un texte (Lu 2012).

Figure 3 : Calcul de densité lexicale

$$\text{Densité lexicale} = \frac{\text{Nombre total}_{\text{vocables lexicaux}}}{\text{Nombre total}_{\text{occurrences}}} \times 100$$

Dans cette étude, elle a été calculée automatiquement à l'aide du logiciel *UAM corpus tool* (O'Donnell 2016). Afin d'obtenir une comparaison pertinente des données, nous avons retenu huit productions par méthode de traduction dans le prétest et sept productions par méthode dans le test. Cela explique que nous n'obtenons pas des valeurs identiques dans les tableaux (1) et (2) pour les TS alors qu'il s'agit bel et bien des mêmes TS dans le prétest et le test. Les résultats sont repris par expérience dans ces tableaux.

5.1.1.3. Résultats

Il apparaît très nettement dans les tableaux (1) et (2) que les textes sources sont moins denses lexicalement que les autres productions, mais il se peut que cela reflète uniquement des divergences entre l'anglais et le français. En revanche, les valeurs obtenues en matière de densité lexicale pour les TH, les TA brutes et les PE sont très similaires, ce qui infirme notre hypothèse et contredit les résultats de précédents

³ Allemand-anglais ; anglais-allemand ; espagnol-allemand ; anglais-français et chinois-anglais.

⁴ Unité de base du lexique, unité minimale de signification ; morphème lexical libre ou lié (Antidote).

travaux sur lesquels nous nous sommes appuyée (Toral 2019 ; Volkart & Bouillon 2022). Malgré des valeurs très proches, nous pouvons affirmer que, dans le prétest, tout comme dans le test, c'est en TH que le taux de densité lexicale est très légèrement plus élevé (67 % et 67,85 %). De plus, les résultats du test révèlent que les TA brutes sont légèrement moins denses lexicalement (65,49 % et 65,15 %) que les TH et que les PE. Ces résultats confirment l'hypothèse de travail qu'ont formulée Castilho et Resende (2022).

Tableau 1 : Densité lexicale – Prétest

Prétest	Lexèmes/phrased	Densité lexicale (%)
TS	10,6	52,12
TH	14,4	67
TAS	15	66,21
TAN (<i>DeepL</i>)	14,3	66,19
PE de TAS	15	66
PE de TAN (<i>DeepL</i>)	14,6	66,59

Tableau 2 : Densité lexicale – Test

Test	Lexèmes/phrased	Densité lexicale (%)
TS	10,6	51,79
TH	13,7	67,85
TA (<i>Google</i>)	14,6	65,49
TA (<i>DeepL</i>)	14,9	65,18
PE (<i>Google</i>)	13,8	67
PE (<i>DeepL</i>)	14	67,39

5.1.2. La diversité lexicale ou variation lexicale

Bonvin et Lambelet (2019 : 66) définissent la diversité lexicale comme

la mesure à la fois la plus connue et la plus ancienne de richesse lexicale. Elle se base sur le principe de (non-)répétition de vocables (type en anglais), l'idée sous-jacente étant qu'une répétition de mêmes mots (autrement dit, la présence de plusieurs occurrences du même vocable) est le signe de compétences lexicales basses (ou inversement, qu'une variation du choix des vocables reflète de bonnes compétences) (cf. Daller, 1999 : 121).

Par conséquent, dans cette étude, nous avons mesuré la diversité lexicale en estimant que plus ce rapport vocables/occurrences est élevé, moins il y a de redondance lexicale et *a fortiori*, plus la richesse lexicale est grande (Toral & Way

2018). Précisons qu'à l'instar de Toral (2019)⁵ entre autres, nous avons supposé une corrélation entre richesse lexicale et qualité linguistique. En d'autres termes, nous jugeons que, de manière générale, plus une traduction est variée et dense sur le plan lexical, plus elle peut être considérée de bonne qualité, tout en reconnaissant que, pour certains types de textes, il peut être utile de tendre davantage vers une uniformité lexicale.

L'indice de diversité lexicale a déjà été largement employé dans le domaine de la traduction. Il ressort de plusieurs travaux révélant l'existence d'un *translationese* que les textes traduits – quelles que soient les langues source et cible – présentent une perte de richesse lexicale par rapport aux textes en langue originale qui se manifeste, entre autres, par un vocabulaire moins varié (Castilho & Resende 2022). Et comme le confirment plusieurs études récentes (De Clercq *et al.* 2021 ; Hansen & Esperança-Rodier 2023 ; Martikainen 2019 ; Toral 2019), la TA entraîne une perte de richesse lexicale par rapport au *translationese*, lui-même moins riche que la langue originale.

Effectivement, Martikainen évoque en TA statistique un effet de lissage des solutions de traductions qui tend à éliminer la variation propre à la traduction humaine « en gommant les solutions minoritaires et en privilégiant les solutions de traduction majoritaires » (Martikainen 2019 : 206).

Après avoir entraîné trois systèmes de TA sur environ 20 millions de paires de phrases (anglais-français et anglais-espagnol) extraites du corpus parallèle *Europarl*, Vanmassenhove, Shterionov et Way (2019) ont démontré que, pour les deux combinaisons de langues, d'une part, aucun système de TA ne parvient à atteindre le degré de richesse lexicale mesuré en TH et, d'autre part, que les systèmes statistiques produisent des sorties plus riches sur le plan de la diversité lexicale que les systèmes neuronaux.

