

Cotutelle de thèse internationale entre
l'ULiège et l'UNIGE

La post-édition de traduction automatique en contexte d'apprentissage

Effets sur la qualité et défis pour l'enseignement de la traduction

Thèse présentée en vue de l'obtention du titre
de Docteur en Langues, Lettres et Traductologie à l'Université de Liège
et de Docteur en Traitement Informatique Multilingue à l'Université de Genève
par

Perrine Schumacher

Jury

D^{re} Valérie Bada, FACPHE, Université de Liège (Directrice de thèse)
Prof. Pierrette Bouillon, FTI, Université de Genève (Directrice de thèse)
Prof. Dominique Longrée, FACPHE, Université de Liège (Président du Jury)
D^{re} Hanna Martikainen, ESIT, Université Sorbonne Nouvelle (Jurée externe)
Prof. Mathilde Fontanet, FTI, Université de Genève (Jurée)
D^r Thierry Fontenelle, Banque européenne d'investissement - BEI (Juré externe)

Présentée et soutenue à l'Université de Liège le 30 août 2023

Depuis plusieurs années, le secteur de la traduction connaît un véritable bouleversement technologique induit en grande partie par l'avènement de nouvelles technologies de la traduction, telles que la traduction automatique (TA) neuronale. Par conséquent, les pratiques professionnelles sont amenées à évoluer et l'actualisation des programmes de formation en traduction se révèle un enjeu crucial à la fois pour l'enseignement de la traduction et pour l'avenir même de la profession. Dans ce contexte, nous consacrons cette thèse à l'étude d'une pratique en plein essor dans le secteur des services langagiers : la post-édition (PE) de TA. Premièrement, l'objectif majeur de ce travail est d'explorer les effets de cette pratique, en contexte universitaire, sur la qualité finale de productions. Deuxièmement, nous cherchons à évaluer les connaissances et les perceptions des étudiants en traduction à l'égard de la TA et de la PE. Troisièmement, notre volonté est enfin de contribuer à la réflexion sur l'intégration d'une formation en TA et en PE dans les cursus des apprenants en traduction. Pour mener à bien ces objectifs, nous avons mené deux expériences contrôlées impliquant des étudiants en traduction afin de comparer les produits de deux processus de traduction (anglais-français) : la traduction humaine (TH) et la post-édition de TA, tant statistique que neuronale. Pour l'exploitation des données empiriques recueillies, nous combinons une approche qualitative (évaluation humaine de la qualité et analyse linguistique des erreurs contenues dans les productions) à une approche quantitative (statistique descriptive, statistique inférentielle et mesures linguistiques automatiques). Nos résultats montrent que la PE de TA, qu'elle soit statistique ou neuronale, permet d'obtenir des productions de qualité globale comparable à la qualité en TH, voire de qualité supérieure pour la PE de TA neuronale. Par ailleurs, l'analyse statistique détaillée indique principalement que les textes post-édités contiennent davantage de calques par rapport aux TH et que les PE de TA neuronale contiennent moins d'erreurs de grammaire et de syntaxe comparativement aux TH. De plus, nous avons fait le constat que la qualité d'une PE dépend du paradigme de TA (statistique ou neuronal), ainsi que du moteur de TA neuronale (*Google Traduction* ou *DeepL*). Nos travaux révèlent également l'existence d'un effet nivelant en PE de TA neuronale sur la qualité des textes cibles qui atteste du fait que plus l'étudiant a de faibles compétences en TH, plus la PE lui sera bénéfique et à l'inverse, plus il a des compétences élevées en TH, plus la PE entraîne une baisse de qualité de sa production finale. De plus, une partie de nos résultats témoigne de la présence de caractéristiques propres à une langue de PE (*post-editese*) qui la distinguent de la langue traduite humainement. Dans la dernière partie de cette thèse, nous étayons notre position en faveur de l'adoption d'une approche raisonnée des outils de TA afin de former des utilisateurs avertis, autonomes et responsables. Enfin, en nous appuyant sur ces résultats et sur notre expérience pédagogique, nous mettons en lumière les principaux enjeux et atouts de la PE et définissons sept grands défis sur lesquels il convient de mettre l'accent dans l'apprentissage de la PE.

Mots-clés : post-édition, traduction automatique neuronale, traduction humaine, expérience contrôlée, qualité en traduction, formation en post-édition, enseignement de la traduction, *post-editese*, effet nivelant en TA neuronale, étude comparative.

For the past few years, the translation industry has been shaped by a technological boom, largely driven by the advent of new translation technologies, such as neural machine translation (NMT). Hence, professional practices are changing, and adapting translation training programs has become a critical challenge both for translation training and for the future of the profession. In such context, this thesis is devoted to the study of an increasingly common practice in the language services industry: machine translation post-editing (MTPE). Firstly, the main objective of our research is to investigate the effects of MTPE in an academic context on final text quality. Secondly, we seek to assess translation students' knowledge and perceptions regarding MT and PE. Thirdly, we aim to contribute to the debate on the introduction of MT and PE training courses in translation curricula. To meet these goals, we conducted two controlled experiments involving translation students to compare the outputs of two translation processes (English into French): human translation (HT) and post-editing using both statistical and neural MT. To analyse the empirical data gathered, we combined a qualitative approach (human quality assessment and linguistic analysis of errors contained in students' products) with a quantitative approach (descriptive statistics, inferential statistics, and automatic linguistic analysis). Our findings show that PE of MT, whether statistical or neural, leads to products of an overall quality comparable to human-translated texts, or even of higher quality in the case of neural MTPE. Moreover, the fine statistical analysis mainly indicates that post-edited texts contain more calques than human-translated texts, and that NMT post-edited language contains fewer grammar and syntax errors compared to human-translated language. In addition, we note that PE quality depends on MT paradigm (statistical or neural), as well as on the neural MT engine used (*Google Translate* or *DeepL*). Our work also uncovers a leveling effect in neural MTPE on the quality of target texts, which demonstrates that the poorer the student's translation skills, the more they will benefit from PE, and conversely, the higher their translation skills, the poorer the quality of their post-edited product. Further, part of our results suggests the existence of typical features that set post-edited language apart from human-translated language (i.e. Post-Editese). In the last part of this thesis, we substantiate our position in favour of MT Literacy training to educate informed, autonomous, and responsible users. Finally, based on our results and our pedagogical experience, we highlight the main issues and opportunities of MTPE and outline seven major challenges that need to be addressed in PE training.

Keywords: post-editing, neural machine translation, human translation, controlled experiment, translation quality, post-editing training, translation training, *post-editese*, NMT leveling effect, comparative study.

Je voudrais tout d'abord adresser un immense merci à mes deux directrices de thèse, D^{re} Valérie Bada et Prof. Pierrette Bouillon, pour avoir accepté de m'accompagner dans cette entreprise audacieuse. Je les remercie infiniment pour leur confiance, leurs précieux conseils et leur soutien infailible.

Ensuite, je tiens à remercier grandement les membres de mon comité de thèse, pour leur disponibilité et leurs commentaires enrichissants. Merci à Thierry Fontenelle pour sa bienveillance et pour ses réflexions stimulantes, ainsi qu'au Prof. Dominique Longrée d'avoir accepté d'être président de mon jury de soutenance et de m'avoir prodigué de précieux conseils.

Merci évidemment aux étudiantes de 3^e bachelier (2017-2018) et de 2^e master (2020-2021) de la filière traduction et interprétation de l'Université de Liège d'avoir bien voulu se plier au jeu en participant à cette étude expérimentale. Je tiens à remercier par avance les *lectrices*, qu'elles soient membres du jury, étudiantes, enseignantes, scientifiques, collègues ou simples curieuses pour leur intérêt dans ce travail. Je suis également reconnaissante envers Vincent François dont les recherches en intelligence artificielle et dont l'humour narquois ont constitué le point de départ de cette thèse.

J'adresse un tout grand merci à mon collègue, Damien Hansen (CIRTI – ULiège), avec qui les échanges ont permis de guider mes réflexions et mes choix, mais surtout de préserver ma motivation et ma santé mentale ! Ses connaissances en linguistique, en informatique et sa curiosité scientifique ont été d'une aide précieuse.

Par ailleurs, je voudrais remercier particulièrement le D^r Antonio Sutura (Département d'électricité, électronique et informatique de l'ULiège) pour sa patience et sa générosité scientifique sans bornes. J'adresse aussi mes remerciements au D^r Marc-Olivier Boldi (Research Center for Statistics de l'Université de Genève) qui m'a aidée à déterminer le test statistique approprié afin d'exploiter valablement les résultats de cette recherche empirique. Merci également à Michaël Koutsoudakis pour sa serviabilité et son assistance dans la gestion de certains aspects techniques.

Un énorme merci à toutes les personnes qui ont égayé mon quotidien et nourri mon parcours tant personnel que doctoral : qu'il s'agisse de mes anciennes enseignantes, devenues mes chères et tendres collègues, Bénédicte, Céline, Laurence, Mathilde, Mélanie, Pierre G., Valérie M. ; d'autres membres de cette incroyable filière TI où il fait si bon travailler, Clémence, Magali, Maud, Pierre R., François, Julien, Raphaël, Patricia, Myriam, Muriel ; ou simplement de belles rencontres, Christina, Valérie-Anne, Marianne, Jonathan, François Tine, Prof. Germain Simons, Loïc, Sabrina.

Je souhaite également remercier tendrement mes amies et mes proches, tout particulièrement mes grands-parents bruxellois chéris, ainsi que ma maman, pour leur relecture complète et minutieuse, et pour leur soutien inconditionnel. Merci enfin à Charly d'avoir enduré mes sautes d'humeur, mes peines et mes joies de ces dernières années et de m'avoir amoureusement épaulée dans cette fin d'aventure doctorale.

Tables des matières

LISTE DES FIGURES.....	XIV
LISTE DES TABLEAUX.....	XVI
LISTE DES ABRÉVIATIONS.....	XIX
1 INTRODUCTION	1
1.1 CONTEXTE ET MOTIVATION	1
1.1.1 <i>L'essor de la traduction automatique</i>	1
1.1.2 <i>Une pression accrue sur le secteur de la traduction</i>	3
1.2 OBJECTIFS ET MÉTHODOLOGIE	5
1.2.1 <i>Questions et sous-questions de recherche</i>	6
1.3 PLAN DE THÈSE	8
1.4 PUBLICATIONS SCIENTIFIQUES.....	9
2 ÉTAT DE L'ART	11
2.1 LA TRADUCTION AUTOMATIQUE	12
2.1.1 <i>Survol historique</i>	12
2.1.1.1 La TA à base de règles	13
2.1.1.2 La révolution statistique.....	14
2.1.1.3 Le tournant neuronal	15
2.1.2 <i>La TA neuronale : principe de fonctionnement</i>	16
2.1.2.1 Un prolongement de la TA statistique (De Faria Pires, 2020)	16
2.1.2.2 Une technologie neuronale d'inspiration biologique ?.....	18
2.1.2.3 Le plongement lexical ou <i>word embedding</i>	18
2.1.2.4 Les architectures neuronales.....	19
2.1.2.4.1 Apprentissage profond.....	19
2.1.2.4.2 Architecture encodeur-décodeur.....	20
2.1.2.4.3 Les réseaux de neurones récurrents (<i>recurrent neural networks</i> – RNN)	21
2.1.2.4.4 Les réseaux de neurones <i>Transformers</i>	23
2.1.3 <i>La TA neuronale : principaux écueils</i>	24
2.2 LA POST-ÉDITION.....	29
2.2.1 <i>Historique et définition</i>	29
2.2.2 <i>Post-édition et traduction</i>	30
2.2.3 <i>Post-édition et révision</i>	31
2.2.4 <i>La qualité en post-édition</i>	33
2.2.4.1 Niveaux de post-édition	33
2.2.5 <i>Les compétences en post-édition</i>	38
2.2.6 <i>La recherche en post-édition</i>	40
2.2.6.1 Étude du processus.....	40
2.2.6.1.1 Méthodes de mesure de l'effort de PE.....	42
2.2.6.2 D'abord la source ou d'abord la cible ?	45
2.2.6.3 Étude du produit.....	46
2.2.6.3.1 Qualité comparable, voire meilleure.....	47
2.2.6.3.2 Études sur la TA neuronale.....	50
2.2.6.3.3 Effets négatifs sur la qualité	51
2.3 ÉVALUATION DE LA QUALITÉ EN TRADUCTION.....	54
2.3.1 <i>La notion de qualité en traduction</i>	54
2.3.2 <i>Tentatives de définition</i>	56
2.3.3 <i>Une subjectivité intrinsèque</i>	58
2.3.4 <i>L'évaluation orientée produit et l'évaluation orientée processus</i>	61
2.3.5 <i>Approche typologique</i>	62

2.3.6	<i>Évaluation de la qualité dans l'enseignement de la traduction</i>	64
2.3.6.1	Quel est le rôle de l'évaluation dans l'enseignement ?	65
2.3.6.2	Approche typologique en didactique de la traduction	66
2.3.6.3	Distinction entre erreurs de langue et erreurs de traduction	67
2.3.6.4	L'évaluation en contexte d'apprentissage vs en contexte professionnel	68
2.3.7	<i>Évaluation de la TA</i>	69
2.3.7.1	Évaluation humaine de la qualité en TA	71
2.3.7.1.1	Approches classiques en TA.....	71
2.3.7.1.2	Approche typologique en TA	72
2.3.7.1.3	Discussion.....	74
2.3.7.2	Automatisation de l'évaluation en TA	74
2.3.7.2.1	Métriques d'évaluation automatique	74
2.3.7.2.2	Méthodes de classification automatique	81
2.3.7.3	Stratégies d'évaluation axées sur la tâche.....	82
2.3.8	<i>Évaluation des nouveaux outils</i>	83
3	MÉTHODOLOGIE	85
3.1	MÉTHODE EXPÉRIMENTALE	86
3.1.1	<i>Variables</i>	86
3.1.1.1	Variable dépendante (V.D.)	86
3.1.1.2	Variables indépendantes (V.I.).....	86
3.1.1.3	Variables parasites (V.P.).....	87
3.1.2	<i>Plan expérimental</i>	87
3.1.3	<i>Participant</i> es	88
3.1.3.1	Participantes au prétest	88
3.1.3.2	Participantes au test	89
3.1.3.3	Échantillonnage	89
3.1.3.4	Consentement libre et éclairé	90
3.1.4	<i>Tâches</i>	90
3.1.4.1	Moteurs de TA.....	91
3.1.4.1.1	Prétest.....	91
3.1.4.1.2	Test	92
3.1.4.2	Qualité attendue en post-édition.....	93
3.1.5	<i>Corpus</i>	94
3.1.5.1	Sélection des textes sources.....	94
3.1.5.2	Score BLEU	96
3.1.6	<i>Répartition des tâches</i>	97
3.1.7	<i>MateCat</i>	98
3.2	DÉROULEMENT DES EXPÉRIENCES	99
3.2.1	<i>Prétest</i>	100
3.2.2	<i>Test</i>	100
3.2.3	<i>Exclusion de données</i>	101
3.3	EXPLOITATION DES DONNÉES.....	101
3.3.1	<i>Combinaison d'approches qualitative et quantitative</i>	101
3.3.2	<i>Approche qualitative : évaluation humaine de la qualité</i>	102
3.3.2.1	Typologie d'erreurs.....	103
3.3.2.1.1	Phase d'acceptabilité	104
3.3.2.1.2	Phase d'adéquation	105
3.3.2.1.3	Perméabilité entre acceptabilité et adéquation	105
3.3.2.2	Limitation de l'arbitraire	106
3.3.2.3	Phase de consolidation	107
3.3.2.4	Notation	107
3.3.2.5	Mesures d'accord inter-évaluateur.....	108
3.3.2.5.1	Accord observé (%)	109
3.3.2.5.2	Coefficient kappa de Cohen.....	111
3.3.2.5.3	Coefficient de corrélation intra-classe.....	111

3.3.3	<i>Approche quantitative : étude statistique</i>	117
3.3.3.1	Statistique descriptive.....	118
3.3.3.2	Statistique inférentielle.....	118
3.3.4	<i>Analyse complémentaire</i>	120
3.3.4.1	Étude du processus.....	120
3.4	CONCLUSION.....	120
4	RECHERCHE EXPÉRIMENTALE – PRÉTEST	121
4.1	INTRODUCTION.....	121
4.2	UNE ÉTAPE MÉTHODOLOGIQUE.....	121
4.3	HYPOTHÈSES DE RECHERCHE	122
4.3.1	<i>Hypothèse HA1</i>	122
4.3.2	<i>Hypothèse HA2</i>	122
4.3.2.1	Sous-hypothèse SHA2.1.	123
4.3.2.2	Sous-hypothèse SHA2.2.	123
4.3.2.3	Sous-hypothèse SHA2.3.	124
4.3.2.4	Sous-hypothèse SHA2.4.	124
4.3.2.5	Sous-hypothèse SHA2.5.	125
4.3.3	<i>Hypothèse HA3</i>	125
4.4	RÉSULTATS	126
4.4.1	<i>Statistiques descriptives</i>	126
4.4.1.1	Distribution des erreurs (diagramme circulaire).....	126
4.4.1.2	Distribution des erreurs (histogramme).....	129
4.4.1.3	Comparaison de la qualité par mode de traduction.....	130
4.4.1.3.1	Moyenne des notes et taux de réussite	130
4.4.1.4	Moyenne des notes et écart-type.....	133
4.4.1.5	Dispersion des notes autour de la médiane	134
4.4.1.6	Comparaison PE de TA statistique et PE de TA neuronale.....	135
4.4.1.7	Performances par étudiante.....	136
4.4.1.7.1	Nombre moyen d’erreurs.....	137
4.4.1.7.2	Notation	138
4.4.2	<i>Statistiques inférentielles</i>	139
4.4.2.1	Qualité globale	141
4.4.2.1.1	Effets sur les erreurs en phase d’acceptabilité	141
4.4.2.1.2	Effets sur les erreurs en phase d’adéquation	142
4.4.2.1.3	Effets sur les notes.....	144
4.4.2.2	Qualité par catégorie d’erreurs.....	147
4.4.2.2.1	Catégories en phase d’acceptabilité.....	148
4.4.2.2.2	Catégories en phase d’adéquation.....	152
4.4.2.3	Tableaux récapitulatifs.....	156
4.4.2.3.1	Influence du mode de traduction sur le nombre d’erreurs.....	156
4.4.2.3.2	Influence du texte source sur le nombre d’erreurs.....	157
4.4.2.3.3	Influence sur les notes	158
4.4.3	<i>Résultats complémentaires</i>	159
4.4.3.1	Étude du processus : Productivité.....	159
4.4.3.1.1	Productivité par texte	159
4.4.3.1.2	Productivité par étudiante.....	160
4.4.3.1.3	Productivité et qualité.....	161
4.4.3.2	Étude du processus : Effort de PE	163
4.4.3.2.1	Effort de PE et système de TA.....	164
4.4.3.2.2	Effort de PE et qualité	165
4.5	SYNTHÈSES ET RÉPONSES AUX HYPOTHÈSES	167
4.5.1	<i>Statistiques descriptives</i>	168
4.5.2	<i>Statistiques inférentielles</i>	169
4.5.2.1	Hypothèse HA1	169
4.5.2.2	Hypothèse HA2.....	169
4.5.2.2.1	Sous-hypothèse SHA2.1 – Erreurs d’adéquation.....	170

4.5.2.2.2	Sous-hypothèse SHA2.3 – Fluidité	170
4.5.2.2.3	Sous-hypothèse SHA2.2 – Calques fautifs	170
4.5.2.2.4	Sous-hypothèse SHA2.4 – Fautes d’orthographe, de grammaire et de syntaxe	171
4.5.2.2.5	Sous-hypothèse SHA2.5 – Texte source	171
4.5.2.3	Hypothèse HA3.....	171
4.5.3	<i>Étude du processus</i>	171
4.6	CONCLUSION	172
5	RECHERCHE EXPÉRIMENTALE – TEST	175
5.1	INTRODUCTION.....	175
5.2	HYPOTHÈSES DE RECHERCHE	175
5.2.1	<i>Hypothèse HA4</i>	175
5.2.2	<i>Hypothèse HA5</i>	176
5.2.2.1	Sous-hypothèse SHA5.1.....	176
5.2.2.2	Sous-hypothèse SHA5.2.....	177
5.2.2.3	Sous-hypothèse SHA5.3.....	177
5.2.2.4	Sous-hypothèse SHA5.4.....	177
5.2.3	<i>Hypothèse HA6</i>	177
5.2.4	<i>Hypothèse HA7</i>	178
5.2.5	<i>Améliorations méthodologiques</i>	178
5.2.5.1	Tâches.....	178
5.2.5.2	Corpus.....	179
5.2.5.3	Comparaison des versions <i>DeepL</i>	180
5.2.5.4	Exploitation du corpus	182
5.2.5.4.1	Approche qualitative : Évaluation humaine de la qualité	182
5.3	RÉSULTATS	186
5.3.1	<i>Statistiques descriptives</i>	186
5.3.1.1	Distribution des erreurs (diagramme circulaire)	186
5.3.1.2	Distribution des erreurs (histogramme)	188
5.3.1.3	Comparaison de la qualité par mode de traduction.....	189
5.3.1.3.1	Moyenne des notes	189
5.3.1.3.2	Taux de réussite.....	190
5.3.1.4	Moyenne des notes et écart-type	191
5.3.1.5	Dispersion des notes autour de la médiane	192
5.3.1.6	PE de <i>Google</i> vs PE de <i>DeepL</i>	193
5.3.1.7	Performances par étudiante.....	194
5.3.1.7.1	Nombre moyen d’erreurs.....	194
5.3.1.7.2	Notation	196
5.3.1.8	Qualité des TA brutes.....	198
5.3.2	<i>Statistiques inférentielles</i>	200
5.3.2.1	Qualité globale	201
5.3.2.1.1	Effets sur les erreurs en phase d’acceptabilité	201
5.3.2.1.2	Effets sur les erreurs en phase d’adéquation	203
5.3.2.1.3	Effets sur les notes.....	203
5.3.2.2	Qualité par catégorie d’erreurs.....	206
5.3.2.2.1	Catégories en phase d’acceptabilité.....	206
5.3.2.2.3	Catégories en phase d’adéquation.....	210
5.3.2.3	Tableaux récapitulatifs.....	213
5.3.2.3.1	Influence du mode de traduction sur le nombre d’erreurs.....	213
5.3.2.3.2	Influence du texte source sur le nombre d’erreurs	214
5.3.2.3.3	Influence sur les notes	215
5.3.3	<i>Résultats complémentaires</i>	216
5.3.3.1	Étude du processus : Productivité.....	216
5.3.3.1.1	Productivité par texte	216
5.3.3.1.2	Productivité par étudiante.....	217
5.3.3.1.3	Productivité et qualité.....	218
5.3.3.2	Étude du processus : Effort de PE	220

5.3.3.2.1	Effort de PE et moteur de TA.....	220
5.3.3.2.2	Effort de PE et qualité	221
5.3.3.3	Recueil d'impressions : post-édition ou traduction ?	223
5.3.3.3.1	Facteurs positifs, neutres ou négatifs	224
5.4	SYNTHÈSES ET RÉPONSES AUX HYPOTHÈSES	227
5.4.1	<i>Statistiques descriptives et inférentielles</i>	227
5.4.1.1	Hypothèse HA4.....	227
5.4.1.2	Hypothèse HA5.....	228
5.4.1.2.1	Sous-hypothèse SHA5.1	228
5.4.1.2.2	Sous-hypothèse SHA5.2	228
5.4.1.2.3	Sous-hypothèse SHA5.3	228
5.4.1.2.4	Sous-hypothèse SHA5.4	228
5.4.1.3	Hypothèse HA6.....	229
5.4.1.4	Hypothèse HA7.....	229
5.4.2	<i>Étude du processus</i>	230
5.5	CONCLUSION.....	230
5.5.1	<i>Hypothèses explicatives (Prétest & Test)</i>	231
5.5.1.1	Accès au sens du texte source	231
5.5.1.2	Effet nivelant en PE de TA neuronale	232
5.5.1.3	Maîtrise de l'orthographe et de la grammaire.....	233
5.5.1.4	Effet fantôme de la TA	233
6	VOUS AVEZ DIT POST-EDITESE ?	235
6.1	MESURES LINGUISTIQUES AUTOMATIQUES	235
6.1.1	<i>Richesse lexicale</i>	236
6.1.1.1	La densité lexicale et la diversité lexicale	237
6.1.1.1.1	La densité lexicale	237
6.1.1.1.2	La diversité lexicale ou variation lexicale.....	238
6.1.1.2	Hypothèse	240
6.1.1.3	Densité lexicale : Calculs et résultats	240
6.1.1.3.1	Calculs	240
6.1.1.3.2	Résultats.....	241
6.1.1.4	Diversité lexicale : calculs et résultats	241
6.1.1.4.1	Calculs	241
6.1.1.4.2	Résultats.....	242
6.1.2	<i>Longueur moyenne des phrases et coefficient de foisonnement</i>	244
6.1.2.1	Hypothèse	246
6.1.2.2	Calculs	246
6.1.2.3	Résultats.....	247
6.1.3	<i>Équivalence syntaxique</i>	248
6.1.3.1	Hypothèse	249
6.1.3.2	Calculs	249
6.1.3.3	Résultats.....	249
6.1.4	<i>Synthèse et discussion</i>	250
6.2	ANALYSE QUALITATIVE DES ERREURS	256
6.2.1	<i>Observations générales</i>	256
6.2.2	<i>Effets de la (PE de) TA neuronale sur la qualité finale</i>	257
6.2.2.1	Effets positifs.....	257
6.2.2.1.1	Calques fautifs évités	257
6.2.2.1.2	Régionalismes évités.....	260
6.2.2.1.3	Maîtrise des règles d'usage du français (grammaire et orthographe).....	260
6.2.2.2	Effets négatifs.....	262
6.2.2.2.1	Marques de littéralité	262
6.2.2.2.2	Effacement d'informations.....	267
6.2.2.2.3	Rupture de cohésion textuelle et biais de genre.....	268
6.2.2.2.4	Rupture de cohérence - Irrationalité	271
6.2.2.2.5	Incohérence temporelle.....	272

6.2.2.2.6	Irrégularité terminologique	274
6.2.2.2.7	Perte de richesse linguistique	275
6.2.3	<i>Synthèse et discussion</i>	278
6.3	CONCLUSION	278
6.3.1	<i>Mesures linguistiques automatiques</i>	278
6.3.1.1	Richesse lexicale	278
6.3.1.2	Longueur moyenne des phrases	278
6.3.1.3	Équivalence syntaxique	279
6.3.2	<i>Analyse qualitative des erreurs</i>	279
7	CONNAISSANCES ET PERCEPTIONS DES ÉTUDIANTES	281
7.1	INTRODUCTION	281
7.2	HYPOTHÈSES	282
7.3	REVUE DE LA LITTÉRATURE	285
7.4	MÉTHODOLOGIE	297
7.4.1	<i>Composition du sondage pré-expérience</i>	298
7.4.2	<i>Composition du sondage post-expérience</i>	298
7.4.3	<i>Participant·es</i>	299
7.4.3.1	Fournée 2018	299
7.4.3.1.1	Sondage pré-expérience	299
7.4.3.1.2	Sondage post-expérience	299
7.4.3.2	Fournée 2021	299
7.4.3.2.1	Sondage pré-expérience	299
7.4.3.2.2	Sondage post-expérience	299
7.5	RÉSULTATS	300
7.5.1	<i>Sondage pré-expérience</i>	301
7.5.2	<i>Sondage post-expérience</i>	331
7.5.2.1	Conclusion (questions 1 – 2A – 2B – 3)	335
7.6	SYNTHÈSE ET VÉRIFICATION DES HYPOTHÈSES	356
7.6.1	<i>Sondages pré-expérience</i>	356
7.6.2	<i>Sondages post-expérience</i>	359
7.7	CONCLUSION ET DISCUSSION	361
8	INTÉGRATION DE LA POST-ÉDITION AUX CURSUS DE FORMATION EN TRADUCTION : ENJEUX, ATOUTS ET DÉFIS	363
8.1	ENJEUX D'UNE FORMATION À LA TA ET À LA PE	363
8.1.1	<i>Quels atouts ?</i>	366
8.1.2	<i>Quand prévoir cette formation ?</i>	369
8.1.2.1	Prérequis de formation	370
8.1.2.2	En début de parcours ?	371
8.1.3	<i>Quel contenu ?</i>	372
8.1.3.1	Acquisition de compétences spécifiques	374
8.1.3.2	Trois piliers de la formation	375
8.1.3.2.1	Apprécier les avantages et les limites de la TA	375
8.1.3.2.2	Sensibiliser aux erreurs récurrentes	376
8.1.3.2.3	Faire ressortir les spécificités humaines	378
8.2	POUR UNE APPROCHE RAISONNÉE DE LA TA NEURONALE	379
8.2.1	<i>Responsabilisation</i>	381
8.2.2	<i>Autonomisation</i>	382
8.2.3	<i>Conscientisation</i>	383
8.3	LES DÉFIS EN POST-ÉDITION	384
8.3.1	<i>Développer la capacité de discernement</i>	385
8.3.2	<i>Trouver le juste équilibre entre sur-édition et sous-édition</i>	386
8.3.3	<i>Assurer la cohésion et la cohérence textuelles</i>	389
8.3.4	<i>Adapter la qualité aux modalités de PE</i>	391

8.3.5	<i>Atténuer l'effet fantôme</i>	392
8.3.6	<i>Déjouer l'illusion de fluidité</i>	394
8.3.7	<i>Revaloriser le rôle de l'humain</i>	396
8.4	QUEL AVENIR POUR LA TRADUCTION PROFESSIONNELLE ?	399
8.4.1	<i>Les outils de TA : une réelle menace pour la profession ?</i>	399
8.4.1.1	L'humain, l'humain et encore l'humain.....	401
8.4.2	<i>Traductrice augmentée ou traductrice diminuée ?</i>	402
8.4.2.1	TA interactive et TA adaptative	404
8.4.3	<i>Quel rôle pour l'humain à l'avenir ?</i>	406
9	CONCLUSIONS ET PERSPECTIVES	409
9.1	RÉCAPITULATIF DES RÉSULTATS	410
9.1.1	<i>Résultats empiriques</i>	410
9.1.1.1	Étude du produit.....	410
9.1.1.1.1	Prétest.....	410
9.1.1.1.2	Test	411
9.1.1.2	Étude du processus	411
9.1.1.2.1	Prétest.....	412
9.1.1.2.2	Test	412
9.1.2	<i>Vous avez dit post-editeuse ?</i>	412
9.1.2.1	Mesures linguistiques automatiques	412
9.1.2.1.1	Richesse lexicale	412
9.1.2.1.2	Longueur moyenne des phrases	412
9.1.2.1.3	Équivalence syntaxique	413
9.1.2.2	Analyse qualitative des erreurs.....	413
9.1.3	<i>Étude des connaissances et perceptions</i>	413
9.1.4	<i>Enjeux, atouts et défis de l'apprentissage de la PE</i>	415
9.2	LIMITES EXPÉRIMENTALES	416
9.3	PERSPECTIVES	417
	BIBLIOGRAPHIE	419
	ANNEXE A	455
	TEXTES SOURCES ET TA BRUTES.....	455
	ANNEXE B	467
	CONSIGNES.....	467
	<i>B.1 Consignes étudiantes</i>	467
	<i>B.2 Consignes évaluatrices</i>	469
	ANNEXE C	476
	STATISTIQUES INFÉRENTIELLES – PRÉTEST	476
	ANNEXE D	489
	STATISTIQUES INFÉRENTIELLES – TEST	489
	ANNEXE E	501
	MESURES LINGUISTIQUES AUTOMATIQUES	501
	ANNEXE F	503
	RECUEIL IMPRESSIONS ÉVALUATRICES	503

Liste des figures

FIGURE 1 : <i>MACHINE TRANSLATION LANDSCAPE</i> (SAVENKOV ET LOPEZ, 2022, P. 36)	16
FIGURE 2 : PLONGEMENT LEXICAL (GEITGEY, 2016, S.P.)	19
FIGURE 3 : RÉSEAUX DE NEURONES AVEC UNE COUCHE CACHÉE (KOEHN, 2020, P. 69)	20
FIGURE 4 : ARCHITECTURE ENCODEUR-DÉCODEUR (CHO ET AL., 2014, P. 105)	21
FIGURE 5 : RNN – « MACHINE LEARNING IS FUN » (GEITGEY, 2016, S.P.)	21
FIGURE 6 : RÉSEAU DE NEURONES RÉCURRENTS (KOEHN, 2020, P. 110)	22
FIGURE 7 : CONTINUUM PE (DE FARIA PIRES, 2020, P. 56)	36
FIGURE 8 : MODÈLE DE COMPÉTENCES EN PE (NITZKE ET HANSEN-SCHIRRA, 2021, P. 70)	39
FIGURE 9 : SUBJECTIVITÉ DANS L'ÉVALUATION (DE KETELE ET ROEGIER, 1993, P. 77)	59
FIGURE 10 : PERFORMANCES DES MOTEURS DE TA (SAVENKOV ET LOPEZ, 2022)	92
FIGURE 11 : CONSTITUTION DU CORPUS (PRÉTEST)	95
FIGURE 12 : CONSTITUTION DU CORPUS (TEST)	96
FIGURE 13 : PRINCIPE DE RÉPARTITION DES TÂCHES (PRÉTEST)	98
FIGURE 14 : INTERFACE <i>MATECAT</i>	98
FIGURE 15 : INTERFACE <i>MATECAT</i> – MYMEMORY	99
FIGURE 16 : INTERFACE <i>MATECAT</i> – MACHINE TRANSLATION	99
FIGURE 17 : CONFIGURATION DU PRÉTEST	100
FIGURE 18 : PHASE D'ACCEPTABILITÉ – PRÉTEST	104
FIGURE 19 : PHASE D'ADÉQUATION – PRÉTEST	105
FIGURE 20 : FORMULE ACCORD OBSERVÉ (SHWETA BAJPAI ET CHATURVEDI, 2015)	109
FIGURE 21 : DIAGRAMME CIRCULAIRE (ACCEPTABILITÉ) – PRÉTEST	127
FIGURE 22 : DIAGRAMME CIRCULAIRE (ADÉQUATION) – PRÉTEST	128
FIGURE 23 : HISTOGRAMME (ACCEPTABILITÉ) – PRÉTEST	129
FIGURE 24 : HISTOGRAMME (ADÉQUATION) – PRÉTEST	130
FIGURE 25 : MOYENNE DES NOTES – PRÉTEST	131
FIGURE 26 : TAUX DE RÉUSSITE – PRÉTEST	131
FIGURE 27 : MOYENNE DES NOTES ET ÉCART-TYPE – PRÉTEST	134
FIGURE 28 : DISPERSION DES NOTES AUTOUR DE LA MÉDIANE – PRÉTEST	135
FIGURE 29 : COMPARAISON DES SYSTÈMES DE TA (STATISTIQUE ET NEURONAL)	136
FIGURE 30 : PERFORMANCES PAR ÉTUDIANTE POUR LA PHASE D'ADÉQUATION	137
FIGURE 31 : PERFORMANCES PAR ÉTUDIANTE POUR LA PHASE D'ACCEPTABILITÉ	138
FIGURE 32 : NOTES OBTENUES PAR ÉTUDIANTE	138
FIGURE 33 : PRODUCTIVITÉ-QUALITÉ – TRADUCTION HUMAINE (PRÉTEST)	161
FIGURE 34 : PRODUCTIVITÉ-QUALITÉ – PE DE TAS	162
FIGURE 35 : PRODUCTIVITÉ-QUALITÉ – PE DE TAN (DEEPL)	162
FIGURE 36 : ÉCHELLE D'INTERPRÉTATION <i>r</i> (RUMSEY, 2003)	163
FIGURE 37 : EFFORT DE PE ET QUALITÉ – PE DE TA STATISTIQUE – PRÉTEST	166
FIGURE 38 : EFFORT DE PE ET QUALITÉ – PE DE TA NEURONALE – PRÉTEST	166
FIGURE 39 : CAPTURE D'ÉCRAN – CENTRE DE CONTRÔLE ENSEIGNANT (TELEVIC EDUCATION, 2018)	179
FIGURE 40 : CAPTURE D'ÉCRAN – RECHERCHE DOCUMENTAIRE (TELEVIC EDUCATION, 2018)	179
FIGURE 41 : PHASE D'ACCEPTABILITÉ – TEST	184
FIGURE 42 : PHASE D'ADÉQUATION – TEST	184
FIGURE 43 : DIAGRAMME CIRCULAIRE (ACCEPTABILITÉ) – TEST	186
FIGURE 44 : DIAGRAMME CIRCULAIRE (ADÉQUATION) – TEST	187
FIGURE 45 : HISTOGRAMME (ACCEPTABILITÉ) – TEST	188

FIGURE 46 : HISTOGRAMME (ADÉQUATION) – TEST	188
FIGURE 47 : MOYENNE DES NOTES – TEST	189
FIGURE 48 : TAUX DE RÉUSSITE SELON LE MODE DE TRADUCTION – TEST.....	190
FIGURE 49 : MOYENNE DES NOTES ET ÉCART-TYPE – TEST.....	191
FIGURE 50 : DISPERSION DES NOTES AUTOUR DE LA MÉDIANE – TEST.....	192
FIGURE 51 : PE DE <i>GOOGLE</i> VS PE DE <i>DEEPL</i>	193
FIGURE 52 : COMPARAISON NOMBRE D’ERREURS PAR ÉTUDIANTE (ADÉQUATION) <i>GOOGLE</i>	194
FIGURE 53 : COMPARAISON NOMBRE D’ERREURS PAR ÉTUDIANTE (ADÉQUATION) <i>DEEPL</i>	195
FIGURE 54 : COMPARAISON NOMBRE D’ERREURS PAR ÉTUDIANTE (ACCEPTABILITÉ) <i>GOOGLE</i>	195
FIGURE 55 : COMPARAISON NOMBRE D’ERREURS PAR ÉTUDIANTE (ACCEPTABILITÉ) <i>DEEPL</i>	196
FIGURE 56 : EFFET NIVELANT EN PE DE TAN (<i>GOOGLE</i>)	197
FIGURE 57 : EFFET NIVELANT EN PE DE TAN (<i>DEEPL</i>).....	198
FIGURE 58 : PRODUCTIVITÉ-QUALITÉ – TRADUCTION HUMAINE – TEST.....	218
FIGURE 59 : PRODUCTIVITÉ-QUALITÉ – PE DE <i>GOOGLE</i> – TEST	219
FIGURE 60 : PRODUCTIVITÉ-QUALITÉ – PE DE <i>DEEPL</i> – TEST.....	219
FIGURE 61 : EFFORT DE PE ET QUALITÉ – PE DE <i>GOOGLE</i>	222
FIGURE 62 : EFFORT DE PE ET QUALITÉ – PE DE <i>DEEPL</i>	222
FIGURE 63 : CALCUL DE DENSITÉ LEXICALE.....	240
FIGURE 64 : CALCUL DE DIVERSITÉ LEXICALE	242
FIGURE 65 : CALCUL DE LONGUEUR MOY. DES PHRASES.....	246
FIGURE 66 : CALCUL DU FOISONNEMENT	246
FIGURE 67 : EXEMPLE D’ARBRE DE DÉPENDANCE (VANROY ET AL., 2021, P. 264).....	248
FIGURE 68 : BIAIS DE GENRE (VANMASSENHOVE, 2020, P. 132)	268
FIGURE 69 : LA TA DANS LA PRESSE ET LES MÉDIAS (SCHUMACHER, 2020A)	289
FIGURE 70 : LE MIRACLE DE LA TA NEURONALE (ROSSI, 2019B, DIAPOSITIVE 3).....	289
FIGURE 71 : RAISONS DU REFUS DE LA POST-ÉDITION (SFT, 2022, DIAPOSITIVE 27).....	292
FIGURE 72 : LES USAGES ACTUELS (ROSSI, 2019B, DIAPOSITIVE 8).....	296
FIGURE 73 : CAPTURE D’ÉCRAN – SONDAGE VIA LIMESURVEY	298
FIGURE 74 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 1)	301
FIGURE 75 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 2.1)	304
FIGURE 76 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 2.3).....	308
FIGURE 77 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 3)	310
FIGURE 78 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 4)	312
FIGURE 79 : SONDAGE POST-EXPÉRIENCE (QUESTION 1)	331
FIGURE 80 : SONDAGE POST-EXPÉRIENCE (QUESTION 2A)	332
FIGURE 81 : SONDAGE POST-EXPÉRIENCE (QUESTION 2B)	333
FIGURE 82 : SONDAGE POST-EXPÉRIENCE (QUESTION 3)	334
FIGURE 83 : SONDAGE POST-EXPÉRIENCE (QUESTION 10)	345
FIGURE 84 : RAPPORT NIMDZI 2023 (HICKEY, 2023)	400
FIGURE 85 : LA TRADUCTION HUMAINE AUGMENTÉE (LOMMEL, 2020)	404
FIGURE 86 : CONTINUUM DE L’INTERVENTION HUMAINE EN PE (VIEIRA, 2019, P. 328).....	405

Liste des tableaux

TABLEAU 1 : SCORES BLEU	97
TABLEAU 2 : PONDÉRATION NOTATION SUR 20 – PRÉTEST	108
TABLEAU 3 : DONNÉES <i>UAM CORPUS TOOL</i>	109
TABLEAU 4 : TAUX D’ACCORD INTER-ÉVALUATEUR (%) – PRÉTEST 2018	110
TABLEAU 5 : TAUX D’ACCORD INTER-ÉVALUATEUR (%) – TEST2021	110
TABLEAU 6 : TAUX D’ACCORD INTER-ÉVALUATEUR (%) (STYLE & CALQUE EXCLUS) – TEST2021	111
TABLEAU 7 : COEFFICIENT DE CORRÉLATION INTRA-CLASSE (TEXTE 2 – ACCEPTABILITÉ)	112
TABLEAU 8 : COEFFICIENT DE CORRÉLATION INTRA-CLASSE (TEXTE 2 – ADÉQUATION)	113
TABLEAU 9 : COEFFICIENT DE CORRÉLATION INTRA-CLASSE (TEXTE 3 – ACCEPTABILITÉ)	113
TABLEAU 10 : COEFFICIENT DE CORRÉLATION INTRA-CLASSE (TEXTE 3 – ADÉQUATION)	113
TABLEAU 11 : COEFFICIENT DE CORRÉLATION INTRA-CLASSE (TEXTE 1 – ACCEPTABILITÉ)	115
TABLEAU 12 : COEFFICIENT DE CORRÉLATION INTRA-CLASSE (TEXTE 1 – ADÉQUATION)	115
TABLEAU 13 : COEFFICIENT DE CORRÉLATION INTRA-CLASSE (TEXTE 2 – ACCEPTABILITÉ)	115
TABLEAU 14 : COEFFICIENT DE CORRÉLATION INTRA-CLASSE (TEXTE 2 – ADÉQUATION)	116
TABLEAU 15 : COEFFICIENT DE CORRÉLATION INTRA-CLASSE (TEXTE 3 – ACCEPTABILITÉ)	116
TABLEAU 16 : COEFFICIENT DE CORRÉLATION INTRA-CLASSE (TEXTE 3 – ADÉQUATION)	116
TABLEAU 17 : ANOVA — DISTRIBUTION DES ERREURS EN PHASE D’ACCEPTABILITÉ	141
TABLEAU 18 : ANOVA — DISTRIBUTION DES ERREURS EN PHASE D’ADÉQUATION	143
TABLEAU 19 : ANOVA — DISTRIBUTION DES NOTES EN PHASE D’ACCEPTABILITÉ	144
TABLEAU 20 : ANOVA — DISTRIBUTION DES NOTES EN PHASE D’ADÉQUATION	145
TABLEAU 21 : ANOVA — DISTRIBUTION DES NOTES GLOBALES	146
TABLEAU 22 : INFLUENCE DU MODE DE TRADUCTION SUR LE NOMBRE D’ERREURS - PRÉTEST	156
TABLEAU 23 : INFLUENCE DU TEXTE SOURCE SUR LE NOMBRE D’ERREURS – PRÉTEST	157
TABLEAU 24 : INFLUENCE DU MODE DE TRADUCTION SUR LES NOTES – PRÉTEST	158
TABLEAU 25 : GAINS DE PRODUCTIVITÉ PAR TEXTE – PRÉTEST	159
TABLEAU 26 : GAINS DE PRODUCTIVITÉ PAR ÉTUDIANTE – PRÉTEST	160
TABLEAU 27 : EFFORT MOYEN DE PE PAR PRODUCTION – PRÉTEST	164
TABLEAU 28 : DÉTAILS EFFORT MOYEN DE PE – PRÉTEST	165
TABLEAU 29 : COMPARAISON <i>DEEPL</i> JAN 2018 vs <i>DEEPL</i> SEPT 2020	182
TABLEAU 30 : PONDÉRATION NOTATION SUR 20 – TEST	185
TABLEAU 31 : NOTES ATTRIBUÉES AUX TA BRUTES	199
TABLEAU 32 : DISTRIBUTION ERREURS EN TA BRUTE (ACCEPTABILITÉ ET ADÉQUATION)	200
TABLEAU 33 : ANOVA — DISTRIBUTION DES ERREURS EN PHASE D’ACCEPTABILITÉ	201
TABLEAU 34 : ANOVA – DISTRIBUTION DES ERREURS EN PHASE D’ADÉQUATION	203
TABLEAU 35 : ANOVA — DISTRIBUTION DES NOTES EN PHASE D’ACCEPTABILITÉ	204
TABLEAU 36 : ANOVA — DISTRIBUTION DES NOTES EN PHASE D’ADÉQUATION	205
TABLEAU 37 : ANOVA — DISTRIBUTION DES NOTES GLOBALES	205
TABLEAU 38 : INFLUENCE DU MODE DE TRADUCTION SUR LE NOMBRE D’ERREURS – TEST	213
TABLEAU 39 : INFLUENCE DU TEXTE SOURCE SUR LE NOMBRE D’ERREURS – TEST	214
TABLEAU 40 : INFLUENCE SUR LES NOTES – TEST	215
TABLEAU 41 : GAINS DE PRODUCTIVITÉ PAR TEXTE – TEST	216
TABLEAU 42 : GAINS DE PRODUCTIVITÉ PAR ÉTUDIANTE – TEST	217
TABLEAU 43 : EFFORT MOYEN DE PE PAR PRODUCTION – TEST	220
TABLEAU 44 : DÉTAILS EFFORT MOYEN DE PE EN POURCENTAGE – TEST	221
TABLEAU 45 : RECUEIL IMPRESSIONS TH/PE	224
TABLEAU 46 : FACTEURS POSITIFS, NEUTRES OU NÉGATIFS – TH vs PE	226

TABLEAU 47 : DENSITÉ LEXICALE – PRÉTEST	241
TABLEAU 48 : DENSITÉ LEXICALE – TEST	241
TABLEAU 49 : DIVERSITÉ LEXICALE – PRÉTEST	242
TABLEAU 50 : DIVERSITÉ LEXICALE – TEST	242
TABLEAU 51 : LONGUEUR DES PHRASES ET FOISONNEMENT – PRÉTEST	247
TABLEAU 52 : LONGUEUR DES PHRASES ET FOISONNEMENT – TEST	247
TABLEAU 53 : ÉQUIVALENCE SYNTAXIQUE – PRÉTEST	249
TABLEAU 54 : ÉQUIVALENCE SYNTAXIQUE – TEST	249
TABLEAU 55 : CALQUES FAUTIFS ÉVITÉS (1)	258
TABLEAU 56 : CALQUES FAUTIFS ÉVITÉS (2)	258
TABLEAU 57 : CALQUES FAUTIFS ÉVITÉS (3)	259
TABLEAU 58 : RÉGIONALISMES ÉVITÉS	260
TABLEAU 59 : MAÎTRISE DES RÈGLES D'USAGE DU FRANÇAIS (1)	261
TABLEAU 60 : MAÎTRISE DES RÈGLES D'USAGE DU FRANÇAIS (2)	261
TABLEAU 61 : MARQUES DE LITTÉRALITÉ ET CALQUES FAUTIFS	263
TABLEAU 62 : MARQUES DE LITTÉRALITÉ ET GLISSEMENTS DE SENS ET CONTRESSENS	264
TABLEAU 63 : LA PLACE DES COMPLÉMENTS CIRCONSTANCIELS	265
TABLEAU 64 : PONCTUATION CALQUÉE	266
TABLEAU 65 : EFFACEMENT D'INFORMATIONS	267
TABLEAU 66 : RUPTURES DE COHÉSION TEXTUELLE ET BIAIS DE GENRE	270
TABLEAU 67 : IRRATIONALITÉ	271
TABLEAU 68 : INCOHÉRENCE TEMPORELLE (1)	272
TABLEAU 69 : INCOHÉRENCE TEMPORELLE (2)	273
TABLEAU 70 : IRRÉGULARITÉ TERMINOLOGIQUE	274
TABLEAU 71 : PERTE DE RICHESSE LINGUISTIQUE EN PE (1)	276
TABLEAU 72 : PERTE DE RICHESSE LINGUISTIQUE EN PE (2)	277
TABLEAU 73 : DEGRÉ DE SATISFACTION PAR RAPPORT À LA TÂCHE DE PE (CARL ET AL., 2015, P. 155)	293
TABLEAU 74 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 1)	302
TABLEAU 75 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 2)	303
TABLEAU 76 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 2.1)	305
TABLEAU 77 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 2.2)	307
TABLEAU 78 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 2.3)	309
TABLEAU 79 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 3)	311
TABLEAU 80 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 4)	313
TABLEAU 81 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 5)	315
TABLEAU 82 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 6)	315
TABLEAU 83 : SONDAGE PRÉ-EXPÉRIENCE (GAIN DE TEMPS)	317
TABLEAU 84 : SONDAGE PRÉ-EXPÉRIENCE (ACCÈS AU SENS)	318
TABLEAU 85 : SONDAGE PRÉ-EXPÉRIENCE (COMPARAISON)	318
TABLEAU 86 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 7)	319
TABLEAU 87 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 8)	320
TABLEAU 88 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 8.1)	321
TABLEAU 89 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 9)	322
TABLEAU 90 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 9.1.) ARGUMENTS FAVORABLES (A)	324
TABLEAU 91 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 9.1.) ARGUMENTS FAVORABLES (B)	325
TABLEAU 92 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 9.1.) ARGUMENTS FAVORABLES (C)	325
TABLEAU 93 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 9.1.) ARGUMENTS FAVORABLES (D)	326
TABLEAU 94 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 9.1.) ARGUMENTS FAVORABLES (E)	327
TABLEAU 95 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 9.1.) ARGUMENTS FAVORABLES (F)	327
TABLEAU 96 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 9.1.) ARGUMENTS DÉFAVORABLES (A)	328
TABLEAU 97 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 9.1.) ARGUMENTS DÉFAVORABLES (B)	328

TABLEAU 98 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 9.1.) ARGUMENTS DÉFAVORABLES (C)	329
TABLEAU 99 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 9.1.) ARGUMENTS DÉFAVORABLES (D)	329
TABLEAU 100 : SONDAGE PRÉ-EXPÉRIENCE (QUESTION 10)	330
TABLEAU 101 : SONDAGE POST-EXPÉRIENCE (QUESTION 4)	336
TABLEAU 102 : SONDAGE POST-EXPÉRIENCE (QUESTION 5A)	337
TABLEAU 103 : SONDAGE POST-EXPÉRIENCE (QUESTION 5B)	338
TABLEAU 104 : SONDAGE POST-EXPÉRIENCE (QUESTION 6)	339
TABLEAU 105 : SONDAGE POST-EXPÉRIENCE (QUESTION 7)	341
TABLEAU 106 : SONDAGE POST-EXPÉRIENCE (QUESTION 8A)	342
TABLEAU 107 : SONDAGE POST-EXPÉRIENCE (QUESTION 8B)	343
TABLEAU 108 : SONDAGE POST-EXPÉRIENCE (QUESTION 9)	344
TABLEAU 109 : SONDAGE POST-EXPÉRIENCE (QUESTION 10.1-A)	347
TABLEAU 110 : SONDAGE POST-EXPÉRIENCE (QUESTION 10.1-B)	348
TABLEAU 111 : SONDAGE POST-EXPÉRIENCE (QUESTION 10.1-C)	348
TABLEAU 112 : SONDAGE POST-EXPÉRIENCE (QUESTION 10.1-D)	349
TABLEAU 113 : SONDAGE POST-EXPÉRIENCE (QUESTION 11)	349
TABLEAU 114 : SONDAGE POST-EXPÉRIENCE (QUESTION 12)	350
TABLEAU 115 : SONDAGE POST-EXPÉRIENCE (QUESTION 13)	350
TABLEAU 116 : SONDAGE POST-EXPÉRIENCE (QUESTION 13.1)	351
TABLEAU 117 : SONDAGE POST-EXPÉRIENCE (QUESTION 13.2)	352
TABLEAU 118 : SONDAGE POST-EXPÉRIENCE (QUESTION 14)	352
TABLEAU 119 : SONDAGE POST-EXPÉRIENCE (QUESTION 14.1)	353
TABLEAU 120 : SONDAGE POST-EXPÉRIENCE (QUESTION 15)	354
TABLEAU 121 : SONDAGE POST-EXPÉRIENCE (QUESTION 16)	355

Liste des abréviations

IA	intelligence artificielle
NMT	<i>Neural Machine Translation</i>
PE	post-édition(s)
TA	traduction automatique
TAL	traitement automatique du langage
TAN	traduction automatique neuronale
TAO	traduction assistée par ordinateur
TAS	traduction automatique statistique
TH	traduction(s) humaine(s)
TC	texte cible
TQA	<i>Translation Quality Assessment</i>
TS	texte source

Avant-propos

Choix terminologiques

Traduction humaine

Dans cette thèse, nous avons délibérément choisi de ne pas faire usage des termes « biotraducteur » et « biotraduction ». En effet, nous sommes d'avis que ces néologismes, apparus pour la première fois dans le roman de science-fiction *Le Revenant de Fomalhaut* par l'auteur Jean-Louis Trudel en 2002 (Froeliger, 2013, p. 20), peuvent être associés à des connotations négatives et que leur usage pourrait laisser entendre « que le traducteur devient l'exception, et que son travail confine à la science-fiction » (Dirand et Rossi, 2019, p. 75). Pour ces raisons, nous employons systématiquement les termes « traducteur/trice » et « traduction humaine » lorsqu'il s'agit de la comparer à la « traduction automatique ». À l'instar de Martikainen (2019c), précisons que, contrairement à ce que ces termes laissent entendre, il ne s'agit pas de comparer une méthode de traduction strictement humaine à une méthode totalement automatisée. Il s'avère effectivement que « les frontières entre ces deux processus sont poreuses et se brouillent en réalité de plus en plus dans les flux de traduction professionnelle » (p. 21) dès lors que ces deux méthodes de traduction impliquent toutes deux une intervention humaine, ainsi que l'utilisation d'outils d'aide à la traduction à divers degrés (*ibid.*).

« Erreur » et « faute »

Nous avons tenu à faire une distinction entre le terme « erreur » et le terme « faute », même si ces deux termes sont fréquemment employés de manière interchangeable en milieu didactique. Comme le souligne J.-P. Astolfi (2008) : « la façon de considérer l'erreur en pédagogie a évolué. On est globalement passé d'une conception de l'erreur tenue pour une faute et donnant lieu à sanction, à une conception nouvelle : l'erreur est un témoin qui permet de repérer les difficultés auxquelles se heurte le processus d'apprentissage ; ce terme revêt aujourd'hui un intérêt didactique particulier » (p. 87). L'erreur révèle aujourd'hui « non pas l'inaptitude de l'élève mais l'existence d'un savoir incomplet, mal assimilé ou mal consolidé. [...] il convient de reconnaître l'erreur comme une non réussite [sic] signifiante et comme l'un des pivots de la démarche pédagogique » (Bédoire et al., 2001, p. 39). Pour ces raisons, nous avons choisi de privilégier l'emploi du mot « erreur », notamment dans l'expression « erreur de traduction »¹ ; nous employons toutefois le terme « faute » dans le sens d'un « manquement à une norme, à un principe dans une discipline donnée : faute d'inattention, de frappe, de syntaxe, d'orthographe, de grammaire » (Druide, 2022).

¹ Notons que la tournure « faute de traduction » apparaît à plusieurs reprises dans ce document lorsque nous citons notamment Delisle (2013).

Conventions typographiques

La question de l'écriture inclusive méritait que l'on s'y attarde dès lors que le débat refait surface en force ces dernières années et que les deux institutions auxquelles je suis rattachée ont émis des recommandations en la matière à destination de leurs membres.

Tout d'abord, les recommandations formulées en 2019 par l'Université de Liège (ULiège) dans son guide pratique « l'égalité dans la langue »² et envoyées à l'ensemble des membres du personnel de l'ULiège préconisent de donner une meilleure visibilité aux femmes dans le monde du travail, « en intégrant dans les pratiques des formes linguistiques plus respectueuses de l'identité féminine » (s.p.). Quant à l'Université de Genève (UNIGE), l'institution encourage également l'utilisation d'un langage épïcène et inclusif dans une directive publiée en mars 2020³. Aussi, d'abord par conviction personnelle, mais également pour des raisons de lisibilité, j'ai décidé, pour l'ensemble de la thèse (à l'exception du résumé), de ne pas utiliser le masculin générique. Parmi les pistes empruntées, j'ai choisi de recourir à des formulations épïcènes et d'expérimenter le féminin dit générique et italisé (les « traductrices » pour désigner « les traducteurs et les traductrices »). Cet usage du féminin générique est recommandé, entre autres, par Gisèle Sapiro (directrice de recherche au CNRS et directrice d'études EHESS) et par Olivier Godechot (directeur de recherche au CNRS et professeur à Sciences Po), « soucieuses de nous faire économiser de l'encre et d'épargner les règles de grammaire » (Grunenwald, 2021, p. 97). Cet emploi du féminin générique italisé est apparu d'autant plus pertinent que les groupes d'étudiantes et d'intervenantes (enseignantes et évaluatrices) dans cette étude sont constituées uniquement, ou presque exclusivement, de femmes.

² https://www.uliege.be/cms/c_11196151/fr/l-egalite-dans-la-langue

³ https://www.unige.ch/rectorat/egalite/files/7215/8695/9897/Directive_redaction_inclusive_epicene_UNI_GE_Web_mars_2020.pdf

1 Introduction

In looking ahead, what remains unclear is the particular roles that translators and everyday users of translation will play in an increasingly technology-dependent globalized society. As translation technologies intersect and sometimes subsume the translation process entirely, an important factor in moving toward the effective use of these technologies and in preparing for future changes is a critical and informed approach in understanding what such tools can and cannot do and how users should use them to achieve the desired result.

STEPHEN DOHERTY (2016)

1.1 Contexte et motivation

1.1.1 L'essor de la traduction automatique

Avec la révolution numérique s'est amorcé le basculement de notre civilisation du signe et du texte vers une civilisation de l'algorithme (Rouvroy, 2018, p. 417). Grâce à cette évolution de l'algorithmique et aux progrès incontestables de la recherche en intelligence artificielle (IA), le grand public assiste, sceptique ou admiratif, à l'automatisation progressive de nombreuses tâches complexes de la vie quotidienne, telles que la conduite, le diagnostic médical, la reconnaissance faciale ou encore la traduction, qui a aujourd'hui le vent en poupe. En 2014, l'application des réseaux de neurones artificiels (*deep learning*) au domaine du traitement automatique du langage (TAL) annonçait un réel changement de paradigme en traduction automatique (TA) (Dirand et Rossi, 2019 ; Massardo et van der Meer, 2017 ; Martikainen, 2019c). Considérée comme l'un des objectifs historiques (Hartley et Popescu-Belis, 2004) et l'un des plus vieux rêves de l'IA (Poibeau, 2016), la TA neuronale (*Neural Machine Translation*) a fait une arrivée en force sur le marché fin 2016 et a très rapidement supplanté l'approche dite « statistique ». L'avènement de cette nouvelle approche a réellement changé la donne dans le secteur de la traduction, venant bouleverser le modèle économique et les pratiques

professionnelles (Loock et Léchaugette, 2021, p. 206). Comme tant d'autres secteurs d'activité, les métiers de la traduction connaissent des transformations majeures (Alonso et Vieira, 2020⁴ ; Deneufbourg, 2021 ; Doherty, 2016 ; Loock, 2019) avec l'émergence de nouvelles pratiques, de nouvelles tâches (localisation de jeux vidéo, pré- et post-édition, transcréation, etc.) issues du développement des nouvelles technologies de la traduction. Les outils de traduction assistée par ordinateur (TAO) et de TA font désormais partie intégrante du poste de travail des traductrices (Alonso et Vieira, 2020, p. 391 ; Martikainen, 2022a) et plus que jamais, ces professionnelles sont amenées à acquérir de nouvelles compétences afin de s'adapter aux mutations de la profession et à la révolution technologique qui se déroule sous leurs yeux (Fontenelle, 2019).

Pourtant, le prétendu « miracle de la traduction automatique neuronale » (Dirand et Rossi, 2019, p. 71) ne fait pas l'unanimité. Depuis son introduction, l'approche neuronale suscite une abondance de critiques élogieuses et de prévisions alarmistes sur le remplacement de l'humain par la machine (O'Brien et Ehrensberger-Dow, 2020)⁵ qui ne manquent pas d'alimenter le débat polarisé sur le devenir de la profession. En effet, beaucoup y voient une technologie révolutionnaire, promise à un avenir radieux et vont jusqu'à prétendre qu'elle est capable d'atteindre la parité humaine (*human parity*) (Hassan et al., 2018⁶ ; Shoshan, 2018), voire une « qualité surhumaine » (*super human performances*) (Barrault et al., 2019)⁷. Tandis que d'autres déplorent une conception agonistique qui oppose l'humain à la machine (Dirand et Rossi, 2019, p. 64) et dénoncent un « battage médiatique autour de la TA » (Hansen et al., 2022, p. 2 ; Moorkens, 2018) en accusant les médias grand public et spécialisés, ainsi que les développeuses de surestimer les performances des moteurs de TA (Loock, 2019, p. 59) et de dévaloriser le rôle de l'humain :

[P]ublished results based on automatic evaluation and crowd work are almost always reported cursorily, devaluing human translation and creating an unrealistic and uncritical perception of MT among the general public, including translation clients. This perception increases the likelihood of MT being introduced into professional workflows. (Moorkens, 2022a, p. 129)

⁴ « As generic and translation-specific technologies are incorporated into translation production chains, the role played by translators is undergoing a series of shifts » (Alonso et Vieira, 2020, p. 391).

⁵ « There is an increasing supposition that MT is now good enough and can replace the work of professional translators » (O'Brien et Ehrensberger-Dow, 2020, p. 146).

⁶ « We find that our latest neural machine translation system has reached a new state-of-the-art, and that the translation quality is at human parity when compared to professional human translations » (Hassan et al., 2018, p. 1).

⁷ Voir l'article « On "Human Parity" and "Super Human Performance" in Machine Translation Evaluation » (Poibeau, 2022).

Vient s'ajouter à cela la problématique des biais dans les méthodes d'évaluation de ces performances (voir notamment Läubli et al., 2020). Or, selon la grande majorité des spécialistes en la matière et malgré des progrès impressionnants, les systèmes actuels de TA sont loin d'être parfaits (Koehn, 2020) et loin d'égaliser ou de surpasser la qualité en TH (Grass, 2022 ; Loock et Léchaugette, 2021 ; Poibeau, 2022) :

Claims that fully automatic high-quality machine translation would be reached in, say, just five years have been repeatedly proven wrong. But over the last decade or so, machine translation has actually become useful [...] It is an admission that machine translation technology today and in the foreseeable future is not perfect and will make mistakes. (Koehn, 2020, pp. 41-42)

Toujours est-il que l'utilisation de la TA à des fins très diverses (O'Brien et Ehrensberger-Dow, 2020) est de plus en plus répandue dans le grand public, mais aussi dans le monde professionnel. L'année 2018 marque d'ailleurs le dépassement d'un seuil symbolique en Europe : Plus des 50 % des entreprises de services linguistiques européennes et des traductrices professionnelles ont déclaré avoir recours à la TA (*Language Industry Survey*, 2018). Cette proportion, en évolution constante, confirme la montée en force de la TA et l'intérêt croissant que cette technologie suscite auprès des différentes protagonistes du marché. Dans un tel contexte, nous avons voulu étudier l'une des pratiques dont l'essor dans le secteur des services langagiers est directement lié aux progrès en TA (Poirier, 2022)⁸ : la post-édition. Il convient d'emblée de préciser que nous entendons par « post-édition » le processus de révision par un humain d'un texte traduit automatiquement par une machine.

1.1.2 Une pression accrue sur le secteur de la traduction

En dépit des « discours affirmant l'inutilité [voire la disparition] de la traduction professionnelle [qui] se font de nouveau entendre chez les non-spécialistes » (Loock, 2020a, p. 270), les traductrices spécialisées sont aujourd'hui plus indispensables que jamais. En témoigne la croissance continue du marché mondial de la traduction (Hickey, 2023). Et pour Robert (2013), il ne fait aucun doute que les progrès réalisés en TA et l'émergence de la PE sont intrinsèquement liés à cette croissance de la demande en traduction. Depuis quelques dizaines d'années, le contexte de mondialisation néolibérale entraîne une augmentation exponentielle de la demande en traduction, principalement en raison de la multiplication des échanges commerciaux, du développement des technologies de l'information et de la communication avec l'avènement d'Internet et la multiplication des sites Web, ainsi que de l'augmentation du nombre d'organismes internationaux et de forums de coopération

⁸ « Cette progression fulgurante de la PE comme révision de textes traduits par la machine est indissociable des progrès accomplis par la traduction automatique (TA) en matière de qualité » (Poirier, 2022, p. 2).

internationale. L’obligation de traduction imposée par certaines lois (la loi Toubon en France ; la directive 2006/42/CE du Parlement européen du 17 mai 2006 qui oblige les fabricantes à fournir une documentation technique en plusieurs langues pour les machines destinées au marché européen, etc.) a également stimulé cette demande en traduction. Cette situation, conjuguée à l’arrivée de la TA neuronale, amène les différentes protagonistes du secteur des services langagiers à se conformer à la maxime « le temps, c’est de l’argent », soumettant les traductrices à des impératifs de rentabilité et de flexibilité sans cesse grandissants (Guerberof Arenas, 2008 ; Moorkens, 2017) et toujours plus rudes ; délais de livraison de plus en plus courts, forte pression sur les tarifs exercés, etc.

The recent advances in artificial intelligence, natural language processing and ready access to freely available online translation tools have not gone unnoticed by the general public, resulting in increased pressure on language service providers to lower their prices and on translation training institutions to justify their existence. (O’Brien et Ehrensberger-Dow, 2020, p. 148)

D’ailleurs, l’intérêt premier de la PE n’est-il pas un gain de temps et, donc, de productivité par rapport à la TH (Carl et al., 2011⁹ ; De Faria Pires, 2020 ; Koehn, 2020¹⁰) pour répondre aux besoins du marché (Robert, 2013) ? « Much of the interest, particularly on the industry side, focuses on the potential for saving time (and thereby money) through the use of PEMT [post-editing of machine translation] » (Koponen, 2016, p. 21). En bref, on attend de la post-éditrice de traduire plus avec moins : « do more, of lower quality, for much less pay » (O’Brien, 2010, p. 23).

Toujours est-il que, d’après les résultats de la nouvelle enquête 2022 sur les pratiques professionnelles des métiers de la traduction menée par la commission Statistique de la Société Française des Traducteurs (SFT) auprès de 1204 traductrices, quatre personnes interrogées sur dix proposent la PE dans leurs prestations. Et pour ces prestataires, cette tâche représente en moyenne 20 % de leur activité (SFT, 2022) ; une tendance qui n’est d’ailleurs pas près de s’inverser (Deneufbourg, 2019 ; *Language Industry Survey*, 2019).

Dans cette conjoncture économique, le recours exponentiel à la TA neuronale (Volkart et Bouillon, 2022, p. 78) et l’augmentation constante de la demande en post-édition (Guerberof Arenas et Moorkens, 2019) sur l’ensemble du secteur soulèvent inévitablement de nouveaux enjeux et questionnements concernant la pratique de la traduction professionnelle, mais aussi concernant l’enseignement de la traduction, qui constitue la toile de fond de notre thèse. Nous

⁹ « [O]ne of the most obvious reasons for engaging in post-editing is the desire to save time » (Carl et al., 2011, p. 137).

¹⁰ « [P]roductivity increase is often touted as a metric by vendors of machine translation solutions who promise 50 %, 100 %, or even 200 % productivity increase with their technology » (Koehn, 2020, p. 44).

verrons que la nécessité de repenser les programmes de formation destinés aux futures traductrices en tenant compte des évolutions technologiques de la profession, ressort clairement de la littérature scientifique (voir notamment De Faria Pires, 2018 ; Guerberof Arenas et Moorkens, 2019 ; Loock et Léchauguette, 2021 ; Nitzke et Hansen-Schirra, 2021 ; Volkart et al., 2022). D'ailleurs, depuis 2017, l'interaction avec la TA figure dans le référentiel de compétences du réseau européen des masters en traduction (EMT)¹¹ et la plupart des formations en traduction comprennent désormais un volet consacré à la TA et à la PE (Loock et Léchauguette, 2021, p. 208).

Ainsi, le choix du sujet de cette thèse tient largement au bouleversement technologique que connaît le monde de la traduction, et plus particulièrement au changement de paradigme en TA qui conduit, entre autres, à la nécessité de repenser la formation initiale des traductrices afin de les préparer au mieux aux réalités de la pratique professionnelle. En outre, en tant que doctorante et assistante au sein de la filière traduction-interprétation de l'Université de Liège, nous avons spontanément choisi de nous intéresser à la post-édition en contexte d'apprentissage.

1.2 Objectifs et méthodologie

Dans cette thèse, nous visons à élargir le champ des connaissances relatives à la post-édition en poursuivant trois objectifs principaux.

Premièrement, l'objectif majeur de cette étude est de comparer les produits de deux processus de traduction en contexte d'apprentissage : la traduction humaine et la post-édition. Bien que la technologie qui sous-tend la TA ait considérablement évolué et que les sorties des moteurs de TA soient de plus en plus exploitables (Deneufbourg, 2021) et exploitées, nous considérons qu'il est effectivement légitime de s'interroger sur la qualité du produit fini après PE. Pour atteindre cet objectif, nous avons adopté une démarche expérimentale en menant deux expériences contrôlées (prétest et test) avec des étudiantes en traduction pour la combinaison de langues anglais-français. Notre étude est orientée produit puisque nous souhaitons explorer principalement les effets de la PE en mettant en lumière les potentielles différences en matière de qualité de produit fini entre la TH et la PE.

Deuxièmement, nous cherchons à rendre compte de l'opinion que portent aujourd'hui les étudiantes en traduction sur la (PE de) TA étant donné que celle-ci influencerait directement

¹¹ « Le réseau des programmes de master européen en traduction (*European Master's in Translation, EMT*) a publié son référentiel de compétences de traducteur et de traduction pour la première fois en 2009, et il l'a revu en profondeur en 2017. Ce référentiel est devenu l'une des principales références pour la formation des traducteurs dans l'Union européenne mais aussi au delà, tant dans les milieux universitaires que dans l'industrie » (Groupe d'experts EMT, 2022, p.2).

leur usage de la TA, ainsi que leurs compétences en PE (O’Brien, 2012b et 2022 ; Stasimioti et Sosoni, 2019). Pour ce faire, nous avons opté pour une méthode de recherche par sondage.

Troisièmement, ce projet s’inscrit dans une volonté de contribuer à une discussion plus large portant sur l’intégration des outils de TA à la formation initiale des traductrices et sur les défis majeurs que pose aujourd’hui la PE de TA aux (futurs) traductrices, en l’état actuel de cette technologie. Nous espérons ainsi faire valoir la nécessité de repenser la formation des étudiantes en faisant ressortir la valeur ajoutée de l’humain par rapport à la machine. Pour ce faire, de plus en plus de voix s’élèvent en faveur d’une *Machine Translation Literacy* (Bowker et Ciro, 2019), concept que R. Loock (2020a) propose de traduire en « utilisation raisonnée des outils de TA » (p. 270) et auquel il est primordial de sensibiliser aujourd’hui les étudiantes tant il est urgent que celles-ci soient conscientes des enjeux et des limites de cette technologie (*ibid.*). Néanmoins, nous tenons à préciser que nous ne prétendons nullement détenir les clés d’un enseignement idéal de la TA et de la PE. Aussi, nous n’avons pas inclus dans ce travail de proposition détaillée concernant la manière dont il conviendrait de faire évoluer la formation des apprenantes en traduction, mais proposons plutôt nos réflexions en la matière.

Cette thèse s’inscrit dans la continuité des travaux préexistants sur la PE en intégrant le recours à la nouvelle génération de TA : la traduction automatique neuronale. Dans la littérature spécialisée, nous avons relevé peu d’études comparatives portant sur les effets de la post-édition sur la qualité du produit fini pour la combinaison de langues anglais-français et pour des textes généraux, et d’autant moins en contexte d’apprentissage, à savoir lorsque cette tâche est effectuée par des étudiantes en traduction. Outre une analyse statistique de notre corpus, nous apportons un regard traductologique et linguistique sur les productions recueillies dans des conditions expérimentales.

1.2.1 Questions et sous-questions de recherche

Nous entendons répondre aux questions de recherche suivantes :

- Question A – Quels sont les effets de la post-édition (anglais-français) en contexte d’apprentissage sur la qualité d’un texte cible ?

Sous-questions – Prétest

A1. La post-édition (anglais-français) en contexte d’apprentissage permet-elle d’obtenir des textes de qualité comparable à la qualité en traduction humaine ?

A2. La post-édition (anglais-français) en contexte d’apprentissage présente-t-elle une distribution d’erreurs comparable à la distribution des erreurs en traduction humaine ?

A3. La qualité de textes post-édités (anglais-français) en contexte d'apprentissage varie-t-elle en fonction de la technologie de traduction automatique utilisée (TA statistique ou TA neuronale) ?

Sous-questions – Test

A4. La post-édition de traduction automatique neuronale (anglais-français) en contexte d'apprentissage permet-elle d'obtenir des textes de qualité comparable à la qualité en traduction humaine ?

A5. La post-édition de traduction automatique neuronale (anglais-français) en contexte d'apprentissage présente-t-elle une distribution d'erreurs comparable à la distribution des erreurs en traduction humaine ?

A6. La qualité de textes post-édités (anglais-français) en contexte d'apprentissage varie-t-elle en fonction du moteur de traduction automatique neuronale utilisé (*Google Traduction* ou *DeepL*) ?

A7. Comparativement à la traduction humaine, existe-t-il un effet nivelant en post-édition de traduction automatique neuronale sur la qualité des textes cibles ?

- Question B – Quelles sont les connaissances et les perceptions des étudiantes en traduction sur la traduction automatique et sur la post-édition ?

Sous-questions – Prétest et Test

B1. Quel usage font les étudiantes de la traduction automatique ?

B2. Quel est le rapport des étudiantes à la post-édition ?

B3. Comment les étudiantes considèrent-elles les performances actuelles en TA et comment imaginent-elles ces performances à l'avenir ?

B4. Quelle est la perception des étudiantes par rapport au processus de PE et à la qualité qui en résulte ?

B5. Les étudiantes jugent-elles l'intégration d'une formation à la (PE de) TA dans leur cursus nécessaire ?

B6. Les étudiantes considèrent-elles l'un des modes de traduction comme moins exigeant ?

B7. Les étudiantes montrent-elles une préférence pour l'un des modes de traduction ?

B8. Quel est le degré de satisfaction des étudiantes en PE et en TH ?

B9. L'opinion des étudiantes a-t-elle changé avec l'expérience ?

1.3 Plan de thèse

Cette thèse est organisée de la manière suivante :

Le **chapitre 2** est consacré à l'état de l'art. Nous proposons d'abord un survol historique de la traduction automatique, dans lequel nous retraçons l'évolution de la recherche en TA en présentant les trois principales approches : la TA à base de règles, la TA statistique et la TA neuronale, dont nous traitons plus en détail. Ensuite, nous dressons le panorama de la pratique au cœur de ce travail de thèse : la post-édition. La dernière section de ce chapitre porte sur les différents aspects de l'évaluation de la qualité en traduction (TH et TA).

Au **chapitre 3**, nous présentons la méthodologie de recherche qui nous a permis de comparer les produits de la traduction humaine et ceux de la post-édition en contexte d'apprentissage pour la combinaison de langues anglais-français. Nous y examinons les différents paramètres d'expérience et la méthode de constitution du corpus avant de décrire les stratégies de collecte et d'analyse des données.

Le **chapitre 4** porte sur notre première recherche expérimentale (prétest). Après avoir expliqué la raison de ce prétest et formulé les hypothèses de travail spécifiques à cette expérience, nous présentons les résultats en trois sous-sections : statistiques descriptives, statistiques inférentielles et résultats complémentaires (étude du processus de PE).

La structure du **chapitre 5** est analogue à celle du chapitre précédent. Nous posons tout d'abord les hypothèses spécifiques à cette deuxième expérience (test) et commentons les améliorations méthodologiques apportées par rapport au prétest. Ensuite, les résultats sont à nouveau présentés en trois sous-sections : statistiques descriptives, statistiques inférentielles et résultats complémentaires (étude du processus de PE). Nous concluons ce chapitre en tentant de dégager des hypothèses explicatives qui s'appliquent aux résultats à la fois du prétest et du test.

Au **chapitre 6**, nous cherchons à approfondir l'analyse des divergences entre textes traduits humainement et textes post-édités, et vérifions l'existence d'un *post-editeuse* dans nos corpus. Dans la première partie de ce chapitre, nous avons recours à trois mesures linguistiques automatiques pour analyser nos données : la richesse lexicale, la longueur moyenne des phrases et l'équivalence syntaxique. Dans la deuxième partie, nous menons une analyse qualitative comparative des erreurs en TH et en PE relevées dans nos corpus en mettant en évidence certains des effets tant positifs que négatifs de la (PE de) TA neuronale.

Le **chapitre 7** est consacré à l'état des lieux des connaissances et des perceptions qu'ont les étudiantes en traduction sur la TA et sur la PE. Nous présentons et interprétons les réponses recueillies dans les sondages que nous leur avons soumis.

Comme nous l'avons déjà évoqué, notre volonté est aussi de contribuer au débat sur l'actualisation des formations en traduction, à laquelle nous consacrons le **chapitre 8**. Dans un premier temps, nous passons en revue la littérature scientifique et avançons divers arguments pour nous permettre d'étayer notre position en faveur de l'intégration de la PE à la formation initiale en traduction. Dans un deuxième temps, nous proposons une réflexion sur l'adoption d'une « approche raisonnée des outils de TA » (Loock, 2020d, s.p.) et dressons une liste des défis majeurs que pose la (PE de) TA aux (futurs) traductrices, ainsi qu'aux formatrices. Dans un troisième et dernier temps, nous tentons de répondre aux diverses interrogations et inquiétudes qui planent sur l'avenir de la traduction professionnelle.

Enfin, au **chapitre 9** « Conclusions et perspectives », nous résumons nos travaux et présentons un récapitulatif des différents résultats. Nous étayons ensuite notre position concernant l'introduction de la TA et de la PE dans les programmes de traduction, ainsi que l'évolution de la traduction professionnelle. Nous tirons aussi les implications de nos résultats pour déterminer les principaux enjeux et atouts de la PE et pour définir une liste de défis majeurs à relever dans l'apprentissage de la PE. Nous terminons cette thèse en soulignant les limites de notre étude et en proposant plusieurs pistes de recherches ultérieures.

1.4 Publications scientifiques

Les deux expériences contrôlées qui sont au cœur de cette thèse et qui constituent l'essentiel des chapitres 2, 3, 4, 5 et 6, de même que les implications de nos résultats et notre réflexion sur l'enseignement de la PE ont fait l'objet de plusieurs publications dans diverses revues et ouvrages traductologiques. Les avantages et les limites de la PE en contexte académique ont été mis en avant dans Schumacher (2019) à partir des résultats issus de l'analyse qualitative du corpus de prétest. Certains des défis majeurs de l'apprentissage de la PE ont été discutés dans Schumacher (2020b). Les premiers résultats du prétest ont été présentés dans Schumacher (2020c), tandis que les résultats des statistiques inférentielles du prétest ont paru dans l'ouvrage collectif *Theory and practice of translation as a vehicle for knowledge transfer*, voir Schumacher et Sutera (2022). Enfin, les résultats des statistiques descriptives du test paraîtront prochainement dans le volume 68 (3) de *Meta* (Schumacher, 2023).