Et la même tendance persiste manifestement en PE. En effet, il ressort de l'étude comparative entre la PE de TA statistique (*Google*) et la TH menée par Čulo et Nitzke (2016) que, d'un côté, il y aurait moins de variation terminologique en TA qu'en TH et que, d'un autre côté, l'usage terminologique qui est fait en PE se rapprocherait fortement de l'usage terminologique en TA.

En 2018, Farrell s'est penché sur la comparaison de la TH et de la PE de TA (*Microsoft Translator*) d'articles Wikipédia pour la paire de langues anglais-italien. Ses observations indiquent une baisse de variation lexicale en PE comparativement à la TH. Dans ses conclusions, Farrell (2018) va jusqu'à déconseiller l'usage de la PE pour la traduction de textes ayant une composante créative :

On account of the findings reported herein, the use of PEMT for texts where variety, originality and inventiveness are quality factors would appear to be unadvisable with the MT technology currently available. (Farrell, 2018 : 58)

Outre une perte de diversité lexicale en TA comparativement à la TH, Toral (2019) a, lui aussi, constaté une perte de diversité en PE dans une moindre proportion : « In all cases the lexical variety in PE is lower than in HT, and again in

⁵ « [W]e have assumed that lexical diversity and density correlate directly with translation quality; *i.e.* the more diverse and dense a translation the better » (Toral, 2019 : 280).

all cases, that of MT is lower than that of PE » (Toral, 2019 : 276). Toujours selon lui, l'intervention humaine en PE permettrait certes d'ajouter de la diversité lexicale par rapport à la TA brute, mais pas d'aboutir à une production aussi variée qu'en TH en raison du haut degré d'interférence de la TA :

Our interpretation is that a post-editor improves the initial MT output in terms of [lexical] variety [...], but due to being primed by the MT output, the result cannot attain the level of HT, and the footprint of the MT system remains in the resulting PE. (*ibid.* : 280)

Enfin, les résultats obtenus par Volkart et Bouillon (2022) sont venus confirmer leur hypothèse de départ : les PE de TA (anglais-français) se sont révélées plus pauvres (moins variées) sur le plan lexical que les TH.

5.1.2.1. Hypothèse – diversité lexicale

En nous fondant sur les travaux résumés *supra*, nous sommes d'avis que le processus de PE ne permet pas de compenser cette perte de richesse lexicale en TA, du moins pas totalement. Par conséquent, nous nous attendons à ce que l'indice de diversité lexicale calculé sur nos corpus soit plus faible pour les textes post-édités que pour les textes traduits humainement. En d'autres termes, nous pensons observer une perte de richesse lexicale en PE par rapport à la TH. Enfin, en nous appuyant sur les observations de Vanmassenhove, Shterionov et Way (2019) qui concernent uniquement la TA, ainsi que sur celles de Toral (2019), nous postulons que la PE de TA neuronale se révèlera moins riche lexicalement que la PE de TA statistique et, en ce sens, plus proches des TS.

5.1.2.2. Méthode de calcul

La mesure la plus fréquemment utilisée pour mesurer la diversité lexicale (Figure 4) est le *Type-Token Ratio* (TTR) (Templin 1957) ; il s'agit de calculer le rapport entre le nombre de vocables (formes lemmatisées) d'un texte (*types*) et le nombre total d'occurrences (*token*).

Figure 4 : Calcul de diversité lexicale

$$TTR = \frac{\text{Nombre total}_{\text{vocables}}}{\text{Nombre total}_{\text{occurrences}}}$$

Toutefois, il a été démontré dans diverses études (Brezina 2018 ; Tezcan, Daems & Macken 2019) que le calcul de cette mesure est fortement influencé par la taille de l'échantillon, puisque comme l'expliquent Bonvin et Lambelet (2019 : 66),

[p]lus un texte est long, moins il est probable que de nouveaux mots surviennent et donc, plus les chances de répétitions de mots augmentent car le nombre d'occurrences augmente plus rapidement que le nombre de vocables.

Pour corriger ce biais, nous avons également eu recours à deux versions améliorées de cette mesure qui sont plus résistantes aux variations de taille de

l'échantillon : le *Mean Segmental Type-Token Ratio* (MSTTR)⁶ (Johnson 1944), ainsi que le *Moving Average Type-Token Ratio* (MATTR)⁷ (Covington & McFall 2010). Ainsi, pour comparer la diversité lexicale en fonction de chaque méthode de traduction, nous avons calculé les mesures TTR, MSTTR et MATTR en utilisant le module Python *LexicalRichness*⁸.

5.1.2.3. Résultats

Les résultats qui figurent dans les tableaux (3) et (4) viennent confirmer nos suppositions. Tout d'abord, nous constatons que, dans les deux tableaux, les mesures de diversité lexicale sont systématiquement inférieures en PE par rapport à la TH. Nous pouvons dire que dans nos deux expériences, ce sont les textes traduits humainement qui présentent le plus haut degré de diversité lexicale et que nous considérons donc comme les plus riches sur le plan lexical. Cette conclusion corrobore les résultats de précédentes études dont celles de Farrell (2018), de Martikainen (2023), de Toral (2019) et de Volkart et Bouillon (2022).