2 État de l'art

Ce chapitre consacré à l'état de l'art s'articule en trois parties. Dans un premier temps, nous proposons un survol historique de la traduction automatique centré sur les deux approches de ce travail : la TA statistique et la TA neuronale. Nous nous attardons sur le principe de fonctionnement de l'approche neuronale en analysant les différents concepts mathématiques et algorithmiques et en décrivant les différentes architectures d'apprentissage profond. Nous discutons également des principaux écueils des modèles de TA neuronale, en l'état actuel de la technologie. La deuxième partie de ce chapitre est consacrée à la post-édition. Après avoir défini cette pratique, nous relevons les différences et les similitudes entre la PE et la traduction, ainsi qu'entre la PE et la révision. Nous définissons ensuite les différents niveaux de qualité reconnus en PE et initiions une discussion sur les compétences indispensables en PE, que nous continuerons à développer à la fin de cette thèse (chapitre 8). Bien que le pan de la recherche en traductologie consacré à la PE soit relativement jeune, nous avons pu recenser bon nombre d'études sur le sujet. Nous examinons les études qui analysent le processus de PE et parcourons ensuite différents travaux consacrés au produit de la PE. Nous avons décidé de consacrer la troisième et dernière partie de ce chapitre à l'évaluation de la qualité en traduction, étant donné qu'il s'agit d'un volet essentiel de nos recherches expérimentales. En guise d'introduction, nous distinguons l'évaluation orientée produit de l'évaluation orientée processus et nous nous penchons plus spécifiquement sur l'approche typologique, car il s'agit de la stratégie d'évaluation qui a été choisie pour exploiter nos données empiriques. La section suivante porte sur l'évaluation de la qualité dans l'enseignement de la traduction étant donné que nos expériences se déroulent en contexte d'apprentissage. Nous nous intéressons au rôle de l'évaluation dans l'enseignement, à l'approche typologique en didactique de la traduction, à la distinction entre erreurs de langue et erreurs de traduction et aux principales divergences entre l'évaluation en contexte d'apprentissage et l'évaluation en contexte professionnel. Enfin, nous discutons des diverses méthodes d'évaluation de la TA (évaluation humaine, évaluation automatisée et évaluation axée sur la tâche).

2.1 La traduction automatique

2.1.1 Survol historique

L’intérêt pour la traduction automatique (TA) n’est pas nouveau et remonterait à la fin de la Seconde Guerre mondiale. Au cours des dernières décennies, la technologie a beaucoup évolué ; succédant aux approches linguistiques (traduction mot à mot et TA à base de règles), les approches statistiques opèrent désormais au niveau de la phrase entière, ce qui implique une meilleure prise en compte de la notion de contexte (Poibeau, 2019, p. 198) sur laquelle nous reviendrons. En 2016, l’avènement de la TA neuronale et particulièrement l’introduction du moteur de TA neuronale de *Google* (GNMT) (Wu et al., 2016) a marqué un véritable tournant dans le domaine de la TA. Par ailleurs, l’usage quotidien de cette technologie par des millions d’utilisatrices à travers le monde témoigne indéniablement des améliorations qualitatives par rapport aux anciennes approches : « Whatever the perception, there has been indeed good solid progress in the quality of machine translation systems that are available to real users » (Koehn, 2020, p. 30). Les performances des systèmes neuronaux sont devenues tellement satisfaisantes que cette technologie se profile aujourd’hui comme l’un des outils incontournables de la traductrice professionnelle comme nous le montrerons au chapitre 8.

« Enjeu scientifique, technique et économique sans précédent » (Léon, 2002, p. 2), la traduction automatique est définie par Bouillon et Clas (1993) dans leur ouvrage sur la traductique comme « l’application de l’informatique à la traduction des textes d’une langue naturelle de départ (ou langue source LS) dans une langue d’arrivée (ou langue cible LC) » (p. 15) ou plus récemment par Kenny (2022) comme « the automatic production of a target-language text on the basis of a source-language text » (p. 32). La TA recouvrirait deux extrêmes : la traduction automatique assistée par l’humain (TAAH) « qui ne peut se dérouler sans intervention humaine » (Bouillon et Clas, 1993, p. 15) et à l’autre extrême, la « traduction entièrement automatique de haute qualité (TEAHQ) » (*ibid.*), véritable fantasme inassouvi des développeuses. C’est sur le premier de ces deux extrêmes que portera cette thèse. Tout comme Bouillon et Clas (*ibid.*) le précisent, notre emploi du terme « TA » n’inclut pas les outils informatiques d’aide à la traduction.

Nous allons nous concentrer sur les trois approches que sont, chronologiquement :

- la TA à base de règles ;
- la TA statistique ;
- la TA neuronale.

Pour un historique plus détaillé sur l’évolution de la recherche en TA, se référer à Kenny (2022), Hutchins (2000) ou encore à Poibeau (2019).

2.1.1.1 La TA à base de règles

Les systèmes dits « à base de règles » (*rule-based machine translation – RBMT*) constituent ce qu’on appelle les architectures classiques en TA¹². Ces systèmes sont apparus dans les années 1950-1960 et, à l’époque, il s’agissait principalement d’apprendre à l’ordinateur la grammaire et les relations entre les langues source et cible au moyen de l’intégration de dictionnaires bilingues et de la programmation de règles formulées par des linguistes ou spécialistes en traduction.

Dans ce bref panorama historique, évoquons tout d’abord la ligne de recherche « pragmatique » (Poibeau, 2016). Les systèmes de 1^{re} génération conçus par ces équipes de recherche reposent sur une approche de traduction directe ou mot-à-mot (Poibeau, 2019). Combinant dictionnaires bilingues et règles de réordonnancement, ces moteurs ne permettaient pas de tenir suffisamment compte de la syntaxe et de la sémantique. En dépit de l’ajout progressif d’informations contextuelles, les performances de ces moteurs de traduction mot-à-mot demeurèrent médiocres, à tel point que les recherches ont été brutalement interrompues dans les années 1960 dans le monde anglo-saxon (Poibeau, 2016 ; 2019) après la parution du célèbre rapport de l’*Automatic Language Processing Advisory Committee* (ALPAC, 1966). Poibeau (2019) analyse les raisons de cet échec :

[C]ette approche a largement échoué parce que l’information contenue dans les textes est souvent parcellaire, parfois contradictoire d’un texte à l’autre, et surtout les connaissances (sur la langue ou sur le monde) sont par nature contextuelles, donc difficiles à représenter et à manipuler. (p. 197)

Après la publication de ce rapport, « la TA perd son ambition de technologie de pointe autonome » (Léon, 2015, p. 17) pour devenir une branche du TAL au profit du champ interdisciplinaire de la linguistique computationnelle (*ibid.*).

En parallèle, les partisans d’une ligne de recherche fondamentale développèrent des systèmes indirects (ou systèmes de 2^e génération) :

Plutôt que de traduire mot à mot, les systèmes indirects vont effectuer une analyse approfondie du texte source, pour en extraire une représentation qui servira de base pour la traduction. On distingue deux approches parmi les systèmes indirects : les systèmes de transfert et ceux par interlangue. (Gerlach, 2008, p. 5)

Sommairement, dans les systèmes de transfert, la traduction se fait par l’intermédiaire d’une représentation syntaxique du TS spécifique à la langue source (*ibid.*, p. 6). Ces systèmes se fondent sur une analyse complète des phrases à traduire et procèdent par l’application de

¹² *Classical MT*, voir Jurafsky et Martin (2008).

règles de transfert (Poibeau, 2016 ; 2019) qui « contiennent les informations sur les équivalences lexicales et structurelles pour une paire de langues » (Gerlach, 2008, p. 6). Alors que dans les systèmes par interlangue, « l'analyse du texte source est poussée encore plus loin, jusqu'à l'extraction du sens [...] [L]a traduction se fait par l'intermédiaire d'une représentation sémantique abstraite et idéalement indépendante des langues naturelles, appelée interlangue » (*ibid.*, p. 8). Notons que l'on retrouve le même type d'architecture (passage par une interlangue) dans les systèmes neuronaux que nous détaillons au point 2.1.2.

Apparaissent alors les premiers systèmes commerciaux, ainsi que les premières entreprises spécialisées en TA dont la société Systran qui, après des années de recherche et de développement, a pu proposer dès les années 1980 un moteur de traduction généraliste de qualité acceptable (Poibeau, 2016, p. 80). En 1997, des millions d'utilisatrices ont accès à cette technologie grâce au premier logiciel gratuit de TA en ligne, *Babel Fish*, proposé par *AltaVista* (Kenny, 2022).

Les systèmes à base de règles resteront l'approche dominante jusque dans les années 1980 et ont même survécu jusqu'à aujourd'hui. Poibeau (2019) précise que les systèmes actuels à base de règles n'utilisent pas ou que peu de statistiques et qu'ils sont encore employés dans des projets, comme le projet Apertium¹³, visant à promouvoir des langues rares ou en danger pour lesquelles il est souvent difficile, voire impossible, de disposer de grands corpus bilingues.

2.1.1.2 La révolution statistique

Après quelques tentatives de mise au point de systèmes mixtes dans lesquels « [d]es modules "intelligents" et des modules probabilistes viennent compléter les modules d'analyse linguistique traditionnels » (Léon, 2002, pp. 4-5), la recherche en TA connaît un renouveau « suscité par le succès des méthodes stochastiques dans le traitement du signal et la reconnaissance de la parole dans les années 1970 » (*ibid.*, p. 5). En effet, grâce aux progrès fulgurants de l'informatique (De Faria Pires, 2020) et à l'accès à d'énormes quantités de données textuelles sur support numérique (Poibeau, 2019), on assiste dès la fin des années 1980 à la révolution statistique avec l'émergence de l'approche dite de TA statistique (*statistical machine translation* – SMT). Dans cette approche fondée sur l'utilisation de corpus parallèles, les moteurs de TA sont entraînés sur cette masse de textes avec leur traduction afin de pouvoir « repérer les cooccurrences les plus fréquentes entre langue-source et langue-cible, et [...] en inférer des traductions potentielles au niveau des mots » (Poibeau, 2016, p. 81). Durant les années 1990 et 2000, le paradigme statistique « rose from obscurity to dominance and made machine translation a useful tool for many applications, from information gisting to increasing the productivity of professional translators » (Koehn, 2020, p. 39). Apparaissent

¹³ www.apertium.com

alors les premiers systèmes commerciaux de TA statistique développés par des sociétés telles que *Language Weaver* (qui sera rachetée par *SDL Trados*), *IBM*, *Microsoft* ou *Google* (*ibid.*, p. 38). Ce n’est toutefois qu’en l’an 2000 que la TA statistique s’imposa sur le marché (Koehn, 2010). Devenue le paradigme dominant, la TA statistique s’est vue progressivement intégrée aux solutions destinées aux professionnelles (par ex. au logiciel *SDL Trados*) avant d’être supplantée par le paradigme neuronal dès 2017.

Parmi les inconvénients de cette approche statistique, relevons qu’il n’y a pas d’analyse du contenu du texte au niveau sémantique, et que l’alignement de mots est une tâche « infiniment plus complexe que l’alignement de phrases » (Poibeau, 2019, p. 91). Une autre limite évidente, qui persiste encore avec la technologie actuelle, est que le processus d’entraînement de ces systèmes requiert inévitablement la disponibilité d’immenses corpus bilingues alignés : « [C]’est le critère quantitatif qui continue de prévaloir. De fait, il a été démontré que les performances des systèmes s’amélioraient régulièrement en fonction de la masse de données disponibles pour les mettre au point » (*ibid.*, p. 71). Par conséquent, il est difficile, voire impossible, d’entraîner ces moteurs pour qu’ils génèrent des traductions de qualité satisfaisante lorsqu’une des langues ou les deux langues à traiter sont dites « peu dotées », c’est-à-dire pour lesquelles peu de données sont disponibles sur Internet.

2.1.1.3 Le tournant neuronal

La traduction automatique neuronale (*Neural Machine Translation* – NMT) constitue la dernière approche de TA à ce jour. Il s’agit du fruit des progrès – fulgurants (Bénard et al., 2022) et même parfois spectaculaires (Poibeau, 2019) – accomplis dans le domaine de l’intelligence artificielle et plus précisément en *deep learning* ou « apprentissage profond ». Le principe du *deep learning* est connu depuis les années 1980, mais son utilisation ne s’est généralisée qu’à partir de 2012 pour la reconnaissance d’images et de 2014 pour la TA. Le *deep learning* est une branche du *machine learning* qui consiste en un ensemble d’algorithmes dont l’objectif est d’apprendre à représenter le monde. On dit de ces algorithmes qu’ils apprennent à modéliser des représentations de haut niveau (sons, images, textes). En 2014, Cho et al. (2014) et Sutskever et al. (2014) sont parvenus, avec succès, à appliquer ce type d’algorithmes à la traduction automatique ; dans ce cas, les algorithmes apprennent à représenter la langue source en caractéristiques complexes. La TA neuronale s’est alors imposée très rapidement dans le milieu scientifique : « Within a year or two, the entire research field of machine translation went neural » (Koehn, 2020, p. 40), déferlant sur le marché en 2016. Le paradigme neuronal est désormais utilisé par la très grande majorité des outils de TA (*DeepL*, *Google*, *Microsoft*, *Facebook* et bien d’autres, voir Figure 1), supplantant de ce fait la TA statistique : « À l’automne 2016, tous les acteurs majeurs du domaine (*Google*, *BingTranslator*, *Facebook*, *Systran*, etc.) sont passés d’un coup à l’apprentissage profond, avec dans la plupart des cas une amélioration des performances immédiatement visible » (Poibeau, 2019, p. 121).

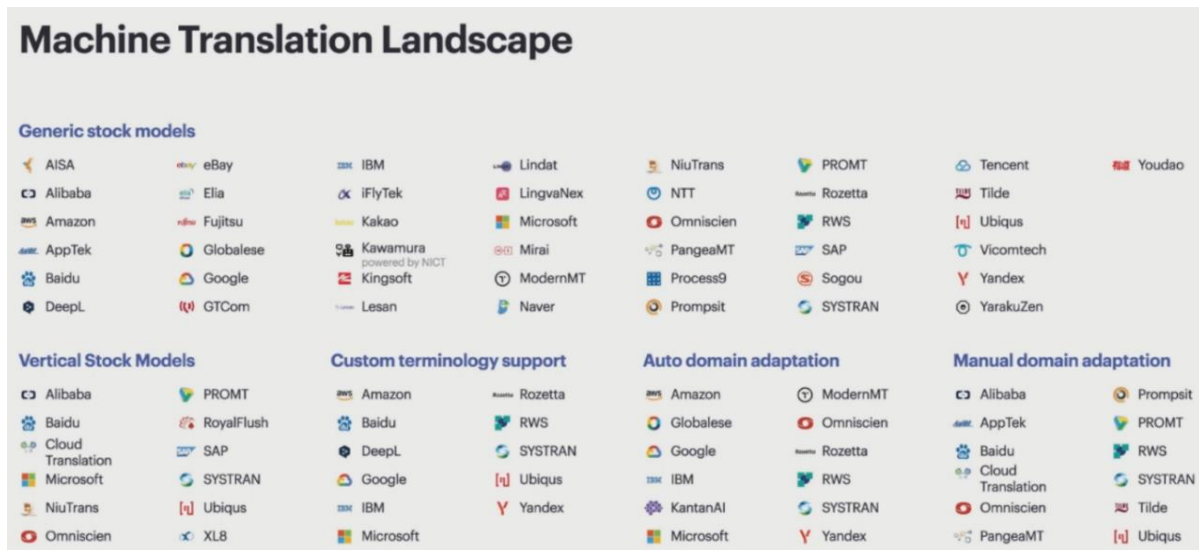


Figure 1 : *Machine Translation Landscape* (Savenkov et Lopez, 2022, p. 36)

2.1.2 La TA neuronale : principe de fonctionnement

2.1.2.1 Un prolongement de la TA statistique (De Faria Pires, 2020)

Si l’on a aujourd’hui abandonné l’adjectif qualificatif « statistique »¹⁴, il n’en demeure pas moins que la TA neuronale est une nouvelle génération de systèmes de TA fondés sur corpus (Forcada, 2017, p. 292) qui « reprend certains des éléments fondamentaux de la TA statistique [...] notamment quand il est question de réseaux de neurones » (Poibeau, 2019, p. 91 et p. 137).

NMT is still mainly based on knowledge inferred from large collections of parallel data, where one sentence in the source language corresponds to one sentence in the target language. From this point of view, NMT is a direct continuation of the previous segment-based approaches (Poibeau, 2022, p. 6019).

Fondamentalement, l’objectif de ces deux approches statistiques reste le même : « formuler la meilleure proposition de traduction automatique statistique possible pour un segment source donné » (De Faria Pires, 2020, p. 17). Et pour ce faire, les modèles statistiques procèdent par probabilité d’occurrence, « [i]n both NMT and SMT, a target sentence is a translation of a source sentence with a certain probability of likelihood [...] In both NMT and

¹⁴ « Despite the fact that often phrase-based SMT is labeled as ‘statistical’ and contrasted to ‘neural’ MT or NMT, we ought to stress that both approaches are in fact *statistical* » (Vanmassenhove et al., 2019, p. 224).

SMT, decoding (the same name is used) selects the most likely target sentence » (Forcada, 2017, pp. 300-301).

Data-driven statistical MT paradigms are concerned with (i) identifying the most probable target words, phrases, or sub-word units given a source-language input sentence and the preceding decoded information, via the translation model, and (ii) chaining those words, phrases or sub-word units in a way that maximizes the likelihood of the generated sentence with respect to the grammatical and stylistic properties of the target language, via the language model. (Vanmassenhove et al., 2019, p. 224)

D’ailleurs, tout comme pour les moteurs de TA statistique, les moteurs neuronaux nécessitent d’être entraînés sur d’énormes ensembles de données alignés : « It is trained on huge corpora of pairs of source-language segments (usually sentences) and their translations, that is, basically from huge translation memories containing hundreds of thousands or even millions of translation units » (Forcada, 2017, p. 292).

Ce qui distingue essentiellement la « TA par apprentissage profond » (Poibeau, 2019) de la TA statistique est que cette technologie est fondée sur une approche faisant appel à des réseaux de neurones¹⁵ : « Neural machine translation [...] is similar to the statistical machine translation technology that was the state of the art until very recently, but uses a completely different computational approach: neural networks » (*ibid.*, p. 292).

D’après Poibeau (2019), la grande nouveauté avec la TA par apprentissage profond est l’analyse contextuelle globale : « [À] tous les niveaux d’analyse, le sens des mots ou des groupes de mots est représenté en tenant compte de leur contexte, c’est-à-dire en tenant compte des éléments linguistiques présents autour d’eux » (p. 122). Par rapport aux systèmes précédents, les moteurs de TA neuronale permettent ainsi de traduire en tenant compte plus largement du contexte grâce à la méthode de « plongement lexical » que nous détaillons un peu plus loin.

Enfin, comme le souligne De Faria Pires (2020), le passage du paradigme statistique au paradigme neuronal s’accompagne de « l’apparition d’une pléthore de nouvelles notions mathématiques et algorithmiques jusque-là absentes des approches statistiques classiques » (p. 22) que nous allons tenter de décrypter dans cette section.

¹⁵ « A neural network is a machine learning technique that takes a number of inputs and predicts outputs » (Koehn, 2020, p. 67).

2.1.2.2 Une technologie neuronale d’inspiration biologique ?

On entend souvent dire que les algorithmes de *deep learning* sont conçus pour reproduire le fonctionnement du cerveau humain, ce qui explique l’origine des termes de « TA neuronale » et de « réseaux de neurones ». D’ailleurs, pour Dirand et Rossi (2019), ce « nouveau paradigme de l’intelligence artificielle suggère, par son nom même, que les systèmes neuronaux soient un analogue des intelligences humaines » (p. 71). En réalité, ces algorithmes ne s’apparentent que de loin au cerveau biologique en ce qu’ils sont au départ une tentative de modélisation mathématique du cerveau ; le fonctionnement des réseaux de neurones artificiels étant à l’origine inspiré du traitement de l’information dans notre cerveau, comme l’explique Forcada (2017) pour la TA neuronale :

Neural machine translation has, in spite of its name, only a very vague connection to neurons or to the way people’s brains (or translators’ brains) work. The name comes from the fact that the neural networks (which should properly be called artificial neural networks) on which NMT is based are composed of thousands of artificial units that resemble neurons in that their output or activation (that is, the degree to which they are excited or inhibited) depends on the stimuli they receive from other neurons and the strength of the connections along which these stimuli are passed. (p. 292)

Résumant les divergences entre neurones artificiels et neurones naturels, Koehn (2020) admet à cet égard ne pas être surpris par la réticence des scientifiques non seulement à établir « une analogie naïve avec les processus humains » (Poibeau, 2019, p. 198), mais aussi à utiliser le terme « réseaux de neurones », « [b]ut the term [neural networks] has stuck, even despite current rebranding efforts to *deep learning* » (Koehn, 2020, p. 31).

2.1.2.3 Le plongement lexical ou *word embedding*

L’approche neuronale qui sous-tend les moteurs de TA actuels consiste à recourir à un ensemble d’algorithmes d’apprentissage (réseau de neurones artificiels) pour apprendre à représenter la langue source en caractéristiques complexes (*high-level features*). En TA, ces représentations sont obtenues grâce à la méthode de « plongement lexical » (*word embedding*) et prennent la forme de vecteurs composés de nombres réels (voir Figure 2). La machine génère ces séries de vecteurs pour trouver la meilleure proposition, c’est-à-dire celle qui reproduit au mieux les valeurs d’apprentissage.

Le plongement permet la prise en compte de segments bien plus longs que ne pouvaient le faire les modèles statistiques, puisque chaque unité reçoit en fait une représentation vectorielle qui permet la prise en compte d’informations beaucoup plus riches que les probabilités d’occurrences de mots ou groupes de mots (Forcada 2017 : 293 et sq.). (cité dans Rossi, 2018, p. 102).

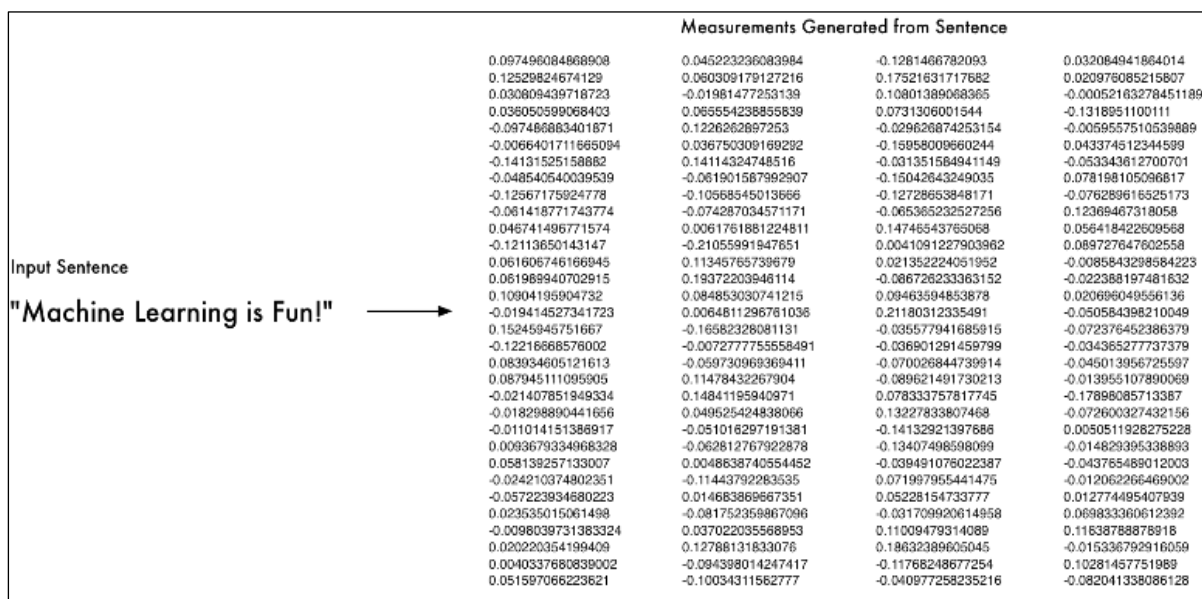


Figure 2 : Plongement lexical (Geitgey, 2016, s.p.)

Dans le cas où les systèmes sont entraînés pour traduire dans toutes les combinaisons de langues, ces séries de nombres réels sont en quelque sorte le « langage » interne propre aux algorithmes de l’intelligence artificielle. Il s’agit d’une sorte d’interlangue servant d’intermédiaire entre la langue source et la langue cible, non sans rappeler les systèmes linguistiques de 2^e génération (systèmes par interlangue). De plus, « [l]e fait que des informations diverses, riches et contextuelles, sont prises en compte simultanément grâce aux représentations sous forme de vecteurs signifie que les risques de propagation des erreurs sont limités » (Poibeau, 2019, p. 198).

2.1.2.4 Les architectures neuronales

2.1.2.4.1 Apprentissage profond

En TA neuronale, les algorithmes sont des algorithmes d’apprentissage dit « profond », car les modèles de réseaux de neurones sont composés de couches de traitement successives. Comme l’expliquent Pérez-Ortiz et al. (2022), les spécialistes ont découvert dès les années 1960 qu’il était possible en *deep learning* de résoudre des tâches plus complexes en ajoutant

plusieurs couches de neurones, appelées « couches cachées » (*hidden layers*) : « Instead of computing the output value directly from the input values, hidden layers are introduced. They are called hidden because we can observe inputs and outputs in training instances but not the mechanism that connects them » (Koehn, 2020, p. 68). Ces couches sont interconnectées via des nœuds de traitement (processing nodes) ou neurones (voir Figure 3). Ainsi, dans cette architecture multicouche, le traitement se fait couche par couche ; chaque couche traite l’information et passe son résultat à la couche suivante (Yvon, 2020).

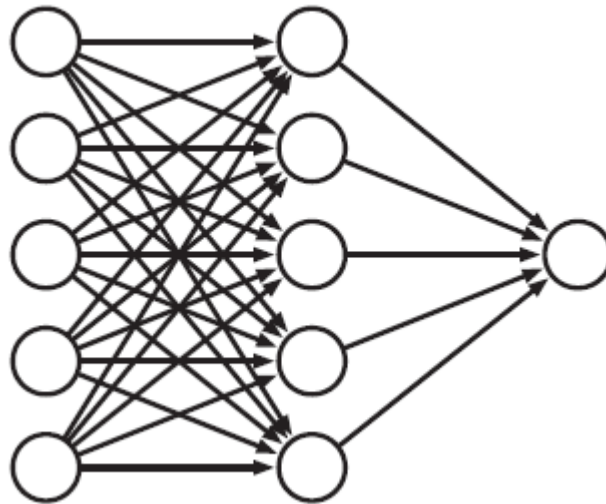


Figure 3 : Réseaux de neurones avec une couche cachée (Koehn, 2020, p. 69)

2.1.2.4.2 Architecture encodeur-décodeur

Tout comme en TA statistique, les modèles neuronaux sont constitués d’une architecture encodeur-décodeur, parfois appelée « séquence à séquence » (Sutskever et al., 2014) : « The core components of neural machine translation are the encoder, which takes input words and converts them into a sequence of contextualized representations, and the decoder, which generates an output sequence of words » (Koehn, 2020, p. 133). Autrement dit, « l’encodeur se charge d’extraire un vecteur d’une longueur fixe qui représente une phrase d’entrée de longueur variable. À partir de cette représentation, le décodeur génère une traduction en langue cible d’une longueur variable » (traduction libre de Cho et al., 2014, p. 103). Ce principe d’encodage et de décodage est illustré par la Figure 4. De nos jours, les architectures de TA sont devenues bien plus complexes, mais elles se fondent toujours sur cette architecture encodeur-décodeur (Poibeau, 2019).

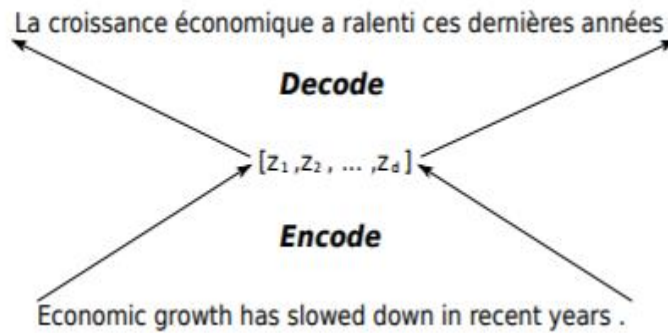


Figure 4 : Architecture encodeur-décodeur (Cho et al., 2014, p. 105)

2.1.2.4.3 Les réseaux de neurones récurrents (recurrent neural networks – RNN)

Améliorant les premières approches, Cho et al. (2014) et Sutskever et al. (2014) proposèrent les premiers modèles performants utilisant des réseaux de neurones récurrents : « The recurrent decoder–encoder architecture with attention [...] considered to be the bread-and-butter of NMT in 2017 » (Forcada, 2017, p. 299). La notion de « réseau récurrent » renvoie « à la capacité du moteur de se référer à tout moment aux informations qu’il a déjà traitées » (De Faria Pires, 2020, p. 26), comme illustré à la Figure 5 qui prend en exemple la traduction en espagnol du syntagme « Machine learning is fun ». Comme l’explique Bussy (2014), « la représentation apprise par une couche est alors utilisée comme entrée de la couche suivante » (p. 3). Concrètement, il s’agit de réutiliser la couche cachée qui vient de servir à prédire le mot w_n et de l’utiliser comme entrée supplémentaire pour prédire le mot w_{n+1} (Koehn, 2020), voir Figure 6. Pour une description plus exhaustive du fonctionnement des réseaux de neurones récurrents, on consultera Koehn (2020).

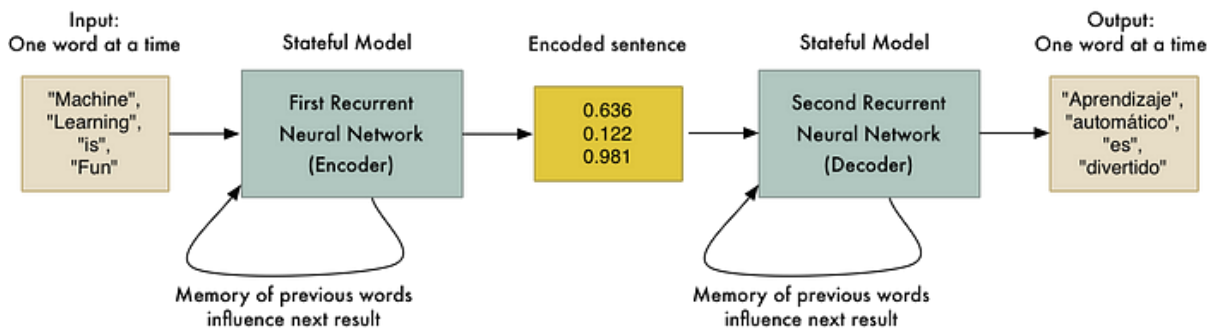


Figure 5 : RNN – « Machine learning is fun » (Geitgey, 2016, s.p.)

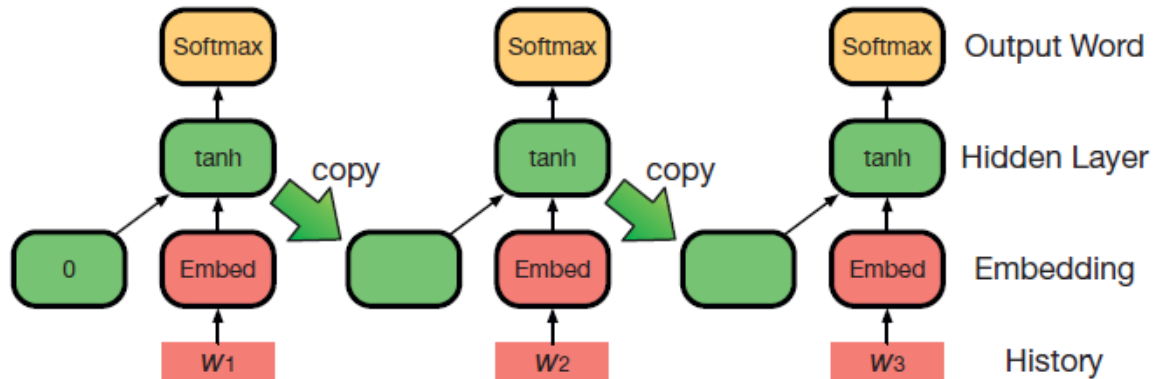


Figure 6 : Réseau de neurones récurrents (Koehn, 2020, p. 110)

Mécanisme d'attention

L'ajout d'un mécanisme d'attention (*attention mechanism*, voir Bahdanau et al., 2015) a permis d'améliorer les résultats de ces systèmes. Il s'agit d'un modèle d'alignement (Koehn, 2020)¹⁶ qui

frees the model from having to encode a whole source sentence into a fixed-length vector, and also lets the model focus only on information relevant to the generation of the next target word. [...] [T]he proposed approach of jointly learning to align and translate achieves significantly improved translation performance over the basic encoder–decoder approach. The improvement is more apparent with longer sentences, but can be observed with sentences of any length. (Bahdanau et al., 2015, p. 9 et p. 2)

Ainsi, à l'architecture classique encodeur-décodeur est venu rapidement s'ajouter le principe d'un mécanisme d'attention permettant de « guider l'analyse en se focalisant sur les portions de texte les plus pertinentes à un moment donné (le système de traduction procède sinon strictement suivant l'ordre des mots de la phrase) » (Poibeau, 2019, p. 136).

Penchons-nous à présent sur le nouveau type d'architecture de réseau de neurones qui est aussi l'architecture la plus populaire de nos jours : les réseaux de neurones *Transformers*.

¹⁶ « The seminal work by Bahdanau et al. (2015) adds an alignment model (so-called attention mechanism) to link generated output words to source words, which includes conditioning on the hidden state that produced the preceding target word. Source words are represented by the two hidden states of recurrent neural networks that process the source sentence left to right and right to left » (Koehn, 2020, p. 141).

2.1.2.4.4 Les réseaux de neurones *Transformers*

L’architecture *Transformer* (Vaswani et al., 2017) est une autre catégorie d’architecture neuronale axée entièrement sur les mécanismes d’attention : « [T]he Transformer [is] a model architecture eschewing recurrence and instead relying entirely on an attention mechanism to draw global dependencies between input and output. The Transformer [...] can reach a new state of the art in translation quality » (p. 2). Ces modèles de type *Transformer* ont effectivement permis d’améliorer significativement les performances des modèles de traitement automatique du langage (TAL) (Simoulin et Crabbé, 2021, p. 246), car il a notamment été montré « qu’ils pouvaient capturer des dépendances linguistiques à plus grande échelle [...] [ce qui est venu] rehausser d’un cran supplémentaire la qualité des traductions » (Hansen et al., 2022, p. 4). Il peut s’agir de modèles pré-entraînés qui sont soit spécifiques à une langue (Simoulin et Crabbé, 2021), soit multilingues dans de rares cas, tels que mBERT ou mT5 (Bender et al., 2021) et qui sont la plupart du temps accessibles et téléchargeables gratuitement en ligne à partir de bibliothèques. L’architecture *Transformer* reste la plus utilisée comme en témoignent Pérez-Ortiz et al. (2022) :

Transformers are currently the paradigm of choice if enough parallel corpora are available for training, because they require shorter training times and allow subtle quality improvements in comparison to recurrent neural networks, but the picture may change dramatically at any time. (p. 162)

C’est d’ailleurs sur cette architecture *Transformer* que se fondent les modèles de langue GPT ; « GPT » étant l’acronyme de *Generative Pre-trained Transformer*. À cet égard, Hansen et al. (2022) sont convaincues que ce sont ces réseaux de neurones *Transformers* qui

ont véritablement lancé les discussions sur la TA en traductologie et qui ont aussi attiré une très vive attention du public, puisqu’ils sont à la base des modèles de langues GPT-2 et GPT-3, par exemple, qui ont été abondamment diffusés dans la presse et sur les réseaux sociaux. (p. 4)

Un modèle de langue

En traitement automatique du langage, un modèle de langue est un « modèle statistique qui modélise la distribution de séquences de mots, plus généralement de séquences de symboles discrets (lettres, phonèmes, mots), dans une langue naturelle. Typiquement, un modèle de langue est capable de prédire le prochain mot d’un texte » (Datafranca, 2023). BERT et les modèles GPT (GPT-2 ; GPT-3 ; GPT-4) sont les exemples les plus connus de modèles de langue basés sur la catégorie d’architecture *Transformer*. Dans le cas du dernier-né d’OpenAI, GPT-4, il s’agit d’un modèle de langue multimodal, c’est-à-dire qu’il est capable de traiter du texte, mais aussi de l’image.

2.1.3 La TA neuronale : principaux écueils

Dans cette section, nous relevons plusieurs écueils en TA neuronale sans que cette liste ne se veuille exhaustive.

Une technologie gourmande

La technologie neuronale est onéreuse, extrêmement gourmande en données d’entraînement, en puissance de calcul et en énergie : « Neural MT is particularly resource intensive, requiring powerful GPUs (*Graphical Processing Units*) for training and large amounts of power » (Moorkens, 2022a, p. 133).

Comme Zhang et al. (2017) le soulignent, l’entraînement est véritablement le talon d’Achille des moteurs de TA neuronale :

Training efficiency is one of the main problems for Neural Machine Translation (NMT). Deep networks need for very large data as well as many training iterations to achieve state-of-the-art performance. This results in very high computation cost, slowing down research and industrialisation. (p. 271)

Rappelons que pour être efficace, le système de TA doit se nourrir d’énormes quantités de données numériques qui sont aujourd’hui disponibles à l’ère du Big Data, mais auxquelles relativement peu de monde a accès : « NMT systems are hard to train, even harder than SMT systems, which already required parallel corpora that are not usually available to individual translators or even small translation agencies » (Forcada, 2017, p. 302). Il convient d’entraîner les réseaux de neurones à partir de gigantesques corpus composés de segments en langue source et de leurs traductions qui représentent des centaines de milliers, voire des millions, d’unités de traduction (*ibid.*) et c’est toujours le cas pour les systèmes actuels : « To train an NMT system, one needs thousands or even millions of examples of source sentence–target sentence pairs » (Pérez-Ortiz et al., 2022, p. 162). Le célèbre adage que l’on attribue à R. Mercer – *There’s no data like more data* (litt. « plus il y a de données, mieux c’est ») et qui vaut au départ pour la TA statistique, s’applique maintenant à la TA neuronale ; plus le corpus d’entraînement est gigantesque, meilleure sera la qualité des résultats générés par la machine et plus grande sera la capacité du système à s’autoaméliorer.

Neural approaches in general, and large, Transformer LMs in particular, have rapidly overtaken the leaderboards on a wide variety of benchmarks and once again the adage “there’s no data like more data” seems to be true. It may seem like progress in the field, in fact, depends on the creation of ever larger language models (and research into how to deploy them to various ends). (Bender et al., 2021, p. 619)

Le mot d’ordre, c’est la quantité ! Pour reprendre les termes de B. Cassin (2016) dans son *Éloge de la traduction* : « la qualité est, pour de bon et sans ironie, une propriété émergente de la

quantité » (p. 143). Et cette tendance n’est manifestement pas près de s’arrêter, du moins tant qu’elle sera synonyme de gain de performance :

As increasingly large amounts of text are collected from the web in datasets [...], this trend of increasingly large LMs can be expected to continue as long as they correlate with an increase in performance. (Bender et al., 2021, p. 611)

Par conséquent, le fonctionnement des moteurs de TA neuronale nécessite une énorme puissance de calcul informatique : « NMT usually requires very large training corpora, typically as large as those used in good old SMT, and its training [...] is computationally very demanding » (Forcada, 2017, p. 295). Et c’est là que le bât blesse, puisque le traitement de ces énormes quantités de données requiert l’utilisation de GPU (processeurs graphiques) de plus en plus puissants et onéreux (Li et al., 2020) qui étaient conçus initialement pour le domaine des jeux vidéo et qui se sont révélés particulièrement performants pour les calculs des réseaux neuronaux (Kenny, 2017).

De plus, l’entraînement de ces systèmes est connu pour être très chronophage (Kenny, 2017 ; Li et al., 2020) puisque la durée d’entraînement peut régulièrement atteindre plusieurs jours, voire parfois plusieurs mois : « most NMT training resorts to using dedicated number-crunching hardware evolved from graphics processors, with typical training times ranging from days to months » (Forcada, 2017, p. 295). Toutefois, dans le cas des modèles de type *Transformer*, la durée d’entraînement serait beaucoup moins longue que celle des réseaux de neurones récurrents (Pérez-Ortiz et al., 2022 ; Vaswani et al., 2017).

Ce n’est pas tout puisque malgré les efforts d’amélioration de ces dernières années, la consommation en énergie de ces processeurs reste problématique :

[W]hile hardware becomes more powerful and costly to engineer and produce, optimization of power consumption and the potential to run massive amounts of parallel processes mean that the power required for training is dropping. Nonetheless, it remains the case that training an NMT system is costly and requires a good deal of power. (Moorkens, 2022a, p. 134)

Effectivement, l’entraînement d’un modèle – en particulier les modèles de type *Transformer* – est extrêmement énergivore comme le montre l’analogie établie par Strubell et al. (2019) : « [M]odels emit substantial carbon emissions; training BERT on GPU is roughly equivalent to a trans-American flight » (p. 3648). À cet égard, on observe ces dernières années l’émergence d’un courant de réflexion critique sur les liens entre traduction et écologie avec la mise en avant de pratiques de traduction écoresponsables, citons l’ouvrage de M. Cronin (2017a), *Eco-Translation – Translation and Ecology in the Age of the Anthropocene*, dans lequel il développe le concept d’« écotraduction », mais aussi de tout un débat autour de la durabilité écologique et sociale de l’intelligence artificielle et de ses nombreuses applications, avec entre autres van

Wynsberghe (2021) qui est à l’origine du concept de *Sustainable AI* (IA durable) : « Sustainable AI is a movement to foster change in the entire lifecycle of AI products (i.e. idea generation, training, re-tuning, implementation, governance) towards greater ecological integrity and social justice » (p. 217).

Des perroquets stochastiques

L’une des critiques fréquemment émises à l’égard de ces modèles par apprentissage profond est le fait qu’ils sont incapables de raisonner et d’avoir une compréhension profonde des textes, restant en quelque sorte en surface. Le système de TA « ne fait que retrouver des équivalents traductionnels à partir de grandes masses de données textuelles, sur une base essentiellement statistique, et le plus souvent sans tenir compte des liens logiques entre phrases » (Poibeau, 2019, p. 197). Pour Grass (2022), on ne peut parler d’intelligence puisque comme il n’y a pas d’analyse globale du sens, « tout au plus une adaptation du modèle au domaine considéré (droit, économie, etc.), [u]ne phrase peut [...] être parfaitement bien construite mais non pertinente du point de vue sociolectal, du domaine, du niveau de formalisme, etc. » (pp. 10-11).

D’ailleurs, les erreurs récurrentes en TA relèvent toujours plus de la sémantique (Forcada, 2017) que de la syntaxe ou de la grammaire. Si le transcodage d’une langue à une autre est assuré dans l’ensemble (Grass, 2022), peut-on réellement parler de traduction dès lors qu’« [i]l n’y a pas de représentation formelle du contenu sémantique du texte à traduire, pas d’inférence et encore moins de raisonnement » (Poibeau, 2019, p. 197) ?

Comparing human translation ability to a non-conscious algorithm is problematic, as contemporary MT systems merely attempt to replicate the patterns found in their training data. (Moorkens et al., 2018b, p. 256)

De plus, Poibeau (2019) explique que, par rapport aux systèmes précédents, les approches actuelles sont « très techniques, à base d’apprentissage artificiel, ce qui a relégué les préoccupations linguistiques au second plan » (p. 147). De ce fait, certains phénomènes linguistiques sont toujours très mal analysés par les modèles statistiques actuels, tels que les pronoms, les coordinations, etc. qui « exigent une analyse prenant en compte des séquences de textes relativement longues, s’étendant parfois au-delà de la phrase simple » (*ibid.*, p. 148). Ce manque de prise en compte d’un contexte plus large entraîne des problèmes de cohésion et de cohérence textuelles dont nous discuterons plus longuement aux chapitres 5 et 7. Précisons tout de même qu’il y a tout un pan important de la recherche en TA qui vise à améliorer ces systèmes grâce à une meilleure prise en compte du contexte (Poibeau, 2022), voir par exemple Lupo et al. (2022), Popel et al. (2020) et Popescu-Belis (2019).

Mais pour l’heure, les modèles neuronaux sont dépourvus de sens commun¹⁷ et ne modélisent toujours pas correctement les liens de causalité (Grass, 2022) : « [c]ertaines erreurs récurrentes sont imputables à [...] l’incapacité des systèmes à détecter des erreurs de sens relevant de la connaissance du monde et de rapports de causalité » (p. 17). Les modèles de langue sur lesquels se fondent les systèmes de TA actuels sont d’ailleurs souvent qualifiés de simples « perroquets stochastiques¹⁸ » :

Contrary to how it may seem when we observe its output, a LM [language model] is a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning: a stochastic parrot. (Bender et al., 2021, p. 617)

Pourtant, selon Poibeau (2019), ces modèles statistiques sont loin d’exclure la sémantique puisqu’ils permettent de déterminer « différentes granularités de sens (c’est-à-dire définir un nombre plus ou moins grand de sens pour un mot donné [...]), ce qui semble bien correspondre à la réalité linguistique, plus que le découpage fixe offert par les dictionnaires courants en tout cas » (p. 151). Pour une réflexion plus approfondie sur le lien entre la TA et la sémantique, se référer à l’article « Traduire sans comprendre ? La place de la sémantique en traduction automatique » (Poibeau, 2016).

Parmi les effets déplorables et les dangers de tels perroquets stochastiques, il a été démontré notamment par Bender et al. (2021) et par Vanmassenhove (2020) qu’ils reproduisent et exacerbent les différents biais présents dans les données d’entraînement. Mais là encore, nous rediscuterons de cette problématique au chapitre 6.

The size of data available on the web has enabled deep learning models to achieve high accuracy on specific benchmarks in NLP [Natural Language Processing] [...]. However [...] the training data has been shown to have problematic characteristics [...] resulting in [language] models that encode stereotypical and derogatory associations along gender, race, ethnicity, and disability status. (Bender et al., 2021, p. 613)

Les risques liés à l’usage de perroquets stochastiques et à cette soif insatiable de données conduisant à des modèles de langue toujours plus fournis sont exposés par Bender et al. (2021) dans leur article « On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? » :

¹⁷ « Capacité de bien juger, sans passion, des questions qui ne peuvent pas se résoudre par un raisonnement logique rigoureux » (Druide, 2022).

¹⁸ « Calcul des probabilités appliqué au traitement des données statistiques » (Le Petit Robert, s.d.).

[T]he tendency of training data ingested from the Internet to encode hegemonic worldviews, the tendency of LMs to amplify biases and other issues in the training data, and the tendency of researchers and other people to mistake LM-driven performance gains for actual natural language understanding — present real-world risks of harm, as these technologies are deployed. (p. 616)

Une technologie opaque

Le dernier écueil majeur que nous souhaitons aborder est celui du caractère opaque et insondable des réseaux de neurones, comme Bentivogli et al. le soulevaient déjà en 2016 :

[T]he NMT process is less transparent than previous paradigms. Indeed, it represents “a further step in the evolution from rule-based approaches that explicitly manipulate knowledge, to the statistical/data-driven framework, still comprehensible in its inner workings, to a sub-symbolic framework in which the translation process is totally opaque to the analysis. (p. 1)

Poibeau (2019) également admet que le manque de lisibilité des réseaux de neurones est un de leurs défauts majeurs : « il est difficile de comprendre ce que "fait" exactement un réseau de neurones, et sur quoi il se fonde pour produire une analyse par exemple » (p. 137). Il apparaît que les scientifiques ne parviennent pas à expliquer clairement ce qu’il se passe au niveau des « couches cachées qui, de ce point de vue, portent bien leur nom ! » (*ibid.*, p. 151) et sont incapables d’expliquer pourquoi cela fonctionne correctement ou pas. La TA neuronale est, en effet, tellement opaque qu’elle en deviendrait presque « magique » : « contemporary MT is marked by its extreme opacity, with Neural MT becoming so opaque in its inner workings that it is often described as “magic” » (Kenny, 2017, s.p.). Poibeau (2019) va jusqu’à qualifier la TA neuronale de « boîte noire » : on connaît les données en entrée et les données en sortie, mais on ne comprend pas le fonctionnement interne des réseaux de neurones. Et F. Yvon (2020) d’insister sur l’importance de parvenir à ouvrir cette boîte noire de manière à augmenter la qualité des sorties de TA ; le rêve absolu, selon lui, étant d’arriver à concevoir des outils de TA qui soient capables d’expliquer leur prise de décision.

2.2 La post-édition

2.2.1 Historique et définition

Sujet évoqué dès la fin des années 1950, la post-édition est une pratique qui serait apparue avec l’émergence de la TA elle-même (García, 2012 ; Massardo et van der Meer, 2017) : « From the outset it was assumed that MT would require assistance, with the concepts of pre-editing and postediting both already mentioned in pre-*Mechanical Translation*¹⁹ literature » (García, 2012, p. 294). Quelques décennies plus tard, avec l’amélioration des systèmes et l’amélioration de la qualité des sorties (Koehn, 2020 ; Poibeau, 2019), la post-édition fait son entrée dans le monde de la traduction professionnelle : « the term ‘post-editing’ entered the industrial world of translation when MT systems started to produce content in target languages that was deemed to be of good enough quality to be edited and improved by translators » (do Carmo et Moorkens, 2020, p. 36).

La PE est effectivement loin d’être un domaine de recherche nouveau puisque les premières études sur le sujet remontent aux années 1980. Parmi les pionnières, citons Muriel Vasconcellos, Anne-Marie Loffler-Laurian, Emma Wagner et Jeff Allen. Toutefois, c’est avec les recherches expérimentales de H. P. Krings, menées dans les années 1990, que cette discipline devient un réel objet de recherche universitaire (Brunette et O’Brien, 2011). L’étude de Krings, dont les résultats ont été traduits en anglais et publiés par Koby en 2001, explore la PE d’un point de vue psycholinguistique en comparant les processus cognitifs à l’œuvre par rapport à ceux en traduction humaine. Si l’intérêt pour cette pratique n’est pas récent, ce n’est que depuis que la TA est capable de fournir des sorties brutes exploitables en contexte professionnel que la PE suscite un intérêt accru en traductologie (De Faria Pires, 2020). Comme le signalent Alonso et Vieira (2020), la PE est devenue un service à part entière qui fait l’objet d’une norme internationale spécifique (ISO 18587) depuis 2017 dans laquelle le processus de post-édition est défini comme suit : « modifier et corriger un texte résultant d’une traduction automatique » (Organisation internationale de normalisation, 2017, s.p.).

Nombreuses sont les définitions de la PE que l’on retrouve dans la littérature scientifique, mais nous en avons retenu trois. La première est celle formulée par Gouadec (2007) : « Post-editing means checking, proof-reading and revising translations carried out by any kind of automatic translation engine » (p. 26). Citons également la définition, relativement large, qui en est donnée par García (2012), car elle apporte une précision – quoique subjective – sur la qualité du produit fini : « [Post-editing] is the process whereby humans amend machine-generated translation output to achieve an acceptable final product » (p. 309). Bien que Garcia (*ibid.*) reconnaisse que cette définition soit simple en apparence, il considère qu’il est nettement plus

¹⁹ Ancienne appellation de la TA, aujourd’hui remplacée par *Machine Translation*.

complexe de déterminer les compétences requises pour être *post-éditrice*, ainsi que de juger de l'acceptabilité de la qualité finale en PE, mais nous y reviendrons. La troisième définition est celle d'O'Brien (2022) qui précise que cette tâche bilingue est confiée essentiellement à une *traductrice professionnelle* alors appelée « *post-éditrice* » : « Post-editing is a bilingual language processing task. It is typically, though not exclusively, undertaken by experienced professional translators. When a person is engaged in this task, they are usually referred to as a *post-editor* » (p. 106).

2.2.2 Post-édition et traduction

Même si cela semble évident, nous tenons à rappeler que la PE est un processus qui diffère de la traduction « humaine » ; « manuelle » ou « conventionnelle » : « post-editing skills are different from translation skills and we cannot assume that a qualified translator will be a successful post-editor » (O'Brien, 2002, p. 100). Les objectifs en matière de qualité finale peuvent diverger en fonction de la tâche comme le soulignait déjà Senez (1998) : « A translator will always strive to disguise the fact that the text has been translated. In the case of post-editing, it is enough for the text to conform to the basic rules of the target language, even if it closely follows the source text » (s.p.). Killman (2018) spécifie qu'il est important que les étudiantes perçoivent la différence entre une tâche de traduction et une tâche d'édition : « The fact that we are seeing excessive errors originating from the MT output and cases of unnecessary post-editing means that we need to be training students to be more effective post-editors and to differentiate between translation and editing tasks » (p. 138).

Alors qu'il ne ressort pas clairement de la littérature que la PE soit plus ou moins exigeante sur le plan cognitif que la TH, nous présumons que les facultés cognitives mobilisées dans ces deux pratiques ne sont pas identiques. Plusieurs études ont notamment montré que la présence de la TA dans le processus de traduction perturbe les stratégies cognitives habituelles en traduction (Čulo et al., 2014 ; Martikainen et Kübler, 2016). Cependant, des recherches supplémentaires dans ce domaine sont nécessaires.

Toutefois, ces deux pratiques ne sont pas totalement dissemblables : « PE requires different skills from HT [human translation] although there is some overlap » (Yamada, 2015, s.p.). En effet, plusieurs des qualités attendues en PE rejoignent celles essentielles en traduction spécialisée : une bonne maîtrise des langues source et cible, des compétences extralinguistiques – « a posteditor also needs to have general world knowledge as well as the relevant domain knowledge in order to properly understand the thematic subject of the source text » (Nitzke et Hansen-Schirra, 2021, p. 70) – ainsi que des compétences en recherche documentaire (*ibid.*). En PE, tout comme en TH, il est également indispensable d'être capable

de déterminer la finalité du texte cible, ainsi que les attentes du client et de public cible (O’Brien, 2002)²⁰.

2.2.3 Post-édition et révision

Alors que beaucoup s’accordent sur le fait que la post-édition est une activité étroitement apparentée à la révision, peut-on aller jusqu’à dire que ces deux activités sont identiques ? Les avis divergent sur ce point.

Si l’on s’en réfère à Robert (2013), la PE est perçue comme un exercice bien particulier qui « implique une gymnastique cérébrale bien spécifique » (p. 33) et qui n’est ni de la traduction, ni de la révision : « La post-édition ne consiste en effet pas à traduire puisqu’une pré-traduction sert de base au travail à effectuer, ni à réviser puisqu’il ne s’agit pas d’une traduction humaine, mais d’une pré-traduction » (*ibid.*). Toujours selon Robert, il est nécessaire de « développer des schémas mentaux inédits [...] [car] la post-édition est un exercice de haute voltige intellectuelle qui nous amène à disséquer et à analyser les textes différemment et à réagir et à fonctionner différemment » (*ibid.*).

Dans le cadre de ses travaux de recherche menés au sein du service de traduction de la Commission européenne (DGT), De Faria Pires (2020) postule que « la post-édition ne constitue pas une révision. [Il s’agit de] deux étapes distinctes du processus de traitement linguistique des textes [puisque] la révision constitue une étape ultérieure à [la PE] » (pp. 46-47). Quant à do Carmo et Moorkens (2020), ils considèrent que la PE s’apparente davantage à une forme de traduction qu’à une forme de révision :

[PE] should be considered a type of translation. Not only because PE represents an evolution of industrial translation processes and because it fulfils the same purpose as translation (to produce a good target text in an efficient and effective way), but also because it requires advanced writing and reading skills in two different languages. (p. 42)

De notre côté, nous établissons une distinction entre PE et révision bien plus ténue que celle qu’en font les autrices précitées, ainsi que Krings (2001) ou encore Loffler-Laurian (1984)²¹, puisque nous considérons ces deux pratiques comme similaires. À l’instar de Toudic et al. (2014), nous pensons qu’en PE, il s’agit bel et bien de réviser les traductions proposées par la

²⁰ « Where translation and post-editing do not differ is in the requirement for ascertaining the target audience’s needs. Translation training programmes train translators to examine the expectations of the source language audience and to compare these to the expectations of the target language audience and to translate accordingly. Post-editors need to perform this task too » (O’Brien, 2002, p. 101).

²¹ « Post-editing is not revision, nor correction, nor rewriting [...] [it] is a new way of considering a text, a new way of working on it, for a new aim » (Loffler-Laurian, 1984, p. 237).

machine, faisant « de l'opérateur humain [...] dans le meilleur des cas, un "post-éditeur", révisant des traductions réalisées par un moteur de traduction, afin d'obtenir un niveau de qualité donné » (pp. 1-2). Tout comme en révision, une grande partie du temps passé en PE est composée de pauses et non de mouvements sur le clavier (Koehn, 2009 ; Ortiz-Martínez et al., 2016), « [t]his suggests that PE is, like revision, more associated with reading than with writing » (do Carmo et Moorkens, 2020, p. 40). Cependant, nous reconnaissons que la tâche de révision et celle de PE ne répondent pas, à fortiori, aux mêmes besoins : premièrement, la post-éditrice est davantage amenée à traduire puisqu'elle n'est pas face à un produit fini, tandis qu'en principe, la réviseuse s'attache elle à corriger les dernières erreurs et coquilles. Deuxièmement, la post-éditrice est amenée, contrairement à la réviseuse, à devoir modifier notamment des erreurs typiquement « machine » c'est-à-dire des erreurs qu'aucun humain ne commettrait. Troisièmement, la réviseuse sait qu'on attend d'elle qu'elle livre une qualité irréprochable. Or, en PE, il est tout à fait possible d'envisager une révision de la production de la post-éditrice. Ainsi, notre avis se rapproche de celui de Gouadec (voir définition au point 2.2.1) et de celui de Čulo et al. (2014) pour qui la post-édition serait « a special type of translation revision » (p. 201).

Néanmoins, si la distinction entre traduction, révision et post-édition est claire et légitime pour certaines, nous pensons qu'elle a de moins en moins lieu d'être dans le contexte professionnel actuel. Nous rejoignons à cet égard O'Brien (2022) lorsqu'elle affirme que de nos jours, traduction, révision et post-édition s'entremêlent dans un même environnement, pour une même tâche confiée très souvent à une seule et même personne :

Within this working environment, the translator might be editing a fuzzy match for one sentence, translating the next sentence, and post-editing the one that follows. Does the translator change from being a revisor, to a translator, to a post-editor for each of these sentences? Not really! Essentially, the task is still all about translation and revision. The main difference lies in the technological input or support the translator is using from one moment to the next. (p. 106)

La PE n'est plus une tâche à part, elle fait désormais partie intégrante du flux de travail en traduction professionnelle :

[I]t may make more sense for the terms 'post-editor' and 'translator' to be merged or used interchangeably depending on the job. In other words, as the need for interacting with MT moves from the peripheries to the centre of the translation industry (see Koponen 2016), post-editing is no longer a separate task undertaken only in MT development circles. Rather, it is increasingly part of what translators do. (Vieira, 2019, p. 330)

2.2.4 La qualité en post-édition

Avant de définir les différents niveaux de PE, précisons que si les travaux traitant de l’influence de la qualité en TA sur la qualité en PE sont peu nombreux (De Faria Pires, 2020), certains ont permis de montrer une corrélation entre la qualité de la TA brute et la qualité d’un texte post-édité (Castilho et al., 2018²² ; De Faria Pires, 2020 ; Vieira, 2017²³).

2.2.4.1 Niveaux de post-édition

La divergence des attentes, que ce soit dans le monde de la recherche ou sur le marché de la traduction (Hu et Cadwell, 2016), donne lieu à différents niveaux attendus de post-édition (O’Brien, 2010 ; Saint André, 2015). Ainsi, il est courant de faire une distinction entre la « post-édition légère », la « post-édition intermédiaire » et la « post-édition complète ». Avant de définir ces niveaux, précisons que, comme ces trois termes peuvent prêter à confusion, ils ne servent pas à déterminer le niveau d’effort de PE que la post-éditrice doit fournir, mais bien la qualité finale du texte cible après PE :

A[...] challenge with the terms “light” and “full” post-editing is that often people misunderstand that these describe how much editing needs to be done, or in other words, how much effort the post-editor should put into the task, rather than what the final translation quality should be [...] it is better to focus on final translation quality requirements, than the vaguer definitions of light and full PE. (Nunziatini et Marg, 2020, pp. 311-312)

Post-édition « légère »

La post-édition dite « légère », « superficielle » (*light post-editing*) ou encore « rapide » (*rapid post-editing*) est définie par la norme ISO 18587 comme le « processus de post-édition [...] permettant d’obtenir un produit qui soit simplement compréhensible, sans tenter de parvenir à un produit comparable à celui obtenu par une traduction humaine » (Organisation internationale de normalisation, 2017, s.p.). Il s’agit essentiellement de rendre le sens du texte source et d’apporter le minimum de modifications à la TA brute : « the main idea of [light post-editing] is to perform a strictly minimal editing on texts in order to remove blatant and significant errors and therefore stylistic issues should not be considered » (Allen, 2003, p. 302).

²² « An important evaluation-oriented use of PE consists of the collection of information that can help to measure MT quality and diagnose translation problems. [...] One of the major concerns in the translation industry is how to quantify the amount of effort that is necessary for MT PE, based on the initial quality of the raw output » (Castilho et al., 2018, p. 29).

²³ « Relationship between editing operations and product quality depends on the level of quality of the raw MT output » (Vieira, 2017, p. 278).

Le groupe de réflexion TAUS²⁴ (*Translation Automation User Society*) définit ce niveau de qualité comme « acceptable » (*good enough*), nous reviendrons dans cette partie sur les niveaux de qualité reconnus par TAUS. La PE rapide permet de répondre à des demandes urgentes de traduction, il s’agit majoritairement de textes qui requièrent d’être traduits exclusivement à titre d’information : « Rapid post-editing (RPE) [...] came into existence to provide translations for urgent texts that are intended merely for information purposes and for restricted circulation, such as working papers for internal meetings, minutes of meetings, technical reports or annexes » (Allen, 2003, p. 302). Pour O’Brien (2010), cette pratique de la PE nécessite que les traductrices désapprennent la plupart de ce qu’elles ont appris en matière de qualité et de professionnalisme, puisque « [d]epending on various factors [...] certain post-editing scenarios (e.g. when light postediting is demanded) expect translators to “temporarily abandon their high standards” (Wagner, 1985, p. 2) » (cité dans Carl et al., 2015, p. 150). L’objectif en PE légère est que le contenu sémantique du texte source soit transmis dans le texte cible, tandis que le style, la terminologie, la grammaire et la syntaxe sont de moindre importance (Nitzke et Hansen-Schirra, 2021, p. 239).

Post-édition minimale ou intermédiaire

La post-édition dite « minimale » ou « intermédiaire » (*minimal post-editing*) est qualifiée par Allen (2003) de « fuzzy and wide-range category » (p. 304), entendez qu’il s’agit d’un niveau de qualité souple, non clairement défini, car il dépend souvent de la manière dont la post-éditrice interprète ce niveau minimal de modifications à apporter (*ibid.*) :

As the term makes it clear, there is reference to a ‘minimum’ number of changes that are applied in order to produce ‘acceptable’ quality. How exactly the ‘minimum’ number and the ‘acceptable’ quality should be defined, is a matter of debate, and a key issue in a production context as well as in a teaching context. (Depraetere, 2010, p. 1)

Il y a une dizaine d’années, Depraetere (*ibid.*) estimait que la plupart des tâches de PE en contexte professionnel étaient des PE de niveau intermédiaire. Il est difficile de dire ce qu’il est en réellement aujourd’hui. Contrairement à ce que l’on pourrait croire, les niveaux de

²⁴ « Fondée en 2005, la TAUS est un groupe de réflexion sur l’innovation et une plateforme partagée de services, de ressources et de recherche à l’intention du secteur mondial de la traduction » (Boers et Liu, 2013, p. 4). TAUS se définit comme « une organisation indépendante et neutre du secteur de la traduction » (Massardo et van der Meer, 2017, p. 34) ; « Nous permettons aux acteurs du secteur de se rassembler, à travers un programme d’organisation d’événements et la mise en place de groupes d’utilisateurs en ligne, ainsi qu’en partageant des connaissances, des indicateurs et des données qui contribuent à la mise sur pied d’un service de meilleure qualité. [...] Les connaissances et les données partagées aident les membres de TAUS dans l’élaboration de stratégies de localisation efficaces. Les indicateurs permettent des processus plus efficaces et l’uniformisation de l’évaluation de la qualité. Les données améliorent l’automatisation de la traduction » (*ibid.*).

qualité attendus en PE légère et en PE intermédiaire ne sont pas forcément évidents pour la traductrice, car cette dernière doit parvenir à faire le deuil de l’excellence : « The greatest challenge in light PE for most professional translators is leaving incorrect grammar and syntax unedited as we are usually used to creating high-quality translations » (Nitzke et Hansen-Schirra, 2021, p. 31). Nous discuterons plus amplement de ce point au chapitre 8.

Post-édition complète

Le dernier niveau de PE est dit « complet » (*full post-editing*) (Fiederer et O’Brien, 2009). En 2011 déjà, García soulignait que ce niveau de PE gagnait du terrain sur les mémoires de traduction : « as MT improves, full post-editing is now encroaching into areas that had been dealt with up to now by translation assisted with TM [translation memory] » (p. 218). Selon Screen (2019), il s’agit du type de PE le plus pertinent en contexte professionnel. Pour beaucoup, il est attendu que la qualité du produit fini en PE complète ne peut être différenciée d’une bonne TH. C’est notamment le cas de Wagner (1985), de Way (2013), mais aussi de Guerra Martínez (2003) : « Full post-editing : The attempt to convert raw machine translation output into a product indistinguishable from human translation » (p. 80). C’est également de cette manière que le concept de PE complète est défini dans la norme ISO 18587 : « processus de post-édition permettant d’obtenir un produit comparable à un produit obtenu par une traduction humaine » (Organisation internationale de normalisation, 2017). Dans sa définition, DePalma (2013) précise même que cette qualité ne peut être différenciée de la qualité d’une TH, que ce soit sur le plan linguistique ou sur le plan stylistique : « Full post-editing [...] is meant to produce human-quality output. The goal is to produce stylistically appropriate, linguistically correct output that is indistinguishable from what a good human translator can produce » (s.p.).

D’autres, pourtant, partagent une conception moins exigeante. C’est le cas de Nitzke et Hansen-Schirra (2021), qui s’appuient sur les consignes élaborées par TAUS pour formuler une définition plus souple de la qualité attendue en PE complète. Selon cette définition, il est acceptable que la qualité en PE complète ne soit pas aussi bonne que celle obtenue par une traductrice humaine : « Besides being comprehensible and accurate [...], stylistic quality is also important for full PE. However, it may still not be as good as it would be when translated from scratch [...] it is still important to remember to use as much of the raw MT output as possible and not to get too lost in the fine-tuning » (*ibid.*, pp. 31-32).

Dans le cadre des présentes recherches, nous avons demandé aux étudiantes de produire une post-édition complète et leur précisant que la qualité de leur production devait être comparable à la qualité attendue en TH.

Les trois niveaux de PE susmentionnés sont représentés sur la Figure 7.



Figure 7 : continuum PE (De Faria Pires, 2020, p. 56)

Toutefois, ce continuum ne fait pas consensus dans le milieu et certaines préfèrent une distinction purement binaire. Ainsi, dans son référentiel de consignes en PE, TAUS/CNGL (2010) fait la distinction entre deux niveaux de qualité attendue en PE : une qualité « acceptable » (*good enough quality*) et une qualité « comparable ou égale à une traduction humaine » (*quality similar or equal to human translation*). Elle définit ce niveau « acceptable » comme

compréhensible (vous pouvez comprendre le contenu principal du message) et exact (sa signification est identique à celle du texte source), mais sa forme stylistique est discutable. Le texte traduit peut sembler artificiel, la syntaxe peut s’avérer inhabituelle et la grammaire imparfaite, mais le message est transmis. (TAUS/CNGL, 2010, p. 4)

Martikainen et Kübler (2016) nous précisent que pour parvenir à ce niveau de qualité dit « acceptable », il convient de n’effectuer que des corrections essentielles :

[L]es consignes comme celles de TAUS (2016), largement adoptées par l’industrie, préconisent notamment d’exploiter au maximum le résultat brut de la traduction automatique et d’éviter d’effectuer des corrections d’ordre stylistique ou de modifier la structure des phrases dans le seul but d’améliorer la fluidité du texte. (p. 2)

Tandis que la qualité « comparable ou égale à une traduction humaine » peut se définir comme

compréhensible (un lecteur comprend parfaitement le contenu du message), exact[e] (la traduction a le même sens que le texte source) et stylistiquement correct[e], même si le style n’est pas nécessairement aussi bon que celui obtenu par un traducteur humain dont la langue maternelle est la langue cible. La syntaxe est normale, la grammaire et la ponctuation sont correctes. (TAUS/CNGL, 2010, p. 4)

Remarquons que ces distinctions de niveaux en PE sont à la fois simplistes et controversées : « the traditional dichotomy of 'light' versus 'heavy' (or 'full') post-editing is no longer sufficient » (Way, 2013, p. 2). Tout comme en traduction, les exigences de qualité en matière de PE sont fonction de bien des facteurs tels que la finalité du texte cible, la qualité de la sortie de TA ou encore des attentes et des consignes du client. Van Egdom, Vieira et al. (2018) soutiennent qu’à chaque situation de communication correspond un niveau de qualité et donc une approche différente en PE (p. 119).

D’aucunes seraient tentées de dire que la PE, quel que soit le niveau attendu, nous impose à jamais de tirer un trait sur « la notion d’une qualité optimale » (De Faria Pires, 2020, p. 55) or nous partageons pleinement l’avis de De Faria Pires selon lequel ces « différents niveaux possibles de post-édition [...] ne signifient pas une altération de la qualité dans l’absolu, mais plutôt une adéquation à l’usage qui sera fait de chaque document cible, visant à maximiser le rendement des post-éditeurs » (*ibid.*).

Post-édition brute et post-édition évoluée professionnelle

A.-M. Robert (2010) ajoute, quant à elle, des précisions en distinguant la « post-édition brute » de la « post-édition évoluée professionnelle », une distinction que reprendra Peraldi dans son analyse comparative menée en 2016. La post-édition brute, explique Robert (2010), « consiste à compléter, modifier, corriger, remanier, réviser et relire directement le texte produit à l’état brut par un moteur de traduction automatique » (p. 139). Tandis que la post-édition évoluée professionnelle consiste à « compléter, modifier, corriger, remanier, réviser et relire le texte produit par un processus qui associe diverses technologies de TA et de TAO [...] Il s’agit d’associer des outils et des savoir-faire, de combiner traduction automatique, mémoires de traduction et traduction humaine » (*ibid.*, pp. 139-140). Concrètement, en PE évoluée, la technologie de TA est intégrée aux mémoires de traduction et la proposition de la TA va apparaître dans le segment vide lorsqu’il n’y a pas de remontées de mémoire de traduction à 100 % (*full matches*) ou lors de remontées de mémoire partielles (*fuzzy matches*). Il est par ailleurs possible de définir le seuil à partir duquel la proposition de TA s’affiche à la place de la remontée de mémoire partielle. Il est également possible d’afficher la sortie de TA au côté des différentes remontées de mémoire pour un même segment et de sélectionner la proposition qui semble la plus adéquate à post-éditer. Une fois le segment validé, il est automatiquement enregistré dans la mémoire de traduction. Notons qu’à l’heure actuelle, il est relativement compliqué de proposer aux apprenantes en traduction une formation à la post-édition évoluée professionnelle, puisqu’il est nécessaire d’avoir accès à des logiciels très souvent payants et onéreux, ou d’établir un partenariat avec des agences de traduction par exemple. Cette situation est regrettable, d’autant plus que ces logiciels entraînés hybrides qui associent technologies de TAO et de TA sont généralement plus performants que les outils de traduction automatique non spécialisés accessibles gratuitement.

2.2.5 Les compétences en post-édition

Comme nous l'avons évoqué, la tâche de PE est principalement confiée à des traductrices, et ce, depuis le début. Prenons Brunette et O'Brien (2011)²⁵, elles estiment que les traductrices sont les mieux placées pour pratiquer la post-édition « entre autres, grâce aux compétences en transfert linguistique et aux ressources techniques dont [elles] disposent (McElhaney et Vasconcellos, 1988; Vasconcellos, 1987; Krings, 2001 : 12) » (cités dans Saint-André, 2015, p. 34). On retrouve également cette idée chez Popović (2018) : « Of course, post-editing is a resource-intensive task that has to be performed by qualified translators » (p. 154), ainsi que chez Nitzke et Hansen-Schirra (2021), « [i]t is essential that post-editors are skilled translators, because they need the same basic skill set » (pp. 69-70).

En 2010, De Almeida et O'Brien (2010) mettaient en avant trois compétences à valoriser chez la post-éditrice dans lesquelles elles font la distinction entre « changements indispensables » et « changements préférentiels ». La première de ces compétences est la capacité de la post-éditrice à effectuer les changements indispensables (*essential changes*), c'est-à-dire la capacité à repérer les points en TA brute qui doivent impérativement être modifiés et à les modifier adéquatement. La post-éditrice doit aussi être capable de travailler à un rythme qui permet d'atteindre le taux de productivité attendu en PE (environ 5000 mots par jour en moyenne)²⁶. Enfin, la post-éditrice doit pouvoir respecter les consignes de PE et limiter les changements préférentiels (*preferential changes*) au strict minimum.

Parmi les quelques référentiels de compétences en PE qui ont été proposés ces dernières années, nous en abordons deux : la norme ISO 18587 et le modèle de compétences de Nitzke et Hansen-Schirra (2021).

La norme ISO 18587 définit six compétences indispensables en post-édition qui rejoignent les compétences fondamentales en traduction : Outre des compétences en transfert linguistique et en recherche documentaire, ainsi que la maîtrise des langues source et cible, la post-éditrice se doit d'avoir des compétences culturelles pour intégrer le texte à la culture cible, des compétences techniques pour pouvoir utiliser les outils adéquats et enfin, des connaissances sur le domaine, sur la thématique du texte source sont également nécessaires. En matière de qualifications, la norme précise que la post-éditrice doit être détentricice d'un diplôme de traduction, ou de tout autre diplôme universitaire à condition d'avoir deux ans d'expérience

²⁵ « [C]e sont les traducteurs qui semblent les plus aptes à assurer des postéditions de qualité » (Brunette et O'Brien, 2011, p. 6).

²⁶ Cette valeur de productivité n'est pas souvent renseignée dans la littérature spécialisée et varie en fonction des sources. Dans une communication présentée à l'Université de Liège fin 2016, la présidente de la FIT Europe, N. Dalügge-Momme (2016), évoquait plutôt une productivité de 3500 mots/jour en PE par rapport à 2000 mots/jour en TH.

professionnelle à temps plein dans le domaine de la traduction ou de la PE ou encore d’avoir cinq ans d’expérience professionnelle à temps plein dans le domaine de la traduction ou de la PE.

Dans tous les cas, la PE est une tâche complexe qui requiert des compétences spécifiques (Nitzke et Hansen-Schirra, 2021). Dans leur ouvrage *A short guide to post-editing*, Nitzke et Hansen-Schirra (*ibid.*) proposent leur propre modèle de compétences de la post-éditrice (Figure 8). Le socle de ce modèle est constitué des compétences fondamentales de la traductrice auxquelles s’ajoutent des compétences spécifiques à la PE, ainsi que des compétences plus générales ou « soft skills ». Ce modèle comprend tout d’abord des compétences en gestion des erreurs (repérage, catégorisation et correction) : « [p]ost-editors must be very sensitive to error spotting and error classification, meaning that they have to be able to decide whether an error has to be corrected according to the guidelines, and of course the efficient correction of the errors » (*ibid.*, pp. 74-75), ainsi que des compétences technologiques (*MT engineering*) pour pouvoir entraîner et évaluer les systèmes de TA et enfin des compétences en évaluation des risques et en apport de conseils, ce qu’elles regroupent sous l’appellation « *consulting competence* » :

*[A] certain consulting competence is essential. Many clients might not be aware of the pros and cons of using machine translation systems for the translation process. Even clients that often work with translation professionals and language service providers might not be aware of risks and strategic processes as PE has only recently become established on the market. Hence, a post-editor has to inform the customer or project manager about potential risks as well as problem-solving strategies, respectively, i.e. the risk assessment should enable the post-editor to give advice on these questions, even if the post-editor is not fully responsible for the decisions regarding the overall project. (*ibid.*, p. 72)*

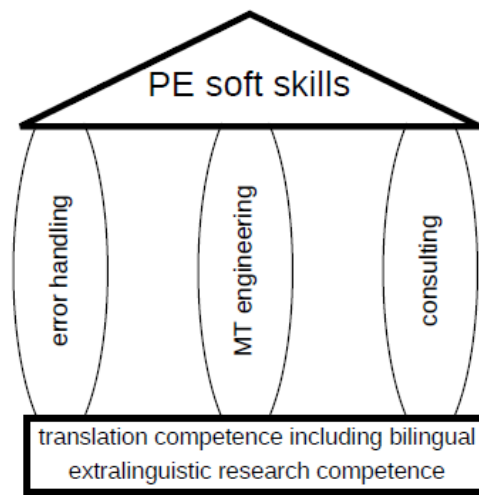


Figure 8 : Modèle de compétences en PE (Nitzke et Hansen-Schirra, 2021, p. 70)

Elles ajoutent que la post-éditrice doit également être en mesure de calculer ses prix en toute transparence et en toute connaissance de cause en fonction de la qualité de la TA et de l'effort de PE nécessaire (*ibid.*)²⁷.

In summary, the post-editor should know the translation market, including all aspects of MT and PE, and should be able to negotiate with the customer on an equal footing. The post-editor should be able to match the needs of the customer with the set-up and conditions of the PE task as well as with the resources available to be able to make an appropriate offer that calculates a realistic time and cost frame for the job. (ibid., p. 73)

Pour notre part, nous discuterons plus amplement au chapitre 8 de ce que nous envisageons comme compétences clés à développer chez les apprenantes avec l'objectif d'en faire des post-éditrices critiques et éclairées.

2.2.6 La recherche en post-édition

Nous l'avons évoqué dans la partie introductive, la recherche en TA et par extension en PE fait l'objet d'un regain d'intérêt dans le monde scientifique depuis plusieurs années déjà. Nous allons à présent dégager les principales tendances en matière d'étude du processus et du produit en post-édition qui ressortent de la littérature scientifique. Comme nous le verrons, ces études privilégient divers angles d'approche qui peuvent parfois être combinés.

2.2.6.1 Étude du processus

Plusieurs chercheuses se sont intéressées au processus même de PE et à la notion d'effort en PE.

D'après Krings (2001), cet effort de PE comprend trois dimensions distinctes, mais interdépendantes :

- L'effort temporel
- L'effort technique
- L'effort cognitif

²⁷ « In the field of PE, service competence means that the post-editor should be able to calculate *prices* competently, consciously, and transparently considering the quality of the MT output and the necessary PE effort, even though measuring and estimating PE effort is challenging (there has been a lot of research in this area, see e.g. Specia 2011; Moorkens et al. 2015; Schaeffer & Carl 2014) » (Nitzke et Hansen-Schirra, 2021, p. 72).

L’effort temporel permet de mesurer le temps requis pour post-éditer une TA brute. Il s’agit du type d’effort qui est le plus visible et le plus important pécuniairement parlant, c’est aussi celui qui est le plus facilement mesurable (Krings, 2001, p. 178).

L’effort technique permet de mesurer le nombre d’opérations d’édition, c’est-à-dire le nombre d’insertions, de suppressions et de réagencements nécessaires lors de la post-édition.

L’effort cognitif est l’effort le plus difficilement mesurable, car il n’est pas observable directement. Il représente « the type and extent of those cognitive processes that must be activated in order to remedy a given deficiency in a machine translation » (*ibid.*, p. 179).

L’interdépendance de ces trois indicateurs est bien détaillée dans la littérature. Dans leur étude, Moorkens et al. (2015) ont mis en évidence l’existence d’une forte corrélation entre l’effort temporel et l’effort technique en PE. Pour Koponen (2016), l’effort temporel est la combinaison des efforts technique et cognitif : « The temporal aspect of post-editing is formed by the technical operations (insertions, deletions and reordering of words) necessary for corrections, as well as the cognitive effort involved in detecting the errors and planning the corrections » (p. 139).

Toutefois, l’effort cognitif ne serait pas toujours directement proportionnel à l’effort technique (Koponen et al., 2012), « puisque le choix de conserver une partie de la sortie de TA peut être tout aussi complexe que celui qui consistera à la modifier ou la déplacer » (Rossi, 2018, p. 84). En outre, il ressort de plusieurs études (Carl et al., 2015 ; Koglin, 2015) que l’effort cognitif de PE serait influencé par la qualité de la TA brute. De plus, selon Wilms (1981), l’effort général de PE serait même inversement proportionnel à la qualité de la TA brute. Koglin (2015) partage cet avis lorsqu’il affirme que l’effort de PE est directement lié à la qualité de la TA brute : « the poorer the machine output quality, the more post-editing effort will be needed – provided that the text is expected to be fully post-edited » (p. 126). De même selon Lacruz (2017), plus la complexité textuelle est élevée ou plus la TA brute est de mauvaise qualité, plus cet effort cognitif sera élevé :

Cognitive effort refers to the mental effort involved in reading the texts, thinking about how to translate and how to correct mistranslations, selecting the desired product, and reflecting on the chosen solutions. A post-editor faced with more difficult texts or texts that have been less successfully machine translated must exert more cognitive effort, but there is no direct way to measure this effort. (p. 386)

Dans le secteur des services langagiers, le calcul de cet effort de PE est aujourd’hui devenu indispensable, non seulement pour déterminer le prix des services, mais aussi pour optimiser l’ensemble du processus de traduction :

One of the major concerns in the translation industry is how to quantify the amount of effort that is necessary for MT PE [...] The purpose of this is to pre-determine whether MT output PE would be time- and cost-effective when compared to translating the text from scratch, which in turn would guide pricing decisions, while also optimising the turnaround time; these are all crucial factors to be competitive in an increasingly pressurised translation market, in which budget-conscious clients regularly ask translators to meet tight deadlines, but are hardly ever prepared to compromise on the final quality of the translation product they receive. The possibility of reliably assessing the expected PE effort has become indispensable, given that it enables LSPs [language service provider] to optimise their translation processes. (Castilho et al., 2018, p. 29)

Néanmoins, parvenir à mesurer et à estimer cet effort de PE n'est pas une mince affaire (Nitzke et Hansen-Schirra, 2021, p. 72), notamment car « si les aspects temporels et techniques sont plus facilement calculables, il existe encore beaucoup de divergences sur une méthode d'évaluation objective et universelle de l'effort cognitif » (Imbres, 2018, p. 33).

L'analyse du processus même de post-édition peut se faire de diverses manières. À cette fin, les principales méthodes employées font appel à des outils de saisie de frappes, à l'oculométrie ou encore à la mesure de la durée de PE. Pour une comparaison de ces différentes méthodes en fonction du type d'effort visé, consulter Cumbreño et Aranberri (2021), et pour un aperçu des méthodes d'observation de l'effort cognitif en PE, consulter Lacruz (2017).

2.2.6.1.1 Méthodes de mesure de l'effort de PE

Outils de saisie de frappes

Les outils de saisie de frappes (*keystroke logging*) sont « des logiciels [qui] permettent d'enregistrer [...] l'utilisation du clavier, en minutant très précisément toutes les frappes, les mouvements du curseur et les pauses » (Lavault-Olléon et Carré, 2012, p. 4).

Oculométrie

L'oculométrie (*eye tracking*) est « une technique récente associant capteurs et caméras » (*ibid.*, p. 4) qui permet d'enregistrer les points de fixation du regard d'une personne, ainsi que les mouvements des yeux d'un point à un autre (Saldanha et O'Brien, 2014, p. 136). De manière générale, les chercheuses considèrent que « plus les pupilles sont dilatées et plus le regard reste fixé longtemps sur un même point, plus la difficulté rencontrée par le sujet est grande » (Sharmin et al. 2008 cités par Lavault-Olléon et Carré, 2012, p. 4), autrement dit, plus l'effort cognitif est important. La tendance générale qui se dégage des études faisant appel à l'oculométrie est qu'en PE, le regard se pose moins longtemps sur le TS qu'en TH et qu'il se concentre davantage sur le texte cible. C'est notamment ce qu'a pu observer Mesa-Lao (2014) : « Gaze time in the source area was considerably shorter in the case of post-editing where much of the gaze activity involved in the task took place in the target text

area » (p. 229). En outre, la PE n’impliquerait pas la même profondeur de lecture et de compréhension du texte source que la TH. En effet, la durée globale de fixation du texte cible serait plus élevée par rapport à la durée de fixation du TS : « Manual translation seems to imply a deeper understanding of the ST, requiring more effort and thus longer fixations » (Carl et al., 2011, p. 140). Même si certains travaux aboutissent à des résultats opposés (voir Cadwell et al., 2018 ; Moorkens et O’ Brien, 2015), plusieurs études laissent entendre que la charge cognitive en post-édition serait moins élevée qu’en traduction humaine (Carl et al., 2015 ; Daems, 2016 ; Jia et al., 2019²⁸ ; Koglin, 2015²⁹ ; Krings, 2001 ; Mesa-Lao, 2014 ; Screen, 2019). Prenons Krings (2001) dont les recherches expérimentales ont permis de constater qu’en PE, le processus de comparaison entre la TA et le texte source aurait un effet réducteur au niveau de la charge cognitive (*cognitive pull-down effect*) sur le texte source : « [Post-editing] forces the translator [to] deal with lower level text comprehension processes in the source text to a much greater extent than would be the case in a normal translation task » (*ibid.*, p. 360). Daems (2016) a montré dans son étude comparative que les étudiantes considèrent la PE comme un processus moins fastidieux que la TH, notamment en raison du fait qu’elles envisagent la traduction comme une tâche linguistique :

Professional translators experienced no obvious difference, whereas students seemed to consider post-editing the least tiring method of translation. This might be explained in part by the findings by Tirkkonen-Condit (1990) that non-professional participants treat translation as a linguistic task, and they rely mostly on dictionaries to solve problems. In a post-editing condition, lexical information is already provided by the MT output, which might reduce the need to look for additional information, and thus make the students experience the process as less tiring than regular human translation. (p. 114)

Par ailleurs, O’Brien et Ehrensberger-Dow (2020) sont d’avis qu’en prônant une approche raisonnée des outils de TA (*MT literacy*), dont nous reparlerons au chapitre 8, il serait possible de diminuer davantage cette charge cognitive en PE : « With respect to the task of post-editing MT, we assume that MT literacy can lower cognitive load by making it easier for the individual performing the task to evaluate the output and identify errors or problems » (O’Brien et Ehrensberger-Dow, 2020, p. 148).

De plus, il est intéressant de noter que, d’après les travaux de Carl et al. (2015), les post-éditrices ne travaillent pas toutes de la même manière : « [certaines] consultent systématiquement le texte source avant de procéder à la post-édition, alors que d’autres post-

²⁸ « Post-editing triggered significantly lower pause density and shorter pause duration than from-scratch translation for both text types, which indicates that post-editing is cognitively less demanding than from-scratch translation » (Jia et al., 2019, p. 79).

²⁹ « The preliminary findings of this study suggest that post-editing could be less effortful than manual translation when the following variables: pauses, task duration and insertions are taken into account » (Koglin, 2015, p. 136).

éditent le texte cible sans prêter attention au texte source, ne le faisant que lorsqu'un problème survient dans la TA brute (Carl et al. 2015 : 159) » (citées dans De Faria Pires, 2020, p. 74).

En outre, selon plusieurs études, il semblerait que « la post-édition mobilise [...] des efforts cognitifs divergents, voire supérieurs comparativement à une relecture classique (Schäfer, 2003 ; O'Brien, 2004 ; Lacruz et al., 2014) » (cités dans Peraldi, 2016, p. 69).

Mesure de la durée de PE

À la lecture de la littérature scientifique se dégage un quasi-consensus sur le gain de temps et, par conséquent, sur le gain de productivité conféré par l'activité de PE par rapport à la TH. Cette productivité et efficacité en PE dépendraient logiquement considérablement de la qualité de la TA brute : « As research has shown, the quality of the MT output is essential for productivity and hence cost efficiency of post-editing [...] High MT quality may increase productivity, low quality can be very tedious and may lead to frustration » (Carl et al., 2015, p. 149).

Bon nombre d'études ont ainsi mis au jour un gain de productivité entre les deux méthodes de traduction, que ce soit chez les traductrices professionnelles (Guerberof Arenas, 2008 ; Screen, 2019 ; Toral et al., 2018) comme chez les étudiantes en traduction (Daems, 2016 ; Depraetere et al., 2014 ; Jia et al., 2019 ; Martikainen et Mestivier, 2020 ; Samman, 2022).

Chez les traductrices professionnelles, Guerberof Arenas (2008) a constaté un gain de productivité qui varie entre 10 % et 18 %. Toral et al. (2018) ont étudié ce phénomène pour la traduction littéraire (roman de fantasy) anglais-catalan et il en ressort que la PE de TA (statistique comme neuronale) permettrait d'augmenter la productivité de la traductrice par rapport à la TH : +18 % pour la PE de TA statistique et +36 % pour la PE de TA neuronale. Quant à Screen (2019), il a comparé les résultats de 20 études pertinentes menées pour diverses paires de langues et dans des conditions divergentes. Sur ces 20 études, 17 ont révélé un gain de productivité entre la PE complète de TA statistique et la TH.

Chez les étudiantes, les travaux de Depraetere et al. (2014) font état d'une augmentation moyenne de productivité entre la PE de TA statistique et la TH de 21,5 %, tandis que chez Martikainen et Mestivier (2020), le gain mesuré entre la PE de TA neuronale et la TH s'élève entre 20 % et 30 %. Jia et al. (2019) ont mesuré une amélioration significative de productivité entre la PE de TA neuronale (*Google Traduction*) uniquement dans le cas de textes spécialisés et non pour les textes généraux. Enfin, Samman (2022) a récemment étudié le taux de productivité chez des étudiantes de bachelier en traduction et les résultats statistiques ont montré que la PE (anglais-arabe) permettait une augmentation non négligeable de la productivité par rapport à la TH.

Il est aisé de comprendre que ce gain de temps découle, entre autres, du fait que *la post-éditrice* dispose déjà d’un texte dactylographié en langue cible, contrairement à une *traductrice* qui part, *elle*, généralement, d’une page blanche. Mais la PE de TA permet également, par moments, de faire gagner du temps à la *traductrice* en lui fournissant directement des propositions de traduction ou des informations lexicales (Daems, 2016 ; Elming et al., 2014), et réduisant ainsi certaines vérifications terminologiques au strict nécessaire. Même constat chez Jia et al. (2019), où les étudiantes ont trouvé que la PE leur a permis d’économiser leurs efforts et de gagner du temps : « post-editing saved them time and effort due to not having to type the whole translation or not having to consult external resources to check all lexical information they did not know » (p. 78).

Par ailleurs, dans l’expérience pédagogique que nous avons menée en 2019 (Schumacher, 2020b), certaines étudiantes ont déclaré que la TA leur a permis d’utiliser des termes, souvent spécialisés, auxquels *elles* n’auraient pas pensé ou qui leur étaient inconnus et qui sont corrects dans cet usage et dans ce contexte ; la TA permettrait-elle d’élargir les connaissances de *la traductrice* ? En outre, toujours dans l’expérience menée en 2019 (*ibid.*), la liste déroulante de synonymes en contexte que propose *DeepL* lorsque l’on clique sur un terme ou syntagme a été fort appréciée par les étudiantes et semble également permettre d’économiser du temps.

Notons toutefois qu’en PE, ce gain de temps est à mettre en balance avec la « perte » de temps plus ou moins grande engendrée par les modifications qui doivent être apportées à la TA brute. En effet, même si la PE de TA semble à première vue faire gagner du temps par rapport à la traduction humaine, il se peut que cela ne soit pas toujours le cas. Il arrive qu’en raison de la piètre qualité de la sortie de la TA, *la traductrice* ait davantage intérêt à traduire certains passages ou même certains textes entiers « à partir de zéro », car la PE serait trop fastidieuse et chronophage.

2.2.6.2 D’abord la source ou d’abord la cible ?

Récemment, Volkart et al. (2022) ont mené une expérience avec 20 étudiantes du master en traduction de l’Université de Genève dans le but d’étudier l’influence de deux stratégies de post-édition (anglais-français et anglais-italien) ; à savoir lorsque le segment source est présenté par défaut à l’étudiante ou bien, au contraire, lorsque c’est le segment cible auquel l’étudiante a d’abord accès. Leur analyse statistique n’a pas révélé d’influence significative de l’une ou de l’autre stratégie, ni sur le taux d’erreurs post-éditées ni sur la durée de PE. Leurs résultats indiquent néanmoins que le taux de modifications optionnelles est légèrement plus élevé dans le cas du scénario où l’étudiant a d’abord accès au segment source, ce qui les conduit à supposer que plus l’étudiante accorde de l’attention au TS, plus sa production se détachera de la sortie de TA brute.

2.2.6.3 Étude du produit

Alors que le gain de temps a été largement démontré dans de nombreuses études récentes. Comme nous venons de le voir, il est bien légitime de s’interroger sur la qualité du produit fini. Qu’en est-il des effets de cette méthode de traduction sur la qualité du produit fini ?

Plusieurs travaux ont déjà été consacrés à cette question, les résultats étant toutefois souvent divergents. Nous présentons dans cette section les principales études qui se sont attachées à explorer le produit de la PE d’un point de vue qualitatif. Plusieurs scientifiques se sont, en effet, intéressées à la question de la qualité en PE. Certaines se sont attelées à évaluer la qualité de textes post-édités, tandis que d’autres se sont attachées à comparer le produit de la TH au produit de la PE à l’aide de diverses méthodes et dans divers contextes. De Faria Pires (2020) relève deux tendances majeures : « une évaluation de la qualité des contenus post-édités est envisageable au travers d’une simple addition des erreurs [...] comme l’ont notamment fait Guerberof Arenas (2014) et Yamada (2019), mais il est également possible d’aller plus loin et de tenir compte de la gravité perçue des phénomènes relevés, comme l’ont fait Daems (2016) et Vieira (2017) » (p. 91). Dans ses propres recherches menées auprès de traductrices professionnelles travaillant à la direction générale de la traduction (DGT) de la Commission européenne, De Faria Pires (2020) a également tenu compte de la gravité perçue des phénomènes relevés.

Remarquons que dans la plupart des cas, la méthodologie choisie, la paire de langues, l’échantillon sélectionné ou encore la taille du corpus constituent autant de facteurs qui diffèrent d’une étude à l’autre, ce qui rend les résultats obtenus souvent difficilement comparables. Malgré cette base de comparaison peu favorable, plusieurs travaux ont permis de nourrir et d’orienter notre réflexion. Les recherches que nous abordons ci-après ont été menées tantôt auprès d’étudiantes, tantôt auprès de professionnelles. Nous estimons qu’il est pertinent de ne pas nous limiter à citer des études menées auprès d’apprenantes, étant donné qu’il a été démontré que l’expérience en traduction ne joue pas toujours sur la qualité du produit fini :

Supporting the findings by Jääskeläinen and Tirkkonen-Condit (1991); Kiraly (1995), and Jääskeläinen (1996), we found that the more experienced translators are not necessarily the more successful translators, with students producing products of comparable overall quality, at least in this particular experiment. (Daems, 2016, p. 113)

Nous commençons par présenter une série de travaux qui attestent une qualité comparable, voire meilleure, entre le produit de la PE et de la TH avant d’examiner plusieurs études qui montrent les effets négatifs de la PE sur la qualité du texte cible.

2.2.6.3.1 Qualité comparable, voire meilleure

Certaines de ces études montrent que la PE, que ce soit par des étudiantes en traduction ou par des professionnelles de la traduction, permet d’obtenir des textes de qualité comparable à des TH. La PE n’entraînerait donc pas de baisse de qualité en comparaison à la TH (Daems, De Clercq et al., 2017 ; Depraetere et al., 2014 ; García, 2011 ; Jia et al., 2019 ; Screen, 2019³⁰).

Parmi les diverses études dont les résultats vont dans ce sens, citons d’abord les deux études plus anciennes menées par García (2010, 2011) avec des apprenantes en traduction. L’outil de TA employé est le Kit du traducteur de *Google* (*Google Translator Toolkit*), un environnement de TA statistique qui permettait de créer et de partager des traductions, d’intégrer des glossaires et des mémoires de traduction, et qui a été définitivement arrêté fin 2019. Dans les deux études, les productions des participantes ont été jugées par des évaluatrices externes selon les critères de qualité du NAATI, un organisme officiel d'accréditation des traductrices et interprètes en Australie. Il ressort des résultats de la 1^{re} étude (García, 2010) que, dans 59 % des cas, les segments post-édités (anglais-chinois) ont été préférés par les évaluatrices aux segments traduits humainement. Pour la deuxième étude (García, 2011), les résultats statistiques de cette étude font état d’une amélioration de la qualité en PE de TA par rapport à la TH, indépendamment du sens de traduction (anglais-chinois/chinois-anglais), de la difficulté du texte source et du niveau des apprenantes. En outre, cette recherche a permis de mettre en lumière que la PE était particulièrement bénéfique aux apprenantes faibles lorsqu’elles traduisent vers leur langue maternelle. Au vu de ces résultats, García (2010, 2011) a considéré que la PE pouvait constituer une solution alternative à la traduction conventionnelle.

Fiederer et O’Brien (2009) se sont également penchées sur la qualité en PE de TA (anglais-allemand). Issus d’un texte technique, les segments qui constituent leur corpus sont repris d’une étude antérieure réalisée par O’Brien (2006) avec des traductrices professionnelles, pour laquelle les textes cibles ont été générés en utilisant *IBM WebSphere* (système de TA à base de règles). Les productions ont été évaluées selon trois critères : clarté (*clarity*), précision (*accuracy*) et style. Leurs résultats montrent que, par rapport à la TH, les segments post-édités obtiennent des scores légèrement supérieurs en matière de clarté, et significativement meilleurs en matière de précision. Par ailleurs, les TH ont été jugées meilleures en matière de style.

³⁰ « Following these results, it can be argued that the use of MT followed by full post-editing to translate a standard text from the public sector did not negatively affect the reading processes of the participants recruited for this study. In terms of their subjective perceptions regarding readability and comprehensibility, it does not appear that MT post-editing affected this aspect in a negative way either, given that there appeared to be no difference between the scores given by both groups. The implications of this for the translation industry is that post-edited texts, given these results, are perceived by end users to be just as readable and comprehensible as translated ones » (Screen, 2019, p. 147).

Guerberof Arenas (2009) a comparé les produits de trois méthodes de traduction (anglais-espagnol) : la TH, la PE de TA statistique (*Language Weaver*) et la traduction à partir de segments issus d'une mémoire de traduction dans une expérience menée avec plusieurs traductrices professionnelles. L'analyse des erreurs grâce à la grille LISA³¹ a montré que les segments traduits à partir d'une mémoire de traduction contenaient bien plus d'erreurs qu'en TH et qu'en PE. Il ressort également de son analyse que « [m]istranslation is the highest value in New segments [TH], but it is very low in MT segments. This could indicate that MT clarifies difficult aspects of the source texts, although more data is needed to explore this trend » (*ibid.*, s.p.). En revanche, les segments post-édités affichent de moins bons résultats qu'en TH pour les catégories suivantes : fidélité (*accuracy*), terminologie, erreurs de langue et régularité (*consistency*).

Mentionnons également l'étude pilote effectuée par Carl et al. (2011) auprès de traductrices professionnelles et d'étudiantes de master qui porte sur la comparaison de TH et de PE de TA (*Google Traduction*) de textes issus de la presse pour la paire de langues anglais-danois. Leurs résultats ont permis de démontrer une augmentation de la productivité en PE, ainsi qu'une très légère augmentation de la qualité en PE. Cette augmentation de qualité est toutefois à considérer avec prudence, car les chercheuses reconnaissent ne pas avoir pu tirer de conclusions probantes en matière de qualité. En effet, l'évaluation par plusieurs évaluatrices humaines s'est révélée extrêmement complexe : « We believe that this difficulty derived in large part from the fact that evaluators were asked to perform relative evaluations, and nearly all the translations were of very high quality » (*ibid.*, p. 141).

Depraetere et al. (2014) ont conduit une étude avec des étudiantes de 2^e master de l'Université de Lille 3 suivant le cours de traduction assistée par ordinateur et gestion de projets. Elles ont utilisé un système de TA à base de règles, en l'occurrence, *Systran 7*. La qualité a été mesurée de deux manières distinctes : leur corpus composé de traductions humaines et de segments post-édités (anglais-français) a tout d'abord été évalué à l'aveugle par plusieurs évaluatrices humaines au moyen d'une échelle de Likert allant de 1 à 5. Alors que les résultats indiquent que la qualité des TH a systématiquement été jugée supérieure à celle des segments post-édités, les différences ont été considérées comme étant minimes. Pour chaque segment, les autrices ont également calculé un score de similarité par rapport à une traduction humaine de référence disponible sur Internet. Une fois de plus, les différences observées entre les

³¹ « LISA defines categories of errors. These are: Mistranslation, Accuracy, Terminology, Language, Style, Country, Consistency and Format. Mistranslation refers to the incorrect understanding of the source text; Accuracy to omissions, additions, cross-references, headers and footers and not reflecting the source text properly; Terminology to glossary adherence, Language to grammar, semantics, spelling, punctuation; Style to adherence to style guides; Country to country standard and local suitability; Consistency to coherence in terminology across the project and Format to correct use of tags, correct character styles, correct footnotes translation, hotkeys not duplicated, correct flagging, correct resizing, correct use of parser, template or project settings file » (Guerberof Arenas, 2009, s.p.).

segments traduits humainement et les segments post-édités sont minimales (entre 0 % et 4 %). Les chercheuses ont ainsi conclu que l’usage de la TA dans leur expérience n’avait pas eu d’effets négatifs sur la qualité des traductions finales.

Les conclusions des chercheuses Daems, Vandepitte, Hartsuiker et Macken (2017), vont, elles aussi, dans le même sens. Menée avec d’une part des traductrices professionnelles et d’autre part des étudiantes en traduction, leur recherche empirique leur a permis de faire une étude comparative de textes post-édités et de TH dans le cas de textes généraux (articles de presse) pour la combinaison de langues anglais-néerlandais. Elles se sont intéressées au processus même de traduction, à la qualité des textes cibles et à l’opinion des participantes sur la PE. Le moteur de TA employé était *Google Traduction*, toujours en TA statistique à l’époque. Elles ont premièrement conclu que la PE permettait non seulement de traduire plus rapidement qu’en TH, mais aussi d’aboutir à des productions de qualité globale comparable aux productions humaines. Leurs résultats montrent toutefois que les textes post-édités contenaient davantage d’erreurs de vocabulaire (*word sense errors* – mauvaise acception d’un terme dans ce contexte) par rapport aux TH, et que les productions des étudiantes contenaient davantage de problèmes de fluidité (*disfluent constructions* – la construction est correcte d’un point de vue grammatical, mais une traduction plus idiomatique était possible) que celles des traductrices professionnelles, particulièrement dans les TH. Deuxièmement, leurs constatations les ont amenées à affirmer qu’il y a, en fin de compte, peu de choses qui distinguent les traductrices professionnelles des étudiantes en traduction : « We found students and professional translators to be more alike than often thought, even when working with different translation methods (human translations and post-editing) » (Daems, Vandepitte et al., 2017, p. 265). Enfin, les autrices de l’étude sont convaincues, à l’instar de García (2010, 2011), que la PE constitue un processus de traduction viable, et ce, même pour les textes non spécialisés ; il s’agirait d’une méthode de traduction plus rapide qui n’entraînerait pas de baisse de qualité par rapport à la TH et qui serait, de surcroît, moins exigeante sur le plan cognitif.

L’étude de Daems, De Clercq et al. (2017) a pour objet de mesurer la présence de différences entre langue traduite et langue post-éditée. Leur corpus est repris de l’étude comparative conduite par Daems (2016), il est composé de productions (anglais-néerlandais) réalisées à la fois par des traductrices professionnelles et par des étudiantes de master en traduction. Il leur a été demandé de fournir des productions de qualité publiable, que ce soit en TH comme en PE. Il est d’abord très intéressant de noter qu’en termes de qualité, aucune différence quantifiable entre les traductions des traductrices professionnelles et celles des étudiantes n’a pu être observée dans leur expérience. Leurs résultats montrent que ni les participantes, ni l’ordinateur n’ont été capables de faire la distinction entre PE et TH. Ainsi, les conclusions de leurs recherches ne permettent pas de confirmer l’existence de ce que les autrices nomment

« *post-editesse*³² », car il serait impossible de faire la distinction entre une PE complète et une TH, que ce soit en matière de qualité, de perception par la lectrice, au niveau lexical, syntaxique ou encore sémantique.

Toutefois, dans la littérature spécialisée, l’inexistence du *post-editesse* ne fait pas consensus. Plusieurs études ont effectivement permis de mettre au jour des caractéristiques propres à une langue de PE. C’est le cas des travaux de Castilho et al. (2019), de Martikainen et Mestivier (2020), de Toral (2019) et de Volkart et Bouillon (2022) que nous développons au chapitre 6.

2.2.6.3.2 Études sur la TA neuronale

Intéressons-nous maintenant aux études plus spécifiques à la TA neuronale. Dans l’étude de Jia et al. (2019) menée avec 30 étudiantes chinoises en master de traduction sans expérience en post-édition, les évaluatrices humaines ont d’abord dû juger la fluidité des phrases cibles avant de juger l’adéquation par rapport au TS ; les critères d’évaluation employés sont ceux d’adéquation (*adequacy*) et de fluidité (*fluency*) détaillés par TAUS/CNGL (2010). Les résultats indiquent que la PE de TA neuronale (en l’occurrence *Google Traduction*) permet, d’une part, d’augmenter la qualité de la TA brute et d’autre part, d’obtenir des productions (anglais-chinois) tout aussi fluides et fidèles que les TH, que ce soit pour des textes généraux ou spécialisés.

Yamada (2019) a mené une étude comparative entre la PE de TA statistique et la PE de TA neuronale auprès d’étudiantes universitaires. L’outil de TA sous la loupe est à nouveau *Google Traduction* et la paire de langues est anglais-japonais. Les observations de Yamada indiquent que, globalement, les textes issus de la PE de TA neuronale sont meilleurs et contiennent, en moyenne, moins d’erreurs que ceux issus de la PE de TA statistique. Il souligne cependant que les étudiantes ont moins bien post-édité la TA neuronale que la TA statistique : le taux de correction des erreurs est, en effet, plus faible en PE de TA neuronale (68 % contre 77 % pour la TA statistique). Yamada laisse entendre que cette différence serait probablement due au fait que, comparativement à l’ancienne génération de TA, la technologie neuronale traduit davantage comme un humain et produit des erreurs qui ressemblent davantage à des fautes qu’un humain commettrait et qui sont, selon lui, plus difficilement repérables par les étudiantes.

Quant à Martikainen et Mestivier (2020), elles ont récemment conduit une étude comparative pour la traduction (anglais-français) de textes spécialisés avec des étudiantes de 1^{er} et 2^e master et ont constaté une amélioration de la qualité en PE de TA neuronale (en l’occurrence,

³² Le *post-editesse* peut être considéré comme le pendant du *translationese* qui témoignerait de la présence de divergences linguistiques objectives entre des textes traduits humainement et des textes post-édités : « *Post-editesse* is the difference between the characteristics of human-translated texts (HT) and the post-edited (PE) versions, in relation to the raw MT output » (Castilho et Resende, 2022, p. 4).

DeepL) avec une réduction moyenne du nombre d’erreurs de moitié par rapport à la TH. Par ailleurs, leurs résultats confirment l’existence d’un *post-editese* qui serait caractérisé par « la reprise des structures syntaxiques du texte de départ, par le calque et par l’uniformisation des solutions de traduction » (*ibid.*, s.p.).

Nous terminons cette section en mentionnant le travail de thèse de Samman (2022) dont les conclusions statistiques permettent d’affirmer que, quelle que soit la méthode (PE de *Google Traduction* ou TH), les étudiantes de bachelier en traduction ont abouti à des productions de qualité comparable : « the results of the translation quality score comparison support the assumption that students who use MTPE would show similar results when compared with students who use HT in terms of translation quality scores in the post-test » (p. 154).

2.2.6.3.3 Effets négatifs sur la qualité

Dans la littérature empirique récente, certaines scientifiques tirent cependant des conclusions qui vont en sens inverse des études que nous venons de parcourir, en mettant en lumière certains écueils en PE et en TA. En effet, plusieurs publications ont mis en évidence que la PE entraînait des répercussions négatives sur la qualité des textes cibles (Čulo et al., 2014 ; Deneufbourg, 2019 ; Depraetere, 2010 ; Martikainen et Kübler, 2016 ; Sycz-Opoń et Gałuskina, 2017).

Dans leur étude comparative réalisée avec 12 traductrices professionnelles et 12 étudiantes, Čulo et al. (2014) ont analysé les divergences entre la PE et la TH sur les plans lexical et grammatical pour la combinaison de langues anglais-allemand. La version statistique de *Google Traduction* a été utilisée comme outil de TA. Leurs observations montrent que le phénomène d’interférence avec la langue source est plus marqué en PE qu’en TH et que cela se traduit, en autres, par l’occurrence de tournures non idiomatiques et de constructions non grammaticales dans les textes post-édités. Ces effets d’interférence qui apparaissent lorsque les structures grammaticales et lexicales de la langue source laissent une trace dans la langue du TC seraient induits par la présence de la sortie de TA, autrement dit, par le processus même de PE (*ibid.*).

L’étude de cas de Depraetere (2010) a pour but d’observer les stratégies intuitives de PE (anglais-français) chez des étudiantes de 2^e master et de formuler des recommandations pédagogiques pour l’enseignement de la PE. Deux systèmes de TA ont été étudiés : un système à base de règles et un système de TA statistique. Depraetere (2010) a conclu que les étudiantes n’ont, de prime abord, pas de problèmes à se satisfaire de propositions imparfaites, qui ne sont pas idiomatiques à 100 % et qu’elles ont tendance à apporter peu de modifications stylistiques et phraséologiques³³ à la TA brute. De plus, les résultats de son étude indiquent

³³ Définition phraséologie : « une unité psychologique s’étend sur plusieurs mots ; le mot n’est qu’un élément de l’unité réelle, qui se trouve être alors une locution composée ; si, dans un groupe de mots, chaque unité graphique perd une partie de sa signification individuelle ou n’en conserve aucune, si la combinaison de ces éléments se

que la plupart des erreurs recensées dans le corpus étaient des calques ou des pertes d’informations (omissions) (*ibid.*, p. 1).

Dans leur étude comparative qualitative et statistique sur un corpus constitué des traductions (anglais-français) de résumés de revues médicales, Cochrane, Martikainen et Kübler (2016) ont observé deux tendances dans les textes issus de la PE de TA statistique : tout d’abord, au niveau terminologique, elles ont constaté une « tendance à la traduction par équivalence formelle, consistant à privilégier la traduction la plus proche, par la forme, de l’original anglais » (p. 8). Ensuite, au niveau phraséologique, elles ont noté une surreprésentation de traductions littérales ou mot-à-mot par rapport aux TH. Leur analyse leur a également permis de mettre au jour une tendance centralisatrice de la TA qui « consiste à privilégier une traduction donnée au détriment de la diversité propre à la traduction humaine » (*ibid.*, p. 11). Selon cette étude, la PE de TA statistique aurait ainsi pour effets de produire « une langue traduite aux caractéristiques particulières : plus uniformes sur les plans stylistique et syntaxique, les textes post-édités sont potentiellement moins cohérents aux niveaux textuel et terminologique que les textes issus de la traduction humaine » (*ibid.*, p. 14). À noter qu’il s’agit ici de PE réalisée par des non-professionnelles de la traduction issues, pour la plupart, du domaine médical. Enfin, les autrices sont convaincues que « la tendance à faire confiance à l’outil pour la terminologie et l’accoutumance à la sortie de TA au niveau de la phraséologie » (*ibid.*) sont venues renforcer les phénomènes qu’elles ont pu observer dans leur corpus.

Quant à Sycz-Opoń et Gałuska (2017), leur étude empirique sur l’usage de la TA (*Google Traduction*) du polonais en anglais et du polonais en français leur a permis de constater que, même après PE, les textes cibles des étudiantes contenaient encore beaucoup d’erreurs, soit qui trouvent leur source dans la TA brute et qui auraient dû être corrigées, soit qui ont été commises par les participantes en post-éditant. Elles ont également indiqué que les étudiantes n’étaient pas parvenues à gommer les maladresses stylistiques (*stylistic awkwardness*) imputables à la TA. Ainsi, elles ont conclu que la majorité des textes cibles n’avaient pas atteint un niveau satisfaisant de qualité.

Deneufbourg (2019) a, lui, réalisé une étude de cas auprès de 25 étudiantes et de 10 professionnelles à qui il a demandé de post-éditer un discours de Barack Obama traduit de l’anglais en français avec *DeepL* (TA neuronale). L’un des objectifs de cette recherche était de comparer deux niveaux de PE (*light/full*) qu’il définit comme suit : « Light post-editing : modification du strict nécessaire en évitant toute intervention stylistique ; Full post-editing : intervention maximale pour atteindre un niveau de qualité comparable à une traduction “humaine” » (*ibid.*, s.p.). Cette étude lui a permis de faire plusieurs constats : il a tout d’abord observé que les étudiantes étaient généralement meilleures pour la tâche de post-édition

présente seule avec un sens bien net, on peut dire qu’il s’agit d’une locution composée. [...] C’est l’ensemble de ces faits que nous comprenons sous le terme général de phraséologie » (Bally, 1951, pp. 65-66).

légère et qu’elles rencontraient par contre des difficultés pour atteindre un niveau de post-édition complète, et ce, contrairement aux traductrices professionnelles pour lesquelles il tire des conclusions inverses. Deneufbourg (2019) a également émis une hypothèse qui nous intéresse tout particulièrement ; il ressort de son analyse que la TA neuronale a tendance à créer une illusion de fluidité en langue cible, une « apparente fluidité » qui aurait pour effet dangereux, selon lui, de tromper la post-éditrice, qu’elle soit étudiante ou traductrice professionnelle, et de la faire passer à côté d’erreurs de sens, car elle ferait parfois preuve d’une confiance aveugle dans les propositions de la machine.

Plus récemment, Loock et Léchaugette (2021) ont réalisé une étude de cas avec 169 étudiantes de 3^e année en LEA (Langues étrangères appliquées) de l’Université de Lille n’ayant reçu aucune formation en TAO ni formation spécifique en post-édition. Ces étudiantes ont été invitées à corriger la TA (*DeepL*) en français d’un texte de presse en anglais et à remplir un questionnaire anonyme concernant leur usage des outils de TA. Il ressort de cette expérience que ces étudiantes ne sont pas parvenues à corriger efficacement le texte traduit automatiquement par *DeepL*. De plus, il apparaît qu’elles ont repéré plus aisément les erreurs liées au transfert du contenu du TS (*accuracy errors*) que les erreurs de fluidité du TC (*fluency errors*) (notamment les erreurs de grammaire et les problèmes de typographie) pour diverses raisons³⁴. Et elles n’ont pas toujours été capables de proposer une correction satisfaisante. Leur conclusion est sans appel : les étudiantes (en l’occurrence de 3^e année en LEA) ne font pas un usage raisonné des outils de TA. Pour pallier ce problème, il apparaît indispensable de proposer une formation : « [S]pecific training is therefore necessary for them [language students] to improve their MT literacy and their critical use of this technology later in the professional world » (*ibid.*, p. 215).

À nouveau, nous pensons que la divergence de conclusions que nous venons de résumer dans cette partie tient principalement à la diversité des systèmes d’évaluation employés et des types de textes évalués, à la composition des échantillons étudiés, ainsi qu’aux différentes attentes en matière de qualité. Il convient également de signaler que parmi les études susmentionnées, très peu ont été menées pour la combinaison de langues anglais-français.

³⁴ « [S]tudents tend to overlook fluency errors, including grammar mistakes in the target language, which could be the result of several factors: excessive trust in what the computer provides, the influence of the MT output preventing them from considering alternative solutions (priming), or a lack of proficiency in the target language, in this instance the students’ mother tongue » (Loock et Léchaugette, 2021, p. 217).

2.3 Évaluation de la qualité en traduction

Evaluation is one of the most problematic areas of translation, having been variously described as “a great stumbling block” (Bassnett-McGuire 1991:8), “a complex challenge” (Mahn 1987:44), “a most wretched question” (Malmkjaer 1998:70), and “a thorny problem” (Snell- Hornby 1992:19)

LYNNE BOWKER (2000a)

L’évaluation de la qualité constitue l’un des volets essentiels de notre travail, sur lequel repose l’exploitation des données récoltées dans nos expériences. Pour nos recherches, nous avons besoin d’une méthode d’évaluation qui permette une étude à la fois analytique et comparative de ces données. Nous voulions pouvoir comparer les différents modes de traduction étudiés et analyser en détail les erreurs commises dans les productions (repérage et classification des erreurs et problèmes de traduction). Ainsi, nous avons décidé de privilégier une approche par typologie d’erreurs et d’opter pour une méthode d’évaluation humaine que nous détaillons au chapitre 3.

Dans cette dernière partie de l’état de l’art, nous commençons par sonder la notion de qualité en traduction avant de nous concentrer sur l’évaluation de la qualité dans l’enseignement de la traduction et d’aborder enfin l’évaluation en traduction automatique, en distinguant trois approches qui coexistent : les méthodes d’évaluation humaine ; les méthodes d’évaluation automatique (ou automatisée) et les méthodes d’évaluation axées sur la tâche (Koehn, 2010).

2.3.1 La notion de qualité en traduction

La qualité en traduction est, en soi, un concept vaste, complexe (Deneufbourg, 2013) et incontestablement subjectif (Bowker, 2000a³⁵ ; Larose, 1998 ; Secară, 2005 ; Toudic et al., 2014). Sujette depuis toujours à débat (Doherty, 2017 ; Secară, 2005 ; Massardo et al., 2016³⁶), la notion de qualité en traduction n’est pas clairement définie et est toujours difficilement mesurable malgré la fécondité du champ de recherche spécifique qui lui est consacré : la TQA (*Translation Quality Assessment*) (Loock et al., 2013 ; Secară, 2005), et malgré l’existence de

³⁵ « The primary difficulty surrounding the issue of translation evaluation is its subjective nature » (Bowker, 2000a, p. 183).

³⁶ « The topic of quality has been debated since translation took its place in the world » (Massardo et al., 2016, p. 84).

nombreuses théories « tantôt complémentaires tantôt diamétralement opposées » (De Faria Pires, 2020, p. 50).

Les méthodes d’évaluation de la qualité en traduction sont en effet « loin de faire l’unanimité au sein de la littérature scientifique » (Deneufbourg, 2013, p. 2) et elles sont aussi diverses que variées suivant l’objectif poursuivi : « In practice [...] resource constraints and the adoption of a necessarily pragmatic approach in many areas mean that TQA processes vary considerably and have many limitations » (Doherty, 2017, p. 1)³⁷. Longtemps « la préoccupation des seuls théologiens, traducteurs littéraires ou pédagogues » (Toudic et al., 2014, p. 1), la qualité en traduction représente depuis plusieurs années un domaine de recherche aux enjeux applicatifs croissants (Hartley et Popescu-Belis, 2004) ; l’évaluation étant, entre autres, devenue un enjeu professionnel et commercial majeur sous l’effet de « [l]a croissance rapide du marché de la traduction et [de] sa mondialisation à partir du milieu des années 1990 » (*ibid.*, p. 2). L’évaluation de la qualité en traduction possède désormais de nombreux domaines d’application (Castilho et al., 2018). Outre son usage dans l’enseignement pour l’évaluation des apprenantes, dans le milieu professionnel pour l’évaluation des compétences, le recrutement ou la promotion de traductrices, ainsi qu’en recherche traductologique tant fondamentale qu’appliquée, cette pratique permet aussi aux conceptrices et aux développeuses de mesurer les performances des nouvelles technologies de la traduction (outils de TAO, TA et PE) (Doherty, 2017 ; Martikainen, 2019c).

La diversité des protagonistes du monde de la traduction et la façon d’envisager la traduction vont ainsi donner lieu à différentes manières d’en évaluer la qualité : « Toute évaluation en traduction dépend de la conception que se fait l’évaluateur du concept de traduction ce qui conduit à une multiplicité de réponses » (Martínez Melis, 2001, p. 138). C’est également ce que confirme Kenny (2018) : « There are multiple ways to evaluate both human and machine translation [...] and metrics tend to differ between academia, commercial production environments, and MT research and development laboratories (p. 64) ». De notre point de vue, nous partageons l’avis de Larose (1998) selon lequel l’évaluation d’une traduction ne peut se faire hors d’un rapport de finalité : « Juger d’une traduction, c’est d’abord et avant tout évaluer l’adéquation des résultats obtenus aux objectifs visés en s’appuyant sur le principe selon lequel il n’existe pas d’idéal de traduction hors d’un rapport de finalité » (p. 2).

En réalité, le concept même de « qualité » donne lieu à différentes interprétations (House, 1998) : « [T]he notion of quality has very fuzzy and shifting boundaries » (Bowker, 2000a, p. 183) ; « In translation studies, quality has been notoriously difficult to define and inevitably variable [sic] » (Rossi et Carré, 2022, p. 52).

³⁷ Pour un résumé de ces méthodes d’évaluation, voir Castilho et al. (2018) et Saldanha et O’Brien (2014).

Plusieurs autrices s'interrogent sur les difficultés que soulève la définition de la notion de qualité en traduction. Sur ce point, Troiano et al. (2000) estiment que le problème de la définition de la « qualité linguistique » réside fondamentalement dans le fait que

[l]'écriture [...] est un art, plus que le fruit d'une simple technique. Sa nature mystérieuse, liée aux métamorphoses infinies du logos, ne peut être appréhendée que par une longue et difficile initiation. Sans parler du talent qui, naturellement, demeure toujours le facteur clé de l'écriture, à l'instar de tout autre art. (p. 56)

Selon Castilho et al. (2018), c'est principalement en raison de la nature complexe de la traduction en tant que processus à la fois cognitif, linguistique, social, culturel et technologique qu'il est si difficile de parvenir à opérationnaliser et à évaluer le concept de qualité en traduction. Enfin, Loock et al. (2013) nous éclairent également sur cette question ; elles rejoignent sur ce point House (1998)³⁸ en expliquant que la manière d'envisager et d'évaluer la qualité en traduction va dépendre de la position théorique adoptée :

Cette difficulté à définir ce qu'est une bonne traduction vient du fait que la qualité est un concept très complexe dès qu'il s'agit de l'appliquer à la traduction. En essayant d'évaluer la qualité d'une traduction, on adopte en effet nécessairement une position théorique spécifique, puisque la question du lien entre un texte original et sa traduction se situe au cœur de n'importe quelle théorie de traduction. (Loock et al., 2013, p. 4)

Il n'est dès lors pas étonnant que la « notion de qualité varie [...] selon le public concerné : les approches théoriques, professionnelles et pédagogiques n'envisagent pas la qualité des textes traduits de la même manière [Secară, A. (2005)] » (citée dans *ibid.*).

2.3.2 Tentatives de définition

Vu qu'il s'agit d'une question incontournable en traductologie (*ibid.*, p. 3), nous avons sélectionné ci-dessous plusieurs tentatives de définition de la notion de qualité en traduction proposées par différentes traductologues.

Gouadec (2010) propose dans son ouvrage intitulé *Quality in Translation* une définition de la qualité en traduction qui porte à la fois sur le produit et sur le processus : « Quality in translation is both the quality of an end product (the translated material) and the quality of the transaction (the service provided) » (p. 270). Toujours selon lui, la qualité de ce service est considérée comme « bonne » « if and when both the provider and providee are satisfied with the translation provision process and, of course, its result » (*ibid.*).

³⁸ « [T]ranslation quality assessment presupposes a theory of translation. [...] Thus different views of translation itself lead to different concepts of translation quality, and different ways of assessing it » (House, 1998, p. 197).

Larose (1998), de même qu’O’Brien (2012), insistent sur le fait que la qualité en traduction est une notion à la fois relative et subjective :

Relative parce qu'elle dépend du client, du texte à traduire, de sa destination, du traducteur, etc.; subjective, en raison de la confusion entre une bonne et une belle traduction [...] Juger d'une traduction, c'est d'abord et avant tout évaluer l'adéquation des résultats obtenus aux objectifs visés en s'appuyant sur le principe selon lequel il n'existe pas d'idéal de traduction hors d'un rapport de finalité. (Larose, 1998, p. 2 et p. 22)

La définition formulée par Koby et Melby (2013) intègre la prise en compte des besoins de l'utilisatrice finale : « A quality translation demonstrates required accuracy and fluency for the audience and purpose and complies with all other negotiated specifications, taking into account end-user needs » (p. 178).

L'idée selon laquelle la qualité dépend du contexte ne date pas d'hier puisqu'elle est présente déjà chez Sager (1989) : « [T]here are no absolute standards of translation quality, but only more or less appropriate translations for the purpose for which they are intended » (p. 91). Et on la retrouve plus tard chez Secară (2005) : « The reason why no single standard will suffice is that quality is context dependent » (p. 1). Martikainen (2019c) constate ainsi l'émergence d'une « qualité adaptée à l'usage »³⁹ centrée sur les besoins de l'utilisatrice du produit fini : « De nouveaux modèles d'évaluation ont ainsi vu le jour, dont le point commun est de considérer la qualité comme fortement dépendante du contexte [...] et ainsi de placer les besoins de l'utilisateur du texte traduit au centre de cette notion de qualité adaptée à l'usage » (p. 40).

À la lecture de ces passages, la qualité en traduction apparaît clairement comme une notion relative (Grbić, 2008) ; on a d'ailleurs assisté à « un glissement de la notion de qualité absolue, définie par rapport à une norme "extérieure" à la traduction, vers la notion de seuil d'acceptabilité, défini par rapport à une utilisation ou à une attente particulières » (Toudic et al., 2014, p. 1). Pour une description de l'évolution des notions de qualité et d'évaluation en traduction, voir Martínez et Hurtado (2001) ou encore Toudic et al. (2014). Ainsi, pour l'ensemble du secteur « l'évaluation de la qualité des traductions est [...] devenue la question cruciale, mais non résolue » (*ibid.*, p. 2) ; l'adoption généralisée des outils de TAO et de TA a même ravivé ce sempiternel débat :

³⁹ *fit-for-purpose quality* (voir Bowker, 2020b et O'Brien, 2012b).

The debate on translation quality (e.g., House, 1997; Nord, 1991; Reiss, 2000) was far from being resolved prior to the advent of TM [translation memory] and MT, and, unsurprisingly, the widespread adoption of such translation technologies has only added fuel to a renewed debate on translation quality assessment. (Doherty, 2016, p. 958)

2.3.3 Une subjectivité intrinsèque

La subjectivité étant inhérente à toute activité humaine, elle est une dimension constitutive et inévitable de la traduction : « Subjectivity thus plays a greater role in translation than in non-mediated forms of communication ; it is inevitably involved in the translation process and affects its product » (Zehnalová, 2013, p. 43). Selon De Faria Pires (2020), il est en effet vain de vouloir prétendre à l'objectivité dès lors que la traduction est

un produit libre, au caractère tout sauf cartésien [...] [L]e produit même de l'interprétation humaine et de l'expression de la pensée de tout individu au travers du traitement linguistique d'un texte en langue source : des éléments comme le contexte de chaque tâche, le vécu, les préférences, ou encore la représentation du monde de chaque traducteur ont inévitablement une influence sur les traductions qu'il produit. (p. 51)

En outre, toute évaluation serait également une tâche fondamentalement subjective dès lors que la subjectivité intervient dans de nombreuses étapes de ce processus : du choix des objectifs de l'évaluation au choix de la stratégie (méthode de recueil d'informations), en passant par le choix des critères et des indicateurs (Gérard, 2002, p. 9). La Figure 9 illustre la place de la subjectivité dans l'évaluation.

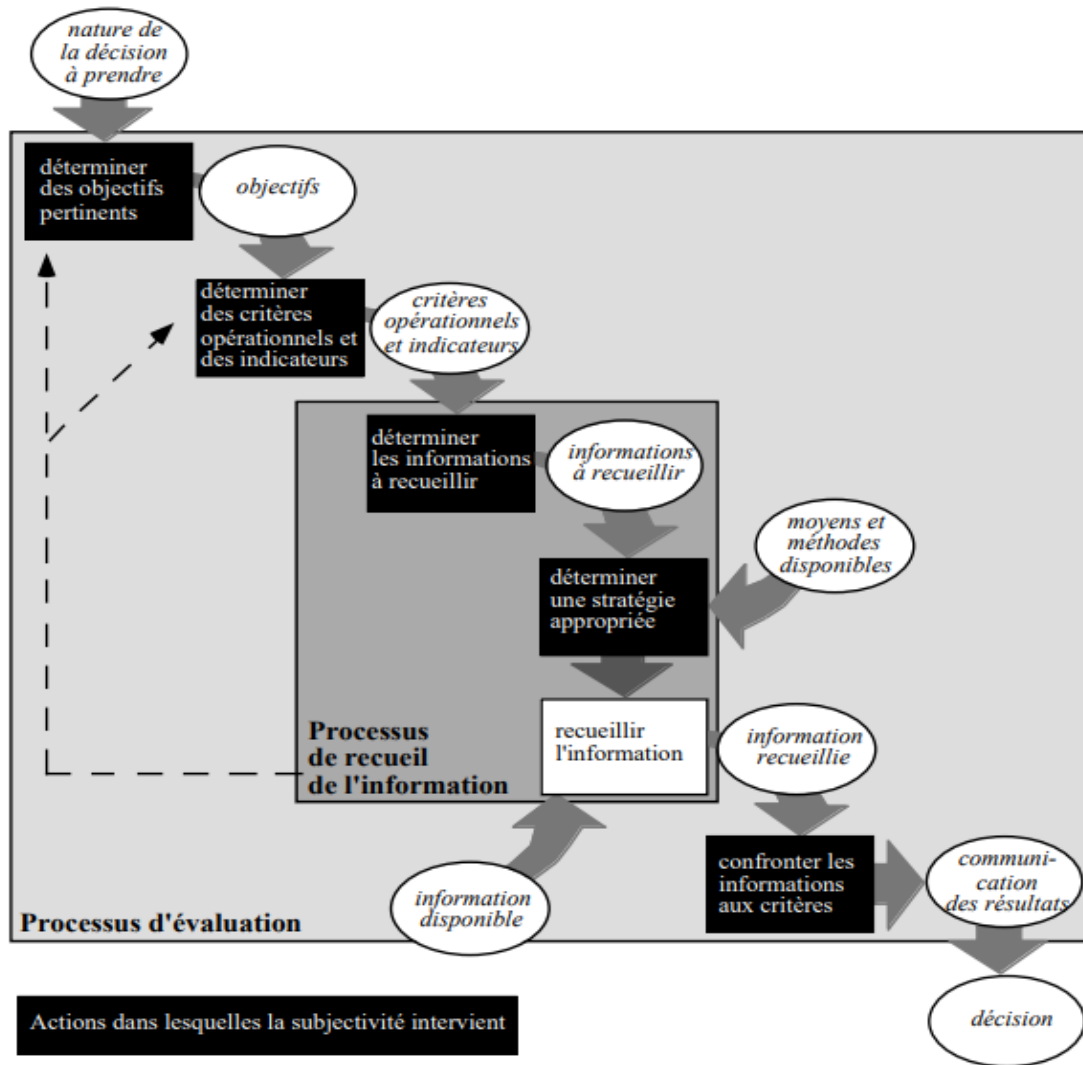


Figure 9 : Subjectivité dans l'évaluation (De Ketele et Roegiers, 1993, p. 77)

Plusieurs chercheuses rendent compte de l'impossibilité d'une évaluation purement objective de la qualité en traduction :

Why should we expect then that a necessarily subjective evaluation of something subjective by its very nature will be perfectly objective? The aim of translation evaluation is not to produce absolute judgements or perfectly objective assessments. Such goals are not attainable. Just as a target text (TT) is not the ideal translation of the source text (ST) and needs to be targeted for a specific situation, specific purpose and a specific audience, translation quality evaluation needs to be targeted in the same way. (Zehnalová, 2013, p. 43)

Pour Doherty (2017) également, cette impossibilité est due à la complexité du processus de traduction, ainsi qu'à la multiplicité des paramètres influençant l'évaluation : « a universal, and entirely objective, approach to TQA would be difficult to argue for – in part, because of the complexity of the translation process, but also because of the diversity of texts, languages, directions and applications used to inform the TQA debate » (p. 132).

Par ailleurs, Troiano et al. (2000) conviennent que dans le domaine des services linguistiques, si l'on veut éviter que « toute affirmation de "qualité" [soit] aléatoire, mythique, sinon illusoire et fallacieuse » (p. 56), il est indispensable que le concept de qualité « soit accompagné de critères d'évaluation à même d'en définir le contenu » (*ibid.*). Mais malgré la définition de critères précis, il est courant que plusieurs évaluatrices ne parviennent pas à s'accorder sur le niveau de qualité d'une traduction donnée (Rossi et Carré, 2022 ; Stymne et Ahrenberg, 2012). Il ressort, en effet, explicitement de la littérature que lorsque l'évaluation de la qualité est confiée à des évaluatrices humaines, elle comporte inévitablement une part de subjectivité (Larose, 1998⁴⁰ ; Poibeau, 2019⁴¹ ; Zehnalová, 2013⁴²) :

[W]hat is a "good" translation for one reviewer might be a "very good" translation for another. If a sentence is very long, but contains one serious error, it might be unclear whether this should be considered a "very bad translation" or a "bad translation". This evaluator subjectivity is one of the main points of criticism on translation quality assessment approaches (Williams, 2009). (cité dans Daems, 2016, p. 22)

Bien entendu, il existe différentes stratégies qui permettent de réduire les biais dans l'évaluation humaine de la qualité. Citons notamment celle avancée par Bayerl et Paul (2011) qui consiste à avoir recours à plusieurs évaluatrices et à viser un haut degré d'accord inter-évaluateur :

[T]he higher the number of annotators who are able to agree, the less bias and distortion can be expected in the data and the more stable and comparable the results are likely to be. Reaching high agreement with a larger number of annotators would thus indicate a higher trustworthiness of annotations. (p. 713)

⁴⁰ « On ne parviendra jamais à éliminer complètement la subjectivité de l'évaluateur, mais on peut la réduire » (Larose, 1998, p. 19).

⁴¹ « Il est très difficile d'évaluer la qualité d'une traduction : cette évaluation comporte forcément une part de subjectivité et dépend étroitement des besoins et du point de vue de l'utilisateur de la traduction » (Poibeau, 2019, p. 153).

⁴² « TQA is carried out by humans, thus the subjective component of the "human factor" is even more pronounced » (Zehnalová, 2013, p. 43).

Effectivement, que ce soit en TH ou en TA, les tâches d’évaluation peuvent être effectuées soit par une seule personne, soit par un groupe d’évaluatrices (ce qui est notre cas) afin de limiter l’arbitraire :

On an individual level, the evaluator typically assesses HT or MT against the given criteria and provides the results. Group-based TQA involves several evaluators conducting the assessment in this way and the averaging of their scores, to moderate strong positive or negative personal biases. (Castilho et al., 2018, p. 23)

Dans le cas où l’évaluation est effectuée par plusieurs personnes, « [i]deally, there should be close similarity between the scoring that each evaluator assigns to each translation » (Doherty, 2017, p. 139). Il est alors recommandé de recourir à des tests statistiques afin d’évaluer ce degré d’accord entre plusieurs évaluatrices. Les méthodes couramment employées pour mesurer cette fiabilité inter-évaluateurs (*inter-rater reliability*) sont le score kappa (kappa de Cohen ou kappa de Fleiss), les coefficients de corrélation de Pearson ou de Spearman (*ibid.*) ou encore le coefficient de corrélation intra-classe (CCI).

Depraetere et al. (2014) soulignent un autre aspect à ne pas négliger : Il s’agit du besoin de former adéquatement les évaluatrices humaines ; « [u]nless there is extensive training and instruction about quality levels, the subjectivity inherent in any evaluation is likely to remain a challenge in the context of human evaluation » (p. 99).

Et si cette subjectivité était nécessaire ?

En fin de compte, posons-nous la question : pourquoi vouloir tant s’efforcer d’éliminer la subjectivité inhérente au processus d’évaluation ? Gérard (2002) adopte un point de vue intéressant sur cette question, en considérant que puisqu’il est vain de chercher à éviter cette subjectivité, « il faut en avoir pleinement conscience, et savoir que c’est parce qu’il y a cette subjectivité que l’on peut parler d’évaluation » (p. 9). Il introduit à cet égard une distinction que nous trouvons tout à fait pertinente ; selon lui, ce n’est pas tant le côté subjectif de l’évaluation qu’il convient d’éviter, mais bien le côté arbitraire : « Une évaluation arbitraire serait celle qui procéderait "n'importe comment", sans se fonder sur une démarche consciente, rigoureuse, critique » (*ibid.*, pp. 1-2). C’est pourquoi, dans ce travail, nous nous sommes attachée avant tout à limiter l’arbitraire dans l’élaboration et dans la mise en œuvre pratique de notre méthode d’évaluation (voir chapitre 3, section 3.3.2.2).

2.3.4 L’évaluation orientée produit et l’évaluation orientée processus

Les modèles d’évaluation diffèrent aussi en fonction de l’objet de l’évaluation (Martikainen, 2019c, p. 33). O’Brien (2012) fait la distinction entre deux grandes catégories d’évaluation :

l'évaluation orientée produit et l'évaluation orientée processus. La première catégorie de modèles se concentre sur les erreurs dans le produit fini : « the error-based models seek to identify errors, classify them, allocate them to a severity level and apply penalty points with a view to deciding whether or not the translation meets a specific pass mark » (pp. 57-58). Tandis que dans la deuxième catégorie, les modèles renvoient à une conception plus large de la qualité puisqu'ils envisagent le processus de production dans sa globalité (Martikainen, 2019c, p. 33) : « The second category views translation quality in a more holistic manner, from the point of view of service provision and the competences, tools and procedures required to produce high quality translation » (O'Brien, 2012a, p. 58).

2.3.5 Approche typologique

L'une des méthodes d'évaluation de la qualité les plus courantes en traduction consiste à utiliser une typologie d'erreurs. Ces typologies « sont utilisées pour les besoins d'évaluation divers dans les différents domaines liés à la traduction » (Martikainen, 2019c, p. 33). C'est d'ailleurs cette approche que nous avons retenue dans le cadre du présent travail. Il s'agit d'une méthode valide pour évaluer la qualité qui consiste à

defining different error categories and marking or counting the errors of each category present in a certain text or sentence. Weightings can be added as well: either the evaluator decides on the weight for each instance of that error type [...] or the error type as a whole receives a fixed weight. (Daems, 2016, pp. 22-23)

Certes, cette approche est très chronophage (*ibid.*), mais elle a bien des avantages, elle permet, entre autres, d'avoir un repérage minutieux des erreurs : « Although the process of error analysis is highly time-consuming, [...] error analysis generates rich data, which is necessary for diagnostic and comparative evaluation of translations » (*ibid.*, p. 23). Avec une approche typologique, il est par exemple possible de mesurer « l'effet de l'erreur sur la qualité de la traduction, en fonction de la nature de celle-ci, de son objectif, de ses destinataires et de son usage » (Toudic et al., 2014, p. 8).

Dans le secteur des services langagiers, on se contente très souvent de recourir à des méthodes typologiques arbitraires et uniques (*one-size-fits-all models*), c'est-à-dire, à des typologies difficilement adaptables en fonction des besoins :

While researchers and academics tend to focus on theoretical and pedagogic concerns related to translation quality, in most sectors of the industry TQA is broadly limited to the application of somewhat arbitrary "one-size-fits-all" error typology models that aim to give quantitative indicators of quality. (Castilho et al., 2018, p. 2)

Parmi les outils d’analyse les plus connus, citons la grille LISA QA⁴³ (*Localisation Industry Standards Association*) et la grille « SAE J2450 Translation Quality Metric⁴⁴ ». Citons également la grille MQM⁴⁵ (*Multidimensional Quality Metrics*) (Lommel et al., 2014), plus flexible que les grilles précédentes, qui a été harmonisée dans le cadre du projet QT21 de l’Union européenne avec le système DQF (*Dynamic Quality Framework*) élaboré par TAUS pour former la typologie MQM-DQF⁴⁶ (Lommel, 2018b).

Ainsi, l’approche typologique se révèle être l’une des méthodes privilégiées dans le secteur de la traduction professionnelle, mais pas seulement, puisqu’elle est aussi largement adoptée dans l’enseignement de la traduction sous d’autres modalités :

[T]he evaluation models used in the industry are predominantly error-based. [...] Interestingly, this is a common assessment approach in academic translator training programmes as well, even though lecturers tend to grade entire translation assignments or exams of their students (as opposed to samples or extracts). (Castilho et al., 2018, p. 14)

⁴³ Outil d’analyse créé en 1990 et modèle d’inspiration pour de nombreux autres modèles d’assurance qualité (O’Brien, 2012a), la grille LISA QA « permet de contrôler tous les composants de la localisation d’un produit, incluant la fonctionnalité, en classant les erreurs par niveau de gravité (*critical, major, minor*) » (Lavault-Olléon et Allignol, 2014, p. 9). http://producthelp.sdl.com/SDL_TMS_2011/en/Creating_and_Maintaining_Organizations/Managing_QA_Models/LISA_QA_Model.htm

⁴⁴ La grille SAE J2450 a été conçue par la *Society of Automotive Engineers* « pour l’évaluation des traductions techniques dans le domaine de l’automobile. Fondée sur sept catégories d’erreurs (Terminologie, Syntaxe, Omission, Structure lexicale, Orthographe, Ponctuation, Divers) et deux niveaux de gravité (majeur et mineur), la grille SAE J2450 a l’avantage de la simplicité, mais l’inconvénient d’être essentiellement adaptée à un seul type de document et de traduction » (Toudic et al., 2014, p. 5).

⁴⁵ La grille MQM « is a framework for analytic Translation Quality Evaluation (TQE). It can be applied to both human translation and machine translation [...] The MQM error typology begins with seven dimensions. Each dimension contains multiple more specific error subtypes. For example, Accuracy contains subtypes such as Addition, Mistranslation, and Omission » (MQM Website, 2022). Pour plus de détails, consulter <https://themqm.org/error-types-2/typology/>

⁴⁶ « Translation quality assessment (TQA) has suffered from a lack of standard methods. Starting in 2012, the Multidimensional Quality Metrics (MQM) and Dynamic Quality Framework (DQF) projects independently began to address the need for such shared methods. In 2014 these approaches were integrated, centring on a shared error typology (the “DQF/MQM Error Typology”) that brought them together. This approach to quality evaluation provides a common vocabulary to describe and categorise translation errors and to create translation quality metrics that tie translation quality to specifications » (Lommel, 2018b, p. 109).

2.3.6 Évaluation de la qualité dans l'enseignement de la traduction

La pédagogie en traduction est, en quelque sorte, le gros œuvre de l'édifice. Les finitions seront toujours laissées au gré de chacun, mais la charpente doit être solide. La pédagogie, c'est aussi le climat dans lequel on apprend, la façon d'apprendre et la façon d'enseigner, la manière d'écouter et d'entendre, bref, de mettre en valeur tout le potentiel de ceux et de celles dont nous avons la responsabilité professionnelle. [...] Elle est un chemin délicat de concertation entre le « maître » et « l'élève » pour viser, et parfois atteindre, l'excellence.

HANNELORE LEE-JAHNKE (2001)

Dans sa thèse, Martínez Melis (2001) reconnaît d'emblée que les méthodes d'évaluation des traductions utilisées dans l'enseignement de la traduction sont extrêmement nombreuses « et peuvent même varier d'un enseignant à l'autre à tel point qu'on peut dire que chacun a son propre système » (p. 107). Aussi, dans cette partie, loin de nous l'idée de rendre compte de la grande diversité de pratiques évaluatives. Nous souhaitons simplement mettre en exergue les principaux défis et enjeux de l'évaluation de la qualité en contexte pédagogique en faisant référence aux principales chercheuses du domaine. Pour un regard approfondi sur le sujet, voir Martínez Melis (2001).

Lee-Jahnke (2001) considère que si « les méthodes d'évaluation se multiplient depuis des années et que tout un chacun essaie de trouver une voie aussi objective que possible » (p. 259), c'est principalement en raison de la subjectivité des critères d'évaluation et parce que les enseignantes et formatrices ne disposent pas d'une terminologie universelle et standardisée :

[N]ous ne nous situons pas dans les sciences exactes où une notation serait relativement "facile", car objectivable. Nous traitons avec une matière qui est délicate à schématiser, en raison de la subjectivité des critères. Il est vrai qu'une terminologie unique, comprise par tous, fait encore défaut dans bien des langues et, même, dans les langues où ce "métalangage" existe, celui-ci est ignoré par bon nombre d'enseignants. (*ibid.*)

Toudic et al. (2014) déplorent que, malgré les efforts menés en ce sens, notamment par Gouadec (1981) et Martínez et Hurtado (2001), il n'existe toujours pas de méthode

d’évaluation consensuelle au sein de l’enseignement universitaire qui soit « suffisamment fidèle aux contraintes professionnelles de la traduction, tout en restant simple d’utilisation et adapté[e] à la mesure des progrès et à la définition de profils d’étudiants traducteurs » (Toudic et al., 2014, p. 3).

Faute d’une définition unique et universelle de la qualité et d’un modèle commun, certaines se fient simplement à leur expérience et à leur intuition, comme l’expliquent Van Egdom, Verplaetse et al. (2018) : « Translators and translator trainers holding a transcendent view of translation quality seem unable to define quality itself or, even, to sum up the concrete characteristics of a quality translation, yet they immediately recognise a translation as being of good or excellent quality » (p. 28).

2.3.6.1 Quel est le rôle de l’évaluation dans l’enseignement ?

Nous retiendrons trois grandes fonctions de l’évaluation en pédagogie : formative, sommative et diagnostique. Proposée par Scriven en 1967, l’évaluation formative « a pour but d’adapter le dispositif pédagogique à la réalité des apprentissages des élèves » (Nunziati, 1990, p. 57). Ainsi, une évaluation sera dite formative « quand le professeur porte un jugement sur un apprentissage à n’importe quel moment *pendant* le processus d’apprentissage, et dans le but d’aider un étudiant à *améliorer l’apprentissage en cours de réalisation* » (Prégent, 1990, p. 53). Alors que l’évaluation sera dite sommative « quand le professeur porte un jugement sur un apprentissage *au terme* de cet apprentissage — comme s’il faisait le bilan, la "somme", des connaissances qu’un étudiant a acquises pendant le cours ou une étape du cours » (*ibid.*, p. 50). Enfin, il existe également l’évaluation à visée diagnostique qui « est effectuée avant un processus d’apprentissage, pour pouvoir situer l’étudiant dans un niveau adéquat ou découvrir la cause de certaines déficiences dans son apprentissage » (Martínez Melis, 2001, p. 50).

Comme le spécifient Toudic et al. (2014), la visée de l’évaluation de la qualité dans l’enseignement de la traduction est essentiellement certificative et formative, « l’évaluation est essentiellement utilisée dans un but de formation et de mesure du niveau et des progrès de l’étudiant en traduction » (p. 3). Et selon *elles*, la difficulté d’aboutir à un modèle d’évaluation approprié à un contexte pédagogique provient de la tension qui existe entre ces deux visées de l’évaluation :

Dans les formations de traducteurs, comme dans d’autres formations, la difficulté de parvenir à un modèle idoine d’évaluation tient à la dualité entre le besoin de sanctionner un certain niveau de compétences par des points (ou des appréciations) et celui d’assurer la progression des étudiants en justifiant l’évaluation de façon objective et circonstanciée. (*ibid.*)

Et force est de constater que toutes les enseignantes et formatrices ne s'entendent pas sur le rôle de l'évaluation : « [p]our beaucoup, elle sert uniquement à sanctionner les travaux à la fin du cours. Pour d'autres, par contre, elle sert à aider les étudiants à s'améliorer dans leurs performances. [...] cette amélioration ne vise pas seulement la réussite dans les examens. Elle doit essentiellement viser l'amélioration de l'aptitude à traduire et, ensuite seulement, le produit fini, à savoir : une traduction de qualité » (Lee-Jahnke, 2001, p. 260).

Or, dans l'enseignement de la traduction, il apparaît que ce sont les modèles d'évaluation sommative qui consistent très souvent à classer les erreurs par catégories et à soustraire des points-fautes qui « ont longtemps prévalu et prévalent encore dans certaines formations » (Toudic et al., 2014, p. 3). Lee-Jahnke (2001) regrette également cette prédominance de l'évaluation sommative :

[Q]uand on pense "évaluation", on désigne en général l'évaluation dite "sommative", c'est-à-dire celle qui sanctionne, et qui, dans la plupart des cas, s'exprime en chiffres, en notes "facilement" attribuées. En réalité, les méthodes et modes de calcul posent d'autres problèmes et maintes tentatives ont été faites, et vont encore être entreprises, afin de rendre cet exercice aussi fiable et objectif que possible. (p. 259)

Elle plaide, en outre, pour l'instauration des modalités d'évaluation formative dès le début de l'enseignement. Et cette situation n'est pas sans conséquences sur l'attitude des étudiantes, comme le constate Allignol (2007), car elles « font confiance à la notation pour s'évaluer et sont perpétuellement en attente de notes » (p. 73).

2.3.6.2 Approche typologique en didactique de la traduction

Comme évoqué précédemment, l'approche typologique est utilisée dans l'évaluation des productions des apprenantes (Martikainen, 2019c). Martikainen nous donne l'exemple de la typologie MeLLANGE⁴⁷ (*Multilingual eLearning in LANGuage Engineering*) (Castagnoli et al., 2011) développée dans le cadre du projet européen du même nom visant l'amélioration des méthodes d'enseignement en traduction spécialisée. Ces typologies « permettent généralement une annotation détaillée des erreurs, puisqu'elles visent une analyse fine des difficultés des apprenants » (Martikainen, 2019c, p. 35).

Si l'on adopte cette approche, il convient de s'entendre sur une typologie clairement définie. À cet égard, Gouadec (1981) souligne la nécessité de disposer de paramètres d'évaluation

⁴⁷ <http://mellange.eila.univ-paris-diderot.fr>

rigoureux, justes et objectifs. Selon lui, une définition claire de ces paramètres permettrait à la traductrice « de parvenir à une conscience claire des objectifs pédagogiques et, surtout, de se trouver avec le "pédagogue" – enseignant ou réviseur – en une relation non ambiguë puisque le pédagogue ne peut se trouver accusé de "subjectivisme" » (*ibid.*, p. 101). En outre, Rossi et Carré (2022) recommandent de veiller à ce que ces paramètres ne soient pas trop nombreux : « Error typologies tend to be rather complex [...] For the sake of simplicity and ease, a limited set of common errors could be used in an evaluation » (p. 58).

Toutefois, si le recours à des catégories d’erreurs et à une pondération peut tendre vers une forme d’objectivité, cette méthode d’évaluation n’est pas sans inconvénient. Elle est relativement chronophage et laisse tout de même une part au jugement subjectif de l’évaluatrice lorsqu’il s’agit de spécifier la catégorie à laquelle l’erreur doit être attribuée, ou de préciser le degré de gravité d’une erreur (mineure ou majeure), et ce, même lorsque les catégories sont définies (Saldanha et O’Brien, 2014, pp. 101-102).

Rappelons que l’approche typologique est la méthode d’évaluation que nous avons décidé d’adopter dans nos deux expériences, car nous voulions reproduire des conditions expérimentales proches du contexte d’apprentissage réel dans lequel les apprenantes évoluent. Nous avons effectivement tenu à employer une méthode d’évaluation qui soit la plus proche possible des pratiques d’évaluation des enseignantes de la filière traduction-interprétation de l’Université de Liège. Toutefois, faute d’une typologie consensuelle parmi le corps enseignant, nous avons élaboré notre propre typologie d’erreurs qui s’appuie sur les principaux critères communément employés au sein de la filière. Nous y revenons plus amplement au chapitre 3.

2.3.6.3 Distinction entre erreurs de langue et erreurs de traduction

« Tout le monde s’accorde, semble-t-il, pour faire la différence entre faute de langue et faute de traduction » (Martínez Melis, 2001, p. 61). Il est, en effet, très fréquent, dans ces typologies, de distinguer les erreurs de langue des erreurs de traduction (Martikainen, 2019c, p. 34). Toudic et al. (2014) expliquent l’une des raisons d’être de cette distinction : « [L]a nécessité d’offrir une évaluation "formative" aux futurs traducteurs a poussé un certain nombre d’universitaires à appliquer des principes plus holistiques ou minimalistes, basés sur une opposition entre transfert du sens et qualité linguistique » (p. 3). Delisle (1993) définit la faute de langue comme une « [e]rreur qui figure dans le texte d’arrivée et qui est attribuable à la méconnaissance de la langue d’arrivée ou de son maniement » (p. 659). Il ajoute que ce type d’erreurs relève d’« une maîtrise insuffisante du vocabulaire, de la grammaire ou des techniques de rédaction » (*ibid.*). Il peut s’agir de barbarismes, de fautes d’orthographe, de

grammaire ou de ponctuation, d'impropriétés, de mauvaises cooccurrences, etc. Tandis que l'erreur de traduction est une « erreur qui résulte d'une mauvaise interprétation d'un terme ou d'un segment du texte de départ, d'un défaut de méthode ou d'un manque de connaissances générales » (Office québécois de la langue française, 2009, s.p.). Il peut s'agir de non-sens, de contresens, de glissements de sens, de calques, etc.

2.3.6.4 L'évaluation en contexte d'apprentissage vs en contexte professionnel

[T]ranslation quality assessment will very much depend on the context in which the translation is done, and the expectations of translator trainers will certainly differ from those of a client who needs a translation for specific purposes.

CAROLINE ROSSI ET ALICE CARRE (2022)

Pour Gile (2005), l'enseignement universitaire classique de la traduction est plutôt orienté produit qu'orienté processus ; les enseignantes s'attachant à évaluer les mérites et les faiblesses des productions des étudiantes, plutôt que le processus même de traduction :

[L]’enseignement prend essentiellement la forme d’exercices de traduction qui sont faits par des étudiants, puis corrigés avec une évaluation du produit de l’exercice que sont les textes d’arrivée. Ce produit est jugé en fonction du nombre et de la gravité des fautes de sens et des fautes et maladroites de langue qu’il contient, et en fonction de la qualité des choix terminologiques, stylistiques et autres faits par les étudiants. À partir des notes attribuées et des commentaires approuvateurs ou désapprouvateurs de l’enseignant, l’étudiant corrige son tir et améliore ses traductions. (pp. 26-27)

Scarpa (2010) souligne que l'évaluation en contexte d'apprentissage adopte une vision essentiellement qualitative de la traduction reposant sur des « critères idéaux et statiques que sont l'exactitude⁴⁸ et l'utilité⁴⁹, lesquels sont principalement d'ordre linguistique et textuel »

⁴⁸ Par exactitude, on se concentre sur « la transmission inaltérée du contenu référentiel du texte de départ » (Scarpa, 2010, p. 239).

⁴⁹ « Pour être utile, une traduction doit être empreinte de transparence pour répondre aux exigences fonctionnelles et stylistiques d'économie, de clarté et de justesse. [...] L'utilité couvre les trois [...] maximes de la quantité (économie : fournir suffisamment d'information sans être redondant, sinon dans le sens du *skopos* de la

(p. 238), tandis que l’évaluation en milieu professionnel se fonde sur une vision essentiellement quantitative de la traduction : « l’exactitude et l’utilité font donc place à des critères de qualité davantage dynamiques, axés sur [...] la justesse de la traduction par rapport aux intentions du donneur d’ouvrage et sur l’acceptabilité de cette traduction par les utilisateurs » (*ibid.*). Toujours selon lui, ces critères idéaux utilisés en contexte d’apprentissage (exactitude et utilité) constituent toutefois une référence valide pour l’évaluation de la qualité en milieu professionnel.

Ajoutons à cela que l’évaluation en contexte d’apprentissage n’est pas du tout axée sur la rentabilité, contrairement à l’évaluation des traductions en contexte professionnel, qui va toujours être assujettie au critère de productivité (Larose, 1998).

Enfin, concernant les exigences en matière de qualité, Lavault-Olléon et Allignol (2014) insistent sur le fait que les étudiantes, parce qu’elles sont en formation,

doivent tendre vers la qualité la plus élevée qu’ils sont capables de fournir [...] [Car] c’est en s’efforçant de fournir toujours le meilleur travail que les étudiants progressent le plus, et c’est cette attitude rigoureuse qui leur permettra plus tard de s’adapter à tout environnement professionnel et à toutes les exigences. Les futurs traducteurs doivent également être conscients qu’un niveau de qualité moyen, même demandé par le commanditaire, peut porter atteinte à leur image de professionnel. (p. 14)

Elles ajoutent qu’il est toutefois indispensable que les formateurs les confrontent « à la réalité professionnelle, et donc à des attentes variées en matière de qualité » (*ibid.*). À cet égard, Toudic et al. (2014) sont convaincus que l’un des défis majeurs pour les enseignantes et formatrices est de parvenir à faire prendre conscience aux étudiantes de la variabilité de la qualité attendue dans le monde professionnel, « tout en maintenant des objectifs de qualité maximale dans la formation et dans l’évaluation » (p. 2).

2.3.7 Évaluation de la TA

Tout comme la traductrice humaine, la machine commet des erreurs tantôt drôles et insignifiantes, tantôt catastrophiques (Kenny, 2022, p. 33). Dans le domaine de la TA, l’évaluation de la qualité n’est pas envisagée comme une fin en soi, mais plutôt comme un moyen d’améliorer les systèmes de TA (Rossi et Carré, 2022)⁵⁰ ou encore d’évaluer

traduction) ; pertinence (justesse : fournir l’information pertinente à la situation) et manière (clarté : fournir l’information de manière claire et ordonnée) » (*ibid.*, p. 241)

⁵⁰ « As far as MT is concerned, quality has been seen more as a means to an end (namely improving systems), and so a pragmatic approach has prevailed » (Rossi et Carré, 2022, p. 52).

l'utilisabilité⁵¹ (*usability*) d'un outil. C'est pourquoi l'évaluation se révèle jouer un rôle central dans la recherche et le développement de solutions de TA (Doherty, 2017) :

Whole branches of research are thus devoted to estimating the quality that given machine translation systems are likely to produce, evaluating particular outputs, designing ways to correct errors by post-editing machine translation output or helping the machine produce better output in the first place, usually by pre-editing source texts to make them easier to translate. (Kenny, 2022, p. 33)

Il n'est d'ailleurs pas rare de demander directement l'avis des utilisatrices des systèmes en les rendant partie prenante du processus d'évaluation :

Inclusion of users in evaluation of MT systems can provide benefits in both directions, such as positive influences on system development and its usability (Flournoy & Callison-Burch, 2001), which in turn lead to better systems and better output, making life easier for specialist users such as student and professional translators. (Doherty et O'Brien, 2014, p. 41).