Tableau 3 : Diversité lexicale – Prétest

Prétest	Occurrences	Vocables	TTR	MSTTR	MATTR
TH	9778	954	0,10	0,38	0,38
PE de TAS	9565	728	0,08	0,36	0,36
PE de TAN (<i>DeepL</i>)	9712	658	0,07	0,36	0,36

Tableau 4 : Diversité lexicale – Test

Test	Occurrences	Vocables	TTR	MSTTR	MATTR
TH	9838	1045	0,11	0,39	0,39
PE (<i>Google</i>)	9405	800	0,09	0,37	0,37
PE (<i>DeepL</i>)	9603	742	0,08	0,37	0,37

⁶ Mean segmental type-token ratio : “In this process the text to be analysed is divided into equal segments in terms of the number of words (normally 100 words per segment). For each segment the TTR is calculated and using an arithmetic mean of the TTR for each segment the MSTTR is obtained” (Torruella & Capsada, 2013 : 449).

⁷ “MATTR is more informative than the mean segment TTR (MSTTR) introduced by Johnson (1944). [...] MSTTR is computed on successive non-overlapping segments of the text whereas MATTR uses a smoothly moving window. Thus MATTR yields a value for every point in the text except for those less than one window length from the beginning, while MSTTR is only a stepwise approximation to this. Thus MATTR is better for tracking changes within texts, and MATTR is not affected by accidental interactions between segment boundaries and text unit boundaries” (Covington & McFall, 2010 : 96).

⁸ Disponible en accès libre sur <https://github.com/LSYS/LexicalRichness>.

Ensuite, si ces résultats ne révèlent pas de différence marquée entre la PE de TA statistique et la PE de TA neuronale, ni entre la PE de *Google* et la PE de *DeepL*, nous pouvons néanmoins remarquer que, dans chaque tableau, le rapport vocables/occurrences (« TTR ») le plus faible est celui obtenu en PE de *DeepL*. D'après ces résultats, les textes post-édités à partir de la TA générée avec *DeepL* seraient moins riches lexicalement par rapport d'une part, aux textes traduits humainement et d'autre part, aux PE de *Google* et aux PE de TA statistique. Une fois de plus, cette observation rejoint la tendance mise en évidence par Toral (2019 : 276 : « In all cases the lexical variety of neural MT is lower than that of statistical MT. Again, the same trend shows when we look at their PEs ». Cela pourrait découler du fait que les post-éditeurs apportent davantage de modifications en TA statistique qu'en TA neuronale. De plus, ce constat est interpellant, d'autant plus vu l'usage de plus en plus répandu de l'outil *DeepL* par le grand public et l'intégration de la TA neuronale dans le flux de travail des traducteur-ices professionnel-les, avec notamment l'implémentation récente dans les outils de TAO les plus populaires sur le marché comme *SDL Trados Studio* ou *memoQ*.

Remarque

Les résultats obtenus pour ces deux indices de richesse lexicale peuvent paraître contradictoires à première vue. Pourtant, le fait que nous ayons conclu à une plus grande diversité lexicale en TH qu'en PE, alors que nous avons constaté que les PE étaient presque tout aussi denses lexicalement que les TH, prouve que l'indice de diversité lexicale est indépendant de celui de densité lexicale et qu'il était dès lors pertinent de ne pas se limiter au calcul de cette dernière. Il ressort de l'analyse linguistique que tout en étant de densité lexicale comparable aux TH, les textes post-édités s'avèrent moins variés lexicalement.

5.2. Longueur moyenne des phrases et coefficient de foisonnement

Nous avons cherché à savoir si la méthode de traduction avait eu de l'influence sur la longueur moyenne des phrases (*Sentence Length*) et sur le coefficient de foisonnement (*Expanding Ratio*) dans nos corpus.

Dans son étude portant sur l'existence d'un *post-editese*, Toral (2019) a constaté que, comparativement aux versions humaines, la longueur moyenne des phrases en PE se rapprochait de la longueur moyenne des phrases du TS. L'auteur attribue ce fait au haut degré d'interférence avec le TS en PE :

Sentence length in PEs is more similar to the sentence length of the source texts, than sentence length in HTs. We link this finding to interference [...]: (i) PEs have interference from the source text in terms of length, which leads to translations that follow the typical sentence length of the source language. (*ibid.* : 279)

Similairement, Castilho et Resende (2022) ont montré dans leur expérience sur la TA littéraire que pour l'un des deux extraits de romans étudiés, les versions traduites humainement présentaient une longueur moyenne de phrases plus élevée par rapport au TS, mais aussi par rapport à la TA et aux versions post-éditées.

Directement corrélé avec cet indice, le coefficient de foisonnement en traduction est défini par Durieux (1990 : 55) comme « l'augmentation de volume du texte d'arrivée par rapport au texte de départ ». Il ne s'agit pas du volume en nombre de caractères typographiques, mais bien en nombre de mots. Comme le soutient Cochrane (1995), le foisonnement n'est pas nécessairement négatif, au contraire :

Il arrive parfois que le traducteur fasse des choix qui allongent le texte. Néanmoins, il faut garder à l'esprit que le foisonnement n'est pas nécessairement mauvais. Il peut même être souhaitable dans certains cas pour clarifier des passages, donner des précisions indispensables ou rendre une phrase de manière plus idiomatique. (Cochrane, 1995 : 187)

Pendant, si ce foisonnement est excessif, cela peut nuire à la qualité et à la lisibilité des traductions (Delisle 2013). Comme le fait remarquer Cochrane (1995), le taux de foisonnement moyen en traduction est loin de faire l'unanimité dans la littérature. Ainsi, pour la TH de l'anglais vers le français, Durieux (1990) indique que la Société française des traducteurs (SFT) annonce un coefficient de foisonnement courant de 25 % et De Clercq *et al.* (2021) affirment que généralement, les agences de traduction s'attendent à un foisonnement qui varie entre 20 % et 25 %. Retenons plus largement que le coefficient de foisonnement moyen de l'anglais au français s'étend entre 10 % et 30 % (Cochrane 2000).