Précisons que, tout comme en TH, la notion de qualité en TA n'est pas consensuelle : « when it comes to assessing the quality of MT outputs, different definitions of quality are again used » (Rossi et Carré, 2022, p. 52). De même, il n'existe pas non plus de consensus sur la méthode d'évaluation de la qualité en TA :

Si la majorité des méthodes d'évaluation proposées pour déterminer les progrès des systèmes construits par apprentissage à base de corpus se fondent essentiellement sur des mesures de qualité à partir des sorties du moteur, force est de constater qu'il n'existe pas de véritable consensus quant à la primauté d'une métrique par rapport à une autre (les avis divergent d'autant plus d'une communauté à l'autre ; à titre d'exemple, linguistes computationnels vs traducteurs professionnels). (Peraldi, 2016, p. 72)

Retenons que l'évaluation du produit en TA repose historiquement sur deux grandes approches : l'évaluation de la sortie d'une TA par des juges humaines et l'évaluation automatisée de la qualité (Doherty, 2017, p. 133).

⁵¹ Le concept d'utilisabilité est défini par la norme ISO 9241-11 comme le « degré auquel un système, un produit ou un service peut être utilisé par des utilisateurs spécifiés, pour réaliser des objectifs spécifiés avec efficacité, efficience et satisfaction, dans un contexte d'utilisation spécifié » (Organisation internationale de normalisation, 2018, s.p.).

2.3.7.1 Évaluation humaine de la qualité en TA

L’implication de juges humaines dans le processus d’évaluation remonte aux balbutiements de ce champ de recherche (Fiederer et O’Brien, 2009). Aujourd’hui encore, cette méthode est considérée comme la plus fiable pour évaluer la qualité d’un moteur de TA (Koehn, 2020). La première grande campagne d’évaluation humaine des systèmes de TA fut la campagne ALPAC en 1966 dont l’objectif était de « mesurer la rentabilité de la TA par rapport à la traduction humaine » (Hartley et Popescu-Belis, 2004, p. 325). La méthode d’évaluation utilisée reposait sur deux critères qui seront largement réutilisés ensuite : l’intelligibilité et la fidélité : « [U]ne traduction peut être intelligible et manquer de fidélité ou d’exactitude (précision), et inversement, elle peut être fidèle et manquer d’intelligibilité [...] L’intelligibilité de la traduction était évaluée sans référence à la phrase source » (Blanchon et Boitet, 2007, p. 38).

2.3.7.1.1 Approches classiques en TA

De nos jours, l’évaluation humaine de la TA repose essentiellement sur le modèle *adequacy/fluency* (Doherty, 2016 ; Koehn et Monz, 2006 ; O’Brien, 2012a) ; ces critères furent officiellement adoptés dans le cadre de la 1^{re} tâche partagée lors de la campagne d’évaluation WMT06 (2006 Workshop/Conference on Machine Translation) (Popović, 2020, p. 5060). Il est demandé aux évaluatrices d’attribuer une note selon ces deux paramètres qui sont généralement évalués sur une échelle de 5 valeurs (O’Brien, 2012a, p. 66). Il s’agit de juger la fluidité/lisibilité (*fluency*) du texte cible et l’adéquation/fidélité (*adequacy*) du texte cible par rapport au texte source sans procéder à une analyse détaillée des erreurs commises. Le critère d’adéquation « mesure à quel point une traduction automatique rend compte d’éléments d’information contenus dans un texte de référence [...] Le caractère naturel de la traduction (*fluency*) vise à vérifier la bonne formation des phrases, leur caractère naturel et la facilité qu’a un lecteur à les comprendre » (Poibeau, 2019, p. 156).

Une des grandes critiques en évaluation de la TA concerne ces deux principaux critères employés pour juger les traductions puisqu’ils sont considérés comme très subjectifs et rudimentaires. De plus, Saldanha et O’Brien (2014, p. 104) ajoutent que cette approche a été déforcée en raison de l’utilisation d’autres termes, tels que *intelligibility*, *acceptability* ou *comprehensibility*, qui recourent entièrement ou partiellement les notions de *fluency* et d’*adequacy*.

L’autre méthode usuelle dans le domaine de l’évaluation humaine de la TA consiste à comparer et à classer différentes propositions de TA d’une même phrase ou d’un même passage d’un TS (Daems, 2016). Plutôt que de juger la fluidité et l’adéquation sur une échelle de valeurs absolues,

it is typically easier to rank two or more systems against each other on a sentence-by-sentence basis. In the case of two systems, the question “Is system output A better than system output B, or worse, or indistinguishable?” is typically answered by human evaluators in a more consistent manner than questions about adequacy or fluency. (Koehn, 2010, p. 220)

Une des limites fréquemment évoquées avec ces méthodes est le manque de prise en compte du contexte, en particulier lorsque l’évaluatrice est amenée à juger de la qualité de phrases isolées : « Typically, such evaluation is done sentence by sentence, but a longer document context may be essential to carry out the judgments. For instance, the resolution of pronouns may only be faithfully evaluated in context » (Koehn, 2010, p. 218, voir également l’étude de Toral et al., 2018).

Läubli et al. (2018) plaident à cet égard en faveur d’un changement de protocole d’évaluation de la qualité en TA vers une méthode qui reposerait sur l’évaluation de documents dans leur ensemble :

As machine translation quality improves, translations will become harder to discriminate in terms of quality, and it may be time to shift towards document-level evaluation, which gives raters more context to understand the original text and its translation, and also exposes translation errors related to discourse phenomena which remain invisible in a sentence-level evaluation. (p. 4795)

2.3.7.1.2 Approche typologique en TA

Ces méthodes classiques de notation et de comparaison des sorties de TA ne permettent pas d’avoir une analyse approfondie de la qualité (Daems, 2016), ni de cerner les spécificités de la TA (Martikainen, 2019c). Alors, pour compenser en partie ces lacunes, il peut être envisagé d’adopter une approche typologique, analogue à l’approche typologique en TH, en demandant à des évaluatrices humaines de catégoriser les erreurs rencontrées, comme l’attestent Popović (2021) : « In order to get more details about the actual errors, error classification according to a predefined error scheme is often performed » (p. 163) et Kenny (2022) :

Sometimes human evaluators are asked not just to score a machine translated segment or document using one of the metrics described above, but to say precisely what is wrong with the particular output, by assigning each error in the segment to a category specified in an error typology. (p. 58)

Pour ce faire, il convient à nouveau de « s’entendre sur une typologie des erreurs qui ne soit ni floue ni subjective [Flanagan, 1994] » (citée dans Hartley et Popescu-Belis, 2004, p. 321) puis de décider de la gravité des erreurs et éventuellement d’une pondération⁵².

Bien entendu, cette approche présente son lot de difficultés, qui valent également en TH :

Manual error classification [of errors in MT output] [...] is a very difficult task for several reasons. The main disadvantages are high costs in terms of time and money, as well as low consistency, especially if the error categories are numerous and complex. In addition, defining an appropriate error typology represents a challenging task itself. Ongoing work [...] aims at generalisation by offering a large typology from which an appropriate sub-set could be selected for the task at hand. (Popović, 2018, p. 154)

Peraldi (2016) insiste sur le besoin de définir des frontières précises entre les différentes catégories d’erreurs et d’éviter toute difficulté liée au recensement des erreurs de traduction, car « [n]on seulement plusieurs erreurs peuvent survenir sur un même mot ou segment, mais chaque erreur peut entraîner à son tour une chaîne d’erreurs sur l’ensemble de la phrase source » (p. 74).

Malgré ces difficultés, étant donné que, comme le précise Martikainen (2019c), « l’analyse des erreurs a généralement pour objectif l’amélioration du moteur de traduction automatique » (p. 37), l’approche typologique représente sans nul doute une aide précieuse pour les développeurs : « Categorizing errors is an important step in diagnosing problems in MT output, often in an effort to provide feedback to system developers » (Kenny, 2022, p. 58).

Comme il s’agit d’une tâche difficilement automatisable, la classification des erreurs en TA incombe la plupart du temps à des annotatrices humaines :

[E]rror classification is usually conducted manually because currently available tools are still not able to distinguish detailed error classes, and are prone to confusions between mistranslations, omissions and additions. This task is usually performed by annotators who identify the errors of the MT output with or without a reference translation. (Álvarez-Vidal et al., 2021, p. 134)

⁵² « Cette décision dépendra de la finalité de l’évaluation et va éventuellement attribuer une pondération différente à une même erreur selon son impact sur la compréhension, ou bien sûr le temps d’édition, ou encore sur la difficulté de correction des algorithmes » (Hartley et Popescu-Belis, 2004, pp. 321-322).

2.3.7.1.3 Discussion

Quelle que soit la méthode d’évaluation de la qualité, il est plutôt rare de trouver des études dans lesquelles une même traduction ou une même TA brute est soumise à évaluation par un groupe de traductrices professionnelles ou de linguistes, notamment en raison des coûts que cela représente (Castilho et al., 2018). En outre, Castilho et al. (*ibid.*) déplorent le peu d’études comprenant des explications détaillées sur les tâches d’évaluation humaine de la qualité, ainsi que le fait que faire appel à des traductrices professionnelles pour évaluer la qualité en TA soit l’exception plutôt que la règle. Doherty (2017) confirme cette tendance à faire appel à des évaluatrices non professionnelles pour évaluer les performances des systèmes de TA :

[T]he need for and value of having professional translators who are trained as evaluators is not widely accepted in the MT research community [...] Owing to limitations in project resources (including funding, time and ability to share data), there is, of course a tendency for researchers to ask students, fellow researchers and online communities to carry out TQA in MT research. (p. 141)

Malgré les limites connues de ces méthodes⁵³, l’évaluation humaine de la TA reste la clé de voûte de l’évaluation automatisée de la qualité : « while costly both in time and resources, human evaluation is required to provide a reliable feedback for measuring MT quality and progress, as well as to serve as a gold standard for development of automatic evaluation metrics » (Popović, 2021, p. 163).

2.3.7.2 Automatisation de l’évaluation en TA

2.3.7.2.1 Métriques d’évaluation automatique

Les évaluations humaines étant souvent jugées comme subjectives, lentes et coûteuses (Koehn, 2010⁵⁴ ; Hartley et Popescu-Belis, 2004 ; Rossi et Carré, 2022 ; Saldanha et O’Brien, 2014 ; Snover et al., 2006), « les chercheurs ont raffiné à l’extrême les critères de qualité linguistique, tout en cherchant des méthodes automatiques pour les évaluer » (Blanchon et Boitet, 2007, p. 37). Ainsi, dans le domaine de la recherche en TA, une solution alternative à

⁵³ « Human evaluation of MT is resource-intensive and it can take much longer than human evaluation of human translations, because MT output quality can vary to a greater extent. Within the MT community, there is a growing awareness of other limitations of human evaluations as currently conducted in MT research. These include related to force-choice dichotomies of good vs bad translation [...] and the use of interval scales [...] Moreover, the subjectivity of human evaluators has also been highlighted » (Doherty, 2017, p. 134).

⁵⁴ « Cost is the major disadvantage of evaluation metrics that include human evaluators, especially bilingual evaluators » (Koehn, 2010, p. 220).

l’évaluation humaine de la qualité est le recours à des algorithmes d’évaluation automatique (Rossi et Carré, 2022) : « [I]n MT labs, reference-based automatic metrics—of which there is a multitude [voir González, Giménez and Girona Salgado (2014)] —have tended to prevail in the evaluation of machine output » (Kenny, 2018, p. 65). Koehn (2020) confirme cette tendance : « Much progress has recently been made in this field, to the point that machine translation researchers trust automatic evaluation metrics and design their systems based on the rise and fall of automatic evaluation scores » (p. 53).

2.3.7.2.1.1 Principe de fonctionnement

Ces métriques reposent toutes sur le même principe ; il s’agit de calculer la distance ou la similarité entre la ou les traduction(s) de « référence » produite(s) par un humain et la traduction « candidate » générée par le moteur de TA :

The evaluation tool is fed both a human generated “gold standard” or reference translation, and the system output, known as the candidate translation [...]. It then compares the candidate against the reference translation and computes the similarity or dissimilarity. To take variation across reference translations into account, some evaluation tools can be fed multiple reference translations. (Rossi et Carré, 2022, p. 59)

Selon ce principe, le degré de similarité avec la référence est un indicateur de qualité (Saldanha et O’Brien, 2014). Concrètement, l’outil va comparer des séquences de *n grammes* qui sont des séquences soit de *n* mots, soit de *n* caractères en fonction de la métrique comme l’expliquent Rossi et Carré (2022) :

When we discuss AEMs [automatic evaluation metrics], n-grams are merely n-word sequences in the candidate translation that also occur in the reference translation. More recently, AEMs have been proposed which consider sequences of characters instead of words. N-grams are then understood as sequences of n characters, rather than sequences of n words. (p. 63)

Rossi et Carré (*ibid.*) incitent à la prudence dans l’interprétation des résultats, car les scores ainsi obtenus peuvent être exprimés en décimales ou en pourcentage et la signification de l’échelle de valeurs varie selon la métrique ; le score parfait sera tantôt 1 (avec BLEU par ex.), tantôt 0 (avec TER).

2.3.7.2.1.2 Aperçu des principales métriques

Ces vingt dernières années, on recense bon nombre de métriques d’évaluation automatique – BLEU, NIST, ROUGE, METEOR, chrF – dont nous vous proposons un tour d’horizon. Pour un aperçu plus détaillé de ces métriques, voir notamment Koehn (2020) et Poibeau (2019).

BLEU

Développé par IBM en 2002, BLEU (*Bilingual Evaluation Understudy*) (Papineni et al., 2002) demeure étonnamment la métrique la plus connue et la plus utilisée en TA (Koehn et Monz, 2006 ; Loock, 2018 ; Way, 2018) en dépit des limites connues et des critiques qui ont été faites (Poibeau, 2019) : « [I]t is somewhat surprising that despite over a decade of metrics research, the go-to metric is still the BLEU score, which was the first automatic metric commonly used in data-driven machine translation research » (Koehn, 2020, p. 64). Selon le principe que nous venons de voir, le score BLEU mesure l’adéquation et la fluidité d’une « traduction candidate » (générée par le système de TA que l’on souhaite évaluer automatiquement) par rapport à une ou plusieurs traductions de référence. Un tel score est obtenu grâce à la mesure de la précision (*precision measure*)⁵⁵ en comparant des séquences de n mots appelées *n-grammes* (Papineni et al., 2022) et est toujours compris entre 0 et 1 (parfois rapporté sur 100). Plus le score se rapproche de 1, plus la TA est considérée comme étant de bonne qualité. Les développeuses jugent ordinairement qu’avec un score BLEU en dessous de 0,2, la TA brute ne vaut pas la peine d’être post-éditée (Toral et Way, 2015). De plus, il est généralement admis qu’avec un score au-dessus de 0,3, la TA peut être considérée comme compréhensible et au-dessus de 0,5, celle-ci peut être considérée comme fluide et de bonne qualité (Lavie, 2011). Parmi les nombreuses critiques adressées à cette métrique, il en est une que nous tenons à mentionner ; il est reproché à BLEU de « traiter tous les mots de la même manière sans tenir compte du fait que certains mots peuvent avoir un impact plus important que d’autres sur la qualité de la traduction (l’oubli ou l’ajout d’une négation, par exemple, peut modifier profondément le sens de la phrase) » (Volkart, 2018, p. 29). Pour un résumé des principales critiques émises à l’égard de cette métrique, se référer à Koehn (2020).

NIST

La métrique NIST (*National Institute of Standards and Technology*) (Doddington, 2002) a été développée au même moment que BLEU. Ces deux mesures sont donc très similaires, « [I]a principale différence réside dans la prise en compte d’un facteur d’informativité. Plus un segment est rare, plus son poids est élevé » (Poibeau, 2019, p. 159).

ROUGE

⁵⁵ « [P]recision is usually understood as the fraction of words in the MT output that are correct (as indicated by the gold-standard reference translation) » (Doherty, 2017, p. 134).

La métrique ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) (Lin, 2004) est surtout utilisée pour l’évaluation de résumés générés automatiquement.

METEOR et COMET

La mesure METEOR (*Metric for Evaluation of Translation with Explicit ORdering*) (Banerjee et Lavie, 2005) tente de « mieux prendre en compte les particularités linguistiques des textes analysés » (Poibeau, 2019, p.159). Cette métrique ne repère pas uniquement les correspondances entre *n-grammes*, mais tient aussi compte des variantes morphologiques et des synonymes (Banerjee et Lavie, 2005). La dernière version prometteuse qui succède à cette métrique est l’outil de prédiction COMET (*Crosslingual Optimized Metric for Evaluation of Translation*) (Rei et al., 2020) :

COMET, a neural framework for training multilingual machine translation evaluation models which obtains new state-of-the-art levels of correlation with human judgements. Our framework leverages recent breakthroughs in cross-lingual pretrained language modeling resulting in highly multilingual and adaptable MT evaluation models that exploit information from both the source input and a target-language reference translation in order to more accurately predict MT quality. (ibid., p. 2685)

chrF

La mesure chrF (*character n-gram F-score*) (Popović, 2015) mesure le degré de similarité entre la sortie de la TA et la ou les traduction(s) de référence en calculant un score F^{56} non pas sur des *n-grammes* de mots, mais sur des *n-grammes* de caractères.

2.3.7.2.1.3 Des métriques fiables ?

D’aucunes se montrent perplexes quant à la relative simplicité de telles métriques par rapport à la complexité de la tâche de traduction. Pour Poibeau (2019), cet écart de complexité serait justement le reflet de la principale difficulté qui sous-tend l’évaluation de la qualité en traduction, à savoir la difficulté à « formaliser la notion de bonne traduction et, plus généralement, difficulté à formaliser les notions de sémantique et de style » (p. 161).

Les spécialistes partent du principe qu’une métrique fiable est une métrique dont les résultats corréleront avec l’évaluation humaine (Volkart, 2018). Effectivement, il est communément admis en TA de considérer le degré de corrélation entre les résultats d’une métrique et les

⁵⁶ Le score F combine les mesures de précision et de rappel (*recall*) : « Often precision in metrics is paired with recall, which would compute the ratio of the n-grams in the reference translation that match the machine translation. [Precision and recall] can be combined into a metric called the f-measure » (Koehn, 2020, p. 54).

évaluations faites par des expertes humaines comme le meilleur gage de fiabilité d'une métrique (Poibeau, 2019). Mais il ne s'agit pas du seul critère utilisé pour juger de la qualité d'une métrique. Le critère d'interprétabilité du score est également pris en compte :

I already argued that interpretability is a positive property of a metric. For instance, the inability to communicate what a BLEU score of, say, 34.2 means is a problem for machine translation research. TER scores that can be roughly characterized as ratio of words that need to be changed are much more intuitive. (Koehn, 2020, p. 64)

2.3.7.2.1.4 Des métriques controversées

Les métriques automatiques ont reçu un accueil pour le moins sceptique et ces outils sont toujours fréquemment décriés en particulier dans le monde de la traductologie (Saldanha et O'Brien, 2014) : « automatic evaluation metrics are under constant debate and their true value in distinguishing better and worse systems is often called into question » (Koehn, 2020, p. 53). D'après Popović (2021)⁵⁷, ces métriques ne sont au demeurant qu'un pâle substitut de l'évaluation humaine en dépit des nombreuses améliorations apportées ces dernières années. Loin d'être une panacée et souvent considérées comme moins fiables et moins précises⁵⁸ que les méthodes d'évaluation humaine, ces méthodes d'évaluation automatisée présentent néanmoins l'avantage d'être « plus objectives, beaucoup plus rapides et bien moins coûteuses » (Volkart, 2018, p. 30), ainsi que particulièrement utiles pour la recherche et le développement : « [a]lthough such means of assessing quality is far from the sophistication of human judgment, it provides a *quick and dirty* solution that is especially valuable in research and development » (Doherty, 2016, p. 959), mais nous y reviendrons.

Parmi les nombreuses limites de ces outils, le problème fondamental réside dans le principe sur lequel ils se fondent : la qualité d'une traduction produite par un moteur de TA est évaluée par comparaison avec une ou plusieurs traductions de référence produite(s) par un humain. Pour Peraldi (2016), il s'agit de la faiblesse principale de ces méthodes automatisées, dès lors que « l'utilisation d'une traduction de référence oblige à un positionnement artificiel par rapport à un idéal de traduction, qui est le plus souvent décontextualisé (p. 74). Cette référence à un idéal de traduction se heurte directement à la « variabilité constitutive des traductions » (Braunstein, 2016, p. 63). Et comme l'affirme Way (2018): « [A]s any pair of

⁵⁷ « While better and better automatic metrics are constantly emerging (Mathur et al., 2020; Ma et al., 2019), many of them being based on semantic word representations (embeddings), all of them represent only an approximate substitution for human assessment of translation quality » (Popović, 2021, p. 163).

⁵⁸ « Automatic error analysis is faster, cheaper, and more consistent, yet state-of-the-art tools are still not able to provide many details » (Popović, 2018, p. 154).

translators will tell you, there is no such thing as *the* correct translation » (p. 166). Il arrive d’ailleurs que ces métriques évaluent négativement de bonnes traductions :

The main problem that affects all these metrics is that they rely on comparison with a limited set of reference sentences, and penalize perfectly acceptable translations that are different from the references. This fact accounts for the low scores assigned to human translations, as well as for the low correlation with human judgements observed in specific settings. (Bouillon, 2012, p. 5)

Rossi et Carré (2022) soulignent également ce paradoxe : « there is usually more than one valid solution in translation: the same source text can have several translations, all equally acceptable » (p. 53). Or, il est impossible qu’une traduction obtienne le score parfait de 1 lorsque

plusieurs traductions de référence humaines sont [...] prises en compte dans l’évaluation. De même, la traduction de référence humaine unique utilisée dans la méthode TER n’est évidemment qu’une possibilité parmi d’autres. Enfin, dans HTER et les autres méthodes reposant sur la post-édition, toute modification effectuée par le post-éditeur est comptabilisée comme une erreur lors de l’évaluation automatisée sans tenir compte d’éventuelles modifications préférentielles ou stylistiques. (Martikainen, 2019c, p. 37)

Par conséquent et de manière ironique, « MT developers are forced to (wrongly) assume human translations to be perfect when conducting automatic MT evaluation » (Way, 2013, p. 2).

De plus, Saldanha et O’Brien (2014) font remarquer un autre problème ; il est parfois difficile de savoir si la qualité de cette ou de ces traduction(s) humaine(s) a fait l’objet d’une évaluation avant que celle(s)-ci serve(nt) de référence.

Un autre problème bien connu est le manque de nuances de ces métriques : « [S]tate-of-the-art tools are still not able to provide many details. [...] Manual error classification can provide more detail as human annotators can distinguish a larger number of error classes than state-of-the-art automatic tools » (Popović, 2018, p. 154), ainsi que le manque de pertinence : « metrics like BLEU exploit mainly surface-matching characteristics that are largely insensitive to more subtle nuances » (Álvarez-Vidal et al., 2021, p. 133). Toudic et al. (2014) ont parfaitement résumé ces limites relatives à l’utilisation de critères réducteurs :

Ces critères ne permettent pas en effet de rendre compte de la variabilité de la traduction (les nombreuses manières d’exprimer un sens identique dans la langue cible), des déplacements ou des phénomènes de transposition (distinguer les déplacements de mots ou de groupes de mots qui modifient le sens de la phrase de ceux qui laissent le sens inchangé) ou du fait que tous les mots n’ont pas la même importance dans la compréhension de la phrase et que cette importance peut changer en fonction du contexte applicatif. (p. 25)

Il est difficile de croire que de tels outils d’évaluation soient capables de refléter de quelconques différences sémantiques entre la TA brute et la ou les traduction(s) de référence ou même le texte source (Koehn, 2010, pp. 228-229). Loock (2018) ajoute que ces métriques automatiques sont toujours la cible de critiques « puisque l’évaluation, qui se concrétise sous forme d’un score, est indépendante des langues source et cible, et n’évalue que la forme et non le contenu (Hartley et Popescu-Belis 2004 ; Koehn 2010) par le biais de comparaison entre *n-grammes* (séries de n mots) » (p. 789).

Par ailleurs, l’étude de Shterionov et al. (2018) a montré que les métriques automatiques sous-estimaient la qualité de la TA neuronale par rapport aux évaluations faites par des juges humaines : « we show that F-measure, BLEU and TER scores do not always conform with NMT quality. Rather, they underestimate NMT quality » (p. 15).

Nous souhaitons évoquer une dernière limite de ces métriques qui dépend de la nature des langues évaluées. Poibeau (2019) explique en détail ce biais des systèmes d’évaluation associé aux langues agglutinantes telles que le hongrois, le japonais et le turc par rapport aux langues analytiques telles que le chinois, l’anglais et le français :

[T]outes les mesures automatiques cherchent à repérer des séquences plus ou moins longues identiques à une ou des traductions de référence, les langues agglutinantes sont clairement désavantagées dans la mesure où elles concentrent dans une même forme de multiples informations. Le désavantage est double : les langues analytiques [...] présentent davantage de séquences longues susceptibles d’améliorer les scores d’évaluation [...] tandis qu’à l’opposé les langues agglutinantes présentent des formes simples complexes qui sont par conséquent difficiles à générer de manière complètement exacte. (pp. 166-167)

2.3.7.2.1.5 Des métriques non sans intérêt

L’évaluation de la qualité en TA est déterminante et incontournable comme le souligne Koehn (2020) : « From an engineering perspective, as we embark on considering different machine translation methods, evaluation is also a crucial step. To check if we are making progress, we need to be able to measure quality » (p. 41). C’est pourquoi, en dépit de ces écueils et faute

de mieux (Rei et al., 2020)⁵⁹, les métriques automatiques présentent une utilité certaine pour les conceptrices et les éditrices de logiciels, mais aussi pour les scientifiques puisqu’elles leur permettent de « mesurer les performances et l’évolution des systèmes au cours du temps [...] des évolutions qui correspondent généralement à la perception de la qualité globale des systèmes » (Poibeau, 2019, p. 163).

2.3.7.2.2 Méthodes de classification automatique

On recense aussi certaines tentatives d’automatisation de classification des erreurs en TA, tels que les outils *Hjerson* (Popović, 2011) ou *Addicter* (Zeman et al., 2011), sur lesquelles Popović (2018) porte une attention particulière :

automatic methods for error classification emerged due to resource – and time – intensity, as well as the inconsistency of the manual process. The motivation is the same as for AEMs, namely use a program to compare the translation output with a reference translation. The goal, however, is not to produce a single overall score, but to estimate the amount of different error types. [...] Automatic error analysis is faster, cheaper, and more consistent, yet state-of-the-art tools are still not able to provide many details. In addition, existing tools are prone to confusion between certain error classes, although some of these distinctions are not easy even for human evaluators. (p. 142 et p. 154)

Certes, comme le souligne Popović (2018, p. 154), ces outils sont imparfaits, mais ils peuvent néanmoins constituer une solution alternative à l’évaluation humaine en TA puisqu’ils permettent d’obtenir une répartition des types d’erreurs présents dans une TA ou de comparer différentes sorties de TA. Enfin, toujours selon Popović (*ibid.*), de tels outils s’avèrent également profitables à l’évaluatrice humaine en lui permettant de disposer d’un texte pré-annoté : « Apart from this, [automatic error classification] can facilitate manual error classification by introducing a pre-annotation step; correcting or expanding existing error tags requires less effort and time than assigning error tags to an unannotated text from scratch » (p. 154).

⁵⁹ « While an increased research interest in neural methods for training MT models and systems has resulted in a recent, dramatic improvement in MT quality, MT evaluation has fallen behind. The MT research community still relies largely on outdated metrics and no new, widely-adopted standard has emerged » (Rei et al., 2020, p. 2685).

2.3.7.3 Stratégies d'évaluation axées sur la tâche

Outre les stratégies traditionnelles axées essentiellement sur le développement des systèmes de TA, une nouvelle tendance dans le domaine de l'évaluation des technologies et des applications gagne aujourd'hui en popularité : les stratégies d'évaluation axées sur la tâche.

Depuis quelques années, en raison de l'automatisation croissante des processus de traitement de l'information et de la diffusion des dispositifs techniques vers un plus large public, l'évaluation des technologies et des applications suscite un intérêt accru correspondant à l'enjeu de l'acceptabilité sociale de la technique. Dépassant le cadre de l'évaluation de la performance intrinsèque des systèmes, la question de l'évaluation déborde ainsi sur celle des usages et des comportements individuels et collectifs face à des artefacts techniques. (Chaudiron, 2004, p. 17)

Rémillard (2018) explique que ces stratégies peuvent être conçues pour répondre à des besoins d'évaluation spécifiques ; il s'agit de méthodes « variées, car elles sont déterminées par le contexte d'évaluation et les objectifs particuliers » (p. 54). Ces méthodes se fondent sur différents critères, tels que l'acceptabilité, l'utilité et la productivité, et permettent d'évaluer dans quelle mesure la TA soutient les utilisatrices dans la réalisation d'une tâche spécifique (*ibid.*, p. 54) :

Task-oriented evaluation metrics attempt to directly measure the utility of machine translation for performing a specific task. If the task is to produce high-quality translations, we may measure the postediting time required to correct machine translation output. In a similar vein, the translation error rate [TER] measures the number of editing steps required to reach a reference translation, and the human translation error rate [HTER] measures how many editing steps a human post-editor has to perform to reach an acceptable translation. (Koehn, 2010, p. 241)

TER

Le score TER (*Translation Error Rate* ou *Translation Edit Rate*) (Snover et al., 2006) mesure « le coût minimum de la suite d'opérations d'édition (insertion, suppression, substitution, déplacement) permettant de transformer une traduction candidate en une traduction de qualité suffisante [traduction de référence] pour une certaine tâche » (Blanchon et Boitet, 2007, p. 52).

HTER

HTER (*Human-targeted Translation Error Rate*) (Snover et al., 2009) est une amélioration de TER. Une des spécificités est que cette nouvelle métrique « associe des postéditeurs humains dans le processus ("human-in-the-loop evaluation") » (Blain, 2013, p. 41). Alors que TER fait

appel à une traduction de référence, HTER évalue la distance d’édition entre une traduction candidate générée automatiquement et une version post-éditée par une traductrice humaine (qui servira de traduction de référence), ainsi cette méthode consiste à

demander à des traducteurs humains parlant couramment la langue cible, de générer une traduction de référence (dite référence ciblée), la plus proche possible de l’hypothèse de traduction et devant partager la même sémantique que des traductions de références présélectionnées au départ (dites références non ciblées). Les annotateurs ont le choix de repartir de l’hypothèse de traduction ou d’une des traductions de références présélectionnées. (*ibid.*)

Blanchon et Boitet (2007) précisent qu’une autre particularité de la mesure HTER par rapport à une mesure classique de distance d’édition est que « les opérations d’édition usuelles [...] s’appliquent non seulement à des mots, mais à des séquences de mots » (pp. 51-52). Supposée être un indicateur fiable de l’effort de PE, cette mesure comporte bien évidemment des failles (Daems, 2016) :

The underlying assumptions when using HTER are that HTER is an indication of actual post-editing effort (Specia & Farzindar, 2010), which implies that all edits made to MT output by a human are expected to require a comparable amount of effort. However, HTER focuses on the final product, without taking the actual process into account, so its relationship to post-editing effort is questionable. In that sense, HTER is a little bit like a car's GPS without intelligent feedback: it provides you with the shortest way to get from point A to point B, but it does not take any obstructions or personal factors into account. For example, a post-editor can return to the same phrase multiple times during the post-editing process, changing that particular phrase each time, but settling on one specific solution in the end. HTER will indicate how different this final solution is from the original MT output, but it does not take into account all edits made during the process. (p. 116)

2.3.8 Évaluation des nouveaux outils

Avec l’arrivée récente d’outils conversationnels tels que *ChatGPT* (OpenAI, 2022) qui s’appuie sur la dernière version du modèle de langue d’OpenAI : GPT-4 ou encore d’outils d’aide à la rédaction, tels que *DeepL Write* (DeepL GmbH, 2023), se pose la question des modalités d’évaluation. En effet, comment juger de l’exactitude des formulations/traductions proposées ou de la capacité de reformulation de ces nouveaux outils susceptibles d’être intégrés au flux de travail de la traductrice ?

Nous souhaitons conclure cette partie consacrée à l'évaluation de la traduction en faisant référence à Doherty et al. (2018). Dans leur article « *On Education and Training in Translation Quality Assessment* », elles invitent en particulier le milieu universitaire à reconnaître et à combler le manque de considération des diverses métriques d'évaluation de la qualité (humaine et machine) employées dans le milieu professionnel. Or, il est essentiel, argumentent-elles, que les formations en traduction intègrent cet aspect :

regardless of the debates on translation quality in academia, the industry will continue to have its own TQA metrics and models that will, in turn, evolve as the industry changes, largely dependent on market trends and technological developments. The challenge for lecturers and trainers is to understand the dynamics at play in the choice and use of these TQA metrics and models, so that their students can learn to appreciate their value while also being aware of their limitations and potential pitfalls. Equipped with this knowledge, human translators, especially university educated ones, can enter the industry with confidence in their value. (ibid., p. 100)

3 Méthodologie

Dans cette thèse, afin de répondre à notre première question de recherche (A)⁶⁰, nous avons choisi d'adopter une méthode expérimentale. Notre but étant d'étudier des relations de cause à effet, de tester des hypothèses en cherchant à démontrer ou à réfuter un lien de causalité entre un facteur et un résultat observé (Oates, 2005, p. 35), nous avons mené une première expérience contrôlée en 2018 (ci-après « prétest ») et une deuxième expérience contrôlée en 2021 (ci-après « test ») avec des étudiantes en traduction. En traductologie, l'expérience contrôlée permet au scientifique d'explorer des relations de causalité dans un environnement contrôlé :

[I]t compares two groups and their properties of behaviour when certain variables are manipulated. It may be designed in such a way that there is an "experimental group" [...] and a "control group" [...]. Note that the groups are not necessarily populated by humans, but can also be composed of texts, for example. (Saldanha et O'Brien, 2014, p. 15)

En outre, nous avons également adopté une méthode de recherche par sondage pour nous permettre de répondre à notre deuxième question de recherche (B)⁶¹. Pour explorer les connaissances et les perceptions qu'ont les étudiantes en traduction sur la TA et sur la PE, nous leur avons soumis deux courts sondages combinant questions ouvertes et questions fermées qui feront l'objet du chapitre 7.

Dans ce chapitre méthodologique, nous nous concentrons sur la méthode expérimentale développée afin de comparer les produits de la traduction humaine aux produits de la post-édition en contexte d'apprentissage pour la combinaison de langues anglais-français. Nous commençons par expliciter les différents paramètres d'expérience, notamment en identifiant les variables intervenant dans nos expériences, en décrivant le profil des participantes et en exposant la méthode de constitution du corpus. Puis nous décrivons le déroulement de chaque expérience et terminons par présenter la stratégie d'exploitation des données récoltées pour laquelle nous avons opté.

⁶⁰ Question A - Quels sont les effets de la post-édition (anglais-français) en contexte d'apprentissage sur la qualité d'un texte cible ?

⁶¹ Question B - Quelles sont les connaissances et les perceptions des étudiantes en traduction sur la traduction automatique et sur la post-édition ?

3.1 Méthode expérimentale

3.1.1 Variables

3.1.1.1 Variable dépendante (V.D.)

En science empirique, la variable dépendante est définie comme « the core concept we are trying to assess in our research question. We expect it to change when it is exposed to varying treatment » (Saldanha et O'Brien, 2014, p. 25). Dans la configuration de nos expériences, la qualité des productions des étudiantes constitue la variable dépendante. Cette variable est mesurée tantôt par le nombre d'erreurs relevées dans les productions des étudiantes, tantôt par les notes attribuées aux productions des étudiantes selon une pondération prédéfinie.

3.1.1.2 Variables indépendantes (V.I.)

Les variables indépendantes « are things that we manipulate in order to see what the effect is on our dependent variables » (*ibid.*).

Dans nos expériences, les variables indépendantes sont au nombre de trois :

- le mode de traduction ;
- la thématique⁶² du texte source ;
- l'étudiante

Ces variables indépendantes n'ont toutefois pas le même statut ; nous avons deux facteurs contrôlés (*fixed factors*) : le mode de traduction (TH et PE) et la thématique du texte source, et un facteur aléatoire (*random factor*) : les étudiantes. Effectivement, nous nous attendons à ce que la qualité des productions dépende logiquement de variations individuelles d'une étudiante à l'autre (telles que le niveau de maîtrise de l'anglais et du français, le niveau d'expérience en traduction et en PE, le degré de familiarité avec la thématique des textes sources, etc.).

⁶² Dans notre étude, les textes sources relèvent du même type textuel (au sens de Reiß et Vermeer (2014), voir section 3.1.5.1), nous les distinguons donc en fonction de leur thématique.

3.1.1.3 Variables parasites (V.P.)

Outre la variable dépendante et plusieurs variables indépendantes, notre recherche présente également toute une série de variables parasites qui peuvent influencer la validité interne⁶³ de notre étude. Il s'agit

[d']autres faits ou d'autres facteurs, invisibles, mal identifiés ou plus ou moins bien contrôlés, [qui] peuvent intervenir dans notre modèle de relation causale, pervertir le sens des résultats et réduire la portée explicative de la ou des variables indépendantes. Ces déterminants ou variables parasites (V.P.) peuvent provenir des caractéristiques de la population étudiée, de l'organisation des tâches dans lesquelles on l'engage, et plus généralement des conséquences du pluri ou multi déterminisme des comportements : on considère en effet qu'il est impossible de faire l'inventaire exhaustif des causes multiples qui déterminent un état biologique complexe ou un comportement. (Anceaux et Sockeel, 2006, p. 75)

Nous sommes parvenue à identifier plusieurs variables parasites potentielles dans nos expériences sans pour autant que la liste ci-dessous soit exhaustive :

- la complexité du texte source ;
- les perceptions de l'étudiante à l'égard de la traduction automatique et de la post-édition (enthousiaste ou réfractaire ?) ;
- la pression liée à la participation à une recherche expérimentale et à l'annotation de leurs erreurs ; ce phénomène est appelé l'effet Hawthorne :

which occurs when people alter (usually improve) their normal behaviour because they are aware that they are being studied. Participants will often respond according to how they think the researcher would like them to respond or they might select what they think is the "nicest" answer because it will reflect well on them. (Saldanha et O'Brien, 2014, p. 153)

- ...

3.1.2 Plan expérimental

Étant donné la présence de plusieurs variables indépendantes, notre étude a été menée selon un plan factoriel à mesures répétées. Légal (2008) définit ce type de plan factoriel comme suit :

⁶³ « [L]a validité interne étant la possibilité d'attribuer les valeurs obtenues de la V.D. aux variations de la V.I. ; c'est-à-dire la possibilité de définir une relation causale entre la V.I. et la V.D. » (Gangloff, 2008, p. 15).

Dans un contexte naturel, un comportement, un état mental ou un processus mental est rarement le produit d'une seule cause, mais plutôt celui de la combinaison de plusieurs causes ou facteurs ou variables. À partir du moment où le chercheur va vouloir s'intéresser aux éventuelles interactions des variables qu'il manipule, il va avoir recours à ce que l'on appelle un plan factoriel (ou fischérien). [...] Un plan factoriel étant caractérisé par la présence d'au moins deux VI [variables indépendantes], il peut se décliner selon trois formes en fonction du type des groupes de mesures. (p. 18)

Dans notre cas, il s'agit d'un plan factoriel dit « à mesures répétées » (ou « à groupes appariés »). On parle de plan à mesures répétées quand « tous les sujets passent par toutes les conditions expérimentales. La comparaison s'effectue sur un même groupe de participantes qui se prête à toutes les modalités de la V.I. Autrement dit, lorsque tous les sujets subissent toutes les modalités de[/des] la V.I. » (*ibid.*, p. 17). Effectivement, dans notre recherche, nous pouvons dire que nos échantillons sont appariés (non indépendants), car dans chaque expérience contrôlée, ce sont les mêmes étudiantes qui ont systématiquement été exposés à toutes les modalités des V.I. (mode de traduction et texte source).

3.1.3 Participantes

Le profil des participantes au prétest diffère expressément du profil des participantes au test, ceci dans le but de voir si l'expérience en traduction des étudiantes joue effectivement un rôle sur nos observations. Bien que, comme nous le verrons, certaines participantes aient une certaine expérience en PE, nous sommes partie du principe que toutes les étudiantes participantes sont bien plus familières de la TH que de la PE.

3.1.3.1 Participantes au prétest

Les participantes au prétest sont 28 étudiantes inscrites en 3^e année de bachelier traduction-interprétation de l'Université de Liège en 2017-2018. Leur langue maternelle est le français, à l'exception d'une étudiante germanophone. Leur participation s'est faite sur une base volontaire à leur retour de séjour Erasmus. Ces 28 étudiantes avaient peu d'expérience en PE, voire pas du tout. En effet, aucune étudiante n'a reçu de formation spécifique dans ce domaine avant de prendre part à cette recherche expérimentale, l'une des raisons étant que l'Université de Liège n'offrait à ce moment-là aucune initiation à la TA et à la PE. Cette expérience reflète donc un usage individuel sans encadrement. Comme le recommande Depraetere (2010)⁶⁴, il nous a néanmoins semblé opportun d'examiner les stratégies intuitives de PE de ces

⁶⁴ « [B]efore teaching students do's and don'ts [...] it is useful to look at what they [students] intuitively get right and where they go wrong and in this way give due attention to potential post-editing pain points » (Depraetere, 2010, p. 3).

étudiantes, et ce, d'autant plus que, parmi les traductrices professionnelles à qui l'on confie aujourd'hui des tâches de PE, elles sont peu nombreuses à avoir reçu une formation spécifique en la matière⁶⁵. Il convient d'ajouter que, dans le but d'initier ces étudiantes à la TA et à la PE, nous avons prévu une courte séance d'introduction de deux heures en amont de cette première expérience (survol historique de la TA et sensibilisation aux limites de cette technologie).

3.1.3.2 Participantes au test

Avant de détailler le profil des participantes au test, mentionnons l'intégration progressive de la TA et de la PE dans l'enseignement de la traduction à l'Université de Liège dès l'année académique 2019-2020 (soit durant la période qui sépare nos deux expériences) dont nous reparlerons au chapitre 8.

Le test a été mené cette fois avec 24 étudiantes de 2^e année de master en traduction de l'Université de Liège inscrites en 2020-2021, dont la langue maternelle est le français. Sans être des post-éditrices aguerries, ces étudiantes en fin de parcours universitaire ont déjà toutes eu l'occasion de faire de la PE au cours de leur cursus (que ce soit dans le cadre d'un cours, d'un séminaire ou de leur propre initiative) sans pour autant avoir reçu de formation spécifique. Toutefois, il est difficile d'estimer leur niveau d'expérience réel étant donné que l'intégration d'exercices de PE dans leur cursus s'est faite progressivement et à tâtons, et uniquement dans certains cours.

3.1.3.3 Échantillonnage

La méthode parfaite d'échantillonnage n'existant pas, nous avons cherché à constituer des échantillons pertinents et représentatifs tels que Saldanha et O'Brien (2014) le recommandent : « Obtaining an ideal sample is very difficult, if not impossible, but the goal of the researcher is to reduce bias and increase validity by using the most appropriate sampling methods under the circumstances » (p. 35). Pour le prétest, comme pour le test, nous avons procédé à un échantillonnage non probabiliste : « this occurs when participants are selected purely on the basis of easy accessibility to the researcher » (*ibid.*, p. 164). Le choix des participantes s'est imposé naturellement puisque nous avons plusieurs cohortes d'étudiantes à notre disposition. Plus spécifiquement, dans notre cas, il s'agit d'un échantillonnage

⁶⁵ « As the development of technology has been very rapid, translators have often had to acquire new technical skills such as post-editing machine translations in continuous education or as part of informal, lifelong learning » (Salmi, 2021, p. 113).

consécutif, car nous avons recruté l'ensemble des étudiantes de 3^e bachelier pour le prétest et l'ensemble des étudiantes de 2^e master pour le test :

Consecutive sampling is very similar to convenience sampling except that it seeks to include ALL accessible subjects as part of the sample. This non-probability sampling technique can be considered as the best of all non-probability samples because it includes all subjects that are available that makes the sample a better representation of the entire population. (Castillo, 2009, s.p.)

3.1.3.4 Consentement libre et éclairé

Avant de prendre part à l'expérience, chaque participante a été invitée à prendre connaissance d'un formulaire de consentement⁶⁶ qui reprend les différentes modalités du projet d'étude (objectif du projet ; raison et nature de la participation ; droit de retrait sans préjudice de la participation ; confidentialité, partage, surveillance et publications), ainsi que les exigences et dispositions du Règlement Général relatif à la Protection des Données (RGPD) de l'Université de Liège, ainsi qu' à donner son accord de participation en signant ce document. Ce projet étant soumis aux règles d'application du RGPD de l'Université de Liège, la présente recherche empirique a fait l'objet d'une demande auprès de M. le Délégué à la protection des données de l'Université de Liège, Jean-François Pirlet, qui a procédé à son enregistrement dans le registre des traitements des données à caractère personnel de l'Université de Liège en date du 27 août 2020.

3.1.4 Tâches

Dans le but de comparer les produits de la TH aux produits de la PE de TA, il a été demandé à chaque étudiante de fournir trois traductions selon trois modes de traduction différents.

Dans le prétest, les étudiantes de 3^e bachelier ont produit :

- une traduction humaine ;
- une post-édition de TA statistique (*BingTranslator*) ;
- une post-édition de TA neuronale (*DeepL*).

Dans le test, les étudiantes de 2^e master ont produit :

- une traduction humaine ;
- une post-édition de traduction automatique neuronale (*Google Traduction*) ;
- une post-édition de traduction automatique neuronale (*DeepL*).

⁶⁶ Accessible sur <https://doi.org/10.7910/DVN/R8EJKN>

Pour la tâche de TH, l'étudiante pouvait disposer de toutes les ressources qu'elle jugeait utiles et nécessaires (dictionnaires explicatifs et traductifs, glossaires, Internet, etc.). Il lui était toutefois interdit d'utiliser tout moteur de TA et tout logiciel de TAO.

Pour les tâches de PE, les textes sources ont été traduits préalablement par le moteur de TA et nous avons demandé aux participantes de post-éditer ces sorties de TA.

3.1.4.1 Moteurs de TA

Concernant la sélection des moteurs de TA, nous avons tenu à travailler à partir de sorties d'outils de TA disponibles gratuitement en ligne, principalement car nos expériences ont été réalisées en contexte pédagogique.

Les différentes TA ont été générées en janvier 2018 pour le prétest et en septembre 2020 pour le test ; l'ensemble des TA brutes figure en Annexe A. Avec un rythme de recherche quasi continu (Poibeau, 2019, p. 148), le domaine de la TA ne cesse d'évoluer à une vitesse phénoménale (Breyel et Grass, 2021 ; Way, 2019) et nul ne peut garantir que les moteurs utilisés produiraient les mêmes TA brutes au moment où nous écrivons ces lignes.

3.1.4.1.1 Prétest

Dans notre prétest, nous entendons comparer la qualité de textes post-édités à partir de deux technologies de TA (statistique et neuronale) afin, notamment, de répondre à notre sous-question de recherche A3⁶⁷. Nous avons choisi d'utiliser deux moteurs de TA génériques grand public : *Bing Translator* comme système de TA statistique et *DeepL* comme système de TA neuronale. Contrairement aux outils spécialisés, les outils génériques ne sont pas entraînés sur la traduction d'un type de texte bien spécifique (Mion, 2021) : « generic MT, that is, MT based on engines trained to cover a wide range of topics, styles and genres, and not specialized in any particular domain » (Ramírez-Sánchez, 2022, p. 165).

Bing Translator (anciennement *Windows Live Translator*) est un service de TA fourni par Microsoft depuis 2007 qui propose des traductions dans plus de 100 langues différentes (au 25 janvier 2023). Jusqu'en mai 2018, *Bing Translator* était encore l'un des rares moteurs gratuits à proposer l'approche statistique, ce qui nous a permis de l'utiliser dans le cadre de notre prétest en février 2018. Ce service est désormais passé entièrement à la TA neuronale.

DeepL Traducteur est un moteur de TA lancé publiquement fin août 2017 par la société Linguee GmbH (aujourd'hui *DeepL GmbH*) dont le siège social est situé à Cologne (Allemagne). Disponible à l'origine uniquement pour 7 langues européennes, *DeepL* est un service de TA

⁶⁷ A3. La qualité de textes post-édités (anglais-français) en contexte d'apprentissage varie-t-elle en fonction de la technologie de traduction automatique utilisée (TA statistique ou TA neuronale) ?

neuronale mondialement connu qui supporte aujourd’hui 31 langues et plus de 800 combinaisons possibles (au 21 avril 2023). Nous avons décidé d’utiliser ce moteur de TA, car déjà en 2018, ses performances devançaient notamment les solutions similaires de *Google*, *Microsoft* et *Facebook* (Smolentceva, 2018) et aujourd’hui, il s’agirait du moteur de TA grand public le plus performant sur le marché (Savenkov et Lopez, 2022, voir Figure 10). En témoignent les résultats des évaluations humaines et automatiques effectuées par la société elle-même qui « montrent une nette préférence des traducteurs humains pour *DeepL* ainsi qu’un score BLEU nettement supérieur par rapport à d’autres systèmes » (Volkart, 2018, p. 42). Se considérant elle-même comme le « leader des avancées scientifiques dans le domaine de la traduction automatique », la société *DeepL* affirme en toute modestie proposer « le meilleur traducteur automatique au monde [en offrant] des résultats supérieurs à ceux de la concurrence (preuves à l’appui) » (*DeepL GmbH*, 2022, s.p.).

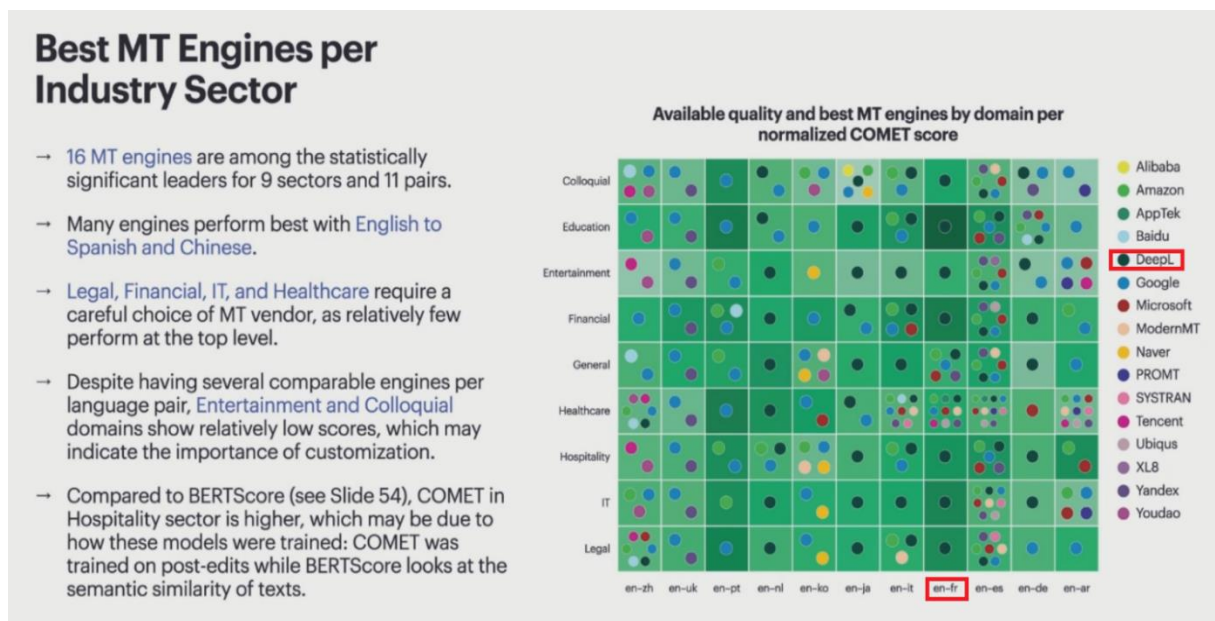


Figure 10 : Performances des moteurs de TA (Savenkov et Lopez, 2022)

3.1.4.1.2 Test

Dans notre test, afin de répondre à notre sous-question de recherche A6⁶⁸, nous entendons vérifier si la qualité des textes post-édités dépend du système de TA, en comparant indirectement les performances de deux outils génériques de TA neuronale disponibles en ligne. Pour ce faire, nous avons sélectionné *Google Traduction* et, à nouveau, *DeepL* ; les deux

⁶⁸ A6. La qualité de textes post-édités (anglais-français) en contexte d’apprentissage varie-t-elle en fonction du moteur de traduction automatique neuronale utilisé (*Google Traduction* ou *DeepL*) ?

moteurs de TA grand public les plus connus aujourd'hui et, sans surprise, les moteurs les plus fréquemment utilisés par les étudiantes (Loock et Léchaugette, 2021).

Google Traduction est le service de traduction automatique proposé par *Google* dès 2006. Il s'agissait au départ d'un système de TA statistique qui est passé au paradigme neuronal en novembre 2016 avec l'outil d'intelligence artificielle *Google Neural Machine Translation* (GNMT), entraînant d'importants progrès. *Google Traduction* supporte aujourd'hui 133 langues (au 21 avril 2023) et permet de traduire aussi bien du texte que des pages Web.

Précisons que, dans les deux expériences, nous avons eu recours à la version *DeepL* disponible en ligne gratuitement et non à la version professionnelle *DeepL Pro*.

3.1.4.2 Qualité attendue en post-édition

En ce qui concerne la qualité attendue en PE, il a été demandé aux étudiantes de réaliser une post-édition complète, telle que définie par la norme ISO 18587 (cf. chapitre 2), dans le but d'aboutir à une production de qualité comparable à la qualité d'une TH. Ceci nous permet d'assurer une certaine validité écologique⁶⁹ puisqu'en conditions pédagogiques réelles, il est attendu que les étudiantes produisent des traductions abouties, de haute qualité. Les étudiantes participantes ont reçu, pour ce faire, une liste de consignes de PE que nous avons reprises de TAUS (cf. nbp 24) et légèrement adaptées aux fins de notre étude (annexe B.1 Consignes étudiantes).

Pour rappel, TAUS/CNGL (2010) définit ce niveau de qualité comme « compréhensible (un lecteur comprend parfaitement le contenu du message), exact (la traduction a le même sens que le texte source) et stylistiquement correct [...]. La syntaxe est normale, la grammaire et la ponctuation sont correctes. » Nous avons toutefois pris le parti d'omettre une partie de la définition de ce niveau de qualité dans nos consignes, tant écrites qu'orales. Voici en gras le passage que nous avons décidé de ne pas mentionner : « [...] et stylistiquement correct, **même si le style n'est pas nécessairement aussi bon que celui obtenu par un traducteur humain dont la langue maternelle est la langue cible** » (*ibid.*, pp.3-4). Nous avons pris cette décision, car, comme l'ont très justement fait remarquer Flanagan et Christensen (2014), cette consigne émanant de TAUS est difficilement interprétable et difficilement applicable par les étudiantes, entre autres parce que le style est une « notion dont la subjectivité en traduction influence inévitablement la manière dont tout étudiant ou traducteur la percevra » (de Faria Pires,

⁶⁹ « [L]a validité qu'on qualifie d'écologique (ou de contextuelle) préconise que l'évaluation évite autant que possible de placer l'individu dans des conditions plus ou moins "artificielles", éloignées du contexte dans lequel les connaissances et compétences évaluées s'exercent et/ou acquièrent leur sens véritable » (IRDP, 2017, s.p.).

2020, p. 63). Ainsi, nous avons préféré ne pas mentionner ce passage soumis à l'arbitraire et jugé trop limitant aux fins de notre étude.

3.1.5 Corpus

En suivant les recommandations de Saldanha et O'Brien (2014), nous avons veillé à travailler sur un échantillon relativement large, avec systématiquement plusieurs versions du même texte source produites par plusieurs étudiantes. Par ailleurs, pour des raisons de comparabilité et de robustesse des résultats, nous avons utilisé les mêmes textes sources pour le prétest et pour le test.

3.1.5.1 Sélection des textes sources

Étant donné que nous entendons explorer les effets de la PE dans un contexte d'apprentissage, nous avons cherché, tant que faire se peut, à reproduire l'environnement des cours de traduction dispensés à l'Université de Liège de façon à obtenir la plus grande validité écologique possible. À cet effet et afin de limiter les biais dans la sélection des textes sources pour nos expériences, nous avons demandé à trois enseignantes de traduction (Université de Liège) de sélectionner chacune un texte d'une longueur comprise entre 300 et 340 mots. L'évaluation de la complexité textuelle en vue d'une traduction dépend de tant de facteurs divers et variés que nous avons préféré nous reposer sur l'expérience de ces enseignantes et leur avons demandé de sélectionner des textes qu'elles pourraient tout à fait traduire ou donner en préparation dans le cadre des cours qu'elles donnent aux étudiantes de la filière traduction-interprétation.

Les trois textes sources sélectionnés par leurs soins appartiennent au même type textuel. Il s'agit de textes généralistes à dominante informative, car ils ont avant tout une visée communicative. Selon Reiß et Vermeer (2014), le type textuel informatif est à distinguer du type expressif et du type opératif et se définit comme suit :

If an author wants the information offer to convey content, i.e. if the text has been composed with the intention of passing on news, knowledge, views, etc. (which can be assigned to the representational function of language), we speak of an informative text type. (p. 182)

Ces trois textes sources relèvent toutefois de trois thématiques distinctes, car nous voulions pouvoir dégager de potentielles variations de performances de la TA en fonction de ce facteur. Il s'agit d'un sujet d'actualité, d'un sujet économique et d'un sujet scientifique. Le premier texte source est un article de presse datant du 9 novembre 2017, il s'intitule « Can Theresa May and her government survive? Our writers' verdicts » (ci-après « texte 1 ») et est tiré du journal britannique *The Guardian*. Le texte économique s'intitule « How will the global

economic order change by 2050? » (ci-après « texte 2 »), il s’agit d’un extrait d’un rapport publié par la société d’audit *PricewaterhouseCoopers* (PWC) en février 2017 sur son site Internet www.pwc.com. Enfin, le troisième texte est un article de vulgarisation scientifique, « What is synesthesia? » (ci-après « texte 3 ») qui a été publié dans la revue américaine *Scientific American* le 11 septembre 2006, et traite de la synesthésie, qui est un trouble de la perception sensorielle. Bien entendu, nous avons vérifié qu’aucune traduction de ces textes n’avait été publiée en ligne au moment des expériences. Les trois textes sources sélectionnés ainsi que les traductions automatiques brutes sont repris en Annexe A.

Avec ces textes, nous nous situons donc dans le champ de la traduction pragmatique⁷⁰ ou « non-littéraire ». Ce champ « regroupe toute forme de traduction ayant une visée de communication » (Froeliger, 2013, s.p.) et avons dès lors demandé aux étudiantes d’adopter une approche de la traduction dite « communicative » :

Today, the ideal would seem to be what we call a ‘communicative’ translation, in which the target language offer of information ‘imitates’ the source text offer of information [...] In this translation type, the target text does not feel like a translation, at least not with regard to the language it uses; rather, it is a translation that can achieve the same functions as the original text and can be used directly for (everyday, literary or aesthetic) communication, as it is (as far as is possible) an equivalent of the original text with regard to all of its dimensions (syntax, semantics, pragmatics). (Reiß et Vermeer, 2014, pp. 124-125)

Constitution du corpus (Prétest) :

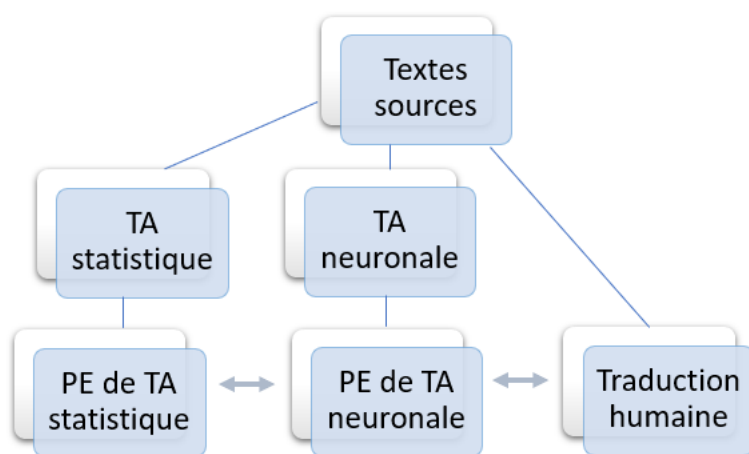


Figure 11 : Constitution du corpus (Prétest)

⁷⁰ « [À] savoir toute traduction non littéraire ou toute traduction se concentrant principalement sur la réception du texte par le destinataire » (Kübler, 2013, p. 4).

Constitution du corpus (Test) :

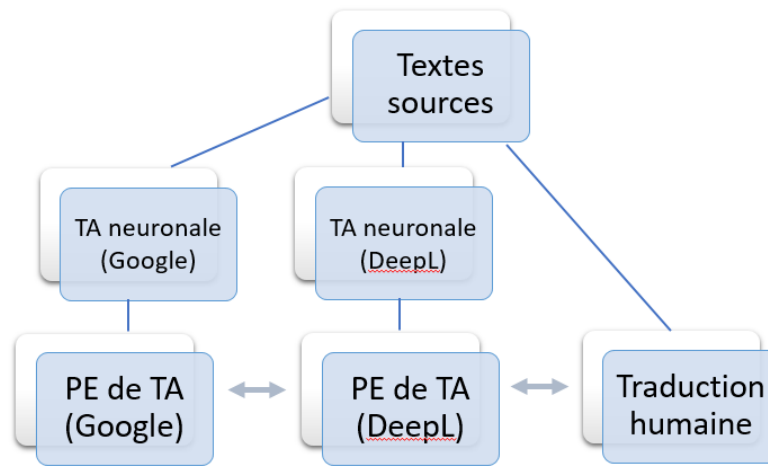


Figure 12 : Constitution du corpus (Test)

L'ensemble du corpus (TS, TA brutes et TC), ainsi que les fichiers d'annotation anonymisés sont stockés sur Dataverse⁷¹ et accessibles via le lien suivant: <https://doi.org/10.7910/DVN/R8EJKN>

3.1.5.2 Score BLEU

À titre informatif, nous avons tenu à calculer le score BLEU (Papineni et al., 2002) pour chaque texte source et pour chaque moteur de TA. Pour les diverses raisons déjà avancées, les scores donnés par ce type de métriques d'évaluation automatique sont en réalité difficilement interprétables et nous disent bien peu de choses sur la qualité réelle de la TA brute (Volkart, 2018) et sur l'effort de post-édition requis. En dépit des limites connues de BLEU, nous avons tenu à indiquer ces résultats à toutes fins utiles, puisque cette mesure est encore largement utilisée dans la pratique (Hansen, 2021 ; Volkart, 2018) et fréquemment citée dans la littérature scientifique portant sur des expériences menées en TA (Way, 2018). Pour une description détaillée de cette métrique, voir la section 2.3.7.2.1.2.

Nous avons calculé ce score en utilisant SacreBLEU⁷² (Post, 2018) et en prenant comme traduction de référence, soit la traduction commune à laquelle les étudiantes ont abouti dans le cadre d'un cours de traduction après la tenue de l'expérience (texte 2), soit une traduction créée par nos soins qui s'appuie sur les meilleures productions (TH comme PE) du corpus, ainsi que sur les commentaires faits par les évaluatrices.

⁷¹ <http://dataverse.harvard.edu>

⁷² BLEU|nrefs:1|case:mixed|eff:no|tok:13a|smooth:exp|version:2.0.0

Le Tableau 1 présente les scores obtenus grâce à la métrique automatique SacreBLEU.

Scores BLEU	Texte 1 (presse)	Texte 2 (économique)	Texte 3 (vulg. scientifique)
<i>Bing</i>⁷³ (jan 2018)	21,71	31,26	22,47
<i>DeepL</i> (jan 2018)	26,80	44,82	27,51
<i>DeepL</i> (sept 2020)	28,25	43,86	34,54
<i>Google Trad</i>⁷⁴ (sept 2020)	27,77	40,89	26,66

Tableau 1 : Scores BLEU

Sans grande surprise, les scores BLEU obtenus pour la TA statistique (*BingTranslator*) sont systématiquement inférieurs à ceux obtenus en TA neuronale pour les trois textes sources. Remarquons également une augmentation du score BLEU (jusqu'à 7 points pour le texte 3) entre la TA brute générée par *DeepL* en 2018 et celle générée par *DeepL* en 2020 (à l'exception du texte 2) alors qu'à notre connaissance, aucune traduction de ces textes n'a été publiée entre nos deux expériences. En supposant que cette amélioration au niveau des scores BLEU reflète bel et bien une augmentation de qualité, il est légitime de considérer que la TA neuronale est une technologie dont les performances ne cessent d'être améliorées comme l'attestent Breyel et Grass (2021) : « L'évolution de la TA neuronale est fascinante, et les systèmes progressent considérablement d'une année sur l'autre. Il suffit pour s'en convaincre de faire traduire le même texte par *DeepL* à un an d'intervalle » (p. 98).

3.1.6 Répartition des tâches

Étant donné que nos expériences ont été menées selon un plan factoriel à mesures répétées, il nous a fallu veiller particulièrement aux éventuels biais et effet d'ordre (Légal, 2008). Par conséquent, afin de minimiser l'effet d'ordre et afin que chaque texte soit également traduit selon les trois modes de traduction différents, nous avons envisagé la répartition des tâches selon un plan croisé (principe du carré latin) entre les facteurs suivants : textes sources, modes de traduction et ordre dans lequel les tâches devaient être accomplies (Figure 13).

⁷³ Dans sa version statistique, avant la mise à jour de mai 2018.

⁷⁴ Dans sa version neuronale (GNMT).

Chaque étudiante s'est vu attribuer une des clés de répartition de manière entièrement aléatoire.

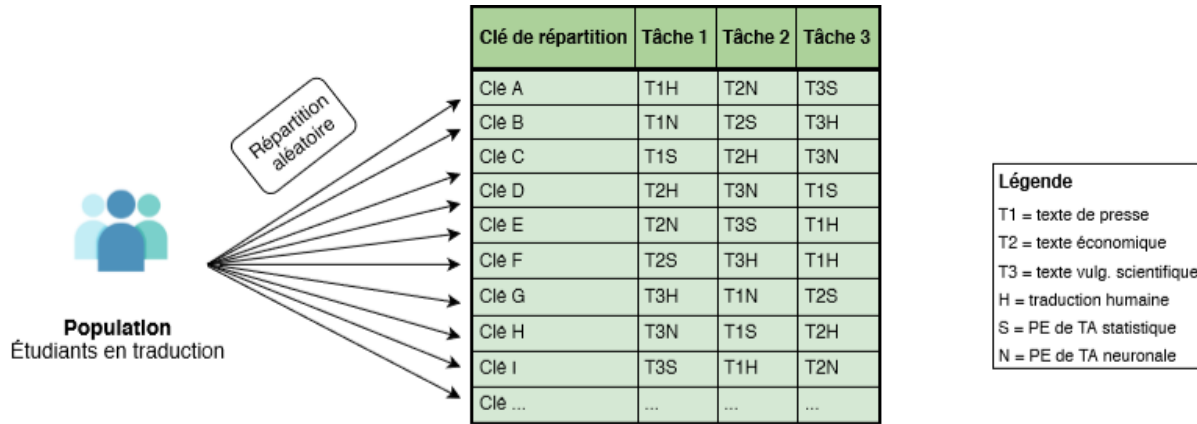


Figure 13 : Principe de répartition des tâches (Prétest)

3.1.7 MateCat

Ces tâches ont été réalisées en utilisant l’outil *MateCat*⁷⁵ (Federico et al., 2014a). Nous avons choisi ce logiciel de TAO, car il s’agit d’un outil gratuit, open source qui est aussi très intuitif et personnalisable. De plus, il nous a permis, par la suite, de disposer d’informations précieuses telles que l’effort de PE en pourcentage et la durée de PE (Federico et al., 2014b).

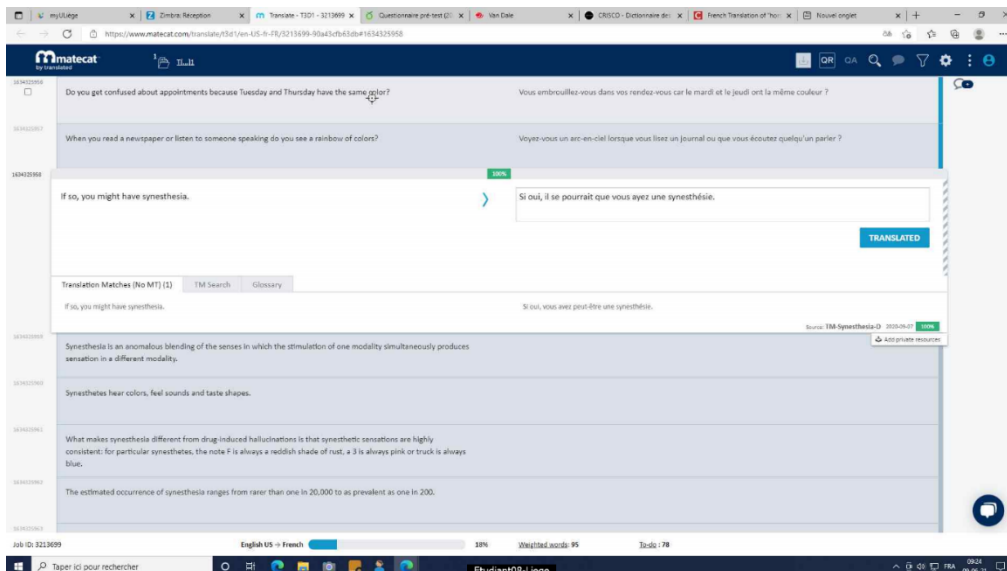


Figure 14 : Interface MateCat

⁷⁵ MateCat est l’acronyme de *machine translation enhanced computer-assisted translation*. Il s’agit d’un outil de traduction assistée par ordinateur accessible gratuitement en ligne à l’adresse : <https://www.matecat.com>

Pour cette expérience, il nous a fallu paramétrer *MateCat* et désactiver les options proposées par défaut par le logiciel ; notamment l'option *MyMemory* qui est une mémoire de traduction collaborative alimentée par les utilisatrices de *MateCat* et l'option *Machine Translation* qui allie *Google Traduction* et *Microsoft Translator* pour offrir des propositions de TA à l'utilisatrice.

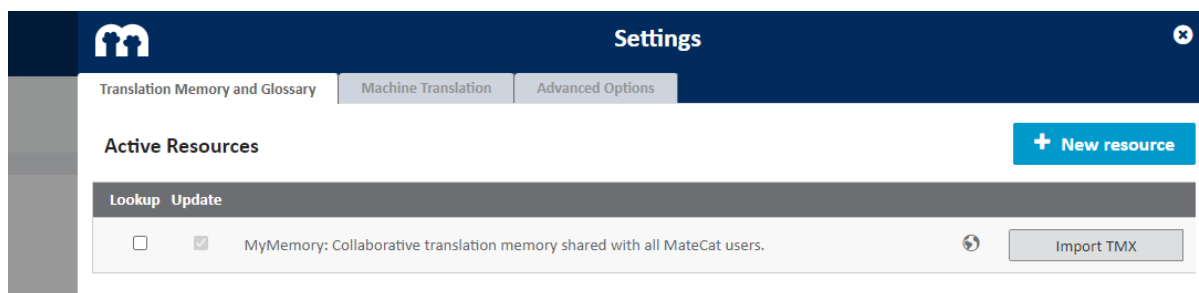


Figure 15 : Interface *MateCat* – MyMemory

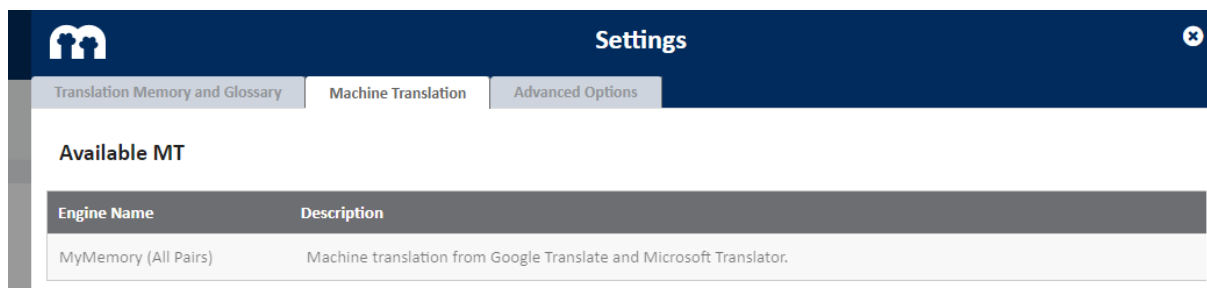


Figure 16 : Interface *MateCat* – Machine Translation

Pour des raisons méthodologiques, nous avons alimenté les mémoires de traduction uniquement avec nos textes sources et leurs TA brutes. Nous avons veillé à créer une mémoire de traduction spécifique à chaque étudiante pour chaque tâche de post-édition afin d'éviter que l'étudiante ne puisse avoir accès aux traductions déjà validées par les autres étudiantes et qui se seraient enregistrées automatiquement dans la mémoire au fur et à mesure de l'expérience. De cette manière, seule la TA brute s'affiche dans la colonne de droite (texte cible).

3.2 Déroulement des expériences

À l'instar de Daems (2016) et de Jia et al. (2019), nous avons soumis aux participantes deux sondages afin de mieux connaître leur opinion sur la TA et sur la PE, ainsi que leur expérience en la matière. Ces deux courts sondages étaient à compléter en ligne via LimeSurvey (Limesurvey GmbH, s.d.) ; le premier était à compléter en début de session et le deuxième en fin de session pour pouvoir avoir un retour sur l'expérience scientifique. L'analyse détaillée des réponses récoltées grâce à ces deux sondages figure au chapitre 7. Durant l'expérience stricto

sensu, les étudiantes ont été amenées à traduire, de l'anglais vers le français, les trois textes sources sélectionnés préalablement selon trois modes de traduction différents. Sur la Figure 17, nous avons schématisé la configuration de la recherche expérimentale (du prétest en l'occurrence) afin de mieux en cerner les différentes étapes.

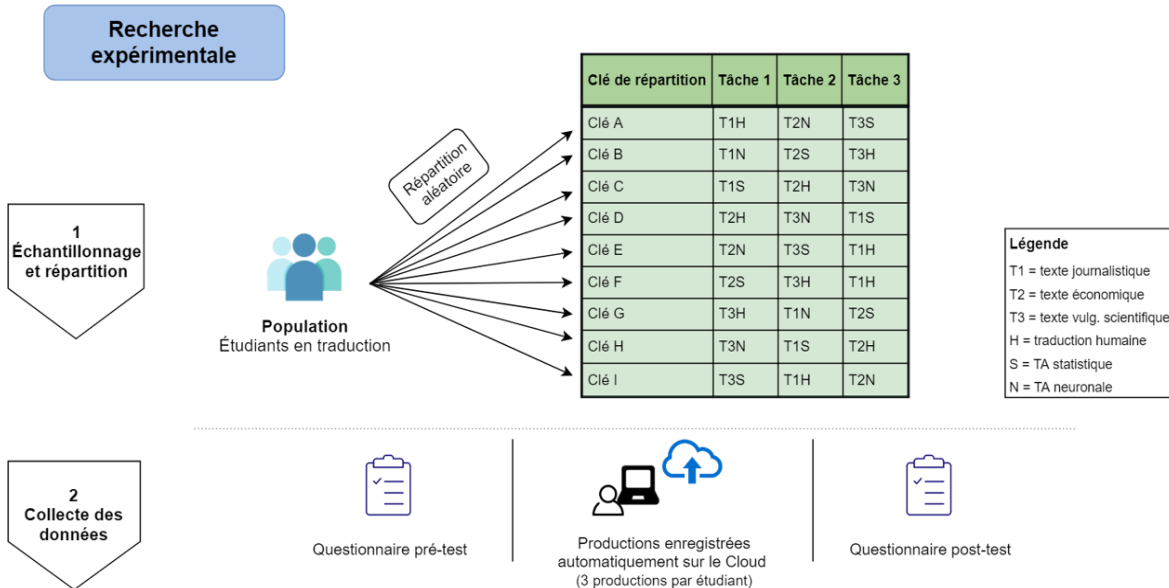


Figure 17 : Configuration du prétest

3.2.1 Prétest

Le prétest a eu lieu en février 2018 à l'Université de Liège et s'est déroulé en une seule session de 3h30 environ. Les 28 étudiantes de 3^e bachelier ont été regroupées dans un amphithéâtre et disposaient de leur ordinateur portable personnel. L'étudiante n'était pas en mesure d'identifier la technologie de TA utilisée (TA statistique ou TA neuronale). Aucune limite de temps par tâche ne leur a été imposée. Toutes les étudiantes ont terminé l'expérience en moins de 3h30.

3.2.2 Test

Cette deuxième expérience contrôlée s'est déroulée en juin 2021 à l'Université de Liège. En raison de la capacité du local dont nous disposions et des différentes exigences sanitaires en vigueur à ce moment-là, l'expérience a dû être divisée en deux sessions de 4h. À nouveau, les participantes n'étaient pas en mesure d'identifier le moteur de TA neuronale (*Google Traduction* ou *DeepL*). Aucune limite de temps par tâche ne leur a été imposée, les étudiantes ont terminé l'expérience en moins de 4h.

3.2.3 Exclusion de données

Au terme de chaque expérience, nous avons récolté trois productions par étudiante. Toutefois, pour le prétest, certaines productions ont dû être exclues du corpus. Il s'agit d'une TH inachevée, d'une production que nous avons jugée illisible en raison du nombre excessif d'erreurs de langue (notamment de grammaire et de syntaxe) et d'une production qu'il nous a été inconcevable de considérer comme étant de la PE, car toutes les propositions de la TA ont été acceptées sans que l'étudiante n'y apporte de modifications. Pour des raisons de fiabilité statistique et de représentativité des différents modes de traduction et des trois thématiques de TS, nous avons été contrainte d'exclure l'ensemble des productions des trois étudiantes concernées par les exclusions que nous venons d'énoncer. Ainsi, après exclusion des productions des étudiantes ET002, ET004 et ET0017, notre corpus final du prétest est composé de 75 productions au total : 25 traductions humaines, 25 PE de TA statistique, 25 PE de TA neuronale et six TA brutes.

Pour le test 2021, aucune production n'a été exclue. Le corpus du test se compose de 24 traductions humaines, de 24 PE de *Google Traduction*, de 24 PE de *DeepL*, ainsi que des six TA brutes.

3.3 Exploitation des données

3.3.1 Combinaison d'approches qualitative et quantitative

Afin de mesurer les effets de la PE sur la qualité d'un texte cible, nous avons procédé à l'étude comparative qualitative et quantitative des productions recueillies. Ces productions ont été traitées de manière totalement anonyme, il nous était impossible de relier la production à l'étudiante, à quelque étape que ce soit.

Pour l'exploitation des données empiriques, nous avons donc opté pour la combinaison d'une étude qualitative (repérage et classifications des erreurs par des évaluatrices humaines ; analyse linguistique des erreurs contenues dans le corpus) et d'une étude quantitative (statistique descriptive, statistique inférentielle et mesures automatiques linguistiques). Certes, nous considérons ces deux types d'approches comme complémentaires, mais nous n'entendons pas mélanger les analyses. Ce choix s'appuie sur les arguments développés par Gondard-Delcroix (2006) :

Booth [2001] s'interroge sur les dangers inhérents à un mixage pur et simple des approches qui risque de travestir leurs soubassements méthodologiques respectifs et de limiter la portée de l'analyse. C'est pourquoi, [sic] la modalité de combinaison la plus développée relève d'une mise en séquence des deux démarches [qualitative et quantitative] au sein du processus de recherche. [...] La complémentarité des domaines d'intelligibilité auxquels renvoient les deux approches est mobilisée pour guider et enrichir le processus de recherche. La démarche est alors séquentielle, les études qualitative et quantitative se succédant en un cycle [...] de façon à renforcer les concepts, les hypothèses et les résultats de la recherche. [...] l'analyse qualitative établit des relations de causalité là où l'étude quantitative perçoit des corrélations ; elle peut participer à l'explication de résultats inattendus. Elle permet également de tester la pertinence des résultats et de préciser ou réorienter les hypothèses de recherche. (p. 57)

Aussi, étant donné que nos analyses quantitative et qualitative ont été effectuées de manière distincte, j'en présenterai les résultats de manière séquentielle également.

3.3.2 Approche qualitative : évaluation humaine de la qualité

[T]here are no absolute standards of translation quality, but only more or less appropriate translations for the purpose for which they are intended.

JUAN CARLOS SAGER (1989)

Comme nous l'avons exposé dans l'état de l'art, l'évaluation de la qualité en traduction se révèle un processus complexe et délicat, inévitablement empreint de subjectivité (Bowker, 2000a ; Larose, 1998 ; Secară, 2005 ; Toudic et al., 2014), dont l'étude approfondie dépasse le cadre de cette thèse. Pour ces raisons, nous nous limitons à décrire le type d'évaluation que nous avons décidé d'adopter dans le cadre de nos expériences. Nous avons choisi d'opter pour une évaluation humaine de la qualité, car, même si cette méthode présente certains inconvénients, elle reste la référence en matière d'évaluation de la qualité en traduction : « human evaluation is still considered the "gold standard" to measure the final quality of the translation » (Vela Valido, 2021, p. 106).

Nous avons privilégié une approche par typologie d'erreurs, un modèle qui est fréquemment utilisé pour évaluer la qualité en traduction, que ce soit en contexte professionnel ou en contexte pédagogique (Gouadec, 1989 ; Saldanha et O'Brien, 2014). Comme le recommandent

Saldanha et O'Brien (2014), nous avons veillé à élaborer des consignes d'évaluation claires incluant des définitions précises des différentes catégories d'erreurs ; nous avons également donné la possibilité aux différentes évaluatrices de poser leurs éventuelles questions ou de demander des précisions sur la typologie d'erreurs.

Nous exposons à présent les raisons pour lesquelles nous avons opté pour une évaluation humaine de la qualité effectuée par des enseignantes en traduction au moyen d'une typologie d'erreurs. Au vu des objectifs que nous nous sommes fixés dans cette étude et étant donné qu'il s'agit d'une étude expérimentale en contexte d'apprentissage, il ne nous a pas semblé opportun d'avoir recours à des systèmes d'évaluation élaborés en contexte de pratique professionnelle et qui sont essentiellement développés pour le marché des services langagiers, telles les grilles LISA QA et DQF-MQM. (cf. section 2.3.5).

3.3.2.1 Typologie d'erreurs

Nous avons opté pour une évaluation par repérage et classification d'erreurs et de problèmes de traduction. En outre, il nous a paru pertinent que les productions recueillies soient évaluées selon une typologie d'erreurs qui soit la plus proche possible de la manière dont les étudiantes sont évaluées au cours des études universitaires en traduction à l'Université de Liège.

Force est de constater qu'il n'existe pas, au sein de notre filière, de méthode d'évaluation de la traduction qui fasse l'objet d'un consensus parmi le corps enseignant, chaque enseignante ayant développé sa propre façon de faire au-delà d'un socle commun de critères d'évaluation. Cette absence de méthode consensuelle nous a conduite à concevoir notre propre typologie d'erreurs. Pour ce faire, nous nous sommes inspirée, d'une part, des principaux critères communément employés au sein de la filière traduction-interprétation de l'Université de Liège et, d'autre part, de la méthode d'évaluation de la qualité développée par Daems (2016) dans le cadre de sa thèse de doctorat, car elle nous a semblé particulièrement adaptée à notre contexte de recherche et à la pratique pédagogique. Tout comme il est fréquemment d'usage dans les typologies d'erreurs professionnelles (Martikainen, 2019c), Daems (2016) a conçu une méthode d'évaluation humaine de la qualité qui permet de faire la distinction entre erreurs de langue et de contenu. Cette approche typologique consiste en deux phases consécutives : acceptabilité et adéquation. Ces deux termes font directement écho aux notions d'*acceptability* et d'*adequacy* proposées par Toury (1995) qui en donne la définition suivante : « whereas adherence to source norms determines a translation's adequacy as compared to the source text, subscription to norms originating in the target culture determines its acceptability » (pp. 56-57). Daems (2016) accorde une importance particulière à cette dichotomie et nous avons trouvé cette évaluation phasée très judicieuse notamment,

car elle permet une analyse fine des erreurs et de leur origine. Nous avons donc repris cette idée et avons prévu deux phases de correction par texte cible :

- une 1^{re} phase qui permet d'évaluer « l'acceptabilité » du texte cible ;
- une 2^e phase qui sert à évaluer « l'adéquation » du texte cible par rapport au texte source.

3.3.2.1.1 Phase d'acceptabilité

Par « acceptabilité » du texte cible, nous entendons le respect des normes et usages de la langue cible. Il s'agit donc essentiellement de repérer les fautes de langue (fautes de syntaxe, de grammaire, d'orthographe, barbarisme, impropriété, etc.), erreurs de style (répétition, longueur de phrase, fluidité, etc.) et problèmes de cohérence (logique interne, irrégularité dans l'emploi de termes, concordance des temps, etc.) (Figure 18). Pour cette étape, l'évaluatrice humaine dispose uniquement du texte cible⁷⁶ en français afin de pouvoir se placer dans la peau de la réceptrice.

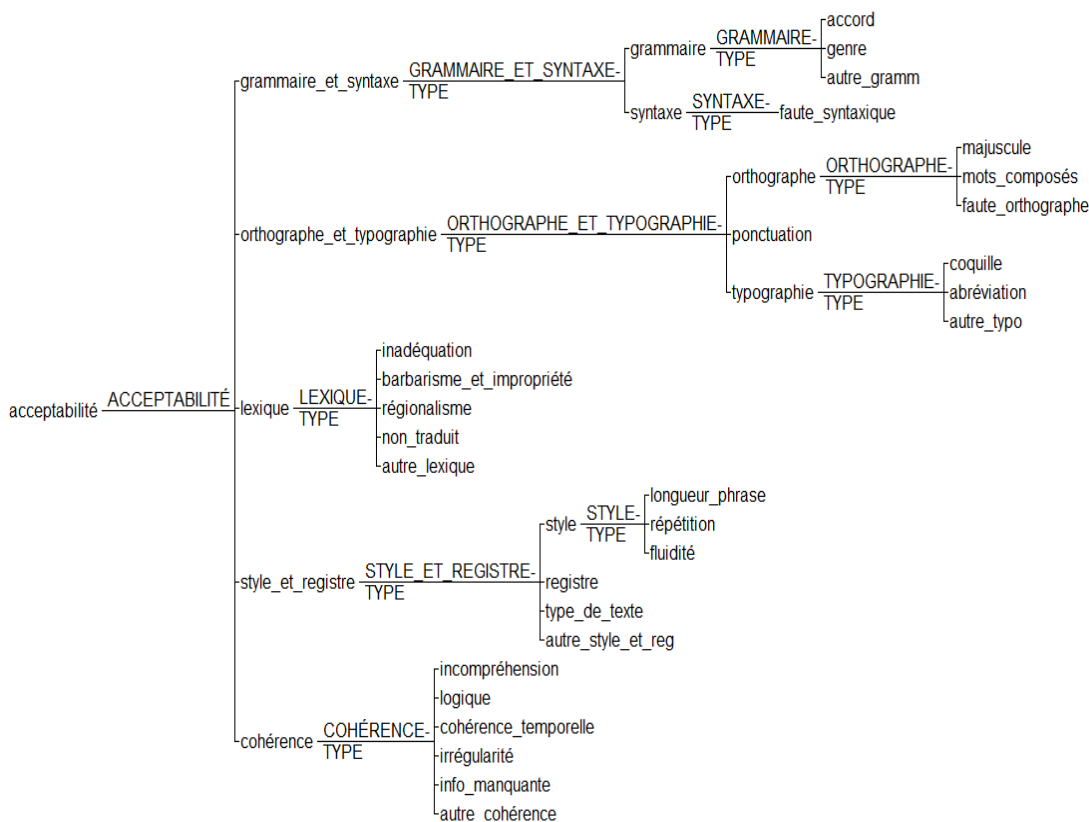


Figure 18 : Phase d'acceptabilité – Prétest

⁷⁶ « In principle, fluency judgements can be made without the evaluator even looking at the source segment » (Rossi et Carré, 2022, p. 57).

3.3.2.1.2 Phase d'adéquation

La deuxième phase d'évaluation permet de juger de « l'adéquation » du texte cible par rapport au texte source. Il s'agit ici de repérer les erreurs de traduction, c'est-à-dire les erreurs de transfert de sens (contresens, glissement de sens, calque, ajout, omission, etc.) (Figure 19). Pour cette étape, l'évaluatrice dispose à la fois du texte cible et du texte source.

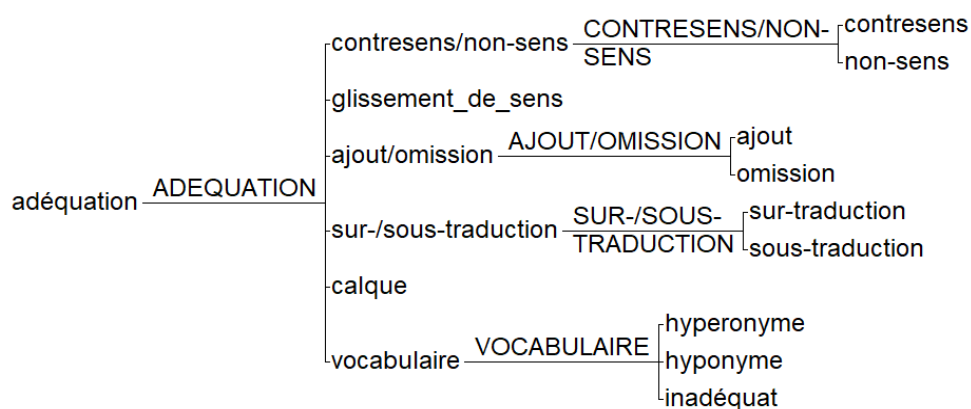


Figure 19 : Phase d'adéquation – Prétest

Autrement dit et à l'instar de la typologie conçue par Daems (2016), si l'évaluatrice repère l'erreur en ayant uniquement le texte cible sous les yeux, celle-ci sera annotée dans la phase d'acceptabilité, tandis que si l'erreur n'apparaît que lorsque l'évaluatrice a le texte cible et le texte source sous les yeux, elle sera prise en compte dans la phase d'adéquation. Ainsi, l'évaluation en phase d'acceptabilité ne requiert de maîtriser que la langue cible, tandis que pour la phase d'adéquation, l'évaluatrice doit avoir un certain degré de connaissance des langues source et cible (Castilho et al., 2018).

3.3.2.1.3 Perméabilité entre acceptabilité et adéquation

Toutefois, même en définissant clairement les catégories d'erreurs, il arrive que la frontière entre phase d'acceptabilité et phase d'adéquation ne soit pas hermétique, comme l'explique Martikainen (2019c) :

[L]es frontières entre erreurs de langue et de contenu sont perméables : en effet, une erreur de langue portant sur la syntaxe ou la terminologie, par exemple, peut bien sûr également constituer un contresens ou un non-sens, qui sont des catégories fondées sur le sens, donc le contenu et son interprétation par le lecteur.
(p. 39)

C'est pourquoi nous avons précisé aux évaluatrices de procéder dans ces cas-là à l'annotation dans les deux phases d'évaluation. Voici les explications que nous leur avons données pour clarifier ce point (voir remarques générales en annexe B.2 Consignes évaluatrices) :

Il est tout à fait possible qu'un même terme/segment doive être souligné au cours des deux phases d'évaluation.

Ex. : *To near-universal shock, however, he won the election.*

→Cependant, **contre tout attente**, il remporta les élections.

= Acceptabilité → grammaire - accord

= Adéquation → glissement de sens

3.3.2.2 Limitation de l'arbitraire

Nous avons tenu à limiter l'arbitraire inhérent à tout processus d'évaluation en limitant la présence de biais et en garantissant robustesse et fiabilité des résultats. Pour ce faire, les productions recueillies durant l'expérience ont été évaluées et annotées selon la typologie susmentionnée par plusieurs évaluateuses humaines (Bayerl et Paul, 2011), à raison de deux évaluateuses par texte source. Dans le cas d'une recherche axée sur l'évaluation de la qualité en traduction en contexte d'apprentissage, Saldanha et O'Brien (2014) recommandent de faire appel à des évaluateuses ayant de l'expérience dans ce type d'évaluation en contexte pédagogique. Dans notre expérience, il s'est agi d'enseignantes en traduction de notre filière ayant près de dix ans d'expérience dans l'enseignement de la traduction (bachelier et master) et de nombreuses années d'expérience en tant que traductrices indépendantes. Notons que dans le prétest, l'autrice de cette étude a intégré l'équipe d'évaluateuses, ce qui constitue un biais évident que nous avons corrigé dans le test 2021.

Ces évaluateuses ont systématiquement évalué chaque production selon les deux phases (acceptabilité et adéquation) au moyen de l'outil d'annotation configurable *UAM Corpus Tool* (O'Donnell, 2016). Il convient de préciser que les deux phases d'évaluation se sont, bien entendu, faites à l'aveugle ; les évaluateuses ne savaient pas si la production qu'elle jugeait était une TH ou une PE. L'évaluatrice avait sous les yeux le texte cible dans son intégralité et non une suite de segments à évaluer. Il leur a été demandé de souligner chaque erreur dans le texte cible et de les classer dans les différentes catégories et sous-catégories préétablies (Figure 18 et Figure 19), selon leur nature et leur gravité, une même erreur ne pouvant être soulignée qu'une seule fois (par ex. : une même faute d'orthographe ne sera pénalisée qu'une seule fois). Un extrait des consignes qui ont été distribuées aux évaluateuses et qui décrivent les différentes catégories d'erreur avec exemples sont reprises en annexe B.2 Consignes évaluateuses.

3.3.2.3 Phase de consolidation

Même si nous avons demandé aux différentes *évaluatrices* de suivre scrupuleusement les consignes d'évaluation que nous leur avons distribuées, nous sommes pleinement consciente que ce type de tâche conduit inévitablement à des divergences d'annotation plus ou moins grandes entre les juges, et ce, notamment, en raison des caractéristiques intrinsèquement subjectives de toute évaluation, et d'autant plus de ce qui relève de l'évaluation de la qualité en traduction. Pour tenter de pallier ce problème, nous avons prévu une étape de mise en commun entre les deux *évaluatrices* de chaque texte, comme Daems (2016) et Stymne and Ahrenberg (2012) le recommandent. Cette phase de consolidation offre aux *évaluatrices* la possibilité de discuter des éventuelles divergences d'évaluation et d'adapter, au besoin, certaines annotations. En effet, il est apparu que certaines erreurs n'avaient été repérées que par une seule *évaluatrice*. Il est arrivé également que la même erreur n'ait pas été étiquetée de la même manière par les deux enseignantes. Prenons l'exemple du terme « dépendre » qui apparaît sous la forme « dépendre » dans une des productions ; cette erreur a été étiquetée dans la catégorie « orthographe » par une *évaluatrice* et dans la catégorie « typographie et ponctuation » par une autre. Nous tenons à souligner qu'il n'était nullement question ici de gommer l'ensemble des divergences d'annotation entre les *évaluatrices*, ce qui n'aurait aucun sens, mais bien de leur laisser l'occasion d'harmoniser à leur guise les annotations si *elles* le jugeaient nécessaire.

3.3.2.4 Notation

Une fois cette mise en commun terminée, les différentes erreurs ont été comptabilisées pour les deux phases en veillant à calculer la moyenne des erreurs entre les annotations des différentes *évaluatrices*, toujours dans l'objectif d'atténuer les biais positifs et négatifs propres à chaque *évaluatrice*⁷⁷. Ensuite, une note globale sur 20 a été attribuée à chaque production et il a fallu pour ce faire définir une pondération. Il va de soi qu'établir une pondération dans une évaluation dépend de nombreux critères (expérience de l'étudiant, longueur et thématique du texte, etc.) ; de plus, rares sont les réviseuses et correctrices qui s'entendent sur la gravité des erreurs (Dussart, 2005). Aussi, dans un souci de limitation des biais et de validité écologique, nous nous sommes reposée à nouveau sur l'expérience de plusieurs enseignantes de notre filière, dont les correctrices, et leur avons soumis un tableau reprenant les différentes catégories d'erreur en leur demandant d'attribuer une pénalité à chacune de ces catégories en fonction de ce qu'*elles* ont l'habitude d'appliquer dans leur pratique. Nous avons obtenu quatre réponses et avons décidé de calculer systématiquement la moyenne de

⁷⁷ « Group-based TQA involves several evaluators conducting the assessment in this way and the averaging of their scores, to moderate strong positive or negative personal biases » (Castilho et al., 2018, p. 23).

ces quatre pénalités par catégorie pour constituer la pondération de nos expériences (Tableau 2). En outre, toujours dans l'objectif de reproduire le contexte d'enseignement dans lequel évoluent les étudiantes en traduction à l'Université de Liège, nous avons jugé bon d'accorder un poids plus important aux erreurs de traduction (phase d'adéquation) par rapport aux erreurs de langue (phase d'acceptabilité). Ainsi, la note obtenue en phase d'acceptabilité compte pour 40 pour cent de la note finale, tandis que la note obtenue en phase d'adéquation vaut pour 60 pour cent de la note finale. Pour calculer la note, les erreurs relevées pour chaque catégorie ont été pondérées à l'aide de la pénalité correspondante, puis additionnées. Ce score a alors été soustrait à la note de 20. Le Tableau 2 présente en détail la pénalité qui a été attribuée à chaque catégorie d'erreurs.

Pondération

Note sur 20

Acceptabilité (40 %)	Catégorie	Grammaire/ Syntaxe	Orthographe	Style/ registre	Lexique	Typographie/ Ponctuation	Cohérence (incompréhension)
	Pénalité	1	1	0,3	0,3	0,2	1,3
Adéquation (60 %)	Catégorie	Contresens/ Non-sens	Glissement de sens	Ajout/ Omission	Sur/sous- traduction	Calque	Vocabulaire
	Pénalité	1,3	0,7	0,7	0,5	0,5	0,5

Tableau 2 : Pondération notation sur 20 – Prétest

3.3.2.5 Mesures d'accord inter-évaluateur

À la suite de la phase de mise en commun entre évaluateurices, nous disposons de deux annotations par production. Nous avons donc tenu à mesurer le degré de concordance entre les différentes évaluateurices afin de garantir la fiabilité des annotations récoltées. L'accord inter-évaluateur ou fiabilité inter-observateur est devenu pratiquement la norme pour tester la fiabilité d'une méthode manuelle d'annotation :

Inter-annotator (or inter-coder) agreement has become the quasi-standard procedure for testing the accuracy of manual annotations. This process is based on the assumption that if multiple coders agree in their coding decisions of the same material we can be certain that—at least for this set of data and this set of coders—annotations are free of unsystematic and distorting variations. (Bayerl et Paul, 2011, p. 700)

3.3.2.5.1 Accord observé (%)

La méthode la plus simple et directe pour mesurer ce degré de concordance est de calculer l'accord observé, c'est-à-dire le nombre d'éléments sur lesquels les évaluateurs s'accordent, divisé par le nombre total d'éléments évalués (Gerlach, 2015). Le résultat ainsi obtenu est toujours compris entre 0 % et 100 % (Shweta Bajpai et Chaturvedi, 2015).

Grâce à l'enregistrement des annotations sur le logiciel *UAM Corpus Tool*, nous disposons d'informations précieuses sur chaque production, telles que le nombre de segments identifiés comme fautifs par les deux évaluateurs et ceux qui l'ont été par seulement une des deux évaluateurs (Tableau 3).

Segments identified by eval 1:	23	
Segments identified by eval 2:	26	
Segments that match:	19	79.2 %
Eval 1 segments not identified by eval 2:	4	16.7 %
Eval 2 segments not identified by eval 1:	7	25.9 %

Tableau 3 : Données *UAM Corpus Tool*

Pour obtenir le taux d'accord inter-évaluateur en pourcentage, nous avons calculé, pour chaque production d'un même texte, le rapport entre le nombre de segments qui « matchent » (erreurs repérées par les deux évaluateurs) et le nombre total d'erreurs repérées (c.-à-d. la somme des segments qui matchent, des erreurs repérées uniquement par l'évaluateur 1 et, enfin, celle repérées uniquement par l'évaluateur 2) avant de calculer la moyenne de ces rapports.

$$\text{Percent agreement} = \frac{\text{(Number of concordant responses)}}{\text{(Total number of responses)}} \times 100$$

Figure 20 : Formule accord observé (Shweta Bajpai et Chaturvedi, 2015)

3.3.2.5.1.1 Taux d'accord – Prétest 2018

Les taux d'accord inter-évaluateur moyens calculés après la phase de consolidation sont repris dans le Tableau 4.

Étant donné que, dans notre prétest, les productions récoltées pour le texte source 1 n'ont été évaluées que par une seule et même évaluateur, nous avons uniquement calculé le taux d'accord pour les textes 2 et 3, respectivement le texte économique et le texte de vulgarisation scientifique.

Taux d'accord moyen (%)

	Texte 2 (économique)	Texte 3 (vulg. scientifique)
Acceptabilité	75,2 % ⁷⁸	68,8 % ⁷⁹
Adéquation	68,2 % ⁸⁰	85,4 % ⁸¹

Tableau 4 : Taux d'accord inter-évaluateur (%) – Prétest 2018

Notons que toutes ces valeurs se situent entre 68 % et 85 %, ce qui signifie que nous pouvons considérer l'accord observé comme bon (Stemler, 2004) et ce, quelle que soit la phase d'évaluation (acceptabilité et adéquation).

3.3.2.5.1.2 Taux d'accord – Test 2021

Tout comme pour le prétest, nous avons calculé le taux d'accord inter-évaluateur après la phase de consolidation. Ces taux d'accord inter-évaluateur moyens sont repris dans le Tableau 5.

Taux d'accord moyen (%)

	Texte 1	Texte 2	Texte 3
Acceptabilité	55,5 %	43,5 %	53,4 %
Adéquation	77,0 %	53,0 %	50,1 %

Tableau 5 : Taux d'accord inter-évaluateur (%) – Test2021

Notons que ces taux d'accord sont relativement faibles et que, pour toute une série de raisons que nous ignorons, toutes ces valeurs sont inférieures aux taux d'accord obtenus dans le prétest. Aussi, il nous a semblé pertinent de recalculer cet accord en excluant, pour chaque phase d'évaluation, la catégorie jugée comme laissant le plus de place à la partialité de jugement, en tout cas celle pour laquelle les évaluateurices ont décidé de maintenir des divergences d'annotation. Il s'agit de la catégorie « style »⁸² en acceptabilité et de la catégorie « calque » en adéquation.

⁷⁸ Soit 759 erreurs qui ont été identifiées par les deux évaluateurices sur un total de 987 erreurs.

⁷⁹ Soit 403 erreurs repérées par les deux évaluateurices sur un total de 575 erreurs.

⁸⁰ Soit 418 erreurs en commun sur un total de 591 erreurs.

⁸¹ Soit 378 erreurs repérées par les deux évaluateurices sur un total de 439 erreurs.

⁸² « [T]he judgement of style is somewhat more subjective than for the other two parameters [clarity and accuracy] and, as we have already discussed, professional linguists may be more critical of style » (Fiederer et O'Brien, 2009, p. 63).

	Texte 1	Texte 2	Texte 3
Acceptabilité	70 %	63,6 %	71,9 %
Adéquation	79 %	55,3 %	55,8 %

Tableau 6 : Taux d'accord inter-évaluateur (%) (Style & Calque exclus) – Test2021

Les taux d'accord inter-évaluateur mesurés de cette manière sont plus élevés, et ce, pour chaque texte, ce qui prouve une certaine fiabilité dans les données d'évaluation.

3.3.2.5.2 Coefficient kappa de Cohen

Toutefois, cette mesure d'accord observé est fragile et loin d'être idéale, car d'une part, ce taux est influencé par le nombre de catégories et, d'autre part, il ne tient pas compte des effets du hasard, à savoir de la possibilité que les annotateurs classent un élément quelconque dans une même catégorie par hasard (Fort et al., 2010). « Ce problème est souvent résolu en appliquant des méthodes corrélationnelles (calcul de coefficients de corrélation), dont le choix dépend du nombre de classements dont on dispose (deux ou plus), ainsi que des caractéristiques métriques des échelles utilisées pour effectuer ces classements » (Arrondo, 2012, p. 46).

Le test le plus connu et le plus couramment employé pour déterminer la fiabilité inter-évaluateur est le calcul du coefficient kappa de Cohen (Cohen, 1960). Néanmoins, dans cette expérience, il ne nous a pas été possible de le calculer au moyen de cette méthode. En effet, le calcul du score kappa s'est avéré inadéquat dans cette étude puisqu'il nécessite de mesurer la différence entre la concordance observée et la concordance liée à la chance ou « aléatoire » calculée à partir des effectifs attendus (Elie et Colombet, 2011), à savoir les erreurs que les deux évaluateurs n'ont pas repérées. Or, nous ne disposons pas en l'occurrence d'un relevé exhaustif des erreurs par production en raison une fois de plus du caractère complexe et subjectif de toute évaluation humaine de la qualité en traduction. Par conséquent, nous n'avons pas pu calculer ces effectifs attendus. Il nous a fallu trouver un autre test pour mesurer ce degré de concordance.

3.3.2.5.3 Coefficient de corrélation intra-classe

Étant donné qu'il nous était impossible de recourir au calcul du score kappa pour mesurer la fiabilité inter-évaluateur dans notre expérience, nous avons opté, après des recherches plus approfondies, pour l'*Intraclass Correlation Coefficient* ou coefficient de corrélation intra-classe (CCI) (McGraw & Wong, 1996), notamment parce que ce test statistique n'est pas influencé par la prévalence et tient compte du degré de désaccord (Gerlach, 2015).

En statistique, l'indice de corrélation CCI est une mesure générale d'accord ou de consensus (Gélinas, 2017) qui permet d'évaluer la concordance entre plusieurs séries de mesures

quantitatives (Müller et Büttner, 1994). Ce test permet de mesurer la concordance de variables ordinales, d'intervalles ou de ratios pour des études comprenant plusieurs évaluateurs (Hallgren, 2012). Le CCI peut s'interpréter comme « la part de la variance observée dans les évaluations qui sont dues à des différences systématiques entre les objets évalués, comparée à la variance totale dans les évaluations » (Jilinskaya, 2012, p. 192). Sa valeur est comprise entre 0 et 1 (Koo et Li, 2015). Plus le coefficient obtenu est proche de 1, moins il y a de variations inter-évaluateurs et donc, plus grand est le degré de fiabilité.

Le calcul du CCI nécessite de sélectionner des paramètres en fonction de la configuration expérimentale. Dans cette étude, la fiabilité inter-évaluateur a été calculée grâce au CCI selon le modèle à effets aléatoires à deux facteurs de type cohérence absolue, pour des mesures moyennes (McGraw et Wong, 1996 ; Koo et Li, 2015).

Pour ce faire, nous avons utilisé le logiciel SPSS (IBM SPSS statistics, version 25) et calculé le coefficient pour chaque phase d'évaluation (acceptabilité ; adéquation) pour les textes 2 et 3. Pour les raisons déjà mentionnées plus haut, il n'a pas été possible de calculer ce coefficient pour le texte 1. Les Tableau 7, Tableau 8, Tableau 9 et Tableau 10 présentent le score CCI et les intervalles de confiance pour les différents textes et phases d'évaluation.

3.3.2.5.3.1 Indices CCI –Prétest 2018

Texte 2 – Acceptabilité

	Corrélation intra-classe ^b	Intervalle de confiance à 95 %		Test F avec valeur réelle 0			
		Borne inférieure	Borne supérieure	Valeur	ddl1	ddl2	Sig
Mesures uniques	,724 ^a	-0,072	0,923	23,518	26	26	0,000
Mesures moyennes	0,840	-0,155	0,960	23,518	26	26	0,000

Modèle à effets aléatoires à deux facteurs lorsque les effets sur les personnes et les effets des mesures sont aléatoires.

a. L'estimateur est identique, que l'effet d'interaction soit présent ou non.

b. Coefficients de corrélation intra-classe de type A utilisant une définition de cohérence absolue.

Tableau 7 : Coefficient de corrélation intra-classe (texte 2 – acceptabilité)

Texte 2 – Adéquation

	Corrélation intra-classe ^b	Intervalle de confiance à 95 %		Test F avec valeur réelle 0			
		Borne inférieure	Borne supérieure	Valeur	ddl1	ddl2	Sig
Mesures uniques	,599 ^a	-,041	,886	33,543	26	26	,000
Mesures moyennes	,749	-,086	,940	33,543	26	26	,000

Modèle à effets aléatoires à deux facteurs lorsque les effets sur les personnes et les effets des mesures sont aléatoires.

a. L'estimateur est identique, que l'effet d'interaction soit présent ou non.

b. Coefficients de corrélation intra-classe de type A utilisant une définition de cohérence absolue.

Tableau 8 : Coefficient de corrélation intra-classe (texte 2 – adéquation)

Texte 3 – Acceptabilité

	Corrélation intra-classe ^b	Intervalle de confiance à 95 %		Test F avec valeur réelle 0			
		Borne inférieure	Borne supérieure	Valeur	ddl1	ddl2	Sig
Mesures uniques	,824 ^a	,462	,932	15,349	26	26	,000
Mesures moyennes	,904	,632	,965	15,349	26	26	,000

Modèle à effets aléatoires à deux facteurs lorsque les effets sur les personnes et les effets des mesures sont aléatoires.

a. L'estimateur est identique, que l'effet d'interaction soit présent ou non.

b. Coefficients de corrélation intra-classe de type A utilisant une définition de cohérence absolue.

Tableau 9 : Coefficient de corrélation intra-classe (texte 3 – acceptabilité)

Texte 3 – Adéquation

	Corrélation intra-classe ^b	Intervalle de confiance à 95 %		Test F avec valeur réelle 0			
		Borne inférieure	Borne supérieure	Valeur	ddl1	ddl2	Sig
Mesures uniques	,932 ^a	,858	,968	27,984	26	26	,000
Mesures moyennes	,965	,924	,984	27,984	26	26	,000

Modèle à effets aléatoires à deux facteurs lorsque les effets sur les personnes et les effets des mesures sont aléatoires.

a. L'estimateur est identique, que l'effet d'interaction soit présent ou non.

b. Coefficients de corrélation intra-classe de type A utilisant une définition de cohérence absolue.

Tableau 10 : Coefficient de corrélation intra-classe (texte 3 – adéquation)

Interprétation

Il n'existe pas de convention universelle pour interpréter le coefficient de corrélation intra-classe. Pour interpréter les valeurs obtenues dans ces tableaux, nous avons choisi de nous appuyer sur l'échelle d'interprétation développée par Koo et Li (2015) d'après laquelle le degré de fiabilité inter-évaluateur est considéré comme faible pour un coefficient inférieur à 0,5, comme modéré pour un coefficient compris entre 0,5 et 0,75, comme bon pour un coefficient compris entre 0,75 et 0,9, et enfin, comme excellent pour un coefficient situé au-delà de 0,9.

Nous avons fixé l'intervalle de confiance à 95 %, c'est-à-dire qu'il y a une probabilité de 95 % que l'intervalle de valeurs délimité par la borne inférieure et par la borne supérieure contienne la valeur de notre coefficient de corrélation intra-classe. Étant donné que notre recherche implique plusieurs évaluateuses, la ligne de résultats qui nous intéresse est « mesures moyennes », car nous cherchons à déterminer si les évaluations des évaluateuses sont cohérentes les unes par rapport aux autres en moyenne.

Pour le texte 2, en phase d'acceptabilité (Tableau 7), nous observons un bon degré de fiabilité inter-évaluateur avec un coefficient de 0,84 (IC 95 % : -1,55 – 0,96).

Pour le texte 2, en phase d'adéquation (Tableau 8), la fidélité inter-évaluateur s'avère également bonne avec un coefficient de 0,75 (IC 95 % : -0,09 – 0,94).

Le CCI pour le texte 3 en phase d'acceptabilité (Tableau 9) s'élève à 0,90 (IC 95 % : 0,63 – 0,96) ce qui peut être considéré comme un excellent degré de fiabilité inter-évaluateur.

Enfin, en phase d'adéquation (Tableau 10), nous avons obtenu, pour le texte 3, un coefficient de 0,96 avec un IC 95 % allant de 0,92 à 0,98 ; ce degré de fiabilité peut également être considéré comme excellent.

De manière générale, nous pouvons conclure que, pour les textes 2 et 3 dans ce prétest, la fiabilité inter-évaluateur est bonne, voire excellente, car le CCI est systématiquement égal ou supérieur à 0,75.

3.3.2.5.3.2 Indices CCI – Test 2021

Étant donné qu'il nous était toujours impossible d'utiliser le score kappa pour calculer la fiabilité inter-évaluateurs, nous avons à nouveau calculé le coefficient de corrélation intra-classe au moyen du logiciel SPSS (version 26). Ces scores ont été mesurés pour l'ensemble des données (catégories « style » et « calque » incluses). Dans les Tableau 11, Tableau 12, Tableau 13, Tableau 14, Tableau 15 et Tableau 16 sont reprises les valeurs de CCI et les intervalles de confiance pour les différents textes et phases d'évaluation.

Texte 1 – Acceptabilité

	Corrélation intra-classe ^b	Intervalle de confiance à 95 %		Test F avec valeur réelle 0			
		Borne inférieure	Borne supérieure	Valeur	ddl1	ddl2	Sig
Mesures uniques	,760 ^a	,294	,908	11,348	25	25	,000
Mesures moyennes	,864	,454	,952	11,348	25	25	,000

Modèle à effets aléatoires à deux facteurs lorsque les effets sur les personnes et les effets des mesures sont aléatoires.

a. L'estimateur est identique, que l'effet d'interaction soit présent ou non.

b. Coefficients de corrélation intra-classe de type A utilisant une définition de cohérence absolue.

Tableau 11 : Coefficient de corrélation intra-classe (texte 1 – acceptabilité)

Texte 1 – Adéquation

	Corrélation intra-classe ^b	Intervalle de confiance à 95 %		Test F avec valeur réelle 0			
		Borne inférieure	Borne supérieure	Valeur	ddl1	ddl2	Sig
Mesures uniques	,813 ^a	,629	,911	9,850	25	25	,000
Mesures moyennes	,897	,772	,953	9,850	25	25	,000

Modèle à effets aléatoires à deux facteurs lorsque les effets sur les personnes et les effets des mesures sont aléatoires.

a. L'estimateur est identique, que l'effet d'interaction soit présent ou non.

b. Coefficients de corrélation intra-classe de type A utilisant une définition de cohérence absolue.

Tableau 12 : Coefficient de corrélation intra-classe (texte 1 – adéquation)

Texte 2 – Acceptabilité

	Corrélation intra-classe ^b	Intervalle de confiance à 95 %		Test F avec valeur réelle 0			
		Borne inférieure	Borne supérieure	Valeur	ddl1	ddl2	Sig
Mesures uniques	,689 ^a	,420	,847	5,818	25	25	,000
Mesures moyennes	,816	,592	,917	5,818	25	25	,000

Modèle à effets aléatoires à deux facteurs lorsque les effets sur les personnes et les effets des mesures sont aléatoires.

a. L'estimateur est identique, que l'effet d'interaction soit présent ou non.

b. Coefficients de corrélation intra-classe de type A utilisant une définition de cohérence absolue.

Tableau 13 : Coefficient de corrélation intra-classe (texte 2 – acceptabilité)

Texte 2 – Adéquation

	Corrélation intra-classe ^b	Intervalle de confiance à 95 %		Test F avec valeur réelle 0			
		Borne inférieure	Borne supérieure	Valeur	ddl1	ddl2	Sig
Mesures uniques	,579 ^a	,263	,785	3,830	25	25	,001
Mesures moyennes	,733	,417	,879	3,830	25	25	,001

Modèle à effets aléatoires à deux facteurs lorsque les effets sur les personnes et les effets des mesures sont aléatoires.

a. L'estimateur est identique, que l'effet d'interaction soit présent ou non.

b. Coefficients de corrélation intra-classe de type A utilisant une définition de cohérence absolue.

Tableau 14 : Coefficient de corrélation intra-classe (texte 2 – adéquation)

Texte 3 – Acceptabilité

	Corrélation intra-classe ^b	Intervalle de confiance à 95 %		Test F avec valeur réelle 0			
		Borne inférieure	Borne supérieure	Valeur	ddl1	ddl2	Sig
Mesures uniques	,539 ^a	-,075	,822	6,861	25	25	,000
Mesures moyennes	,701	-,163	,902	6,861	25	25	,000

Modèle à effets aléatoires à deux facteurs lorsque les effets sur les personnes et les effets des mesures sont aléatoires.

a. L'estimateur est identique, que l'effet d'interaction soit présent ou non.

b. Coefficients de corrélation intra-classe de type A utilisant une définition de cohérence absolue.

Tableau 15 : Coefficient de corrélation intra-classe (texte 3 – acceptabilité)

Texte 3 – Adéquation

	Corrélation intra- classe ^b	Intervalle de confiance à 95 %		Test F avec valeur réelle 0			
		Borne inférieure	Borne supérieure	Valeur	ddl1	ddl2	Sig
Mesures uniques	,810 ^a	,622	,910	9,328	25	25	,000
Mesures moyennes	,895	,767	,953	9,328	25	25	,000

Modèle à effets aléatoires à deux facteurs lorsque les effets sur les personnes et les effets des mesures sont aléatoires.

a. L'estimateur est identique, que l'effet d'interaction soit présent ou non.

b. Coefficients de corrélation intra-classe de type A utilisant une définition de cohérence absolue.

Tableau 16 : Coefficient de corrélation intra-classe (texte 3 – adéquation)

Interprétation

Pour une description détaillée de l'échelle d'interprétation et de l'intervalle de confiance auxquels nous faisons référence, se reporter à la Figure 36. Pour rappel, la ligne de résultats qui nous intéresse est « mesures moyennes », car nous cherchons à déterminer si les évaluations des évaluateurices sont cohérentes les unes par rapport aux autres en moyenne.

Pour le texte 1 en phase d'acceptabilité (Tableau 11), nous observons un bon degré de fiabilité inter-évaluateur avec un coefficient de 0,86 (IC 95 % : 0,45 – 0,95).

Pour le texte 1 en phase d'adéquation (Tableau 12), la fidélité inter-évaluateur se révèle très bonne avec un coefficient arrondi de 0,90 (IC 95 % : 0,77 – 0,95).

Pour le texte 2 en phase d'acceptabilité (Tableau 13), nous observons un bon degré de fiabilité inter-évaluateur avec un coefficient de 0,82 (IC 95 % : 0,59 – 0,92).

Pour le texte 2 en phase d'adéquation (Tableau 14), la fiabilité inter-évaluateur est modérée avec un coefficient de 0,73 (IC 95 % : 0,42 – 0,88).

Le CCI pour le texte 3 en phase d'acceptabilité (Tableau 15) s'élève à 0,70 (IC 95 % : -0,16 – 0,90) ; cet indice démontre une fiabilité modérée.

Enfin, pour le texte 3 en phase d'adéquation (Tableau 16), nous obtenons un coefficient arrondi de 0,90 avec un IC 95 % allant de 0,77 à 0,95 ; ce taux d'accord peut être considéré comme excellent.

Nous pouvons conclure que, de manière générale dans le test, la fiabilité de l'évaluation humaine est satisfaisante puisque les différents taux d'accord inter-évaluateur mesurés sont systématiquement supérieurs ou égaux à 0,70.

3.3.3 Approche quantitative : étude statistique

À partir des relevés d'erreurs et des notes attribuées à chaque production lors de l'évaluation humaine, nous avons procédé à une analyse quantitative combinant la statistique descriptive et la statistique inférentielle.

L'une des caractéristiques majeures de notre étude tient au fait qu'elle relève du champ des sciences humaines et que, contrairement à une expérience en sciences fondamentales ou appliquées, nous n'avons pas pu soumettre le même texte source plusieurs fois à une même étudiante. Un tel scénario aurait pourtant été idéal pour l'analyse statistique, mais faire traduire le même texte plusieurs fois par une traductrice aurait introduit naturellement des biais dans notre étude, car celle-ci aurait notamment conservé en mémoire une partie ou l'ensemble des choix de traduction qu'elle a effectués. Il n'aurait donc pas été possible de mesurer de manière valide les potentiels effets du mode de traduction sur la qualité du produit

fini. Par conséquent, il a fallu déterminer les tests statistiques qu'il convenait d'appliquer afin d'exploiter valablement les résultats de ces expériences en fonction de ces spécificités.

3.3.3.1 Statistique descriptive

La statistique descriptive se définit comme « l'ensemble des méthodes de dénombrement, de classement, de synthèse et de présentation de données quantitatives relatives à un ensemble d'individus » (Albarello et al., 2010, p. 11). Baccini (2010) précise qu'habituellement, l'étape descriptive précède l'étape inférentielle : « D'un point de vue méthodologique, on notera que la statistique descriptive précède en général la statistique inférentielle dans une démarche de traitement de données : ces deux aspects de la statistique se complètent bien plus qu'ils ne s'opposent » (p. 6).

Ainsi, nous nous sommes attelée à classer, à décrire et à présenter (en graphique et en histogramme) les données quantitatives tirées de l'évaluation de la qualité. Dans le but de répondre à nos sous-questions de recherche, nous avons exploré les trois modes de traduction en observant la distribution des erreurs par catégorie, la qualité des productions par mode de traduction, la moyenne des notes, le taux de réussite et les écarts-types selon le mode de traduction et selon le TS, mais aussi en comparant les performances de différents systèmes de TA et les performances par étudiante.

3.3.3.2 Statistique inférentielle

La statistique inférentielle est définie comme « l'ensemble des méthodes et des théories permettant de généraliser à une population de référence des conclusions obtenues à partir de l'étude d'un échantillon extrait de cette population » (Albarello et al., 2010, p. 13).

Pour rappel, selon nos paramètres d'expérience, la qualité des productions des étudiantes constitue la variable dépendante, tandis que les variables indépendantes sont le mode de traduction ; la thématique du texte source ; l'étudiante. Pour explorer les relations entre nos variables indépendantes et notre variable dépendante, c'est-à-dire pour tester si la qualité des productions recueillies a été influencée par le mode de traduction ou par le texte source, nous avons cherché à démontrer l'existence de différences statistiquement significatives à deux niveaux : la variable de qualité du texte cible a été mesurée tantôt par le nombre d'erreurs relevées dans les productions des étudiantes, tantôt par les notes attribuées aux productions des étudiantes selon une pondération prédéfinie.

D'un point de vue statistique, en tenant compte des caractéristiques spécifiques de ces expériences, nous avons eu recours à un test d'analyse de la variance (ANOVA). Il s'agit d'une méthode d'analyse fréquemment utilisée pour analyser les résultats d'expériences contrôlées (Legendre et Borcard, s.d.), et qui est particulièrement adaptée à notre cas puisque ce modèle statistique permet de comparer des moyennes d'échantillons pour tester l'influence d'une (ou de plusieurs) variable(s) qualitative(s) sur une variable quantitative :

Une Anova (ou analyse de variance) est un test statistique paramétrique qui permet de comparer des moyennes et de déterminer si ces moyennes diffèrent significativement. L'Anova a pour particularité de comparer les moyennes en tenant compte de la variance dans ses estimations (d'où son nom). Il s'agit en fait de faire un rapport entre variance(s) inter conditions et variance(s) intra conditions. Autrement dit, il s'agit de voir si la dispersion des mesures est plus importante entre les conditions qu'à l'intérieur des conditions. (Légal, 2008, p. 25)

Nous n'avons pas pu effectuer de test ANOVA classique, car celui-ci requiert un degré élevé de contrôle des différentes variables de l'expérience. Or, comme il s'agit d'une expérience en sciences humaines, il nous est difficile de contrôler suffisamment l'ensemble des variables (Balling, 2008). Nous avons donc eu recours à un test statistique plus flexible qui autorise ce manque de contrôle. Étant donné que nos expériences ont été réalisées selon un plan factoriel à mesures répétées, avec deux facteurs de variabilité (le mode de traduction et le texte source) et plus de deux échantillons appariés, nous avons dû envisager l'application d'un test ANOVA à deux facteurs de type III, en mesures répétées (*two-way ANOVA of type III with repeated measures*). Comme son nom le précise, ce type d'analyse de la variance permet de traiter les répétitions de mesures sur les mêmes sujets (les étudiantes).

Cette analyse statistique a été effectuée en collaboration avec le D^r Marc-Olivier Boldi (*Research Center for Statistics* de l'Université de Genève) et le D^r Antonio Sutera (Département d'électricité, électronique et informatique de l'ULiège). Pour le traitement des données, nous avons utilisé le logiciel de statistiques R 3.6.3 (R Core Team, 2020). Dans ce logiciel, nous avons également eu recours aux packages *LmerTest* v3.1-1 (Kuznetsova et al., 2017) et *Multcomp* v1.4-12 (Hothorn et al., 2008).

3.3.4 Analyse complémentaire

3.3.4.1 Étude du processus

Étant donné que nous avons à notre disposition les données nécessaires, nous nous sommes brièvement penchée sur le processus même de post-édition, même si cet aspect ne fait pas l'objet de nos questions de recherche.

Nous avons mesuré les éventuels gains de productivité par texte et par étudiante en fonction des différents modes de traduction, et avons exploré l'existence d'une relation entre la productivité d'une étudiante et la qualité du TC. Enfin, nous avons observé l'effort de PE en fonction du type de TA et avons tenté de mettre au jour une relation entre l'effort moyen de PE et la qualité du TC.

3.4 Conclusion

Dans ce chapitre méthodologique, nous avons présenté dans un premier temps le cadre méthodologique adopté dans ce travail en commençant par exposer les différents paramètres expérimentaux : variables, participantes, tâches, corpus... Dans un deuxième temps, nous nous sommes attachée à résumer et à schématiser le déroulement de chaque expérience. Enfin, dans un troisième temps, nous avons décrit les différentes étapes d'exploitation des données récoltées dans nos expériences. Pour l'analyse qualitative du corpus, les productions seront évaluées à l'aveugle par une équipe d'évaluatrices humaines à l'aide d'une typologie d'erreurs élaborée aux fins de cette recherche. Nous avons prévu deux phases d'évaluation (acceptabilité et adéquation) dont nous avons détaillé les différents critères, ainsi que la grille de pondération qui permet d'attribuer une note sur 20 à chaque production. Nous avons également présenté les résultats des calculs d'accord inter-évaluateur. Les données tirées de l'évaluation humaine feront ensuite l'objet d'une analyse quantitative combinant la statistique descriptive et la statistique inférentielle. La dernière étape d'exploitation consiste en une brève étude du processus de PE.

4 Recherche expérimentale – Prétest

4.1 Introduction

Ce chapitre porte sur notre première expérience contrôlée (prétest). D'une part, grâce à ce prétest, nous avons pu comparer la qualité des produits de trois modes de traduction : la TH, la PE de TA statistique et la PE de TA neuronale. D'autre part, nous avons soumis aux étudiantes deux courts sondages avant et après la tenue de l'expérience afin de récolter les informations que nous souhaitions pour connaître leur opinion sur la TA et sur la PE, ainsi que pour en savoir davantage sur leurs connaissances et perceptions en la matière.

Le chapitre est structuré comme suit : Nous expliquons tout d'abord la raison d'être de ce prétest et développons nos différentes hypothèses de travail spécifiques à cette première expérience. La section consacrée à la présentation des résultats est divisée en trois sous-sections : statistiques descriptives, statistiques inférentielles et résultats complémentaires. Nous proposons ensuite une synthèse et une discussion de ces résultats avant de conclure ce chapitre.

4.2 Une étape méthodologique

Le prétest a été envisagé comme une étape méthodologique en vue de l'élaboration d'une deuxième recherche expérimentale (test) que nous avons menée en 2021 et qui constitue le cœur de cette thèse. Cette première expérience nous a effectivement permis non seulement de tester et d'affiner les différents paramètres d'expérience, mais aussi de reformuler nos hypothèses. Nous avons pu, par exemple, non seulement tester et adapter nos critères d'évaluation de la qualité, mais également vérifier que la thématique du texte source pouvait bel et bien être considérée comme une variable dans cette recherche expérimentale. Le but était aussi de pouvoir se familiariser avec l'interface de traduction et de permettre aux *évaluatrices* humaines de prendre en main l'outil d'évaluation. Par ailleurs, les résultats obtenus dans cette première expérience nous ont permis de reconsidérer et de reformuler, en partie, nos hypothèses de travail et d'envisager toute une série d'améliorations méthodologiques en vue du test principal, que nous détaillons au début du chapitre 5.

4.3 Hypothèses de recherche

En nous fondant sur les études antérieures ayant pour objet la comparaison de qualité entre la TH et la PE, en particulier sur celles menées auprès d'étudiantes en traduction, ainsi que sur notre expérience pédagogique et sur notre intuition, nous formulons plusieurs hypothèses afin de répondre aux questions et sous-questions de recherche. Notons qu'au moment où nous avons développé ces hypothèses (en 2017-2018), la littérature spécialisée recelait peu d'études incluant la TA neuronale.

Question A – Quels sont les effets de la post-édition (anglais-français) en contexte d'apprentissage sur la qualité d'un texte cible ?

4.3.1 Hypothèse HA1

Sous-question A1. La post-édition (anglais-français) en contexte d'apprentissage permet-elle d'obtenir des textes de qualité comparable à la qualité en traduction humaine ?

- Hypothèse nulle (HA1.0.)

La post-édition (anglais-français) en contexte d'apprentissage permet d'obtenir des textes de qualité comparable à la qualité en traduction humaine.

- Hypothèse alternative (HA1.1.)

La post-édition (anglais-français) en contexte d'apprentissage ne permet pas d'obtenir des textes de qualité comparable à la qualité en traduction humaine.

En 2018, nous formulons l'hypothèse principale suivante en nous appuyant, entre autres, sur les travaux de Čulo et al. (2014), Depraetere (2010), ainsi que sur ceux de Sycz-Opoń et Gałuskina (2017) : la PE influe négativement sur la qualité du produit fini et ne permet pas à des étudiantes de produire des textes cibles de qualité comparable à une TH. Étant donné que nous cherchons à confirmer l'hypothèse alternative (HA1.1.), nous nous attendons à ce que les textes post-édités contiennent, en moyenne, un plus grand nombre d'erreurs que les TH et qu'ils obtiennent une moins bonne note globale sur 20 par rapport aux TH.

4.3.2 Hypothèse HA2

Sous-question A2. La post-édition (anglais-français) en contexte d'apprentissage présente-t-elle une distribution d'erreurs comparable à la distribution des erreurs en traduction humaine ?

- Hypothèse nulle (HA2.0.)

La post-édition (anglais-français) présente une distribution d'erreurs comparable à la distribution des erreurs en traduction humaine.

- Hypothèse alternative (HA2.1.)

La post-édition (anglais-français) ne présente pas une distribution d'erreurs comparable à la distribution des erreurs en traduction humaine.

En ce qui concerne la qualité à ce niveau, nous partons de l'hypothèse que nous n'allons pas observer la même distribution d'erreurs en PE qu'en TH.

Nous posons les sous-hypothèses suivantes :

4.3.2.1 Sous-hypothèse SHA2.1.

En nous fondant sur l'étude de cas effectuée par Deneufbourg (2018), ainsi que sur les travaux de Forcada (2017)⁸³ portant sur la TA, nous nous attendons à relever davantage d'erreurs d'adéquation, c'est-à-dire des erreurs de traduction sur le plan sémantique (contresens⁸⁴, glissement de sens⁸⁵, omission, etc.) en PE qu'en TH.

4.3.2.2 Sous-hypothèse SHA2.2.

Nous sommes convaincue qu'en dépit des nombreux progrès accomplis, les outils de TA statistique comme neuronale ont toujours tendance à générer un texte qui reste plus proche linguistiquement du texte source qu'une TH ne le serait. À cet égard, nous partageons l'avis de Daems (2016) : « post-editing effort (and, potentially, its quality) is influenced by machine translation output, which, in turn is influenced by source text characteristics » (p. 25). Effectivement, nous pensons que, comme les textes post-édités passent par le filtre de la TA, ceux-ci présenteront un plus haut degré d'interférence⁸⁶ avec la langue source que les textes

⁸³ « Due to the semantic nature of learned representations, errors are usually semantically motivated » (Forcada, 2017, p. 303).

⁸⁴ « Faute de traduction qui consiste à attribuer à un segment du texte de départ un sens contraire à celui qu'a voulu exprimer l'auteur » (Delisle, 2013, p. 649).

⁸⁵ Par « glissement de sens », nous entendons une ambiguïté sémantique, une perte partielle, voire totale, de sens d'un segment du TS sans pour autant que cela mène à un contresens ou à un non-sens.

⁸⁶ Élaborée par Toury (2012), la loi d'interférence en traduction se définit comme suit : « In its most general form, the law of interference would read: in translation, phenomena pertaining to the make-up of the source text tend to force themselves on the translators and be transferred to the target text, whether they manifest themselves in the form of negative transfer (i.e., deviations from normal, codified practices of the target system), or in the

traduits humainement. À la suite de la lecture des travaux de Depraetere (2010) et de l'étude comparative menée par Martikainen et Kübler (2016), nous supposons que les TA post-éditées présenteront un plus grand nombre de marques de littéralité, en particulier de calques⁸⁷ fautifs par rapport aux TH.

4.3.2.3 Sous-hypothèse SHA2.3.

Nous supposons que la PE aura des répercussions négatives sur la fluidité⁸⁸ générale du texte cible, en partie en raison de cette plus grande occurrence attendue de marques de littéralité, en particulier de calques fautifs. Dès lors, nous pensons que nous allons relever dans les TA post-éditées des faiblesses sur les plans de la phraséologie⁸⁹ et de l'idiomaticité⁹⁰ (Čulo et al., 2014 ; Deneufbourg, 2018 ; Loock, 2018 ; Sycz-Opoń et Gałuskińska, 2017). Nous prévoyons par exemple d'observer la présence de tournures calquées, non idiomaticques ou non collocationnelles dans les PE.

4.3.2.4 Sous-hypothèse SHA2.4.

En nous appuyant sur les observations de Nitzke (2016)⁹¹ et de Daems (2016)⁹², nous présumons que les produits de la PE contiendront, en moyenne, moins d'erreurs de langue,

form of positive transfer (i.e., an increase in the frequency of features which do exist in the target system and can be used anyway) » (pp. 311-312).

⁸⁷ Selon Chuquet et Paillard (1989), « [l]e calque consiste à utiliser des éléments lexicaux qui existent dans une langue donnée avec la construction ou le sens qu'ont ces éléments dans l'autre langue » (p. 223) ou encore, 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).

⁸⁸ Par « fluidité », nous entendons une facilité de lecture et de compréhension induite par une cohérence formelle. « S'il faut relire un passage deux ou trois fois avant de le comprendre, c'est peut-être qu'il y a un problème de charnière ou que l'ordre des mots ou des phrases est trop calqué sur l'original. On ne saurait attribuer le manque de fluidité de sa traduction à la piètre qualité du texte de départ » (Delisle, 2013).

⁸⁹ En linguistique, la phraséologie est l'« ensemble des expressions, locutions, collocations et phrases codées conventionnellement dans la langue générale » (Druide, 2022).

⁹⁰ L'idiomaticité peut se définir comme le « caractère d'un texte traduit qui ne « sent » pas la traduction, donc qui a l'air d'avoir été rédigé dans la langue même où il a en fait été traduit » (« Idiomaticité », 2020).

⁹¹ « [A] few categories have to be added, such as spelling mistakes, which occur only occasionally in MT but quite often in human translation » (Nitzke, 2016, p. 91).

⁹² « While the most common MT errors, such as grammatical errors, were easily corrected by most students during post-editing. [...] MT error analysis [...] All subcategories for style and spelling have been grouped together

c'est-à-dire moins de fautes d'orthographe, ainsi que moins d'erreurs de syntaxe et de grammaire que les TH. Nous nous attendons à ce que cette sous-hypothèse soit davantage vérifiée en PE de TA neuronale.

4.3.2.5 Sous-hypothèse SHA2.5.

Enfin, nous supposons que la thématique du texte source aura une influence sur la distribution des erreurs observées en PE et en TH.

4.3.3 Hypothèse HA3

Sous-question A3. La qualité de textes post-édités (anglais-français) en contexte d'apprentissage varie-t-elle en fonction de la technologie de traduction automatique utilisée (TA statistique ou TA neuronale) ?

- Hypothèse nulle (HA3.0.)

La qualité de textes post-édités ne dépend pas de la technologie de traduction automatique utilisée.

- Hypothèse alternative (HA3.1.)

La qualité de textes post-édités dépend de la technologie de traduction automatique utilisée.

À cet égard, nous supposons que la qualité de textes post-édités sera meilleure avec la TA neuronale qu'avec la TA statistique. Plusieurs études attestent une amélioration de la qualité en TA neuronale par rapport à l'ancienne approche statistique (Castilho et al., 2017 ; Wu et al., 2016). Wu et al. (2016) ont montré qu'après évaluation humaine de la qualité, il apparaît que la version neuronale de *Google* permet de réduire le nombre d'erreurs de 60 % en moyenne par rapport à l'ancien système de TA statistique pour les combinaisons anglais-français et anglais-allemand. Pour leur part, Castilho et al. (2017) ont constaté une amélioration en TA neuronale au niveau des scores des métriques automatiques (BLEU, chrF3, METEOR), au niveau de la fluidité et du nombre de segments nécessitant une PE, ainsi qu'au niveau de la préférence qu'ont manifestée les traductrices à l'égard des propositions issues de la TA neuronale. C'est pourquoi nous nous attendons à ce qu'en TA neuronale, les étudiantes produisent des post-éditions qui contiennent, en moyenne, moins d'erreurs que les PE de TA statistique et que ces productions se voient attribuer de meilleures notes que les PE de TA statistique.

into the main categories, since there were very few instances of these subcategories » (Daems, 2016, p. 72 et pp. 119-120).

Dans le but de pouvoir répondre à ces questions, il nous a fallu comparer en détail les produits de la TH et de la PE. Nous avons procédé à l'évaluation de la qualité des productions qui constituent notre corpus. Pour ce faire, nous avons proposé une approche typologique, ce qui nous a permis de classer les erreurs par catégorie, de les comptabiliser et d'attribuer une note sur 20 à chaque production. Cette méthode d'évaluation a été développée au chapitre 3.

4.4 Résultats

Nous rappelons que notre variable dépendante – la qualité du texte cible – a été mesurée tantôt par le nombre d'erreurs relevées dans les productions des étudiantes, tantôt par les notes attribuées aux productions des étudiantes selon une pondération prédéfinie. Nous commençons par analyser les statistiques descriptives puis les statistiques inférentielles. Enfin, nous présentons des résultats complémentaires portant sur le processus même de post-édition. Il convient de noter que ces résultats sont limités par le fait que l'évaluation des productions pour le texte 1 n'a été effectuée que par une seule évaluatrice, à savoir l'autrice de cette thèse. Nous avons veillé à ne pas reproduire ce biais lors du test principal en 2021.

4.4.1 Statistiques descriptives

Dans un souci de clarté, nous allons parcourir différentes figures qui permettent de visualiser les résultats obtenus à la suite de l'évaluation humaine de la qualité des productions.

4.4.1.1 Distribution des erreurs (diagramme circulaire)

Dans un premier temps, nous avons observé la distribution des erreurs relevées dans notre corpus par catégorie indépendamment du mode de traduction et du texte source (voir Figure 21 et Figure 22).

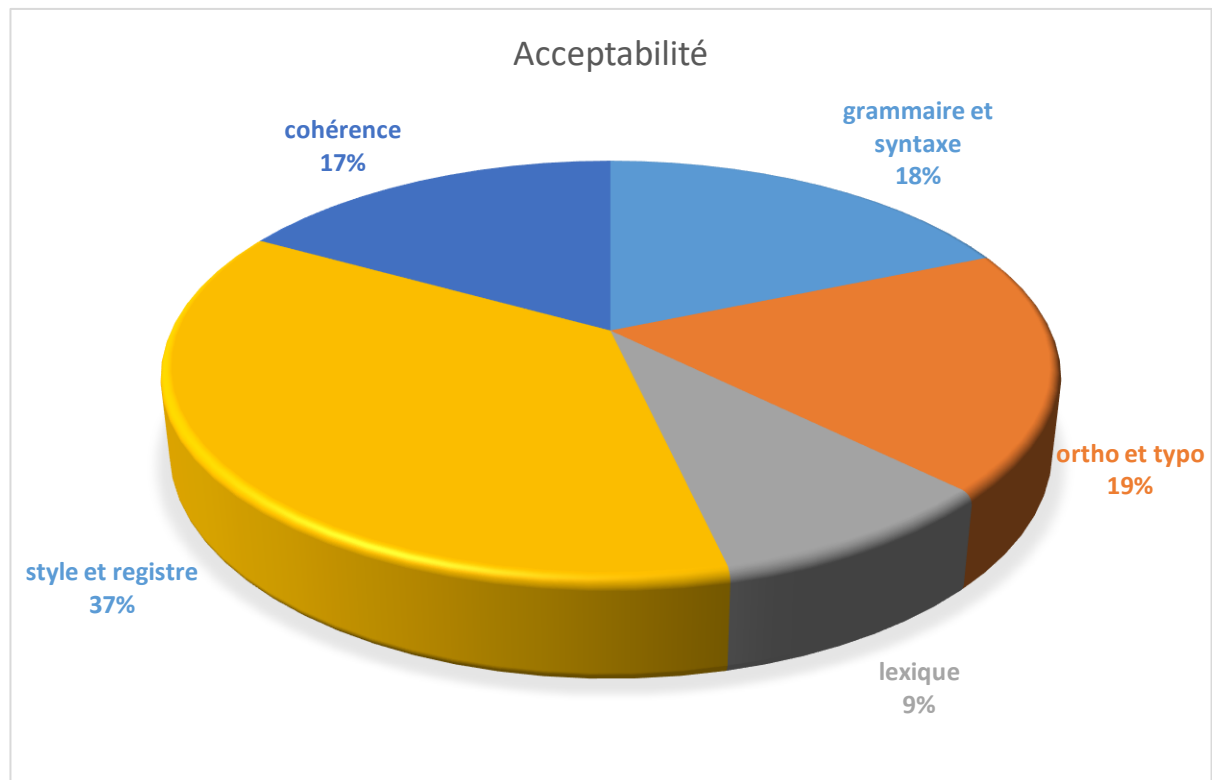


Figure 21 : Diagramme circulaire (acceptabilité) – Prétest

Pour la première phase de l'évaluation (figure 21) — la phase d'acceptabilité (erreurs de langue) — la catégorie d'erreurs la plus représentée est « style et registre » (répétition, fluidité, registre, etc.). Cette catégorie représente 37 % de l'ensemble des erreurs de langue commises par les étudiantes. Les catégories qui arrivent ensuite sont « grammaire et syntaxe » et « ortho et typo » (orthographe et typographie) avec respectivement 18 % et 19 %, suivies de la catégorie « cohérence » (17 %) et enfin « lexique » (9 %).

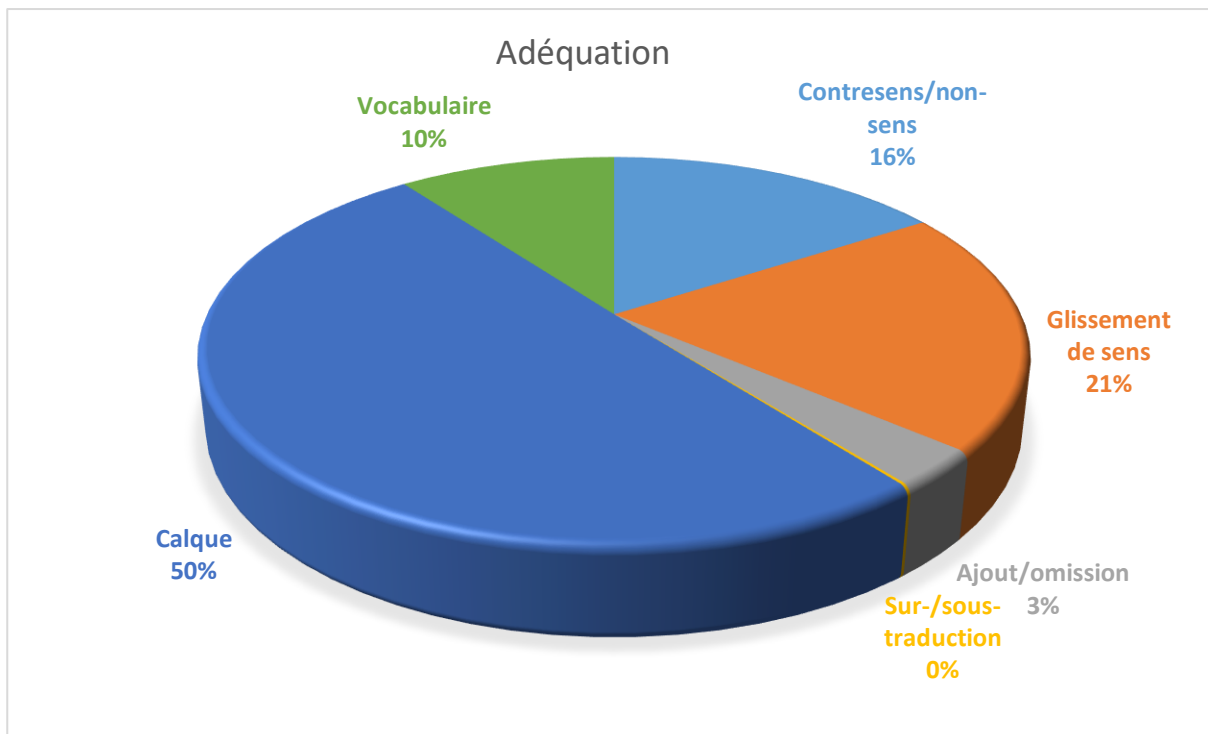


Figure 22 : Diagramme circulaire (adéquation) – Prétest

Pour la deuxième phase de l'évaluation (Figure 22) — la phase d'adéquation (erreurs de sémantique) — la catégorie d'erreurs la plus représentée est celle des « calques ». Cette catégorie représente 50 % de l'ensemble des erreurs de traduction commises par les étudiantes. La deuxième catégorie est celle des « glissements de sens », comptant pour 21 % de l'ensemble des erreurs. Arrivent ensuite les « contresens/non-sens » (16 %), « vocabulaire » (10 %), « ajout/omission » qui représentent seulement 3 %, et enfin la catégorie « sur-/sous-traduction » avec 0 % qui n'est pas représentée.

Les distributions des erreurs selon le mode de traduction et selon la thématique du texte source sont reprises ci-dessous (Figure 23 et Figure 24). Afin de pouvoir tirer des conclusions sur les différences observées sur ces représentations graphiques, nous avons eu recours à la statistique inférentielle (voir section 4.4.2).

4.4.1.2 Distribution des erreurs (histogramme)

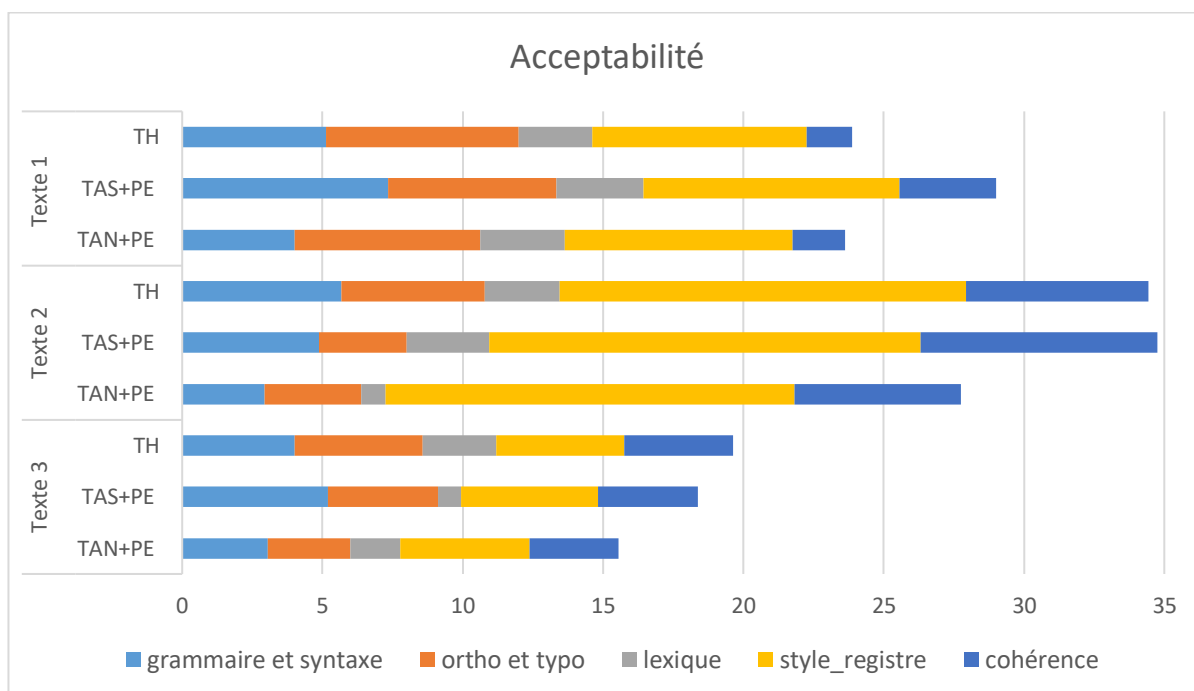


Figure 23 : Histogramme (Acceptabilité) – Prétest

Sur la Figure 23, nous pouvons constater que le texte 3 (sujet scientifique) a généré, en moyenne, nettement moins d'erreurs en phase d'acceptabilité que les deux autres TS, quel que soit le mode de traduction. Nous pouvons également affirmer qu'en PE de TA neuronale, les étudiantes ont commis en moyenne moins d'erreurs en phase d'acceptabilité par rapport aux deux autres modes de traduction, cette observation étant valable pour les trois TS. Enfin, les PE de TA neuronale contiennent manifestement moins d'erreurs de grammaire et de syntaxe que les autres productions, quel que soit le TS.

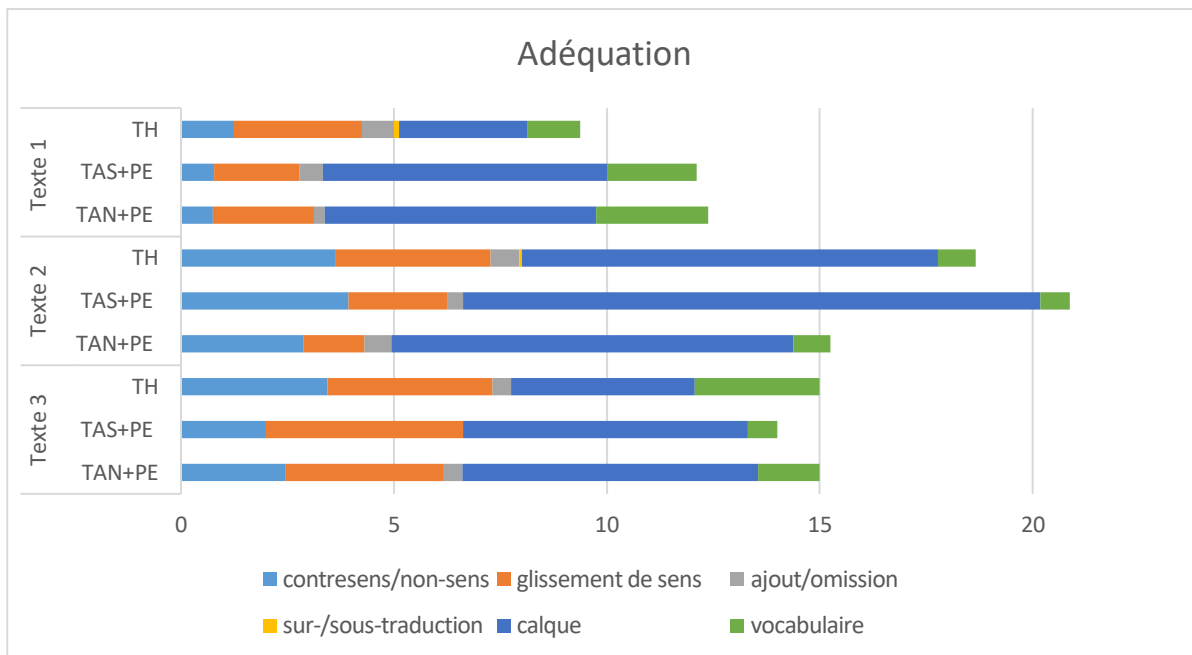


Figure 24 : Histogramme (Adéquation) – Prétest

En phase d'adéquation (Figure 24), le texte source 1 (article de presse) est celui pour lequel les étudiantes ont commis en moyenne le moins d'erreurs par rapport aux deux autres TS. Cet histogramme permet de constater également que la TH est le mode de traduction qui a généré notablement le moins de calques, à tout le moins pour les textes 1 et 3.

4.4.1.3 Comparaison de la qualité par mode de traduction

4.4.1.3.1 Moyenne des notes et taux de réussite

Pour rappel, nous avons calculé, pour chaque production, une note globale selon la pondération que nous avons prédéfinie pour chaque catégorie d'erreurs. Cette note globale combine la note obtenue en phase d'acceptabilité (à raison de 40 pour cent) et la note obtenue en phase d'adéquation (à raison de 60 pour cent). Nous nous sommes également penchée sur le taux de réussite selon le mode de traduction, la note de réussite dans le milieu universitaire étant fixée à 10/20 (contre 12/20 auparavant) (Fédération Wallonie-Bruxelles, 2021). Ce seuil de réussite découle de la réforme de l'enseignement supérieur (hautes écoles et universités), dite du Décret Paysage, qui est en application depuis septembre 2014 dans l'ensemble de la Fédération Wallonie-Bruxelles de Belgique.

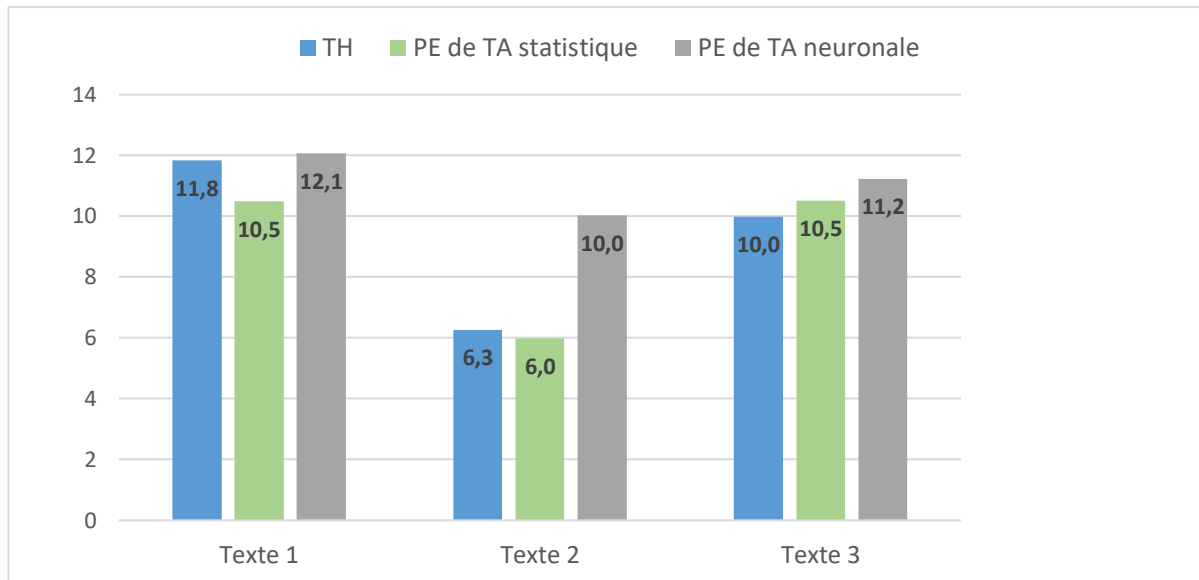


Figure 25 : Moyenne des notes – Prétest

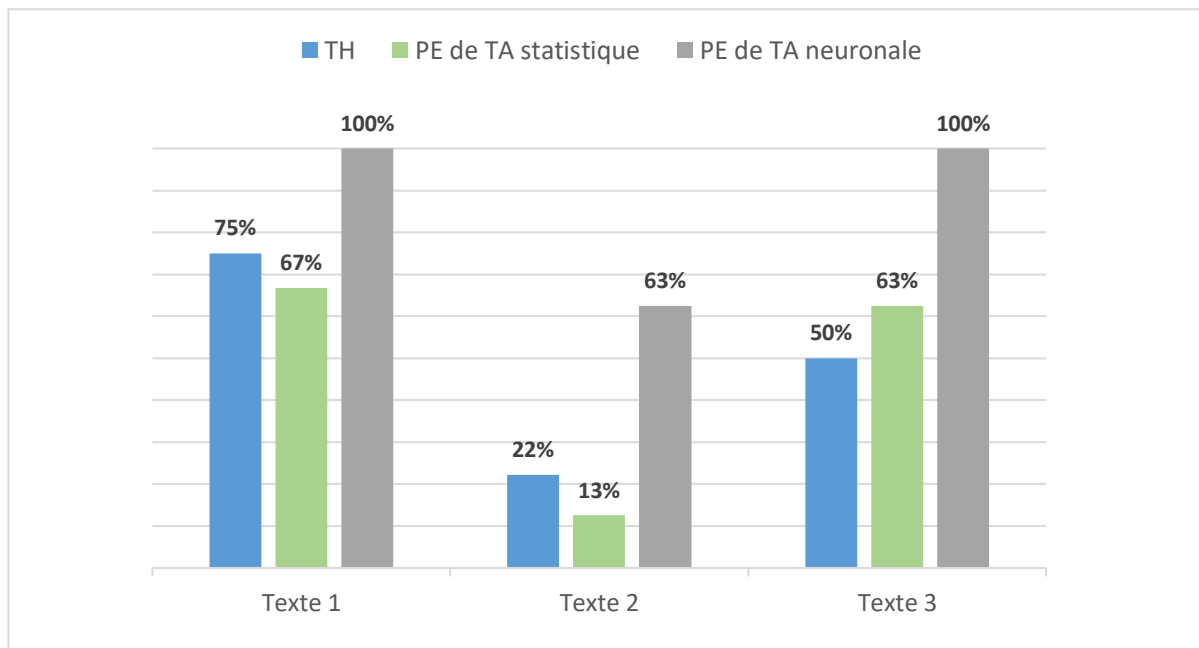


Figure 26 : Taux de réussite – Prétest

Observations générales

Nous pouvons observer sur la Figure 25 que la moyenne des notes la plus élevée est systématiquement obtenue en PE de TA neuronale, et ce pour les trois textes, même s'il faut reconnaître que la différence est parfois minime (cf. texte 3).

Sur la Figure 26, il semblerait que le mode de traduction influe sur le taux de réussite, et ce, pour les trois textes sources. La PE de TA neuronale a permis d'obtenir systématiquement un meilleur taux de réussite qu'en PE de TA statistique ou qu'en TH. Par conséquent, nous pouvons affirmer que, dans cette étude, la probabilité qu'une étudiante obtienne une note de réussite (supérieure ou égale à 10/20) est plus élevée lorsque cette dernière post-édite de la TA neuronale plutôt que lorsqu'elle traduit humainement, ou encore que lorsqu'elle post-édite de la TA statistique.

Observations par texte source

Texte 1 – article de presse

Pour le texte 1, il apparaît que les étudiantes ayant post-édité le texte issu de la TA neuronale ont toutes obtenu une note supérieure ou égale à 10 sur 20. C'est aussi ce groupe qui obtient la moyenne générale la plus élevée (12,1/20). Parmi les étudiantes ayant traduit de manière conventionnelle (TH), le taux de réussite s'élève à 75 % avec une moyenne générale de 11,8 sur 20. Les moins bons résultats sont obtenus en PE de TA statistique avec un taux de réussite de 67 % et une moyenne de 10,5.

Texte 2 – thématique économique

Les résultats obtenus pour le texte 2 montrent des taux de réussite extrêmement bas en TH (13 %) et en PE de TA statistique (22 %). En revanche, le taux de réussite observé en PE de TA neuronale s'élève à 60 % avec une moyenne générale de 10 sur 20.