Cet indice ne met manifestement personne d'accord étant donné que l'étude du foisonnement en TA et en PE fait l'objet de résultats contradictoires dans la littérature dont nous donnons un aperçu *infra*.

Concernant le foisonnement en TA, De Clercq *et al.* (2021) ont effectué une étude comparative entre la langue française originale et la langue française traduite automatiquement de l'anglais à partir de *DeepL*, de *Google Traduction* et de l'outil de la Commission européenne – *eTranslation* – en version statistique et en version neuronale. Elles ont observé la même tendance que pour la TH, c'est-à-dire un coefficient de foisonnement en TA situé entre 20 % et 25 % par rapport au français original.

Martikainen et Kübler (2016) se sont livrées à l'étude comparative de TH et de PE de TA statistique (anglais-français) et ont conclu que la TH était légèrement plus prolifique que la PE, « même si les valeurs sont proches et avoisinent les 30 % de foisonnement dans les deux cas » (*ibid.* : 5). Elles ont également constaté un foisonnement entre la TA et la PE (3,97 %) qu'elles attribuent principalement aux modifications syntaxiques apportées en PE.

La même tendance se dégage des travaux de Castilho, Resende et Mitkov (2019). Elles ont voulu explorer l'existence d'un *post-editese* en comparant des textes traduits humainement, des TA brutes (*Google Traduction*) et des PE. Pour ce faire, elles ont fait appel aussi bien à des traducteur·ices professionnel·les qu'à des étudiant·es en traduction pour post-éditer (anglais-brésilien) des textes tirés de deux corpus : un corpus de presse (*New York Times*) et un corpus littéraire (*Opus corpus*). Elles ont examiné des mesures linguistiques telles que la diversité lexicale et la densité lexicale, ou encore la longueur moyenne des phrases, et ont pu confirmer l'existence d'un *post-editese* :

Post-edited features were found to be reflected as more interference from the original than HT texts and also more interference from the raw MT output. Our results show that the greater the human interference in the raw MT texts, the greater their distance from the original text and, consequently, their distance from the MT output. This is the case when the raw MT is fully post-edited [PF]. In this case, the PF version tends to be closer to HT, and further from MT, PL [PE légère] and original versions, suggesting a great similarity in terms of features between HT and PF. (*ibid.* : 25-26)

En matière de longueur moyenne des phrases et de foisonnement, elles ont abouti à la conclusion que les textes post-édités se rapprochent davantage des textes sources que les versions humaines. Elles concluent leur étude en posant cette question fondamentale : « Can we be sure that the greater the differences between the original and the translation (as the revealed by HT and PF versions), the higher the quality? » (*ibid.* : 26).

Tandis que les conclusions formulées par Volkart et Bouillon (2022), à nouveau pour la combinaison anglais-français, vont dans le sens contraire puisqu'elles ont mis au jour un coefficient de foisonnement nettement plus élevé en PE de TA neuronale (37,18 %) qu'en TH (30,77 %).

5.2.1. Hypothèse – longueur et foisonnement

De notre côté, en nous fondant sur les études de Castilho, Resende et Mitkov (2019), de Castilho et Resende (2022), de Martikainen et Kübler (2016) et de Toral (2019), et en ayant à l'esprit le haut degré d'interférence avec la langue source en PE (Depraetere 2010 ; Toral 2019), nous formulons l'hypothèse que les textes post-édités par les élèves présenteront une longueur moyenne de phrases et un coefficient de foisonnement moins élevés que les TH et donc plus proches des valeurs du TS.

5.2.2. Méthode de calcul

Dans nos expériences, pour chaque méthode de traduction, l'indice de longueur moyenne des phrases a été obtenu en divisant le nombre total d'occurrences par le nombre total de phrases (Figure (5)).

Figure 5 : Calcul de longueur moy. des phrases

$$\text{Longueur moy. des phrases} = \frac{\text{Nombre total}_{\text{occurrences}}}{\text{Nombre total}_{\text{phrases}}}$$

À nouveau, afin d'obtenir une comparaison pertinente des données, seules huit productions ont été retenues par méthode de traduction dans le prétest avec un total de 336 phrases, et sept productions par méthode dans le test avec un total de 294 phrases. C'est pourquoi nous n'obtenons pas deux valeurs identiques dans les tableaux (5) et (6) pour la longueur moyenne des phrases des TS alors qu'il s'agit bel et bien des mêmes TS dans les deux expériences.

Pour le calcul du coefficient de foisonnement, nous avons employé la formule suivante :

Figure 6 : Calcul du foisonnement

$$\text{Foisonnement} = \frac{\text{Longueur totale}_{\text{texte cible}} - \text{Longueur totale}_{\text{texte source}}}{\text{Longueur totale}_{\text{texte source}}} \times 100$$

Dans ce calcul, la longueur totale correspond au nombre total d'occurrences (*token*).