Texte 3 – thématique scientifique

Nous observons à nouveau un taux de réussite de 100 % en TA neuronale avec une moyenne générale de 11,2 sur 20, toutes les étudiantes ayant obtenu une note supérieure ou égale à 10 sur 20 avec ce mode de traduction. En comparaison, le taux de réussite en TH s'élève à 50 % et en PE de TA statistique à 63 %.

Discussion

Comment expliquer ces divergences de résultats en fonction du texte source ? Nous développons ci-dessous une liste non exhaustive d'hypothèses interprétatives qui permettraient d'expliquer, selon nous, ces divergences de résultats entre les textes sources.

1. Tout d'abord, malgré les précautions prises dans la sélection des textes sources, il est probable que la complexité disciplinaire varie sensiblement d'un texte à l'autre ;
2. Il est également probable que la sévérité d'annotation des erreurs soit propre à chaque évaluatrice ; en d'autres termes, les différentes évaluatrices ne font peut-être pas preuve du même degré de tolérance pour certaines catégories d'erreur bien qu'elles aient recours aux mêmes critères d'évaluation. Ainsi, il se peut qu'une telle divergence de sévérité demeure, même après la phase de mise en commun des annotations ;
3. Les étudiantes en 3^e année de traduction de l'Université de Liège jouissent sans nul doute d'une certaine expérience en traduction d'articles de presse. En revanche, il est fort probable qu'elles n'aient pas ou que peu d'expérience en traduction de textes aux thématiques plus « spécialisées » comme c'est le cas du texte de PWC (texte 2) et du texte de vulgarisation scientifique sur la synesthésie (texte 3) ;
4. Enfin, il se peut que la TA ait généré des traductions brutes de qualité variable en fonction du texte source soumis. Dans notre expérience, nous pensons particulièrement au texte 2, qui comprend des données chiffrées et qui est constitué de phrases relativement longues avec enchâssements. Il se pourrait donc que la TA brute du texte source 2 soit plus médiocre que les autres TA brutes.

4.4.1.4 Moyenne des notes et écart-type

Les deux figures suivantes nous permettent de comprendre un peu mieux les résultats obtenus en matière de taux de réussite. Nous avons repris sur la Figure 27 la moyenne des notes obtenues en fonction du mode de traduction et du texte source ainsi que l'écart-type qui, pour rappel, est un « indice de dispersion autour de la moyenne exprimant la plus ou moins grande hétérogénéité des données » (Gana et Broc, 2018).

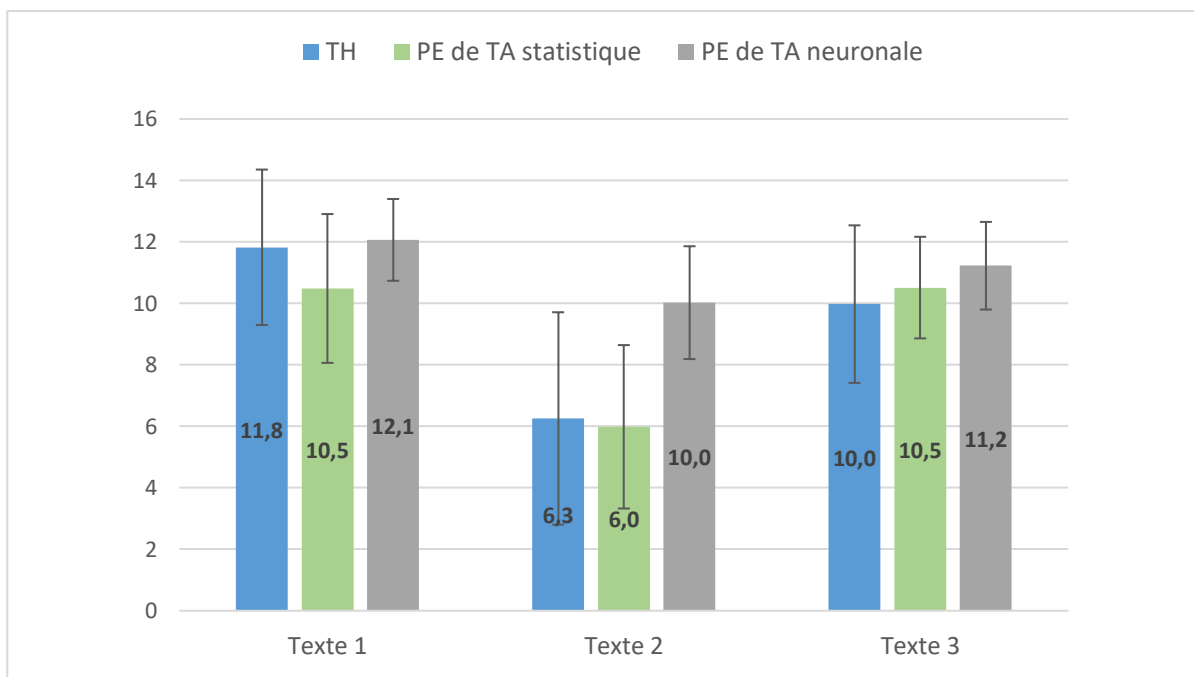


Figure 27 : Moyenne des notes et écart-type – Prétest

Sur la Figure 27, nous pouvons constater une réduction de l'écart-type en PE de TA neuronale particulièrement par rapport à l'écart-type calculé en TH. Autrement dit, ces résultats montrent un (net) resserrement des notes autour de la moyenne en PE de TA neuronale (écart-type parfois divisé par deux par rapport à l'écart-type en TH), et ce, pour les trois textes. Ce resserrement rend compte d'une répartition moins étalée des notes en PE de TA neuronale, ces productions ayant récolté moins de (très) mauvaises notes et moins de (très) bonnes notes par rapport aux deux autres modes.

4.4.1.5 Dispersion des notes autour de la médiane

La Figure 28 en boîte à moustaches illustre clairement ce resserrement des notes. Nous avons repris l'ensemble des notes obtenues selon le mode de traduction, indépendamment du texte source. Pour chaque mode de traduction, la moustache supérieure et la moustache inférieure indiquent respectivement la note la plus élevée et la note la plus basse de la série de notes ; la « boîte » délimitée par le 1^{er} quartile (Q1) et le 3^{ème} quartile (Q3) représente l'intervalle interquartile qui contient au moins 50 pour cent des notes ; le marqueur \times représente la moyenne des notes et la ligne horizontale représente la médiane. Cette figure nous permet de relever que non seulement la moyenne, mais aussi la note médiane sont plus élevées en PE de TA neuronale par rapport aux deux autres modes de traduction. Le resserrement des notes se traduit ici par une réduction de l'écart interquartile (Q3-Q1) : 4,68 en TH ; 4,15 en PE de TA statistique et 1,79 en PE de TA neuronale.

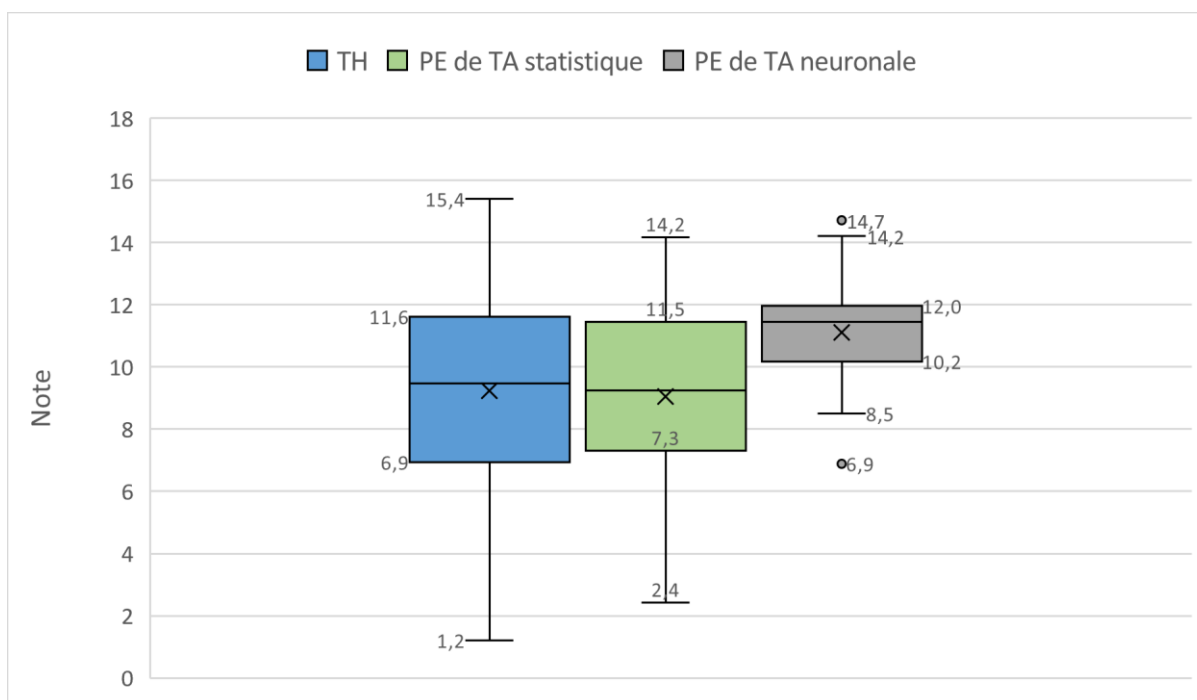


Figure 28 : Dispersion des notes autour de la médiane – Prétest

Ces deux dernières figures nous permettent de dire que l'un des modes de traduction – la PE de TA neuronale – a manifestement généré de faibles écarts dans les notes obtenues comparativement aux deux autres modes, entraînant un resserrement des notes autour de la moyenne et de la médiane. Ainsi, nous formulons l'hypothèse interprétative suivante : le recours à la PE de TA neuronale chez les étudiantes a tendance à homogénéiser l'ensemble des notes. Nous pouvons tirer comme conclusion que, dans notre prétest, la (PE de) TA neuronale entraîne une homogénéisation notable de la qualité des textes cibles. Ces observations nécessitent bien évidemment d'être confirmées par d'autres travaux.

4.4.1.6 Comparaison PE de TA statistique et PE de TA neuronale

Nous avons cherché également à vérifier l'hypothèse HA3, à savoir si la qualité d'un texte post-édité dépend du système de TA employée (TA statistique ou TA neuronale). Pour ce faire, nous avons comparé les productions obtenues en PE de TA statistique aux productions obtenues en PE de TA neuronale au niveau du nombre moyen d'erreurs comptabilisées par texte source.

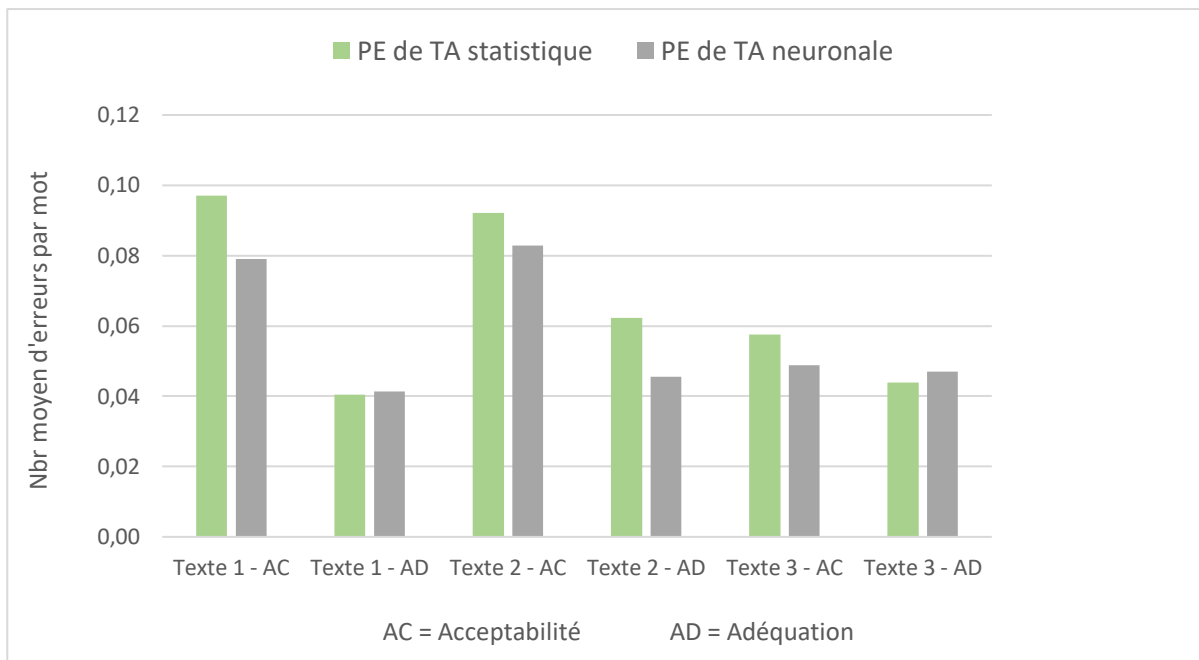


Figure 29 : Comparaison des systèmes de TA (statistique et neuronal)

Sur la Figure 29, nous avons représenté, pour chaque texte, le nombre moyen d'erreurs relevées en phase d'acceptabilité (AC) et le nombre moyen d'erreurs en phase d'adéquation (AD). Afin d'avoir une base de comparaison valable, nous avons tenu à normaliser nos résultats, c'est en prenant comme indicateur le nombre moyen d'erreurs par mot du texte source. Les résultats représentés sur cette figure portent à croire que la PE de TA neuronale permet d'aboutir à des productions de meilleure qualité, c.-à-d. contenant, en moyenne, moins d'erreurs que les PE de TA statistique. Remarquons toutefois que, pour les textes 1 et 3 en phase d'adéquation, les PE de TA neuronale contiennent légèrement plus d'erreurs par rapport aux PE de TA statistique. Nous n'avons pas jugé utile de nous pencher sur ces exceptions étant donné la faible valeur des écarts observés.

4.4.1.7 Performances par étudiante

Au vu des résultats auxquels nous sommes arrivée, nous avons cherché à savoir si les étudiantes, prises individuellement, avaient mieux réussi une des tâches qui leur avaient été assignées ou bien si leurs différentes productions étaient de qualité comparable. Nous nous concentrons désormais uniquement sur la comparaison entre la TH et la PE de TA neuronale, car comme nous venons de le voir, il s'avère que la PE de TA neuronale donne de meilleurs résultats que la PE de TA statistique.

4.4.1.7.1 Nombre moyen d'erreurs

Une fois de plus, nous avons normalisé les résultats en prenant comme référence le nombre d'erreurs par mot du texte source. Sur la Figure 30 et sur la Figure 31, les traductions humaines (TH) sont classées de manière croissante : de la production qui contient le plus grand nombre d'erreurs par mot à celle qui en contient le moins, autrement dit, de la moins bonne TH à la meilleure, indépendamment du texte source.

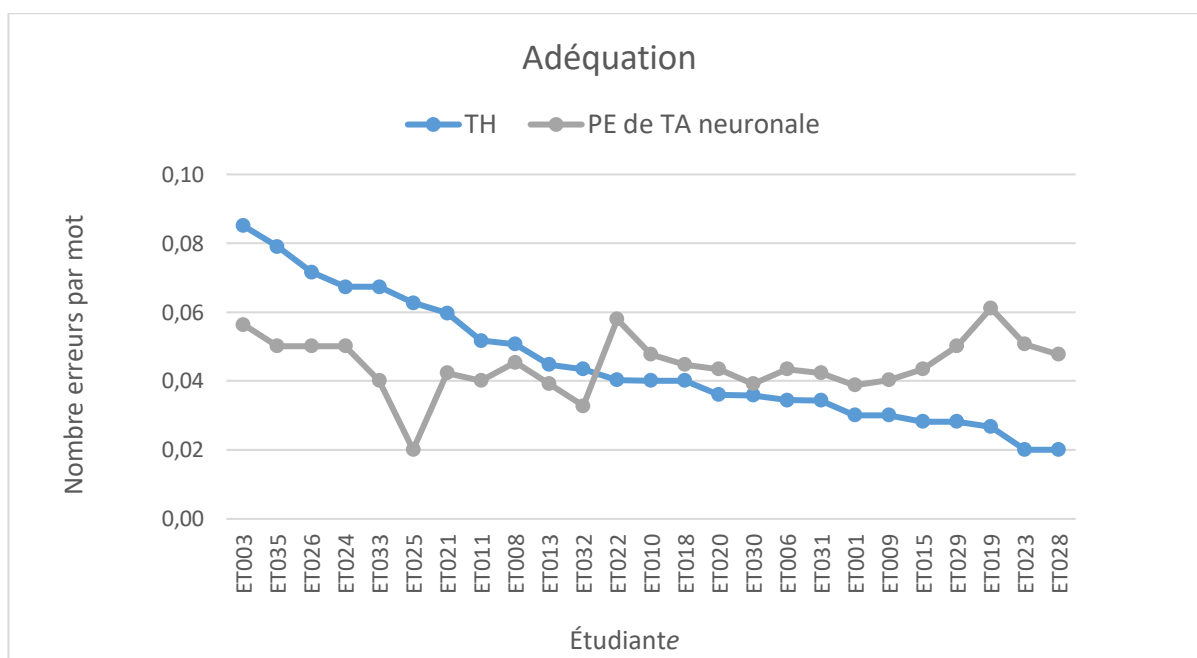


Figure 30 : Performances par étudiante pour la phase d'adéquation

Sur la Figure 30, la TH de l'étudiante 22 (ET022) se situe à la médiane (0,04). Nous pouvons d'emblée constater cette tendance : les étudiantes qui commettent un grand nombre d'erreurs de traduction (phase d'adéquation) dans leur TH commettent systématiquement moins d'erreurs dans leur PE de TA neuronale. En effet, pour toutes les étudiantes dont la TH se situe au-dessus de cette médiane ($> 0,04$), il semblerait que le recours à la PE de TA neuronale ait été bénéfique et leur ait permis d'aboutir à un texte cible final contenant moins d'erreurs que leur TH. À l'inverse, toutes les étudiantes dont la TH se situe en dessous de cette médiane ($< 0,04$) ont produit une PE de TA neuronale qui contient plus d'erreurs par mot que n'en contient leur production humaine.

Nous observons également cette tendance, toutefois de manière moins marquée, pour les erreurs de langue (phase d'acceptabilité) sur la Figure 31. La médiane se situe, cette fois, à 0,06 (ET006 et ET033).

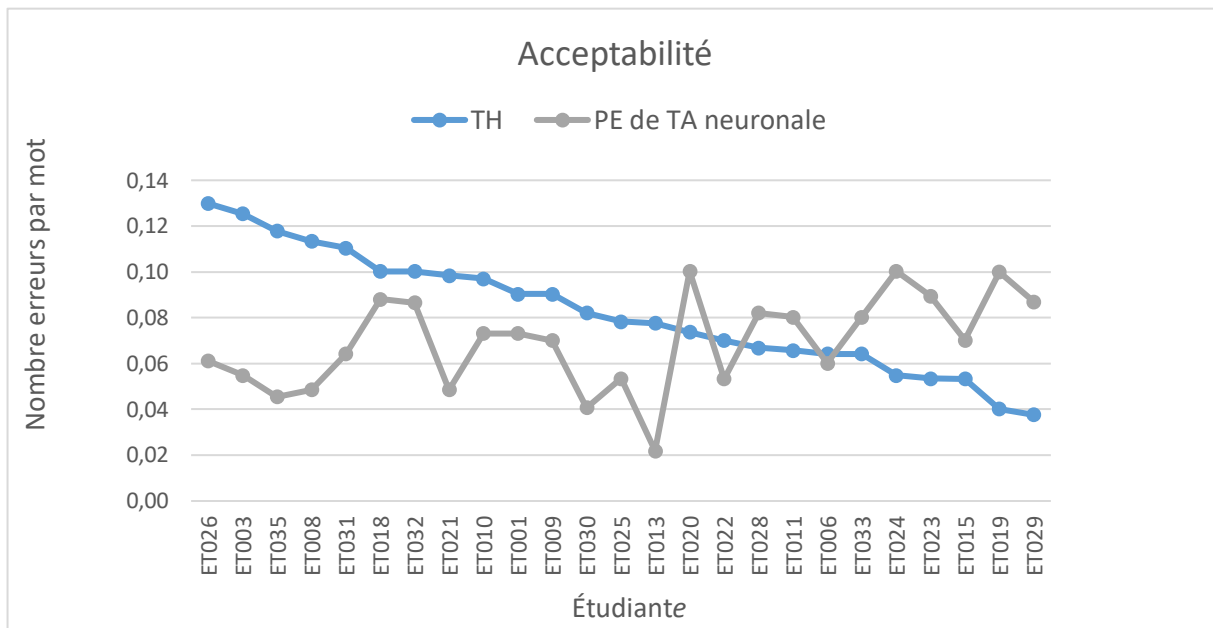


Figure 31 : Performances par étudiante pour la phase d'acceptabilité

4.4.1.7.2 Notation

Nous avons voulu vérifier si cette tendance était également observable au niveau de la notation. Sur la Figure 32, les TH sont classées à présent de manière décroissante : de la meilleure production à la moins bonne.

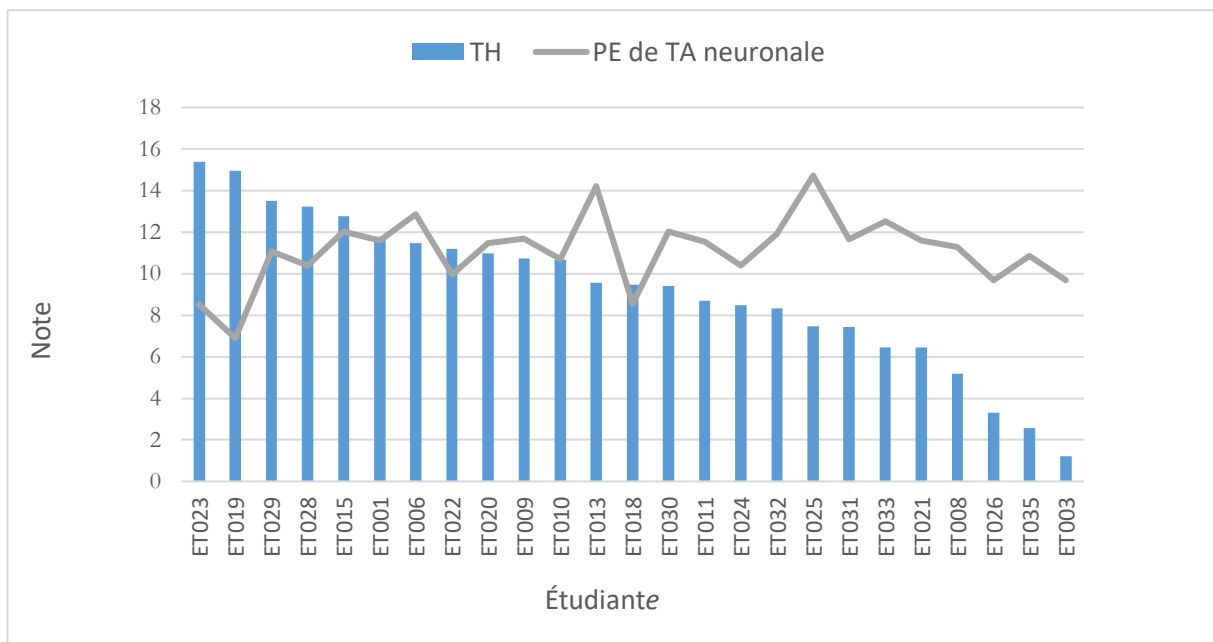


Figure 32 : Notes obtenues par étudiante

Ces résultats ont été obtenus en comparant les notes attribuées en TH et en PE de TA neuronale pour une même étudiante. Les étudiantes que nous considérons comme les plus faibles, à savoir celles dont la note arrondie en TH n'atteint pas le seuil de réussite de 10/20 — ET030 jusque ET003 inclus — sont celles qui obtiennent systématiquement une meilleure note en PE de TA et donc, nous en déduisons que ce sont celles qui bénéficient le plus de ce mode de traduction. À l'inverse, les étudiantes dont nous considérons les productions humaines comme de bonne qualité (note supérieure ou égale à 12/20) — ET023 à ET015 inclus — ont toutes obtenu une note plus basse en PE de TA neuronale. Cette figure nous permet de mettre au jour un « effet nivelant » en PE de TA neuronale (*leveling effect* chez Killman⁹³).

À partir des Figure 30, Figure 31 et Figure 32, nous tirons la conclusion suivante : dans ce prétest, la PE de TA neuronale a eu un effet nivelant sur la qualité des textes cibles. Nous reviendrons sur cet effet nivelant à la section 5.5.1.2.

4.4.2 Statistiques inférentielles

Comme nous l'avons détaillé au chapitre 3, les différentes valeurs obtenues à la suite de l'évaluation humaine des productions ont systématiquement été soumises à un test ANOVA à deux facteurs de type III, en mesures répétées. Nous tentons de déterminer si nos variables indépendantes, à savoir le mode de traduction et le texte source (respectivement « méthode » et « texte » dans les tableaux ci-après), ont eu un effet significatif sur notre variable dépendante, à savoir la qualité du texte cible, mesurée tantôt par le nombre d'erreurs relevées dans les productions des étudiantes (Tableau 17 et Tableau 18), tantôt par les notes attribuées aux productions des étudiantes (Tableau 19, Tableau 20 et Tableau 21).

Avant toute chose, il convient de faire la distinction entre les effets principaux et les effets d'interaction entre les variables :

Les plans factoriels permettent de mettre en évidence d'une part les effets respectifs de chaque VI sur la VD : on parle alors d'effets principaux et d'autre part leurs éventuelles combinaisons en fonction des différentes valeurs ou modalités des VI : on parle alors d'effet d'interaction [...] Un plan factoriel permettra de tester toutes les combinaisons possibles entre les différentes modalités des VI. (Légal, 2008, p. 18)

⁹³ « [P]ost-editing allowed students below, at, and just above the established median to improve the quality of their translation products, especially in cases of students with particularly high HT error ratios and with performances below the median. The HT quality and performances of the translators well above the median are superior to their PET quality and performances [...] Both they and the translators at and just above the median all experienced a reduction in performance levels when post-editing, while those below it all experienced a rise, giving way to a leveling effect » (Killman, 2018, p. 137).

Concrètement, dans les résultats présentés ci-dessous, nous cherchons tout d'abord à voir s'il y a une interaction entre les modes de traduction et la thématique des textes sources⁹⁴. Pour ce faire, nous analysons la ligne « methode:texte » (voir tableaux ci-dessous). Cette ligne considère simultanément l'effet des deux variables. Elle fait l'hypothèse que la valeur de la variable dépendante (qualité du texte cible) dépend de l'interaction des deux variables indépendantes (en l'occurrence, les modes de traduction et les textes sources). Dès lors, une interaction sera significative si la valeur de la variable dépendante est significativement différente pour différentes combinaisons de valeurs des variables indépendantes. Alors que si l'interaction se révèle non significative, nous passons à l'analyse des effets principaux. Pour ce faire, nous examinons à présent les lignes « methode » et « texte » afin de déterminer si nos variables indépendantes, prises séparément, ont eu un effet significatif sur notre variable dépendante. Le cas échéant, nous analysons individuellement les effets de chaque mode de traduction ou de chaque texte source, grâce au test post-hoc de Tukey (ou test de comparaisons multiples de Tukey) pour l'analyse de variance en mesures répétées, ci-après dénommé « test de Tukey ». Pour chaque paire de valeurs d'une même variable indépendante (« methode » ou « texte »), ce test fait l'hypothèse d'une égalité entre les valeurs moyennes obtenues pour la variable dépendante étant donné la valeur de la variable indépendante. S'il y a une différence significative, cela traduit le fait que la variable dépendante a, en moyenne, une valeur différente pour chaque valeur de la variable indépendante. Notons que le seuil de significativité a été fixé à 0,05 pour l'ensemble des tests statistiques, ce qui correspond au seuil de significativité le plus couramment utilisé (Lang, 2007). Si $P > 0,05$, le résultat sera dit statistiquement non significatif tandis que si $P < 0,05$, il sera considéré comme statistiquement significatif ; le risque de se tromper étant inférieur à 5 %.

Les résultats sont présentés comme suit :

Nous analysons les effets des variables indépendantes sur

- A) la qualité globale des productions (4.4.2.1) selon trois niveaux :
- les erreurs en phase d'acceptabilité (4.4.2.1.1) ;
 - les erreurs en phase d'adéquation (4.4.2.1.2) ;
 - les notes globales attribuées aux productions des étudiantes (4.4.2.1.3) ;
- B) la qualité des productions par catégories d'erreur (4.4.2.2).

⁹⁴ Pour rappel : texte 1 : sujet d'actualité ; texte 2 : sujet économique ; texte 3 : sujet scientifique

4.4.2.1 Qualité globale

4.4.2.1.1 Effets sur les erreurs en phase d'acceptabilité

Le premier test ANOVA (Tableau 17) que nous avons effectué permet d'analyser les effets du mode de traduction et du texte source sur le nombre d'erreurs comptabilisées en phase d'acceptabilité (erreurs de langue, problème de fluidité et de cohérence).

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF  F value    Pr(>F)
## methode      341.22  170.61     2 44.000   8.2272 0.0009216 ***
## texte       2609.56 1304.78     2 44.000  62.9191 1.245e-13 ***
## methode:texte  86.47   21.62     4 37.323   1.0425 0.3983988
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau 17 : ANOVA — Distribution des erreurs en phase d'acceptabilité

Interaction mode–texte

Tout d'abord, ce test statistique n'a pas révélé d'interaction significative entre le mode de traduction et le texte (« methode:texte » dans le Tableau 17) ($F= 1,04$; $dfN= 4$; $dfD= 37,32$; $P= 0,3983988$). En d'autres termes, la qualité des productions ne dépend pas significativement de l'interaction entre nos deux variables indépendantes.

Mode de traduction

En revanche, il apparaît que l'effet « mode de traduction » (« methode » dans le Tableau 17) est attesté car $P<0,001$ ($F= 8,22$; $dfN= 2$; $dfD= 44$; $P=0,0009216$), ce qui indique qu'il y a une différence significative entre les distributions de valeurs pour les erreurs d'acceptabilité correspondant à chaque mode de traduction.

L'analyse statistique indique que la distribution des erreurs d'acceptabilité a été influencée par le mode de traduction, et ce, indépendamment du texte source.

Test de Tukey 1

Quant aux résultats du test de Tukey pour les modes de traduction (Tableau C. 1 en annexe), ils nous permettent de mettre au jour un écart statistiquement significatif entre la TA statistique et la TA neuronale ($P<0,001$), ainsi qu'entre la TH et la TA neuronale ($P<0,05$). La TA statistique présente une moyenne supérieure à celle en PE de TA neuronale de 5,06 ; la TH étant en moyenne supérieure à la PE de TA neuronale de 3,67. Cependant, il n'y a pas de différence significative entre la TH et la PE de TA statistique ($P>0,05$).

Dans cette expérience, l'analyse statistique indique que la distribution des erreurs d'acceptabilité a été influencée par le mode de traduction, et ce, indépendamment du texte source. Il apparaît, en effet, que les étudiantes ont commis significativement moins d'erreurs de langue lorsqu'elles ont post-édité de la TA neuronale, qu'elles n'en ont commis en PE de TA statistique et en TH. Ces observations indiqueraient-elles que la TA neuronale brute est davantage exploitable et que par conséquent, les PE de TA neuronale sont les productions qui ont été le moins post-éditées ?

Texte source

L'effet « texte source » (« texte » dans le Tableau 17) est, lui aussi, significatif avec $P < 0,001$ ($F = 62,91$; $dfN = 2$; $dfD = 44$; $P < 0,001$), ce qui indique qu'il y a une différence significative entre les distributions de valeurs pour les erreurs d'acceptabilité correspondant à chaque texte.

Test de Tukey 2

Les résultats du test de Tukey pour les TS (Tableau C. 2 en annexe) montrent l'existence de différences significatives entre tous les textes sources ($P < 0,001$). Ces observations nous amènent à affirmer que, pour la phase d'acceptabilité, la thématique du texte source constitue bel et bien une variable dans cette expérience, au même titre que le mode de traduction. Le texte 3 présente une moyenne inférieure à celle du texte 2 et à celle du texte 1 (respectivement de 14,46 et de 7,64) et le texte 2 présente une moyenne supérieure à celle du texte 1 de 6,81. Remarquons que le texte source pour lequel les étudiantes ont commis significativement le plus grand nombre d'erreurs d'acceptabilité est le texte 2 (sujet économique) et ce, indépendamment du mode de traduction. Tandis que le texte qui a engendré le moins d'erreurs dans le chef des étudiantes est le texte 3 (sujet scientifique).

4.4.2.1.2 Effets sur les erreurs en phase d'adéquation

Le deuxième test ANOVA nous permet d'analyser les effets du mode de traduction et du texte source sur le nombre d'erreurs comptabilisées cette fois en phase d'adéquation (erreurs de traduction).

```

mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## methode      32.08  16.039     2 44.000  1.0966 0.34295
## texte       606.78 303.390     2 44.000 20.7434 4.51e-07 ***
## methode:texte 123.43  30.858     4 38.452  2.1098 0.09831 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Tableau 18 : ANOVA — Distribution des erreurs en phase d'adéquation

Interaction mode–texte

Le test statistique ne révèle pas d'interaction significative entre le mode de traduction et le texte ($F= 2,10$; $dfN= 4$; $dfD= 38,45$; $P=0,09831$). Il ne nous a pas été possible d'inférer de dépendance entre le mode de traduction et le nombre moyen d'erreurs commises en phase d'adéquation.

Mode de traduction

Contrairement aux résultats obtenus pour la phase d'acceptabilité, le test ANOVA effectué pour la phase d'adéquation indique qu'il n'y a pas de dépendance entre le mode de traduction et la distribution des erreurs d'adéquation (erreurs de traduction) : l'effet « mode de traduction » est non significatif ($P>0,05$). Il n'est donc pas nécessaire de passer par un test de Tukey.

Texte source

Il apparaît que l'effet « texte source » est significatif ($F= 20,74$; $dfN= 2$; $dfD= 44$; $P<0,001$).

Test de Tukey

Les résultats du test de Tukey pour les TS (Tableau C. 3 en annexe) montrent que toutes les moyennes entre les textes sont significativement différentes ($P<0,01$). Nous pouvons conclure que la thématique du texte source a influencé la distribution des erreurs d'adéquation dans notre corpus. Le texte 2 présente une moyenne supérieure à celle du texte 1 et à celle du texte 3 (respectivement de 6,97 et de 3,59). Le texte 3 présente un écart significatif par rapport au texte 1 de 3,38. Ainsi, le texte source pour lequel les étudiantes ont commis significativement le plus grand nombre d'erreurs d'adéquation est à nouveau le texte 2 (économique) et ce, indépendamment du mode de traduction. Tandis que le texte qui a engendré le moins d'erreurs d'adéquation de la part des étudiantes est, cette fois, le texte 1 (article de presse).

4.4.2.1.3 Effets sur les notes

La variable dépendante – la qualité du texte cible – est à présent mesurée en termes de notes attribuées aux productions des étudiantes (Tableau 19). Au point 4.4.2.1.3.1, les tests sont menés à partir des notes attribuées en phase d’acceptabilité, au point 4.4.2.1.3.2 à partir de celles attribuées en phase d’adéquation, et au point 4.4.2.1.3.3 à partir des notes globales qui combinent la note obtenue en phase d’acceptabilité (40 %) et la note obtenue en phase d’adéquation (60 %).

4.4.2.1.3.1 Notes d’acceptabilité

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## methode      203.968  101.984     2  44.00  18.2348 1.706e-06 ***
## texte        263.657  131.829     2  44.00  23.5710 1.102e-07 ***
## methode:texte  23.813    5.953     4  36.65   1.0644  0.388
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau 19 : ANOVA — Distribution des notes en phase d’acceptabilité

Le test ANOVA mené pour les notes attribuées en phase d’acceptabilité (Tableau 19) n’indique pas d’interaction significative entre les modes de traduction et les textes sources ($F = 1,06$; $dfN = 4$; $dfD = 36,65$; $P = 0,388$).

Mode de traduction

En revanche, l’effet « mode de traduction » se révèle significatif ($F = 18,23$; $dfN = 2$; $dfD = 44$; $P < 0,001$). Nous pouvons inférer l’existence d’une dépendance entre le mode de traduction et la note obtenue en phase d’acceptabilité.

Test de Tukey 1

Le Tableau C. 4 en annexe reprend les résultats du test de Tukey pour les modes de traduction et fait apparaître que les moyennes sont statistiquement différentes entre la PE de TA statistique et la PE de TA neuronale ($P < 0,001$), ainsi qu’entre la TH et la PE de TA neuronale ($P < 0,001$). La PE de TA neuronale présente une moyenne supérieure à celle en PE de TA statistique de 3,38 et à celle en TH de 3,03. Nous pouvons en conclure que les étudiantes ont obtenu, en moyenne, de meilleures notes en phase d’acceptabilité lorsqu’elles ont post-édité la TA neuronale par rapport aux deux autres modes de traduction.

Texte source

L'effet « texte source » (Tableau 19) est lui aussi significatif ($F = 23,57$; $dfN = 2$; $dfD = 44$; $P < 0,001$). Il existe une influence entre le texte source et les notes attribuées en phase d'acceptabilité.

Test de Tukey 2

Le Tableau C. 5 en annexe fait apparaître que les moyennes des notes sont statistiquement différentes pour les trois textes sources. Il y a un écart faiblement significatif entre le texte 2 et le texte 1 ($P < 0,05$), le texte 1 ayant une moyenne supérieure au texte 2 de 1,76. Ensuite, le texte 3 présente une moyenne significativement supérieure à celle du texte 1 ($P < 0,001$), ainsi qu'à celle du texte 2 ($P < 0,001$) (respectivement de 2,79 et de 4,56).

Au niveau des notes obtenues en phase d'acceptabilité, il ressort que c'est à nouveau le texte 2 qui a récolté, en moyenne, les moins bons résultats, tandis que les productions du texte 3 ont récolté, en moyenne, les meilleures notes, indépendamment du mode de traduction.

4.4.2.1.3.2 Notes d'adéquation

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## methode      14.55   7.275     2 44.000  0.9131    0.4087
## texte       385.21 192.603     2 44.000 24.1738 8.253e-08 ***
## methode:texte  54.16  13.540     4 38.564  1.6995    0.1700
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau 20 : ANOVA — Distribution des notes en phase d'adéquation

Interaction mode–texte

Les résultats de l'ANOVA effectuée sur les notes obtenues en phase d'adéquation (Tableau 20) indiquent une absence d'interaction significative entre les modes de traduction et les textes sources ($F = 1,69$; $dfN = 4$; $dfD = 38,56$; $P = 0,17$).

Mode de traduction

Il ressort également du Tableau 20 que les modes de traduction n'ont pas eu d'influence significative sur les notes en phase d'adéquation ($F = 0,91$; $dfN = 2$; $dfD = 44$; $P = 0,4087$).

Texte source

En revanche, les résultats sont significatifs pour l'effet « texte source » ($F = 24,17$; $dfN = 2$; $dfD = 44$; $P < 0,001$).

Test de Tukey

Le test de Tukey pour les notes obtenues en phase d'adéquation (Tableau C. 6 en annexe) montre que les écarts entre les moyennes de notes pour les différents textes sources sont significatifs ($P < 0,05$). Le texte 2 présente une moyenne inférieure à celle du texte 1, ainsi qu'à celle du texte 3 (respectivement de 5,52 et de 2,22). Enfin, les notes obtenues pour le texte 1 sont en moyenne supérieures à celles obtenues pour le texte 3 de 3,29. Pour la phase d'adéquation, le texte 2 est celui qui a récolté les notes les plus basses, en moyenne, dans notre corpus, indépendamment du mode de traduction. Tandis que le texte 1 est celui qui a posé le moins de difficultés aux étudiantes en phase d'adéquation.

4.4.2.1.3.3 Note globale (acceptabilité et adéquation)

Nous cherchons à présent à savoir si le mode de traduction ou le texte source ont eu une influence significative sur la note globale attribuée aux productions des étudiantes. Pour rappel, la note globale sur 20 comprend la note obtenue en phase d'acceptabilité (40 %) et la note obtenue en phase d'adéquation (60 %).

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF  F value    Pr(>F)
## methode      63.821   31.911     2 44.000   6.6947 0.002895 **
## texte       225.297  112.648     2 44.000  23.6331 1.069e-07 ***
## methode:texte 29.998    7.500     4 38.161   1.5734 0.201092
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau 21 : ANOVA — Distribution des notes globales

Interaction mode–texte

Les résultats de l'ANOVA repris dans le Tableau 21 ne montrent pas d'interaction significative entre les modes de traduction et les textes sources ($F = 1,57$; $dfN = 4$; $dfD = 38,16$; $P = 0,201092$).

Mode de traduction

Les résultats indiquent que l'effet « mode de traduction » est significatif ($F= 6,69$; $dfN= 2$; $dfD= 44$; $P= 0,002895$). Il apparaît que le mode de traduction a eu une influence significative sur la qualité des textes cibles, mesurée ici en notes globales attribuées aux productions des étudiantes et ce, indépendamment du texte source.

Test de Tukey 1

Les résultats du test de Tukey pour les modes de traduction (Tableau C. 7 en annexe) révèlent un écart significatif entre la TA statistique et la TA neuronale ($P<0,01$), ainsi qu'entre la TH et la TA neuronale ($P<0,05$). Nous pouvons tirer comme conclusion que les notes globales des étudiantes sont, en moyenne, meilleures en PE de TA neuronale par rapport aux deux autres modes de traduction, et ce, indépendamment du texte source. Les notes obtenues en PE de TA neuronale sont, en moyenne, supérieures de 1,75 par rapport aux notes obtenues en TH et de 2,11 par rapport à celles obtenues en PE de TA statistique.

Texte source

Pour l'effet « texte source », le résultat de l'ANOVA est significatif ($F= 23,63$; $dfN= 2$; $dfD= 44$; $P<0,001$). On peut dès lors établir l'existence d'une dépendance entre le texte source et la note globale attribuée.

Test de Tukey 2

Le test de Tukey pour les TS (Tableau C. 8 en annexe) indique que les moyennes du texte source 2 sont significativement différentes des moyennes du texte source 1 ($P<0,001$), il en va de même entre le texte 3 et le texte 2 ($P<0,001$). Les textes 1 et 3 présentent une moyenne supérieure à celle du texte 2 (respectivement 4,04 et 3,16). Les notes du texte 2 sont, en effet, inférieures de 4,04 en moyenne par rapport aux notes obtenues pour le texte 1 et inférieures de 3,16 en moyenne par rapport aux productions du texte 3. Au vu de ces observations, nous constatons que le texte 2 (économique), une fois de plus, apparaît comme le texte d'un niveau de complexité plus élevé pour les étudiantes, car il a généré des notes globales significativement plus basses que les deux autres textes sources.

4.4.2.2 Qualité par catégorie d'erreurs

Après avoir observé la qualité globale des productions, nous avons cherché à approfondir nos résultats en appliquant les tests ANOVA et Tukey, cette fois aux différentes catégories d'erreurs.

4.4.2.2.1 Catégories en phase d'acceptabilité

Nous démarrons par l'analyse des résultats pour les catégories en phase d'acceptabilité

4.4.2.2.1.1 Grammaire et syntaxe

Interaction mode–texte

Le test ANOVA pour les erreurs de grammaire et de syntaxe (Tableau C. 9 en annexe) ne révèle pas d'interaction significative entre les modes de traduction et les textes sources.

Mode de traduction

Les résultats du test ANOVA font apparaître une relation de dépendance entre le mode de traduction et le nombre d'erreurs de grammaire et de syntaxe présentes dans les productions des étudiantes ($P < 0,001$) (Tableau C. 9).

Test de Tukey 1

Le test de Tukey pour les modes de traduction (Tableau C. 10 en annexe) révèle deux écarts significatifs : d'une part entre la PE de TA statistique et la PE de TA neuronale ($P < 0,001$) et, d'autre part, entre la TH et la PE de TA neuronale ($P < 0,01$). Pour les erreurs de grammaire et de syntaxe, la PE de TA statistique présente une moyenne significativement supérieure (+2,46) à celle en PE de TA neuronale et la TH présente une moyenne significativement supérieure (+1,59) à celle en PE de TA neuronale. Nous pouvons en conclure qu'en PE de TA neuronale, les étudiantes ont commis, en moyenne, moins d'erreurs de grammaire et de syntaxe par rapport aux deux autres modes de traduction.

Texte source

Le test ANOVA fait également apparaître une relation de dépendance entre le texte source et le nombre d'erreurs de grammaire et de syntaxe recensées dans notre corpus ($P < 0,05$) (Tableau C. 9).

Test de Tukey 2

Quant au test de Tukey sur les textes sources (Tableau C. 11 en annexe), il montre un seul résultat significatif : le texte 1 présente une moyenne supérieure à celle du texte 3 de 1,40 ($P < 0,05$). Il s'avère que les productions finales du texte source 1 (article de presse) contiennent davantage d'erreurs de grammaire et de syntaxe que celles du texte 3 (scientifique), indépendamment du mode de traduction.

4.4.2.2.1.2 Orthographe

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## methode      20.8072  10.4036     2  44.000  16.6276 4.184e-06 ***
## texte        4.5476   2.2738     2  44.000   3.6341 0.03462 *
## methode:texte 1.1977   0.2994     4  37.286   0.4785 0.75123
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interaction mode–texte

Le test ANOVA ne permet pas de mettre au jour une interaction significative entre les modes de traduction et les textes sources (Tableau C. 12 en annexe).

Mode de traduction

Le test ANOVA nous permet d’inférer une relation de dépendance entre le mode de traduction et le nombre de fautes d’orthographe relevées dans les productions des étudiantes ($P < 0,001$) (Tableau C. 12).

Test de Tukey 1

Le test de Tukey pour les modes de traduction (Tableau C. 13 en annexe) révèle un écart significatif entre la TH et la PE de TA neuronale ($P < 0,001$), ainsi qu’entre la TH et la PE de TA statistique ($P < 0,001$). Nous pouvons conclure de ces résultats que lorsque les étudiantes ont traduit « humainement » au cours de cette expérience, *elles* ont commis, en moyenne, davantage de fautes d’orthographe que lorsqu’*elles* ont post-édité la TA que ce soit avec le moteur statistique ou neuronal.

Texte source

Nous pouvons également inférer une relation entre le texte source et le nombre de fautes d’orthographe ($P < 0,05$) (Tableau C. 12).

Test de Tukey 2

Concernant les textes sources, le test de Tukey (Tableau C. 14 en annexe) révèle un écart significatif entre le texte 3 et le texte 2 ($P < 0,05$), le texte 2 présentant une moyenne supérieure au texte 3 de 0,52. Il y a également une différence marginalement significative ($P < 0,1$) entre les textes 1 et 3, le texte 1 présentant une moyenne supérieure au texte 3 de 0,52. Ainsi, le texte pour lequel les étudiantes ont commis, en moyenne, le moins de fautes d’orthographe est le texte 3 (sujet scientifique).

4.4.2.2.1.3 Typographie et ponctuation

Interaction mode–texte

Il n'y a pas d'interaction significative entre les modes de traduction et les textes sources pour cette catégorie d'erreurs (Tableau C. 15 en annexe).

Texte source

Le test ANOVA (Tableau C. 15) ne nous permet d'inférer qu'une seule influence significative entre les textes sources et le nombre d'erreurs de typographie et de problèmes de ponctuation ($P < 0,001$). Le mode de traduction n'a manifestement pas eu d'effet sur cette catégorie d'erreurs.

Test de Tukey

Les résultats du test de Tukey pour les TS (Tableau C. 16 en annexe) indiquent deux écarts significatifs : le texte 1 présente une moyenne supérieure à celle du texte 2 ($P < 0,001$), ainsi qu'à celle du texte 3 ($P < 0,001$). Ainsi, les productions du texte 1 (article de presse) sont celles qui contiennent le plus d'erreurs de typographie et de problèmes de ponctuation, et ce, indépendamment du mode de traduction.

4.4.2.2.1.4 Lexique

Interaction mode–texte

Le test ANOVA révèle que, pour la catégorie « lexique », l'interaction « méthode - texte source » est significative ($P < 0,05$) (Tableau C. 17). Cela signifie que le mode de traduction a eu une influence sur le nombre d'erreurs de lexique, mais uniquement en interaction avec le texte source. Autrement dit, les différences observées entre deux modes de traduction varient d'un texte à l'autre. Étant donné cette interaction, il n'est pas pertinent d'analyser les effets principaux isolément.

4.4.2.2.1.5 Style et registre

Cette catégorie regroupe des erreurs de type : problèmes de fluidité, de registre, phrases trop longues, répétition, etc.

Interaction mode–texte

Au niveau des erreurs de « style et registre », le test ANOVA ne révèle pas d'interaction significative entre les modes de traduction et les textes sources (Tableau C. 18 en annexe).

Mode de traduction

Une fois de plus, le mode de traduction n'a manifestement pas eu d'effet significatif sur cette catégorie d'erreurs (Tableau C. 18).

Texte source

Le test ANOVA ne donne qu'un résultat significatif nous permettant de dire qu'il y a une dépendance entre les textes source et le nombre d'erreurs de style et de registre ($P < 0,001$) (Tableau C. 18).

Test de Tukey

Le test de Tukey pour les TS (Tableau C. 19 en annexe) révèle que toutes les moyennes entre les textes sont significativement différentes ($P < 0,001$). Le texte 2 présente une moyenne supérieure à la moyenne du texte 1 de 6,52, ainsi qu'à la moyenne du texte 3 de 10,12. Enfin, le texte 3 présente une moyenne inférieure à celle du texte 1 de 3,60. Nous pouvons en conclure que le texte 2 (économique) est celui qui a généré en moyenne, le plus d'erreurs de style et de registre, et ce, de manière très notable ; tandis que le texte 3 (scientifique) est celui qui en a généré le moins.

4.4.2.2.1.6 Cohérence

Pour rappel, cette catégorie concerne des erreurs de type : problèmes de logique interne, irrégularité dans l'emploi de termes, concordance des temps, etc.

Interaction mode–texte

Le test ANOVA ne montre pas d'interaction significative entre les modes de traduction et les textes sources (Tableau C. 20 en annexe).

Mode de traduction

Pour cette catégorie, les résultats indiquent que le mode de traduction a eu une influence sur le nombre d'erreurs de cohérence ($P < 0,01$) (Tableau C. 20).

Test de Tukey 1

Le test de Tukey pour les modes de traduction (tableau C.21 en annexe) fait apparaître deux écarts significatifs : le PE de TA statistique présente, en moyenne, un résultat supérieur au résultat obtenu en PE de TA neuronale de 1,48 ($P < 0,01$), ainsi qu'au résultat obtenu en moyenne en TH de 1,14 ($P < 0,05$). Dans cette expérience, nous pouvons affirmer que la PE de TA statistique a engendré davantage d'erreurs de cohérence que les deux autres modes.

Texte source

Les résultats indiquent que le texte source a eu une influence sur les erreurs de cohérence ($P < 0,001$) (Tableau C. 20).

Test de Tukey 2

Le test de Tukey pour les TS (Tableau C. 22 en annexe) révèle que toutes les moyennes entre les textes sont significativement différentes. Le texte 2 présente une moyenne supérieure à la moyenne du texte 1 de 4,64 ($P < 0,001$), ainsi qu'à la moyenne du texte 3 de 3,42 ($P < 0,001$). Enfin, le texte 3 a une moyenne supérieure à celle du texte 1 de 1,21 ($P < 0,05$). Une fois de plus, le texte source qui a généré le plus d'erreurs de cohérence dans le chef des étudiantes est le texte 2 (économique) et celui qui en a généré le moins est le texte 1 (article de presse).

4.4.2.2.2 Catégories en phase d'adéquation

Passons à présent à l'analyse des résultats pour les catégories en phase d'adéquation.

4.4.2.2.2.1 Contresens et non-sens**Interaction mode–texte**

Le test ANOVA ne montre pas d'interaction significative entre les modes de traduction et les textes sources (Tableau C. 23 en annexe).

Mode de traduction

Le mode de traduction n'a manifestement pas eu d'influence sur le nombre de contresens et de non-sens comptabilisés dans les productions finales (Tableau C. 23).

Texte source

Ce test ANOVA ne donne qu'un seul résultat significatif qui nous permet d'inférer une relation entre le texte source et le nombre de contresens et de non-sens relevés dans le corpus ($P < 0,001$) (Tableau C. 23).

Test de Tukey

Les résultats du test de Tukey pour les TS (Tableau C. 24 en annexe) indiquent que le nombre moyen de contresens et de non-sens est statistiquement différent pour les trois textes sources. Le texte 2 présente une moyenne supérieure à celle du texte 1 de 2,54 ($P < 0,001$) et à celle du texte 3 de 0,84 ($P < 0,05$), et enfin, le texte 3 a une moyenne supérieure à celle du texte 1 de 1,70 ($P < 0,001$). Le texte source qui a généré, en moyenne, le plus de contresens et de non-sens, est le texte 2 (économique), et ce, tous modes de traduction confondus ; celui qui en a généré le moins est le texte 1 (article de presse)

4.4.2.2.2 Glissement de sens

Interaction mode–texte

Au niveau des glissements de sens, le test statistique ne révèle pas d'interaction significative entre les modes de traduction et les textes sources (Tableau C. 25 en annexe).

Mode de traduction

Le test ANOVA fait apparaître une dépendance entre le mode de traduction et le nombre de glissements de sens comptabilisés dans le corpus ($P < 0,05$) (Tableau C. 25).

Test de Tukey 1

Le test de Tukey pour les modes de traduction (Tableau C. 26 en annexe) révèle une différence significative au niveau des glissements de sens entre la TH et la PE de TA neuronale ($P < 0,05$). Cette différence peut s'interpréter de la manière suivante : les étudiantes ont commis, en moyenne, plus de glissements de sens lorsqu'elles ont traduit « humainement » que lorsqu'elles ont post-édité de la TA neuronale. En moyenne, il y a 1,00 glissement de sens en plus en TH qu'en PE de la TA neuronale. Il apparaît dans cette expérience que le recours à la PE de TA neuronale a permis de réduire les occurrences de glissements de sens par rapport à la TH.

Texte source

Les résultats repris dans le tableau C. 25 en annexe indiquent également une dépendance entre le nombre de glissements de sens et le texte source ($P < 0,001$).

Test de Tukey 2

Le test de Tukey pour les TS (Tableau C. 27 en annexe) fait apparaître un écart significatif entre le texte 3 et le texte 1 ($P < 0,001$) ; le texte 3 présentant une moyenne supérieure de 1,615, ainsi qu'un écart significatif entre le texte 3 et le texte 2 ($P < 0,001$) ; le texte 3 présentant une moyenne supérieure de 1,60. Nous pouvons dès lors en conclure que le texte 3 (scientifique) est le texte source qui a généré le plus de glissements de sens dans les productions finales, indépendamment du mode de traduction.

4.4.2.2.3 Ajout/Omission

Pour la catégorie ajout/omission, le test ANOVA (Tableau C. 28 en annexe) ne donne aucun résultat significatif, ce qui signifie que ni le mode de traduction ni le texte source n'ont eu d'influence significative sur le nombre d'ajouts ou d'omissions dans notre corpus.

4.4.2.2.2.4 Sur-/sous-traduction

Le nombre d'erreurs relevées dans cette catégorie est trop faible pour pouvoir appliquer les tests statistiques et obtenir des résultats statistiques significatifs.

4.4.2.2.2.5 Calque

Interaction mode–texte

En ce qui concerne les erreurs de calques, il n'y a pas d'interaction significative entre les modes de traduction et les textes sources qui influencerait les moyennes (Tableau C. 29 en annexe).

Mode de traduction

Le test ANOVA (Tableau C. 29) révèle une influence significative entre le mode de traduction et le nombre de calques recensés dans les productions ($P < 0,001$).

Test de Tukey 1

Les résultats du test de Tukey pour les modes de traduction (Tableau C. 30 en annexe) montrent que les moyennes sont statistiquement différentes entre la TH et la PE de TA neuronale ($P < 0,05$), ainsi qu'entre la TH et la PE de TA statistique ($P < 0,001$). La TH présente une moyenne inférieure à celle de la PE de TA neuronale de 1,88, et à celle de la PE de TA statistique de 3,27. Ainsi, nous pouvons affirmer que les étudiantes ont commis, en moyenne, davantage de calques fautifs en PE, que ce soit de TA statistique ou de TA neuronale, qu'elles n'en ont commis en TH, et ce, indépendamment du texte source. En outre, l'écart le plus important étant celui observé entre la TH et la PE de TA statistique, cela nous porte à penser que la PE de TA neuronale génèrerait moins de calques fautifs dans le chef des étudiantes que l'ancienne génération de TA. Ce constat vient confirmer l'une de nos hypothèses (SHA2.3). En effet, nous nous attendions à ce que les PE contiennent un plus grand nombre de calques fautifs que les TH.

Texte source

De même, les résultats montrent une influence significative entre le texte source et le nombre de calques ($P < 0,001$) (Tableau C. 29).

Test de Tukey 2

Le test de Tukey pour les TS (Tableau C. 31 en annexe) révèle un écart significatif entre le texte 2 et le texte 1 ($P < 0,001$), ainsi qu'entre le texte 3 et le texte 2 ($P < 0,001$). Nous pouvons en déduire que le texte source qui a comptabilisé, en moyenne, le plus de calques fautifs est le texte 2 (économique), car il présente une moyenne supérieure à celle du texte 1 et à celle du texte 3 (respectivement de 5,57 et de 4,94).

4.4.2.2.6 Vocabulaire

Interaction mode–texte

Pour la catégorie « vocabulaire », les résultats de l'ANOVA montrent que l'interaction « méthode - texte source » est significative ($P < 0,01$) (Tableau C. 32).

Cela signifie que le mode de traduction a eu une influence, mais uniquement en interaction avec le texte source. Autrement dit, les différences observées entre deux modes de traduction varient d'un texte à l'autre.

Mode de traduction

La variable « mode de traduction », seule, n'a pas eu d'influence sur le nombre d'erreurs de vocabulaire.

Texte source

Il apparaît également dans le Tableau C. 32 que la variable « texte source », seule, a eu une influence significative sur le nombre d'erreurs de vocabulaire ($P < 0,01$).

Test de Tukey

Les résultats du test de Tukey pour les TS (Tableau C. 33 en annexe) révèlent que l'écart entre les textes 2 et 1, ainsi que l'écart entre les textes 3 et 2 sont statistiquement significatifs (respectivement $P < 0,001$ et $P < 0,05$). Il ressort que le texte 2 est celui qui a généré, en moyenne, le moins d'erreurs de vocabulaire (- 1,17 par rapport au texte 1 et -0,87 par rapport au texte 3).

4.4.2.3 Tableaux récapitulatifs

4.4.2.3.1 Influence du mode de traduction sur le nombre d'erreurs

En résumé, dans ce prétest, la PE de TA neuronale a généré significativement moins d'erreurs d'acceptabilité que la TH. Lorsque l'on considère les catégories d'erreurs isolément, la PE (TA statistique et TA neuronale) a généré significativement plus de calques fautifs et moins de fautes d'orthographe que la TH. De plus, la PE de TA neuronale (*DeepL*) a entraîné significativement moins de glissements de sens que la TH, et moins d'erreurs de grammaire et de syntaxe que la PE de TA statistique et la TH. Enfin, la PE de TA statistique a généré significativement davantage d'erreurs de cohérence que la PE de TA neuronale et la TH.

Influence mode de traduction (nombre d'erreurs)	TH		TAS + PE		TAN + PE	
	vs TAS + PE	vs TAN + PE	vs TH	vs TAN + PE	vs TH	vs TAS + PE
Acceptabilité	/	+	/	+	-	-
Adéquation	/	/	/	/	/	/
Catégories acceptabilité						
Grammaire et syntaxe	/	+	/	+	-	-
Fautes d'orthographe	+	+	-	/	-	/
Typographie et ponctuation	/	/	/	/	/	/
Lexique	/	/	/	/	/	/
Style et registre	/	/	/	/	/	/
Cohérence	+	/	-	+	/	-
Catégories adéquation						
Contresens et non-sens	/	/	/	/	/	/
Glissement de sens	/	+	/	/	-	/
Ajout/omission	/	/	/	/	/	/
Sur-/sous-traduction	/	/	/	/	/	/
Calques	-	-	+	/	+	/
Vocabulaire	/	/	/	/	/	/
Légende						
- : Effet négatif – le mode de traduction a eu pour effet de générer significativement moins d'erreurs par rapport à l'autre mode de traduction.						
+ : Effet positif – le mode de traduction a eu pour effet de générer significativement davantage d'erreurs par rapport à l'autre mode de traduction.						
/ : résultats non significatifs ou présence d'une interaction significative entre les modes de traduction et les textes sources. Nous n'avons pas pu conclure à l'existence d'une relation entre le mode de traduction et le nombre d'erreurs.						

Tableau 22 : Influence du mode de traduction sur le nombre d'erreurs - Prétest

4.4.2.3.2 Influence du texte source sur le nombre d'erreurs

Influence texte source (nombre d'erreurs)	Texte 1 (presse)		Texte 2 (économique)		Texte 3 (scientifique)	
	vs texte 2	vs texte 3	vs texte 1	vs texte 3	vs texte 1	vs texte 2
Acceptabilité	-	+	+	+	-	-
Adéquation	-	-	+	+	+	-
Catégories acceptabilité						
Grammaire et syntaxe	/	+	/	/	-	/
Fautes d'orthographe	/	/	/	+	/	-
Typographie et ponctuation	+	+	-	/	-	/
Lexique	/	+	/	/	-	/
Style et registre	-	+	+	+	-	-
Cohérence	-	-	+	+	+	-
Catégories adéquation						
Contresens et non-sens	-	-	+	+	+	-
Glissement de sens	/	-	/	-	+	+
Ajout/Omission	/	/	/	/	/	/
Sur-/sous-traduction	/	/	/	/	/	/
Calques	-	/	+	+	/	-
Vocabulaire	+	/	-	-	/	+
Légende						
- : Effet négatif – le texte source a eu pour effet de générer significativement moins d'erreurs par rapport à l'autre texte source.						
+ : Effet positif – le texte source a eu pour effet de générer significativement davantage d'erreurs par rapport à un l'autre texte source.						
/ : résultats non significatifs ou présence d'une interaction significative entre les modes de traduction et les textes sources. Nous n'avons pas pu conclure à l'existence d'une relation entre le texte source et le nombre d'erreurs.						

Tableau 23 : Influence du texte source sur le nombre d'erreurs – Prétest

4.4.2.3.3 Influence sur les notes

Concernant l'influence de la variable « mode de traduction » sur les notes attribuées aux productions, nous pouvons dire que la PE de TA neuronale a entraîné davantage de meilleures notes d'acceptabilité et de meilleures notes globales (acceptabilité + adéquation) par rapport à la TH et à la PE de TA statistique, quel que soit le TS.

Alors qu'au niveau de l'influence du texte source sur les notes, il apparaît que le texte 2 (économique) est celui pour lequel les productions ont obtenu les notes les moins élevées et à l'inverse, que le texte 1 (presse) a récolté les meilleures notes, quel que soit le mode de traduction.

Influence mode de traduction (notes)	TH		TAS + PE		TAN + PE	
	vs TAS + PE	vs TAN + PE	vs TH	vs TAN + PE	vs TH	vs TAS + PE
Acceptabilité	/	-	/	-	+	+
Adéquation	/	/	/	/	/	/
Acceptabilité + Adéquation	/	-	/	-	+	+
Influence texte source (notes)	Texte 1		Texte 2		Texte 3	
	vs texte 2	vs texte 3	vs texte 1	vs texte 3	vs texte 1	vs texte 2
Acceptabilité	+	-	-	-	+	+
Adéquation	+	+	-	-	-	+
Acceptabilité + Adéquation	+	+	-	-	-	+
Légende						
- : Effet négatif – Le mode de traduction/le texte source a eu pour effet significatif de générer, en moyenne, des notes plus basses par rapport à l'autre mode de traduction/texte source.						
+ : Effet positif – Le mode de traduction/le texte source a eu pour effet significatif de générer, en moyenne, des notes plus élevées par rapport à l'autre mode de traduction/texte source.						
/ : résultats non significatifs ou présence d'une interaction significative entre les modes de traduction et les textes sources. Nous n'avons pas pu conclure à l'existence d'une relation entre le mode de traduction/texte source et les notes attribuées aux productions.						

Tableau 24 : Influence du mode de traduction sur les notes – Prétest

4.4.3 Résultats complémentaires

Concernant le processus même de PE, nous présentons premièrement les résultats du calcul des gains de productivité en fonction des différents modes de traduction (par texte et par étudiante). Et deuxièmement, nous présentons les résultats du calcul d'effort moyen de PE en fonction du type de TA.

4.4.3.1 Étude du processus : Productivité

4.4.3.1.1 Productivité par texte

Le Tableau 25 présente les gains de productivité en fonction du texte source. Pour l'ensemble des productions, nous avons calculé le temps moyen de traduction/PE en secondes par mot du texte source, et comparé ces résultats pour mesurer les effets sur le taux moyen de productivité.

	TH (s/mot)	TAS+PE ⁹⁵ (s/mot)	TAN+PE ⁹⁶ (s/mot)	TAS+PE vs TH	TAN+PE vs TH	TAN+PE vs TAS+PE
Texte 1	8,01	7,88	4,33	+2 %	+46 %	+45 %
Texte 2	8,99	4,55	4,74	+49 %	+47 %	-4 %
Texte 3	7,53	6,85	6,60	+9 %	+12 %	+4 %

Tableau 25 : Gains de productivité par texte – Prétest

Gain de productivité

On observe systématiquement un gain de productivité en PE par rapport à la TH. Les gains de productivité les plus importants ont été observés entre la PE de TA neuronale et la TH (+ 46 % pour le texte 1 ; + 47 % pour le texte 2 et + 12 % pour le texte 3). Les résultats obtenus pour le texte 1 (article de presse) corroborent les observations de Carl et al. (2011), ainsi que celles de Daems (2016) selon lesquelles la PE de TA serait plus rapide que la traduction humaine dans le cas de textes généralistes. En outre, pour le texte 1 et le texte 3, la PE de TA neuronale a même permis d'augmenter le taux de productivité par rapport à celui obtenu en PE de TA statistique (+45 % et +4 %). Nous pouvons affirmer que la PE (TA statistique ou TA neuronale) a permis aux étudiantes d'augmenter leur productivité par rapport à la TH, c'est-à-dire de traduire plus de mots en moins de temps. À titre de comparaison, les travaux de Depraetere et al. (2014)⁹⁷ font état d'une augmentation moyenne de la productivité entre la PE de TA statistique et la TH de

⁹⁵ TAS + PE : post-édition de traduction automatique statistique

⁹⁶ TAN + PE : post-édition de traduction automatique neuronale

⁹⁷ « The productivity increased on average by 21.5 % (post-edited translation vs. translation from scratch). It may be concluded that MT enhances the translator's productivity, even if they are in the initial stages of their careers » (Depraetere et al., 2014, p. 99)

21,5 %.

4.4.3.1.2 Productivité par étudiante

Nous avons voulu également observer les effets du mode de traduction sur le taux de productivité de chaque étudiante. Les résultats sont repris dans le Tableau 26. Nous avons mis en évidence (en gras souligné) les valeurs maximale et minimale pour chaque mode de

	TH (s/mot)	TAS+PE (s/mot)	TAN+PE (s/mot)	TAS + PE vs TH	TAN + PE vs TH	TAN+PE vs TAS+PE
ET001	7,39	6,02	3,03	+18 %	+59 %	+50 %
ET003	9,63	<u>14,52</u>	4,41	-51 %	+54 %	+70 %
ET006	7,72	5,38	4,31	+30 %	+44 %	+20 %
ET008	5,70	4,83	6,04	+15 %	-6 %	-25 %
ET009	18,14	13,68	4,46	+25 %	+75 %	+67 %
ET010	5,76	6,36	4,86	-11 %	+15 %	+24 %
ET011	6,83	2,80	2,96	+59 %	+57 %	-6 %
ET013	5,20	5,24	6,84	-1 %	-32 %	-31 %
ET015	9,09	6,78	5,80	+25 %	+36 %	+14 %
ET018	4,73	5,90	3,80	-25 %	+20 %	+36 %
ET019	5,74	5,54	2,57	+3 %	+55 %	+54 %
ET020	7,33	<u>2,24</u>	2,33	+69 %	+68 %	-4 %
ET021	9,60	7,99	9,97	+17 %	-4 %	-25 %
ET022	<u>4,16</u>	5,85	2,22	-41 %	+47 %	+62 %
ET023	10,09	7,86	9,26	+22 %	+8 %	-18 %
ET024	5,50	5,14	4,14	+7 %	+25 %	+19 %
ET025	6,27	3,47	3,45	+45 %	+45 %	+1 %
ET026	6,92	4,74	3,79	+32 %	+45 %	+20 %
ET028	6,47	3,54	<u>1,55</u>	+45 %	+76 %	+56 %
ET029	6,94	4,92	6,93	+29 %	0 %	-41 %
ET030	11,56	7,30	4,95	+37 %	+57 %	+32 %
ET031	<u>18,13</u>	9,03	<u>11,37</u>	+50 %	+37 %	-26 %
ET032	5,74	5,90	8,33	-3 %	-45 %	-41 %
ET033	10,56	5,66	4,74	+46 %	+55 %	+16 %
ET035	9,99	11,43	9,77	-14 %	+2 %	+15 %

traduction.

Tableau 26 : Gains de productivité par étudiante – Prétest

Pour 18 participantes sur 25 (soit 72 %), la PE de TA statistique a engendré un gain de productivité par rapport à la TH. Cette proportion est encore plus élevée pour la PE de TA neuronale : 20 étudiantes sur 25 (80 %) ont traduit plus rapidement en PE de TA neuronale

qu'en TH. Le gain de productivité moyen s'élève à 17 % en PE de TA statistique et à 32 % en PE de neuronale. Tout comme Shterionov et al. (2018) l'ont déjà montré dans leur étude comparative, nous pouvons tirer comme conclusion que la PE de TA neuronale est le mode de traduction qui entraîne, de manière générale, le meilleur taux de productivité, même si 9 étudiantes sur 25 (36 %) ont traduit plus rapidement en PE de TA statistique qu'en PE de TA neuronale.

4.4.3.1.3 Productivité et qualité

- ❖ Existe-t-il une corrélation entre la productivité d'une étudiante et la note globale attribuée à sa production ?

Nous avons également voulu vérifier l'existence d'une corrélation entre la productivité (mesurée en secondes par mot du texte source) et la note obtenue par l'étudiante pour sa production. Pour ce faire, nous avons généré une figure en nuage de points par mode de traduction, sur laquelle sont reprises les productions des trois TS.

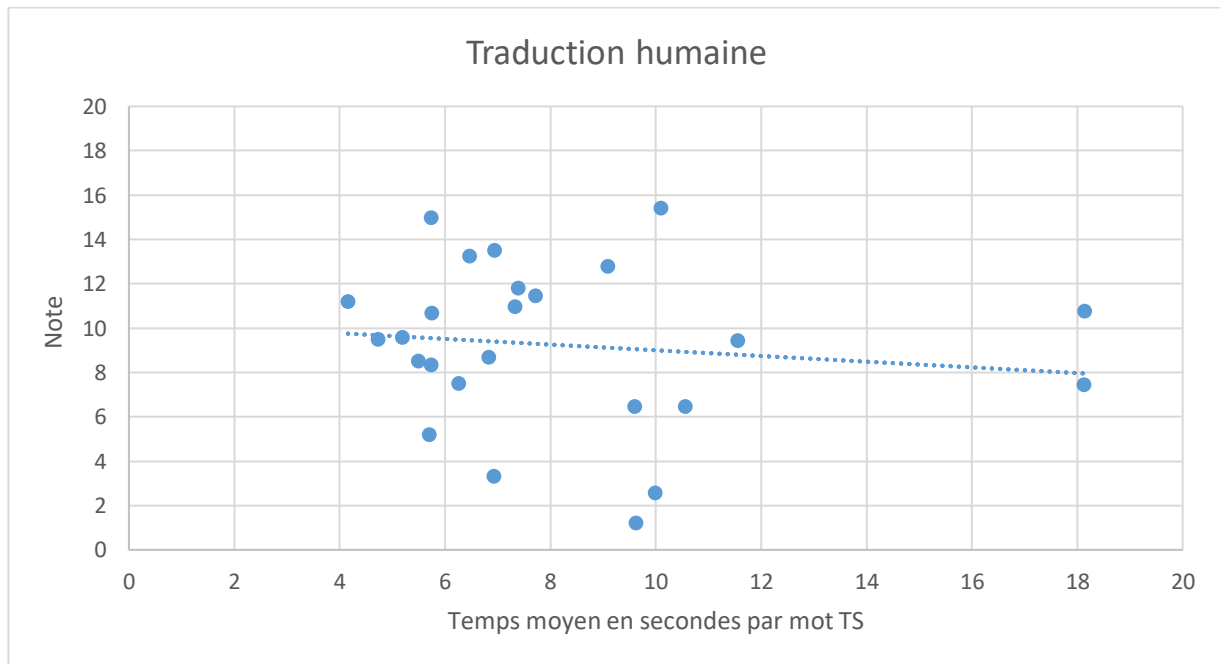


Figure 33 : Productivité-Qualité – Traduction humaine (Prétest)

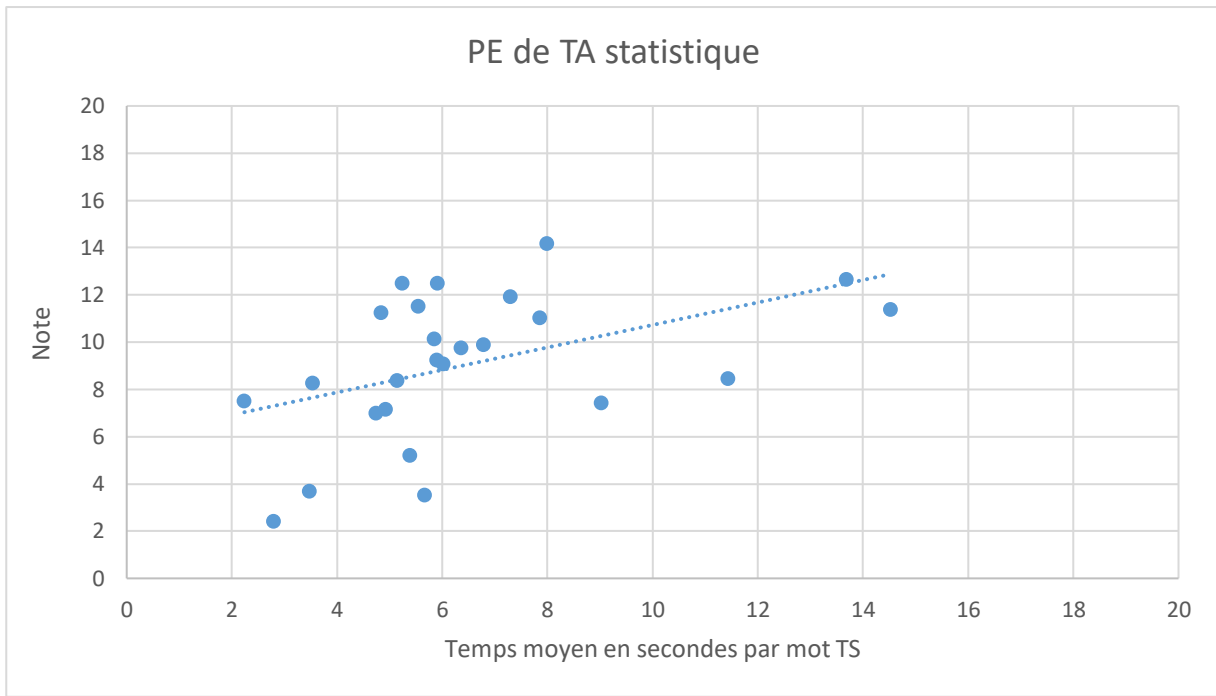


Figure 34 : Productivité-Qualité – PE de TAS

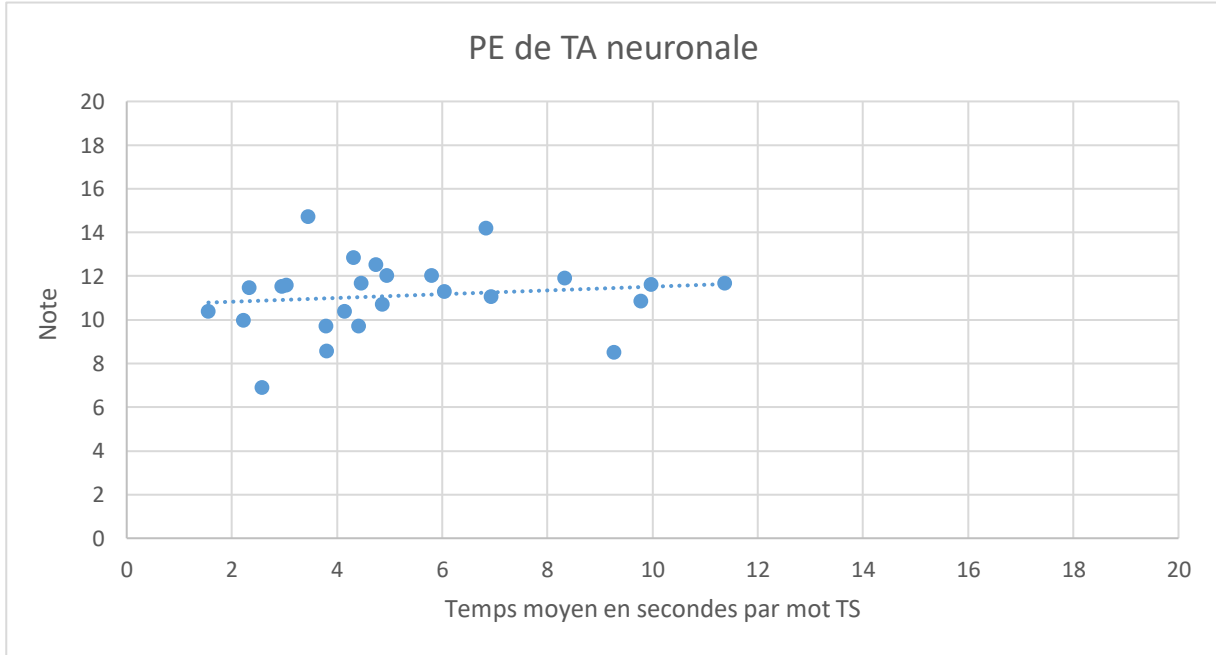


Figure 35 : Productivité-Qualité – PE de TAN (DeepL)

Afin de rejeter ou de confirmer l'hypothèse nulle qui suppose qu'il n'y a pas de corrélation entre le taux de productivité et la note globale obtenue par l'étudiante, nous avons calculé le coefficient de corrélation de Pearson⁹⁸ sur les recommandations du Prof. Sylvain Sardy, professeur associé de la Section de mathématiques de l'UNIGE, spécialisé en statistique et en optimisation. Pour interpréter les valeurs r ainsi obtenues, nous avons pris pour référence la classification suivante :

- **Exactly -1.** A perfect downhill (negative) linear relationship
- **-0.70.** A strong downhill (negative) linear relationship
- **-0.50.** A moderate downhill (negative) relationship
- **-0.30.** A weak downhill (negative) linear relationship
- **0.** No linear relationship
- **+0.30.** A weak uphill (positive) linear relationship
- **+0.50.** A moderate uphill (positive) relationship
- **+0.70.** A strong uphill (positive) linear relationship
- **Exactly +1.** A perfect uphill (positive) linear relationship

Figure 36 : Échelle d'interprétation r (Rumsey, 2003)

L'intervalle de confiance est fixé à 95 %, c'est-à-dire qu'il y a une probabilité de 95 % que l'intervalle de valeurs délimité par la borne inférieure et par la borne supérieure contienne la valeur de notre coefficient de corrélation interclasse.

Les résultats du test de Pearson ne révèlent pas de corrélation en TH, ni en PE de TA neuronale, confirmant ainsi l'hypothèse nulle (respectivement $r = -0,13$ [IC 95 % : $-0,496 - 0,284$] et $r = 0,14$ [IC 95 % : $-0,274 - 0,504$]). Ces résultats montrent toutefois qu'il existe une faible corrélation positive entre le temps moyen de PE et la note obtenue en PE de TA statistique ($r = 0,47$ [IC 95 % : $0,087 - 0,727$]). Cela signifie que plus l'étudiante a passé du temps à post-éditer la TA statistique, meilleure est la note qu'elle a obtenue.

4.4.3.2 Étude du processus : Effort de PE

L'outil *MateCat* nous a permis de disposer pour chaque production du taux moyen d'effort de PE. Cet effort de post-édition correspond au pourcentage moyen de modifications que

⁹⁸ « Le coefficient de Pearson est un indice reflétant une relation linéaire entre deux variables continues. Le coefficient de corrélation varie entre -1 et +1, 0 reflétant une relation nulle entre les deux variables, une valeur négative (corrélation négative) signifiant que lorsqu'une des variable [sic] augmente, l'autre diminue ; tandis qu'une valeur positive (corrélation positive) indique que les deux variables varient ensemble dans le même sens » (Farnir et al., 2021, s.p.).

l'étudiante a apportées à la TA brute (Federico et al., 2014b)⁹⁹. Zaretskaya et al. (2016) précisent la manière dont cet effort est généralement calculé :

[P]ost-editing effort (PEE) for each segment, which a numerical indicator of the amount of editing performed within the segment and is based on fuzzy match algorithms used in CAT tools, as well as the TER metric (Snover et al., 2006). In general, it is calculated as a number of changed words in the sentence divided by the total number of words (p. 76)

4.4.3.2.1 Effort de PE et système de TA

Le tableau 27 reprend l'ensemble des efforts de PE (%) en fonction du système de TA (TA statistique ou TA neuronale) et du texte source.

Texte 1		Texte 2		Texte 3	
PE de TAS	PE de TAN	PE de TAS	PE de TAN	PE de TAS	PE de TAN
28 %	9 %	19 %	13 %	16 %	8 %
35 %	41 %	28 %	6 %	17 %	23 %
27 %	14 %	7 %	10 %	19 %	14 %
39 %	19 %	23 %	11 %	20 %	6 %
13 %	15 %	21 %	4 %	25 %	16 %
26 %	16 %	12 %	3 %	23 %	6 %
37 %	16 %	16 %	24 %	20 %	6 %
30 %	7 %	8 %	6 %	32 %	20 %
23 %					9 %
					21 %
					9 %

Tableau 27 : Effort moyen de PE par production – Prétest

⁹⁹ « Average post-editing effort as a percentage. This indicates the average percentage of changes made to suggestions from the TM or MT system across the whole job » (Federico et al., 2014b, p. 49).

<i>Texte source</i>	Système de TA	Min.	Moyenne	Médiane	Max.	Écart-type
<i>Texte 1</i>	PE de TAS	13 %	29 %	28 %	39 %	0,077
	PE de TAN	7 %	17 %	15 %	41 %	0,103
<i>Texte 2</i>	PE de TAS	7 %	17 %	18 %	28 %	0,074
	PE de TAN	3 %	10 %	8 %	24 %	0,068
<i>Texte 3</i>	PE de TAS	16 %	21 %	20 %	32 %	0,051
	PE de TAN	6 %	13 %	9 %	23 %	0,065

Tableau 28 : Détails effort moyen de PE – Prétest

En nous servant des données du Tableau 27, nous avons calculé la valeur minimale, la moyenne, la médiane, la valeur maximale et l'écart-type par système de TA et par texte source (Tableau 28). Sur le Tableau 28, nous constatons tout d'abord systématiquement une baisse notable de la valeur moyenne et de la valeur médiane en PE de TA neuronale, comparativement aux valeurs obtenues en PE de TA statistique. Pour chaque texte source, les valeurs des deux écarts-types mesurés sont proches, voire très proches, ce qui est plutôt intéressant, car cela signifie que la dispersion des efforts est sensiblement la même pour les deux modes de traduction observés. Il apparaît que le texte qui a nécessité le plus d'effort de PE est l'article de presse (texte 1), que ce soit en PE de TA statistique ou en PE de TA neuronale.

Ces observations nous amènent à conclure que, dans cette expérience, l'effort moyen en PE de TA neuronale fut moindre que l'effort moyen en PE de TA statistique, et ce, pour les trois textes sources. Or, selon Koglin (2015), l'effort de PE serait directement lié à la qualité de la TA brute : « the poorer the machine output quality, the more post-editing effort will be needed — provided that the text is expected to be fully post-edited » (p. 126). En présumant la validité de cette corrélation, nous pourrions légitimement affirmer que, dans notre expérience, la qualité des TA neuronales brutes est meilleure que la qualité des TA statistiques brutes, quel que soit le texte source.

4.4.3.2.2 Effort de PE et qualité

Nous avons cherché à savoir s'il y avait une quelconque corrélation entre cet effort moyen de PE et la note globale obtenue par l'étudiante (Figure 37 et Figure 38). Nous supposons que, plus ce pourcentage d'effort de PE est élevé, meilleure est la qualité de la production et, par conséquent, plus la note obtenue est élevée.

- ❖ Existe-t-il une corrélation entre l'effort moyen de PE et la note globale obtenue par l'étudiante ?

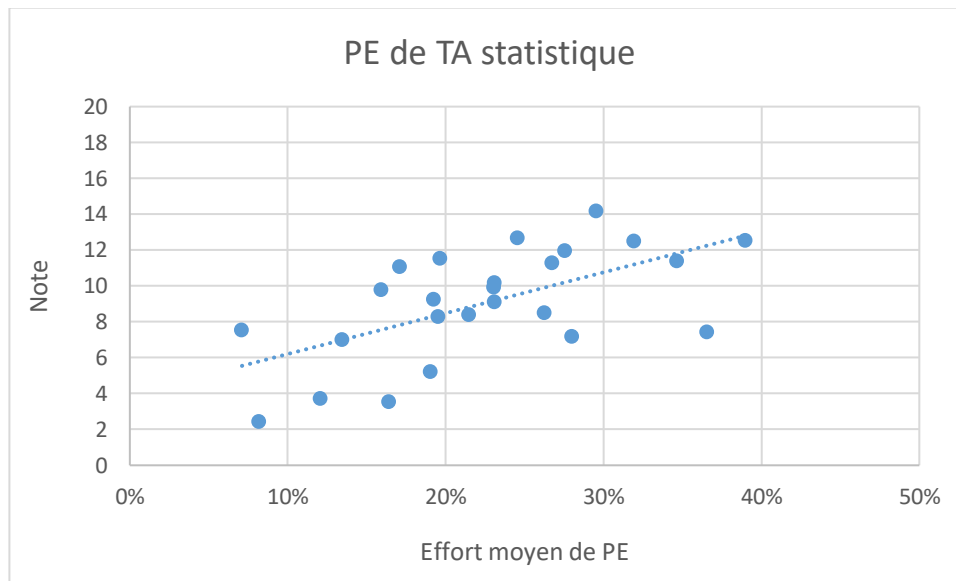


Figure 37 : Effort de PE et qualité – PE de TA statistique – Prétest

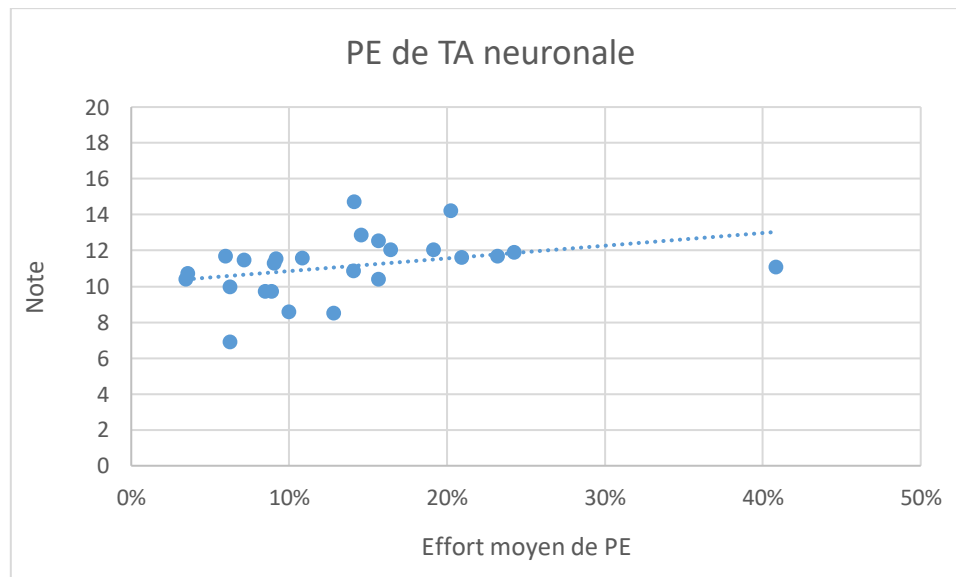


Figure 38 : Effort de PE et qualité – PE de TA neuronale – Prétest

Pour vérifier l'existence d'une corrélation, nous avons à nouveau calculé le coefficient de corrélation de Pearson. Nous pouvons affirmer qu'il existe une corrélation positive modérée ($r = 0,61$ [IC 95 % : 0,285 – 0,810]) entre l'effort moyen de PE en TA statistique et la note obtenue. C'est-à-dire que plus le taux d'effort moyen est élevé, plus la qualité de la production est

élevée, quel que soit le texte source. Les résultats révèlent également une faible relation linéaire positive entre l'effort de PE en TA neuronale et la qualité des productions ($r = 0,34$ [IC 95 % : -0,061 – 0,650]).

De manière similaire aux conclusions concernant le rapport entre le temps moyen en PE statistique et la qualité des productions, nous constatons la présence d'une corrélation positive entre l'effort moyen de PE et la note obtenue. Plus l'étudiante apporte des modifications à la TA brute (statistique et neuronale), meilleure est la note attribuée à sa production finale. De plus, il est apparu que les modifications apportées en TA statistique ont davantage permis de gagner en qualité par rapport aux modifications apportées en TA neuronale. Ces observations rejoignent les conclusions de l'étude menée par Mutal et al. (2019) dont les résultats supposent, d'une part, que les erreurs en TA statistique sont plus aisément repérables qu'en TA neuronale et, d'autre part, que les modifications effectuées en PE de TA neuronale sont, en général, davantage jugées non pertinentes par rapport à celles effectuées en PE de TA statistique :

If we consider all edits together, the evaluators would have post-edited the SMT output more than the NMT output: 68.53 % of the edits would have been modified by a majority of judges in SMT versus 14.69 % in NMT. This confirms our hypothesis that the edits in NMT are more often considered to be non-significant in the post-editing task. (p. 5)

Soulignons que, dans notre cas, la majorité des participantes étaient des post-éditrices novices. Il se peut que les modifications qu'elles ont apportées à la TA brute ne soient pas toujours pertinentes (voire nécessaires).

Notons qu'il serait pertinent d'approfondir toutes ces observations du processus à l'aide d'un modèle de régression mixte.

4.5 Synthèses et réponses aux hypothèses

Dans cette première étude comparative, notre objectif était d'explorer les potentiels effets de la PE de TA statistique et de TA neuronale en contexte d'apprentissage sur la qualité du texte cible pour la paire de langues anglais-français. Nous tenons à souligner que les résultats présentés dans ce chapitre se rapportent uniquement à notre première recherche empirique. Nous sommes bien consciente que la validité de ceux-ci se limite au cadre spécifique de cette expérience menée avec des apprenantes en traduction et pour la combinaison de langues anglais-français. Nous estimons qu'il serait prématuré, à ce stade, d'en tirer de quelconques conclusions plus générales. C'est pourquoi nous avons prévu, dès le départ, de mener une deuxième expérience afin de pouvoir confirmer ou infirmer nos observations.

4.5.1 Statistiques descriptives

Dans ce prétest, il apparaît que les textes post-édités ne contiennent pas, en moyenne, davantage d'erreurs que les TH et n'ont pas obtenu de notes inférieures à celles attribuées aux TH ; c'est même le contraire dans le cas de la PE de TA neuronale. Ces résultats corroborent les conclusions des études de Daems, De Clercq et al. (2017)¹⁰⁰, Daems, Vandepitte et al. (2017), García (2011)¹⁰¹, Martikainen et Mestivier (2020).

Tout d'abord, l'analyse comparative des productions révèle que le taux de réussite atteint en PE de TA neuronale est plus élevé, voire nettement plus élevé, par rapport au taux de réussite atteint en TH, et ce, contrairement à ce que nous avons pressenti (Figure 25).

Ensuite, il est également intéressant de noter que nous avons observé une moyenne générale plus élevée, ainsi qu'un meilleur taux de réussite en PE de TA neuronale qu'en PE de TA statistique (Figure 25 et Figure 26), et ce, pour les trois textes sources. Ces résultats confirment notamment ceux obtenus dans l'étude comparative entre la PE de TA statistique et la PE de TA neuronale menée par Yamada (2019) auprès d'étudiantes universitaires.

Dès lors, nous pouvons soutenir que, dans cette étude, la probabilité qu'une étudiante obtienne une note de réussite (supérieure ou égale à 10/20) est plus élevée en PE de TA neuronale qu'elle ne l'est en TH ou en PE de TA statistique.

Toutefois, nos résultats tendent à montrer l'existence d'un effet nivelant en PE de TA neuronale chez les étudiantes. Cet effet s'est traduit à la fois par un resserrement des notes autour de la moyenne, particulièrement en PE de TA neuronale par rapport à la TH (Figure 27 et Figure 28) ; et par l'établissement d'une relation inverse entre le niveau d'une étudiante en TH et la qualité de sa PE de TA neuronale (Figure 30, Figure 31 et Figure 32). Cette relation peut s'interpréter comme suit : plus l'étudiante est faible en TH, plus la PE de TA va lui permettre d'augmenter la qualité de sa production finale et à l'inverse, plus l'étudiante est compétente, plus la PE de TA neuronale va entraîner une baisse de qualité de sa production. Les étudiantes faibles tireraient donc davantage profit de la PE de TA neuronale que les autres étudiantes. A contrario, les *bonnes* étudiantes, outre le fait de ne tirer aucun avantage de la PE de TA neuronale, seraient desservies par ce mode de traduction qui ne leur permet pas de produire des textes cibles d'aussi bonne qualité qu'en TH.

¹⁰⁰ « [F]ully PE texts are indistinguishable from HT texts with regard to quality, reader perception and traditional, lexical, syntactic and semantic features » (Daems, De Clercq et al., 2017, p. 101).

¹⁰¹ « With regard to quality [...] post-editing produces significantly better statistical results compared to translating manually. These gains in quality are observed independently of language direction, text difficulty or translator's level of performance » (García, 2011, p. 2017).

4.5.2 Statistiques inférentielles

4.5.2.1 Hypothèse HA1

Les statistiques inférentielles ont montré que le mode de traduction a eu une influence significative sur la qualité des productions à deux niveaux : au niveau de la distribution des erreurs d'acceptabilité (erreurs de langue, problèmes de fluidité et de cohérence en langue cible) et au niveau des notes attribuées aux productions. En ce qui concerne la distribution des erreurs d'acceptabilité, ainsi que les notes attribuées aux productions, d'une part, les résultats des différents tests de Tukey ne montrent pas d'écarts significatifs entre les PE de TA statistique et les TH, laissant supposer que ces productions sont de qualité comparable. D'autre part, ces tests révèlent des différences significatives entre les PE de TA neuronale et les TH, ainsi qu'entre les PE de TA neuronale et les PE de TA statistique (voir Tableau 22, Tableau 23 et Tableau 24).

Ainsi, nous pouvons en tirer deux constats :

En PE de TA neuronale,

- A) les étudiantes ont commis significativement moins d'erreurs d'acceptabilité qu'en TH et qu'en PE de TA statistique ;
- B) les étudiantes ont obtenu, en moyenne, de meilleures notes qu'en TH et qu'en PE de TA statistique, et ce, indépendamment du TS.

Ces constats viennent ainsi infirmer notre hypothèse de départ (HA1) à savoir que la PE influe négativement sur la qualité du produit fini et ne permet pas à des étudiantes de produire des textes cibles de qualité comparable à une TH. Pour rappel, nous nous attendions à ce que les textes post-édités contiennent, en moyenne, un plus grand nombre d'erreurs que les TH et qu'ils obtiennent une moins bonne note globale sur 20 par rapport aux TH, ce qui ne fut pas le cas de toute évidence. Nous pouvons même affirmer que la qualité globale des textes post-édités par les étudiantes s'est révélée équivalente à la qualité des TH, voire supérieure dans le cas de la PE de TA neuronale.

4.5.2.2 Hypothèse HA2

Notre hypothèse HA2, qui suppose que la PE de TA ne présente pas une distribution d'erreurs comparable à la distribution d'erreurs observée en traduction humaine, se voit vérifiée. Effectivement, lorsque l'on analyse les catégories d'erreurs isolément, nos résultats montrent que le mode de traduction a eu des effets significatifs sur les catégories suivantes : calque ; glissement de sens ; orthographe ; grammaire et syntaxe ; cohérence.

En synthèse, il ressort de l'analyse statistique que :

- A) la PE (TA statistique et TA neuronale) a généré significativement
 - plus de calques fautifs par rapport à la TH ;
 - moins de fautes d'orthographe par rapport à la TH ;

- B) la PE de TA neuronale a généré significativement
 - moins de glissements de sens par rapport à la TH ;
 - moins d'erreurs de grammaire et de syntaxe par rapport à la PE de TA statistique et à la TH ;

- C) la PE de TA statistique a généré significativement
 - davantage d'erreurs de cohérence par rapport à la PE de TA neuronale et à la TH.

Concernant cette plus grande fréquence d'erreurs de cohérence en PE de TA statistique, il nous est difficile de tirer une conclusion qui tienne la route tant cette catégorie regroupe plusieurs sous-catégories d'erreurs. Pour rappel, il s'agit d'erreurs de type : logique interne, irrégularité dans l'emploi de termes, concordance des temps, etc.

4.5.2.2.1 Sous-hypothèse SHA2.1 – Erreurs d'adéquation

Les tests ANOVA ne nous permettent ni de confirmer, ni d'infirmer notre sous-hypothèse SHA2.1 selon laquelle nous supposons que nous allons observer davantage d'erreurs d'adéquation – erreurs de traduction sur le plan sémantique de type contresens, glissement de sens, etc. – en PE qu'en TH. Effectivement, les résultats étant non significatifs, nous n'avons pas pu inférer de dépendance entre le mode de traduction et la distribution globale des erreurs d'adéquation.

4.5.2.2.2 Sous-hypothèse SHA2.3 – Fluidité

En outre, il ne nous a pas été possible de confirmer la sous-hypothèse SHA2.3 au moyen des statistiques inférentielles. En effet, nous avons obtenu des résultats non significatifs pour la catégorie « style et registre » qui reprend entre autres les problèmes de fluidité (mauvaise cooccurrence, non-idiomaticité, collocation malheureuse, lourdeur, déformation de proverbes, etc.). Contrairement à ce que nous avons postulé, il semble que la PE n'a pas eu d'influence négative sur les erreurs de la catégorie « style et registre ».

4.5.2.2.3 Sous-hypothèse SHA2.2 – Calques fautifs

Les résultats obtenus viennent cependant valider notre sous-hypothèse SHA2.2, car nous avons pu démontrer que les textes post-édités par les étudiantes à partir de la TA statistique

et de la TA neuronale contenaient significativement plus de calques fautifs que les textes traduits humainement.

4.5.2.2.4 Sous-hypothèse SHA2.4 – Fautes d’orthographe, de grammaire et de syntaxe

Notre sous-hypothèse SHA2.4 se trouve également entièrement validée : les évaluatrices ont recensé, en moyenne, moins de fautes d’orthographe dans les textes post-édités (TA statistique et TA neuronale) par les étudiantes que dans les productions humaines ; il semble également que la PE de TA neuronale ait généré moins d’erreurs de grammaire et de syntaxe par rapport aux deux autres modes de traduction.

4.5.2.2.5 Sous-hypothèse SHA2.5 – Texte source

En nous appuyant sur le tableau récapitulatif (Tableau 23), nous pouvons dire que la variable « texte source » a effectivement eu une influence significative sur la plupart des catégories d’erreurs, ce qui appuie notre sous-hypothèse SHA2.5. Nous avons constaté que le texte source a eu des effets significatifs sur la distribution des erreurs d’acceptabilité, des erreurs d’adéquation, ainsi que sur la distribution des erreurs appartenant à toutes les catégories à l’exception de la catégorie « Ajout/Omission » et « Sur-/sous-traduction ». Précisons qu’il est toutefois relativement difficile de pouvoir attribuer ces écarts significatifs au seul facteur « thématique du TS ». En effet, d’autres éléments peuvent également expliquer ces divergences au niveau des distributions d’erreurs tels que la sévérité et la sensibilité propres à chaque binôme d’évaluatrices.

4.5.2.3 Hypothèse HA3

Ces résultats nous amènent à valider également notre troisième hypothèse (HA3), à savoir que la qualité de textes post-édités dépend de la technologie de TA employée (TA statistique ou TA neuronale). Nous nous attendions à obtenir des productions de meilleure qualité en TA neuronale. Et effectivement, nous avons constaté, dans cette expérience, que comparativement à la PE de TA statistique, les PE de TA neuronale ont obtenu de meilleures notes et contiennent significativement :

- moins d’erreurs d’acceptabilité;
- moins de problèmes de cohérence ;
- moins de fautes de grammaire et de syntaxe.

4.5.3 Étude du processus

Nous avons constaté un gain de productivité en PE par rapport à la TH. Nous pouvons effectivement affirmer que la PE (que ce soit de TA statistique ou de TA neuronale) a permis

aux étudiantes d'augmenter leur taux de productivité par rapport à celui observé en TH, autrement dit, de traduire plus de mots en moins de temps ; le meilleur taux de productivité ayant été observé en PE de TA neuronale. Ces résultats confirment les tendances qui avaient déjà été observées dans de nombreuses recherches en la matière, citons les travaux de Daems (2016), de Daems, Vandepitte et al. (2017), de Depraetere et al. (2014).

Nous n'avons pas observé de corrélation entre la durée de traduction ou la durée de PE de TA neuronale et la qualité de la production finale. En revanche, il apparaît que plus l'étudiante a passé du temps à post-éditer la TA statistique, meilleure est la note qu'elle a obtenue (Figure 34), même si cette corrélation positive est faible.

Ensuite, nous avons conclu à l'existence d'une faible corrélation entre le taux d'effort de PE et la qualité des textes post-édités (Figure 37 et Figure 38). En d'autres termes, cette corrélation signifie que, plus l'étudiante a apporté des modifications à la TA brute, meilleure est la note attribuée à sa production finale. De plus, nos résultats attestent une diminution notable de l'effort de PE en TA neuronale par rapport à la PE de TA statistique, ce qui appuie l'hypothèse avancée par Mutal et al. (2019) concernant le caractère davantage facultatif des modifications apportées par la traductrice en PE de TA neuronale par rapport aux changements effectués en PE de TA statistique. Il nous paraît sensé de supposer que cette diminution de l'effort de PE découle principalement des nombreuses améliorations qualitatives entre l'ancienne et la nouvelle génération de TA déjà relevées dans divers travaux (Castilho et al., 2017 ; Koehn, 2020 ; O'Brien et Ehrensberger-Dow, 2020 ; Poibeau, 2019 ; Wu et al., 2016). Outre une amélioration observée en matière de fluidité en langue cible (Loock, 2018 ; Macken et al., 2019), il a notamment été montré dans plusieurs études que les systèmes neuronaux étaient particulièrement performants pour générer des paraphrases et pour employer des synonymes de manière correcte (Mallinson et al., 2017 ; Mutal et al., 2019).

4.6 Conclusion

Les résultats de cette première recherche expérimentale ont infirmé notre hypothèse de départ (HA1) qui était que la PE effectuée par des étudiantes peut avoir une influence négative sur la qualité d'un texte cible. Nous pouvons affirmer que, dans ce prétest, la PE de TA (statistique comme neuronale) n'a pas généré de productions de moindre qualité qu'en TH, ce qui corrobore les résultats obtenus dans de nombreuses études (Daems, 2016 ; Daems, Vandepitte et al., 2017 ; Depraetere et al., 2014 ; García, 2011 ; Jia et al., 2019).

En outre, la PE de TA neuronale a permis d'aboutir à des productions de qualité globale supérieure par rapport aux deux autres modes de traduction ; les étudiantes ont commis significativement moins d'erreurs d'acceptabilité lorsqu'elles ont post-édité la TA neuronale par rapport à la PE de TA statistique et TH. De plus, les PE de TA neuronale ont obtenu de

meilleures notes, en moyenne, par rapport aux deux autres modes de traduction. Toutefois, il n'a pas été possible de déduire une corrélation entre le mode de traduction et la distribution globale des erreurs d'adéquation.

Ce constat vient confirmer les résultats obtenus notamment par Martikainen (2019c)¹⁰² ainsi que les résultats de l'étude comparative menée par Martikainen et Mestivier (2020) avec des étudiantes de 1^{er} et 2^e master dans laquelle il est fait état d'une réduction des erreurs d'environ 50 %, en moyenne, en PE de TA neuronale (*DeepL*) par rapport à la TH. Selon nous, ces constatations jouent en faveur de l'intégration de la PE, et particulièrement de la PE de TA neuronale, dans la formation initiale des futures traductrices dès lors que ce mode de traduction permettrait manifestement d'augmenter la qualité de leurs productions. Ce point fait l'objet d'une réflexion approfondie au chapitre 8.

Par ailleurs, l'analyse détaillée des catégories d'erreurs a abouti à des résultats plus nuancés. Il a été montré que les textes post-édités (TA statistique et TA neuronale) contiennent d'une part, plus de calques, d'autre part, moins de fautes d'orthographe que les TH. De plus, nous pouvons affirmer qu'en PE de TA neuronale, les étudiantes ont commis moins de glissements de sens par rapport à la TH et moins d'erreurs de grammaire et de syntaxe par rapport à la PE de TA statistique et à la TH. Enfin, il apparaît que les textes post-édités à partir de la TA statistique contiennent davantage de problèmes de cohérence par rapport à la PE de TA neuronale et à la TH.

Enfin, concernant le processus même de post-édition, nos résultats attestent un gain de productivité en PE par rapport à la TH. Toutefois, nous n'avons pas pu constater de corrélation entre la durée de traduction/post-édition et la qualité de la production de l'étudiante, sauf en PE de TA statistique (faible corrélation positive). De plus, nous avons pu observer une réduction notable de l'effort de PE en TA neuronale par rapport à la PE de TA statistique, ainsi qu'une faible corrélation positive entre le taux d'effort de PE et la qualité des textes post-édités, qui est davantage marquée en PE de TA statistique.

¹⁰² « [C]hez les apprenants novices en LSP [langues de spécialité] médicale travaillant sur un texte spécialisé, la post-édition de la TA neuronale produite par le moteur *DeepL* prend moins de temps et résulte en une qualité finale meilleure que la traduction humaine de ce même texte. En effet, les textes post-édités de l'échantillon contiennent nettement moins de distorsions que les textes traduits, mais également moins d'erreurs mineures et notamment de fautes d'orthographe, relativement fréquentes dans les textes traduits par les apprenants » (Martikainen, 2019c, p. 301).

5 Recherche expérimentale – Test

5.1 Introduction

Dans la continuité du prétest 2018, nous avons mené en 2021 une deuxième recherche expérimentale comparative de TH et de PE (anglais-français) en nous concentrant sur la PE de TA neuronale (*Google Traduction* et *DeepL*). La distinction principale par rapport au prétest réside dans le profil des participantes : alors que les participantes au prétest sont des étudiantes en 3^e année de bachelier en traduction-interprétation de l'Université de Liège, pour le test, il s'agit d'étudiantes de 2^e année de master. Par rapport au prétest, les étudiantes ayant pris part au test sont déjà *toutes* familières de la TA et ont déjà été amenées à post-éditer dans le cadre de leurs études sans pour autant avoir reçu de formation spécifique en la matière (cf. chapitre 3, section 3.1.3). Précisons qu'aucune élève n'a participé à la fois au prétest et au test.

La structure de ce chapitre est semblable à celle du chapitre 4. Après avoir formulé nos hypothèses spécifiques à cette deuxième expérience, nous commentons les améliorations méthodologiques apportées par rapport au prétest. Ensuite, nous présentons les résultats en trois parties qui correspondent à chaque étape de l'exploitation des données récoltées : statistiques descriptives, statistiques inférentielles et étude du processus. Puis nous synthétisons ces résultats avant de clore ce chapitre en tentant de dégager des hypothèses explicatives qui valent à la fois pour le prétest et pour le test.

5.2 Hypothèses de recherche

En nous appuyant sur les études récentes incluant la TA neuronale, ainsi que sur les résultats obtenus dans le prétest, nous avons reformulé certaines de nos hypothèses.

La question de recherche de départ est identique à celle du prétest :

Question A – Quels sont les effets de la post-édition (anglais-français) en contexte d'apprentissage sur la qualité d'un texte cible ?

5.2.1 Hypothèse HA4

Sous-question A4. La post-édition de traduction automatique neuronale (anglais-français) en contexte d'apprentissage permet-elle d'obtenir des textes de qualité comparable à la qualité en traduction humaine ?

Contrairement à ce que nous avons formulé pour le prétest en 2018, nous avançons l'hypothèse suivante : la PE de TA neuronale permet à des étudiantes de produire des textes cibles de qualité comparable à une TH, voire de meilleure qualité. Nous nous appuyons, cette fois, sur les travaux de Jia et al. (2019), de Martikainen et Mestivier (2020) et de Yamada (2019), ainsi que sur les résultats du prétest. Par conséquent, nous nous attendons à ce que les textes post-édités contiennent, de manière générale, soit le même nombre d'erreurs, soit moins d'erreurs que les TH et qu'ils obtiennent, en moyenne, des notes comparables aux TH, voire meilleures.

- Hypothèse nulle (HA4.0.)

La post-édition de TA neuronale (anglais-français) en contexte d'apprentissage ne permet pas d'obtenir des textes de qualité comparable à la qualité en traduction humaine.

- Hypothèse alternative (HA4.1.)

La post-édition de TA neuronale (anglais-français) en contexte d'apprentissage permet d'obtenir des textes de qualité comparable à la qualité en traduction humaine.

5.2.2 Hypothèse HA5

Sous-question A5. La post-édition de traduction automatique neuronale (anglais-français) en contexte d'apprentissage présente-t-elle une distribution d'erreurs comparable à la distribution des erreurs en traduction humaine ?

- Hypothèse nulle (HA5.0.)

La PE de TA neuronale présente une distribution d'erreurs comparable à la distribution des erreurs en TH.

- Hypothèse alternative (HA5.1.)

La PE de TA neuronale ne présente pas une distribution d'erreurs comparable à la distribution des erreurs en TH.

En ce qui concerne la qualité à ce niveau, nous nous attendons à ce que la distribution des erreurs en PE de TA neuronale ne soit pas comparable à la distribution des erreurs en TH. Pour répondre à cette sous-question A5, nous formulons également les sous-hypothèses suivantes :

5.2.2.1 Sous-hypothèse SHA5.1.

En ce qui concerne la qualité à ce niveau, nous pensons qu'au vu des conclusions du prétest, nous postulons que nous allons relever moins d'erreurs d'acceptabilité (erreurs de langue) en PE de TA neuronale qu'en TH.

5.2.2.2 Sous-hypothèse SHA5.2.

Plus spécifiquement, en nous fondant toujours sur les résultats du prétest, nous présumons que nous allons relever moins d’erreurs de grammaire et de syntaxe, ainsi que moins de fautes d’orthographe en PE qu’en TH.

5.2.2.3 Sous-hypothèse SHA5.3.

En nous appuyant sur les études menées par Fiederer et O’Brien (2009), Loock (2018) et Martikainen et Mestivier (2020), ainsi que sur les résultats du prétest, nous continuons à croire que la PE va engendrer un plus grand nombre de marques de littéralité qu’en TH, et plus spécifiquement de calques fautifs.

5.2.2.4 Sous-hypothèse SHA5.4.

Nous supposons à nouveau que la thématique du texte source aura une certaine influence sur la distribution des erreurs observées en PE et en TH.

5.2.3 Hypothèse HA6

Sous-question A6. La qualité de textes post-édités (anglais-français) en contexte d’apprentissage varie-t-elle en fonction du moteur de traduction automatique neuronale utilisé (*Google Traduction* ou *DeepL*) ?

- Hypothèse nulle (HA6.0.)

La qualité de textes post-édités ne dépend pas du moteur de traduction automatique neuronale utilisé.

- Hypothèse alternative (HA6.1.)

La qualité de textes post-édités dépend du moteur de traduction automatique neuronale utilisé.

Nous pensons que les PE du moteur *DeepL* donnera de meilleurs résultats que les PE du moteur *Google Traduction*. *DeepL* était considéré au moment où nous avons mené ce test (en 2021) comme le moteur de TA générique le plus performant (cf. section 3.1.4.1). De plus, *DeepL* serait particulièrement efficace pour la combinaison anglais-français (Poibeau, 2019, p. 121) par rapport aux moteurs concurrents comme *Google Traduction*, *Amazon* ou *Microsoft*, c’est ce qui ressort en tout cas de l’étude menée en janvier 2020 par la société (*DeepL GmbH*, 2020). Rappelons qu’aux fins de cette étude en contexte d’apprentissage, nous avons eu recours à la version disponible en ligne gratuitement et non à la version professionnelle *DeepL Pro*.

5.2.4 Hypothèse HA7

Sous-question A7. Existe-t-il un effet nivelant en post-édition de traduction automatique neuronale sur la qualité des textes cibles ?

- Hypothèse nulle (HA7.0.)

Il n'existe pas d'effet nivelant en PE de TA neuronale sur la qualité des textes cibles.

- Hypothèse alternative (HA7.1.)

Il existe un effet nivelant en PE de TA neuronale sur la qualité des textes cibles.

En nous fondant sur les travaux de García (2011), de Killman (2018) et de Lee et Liao (2011), ainsi que sur les résultats du prétest, nous nous attendons à constater une nouvelle fois un effet nivelant sur la qualité des textes cibles en PE de TA neuronale chez les apprenantes. Nous pensons que cet effet se révélera à travers l'homogénéisation des notes attribuées aux PE de TA neuronale par rapport aux notes attribuées aux productions humaines. La PE de TA neuronale permettrait ainsi aux étudiantes faibles de s'en sortir en leur évitant d'obtenir de (très) mauvaises notes, mais ce processus aurait également pour conséquence de brider les bons éléments.

5.2.5 Améliorations méthodologiques

Comme nous l'avons déjà évoqué, nous avons envisagé le prétest comme une étape méthodologique en vue de l'élaboration du test qui fait l'objet de ce chapitre. Ces améliorations méthodologiques portent aussi bien sur les conditions de réalisation des tâches et sur la composition du corpus que sur les critères d'évaluation de la qualité.

5.2.5.1 Tâches

Pour la TH, l'étudiante pouvait à nouveau disposer des ressources qu'elle jugeait utiles. Il lui était toutefois interdit d'utiliser tout moteur de TA et tout logiciel de TAO. Pour nous assurer de l'authenticité des productions humaines, cette expérience s'est déroulée en présentiel dans une salle informatique (Labo de Langues du bâtiment L3). Nous avons pris soin de bloquer l'accès aux différents outils de TA disponibles en ligne et, grâce au logiciel AVIDAnet®Live¹⁰³, nous avons pu surveiller l'écran de chaque étudiante tout au long de l'expérience (Figure 39 et Figure 40).

¹⁰³ AVIDAnet® LIVE est une suite logicielle de contrôle de classe permettant à un enseignant de donner des cours dans un environnement enrichi par la technologie (Televic Education, 2018). Ce logiciel est proposé par Televic Education qui est l'une des quatre divisions du groupe Televic.



Figure 39 : Capture d’écran – centre de contrôle Enseignant (Televic Education, 2018)

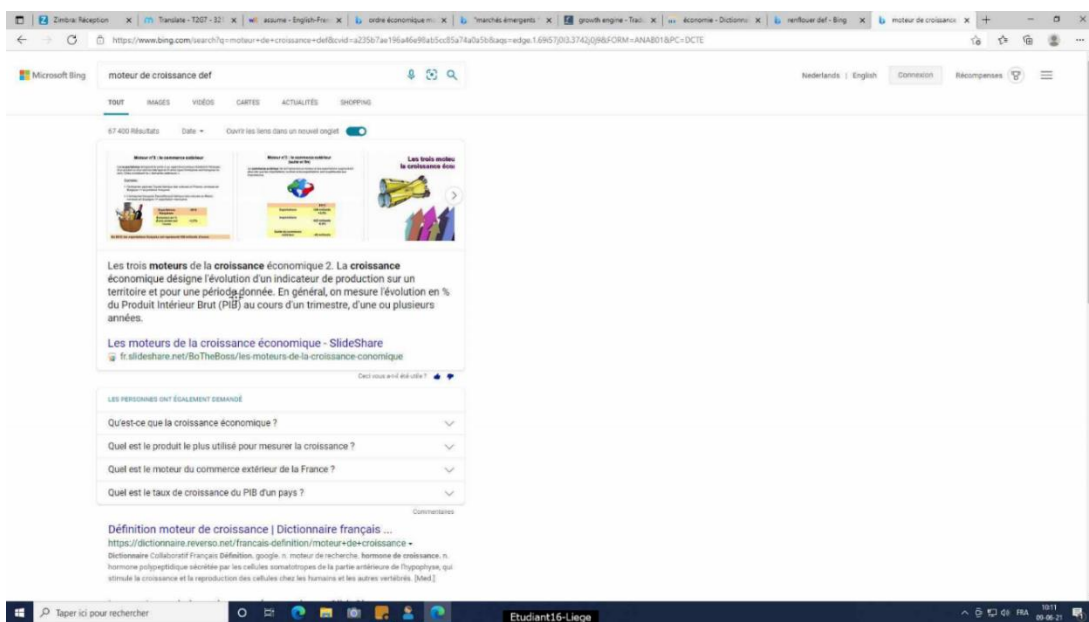


Figure 40 : Capture d’écran – recherche documentaire (Televic Education, 2018)

5.2.5.2 Corpus

Pour des raisons de comparabilité avec le prétest et de robustesse des résultats, nous avons réutilisé dans ce test les mêmes textes sources que dans le prétest. Nous avons toutefois décidé d’intégrer, pour chaque texte source, les deux TA brutes (*Google Traduction* et *DeepL*) au corpus d’évaluation humaine, et ce, à l’insu des évaluatrices. Ces sorties brutes ont donc été évaluées par les évaluatrices, au même titre que les productions des étudiantes, ce qui nous a permis de leur attribuer une note sur 20. Dans un souci de

pertinence et de validité des résultats, ces six productions (deux TA brutes par TS) ont été exclues du corpus dans l'étude quantitative : statistique descriptive, statistique inférentielle et étude du processus.

5.2.5.3 Comparaison des versions *DeepL*

Étant donné que nous avons décidé de réutiliser les mêmes textes sources qu'en 2018, nous avons comparé sommairement les différentes versions de TA brutes générées par le même moteur (*DeepL* janvier 2018 vs *DeepL* septembre 2020). Voici le compte-rendu de cette analyse comparative :

Nous avons observé majoritairement des divergences d'ordre lexical, syntaxique et typographique. Pour les trois textes sources, nous sommes d'avis que la majeure partie des modifications observées dans la version la plus récente sont bénéfiques par rapport à la version 2018. Autrement dit, nous pensons que ces modifications contribuent à l'amélioration de la qualité du TC. Ce constat reflète sans nul doute les progrès considérables réalisés dans le domaine de la TA neuronale. Toutefois, nous estimons que certaines modifications observées dans la version récente sont neutres ; elles n'améliorent pas fondamentalement la qualité, mais ne la diminuent pas non plus. Dans de très rares cas, nous avons même jugé que certaines modifications diminuent la qualité de la TA brute par rapport à la version 2018.

En comparant les deux versions de TA brutes, nous avons observé des modifications de type :

- amélioration de la fluidité ; expressions et tournures plus courantes et idiomatiques (1) ;
- choix lexicaux plus appropriés (2) ;
- réagencement syntaxique, amélioration syntaxique (3) ;
- meilleure application des règles typographiques (espace devant le deux-points, le point d'exclamation, etc.) (4) ;
- élimination de calques (5) ;
- élimination de glissements de sens (6) ;
- élimination d'omissions (7) ;
- élimination d'incohérences temporelles (8) ;
- élimination de cas de redondance (9) ;
- diminution des problèmes de référents et de biais de genre (10).

Le Tableau 29 nous permet d'illustrer chaque type d'amélioration.

Type	Texte source	DeepL janvier 2018	DeepL septembre 2020
1	T2 – « A number of other emerging markets will also take centre stage »	« Un certain nombre d'autres marchés émergents <u>occuperont également une place centrale</u> »	« Un certain nombre d'autres marchés émergents <u>occuperont également le devant de la scène</u> »
2	T3 – « This is called grapheme-color synesthesia ».	« C'est ce qu'on appelle <u>la synesthésie de couleur du graphème</u> »	« C'est ce qu'on appelle <u>la synesthésie graphème-couleur</u> »
3	T2 – « But this will require patience to ride out the storms we have seen recently in economies like, for example, Brazil, Nigeria and Turkey, all of which still have considerable long-term economic potential based on our analysis.»	« [...] dans des économies comme, par exemple, le Brésil, le Nigeria et la Turquie, <u>qui ont toutes encore un potentiel économique considérable à long terme sur la base de notre analyse</u> »	« [...] dans des économies comme, par exemple, le Brésil, le Nigeria et la Turquie, <u>qui, d'après notre analyse, ont encore un potentiel économique considérable à long terme</u> »
4	T1 – « Can Theresa May and her government survive? »	« Theresa May et son gouvernement <u>peuvent-ils survivre?</u> »	« Theresa May et son gouvernement <u>peuvent-ils survivre ?</u> »
5	T2 – « but emerging economies should make good progress towards closing this gap by 2050. »	« mais les économies émergentes devraient bien <u>progresser vers</u> la réduction de cet écart d'ici 2050 »	« mais les économies émergentes devraient bien <u>progresser pour</u> combler cet écart d'ici 2050 »
6	T1 – « It recognises that the referendum decision must be implemented, but also that to implement it with gung-ho gusto, of the kind that Boris Johnson could provide [...] » T1 – « Yet she remains in Downing Street, and the loss of two ministers who were found in different ways to have misbehaved does not change the powerful reasons for keeping her there »	« Il reconnaît que la décision référendaire doit être mise en œuvre, mais aussi que sa mise en œuvre <u>avec une hargne débridée</u> , comme celle que Boris Johnson pourrait fournir [...] » « Pourtant, elle demeure à Downing Street, et la perte de deux ministres <u>qui ont été retrouvés de différentes façons pour avoir mal agi</u> ne change rien aux puissantes <u>raisons de sa présence</u> »	« Il reconnaît que la décision du référendum doit être mise en œuvre, mais aussi que la mettre en œuvre <u>avec un enthousiasme débordant</u> , du type de celui que Boris Johnson pourrait apporter [...] » « Pourtant, elle reste à Downing Street, et la perte de deux ministres <u>qui se sont mal comportés de différentes manières</u> ne change rien aux puissantes <u>raisons qui l'ont poussée à rester là-bas</u> »

7	T3 – « Is a week shaped like a tipped-over “D” »	« Est-ce qu'une semaine <u>est en forme de D</u> [...] »	« Une semaine a-t-elle <u>la forme d'un "D" renversé</u> [...] »
8	T3 – « But synesthetes also report making computational errors because “6” and “8” have the same color »	« Mais les synesthetes rapportent aussi <u>avoir fait des</u> erreurs de calcul parce que 6 et 8 ont la même couleur »	« Mais les synesthésiens déclarent également <u>faire</u> des erreurs de calcul parce que le "6" et le "8" ont la même couleur »
9	T2 – « (including no sustained long-term retreat into protectionism) »	« (y compris l'absence d'un <u>repli durable à long terme</u> vers le protectionnisme) »	« (y compris l'absence de <u>repli durable</u> vers le protectionnisme) »
10	T1 – « Theresa May [...] four Westminster watchers discuss whether the Conservative leader »	« Theresa May [...] quatre observateurs de Westminster se demandent si <u>le</u> chef du Parti conservateur »	« Theresa May [...] quatre observateurs de Westminster se demandent si <u>la</u> chef des conservateurs »

Tableau 29 : Comparaison *DeepL* jan 2018 vs *DeepL* sept 2020

5.2.5.4 Exploitation du corpus

5.2.5.4.1 Approche qualitative : Évaluation humaine de la qualité

5.2.5.4.1.1 Évaluatrices humaines

Premièrement, nous avons tenu à corriger l'un des principaux biais du prétest ; à savoir la participation de l'autrice de cette recherche à l'évaluation de la qualité des productions. Pour ce test, l'équipe d'évaluatrices est à présent uniquement composée d'enseignantes/traductrices extérieures à l'étude. Chaque production a été évaluée par un binôme d'évaluatrices.

5.2.5.4.1.2 Recueil d'impressions

Deuxièmement, à la fin du processus d'évaluation à l'aveugle, nous avons demandé cette fois à chaque évaluatrice de juger selon son intuition si la production était une traduction humaine ou une post-édition, ainsi que de motiver sommairement son choix lorsque cela était possible. Nous discutons de ce recueil d'impressions à la section 5.3.3.3 du présent chapitre.

5.2.5.4.1.3 Typologie d'erreurs

Les productions récoltées dans cette deuxième expérience ont été évaluées et annotées selon la même méthode que lors du prétest. Toutefois, à la suite d'une séance de réflexion avec les étudiantes de la filière et grâce aux retours des différentes évaluatrices, nous avons décidé de modifier quelque peu nos deux grilles d'évaluation (acceptabilité et adéquation)

par rapport au prétest, avec pour objectif principal de rendre l'évaluation moins fastidieuse et moins ambiguë. Nous avons apporté quelques améliorations à la typologie d'erreurs en supprimant notamment l'une ou l'autre catégorie comme le recommandent Bayerl et Paul (2011)¹⁰⁴ afin de renforcer l'accord entre évaluateuses.

Nous avons listé ci-dessous les différents changements par rapport à la typologie utilisée dans le prétest.

Phase d'acceptabilité

- Distinction entre les fautes d'orthographe et les erreurs de typographie/punctuation, car ces erreurs n'ont pas le même degré de gravité ;
- Changement du nom de catégorie « style et registre » en « style » ;
- Isolation de l'étiquette « incompréhension », reprise avant sous « cohérence » pour des raisons de degré de gravité. Une « incompréhension » est une tournure/segment incompréhensible ou absurde qui entrave la compréhension du texte cible.

En raison du caractère fastidieux et chronophage de l'évaluation lors du prétest relevé par les différentes évaluateuses, nous avons décidé de simplifier la catégorisation en supprimant :

- les sous-catégories en grammaire : « accord/genre/autre gramm » ;
- les sous-catégories en orthographe : « majuscule/mots composés/faute orthographe » ;
- les sous-catégories en lexicale : inadéquation ; non traduit ;
- les sous-catégories en typographie : « coquille/abréviation/autre typo » ;
- les sous-catégories en style : type de texte ; autre style et registre.

¹⁰⁴ « To increase likelihood of inter-coder agreement, we suggest: - use few categories; - use annotators with the same level of domain expertise; - provide training, preferably of high intensity » (Bayerl et Paul, 2011, p. 719).

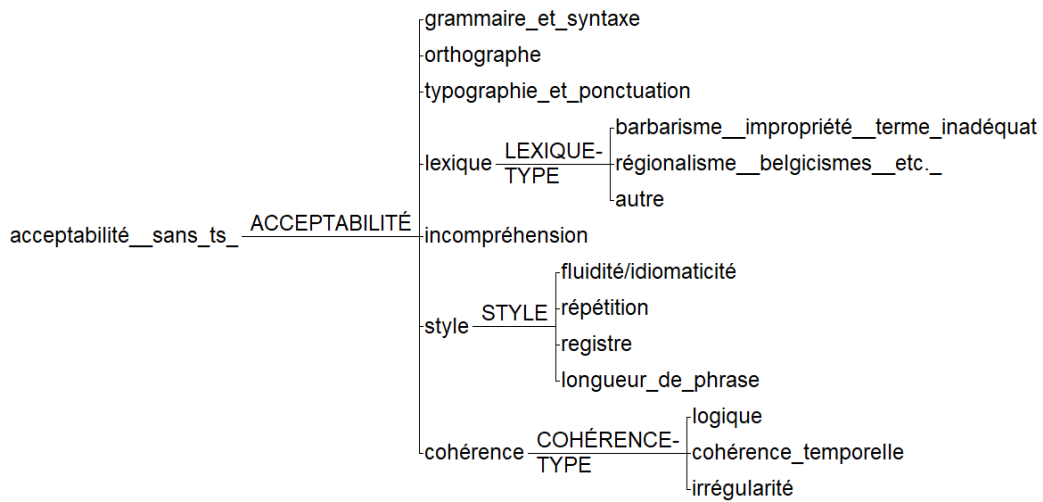


Figure 41 : Phase d'acceptabilité – Test

Phase d'adéquation

À nouveau, dans l'idée de simplifier la classification des erreurs, nous avons apporté les modifications suivantes :

- distinction entre « omission » et « ajout » ;
- ajout d'une catégorie « autre » ;
- suppression de la catégorie « sur-/sous-traduction » car sous-utilisée. On suppose que les erreurs de ce type ont été reprises sous « vocabulaire » ou « glissement de sens »
- suppression des sous-catégories en vocabulaire : « hyperonyme/hyponyme/inadéquat ».

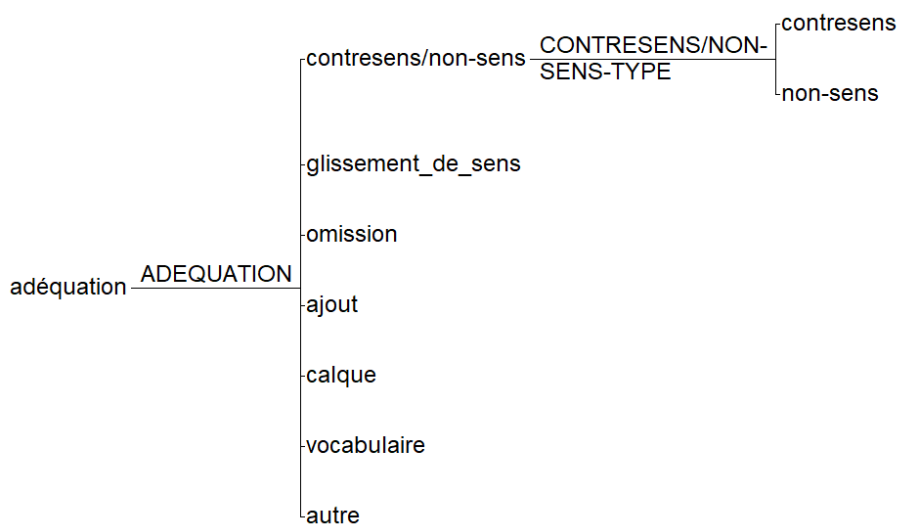


Figure 42 : Phase d'adéquation – Test

Pondération de la note globale

Pondération

Note sur 20

Acceptabilité (40%)	Catégorie	Grammaire/ Syntaxe	Orthographe	Typographie/Ponctuation	Lexique	Incompréhension	Style	Cohérence
	Pénalité	1	1	0,3	0,3	1,3	0,3	0,3
Adéquation (60%)	Catégorie	Contresens/Non-sens	Glissement de sens	Omission	Ajout	Calque	Vocabulaire	Autre
	Pénalité	1,3	0,7	0,7	0,7	0,5	0,5	0,5

Tableau 30 : Pondération notation sur 20 – Test

5.3 Résultats

5.3.1 Statistiques descriptives

5.3.1.1 Distribution des erreurs (diagramme circulaire)

Tout d'abord, la Figure 43 et la Figure 44 nous permettent de visualiser la distribution des erreurs par catégories dans l'ensemble de notre corpus, indépendamment du mode de traduction et du TS.

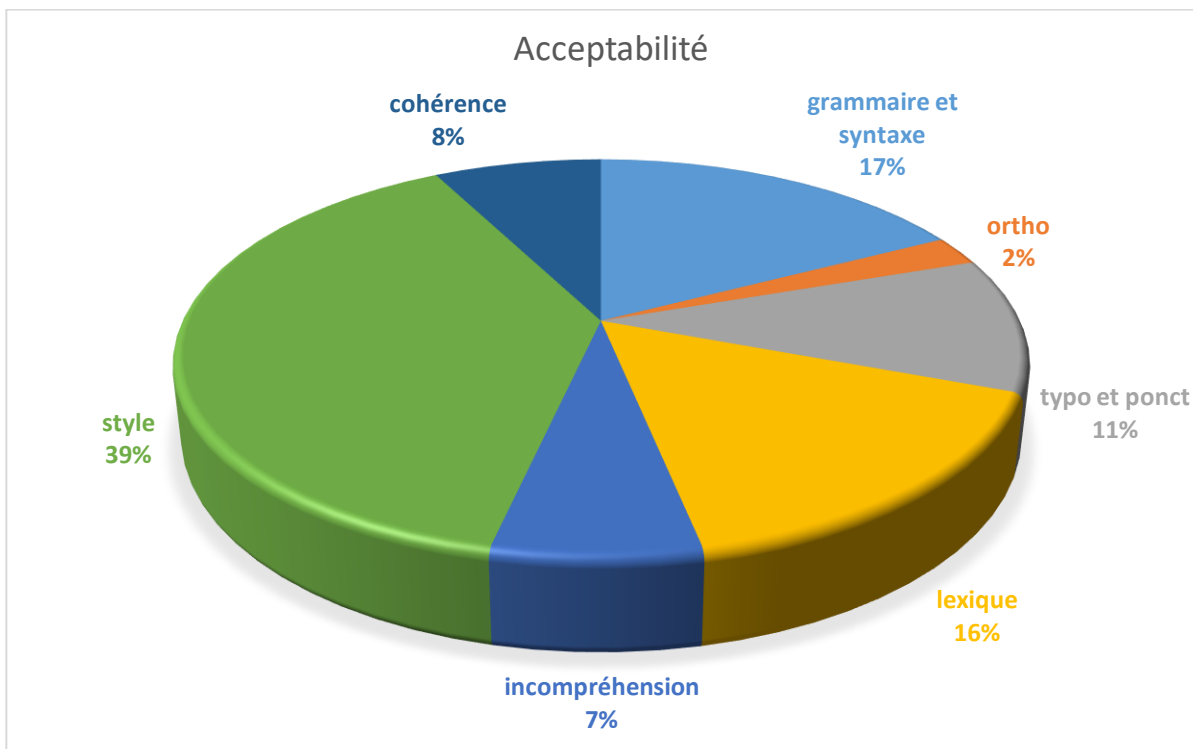


Figure 43 : Diagramme circulaire (Acceptabilité) – Test

Concernant l'acceptabilité du texte cible (Figure 43), nous observons une ventilation semblable à celle du prétest. On peut tout d'abord observer une nette prépondérance des erreurs de « style » (fluidité, idiomatité, répétition, longueur de phrase, registre inadéquat, autre) qui représentent 39 % de l'ensemble des erreurs d'acceptabilité commises par les étudiantes. Les catégories qui arrivent ensuite sont « grammaire et syntaxe » et « lexique » avec respectivement 17 % et 16 %, suivies des catégories « typographie et ponctuation » (11 %), « cohérence » (8 %), « incompréhension » (7 %) et enfin « orthographe » (2 %).

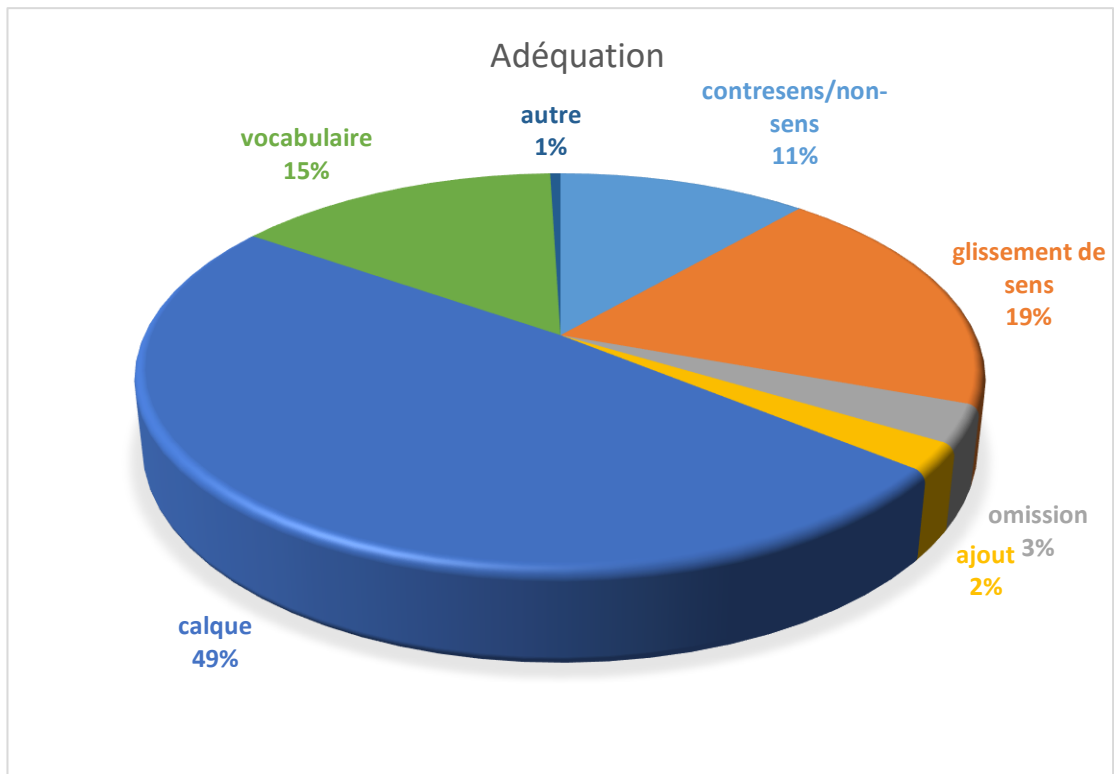


Figure 44 : Diagramme circulaire (Adéquation) – Test

Dans la deuxième phase d'évaluation (Figure 44), la catégorie la plus représentée est celle des calques, tout comme dans le prétest. Dans ce test, les calques représentent près d'une erreur sur deux (49 %). À nouveau, la deuxième catégorie la plus représentée est celle des « glissements de sens », comptant pour 19 % de l'ensemble des erreurs, arrive ensuite la catégorie « vocabulaire » (15 %), « contresens/non-sens » (11 %), « omission » (3 %), « ajout » (2 %) et autre (1 %). Les distributions des erreurs en fonction du mode de traduction et du type de texte sont reprises dans la Figure 45 et la Figure 46.

5.3.1.2 Distribution des erreurs (histogramme)

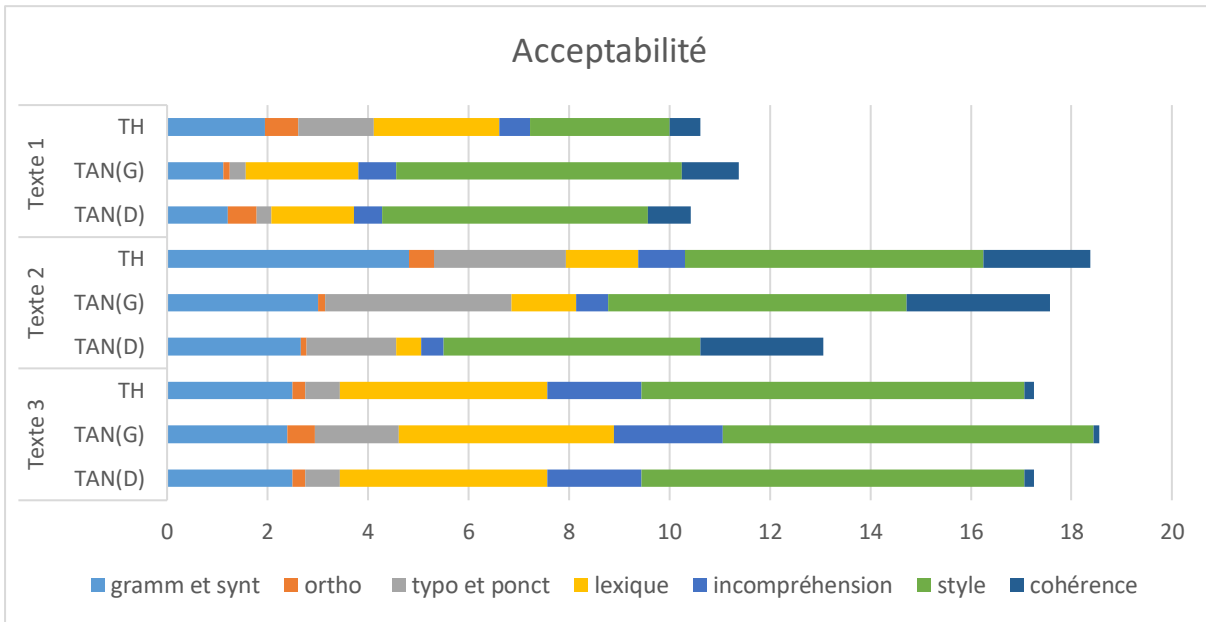


Figure 45 : Histogramme (Acceptabilité) – Test

En analysant la Figure 45, nous pouvons affirmer deux choses : premièrement que le texte 1 (article de presse) a généré moins d'erreurs par rapport aux deux autres TS quel que soit le mode de traduction, et deuxièmement, que pour le texte 2 (texte économique), les étudiantes ont, en moyenne, commis sensiblement moins d'erreurs en PE de TA neuronale générée par *DeepL* (TAN(D)) par rapport aux deux autres modes de traduction.

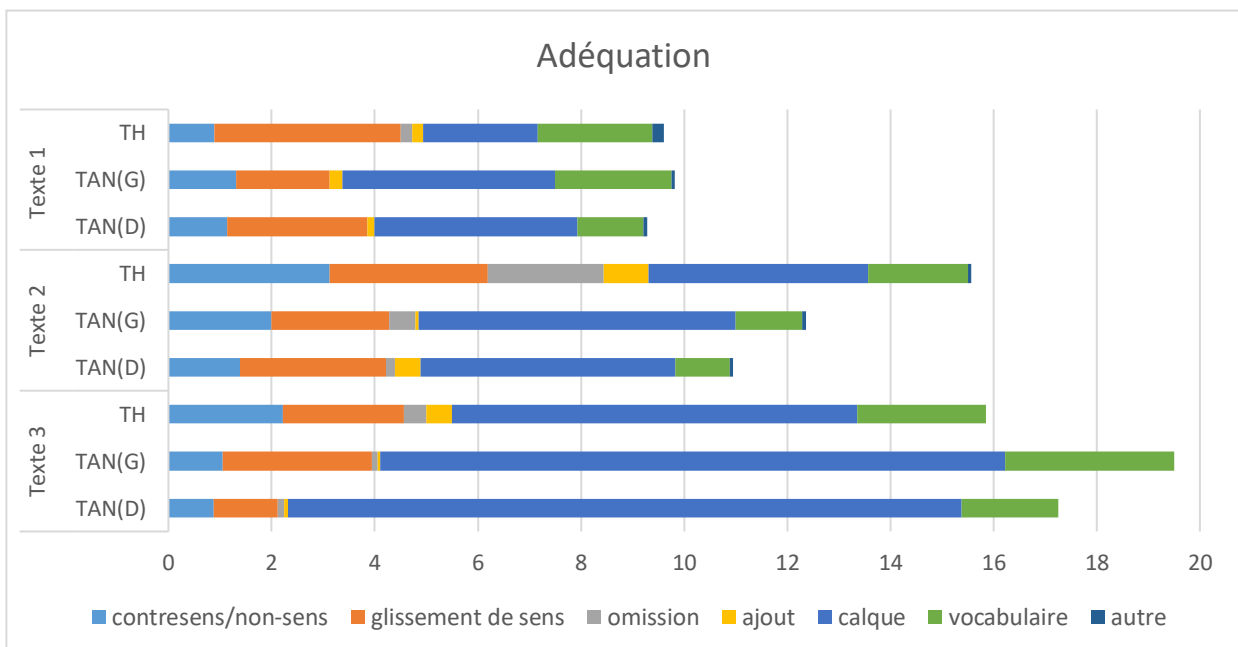


Figure 46 : Histogramme (Adéquation) – Test

En analysant la Figure 46, il apparaît que la TH est le mode de traduction qui a généré le moins de calques, quel que soit le TS. Nous reviendrons sur cette surreprésentation de calques en TA au chapitre 6). Il semble également que la PE de TA générée par *DeepL* ait entraîné, en moyenne, moins d’erreurs de vocabulaire par rapport aux deux autres modes de traduction et enfin, que c’est en PE de TA générée par *Google Traduction* que les étudiantes ont commis le plus d’erreurs en moyenne, quel que soit le TS.

Afin de pouvoir inférer des relations de dépendance entre le mode de traduction et la distribution des erreurs, nous avons à nouveau eu recours à la statistique inférentielle.

5.3.1.3 Comparaison de la qualité par mode de traduction

Rappelons que nous avons calculé, pour chaque production, une note globale selon la pondération que nous avons prédéfinie pour chaque catégorie d’erreurs (Tableau 30). Une fois de plus, cette note globale combine la note obtenue en phase d’acceptabilité (à raison de 40 pour cent) et la note obtenue en phase d’adéquation (à raison de 60 pour cent).

5.3.1.3.1 Moyenne des notes

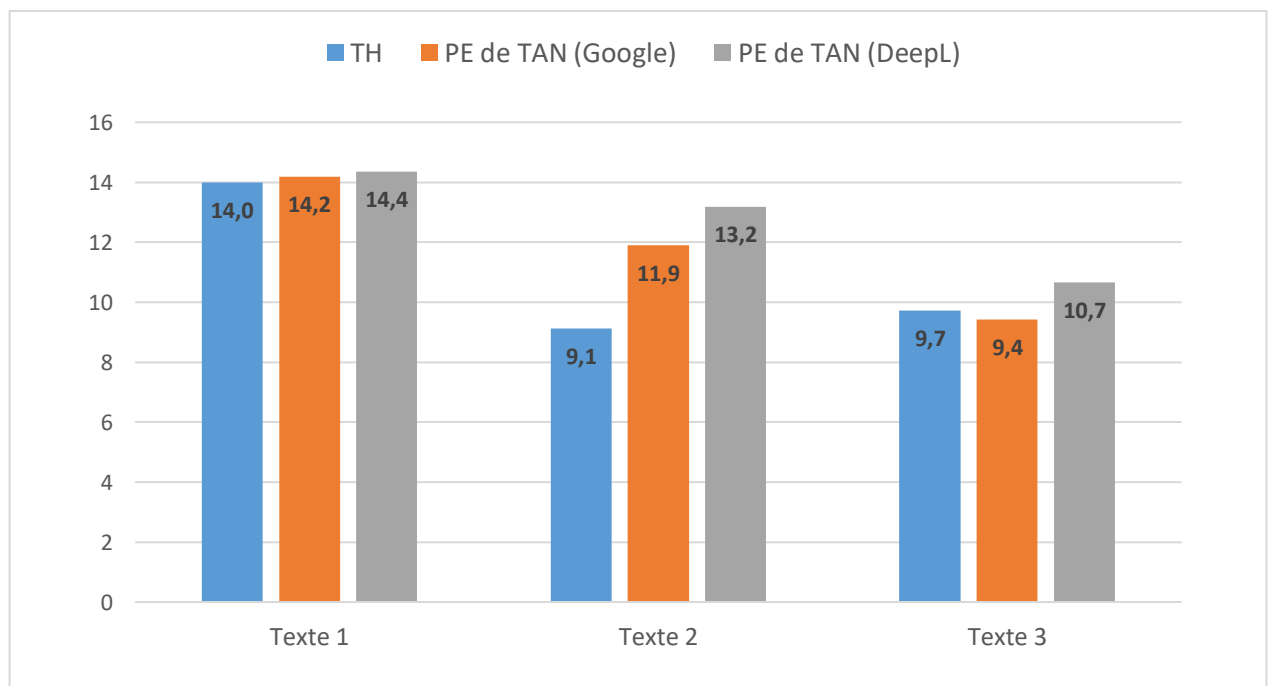


Figure 47 : Moyenne des notes – Test

La Figure 47 nous permet de visualiser la moyenne générale des notes en fonction du mode de traduction et en fonction du TS. Nous constatons que la TH est le mode de traduction pour lequel la moyenne générale des notes est la plus basse par rapport aux deux autres modes, à l’exception du texte 3, tandis que la PE de TA neuronale générée par *DeepL* est le

mode de traduction pour lequel la moyenne des notes est toujours la plus haute, même si la différence est parfois très faible (voir texte 1).

5.3.1.3.2 Taux de réussite

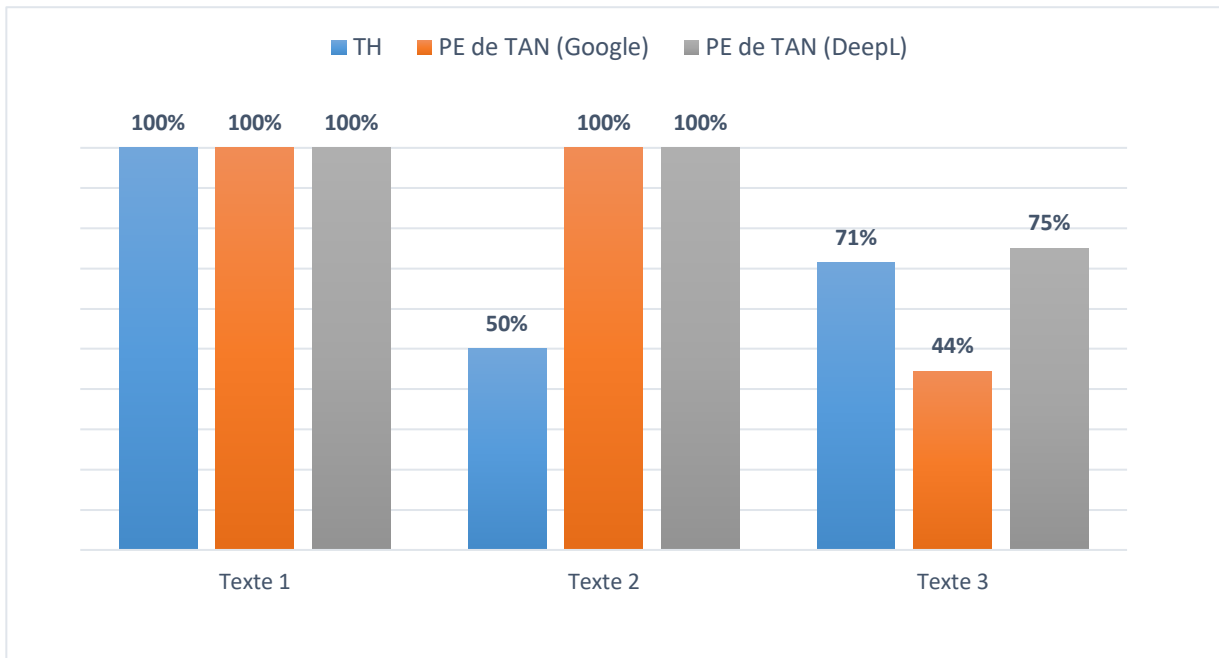


Figure 48 : Taux de réussite selon le mode de traduction – Test

Comme déjà mentionné, dans l'enseignement supérieur en Fédération Wallonie-Bruxelles de Belgique, le seuil de réussite est fixé à 10 sur 20 (Fédération Wallonie-Bruxelles, 2021). Sur la Figure 48, il apparaît que la PE de TA neuronale (*Google Traduction* ou *DeepL*) a permis d'obtenir de très bons taux de réussite pour les trois TS, sauf dans le cas du texte 3 en PE de TA générée par *Google* avec un taux de réussite de 44 %. Il est surprenant de noter que pour le texte 2 (texte économique), la TH n'obtient que 50 % de taux de réussite alors que toutes les étudiantes ayant post-édité ce texte ont obtenu une note de réussite. La TA brute leur aurait-elle permis d'avoir un meilleur « accès » au sens du TS ? Nous imaginons que la TA neuronale peut, dans certains cas, pallier la compréhension lacunaire ou partielle des étudiantes en clarifiant certains passages complexes du TS, ce qui permet d'améliorer la compréhension générale qu'elles ont du TS.

Sans vouloir tirer de conclusions hâtives, nous pouvons affirmer que, dans cette étude, si l'on considère les trois TS, la PE de TA neuronale produite par *DeepL* est le mode de traduction le plus susceptible de faire réussir une étudiante.

Les résultats représentés sur la Figure 47 et sur la Figure 48 suscitent également plusieurs interrogations : peut-on en conclure que le texte 1 (article de presse) était (trop) simple à traduire/post-éditer pour les étudiantes de 2^e master puisque toutes ont obtenu une note égale ou supérieure au seuil de réussite ? Et à l'inverse, que le texte 3 (texte scientifique)

était d'une trop grande complexité à traduire/post-éditer, ou encore que les évaluatrices ont été (trop) clémentes dans leur évaluation des productions du texte 1 et (trop) sévères pour les productions du texte 3 ? Peut-être aussi que la TA brute était de moins bonne qualité pour le texte 3 ? Il s'agit là d'hypothèses explicatives qu'il nous est impossible de confirmer.

5.3.1.4 Moyenne des notes et écart-type

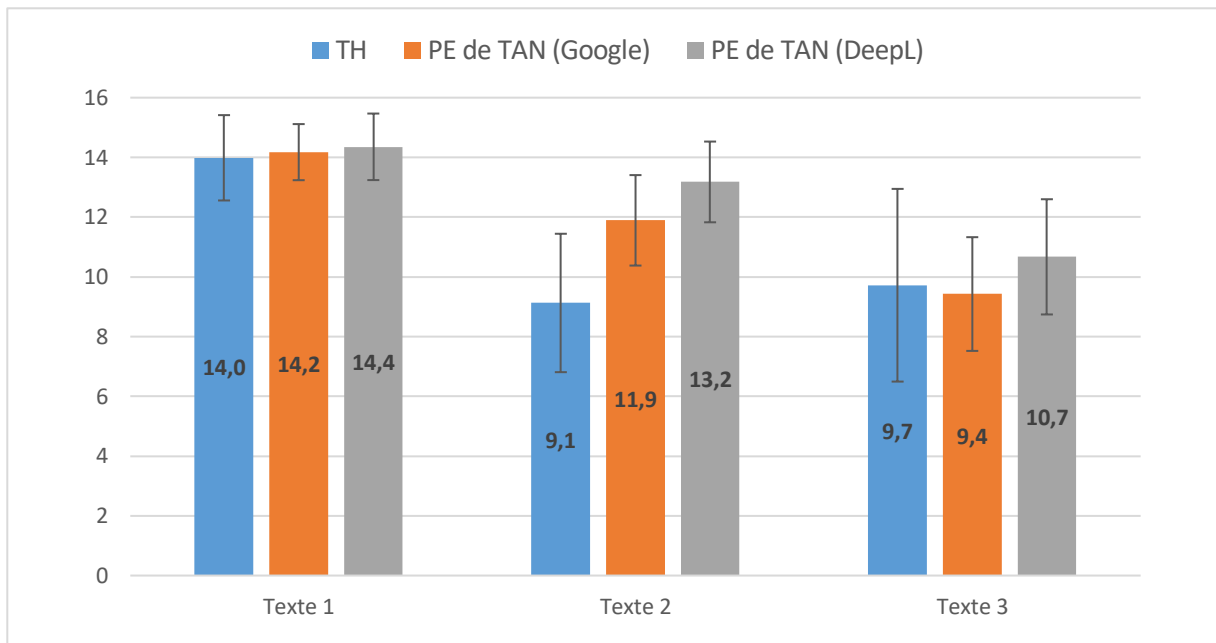


Figure 49 : Moyenne des notes et écart-type – Test

Sur la Figure 49, nous avons repris la moyenne des notes obtenues en fonction du mode de traduction et du TS, ainsi que l'écart-type. D'emblée, nous constatons une réduction de l'écart-type en PE (*Google Traduction* et *DeepL*) par rapport à l'écart-type calculé en TH, quel que soit le TS. Ces résultats, similaires aux résultats du prétest, reflètent un resserrement des notes autour de la moyenne en PE, et ce, pour les trois textes. Ce resserrement traduit une répartition moins étalée des notes en PE ; les textes post-édités ont obtenu moins de (très) mauvaises notes et moins de (très) bonnes notes que les TH.

5.3.1.5 Dispersion des notes autour de la médiane

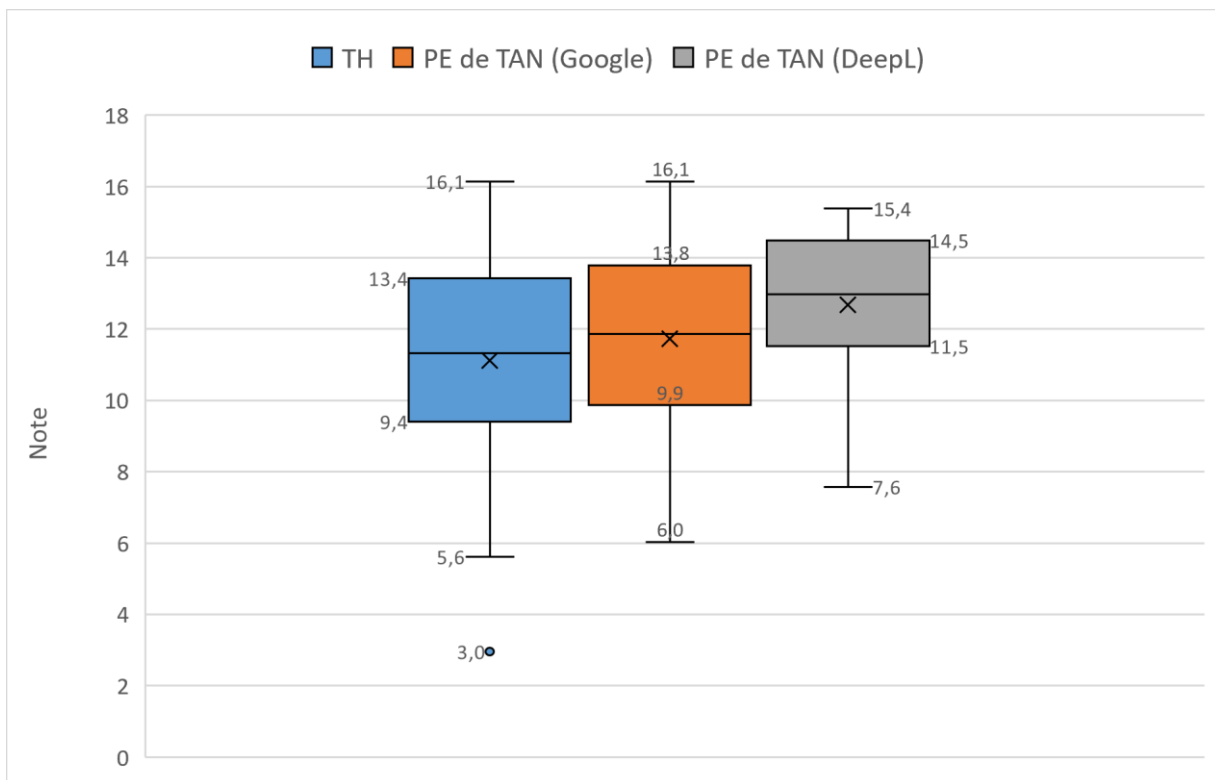


Figure 50 : Dispersion des notes autour de la médiane – Test

La Figure 50 en boîte à moustaches permet de visualiser la dispersion des notes autour de la médiane selon le mode de traduction. Rappelons que la moustache supérieure et la moustache inférieure indiquent respectivement la note la plus élevée et la note la plus basse de la série de notes, et que la « boîte » délimitée par le 1^{er} quartile (Q1) et le 3^e quartile (Q3) représente l'intervalle interquartile qui contient au moins 50 pour cent des notes ; le marqueur x indique la moyenne des notes et la ligne horizontale représente la médiane.

Tout comme dans le prétest, nous pouvons constater un resserrement des notes autour de la médiane. Nous relevons également que, non seulement la moyenne, mais aussi la note médiane, sont plus élevées en PE de TA générée par *DeepL* par rapport à la PE de TA générée par *Google Traduction*, mais davantage encore par rapport à la TH. Le resserrement des notes se traduit par une réduction de l'écart interquartile (Q3-Q1) : 4,02 en TH ; 3,92 en PE de TAN (*Google*) et 2,97 en PE de TAN (*DeepL*). Cette réduction est tout particulièrement marquée en PE de *DeepL*.

Enfin, notons la présence d'une valeur extrêmement basse (3,0) en TH qui se trouve représentée par un point en dessous de la moustache inférieure ; il s'agit d'une valeur extrême dite « aberrante »¹⁰⁵.

5.3.1.6 PE de Google vs PE de DeepL

Nous avons voulu vérifier l'hypothèse HA6 selon laquelle la qualité d'un texte post-édité serait influencée par le moteur de TA neuronale (*Google Traduction* ou *DeepL*). Pour ce faire, nous avons comparé les productions obtenues en PE de *Google Traduction* aux productions obtenues en PE de *DeepL* selon le niveau moyen d'erreurs comptabilisées par texte source.

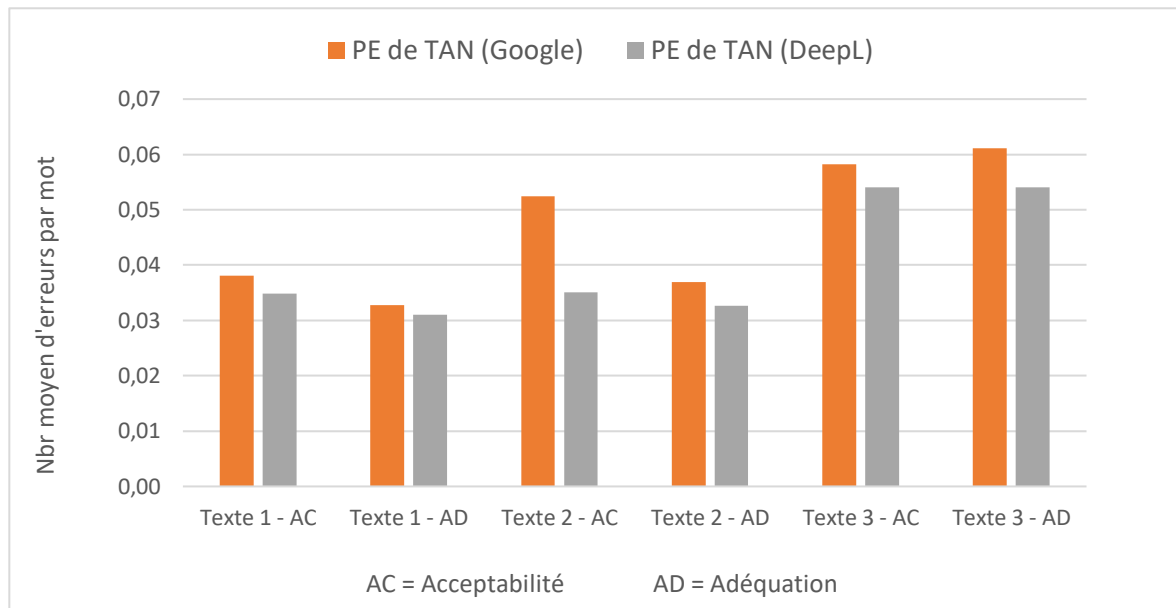


Figure 51 : PE de Google vs PE de DeepL

Sur la Figure 51, nous retrouvons, pour chaque texte, le nombre moyen d'erreurs relevées en phase d'acceptabilité (AC) et le nombre moyen d'erreurs en phase d'adéquation (AD). Afin d'établir une base de comparaison valable, nous avons normalisé nos résultats en prenant comme indicateur le nombre moyen d'erreurs par mot du TS. Les résultats représentés sur cette figure révèlent que la PE de *DeepL* a permis d'aboutir à des productions de meilleure qualité, puisqu'elles contiennent, en moyenne, moins d'erreurs que les PE de TA générée avec *Google Traduction*. Cette conclusion vaut aussi bien pour la phase d'acceptabilité que pour la phase d'adéquation, quel que soit le TS. Ces résultats nous permettent de rejeter l'hypothèse nulle HA6.0.

¹⁰⁵ « Une valeur qui semble dévier de façon marquée par rapport à l'ensemble des autres membres de l'échantillon dans lequel il apparait » (Grubbs, 1969, p. 1).

5.3.1.7 Performances par étudiante

Nous avons à nouveau approfondi ces observations en analysant les performances par étudiante.

5.3.1.7.1 Nombre moyen d'erreurs

Les données ont été, une fois de plus, normalisées en prenant comme référence le nombre d'erreurs par mot du TS. Sur la Figure 52, la Figure 53, la Figure 54 et la Figure 55, les traductions humaines (TH) sont classées de manière croissante : de la production qui contient le plus grand nombre d'erreurs par mot à celle qui en contient le moins, autrement dit, de la moins bonne TH à la meilleure, indépendamment du TS.

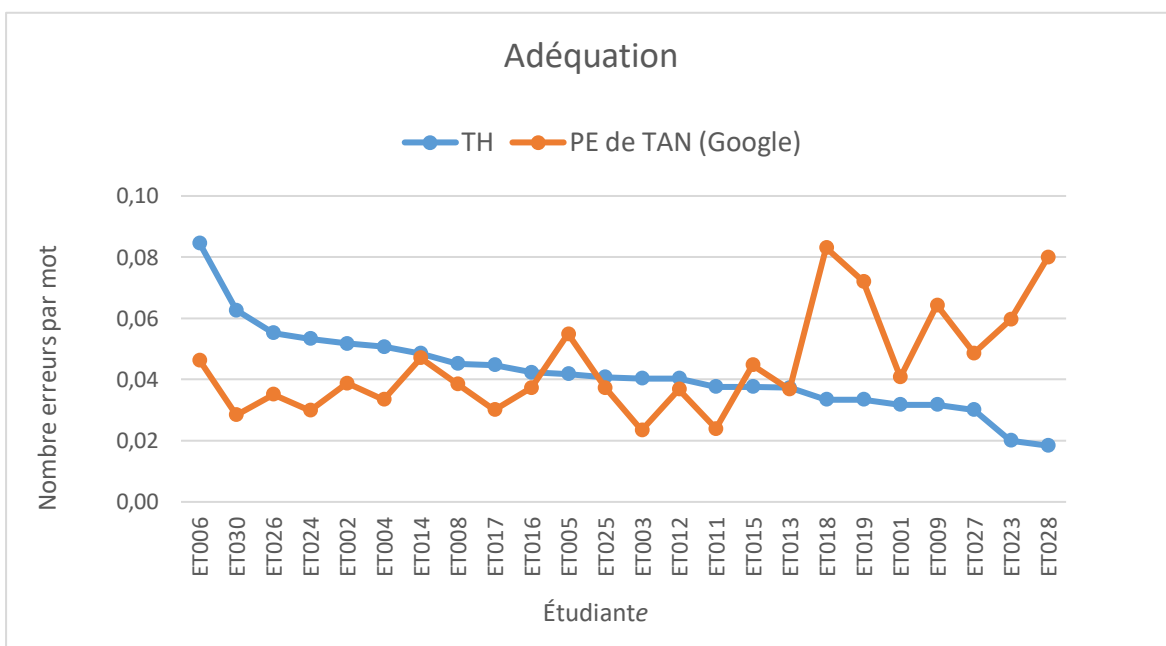


Figure 52 : Comparaison nombre d'erreurs par étudiante (Adéquation) Google

Sur cette Figure 52, la TH de l'étudiante ET025 se situe à la médiane (0,041). Nous constatons d'emblée que les étudiantes qui commettent en moyenne le plus grand nombre d'erreurs d'adéquation en TH commettent systématiquement moins d'erreurs dans leur PE. En effet, pour toutes les étudiantes dont la TH se situe au-dessus de cette médiane ($> 0,041$) (sur la gauche de la figure), nous présumons que le recours à la PE de TA neuronale a été bénéfique et leur a permis d'aboutir à un texte cible final contenant moins d'erreurs en moyenne que leur TH, à l'exception de l'étudiante ET005. Et l'inverse se vérifie également : la majorité des étudiantes dont la TH se situe en dessous de cette médiane ($< 0,041$) ont produit une PE qui contient (nettement) plus d'erreurs par mot du TS par rapport à leur production humaine (sur la droite de la figure).

Les trois figures suivantes nous permettent d’observer ce phénomène en phase d’adéquation avec les PE de TA générée avec *DeepL* (Figure 53), ainsi qu’en phase d’acceptabilité avec les PE de TA produites par *Google* (Figure 54) et par *DeepL* (Figure 55). La même tendance que sur la Figure 52 se dégage, quoique beaucoup moins nettement.

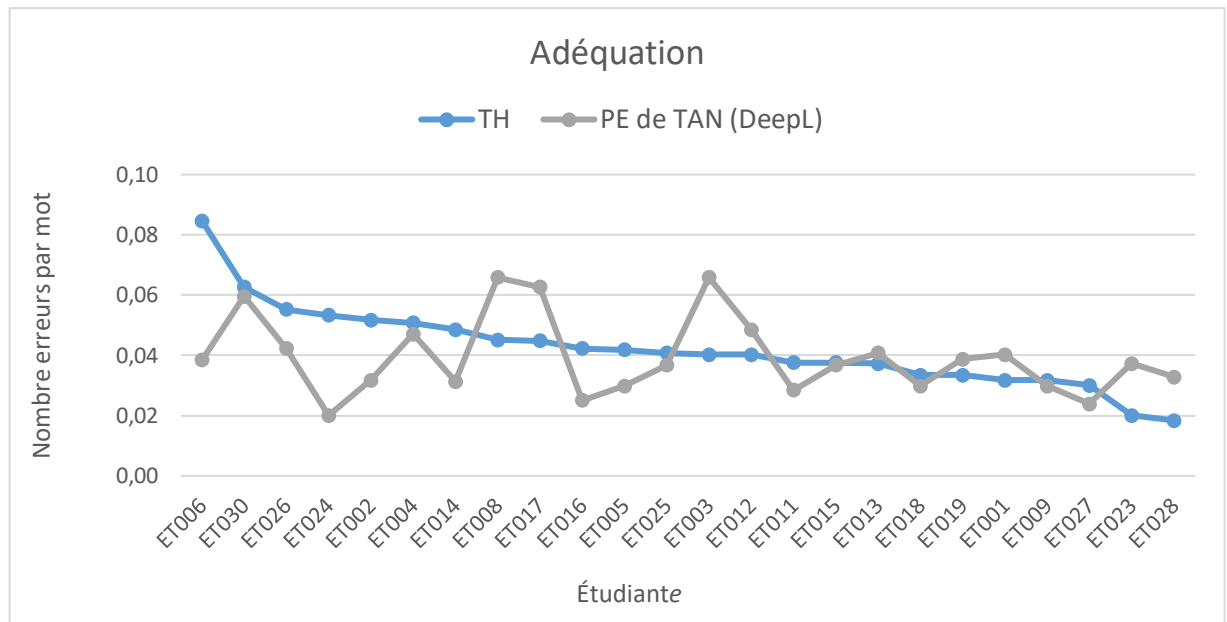


Figure 53 : Comparaison nombre d’erreurs par étudiante (Adéquation) *DeepL*

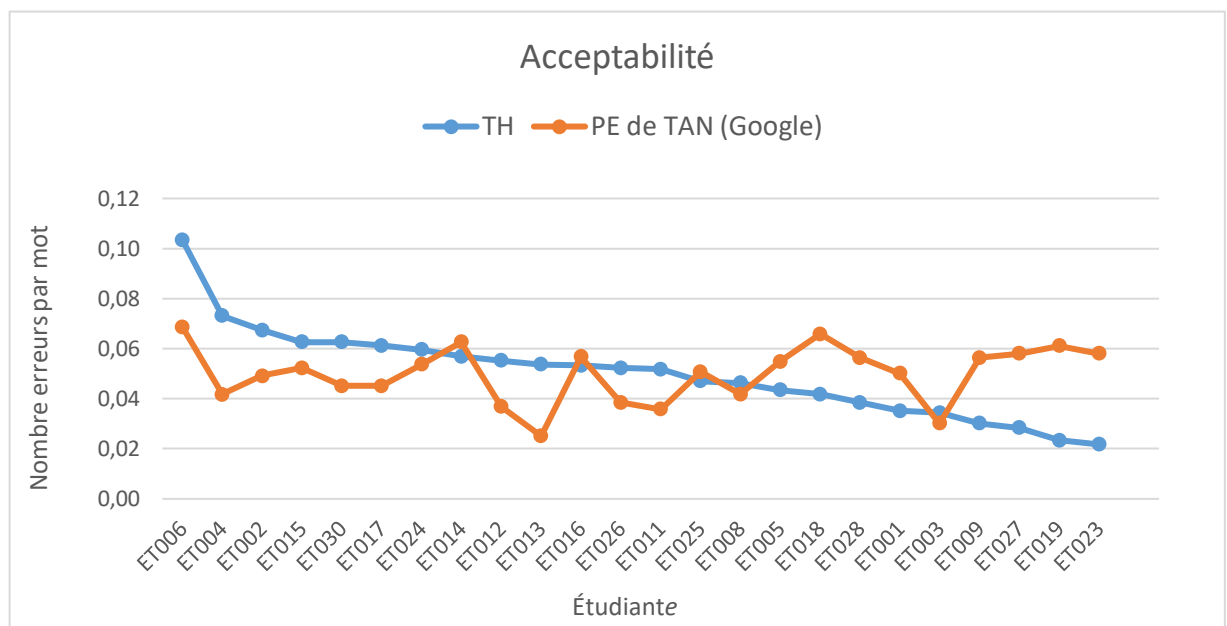


Figure 54 : Comparaison nombre d’erreurs par étudiante (Acceptabilité) *Google*

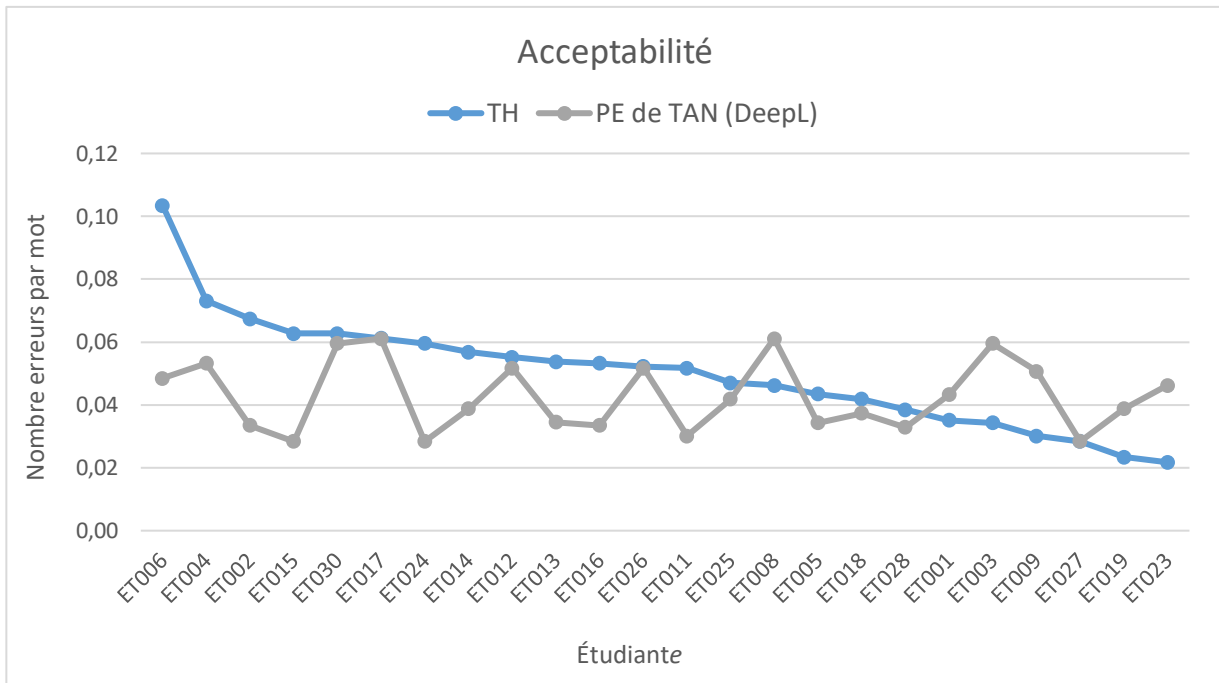
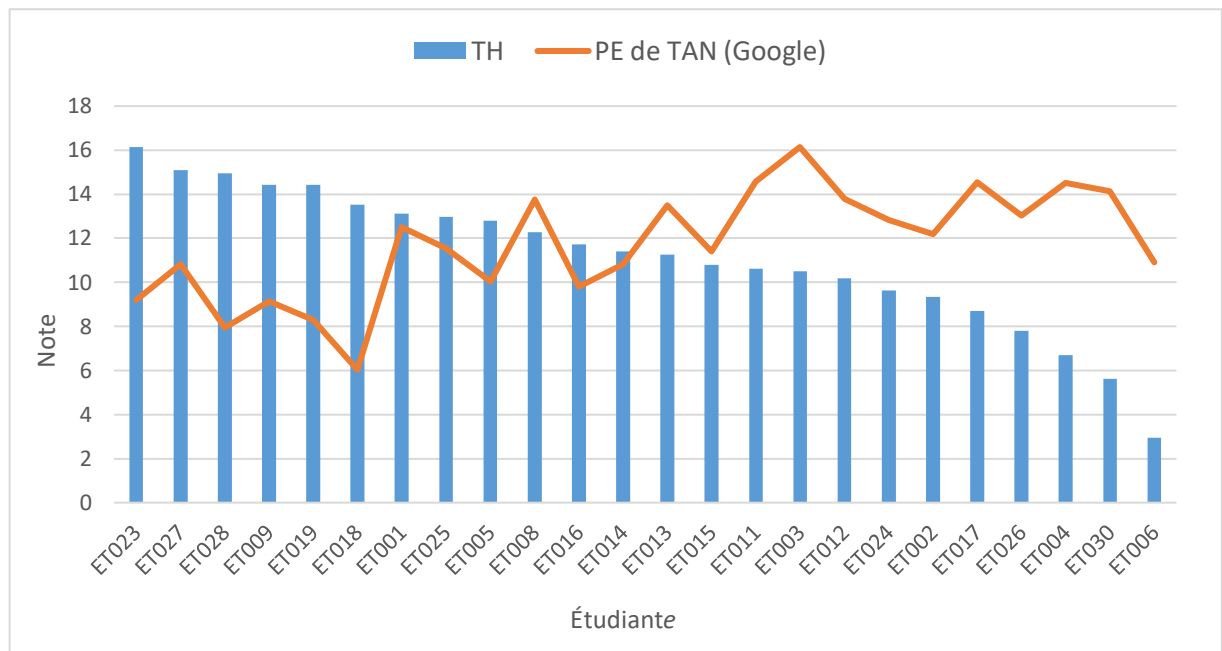


Figure 55 : Comparaison nombre d'erreurs par étudiante (Acceptabilité) *DeepL*

5.3.1.7.2 Notation

Nous avons souhaité vérifier si cette tendance se dégageait également au niveau de la notation. La Figure 56 et la Figure 57 ont été obtenues en comparant, pour une même étudiante, les notes attribuées en TH à celles attribuées en PE de TA (*Google Traduction* pour la Figure 56 et *DeepL* pour la Figure 57). Sur ces deux figures, les TH sont classées à présent de manière décroissante : de la meilleure production à la moins bonne.

Figure 56 : Effet nivelant en PE de TAN (*Google*)

Sur la Figure 56, nous voyons que les étudiantes dont la note en TH est située en deçà de la médiane (11,3) — ET013 jusque ET006 inclus — obtiennent systématiquement une meilleure note en PE de TA générée avec *Google Traduction*. De plus, les étudiantes que nous considérons comme les plus faibles, à savoir celles dont la note arrondie en TH n'atteint pas le seuil de réussite de 10/20 — ET002 jusque ET006 inclus — ont obtenu une note nettement plus élevée en PE. Nous en concluons que ces étudiantes bénéficient particulièrement de ce mode de traduction. À l'inverse, les étudiantes dont les productions humaines sont situées au-delà de la valeur médiane — ET023 à ET014 inclus — ont toutes obtenu une note plus basse en PE de TA neuronale à l'exception de l'étudiante ET008. Cette figure nous permet une nouvelle fois de relever un effet nivelant en PE de TA neuronale.

Sur la Figure 57, nous observons également ce phénomène en PE de TA générée avec *DeepL*, toutefois de manière moins marquée.

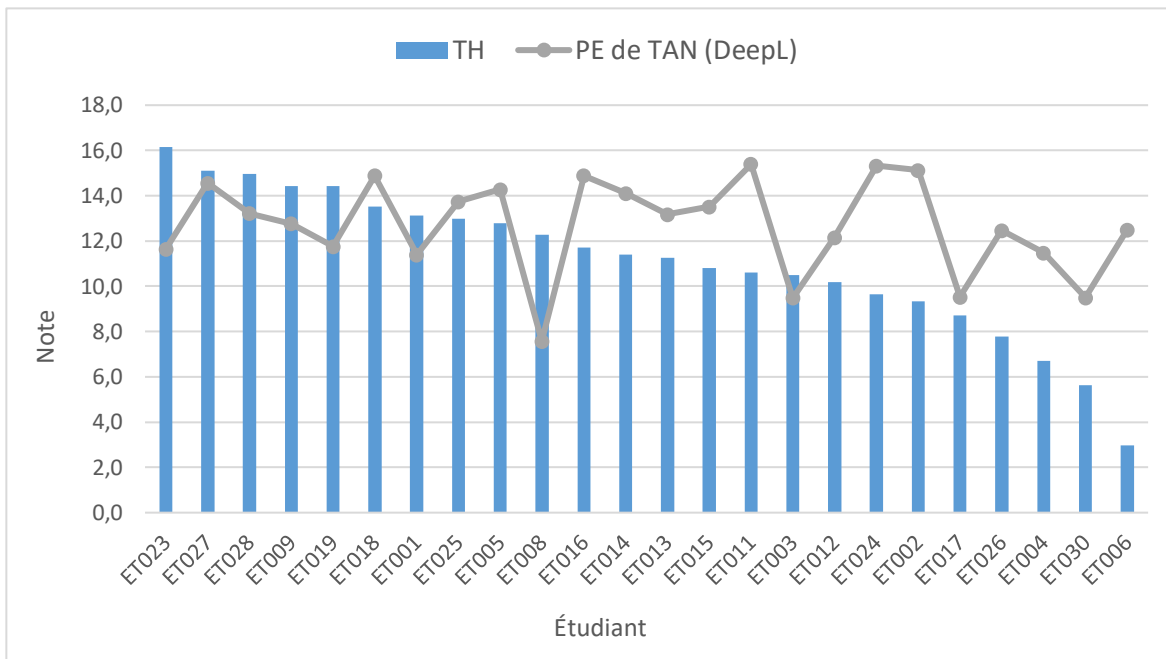


Figure 57 : Effet nivelant en PE de TAN (*DeepL*)

Les résultats que nous venons de présenter, et particulièrement les Figure 55 et Figure 56, nous amènent à tirer la conclusion suivante : la PE de TA neuronale (en particulier *Google Traduction*) chez les étudiantes aurait un effet nivelant sur la qualité des textes cibles. En comparaison à la TH, ce mode de traduction permettrait aux étudiantes faibles de s'en sortir en réduisant le nombre moyen d'erreurs et en leur évitant d'obtenir de (très) mauvaises notes, mais ce processus aurait également pour effet de diminuer les performances des bons éléments.

5.3.1.8 Qualité des TA brutes

Nous avons décidé d'intégrer, pour chaque TS, les deux TA brutes (*Google Traduction* et *DeepL*) au corpus d'évaluation humaine à l'insu des évaluateurices. Chaque sortie brute a donc été évaluée par deux évaluateurices, au même titre que les productions des étudiantes, ce qui nous a permis de leur attribuer une note sur 20 (Tableau 31). Pour rappel, les trois TS n'ont pas été évalués par les mêmes binômes d'évaluateurices. Dans un souci de pertinence et de validité des résultats, ces six TA brutes (deux par TS) ont été exclues de l'analyse quantitative.

Note générale /20		
	<i>Google Traduction</i>	<i>DeepL</i>
Texte 1 (presse)	10,9	10,7
Texte 2 (économique)	10,9	13,4
Texte 3 (vulg. scientifique)	4,3	6,8

Tableau 31 : Notes attribuées aux TA brutes

Sur le Tableau 31, il est très interpellant de constater que quatre sorties de TA sur six ont obtenu une note supérieure au seuil de réussite fixé à 10 sur 20, même s'il faut bien noter que trois d'entre elles sont inférieures à la note de 11 sur 20. C'est la TA produite par *DeepL* pour le texte économique qui reçoit la note la plus élevée : 13,4 sur 20.

En comparaison, les TA du texte de vulgarisation scientifique ont obtenu de très mauvaises notes (4,3 et 6,8 sur 20). On pourrait donc légitimement se demander si le texte 3 ne se prêtait pas moins bien à une traduction machine en raison de plusieurs facteurs (thématique ; complexité syntaxique, terminologique...) ou si le binôme d'évaluatrices du texte 3 n'a pas été plus sévère dans son évaluation de la qualité par rapport aux deux autres textes.

Si l'on compare à présent les moyennes générales (Figure 47) aux notes attribuées par les évaluatrices humaines aux TA brutes (Tableau 31), il apparaît que les textes post-édités ont été jugés, en moyenne, de meilleure qualité que les TA brutes. La PE a donc permis de faire systématiquement remonter la moyenne des notes à une exception : la moyenne générale en PE de TA générée par *DeepL* pour le texte 2 est de 13,2 sur 20, soit une valeur plus basse que la note qui a été attribuée par les évaluatrices à la TA brute (13,4) ! Cette note est également supérieure aux valeurs médianes obtenues en TH, en PE de TA (*Google*) et en PE de TA (*DeepL*) pour le texte 2.

	Texte 1		Texte 2		Texte 3	
	<i>Google</i>	<i>DeepL</i>	<i>Google</i>	<i>DeepL</i>	<i>Google</i>	<i>DeepL</i>
Acceptabilité						
Grammaire/Syntaxe	1,5	2	2	2	4	4
Orthographe	0	1	0	0	0	0
Typographie/ponctuation	0	0,5	3	2	2,5	0
Lexique	5	3,5	1,5	0,5	5,5	7,5
Incompréhension	1	1	1,5	0	2,5	2
Style	12	10	9,5	6	12,5	11,5
Cohérence	2	0,5	3	1,5	0	0
Adéquation						
Contresens/non-sens	1	2	1	1	1,5	1
Glissement de sens	2	1	3	2,5	2,5	0
Omission	0	0	0	0	0	0
Ajout	0	0	0	0	0	0
Calque	10	12	12	9	26	23
Vocabulaire	3,5	1	0	0,5	1,5	2
Autre	0	0	0	0	0	0

Tableau 32 : Distribution erreurs en TA brute (Acceptabilité et Adéquation)

En observant le Tableau 32, nous constatons que les catégories d'erreurs les plus représentées sont celles du « style » et du « lexique » en phase d'acceptabilité. Tandis qu'en phase d'adéquation, les évaluatrices ont jugé que les TA brutes contenaient principalement des calques fautifs, en particulier le texte 3, ce qui corrobore les résultats à la fois du prétest et d'une autre expérience que nous avons menée en contexte universitaire (Schumacher 2020b).

5.3.2 Statistiques inférentielles

Tout comme pour le prétest, nous avons effectué un test ANOVA à deux facteurs de type III, en mesures répétées, ainsi qu'un test post-hoc de Tukey avec les données récoltées lors de l'évaluation humaine. La qualité du texte cible est mesurée tantôt par le nombre d'erreurs relevées dans les productions des étudiantes (Tableau 33 et Tableau 34), tantôt par les notes attribuées aux productions des étudiantes (Tableau 35, Tableau 36 et Tableau

37). Le seuil de significativité est fixé à 0,05 pour l'ensemble des tests statistiques. Si $P > 0,05$, le résultat sera dit statistiquement non significatif, tandis que si $P < 0,05$, il sera considéré comme statistiquement significatif, et les risques de se tromper sont inférieurs à 5 %. Pour une description détaillée du test statistique voir 4.4.2.

Les résultats sont à nouveau présentés comme suit :

- A) la qualité globale des productions (5.3.2.1) selon trois niveaux :
- les erreurs en phase d'acceptabilité (5.3.2.1.1)
 - les erreurs en phase d'adéquation (5.3.2.1.2)
 - les notes globales attribuées aux productions des étudiantes (5.3.2.1.3).
- B) la qualité des productions par catégories d'erreur (5.3.2.2)

5.3.2.1 Qualité globale

5.3.2.1.1 Effets sur les erreurs en phase d'acceptabilité

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## methode      73.26   36.63     2  53.146  5.3106  0.007895 **
## texte       784.60  392.30     2  39.837 56.8785 2.119e-12 ***
## methode:texte  41.72   10.43     4  54.986  1.5122  0.211329
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau 33 : ANOVA — Distribution des erreurs en phase d'acceptabilité

Penchons-nous tout d'abord sur les effets du mode de traduction et du texte source sur le nombre d'erreurs comptabilisées en phase d'acceptabilité.

Interaction mode–texte

Le Tableau 33 : ANOVA — Distribution des erreurs en phase d'acceptabilité permet de constater l'absence d'interaction significative entre le mode de traduction et le texte (« methode:texte ») ($F = 1,51$; $dfN = 4$; $dfD = 54,98$; $P = 0,211329$). Autrement dit, la qualité des productions ne dépend pas significativement de l'interaction entre nos deux variables indépendantes.

Mode de traduction

Il apparaît que l'effet « mode de traduction » (« methode » dans le Tableau 33) est significatif, car $P < 0,01$ ($F = 5,31$; $dfN = 2$; $dfD = 53,14$; $P = 0,007895$), ce qui révèle qu'il y a une différence significative entre les distributions de valeurs pour les erreurs d'acceptabilité correspondant à chaque mode de traduction. Ce résultat signifie que la distribution des erreurs d'acceptabilité a été influencée par le mode de traduction, et ce, quel que soit le texte source.

Test de Tukey 1

Quant aux résultats du test de Tukey pour les modes de traduction (Tableau D. 1 en annexe), ils révèlent un écart statistiquement significatif entre la PE de *Google* et la PE de *DeepL* ($P < 0,05$), ainsi qu'entre la TH et la PE de *DeepL* ($P < 0,01$). La PE de *Google* présente une moyenne supérieure à celle en PE de *DeepL* de 2,24 ; la TH étant en moyenne supérieure à la PE de TA (*DeepL*) de 2,59. Par contre, il n'y a pas de différence significative entre la TH et la PE de *Google* ($P > 0,05$). Ces résultats montrent que la distribution des erreurs d'acceptabilité (erreurs de langue, problèmes de fluidité et de cohérence en langue cible) a été influencée par le mode de traduction, et ce, quel que soit le texte source. Dans cette expérience, les étudiantes ont commis significativement moins d'erreurs d'acceptabilité lorsqu'elles ont post-édité du *DeepL* que dans leur PE de *Google* et leur TH. Ces observations nous amènent à nous demander si les TA brutes générées par *DeepL* sont davantage exploitables que les TA de *Google* et si elles ont par conséquent nécessité moins d'effort de PE que les autres productions.

Texte source

L'effet « texte source » (« texte » dans le Tableau 33) est, lui aussi, significatif avec $P < 0,001$ ($F = 56,87$; $dfN = 2$; $dfD = 39,83$; $P < 0,001$), ce qui révèle qu'il y a une différence significative entre les distributions de valeurs pour les erreurs d'acceptabilité correspondant à chaque texte.

Test de Tukey 2

Les résultats du test de Tukey pour les TS (Tableau D. 2 en annexe) révèlent l'existence de différences significatives entre tous les textes sources. Grâce à ces observations, nous pouvons affirmer que, pour la phase d'acceptabilité, la thématique du texte source constitue bel et bien une variable dans cette expérience, au même titre que le mode de traduction. Le texte 3 présente une moyenne supérieure à celle du texte 1 ($P < 0,001$) et à celle du texte 2 ($P < 0,01$), respectivement de 7,93 et de 2,40. Le texte 2 présente une moyenne supérieure à celle du texte 1 ($P < 0,001$) de 5,53.

Remarquons que le texte source pour lequel les étudiantes ont commis significativement le plus grand nombre d'erreurs d'acceptabilité est le texte 3 (vulgarisation scientifique) et ce, quel que soit le mode de traduction. Tandis que le texte qui a engendré le moins d'erreurs dans le chef des étudiantes est le texte 1 (article de presse).

5.3.2.1.2 Effets sur les erreurs en phase d'adéquation