5.2.3. Résultats

Confirmant ce qui ressort de la littérature, les résultats repris dans ces deux tableaux ont indiqué une longueur moyenne de phrases nettement plus élevée en langue traduite par rapport au TS. Toutefois, et contrairement à ce que nous avons supposé, les valeurs obtenues en PE sont très semblables à celles obtenues en TH. Il apparaît donc que, pour la longueur moyenne des phrases dans nos expériences, les versions post-éditées se rapprochent davantage des TH que des TS. Ces observations vont dans le sens contraire des conclusions dégagées, entre autres, par Castilho et Resende (2022) et Toral (2019) ; peut-être est-ce parce que, contrairement aux échantillons de ces études, les nôtres se composent exclusivement de productions d'élèves.

Tableau 5 : Longueur des phrases et foisonnement – Prétest

Prétest	Longueur moy. des phrases	Coefficient de foisonnement
TS	24,81	–
TH	29,59	+19,27 %
PE de TAS	29,08	+17,20 %
PE de TAN (<i>DeepL</i>)	29,64	+19,48 %

Tableau 6 : Longueur des phrases et foisonnement – Test

Test	Longueur moy. des phrases	Coefficient de foisonnement
TS	24,74	–
TH	30,26	+22,32 %
PE (<i>Google</i>)	29,98	+21,17 %
PE (<i>DeepL</i>)	30,12	+21,75 %

En matière de foisonnement, les taux oscillent entre 17 % et 22 % ce qui correspond au foisonnement communément attendu pour la traduction de l'anglais vers le français. Pour la comparaison entre la PE de TA statistique et la TH dans le prétest, nos résultats montrent que les textes post-édités foisonnent moins que les TH, corroborant les observations de Martikainen et Kübler (2016). En revanche, il apparaît que les taux de foisonnement en PE de TA neuronale (prétest et test) sont comparables aux taux de foisonnement en TH et ce, contrairement à notre hypothèse et aux résultats précédemment obtenus par Castilho et Resende (2022), Toral (2019) et Volkart et Bouillon (2022). Par conséquent, les résultats figurant dans les tableaux 5 et 6 ne révèlent pas l'existence d'un *post-editese*.

Ajoutons finalement qu'il est intéressant de constater que les valeurs obtenues dans le test sont toutes légèrement plus élevées que les valeurs du prétest. Serait-ce dû à la différence d'expérience des participant-es ? Par rapport aux élèves de 3^e bachelier (prétest), les élèves de 2^e master (test) se sont-ils-elles davantage éloigné.es du TS et de la TA brute, auraient-ils-elles eu davantage recours à l'explicitation ?

5.3. Équivalence syntaxique

Nous avons également procédé à une analyse statistique de distance textuelle qui permet d'étudier la proximité syntaxique d'un texte cible par rapport à un texte source.

Trois indices ont été calculés automatiquement :

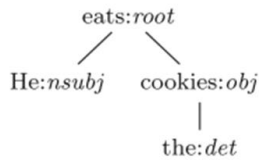
1) *Syntactically Aware Cross* (SACr) : cet indice permet de quantifier les réagencements syntaxiques entre la cible et la source ; « [it] measures the degree of word group reordering by creating syntactically motivated groups of words that are aligned » (Vanroy *et al.*, 2021 : 259) ;

2) *Part-of-Speech (POS) changes* : il s'agit de mesurer les changements en matière d'étiquetage morphosyntaxique⁹ entre les phrases sources et cibles ;

3) *Aligned Syntactic Tree Edit Distance* (ASTrED) : ce troisième indice mesure des divergences structurelles plus profondes en comparant les arbres de dépendance (*dependency tree*) source et cible : « [It] compares the deep linguistic structure of the source and target sentences while taking word alignment into account » (*ibid.* : 291). Vanroy *et al.* (2021) fournissent un exemple d'arbre de dépendance pour la phrase « He eats the cookies » :

Figure 7 : Exemple d'arbre de dépendance (Vanroy *et al.*, 2021 : 264)

⁹ Étiquetage morphosyntaxique : « En traitement du langage naturel, processus qui consiste à identifier la catégorie syntaxique ainsi que des informations morphosyntaxiques (comme la partie du discours, le genre, le nombre, etc.) associées aux occurrences des mots d'un texte ou d'une phrase dans leur contexte d'énonciation, et ce, à l'aide d'un outil informatique » (GLFIA).



Les résultats auxquels ont abouti Hansen et Esperança-Rodier (2023) après le calcul de l'équivalence syntaxique démontrent la même tendance pour les trois indices (SACr, POS changes et ASTrED) : la TH arrive nettement en tête de classement par rapport aux différents systèmes de TA évalués (*Google Traduction* ; *DeepL* et leur système neuronal « maison »), ce qui indique que les TA sont plus proches de la source que les TH.

Concernant la PE, Toral (2019) a montré dans son étude comparative que l'étiquetage morphosyntaxique des textes post-édités est plus proche de l'étiquetage morphosyntaxique du TS comparativement à l'étiquetage en TH.

5.3.1. Hypothèse – Équivalence syntaxique

Nous nous attendons à ce que, dans nos deux expériences, les textes post-édités soient plus proches syntaxiquement des TS par rapport aux TH, et ce, quel que soit le TS.

5.3.2. Méthode de calcul

Cette proposition de mesure d'équivalence syntaxique (Vanroy *et al.* 2021) a été calculée grâce au module Python ASTrED¹⁰ (*Aligned Syntactic Tree Edit Distance*). Ce module utilise la librairie Stanza Python NLP (Qi *et al.* 2020) développée par le Stanford NLP Group et fréquemment utilisée en TAL. En outre, cet outil d'analyse syntaxique automatique figure en première place des résultats obtenus pour le français dans la campagne d'évaluation de la *Conference on Natural Language Learning* (CoNLL) 2017 (Zeman *et al.* 2017).

5.3.3. Résultats

Dans les tableaux (7) et (8)¹¹ pour l'indice SACr, nous constatons un nombre nettement plus élevé de croisements syntaxiques en TH comparativement aux PE. De plus, l'analyse de la PE de TA neuronale (*DeepL*) révèle davantage de réagencements par rapport à la PE de TA statistique (prétest) et par rapport à la PE de *Google* (test).

Pour l'indice *POS changes*, nous observons cette même tendance quoique de manière moins prononcée ; c'est dans les TH que l'on constate le plus de changements d'étiquettes morphosyntaxiques. Par conséquent, les PE sont, une fois

¹⁰ Disponible en accès libre sur <https://github.com/BramVanroy/astred>

¹¹ Les tableaux complets de résultats figurent en annexe.

de plus, plus proches de la structure des textes sources que les TH. Ces résultats confirment notre hypothèse et corroborent à nouveau les résultats des travaux de Toral (2019).

Enfin, l'indice ASTrED confirme cette tendance : ce sont à nouveau les TH qui se détachent nettement. La comparaison des arbres de dépendance sources et cibles a permis de montrer de plus grandes divergences structurelles en TH qu'en PE ; les PE seraient ainsi plus proches syntaxiquement des textes sources que les TH. Les PE de *DeepL* sont à nouveau plus éloignées de la source que les PE de TA statistique (prétest) et que les PE de *Google* (test).

Tableau 7 : Équivalence syntaxique – Prétest

Prétest 2018	SACr	POS changes	ASTrED
TH	1651	1238	4149
PE de TAS	934	1054	3714
PE de TAN (<i>DeepL</i>)	1219	1073	3884

Tableau 8 : Équivalence syntaxique – Test

Test 2021	SACr	POS changes	ASTrED
TH	1397	947	3252
PE (<i>Google</i>)	998	848	2904
PE (<i>DeepL</i>)	1133	825	3027

6. SYNTHÈSE ET CONCLUSION

À l'instar de Toral (2019) et de Volkart et Bouillon (2022), mais contrairement à Daems, De Clercq et Macken (2017), nous pouvons dire qu'une partie de nos résultats confirme l'existence d'un *post-editese*. En effet, les comparaisons en matière de diversité lexicale et d'équivalence syntaxique ont mis en évidence des caractéristiques propres à une langue de PE qui se distingue du *translationese* : les textes post-édités (tout particulièrement en PE de *DeepL*) sont apparus moins riches lexicalement que les TH et plus proches syntaxiquement de la langue source que les TH.

En revanche, et contrairement à notre hypothèse, les indices de densité lexicale, de longueur moyenne des phrases et de foisonnement, n'ont pas révélé de différences notables entre la PE et la TH. Le profil des participant·es aux expériences a pu jouer un rôle. Le fait que les textes post-édités soient presque aussi denses et foisonnants que les TH pourrait être attribué à la participation d'apprenant·es en traduction plutôt que de traducteur·ices professionnel·les, telle est en tout cas notre principale hypothèse explicative.

Nous estimons que ces observations étayent l'influence de la TA neuronale sur la qualité linguistique du texte cible malgré l'intervention d'un·e post-éditeur·ice, faisant penser à de possibles « effets fantômes » de la TA en PE. Selon nous, ces

marques caractéristiques d'un *post-editese* sont en réalité dues à un double phénomène d'interférence : d'une part, à l'interférence indirecte avec la langue source en PE puisque la production passe par le filtre de la TA qui présente un haut degré avéré d'interférence avec le TS, d'autre part, à l'interférence avec la TA : « post-editing effort (and, potentially, its quality) is influenced by machine translation output, which, in turn is influenced by source text characteristics » (Daems, 2016 : 25).

Pour conclure, en dépit des progrès algorithmiques bluffants réalisés en TA, les tendances observées dans cette étude en matière de diversité lexicale et d'équivalence syntaxique confirment la nécessité de renforcer la formation des futur·es traducteur·ices en y intégrant la pratique de la PE. En effet, nous sommes persuadée qu'une fois sensibilisé·es, les étudiant·es seraient davantage en mesure de repérer ces effets d'interférence et de lisser au besoin les marques laissées par la TA, améliorant ainsi la qualité des textes post-édités.

Nous remercions le D^r Damien Hansen pour le partage de connaissances en TAL et le précieux coup de main dans le calcul des mesures linguistiques automatiques.

ANNEXES

Syntactic Equivalence (2021):

Number of SACr crosses in TH:	1397
POS changes in TH:	947
Edit distance (ASTrED) in TH:	3252
# sentences:	200

Number of SACr crosses in DL:	1133
POS changes in DL:	825
Edit distance (ASTrED) in DL:	3027
# sentences:	200

Number of SACr crosses in GT:	998
POS changes in GT:	848
Edit distance (ASTrED) in GT:	2904
# sentences:	200

Syntactic Equivalence (2018):

Number of SACr crosses in TH:	1651
POS changes in TH:	1238
Edit distance (ASTrED) in TH:	4149
# sentences:	259

Number of SACr crosses in TAN:	1219
POS changes in TAN:	1073
Edit distance (ASTrED) in TAN:	3884
# sentences: 259	

Number of SACr crosses in TAS:	934
POS changes in TAS:	1054
Edit distance (ASTrED) in TAS:	3714
# sentences: 259	

Références

Corpus – Dictionnaires

- [GLFIA] *Grand Lexique français de l'intelligence artificielle* (2021), DataFranca Wiki. [<https://datafranca.org/wiki/>]
- [ANTIDOTE] *Antidote (version 10) : Le remède à tous vos mots* (2022), Druide informatique Inc.

Livres – Articles – Communication

- BONVIN A. & LAMBELET A. (2019), « Exploration empirique de la richesse lexicale : la perception humaine », *Linguistik Online* 100 (7), 65-94. [en ligne]
- BREZINA V. (2018), *Statistics in corpus linguistic: A practical guide*, Cambridge, Cambridge University Press.
- BRUNET, E. (2015), « La répétition dans la phrase. Etude statistique », *Semen* 38. [en ligne]
- CASTILHO S., RESENDE N.C.A. & MITKOV R. (2019), “What Influences Post-editing features? A preliminary study”, in *Proceedings of the Second Workshop on Human-Informed Translation and Interpreting Technology*, Varna. [en ligne]
- CASTILHO S. & RESENDE N. (2022), “Post-Editing in Literary Translations”, *Information* 13 (2), 1-22.
- COCHRANE G. (1995), « Le foisonnement, phénomène complexe », *TTR* 8 (2), 175-193. [en ligne]
- COCHRANE G. (2000), *Le foisonnement dans les textes de spécialité, illusion d'optique ou réalité quantifiable ?* [Thèse de doctorat], Québec, Université Laval.
- COVINGTON MICHAEL, A. & MCFALL JOE, D. (2010), “Cutting the Gordian Knot: The Moving-Average Type-Token Ratio (MATTR)”, *Journal of Quantitative Linguistics* 17 (2), 94-100.
- ČULO O. & NITZKE J. (2016), “Patterns of terminological variation in post-editing and of cognate use in machine translation in contrast to human translation”, *Baltic Journal of Modern Computing* 4 (2), 106-114.
- 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* [Thèse de doctorat en philosophie et lettres], Gand, Université de Gand.
- DAEMS J., DE CLERCQ O. & MACKEN L. (2017), “Translationese and Post-editing: How comparable is comparable quality?” *Linguistica Antverpiensia, New Series: Themes in Translation Studies* 16, 89-103. [en ligne]
- DE CLERCQ O., DE SUTTER G., LOOCK R., CAPPELLE B. & PLEVOETS K. (2021), “Uncovering Machine Translationese Using Corpus Analysis Techniques to Distinguish between Original and Machine-Translated French”, *Translation Quarterly* 101, 21-45. [en ligne]
- DELISLE J. (2013), *La traduction raisonnée. Manuel d'initiation à la traduction professionnelle de l'anglais vers le français*, 3^e éd., Ottawa, Presses de l'Université d'Ottawa.
- DENEUFBOURG G. (2021), « Traduction automatique : la dangereuse « sagesse des foules » », *The Conversation*. [en ligne]
- DEPRAETERE I. (2010), “What counts as useful advice in a university postediting training context? Report on a case study”, in *Proceedings of the 14th Annual EAMT Conference*, Saint-Raphaël, EAMT. [en ligne]

- DURIEUX C. (1990), « Le foisonnement en traduction technique d'anglais en français », *Meta* 35 (1), 55-60. [en ligne]
- FARRELL M. (2018), "Machine Translation Markers in Post Edited Machine Translation Output", in *Proceedings of the 40th Conference Translating and the Computer*, Asling, 50-59.
- GELLERSTAM M. (1986), "Translationese in Swedish novels translated from English", in L. Wollin et H. Lindquist (eds), *Translation Studies in Scandinavia* vol. 4, CWK Gleerup, 88-95
- HALLIDAY M. A. K. (1985), *Spoken and written language*, Oxford, Oxford University Press.
- HANSEN D. & ESPERANÇA-RODIER E. (2023), "Human-Adapted MT for Literary Texts: Reality or Fantasy?", in S. Castilho, R. Caro Quintana, M. Stasimioti & V. Sosoni (eds), *Proceedings of the International Conference New Trends in Translation and Technology*, NeTTT 2022, Incoma Ltd, 178-190.
- JOHNSON W. (1944), "Studies in language behavior: A Program of Research", *Psychological Monographs* 56 (2), 1-15. [en ligne]
- LU X. (2012), "The Relationship of Lexical Richness to the Quality of ESL Learners' Oral Narratives", *The Modern Language Journal* 96, 190-208.
- MARTIKAINEN H. (2019), *Sources de distorsion dans les résumés traduits de revues systématiques : Une comparaison de la traduction humaine et de la traduction automatique post-éditée* [Thèse de doctorat en Linguistique], Paris, Université Paris-Diderot.
- MARTIKAINEN, H. (2022). "Ghosts in the machine: Can adaptive MT help reclaim a place for the human in the loop?" *Journal of Data Mining and Digital Humanities. Vers une robotique du traduire* [Hors-série]. <https://hal.science/hal-03548696/document>
- MARTIKAINEN H. (2023), "Investigating the Usability of Automatic Metrics for Characterizing Translated vs Post-edited Texts in the Post-editing Classroom to Further Students' MT Literacy", in N. Froeliger, C. Larssonneur et G. Sofo (eds), *Traduction humaine et traitement automatique des langues. Vers un nouveau consensus ?*, Venice University Press, Fondazione Università Ca' Foscari, 67-79..
- MARTIKAINEN H. & KÜBLER N. (2016), « Ergonomie cognitive de la post-édition de traduction automatique : enjeux pour la qualité des traductions », *ILCEA* 27. [en ligne]
- MULLER CH. (1969), « La statistique lexicale », *Langue française* 2, 30-43. [en ligne]
- O'DONNELL M. (2016), "The UAM CorpusTool 3.3" [Logiciel]. <http://corpustool.com>
- QI P., ZHANG Y., ZHANG Y., BOLTON J. & MANNING C. D. (2020), "Stanza: A Python natural language processing toolkit for many human languages", in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Association for Computational Linguistics, 101-108. [en ligne]
- SCARPA F. (2006), "Corpus-based quality assessment of specialist translation: A study using parallel and comparable corpora in English and Italian", in M. Gotti & S. Šarčević (eds), *Insights into specialized translation-linguistics insights*, Berne, Peter Lang, 155-172.
- SCHUMACHER P. (2019), « Avantages et limites de la post-édition », *Traduire, hors cahier* (241), 108-123.
- SCHUMACHER P. (2023a), *La post-édition de traduction automatique en contexte d'apprentissage. Effets sur la qualité et défis pour l'enseignement de la traduction* [Thèse de doctorat en Traductologie], Liège, Université de Liège. [en ligne]
- SCHUMACHER P. (2023b), « Traduction humaine et postédition : contrôle qualité en contexte académique », *Meta: Journal des Traducteurs* 68 (3), 510-536.

- TEMPLIN M. C. (1957), *Certain language skills in children; their development and interrelationships*, Minneapolis, University of Minnesota Press.
- TEZCAN A., DAEMS J. & MACKEN L. (2019), “When a ‘Sport’ Is a Person and Other Issues for NMT of Novels”, in *Proceedings of the Qualities of Literary Machine Translation*, Dublin, EAMT, 40-49.
- TO V., FAN S. & DAMON T. (2013), “Lexical Density and Readability: A Case Study of English Textbooks”, *The International Journal of Language, Society and Culture* 37, 61-71.
- TORAL A. (2019), “Post-editeese: an Exacerbated Translationese”, in *Proceedings of Machine Translation Summit XVII: Research Track*, Dublin, EAMT, 273-281.
- TORAL A. & WAY A. (2015), “Machine-assisted translation of literary text: A case study”, *Translation Spaces* 4, 240-267. [en ligne]
- TORRUELLA J. & CAPSADA R. (2013), “Lexical Statistics and Tipological Structures: A Measure of Lexical Richness”, *Procedia - Social and Behavioral Sciences* 95, 447-454. [en ligne]
- URE J. (1971), “Lexical Density and Register Differentiation”, in G. Perren & J. Trim (eds), *Applications of Linguistics*, Cambridge, Cambridge University Press, 443-452.
- VANMASSENHOVE E., SHTERIONOV D. & WAY A. (2019), “Lost in Translation: Loss and Decay of Linguistic Richness in Machine Translation”, in *Proceedings of Machine Translation Summit XVII: Research Track*, Dublin, EAMT, 222-232.
- VANMASSENHOVE E., SHTERIONOV D. & GWILLIAM M. (2021), “Machine Translationese: Effects of Algorithmic Bias on Linguistic Complexity in Machine Translation”, in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, Association for Computational Linguistics, 2203-2213.
- VANROY B., DE CLERCQ O., TEZCAN A., DAEMS J. & MACKEN L. (2021), “Metrics of Syntactic Equivalence to Assess Translation Difficulty”, in M. Carl (ed.), *Explorations in Empirical Translation Process Research, Machine Translation: Technologies and Applications*, Springer, 259-294. [en ligne]
- VOLKART L. & BOUILLON P. (2022), “Studying Post-Editese in a Professional Context: A Pilot Study”, in *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation*, Gand, EAMT, 71-79.
- ZEMAN D., POPEL M., STRAKA M., HAJIČ J., NIVRE J., GINTER F., LUOTOLAHTI J., PYYSALO S., PETROV S., POTTHAST M., TYERS F., BADMAEVA E., GÖKIRMAK M., NEDOLUZHKO A., CINKOVÁ S., HAJIČ JR. J., HLAVÁČOVÁ J., KETTNEROVÁ V., UREŠOVÁ ... LI, J. (2017), “CoNLL 2017 shared task: Multilingual parsing from raw text to Universal Dependencies”, in *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, Association for Computational Linguistics, 1-19. [en ligne]