```

mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## methode         23.53   11.77     2 59.264  1.3397  0.26973
## texte          758.41  379.20     2 42.184 43.1768 6.223e-11 ***
## methode:texte  110.14   27.53     4 60.738  3.1351  0.02073 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Tableau 34 : ANOVA – Distribution des erreurs en phase d'adéquation

Nous étudions à présent les effets du mode de traduction et du texte source sur le nombre d'erreurs comptabilisées cette fois en phase d'adéquation (erreurs sémantiques).

Interaction mode–texte

Le test ANOVA (Tableau 34) révèle une interaction significative entre le mode de traduction et le texte ($F= 3,13$; $dfN= 4$; $dfD= 60,73$; $P=0,02073$). Cela signifie que le mode de traduction a eu une influence sur le nombre moyen d'erreurs d'adéquation, mais uniquement en interaction avec le texte source, autrement dit, les effets d'un mode de traduction varient d'un texte à l'autre. Nous pouvons dire que dans cette expérience, la qualité des productions dépend significativement de l'interaction entre nos deux variables indépendantes (modes de traduction et textes sources). Étant donné la présence de cette interaction, il n'est pas pertinent d'analyser isolément les effets des variables indépendantes.

5.3.2.1.3 Effets sur les notes

La qualité du texte cible est à présent mesurée au moyen des notes attribuées aux productions des étudiantes. Dans le Tableau 35, les tests ont été réalisés sur les notes attribuées en phase d'acceptabilité, dans le Tableau 36 sur celles attribuées en phase d'adéquation et dans le Tableau 37 sur les notes globales qui combinent la note obtenue en phase d'acceptabilité (40 %) à la note obtenue en phase d'adéquation (60 %).

5.3.2.1.3.1 Notes d'acceptabilité

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## methode      30.065  15.032     2  57.260  5.5304  0.006369 **
## texte       238.880 119.440     2  39.741 43.9418 8.544e-11 ***
## methode:texte  18.170   4.542     4  59.114  1.6711  0.168729
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau 35 : ANOVA — Distribution des notes en phase d'acceptabilité

Le test ANOVA réalisé sur les notes attribuées en phase d'acceptabilité (Tableau 35) indique l'absence d'interaction entre les modes de traduction et les textes sources ($F = 1,67$; $dfN = 4$; $dfD = 59,11$; $P = 0,168729$).

Mode de traduction

Cependant, l'effet de la variable « mode de traduction » se révèle significatif ($F = 5,53$; $dfN = 2$; $dfD = 57,26$; $P = 0,006369$). Nous pouvons dès lors inférer l'existence d'une dépendance entre le mode de traduction et la note obtenue en phase d'acceptabilité.

Test de Tukey 1

Les résultats du test de Tukey pour les modes de traduction (Tableau D. 4 en annexe) montrent que les moyennes sont statistiquement différentes entre la TH et la PE de *DeepL* ($P < 0,01$). La TH présente une moyenne inférieure de 1,75 à celle obtenue en PE de *DeepL*. Ainsi, nous pouvons affirmer que les étudiantes ont obtenu, en moyenne, de meilleures notes en phase d'acceptabilité lorsqu'elles ont post-édité *DeepL* qu'en produisant la TH.

Texte source

L'effet de la variable « texte source » (Tableau 35) est lui aussi significatif ($F = 43,94$; $dfN = 2$; $dfD = 39,74$; $P < 0,001$). Il existe une relation de dépendance entre le texte source et les notes attribuées en phase d'acceptabilité.

Test de Tukey 2

Les moyennes des notes sont statistiquement différentes pour les trois textes sources (Tableau D. 52 en annexe). Il y a tout d'abord un écart très significatif ($P < 0,001$) entre le texte 2 et le texte 1, le texte 2 ayant une moyenne inférieure au texte 1 de 2,79, ainsi qu'entre le texte 3 et le texte 1 avec un écart de 4,49. Enfin, le texte 3 présente une moyenne significativement supérieure à celle du texte 2 ($P < 0,01$).

Il ressort de ces tests que c'est une fois de plus le texte 3 qui récolte, en moyenne, les moins bonnes notes en phase d'acceptabilité, tandis que les productions du texte 1 ont obtenu, en moyenne, les meilleures notes, quel que soit le mode de traduction.

5.3.2.1.3.2 Notes d'adéquation

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## methode         31.747  15.873     2  56.248  3.9142  0.02561 *
## texte          210.707 105.353     2  41.815 25.9792 4.643e-08 ***
## methode:texte   46.696  11.674     4  57.955  2.8787  0.03041 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau 36 : ANOVA — Distribution des notes en phase d'adéquation

Interaction mode–texte

Sur le Tableau 36, l'analyse des résultats révèle que l'interaction entre les modes de traduction et les textes sources est significativement présente ($F = 2,89$; $dfN = 4$; $dfD = 57,95$; $P = 0,03041$). Le mode de traduction a eu une influence sur les notes obtenues en phase d'adéquation, mais uniquement en interaction avec le texte source. Nous pouvons en conclure une fois de plus que, dans cette expérience, la qualité des productions dépend significativement de l'interaction entre nos deux variables indépendantes (modes de traduction et textes sources). Par conséquent, nous n'approfondirons pas l'analyse étant donné qu'il existe une interaction statistiquement significative entre les modes de traduction et les TS.

5.3.2.1.3.3 Note globale (acceptabilité + adéquation)

Observons à présent si le mode de traduction ou le texte source ont exercé une influence significative sur la note globale attribuée aux productions des étudiantes.

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## methode         27.942  13.971     2  52.831  6.6948 0.002562 **
## texte          219.513 109.756     2  41.187 52.5945 4.56e-12 ***
## methode:texte   30.033   7.508     4  54.485  3.5978 0.011284 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau 37 : ANOVA — Distribution des notes globales

Interaction mode–texte

Les résultats ANOVA repris dans le Tableau 37 révèlent la présence d’une interaction significative entre les deux variables : modes de traduction et textes sources ($F= 3,59$; $dfN= 4$; $dfD= 54,48$; $P= 0,011284$).

Une fois de plus, le mode de traduction a eu une influence sur les notes globales (acceptabilité + adéquation), mais uniquement en interaction avec le texte source. Il n’est dès lors pas pertinent d’analyser les effets principaux isolément.

5.3.2.2 Qualité par catégorie d’erreurs

Les tests ANOVA repris ci-dessous nous ont permis de déterminer les effets du mode de traduction ou du texte source sur la qualité, non plus globale, mais pour chaque catégorie d’erreurs.

5.3.2.2.1 Catégories en phase d’acceptabilité

Commençons par l’analyse des résultats pour les catégories en phase d’acceptabilité.

5.3.2.2.1.1 Grammaire et syntaxe

Interaction mode–texte

Le test ANOVA pour les erreurs de grammaire et de syntaxe (Tableau D. 6 en annexe) indique l’absence d’interaction significative entre les modes de traduction et les textes sources.

Mode de traduction

Toujours dans le Tableau D. 6 en annexe, nous observons une relation de dépendance entre le mode de traduction et le nombre d’erreurs de grammaire et de syntaxe présentes dans les productions des étudiantes ($P<0,001$).

Test de Tukey 1

Le test de Tukey pour les modes de traduction (Tableau D. 7 en annexe) met en évidence deux écarts significatifs : d’une part entre la TH et la PE de *DeepL* ($P<0,001$) et, d’autre part, entre la TH et la PE de *Google* ($P<0,01$). Pour les erreurs de grammaire et de syntaxe, la TH présente une moyenne significativement supérieure (+1,55) à celle en PE de *DeepL* et également supérieure (+1,48) à celle en PE de *Google*. Nous en déduisons qu’en TH, les étudiantes ont commis, en moyenne, plus d’erreurs de grammaire et de syntaxe par rapport à la PE de TA neuronale, quel que soit le TS.

Texte source

Il ressort également du Tableau D. 6 en annexe que la variable « texte source » a influencé le nombre d'erreurs de grammaire et de syntaxe relevées de notre corpus ($P < 0,001$).

Test de Tukey 2

Le test de Tukey pour les textes sources (Tableau D. 8 en annexe) montre également deux écarts significatifs. Premièrement, le texte 2 présente une moyenne supérieure à celle du texte 1 de 2,08 ($P < 0,001$) et deuxièmement, le texte 3 présente une moyenne supérieure au texte 1 de 1,64 ($P < 0,001$). Dans notre expérience, les productions finales du texte source 1 (article de presse) contiennent moins d'erreurs de grammaire et de syntaxe que dans les deux autres TS, quel que soit le mode de traduction.

5.3.2.2.1.2 Orthographe

Le test ANOVA pour les fautes d'orthographe ne révèle pas de résultats significatifs (Tableau D. 9 en annexe).

5.3.2.2.1.3 Typographie et ponctuation

Interaction mode–texte

Il n'y a pas d'interaction significative entre les modes de traduction et les textes sources pour cette catégorie d'erreurs (Tableau D. 10 en annexe).

Mode de traduction

Grâce au test ANOVA (Tableau D. 10), nous constatons une influence significative du mode de traduction sur les erreurs de typographie et de ponctuation ($P < 0,001$).

Test de Tukey

En effet, d'après les résultats du test de Tukey (Tableau D. 11 en annexe), nous remarquons un écart significatif entre la PE de *Google* et la PE de *DeepL* ($P < 0,05$), ainsi qu'entre la TH et la PE de *DeepL* ($P < 0,001$). Quel que soit le TS, la PE de *DeepL* a permis d'obtenir des productions qui contiennent moins d'erreurs de typographie et de ponctuation que la PE de *Google* et que la TH.

Texte source

La variable « texte source » est, elle aussi, associée au nombre d'erreurs de typographie et aux problèmes de ponctuation ($P < 0,001$).

Test de Tukey

Les résultats du test de Tukey pour les TS (Tableau D. 12 en annexe) indiquent des écarts significatifs pour les trois TS : le texte 2 présente une moyenne supérieure à celle du texte 1 ($P < 0,001$), ainsi qu'à celle du texte 3 ($P < 0,05$). Enfin, le texte 3 a une moyenne supérieure

au texte 1 ($P < 0,05$). Ainsi, les productions du texte 2 (texte économique) sont celles qui contiennent le plus d'erreurs de typographie et de problèmes de ponctuation, et ce, quel que soit le mode de traduction.

5.3.2.2.1.4 Lexique

Interaction mode–texte

Il n'y a pas d'interaction significative entre les modes de traduction et les textes sources pour cette catégorie d'erreurs (Tableau D. 13 en annexe).

Mode de traduction

Il apparaît que le mode de traduction n'a pas eu d'influence significative sur le nombre d'erreurs de lexique (Tableau D. 13).

Texte source

Cependant, nous observons un effet significatif de la variable « texte source » sur le nombre d'erreurs de lexique ($P < 0,001$) (Tableau D. 13).

Test de Tukey

Les résultats du test de Tukey pour les TS (Tableau D. 14 en annexe) montrent des écarts significatifs pour les trois TS. Le texte 3 a une moyenne supérieure au texte 1 ($P < 0,001$) et au texte 2 ($P < 0,001$), tandis que le texte 2 présente une moyenne significativement inférieure au texte 1 ($P < 0,05$). Le texte 3 (vulgarisation scientifique) est celui qui a engendré le plus grand nombre de problèmes de lexique en moyenne, quel que soit le mode de traduction.

5.3.2.2.1.5 Incompréhension

Interaction mode–texte

Il n'y a toujours pas d'interaction significative entre les modes de traduction et les textes sources pour cette catégorie d'erreurs (Tableau D. 15 en annexe).

Mode de traduction

Le mode de traduction n'a pas eu d'influence significative sur le nombre d'erreurs de la catégorie « incompréhension » (Tableau D. 15).

Texte source

En revanche, il y a une influence significative de la variable « texte source » sur le nombre de problèmes d'incompréhension ($P < 0,001$) (Tableau D. 15).

Test de Tukey

Le test de Tukey pour les TS (Tableau D. 16 en annexe) nous permet de mettre au jour deux écarts significatifs. Le texte 3 présente une moyenne supérieure au texte 1 ($P < 0,001$), ainsi qu'au texte 2 ($P < 0,001$). Une fois de plus, c'est le texte 3 (vulgarisation scientifique) qui a généré le plus grand nombre d'erreurs d'incompréhension, quel que soit le mode de traduction.

5.3.2.2.1.6 Style

Interaction mode–texte

Les résultats du test ANOVA révèlent l'existence d'une interaction entre les modes de traduction et les textes sources pour la catégorie d'erreurs de style ($F = 3,80$; $dfN = 4$; $dfD = 63$; $P = 0,007812$) voir Tableau D. 17 en annexe. Cela indique que le mode de traduction a eu une influence sur les erreurs de style, mais uniquement en interaction avec le texte source. Étant donné la présence d'une interaction statistiquement significative, il n'est pas pertinent d'analyser séparément les effets des variables indépendantes.

5.3.2.2.1.7 Cohérence

Interaction mode–texte

Il n'y a pas d'interaction significative entre les modes de traduction et les textes sources pour cette catégorie d'erreurs (Tableau D. 18 en annexe).

Mode de traduction

Le mode de traduction n'a pas eu d'influence significative sur le nombre de problèmes de cohérence (Tableau D. 18).

Texte source

Toutefois, la variable « texte source » est associée significativement au nombre de problèmes de cohérence ($P < 0,001$) (Tableau D. 18).

Test de Tukey

D'après les résultats du test de Tukey pour les TS (Tableau D. 19 en annexe), nous remarquons des écarts significatifs pour les trois TS : Le texte 3 présente une moyenne inférieure au texte 1 ($P < 0,01$) et au texte 2 ($P < 0,001$), tandis que le texte 2 présente une moyenne significativement inférieure au texte 1 ($P < 0,001$). Nous pouvons en conclure que le texte 2 (texte économique) est le TS qui a engendré le plus grand nombre de problèmes de cohérence, quel que soit le mode de traduction.

5.3.2.2.3 Catégories en phase d'adéquation

Analysons à présent les résultats obtenus pour les catégories en phase d'adéquation.

5.3.2.2.3.1 Contresens et non-sens

Interaction mode–texte

Le test ANOVA révèle la présence d'une interaction significative entre les modes de traduction et les textes sources pour les contresens et les non-sens ($F= 3,57$; $dfN= 4$; $dfD= 54,84$; $P= 0,011668$), voir Tableau D. 20 en annexe.

Dans cette expérience, le mode de traduction a eu une influence sur la présence de contresens et de non-sens, mais uniquement en interaction avec le texte source. Cela signifie que les différences observées entre deux modes de traduction varient d'un texte à l'autre. Comme nous sommes en présence d'une interaction statistiquement significative, il n'est pas pertinent d'analyser séparément les effets de nos variables indépendantes.

5.3.2.2.3.2 Glissement de sens

Interaction mode–texte

Concernant les glissements de sens, le test statistique nous permet de mettre au jour une interaction entre les modes de traduction et les textes sources ($F= 2,66$; $dfN= 4$; $dfD= 63$; $P= 0,04068$), voir Tableau D. 21 en annexe.

Nous pouvons affirmer que les occurrences de glissement de sens dans les productions dépendent significativement de l'interaction entre nos deux variables indépendantes (modes de traduction et textes sources). Une fois de plus, nous n'approfondirons pas l'analyse étant donné qu'il existe une interaction statistiquement significative entre nos deux variables indépendantes.

5.3.2.2.3.3 Omission

Interaction mode–texte

À nouveau, les résultats confirment l'existence d'une interaction statistiquement significative entre les modes de traduction et les textes sources pour les omissions ($F= 7$; $dfN= 4$; $dfD= 62,53$; $P= 0,00001005$), voir Tableau D. 22 en annexe.

Dans notre corpus, le nombre d'omissions varie significativement en fonction de l'interaction entre les modes de traduction et les textes sources. Il n'est dès lors pas pertinent d'analyser les effets principaux isolément.

5.3.2.2.3.4 Ajout

Interaction mode–texte

Pour les ajouts, le test ANOVA ne révèle pas d'interaction entre les modes de traduction et les TS (Tableau D. 23 en annexe).

Mode de traduction

Le mode de traduction a eu une influence significative sur le nombre d'ajouts ($P < 0,05$).

Test de Tukey

D'après les résultats du test de Tukey pour les modes de traduction (Tableau D. 24 en annexe), nous constatons un écart significatif entre la TH et la PE de *Google* ($P < 0,01$), même si celui-ci est très faible (0,38). Nous pouvons tirer comme conclusion qu'en TH, les étudiantes ont commis, en moyenne, très légèrement plus d'ajouts qu'en PE de *Google*.

Texte source

D'après les résultats du test ANOVA, la variable « texte source » n'est pas associée au nombre d'ajouts (Tableau D. 23).

5.3.2.2.3.5 Calque

Interaction mode–texte

Pour les erreurs de calques, le test ANOVA ne permet pas de mettre au jour d'interaction significative entre les modes de traduction et les textes sources qui influencerait les moyennes (Tableau D. 25 en annexe).

Mode de traduction

En revanche, la variable « mode de traduction » a eu une influence très significative sur le nombre de calques recensés dans les productions ($P < 0,001$), voir Tableau D. 25.

Test de Tukey 1

Les résultats du test de Tukey pour les modes de traduction (Tableau D. 26 en annexe) montrent que les moyennes sont statistiquement différentes entre la TH et la PE de *DeepL* ($P < 0,01$), ainsi qu'entre la TH et la PE de *Google* ($P < 0,001$). La TH présente une moyenne inférieure à celle de la PE de *DeepL* de 2,42 et à celle de la PE de *Google* de 2,59.

Ainsi, tout comme dans le prétest, nous pouvons affirmer que les étudiantes ont commis, en moyenne, davantage de calques fautifs en PE, que ce soit de *DeepL* ou de *Google*, qu'elles n'en ont commis en TH, et ce, quel que soit le TS. Ces résultats nous amènent à la conclusion que la PE de TA neuronale génère davantage de calques fautifs dans le chef des étudiantes que la TH. Ce constat confirme une nouvelle fois l'un de nos postulats de départ puisque nous avons supposé que les PE contiendraient, en moyenne, un plus grand nombre de calques fautifs que les TH.

Texte source

De même, les résultats montrent une influence très significative entre le texte source et le nombre de calques ($P < 0,001$), voir Tableau D. 25.

Test de Tukey 2

Le test de Tukey pour les TS (tableau D. 27 en annexe) révèle un écart significatif entre le texte 3 et le texte 1 ($P < 0,001$), ainsi qu'entre le texte 3 et le texte 2 ($P < 0,001$). Nous pouvons en déduire que le texte source qui a comptabilisé, en moyenne, le plus de calques fautifs est le texte 3 (vulgarisation scientifique), car il présente une moyenne supérieure à celle du texte 1 et à celle du texte 2 (respectivement de 7,62 et de 6,00).

5.3.2.2.3.6 Vocabulaire

Interaction mode–texte

Pour la catégorie « vocabulaire », les résultats de l'ANOVA indiquent l'absence d'interaction significative entre les modes de traduction et les TS (Tableau D. 28 en annexe).

Mode de traduction

Le Tableau D. 28 montre que la variable « mode de traduction » est associée au nombre d'erreurs de vocabulaire ($P < 0,05$).

Test de Tukey 1

Le test de Tukey pour les modes de traduction (Tableau D. 29 en annexe) révèle une différence significative au niveau des erreurs de vocabulaire entre la PE de *Google* et la PE de *DeepL* ($P < 0,05$), ainsi qu'entre la TH et la PE de *DeepL* ($P < 0,05$). Ces écarts significatifs peuvent s'interpréter de la manière suivante : les étudiantes ont commis, en moyenne, moins d'erreurs de vocabulaire en PE de *DeepL* par rapport aux deux autres modes de traduction.

Texte source

Il apparaît également dans le Tableau D. 28, que la variable « texte source » a eu une influence significative sur le nombre d'erreurs de vocabulaire ($P < 0,01$).

Test de Tukey 2

Les résultats du test de Tukey pour les TS (Tableau D. 30 en annexe) nous permettent de constater un écart significatif ($P < 0,01$) entre le texte 3 et le texte 2 de 1,12.

5.3.2.2.3.7 Autre

Le nombre d'erreurs relevées dans cette catégorie est trop faible pour pouvoir appliquer les tests statistiques et obtenir des résultats statistiquement significatifs.

5.3.2.3 Tableaux récapitulatifs

5.3.2.3.1 Influence du mode de traduction sur le nombre d’erreurs

En résumé, dans ce test, la PE de *DeepL* a généré significativement moins d’erreurs d’acceptabilité que la PE de *Google* et que la TH. Lorsque l’on considère les catégories d’erreurs isolément, la PE de *DeepL* a généré moins d’erreurs de grammaire et de syntaxe que la TH, ainsi que moins d’erreurs de vocabulaire et de problèmes de typographie et de ponctuation que les deux autres modes de traduction. En ce qui concerne la TH, nous pouvons affirmer que les productions contiennent, d’une part, davantage d’erreurs de grammaire et de syntaxe et, d’autre part, moins de calques comparativement à la PE de *DeepL* et à la PE de *Google*.

Influence mode de traduction (nombre d’erreurs)	TH		PE de <i>Google</i>		PE de <i>DeepL</i>	
	vs PE <i>Google</i>	vs PE <i>DeepL</i>	vs TH	vs PE <i>DeepL</i>	vs TH	vs PE <i>Google</i>
Acceptabilité	/	+	/	+	-	-
Adéquation	/	/	/	/	/	/
Catégories acceptabilité						
Grammaire et syntaxe	+	+	-	/	-	/
Fautes d’orthographe	/	/	/	/	/	/
Typo et ponctuation	/	+	/	+	-	-
Lexique	/	/	/	/	/	/
Incompréhension	/	/	/	/	/	/
Style	/	/	/	/	/	/
Cohérence	/	/	/	/	/	/
Catégories adéquation						
Contresens et non-sens	/	/	/	/	/	/
Glissement de sens	/	/	/	/	/	/
Omission	/	/	/	/	/	/
Ajout	+	/	-	/	/	/
Calque	-	-	+	/	+	/
Vocabulaire	/	+	/	+	-	-
Autre	/	/	/	/	/	/

Légende
- : Effet négatif – le mode de traduction a eu pour effet de générer significativement moins d’erreurs que l’autre mode de traduction.
+ : Effet positif – le mode de traduction a eu pour effet de générer significativement davantage d’erreurs que l’autre mode de traduction.
/ : résultats non significatifs ou présence d’une interaction significative entre les modes de traduction et les textes sources. Nous n’avons pas pu conclure à l’existence d’une relation entre le mode de traduction et le nombre d’erreurs.

Tableau 38 : Influence du mode de traduction sur le nombre d’erreurs – Test

5.3.2.3.2 Influence du texte source sur le nombre d'erreurs

Influence texte source (nombre d'erreurs)	Texte 1 (presse)		Texte 2 (économique)		Texte 3 (scientifique)	
	vs texte 2	vs texte 3	vs texte 1	vs texte 3	vs texte 1	vs texte 2
Acceptabilité	-	-	+	-	+	+
Adéquation	/	/	/	/	/	/
Catégories acceptabilité						
Grammaire et syntaxe	-	-	+	/	+	/
Fautes d'orthographe	/	/	/	/	/	/
Typographie et ponctuation	-	-	+	+	+	-
Lexique	-	-	+	-	+	+
Incompréhension	/	-	/	-	+	+
Style	/	/	/	/	/	/
Cohérence	-	+	+	+	-	-
Catégories adéquation						
Contresens et non-sens	/	/	/	/	/	/
Glissement de sens	/	/	/	/	/	/
Omission	/	/	/	/	/	/
Ajout	/	/	/	/	/	/
Calque	/	-	/	-	+	+
Vocabulaire	/	/	/	-	/	+
Autre	/	/	/	/	/	/
Légende						
- : Effet négatif – le texte source a eu pour effet de générer significativement moins d'erreurs que l'autre texte source.						
+ : Effet positif – le texte source a eu pour effet de générer significativement davantage d'erreurs que l'autre texte source.						
/ : résultats non significatifs ou présence d'une interaction significative entre les modes de traduction et les textes sources. Nous n'avons pas pu conclure à l'existence d'une relation entre le texte source et le nombre d'erreurs.						

Tableau 39 : Influence du texte source sur le nombre d'erreurs – Test

5.3.2.3.3 Influence sur les notes

Concernant l'influence de la variable « mode de traduction » sur les notes attribuées aux productions, nous pouvons conclure que les productions humaines ont obtenu, en moyenne, des notes plus basses en acceptabilité par rapport aux PE (*DeepL*).

Tandis qu'au niveau de l'influence du texte source sur les notes d'acceptabilité, il apparaît que le texte 3 est celui pour lequel les productions ont obtenu les notes les moins élevées. En revanche, le texte 1 a obtenu les meilleures notes, quel que soit le mode de traduction.

Influence mode de traduction (notes)	TH		PE <i>Google</i>		PE <i>DeepL</i>	
	vs PE <i>Google</i>	vs PE <i>DeepL</i>	vs TH	vs PE <i>DeepL</i>	vs TH	vs PE <i>Google</i>
Acceptabilité	/	-	/	/	+	/
Adéquation	/	/	/	/	/	/
Acceptabilité + Adéquation	/	/	/	/	/	/
Influence texte source (notes)	Texte 1		Texte 2		Texte 3	
	vs texte 2	vs texte 3	vs texte 1	vs texte 3	vs texte 1	vs texte 2
Acceptabilité	+	+	-	+	-	-
Adéquation	/	/	/	/	/	/
Acceptabilité + Adéquation	/	/	/	/	/	/
Légende						
- : Effet négatif – Le mode de traduction/le texte source a eu pour effet significatif de générer, en moyenne, des notes plus basses par rapport à l'autre mode de traduction/texte source.						
+ : Effet positif – Le mode de traduction/le texte source a eu pour effet significatif de générer, en moyenne, des notes plus élevées par rapport à l'autre mode de traduction/texte source.						
/ : résultats non significatifs ou présence d'une interaction significative entre les modes de traduction et les textes sources. Nous n'avons pas pu conclure à l'existence d'une relation entre le mode de traduction et les notes attribuées aux productions.						

Tableau 40 : Influence sur les notes – Test

5.3.3 Résultats complémentaires

5.3.3.1 Étude du processus : Productivité

Comme dans le prétest, nous nous sommes rapidement penchée sur le processus même de post-édition. Nous présentons en premier lieu les résultats relatifs au taux de productivité en fonction des différents modes de traduction (par texte et par étudiant). Et en deuxième lieu, nous présentons les résultats relatifs à l'effort moyen de PE en fonction du moteur de TA et du texte source.

5.3.3.1.1 Productivité par texte

Le Tableau 41 regroupe les gains de productivité en fonction du texte source.

À nouveau, pour l'ensemble des productions, nous avons calculé le temps moyen de traduction/PE en secondes par mot du texte source et comparé ces résultats pour mesurer les effets sur le taux moyen de productivité.

	TH s/mot	PE (<i>Google</i>) s/mot	PE (<i>DeepL</i>) s/mot	PE (<i>Google</i>) vs TH	PE (<i>DeepL</i>) vs TH	PE (<i>DeepL</i>) vs PE (<i>Google</i>)
Texte 1	10,99	8,99	7,66	18 %	30 %	15 %
Texte 2	12,01	5,70	7,21	53 %	40 %	-27 %
Texte 3	10,35	7,49	6,72	28 %	35 %	10 %

Tableau 41 : Gains de productivité par texte – Test

Gain de productivité

Tout comme pour le prétest et sans grande surprise, on observe sur le Tableau 41 systématiquement pour les trois textes un gain de productivité en PE par rapport à la TH. Les gains de productivité les plus élevés ont été enregistrés pour le texte 2 (+ 53 % en PE de *Google* et + 40 % en PE de *DeepL*). C'est déjà ce que nous avons constaté lors du prétest. Remarquons toutefois que pour le texte 2, contrairement aux deux autres textes, les étudiantes ont été moins productives en PE de *DeepL* par rapport à la PE de *Google* (-27 %). Une fois encore, nous pouvons conclure que la PE (que ce soit de *Google* ou de *DeepL*) a permis aux étudiantes d'augmenter leur productivité par rapport à la TH, autrement dit, de traduire plus en moins de temps. À titre de comparaison, Martikainen et Mestivier (2020) ont relevé chez les étudiantes un gain de productivité entre la PE de TA neuronale et la TH (anglais-français) situé entre 20 % et 30 %.

5.3.3.1.2 Productivité par étudiante

Intéressons-nous à présent au taux de productivité par étudiante. Les résultats sont repris dans le Tableau 42. Nous avons mis en évidence les valeurs maximale et minimale pour chaque mode de traduction.

Remarque : les tableaux et figures qui suivent ne reprennent pas les productions de l'étudiante ET002. Nous avons, en effet, dû exclure les trois productions de cette participante, car les données concernant non seulement la durée de PE et de traduction, mais aussi l'effort de PE mesurés pour ces trois tâches, étaient anormalement peu élevées, alors qu'il y a bel et bien eu un certain nombre de modifications apportées au texte brut. Nous supposons que l'étudiante n'a pas travaillé directement dans l'outil *Matecat* et qu'elle a certainement utilisé un traitement de texte avant de copier-coller les différents segments de sa production dans *Matecat* et de valider ses tâches.

Tableau 42 : Gains de productivité par étudiante – Test

	TH (s/mot)	PE (Google) (s/mot)	PE (DeepL) (s/mot)	PE (Google) vs TH	PE (DeepL) vs TH	PE (DeepL) vs PE (Google)
ET001	10,32	7,86	8,28	+24 %	+20 %	-5 %
ET003	13,58	8,48	4,23	+38 %	+69 %	+50 %
ET004	7,20	3,88	5,58	+46 %	+22 %	-44 %
ET005	8,03	12,52	7,55	-56 %	+6 %	+40 %
ET006	11,11	5,56	3,77	+50 %	+66 %	+32 %
ET008	12,00	<u>20,71</u>	9,81	-73 %	+18 %	+53 %
ET009	12,65	10,70	6,60	+15 %	+48 %	+38 %
ET011	<u>6,19</u>	4,38	<u>1,56</u>	+29 %	+75 %	+64 %
ET012	<u>16,41</u>	7,70	9,17	+53 %	+44 %	-19 %
ET013	7,61	10,14	9,07	-33 %	-19 %	+11 %
ET014	10,45	9,89	8,99	+5 %	+14 %	+9 %
ET015	12,94	7,38	8,61	+43 %	+33 %	-17 %
ET016	9,52	4,94	7,61	+48 %	+20 %	-54 %
ET017	16,20	8,42	6,92	+48 %	+57 %	+18 %
ET018	14,10	4,08	8,82	+71 %	+37 %	-116 %
ET019	13,82	6,95	4,81	+50 %	+65 %	+31 %
ET023	9,19	<u>3,83</u>	3,31	+58 %	+64 %	+14 %
ET024	8,69	8,53	9,13	+2 %	-5 %	-7 %
ET025	13,66	7,85	<u>18,71</u>	+43 %	-37 %	-138 %
ET026	9,62	4,94	3,97	+49 %	+59 %	+20 %
ET027	12,67	6,86	10,22	+46 %	+19 %	-49 %
ET028	7,64	4,73	6,35	+38 %	+17 %	-34 %
ET030	13,49	7,67	4,99	+43 %	+63 %	-46 %

Pour 20 participantes sur 23 (soit 87 %), la PE (que ce soit de *Google* ou de *DeepL*) a engendré un gain de productivité par rapport à la TH. Le gain de productivité moyen s'élève à 28 % en PE de *Google* et à 33 % en PE de *DeepL*. Par ailleurs, pour 12 étudiantes sur 23

(soit 52 %), on constate un meilleur taux de productivité en PE de *DeepL* par rapport à la PE de *Google*.

5.3.3.1.3 Productivité et qualité

- ❖ Existe-t-il une corrélation entre la productivité d'une étudiante et la note globale attribuée à sa production ?

Nous avons une fois de plus vérifié l'existence d'une corrélation entre le taux de productivité (mesuré en secondes par mot du texte source) et la note obtenue par l'étudiante pour sa production. Pour ce faire, nous avons généré une figure en nuage de points, par mode de traduction, sur laquelle sont reprises les productions des trois TS.

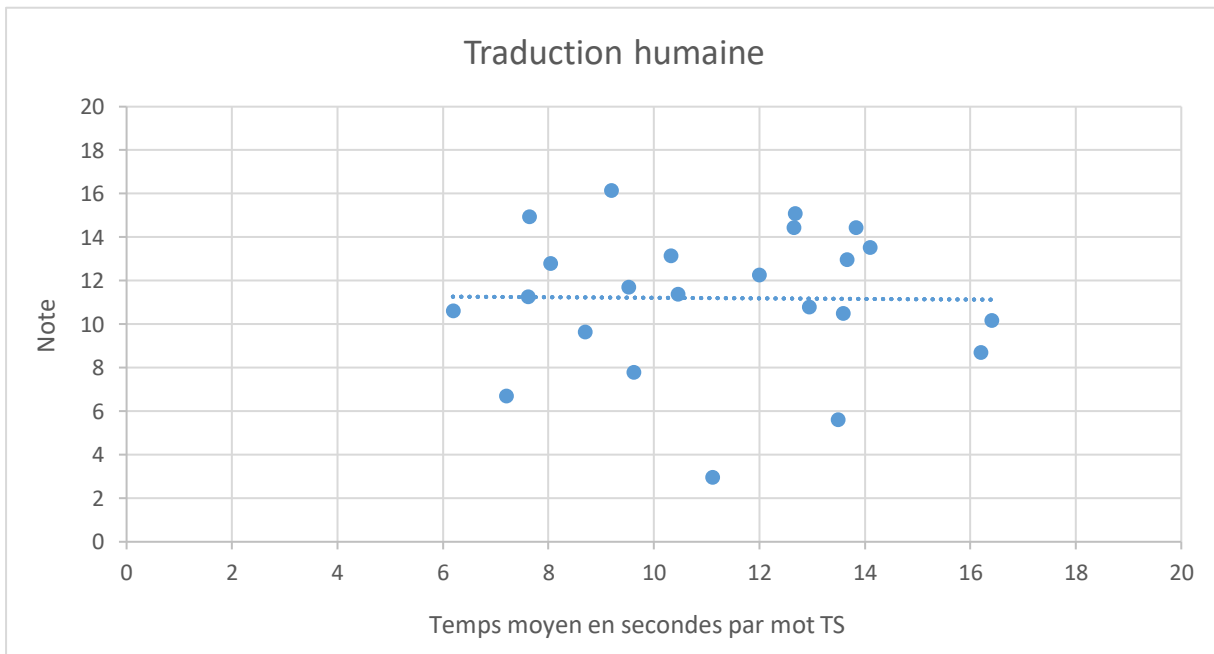
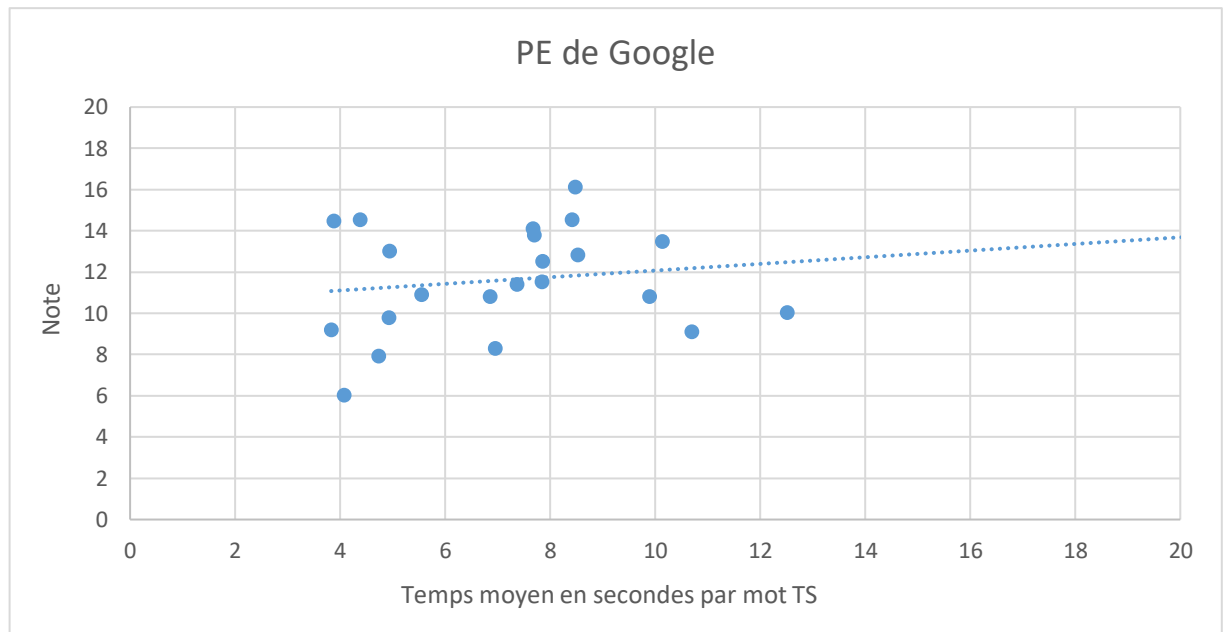
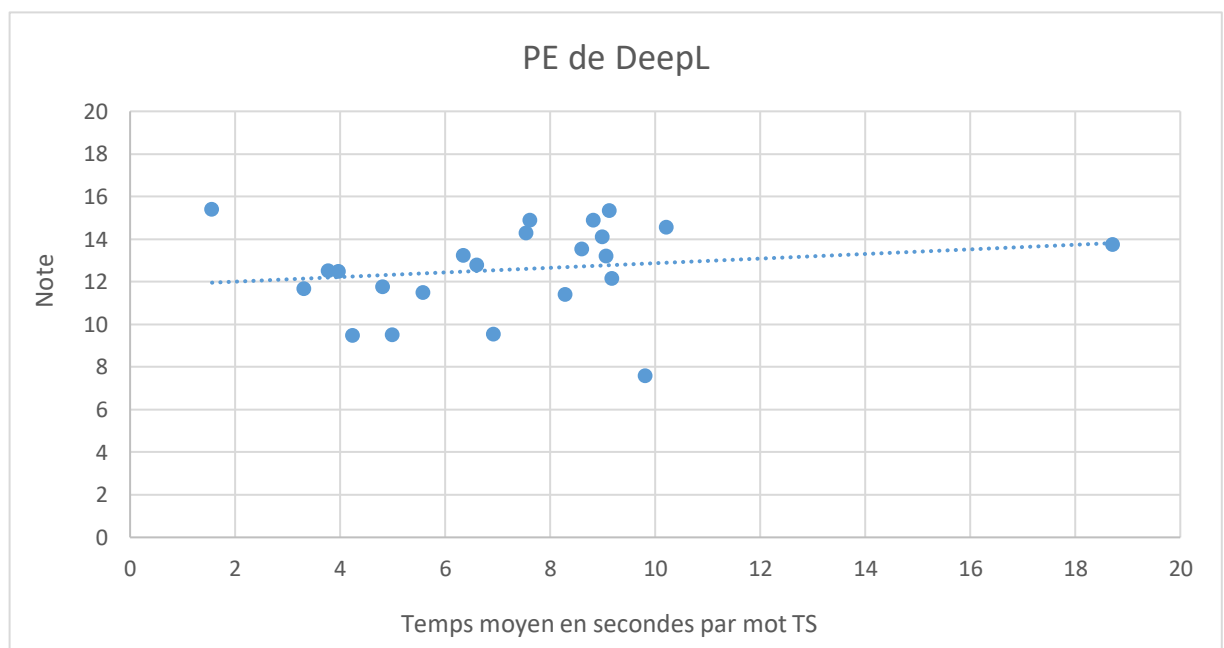


Figure 58 : Productivité-Qualité – Traduction humaine – Test

Figure 59 : Productivité-Qualité – PE de *Google* – TestFigure 60 : Productivité-Qualité – PE de *DeepL* – Test

Pour cette expérience, les résultats du test de Pearson¹⁰⁶ ne nous permettent pas de mettre au jour une corrélation entre la productivité et la qualité, ni en TH ($r = -0,012$ [IC 95 % : $-0,422 - 0,402$]), ni en PE de *Google* ($r = 0,23$ [IC 95 % : $-0,200 - 0,588$]), ni en PE de *DeepL* ($r = 0,18$ [IC 95 % : $-0,252 - 0,550$]). Au vu de ces résultats et contrairement à ce que

¹⁰⁶ Se référer au chapitre 4 – 4.4.3.1.3 Productivité et qualité – pour une description plus détaillée du calcul et de l'interprétation du coefficient de corrélation de Pearson.

l'intuition pourrait laisser penser, nous ne pouvons pas affirmer que plus l'étudiante a pris du temps pour traduire/post-éditer, meilleure est la note qu'elle a obtenue.

5.3.3.2 Étude du processus : Effort de PE

Le Tableau 43 reprend l'ensemble des efforts de PE (%)¹⁰⁷ en fonction du moteur de TA (*Google* ou *DeepL*) et en fonction du texte source.

5.3.3.2.1 Effort de PE et moteur de TA

Texte 1		Texte 2		Texte 3	
PE de Google	PE de DeepL	PE de Google	PE de DeepL	PE de Google	PE de DeepL
27 %	19 %	21 %	15 %	40 %	21 %
27 %	24 %	11 %	29 %	30 %	26 %
45 %	36 %	15 %	12 %	28 %	26 %
20 %	17 %	20 %	19 %	27 %	17 %
38 %	38 %	28 %	22 %	14 %	14 %
25 %	50 %	26 %	16 %	39 %	22 %
19 %	39 %	31 %	26 %	31 %	10 %
42 %			21 %	26 %	9 %
			29 %	14 %	

Tableau 43 : Effort moyen de PE par production – Test

¹⁰⁷ Se référer à nouveau au chapitre 4 – 4.4.3.2 Étude du processus : Effort de PE – pour une description du calcul de l'effort de PE.

<i>Texte source</i>	Système de TA	Min.	Moyenne	Médiane	Max.	Écart-type
<i>Texte 1</i>	<i>PE de Google</i>	19 %	30 %	32 %	45 %	0,100
	<i>PE de DeepL</i>	17 %	32 %	36 %	50 %	0,121
<i>Texte 2</i>	<i>PE de Google</i>	11 %	22 %	21 %	31 %	0,072
	<i>PE de DeepL</i>	12 %	21 %	21 %	29 %	0,061
<i>Texte 3</i>	<i>PE de Google</i>	14 %	28 %	28 %	40 %	0,092
	<i>PE de DeepL</i>	9 %	18 %	19 %	26 %	0,067

Tableau 44 : Détails effort moyen de PE en pourcentage – Test

Grâce aux données du Tableau 43, nous avons pu calculer la valeur minimale, la moyenne, la médiane, la valeur maximale et l'écart-type en fonction du moteur de TA et du TS (voir Tableau 44). Les valeurs obtenues pour les textes 1 et 2 (notamment la valeur moyenne et la valeur médiane) sont très similaires pour les deux systèmes de TA (*Google* et *DeepL*). En revanche, pour le texte 3, on remarque une nette baisse du taux moyen et du taux médian en PE de *DeepL*. Manifestement, dans le cas du texte de vulgarisation scientifique (texte 3), la TA générée par *DeepL* a été moins modifiée par les étudiantes que la TA générée par *Google*. Ce constat laisse supposer que la TA brute générée par *DeepL* était de meilleure qualité que la TA brute générée par *Google*. À cet égard, la différence de près de 8 points observée entre le score BLEU calculé pour la TA *DeepL* et celui calculé pour la TA *Google* (cf. Tableau 1) vient étayer notre supposition.

En résumé, pour les textes 1 et 2, nous n'observons pas de différences majeures au niveau du taux d'effort moyen entre la PE de *Google* et la PE de *DeepL*. Pour le texte 3, en revanche, l'effort moyen de PE fut plus élevé en PE de *Google* qu'en PE de *DeepL*. Dans la lignée des conclusions que nous avons tirées pour le prétest, c'est le texte 1 qui a généré, en moyenne, les taux d'effort de PE les plus élevés.

En comparant ces résultats aux résultats obtenus dans le prétest, il apparaît que les étudiantes de 2^e master ont apporté davantage de modifications aux TA brutes que les étudiantes de 3^e bachelier, et ce, quel que soit le TS. Mais comment interpréter ce constat ? Les étudiantes moins chevronnées se satisferaient-elles plus rapidement de ce que leur propose la machine que les étudiantes plus expérimentées ? Les étudiantes les moins aguerries feraient-elles davantage confiance à la machine ou bien oseraient-elles moins modifier les propositions de la TA comparativement aux étudiantes en fin de parcours universitaire ?

5.3.3.2.2 Effort de PE et qualité

À nouveau, nous avons voulu vérifier l'existence d'une corrélation entre cet effort moyen de PE et la note globale obtenue par l'étudiante (Figure 61 et Figure 62). Nous présumons toujours que plus ce pourcentage d'effort de PE est élevé, meilleure est la qualité de la production et, donc, plus la note obtenue est élevée.

- ❖ Existe-t-il une corrélation entre l'effort moyen de PE et la note globale obtenue par l'étudiante ?

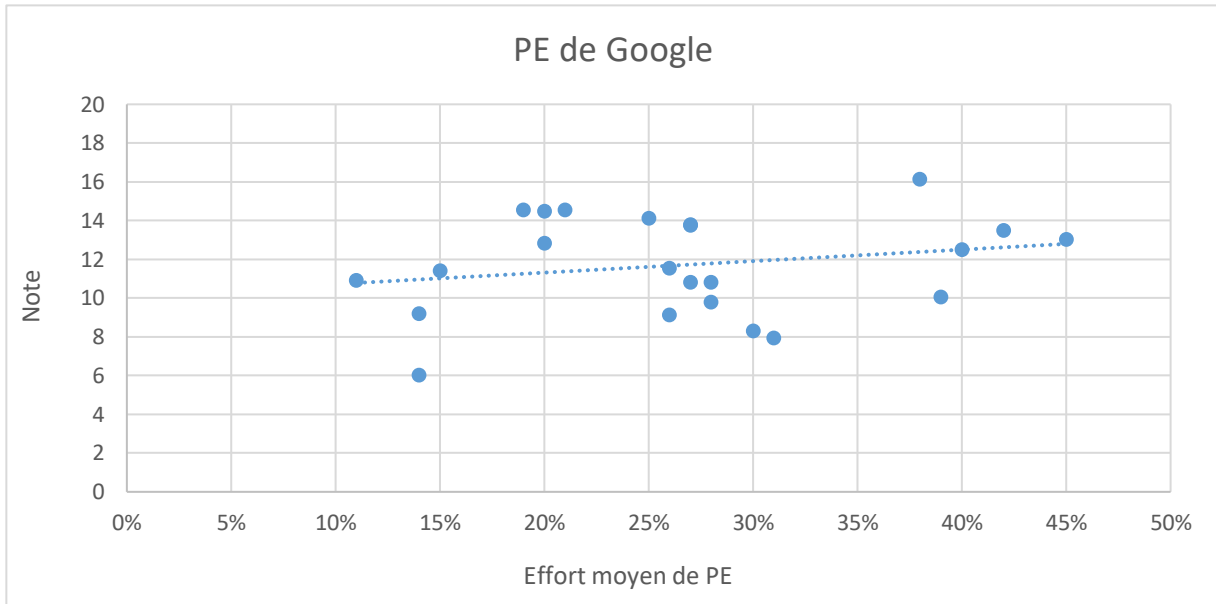


Figure 61 : Effort de PE et qualité – PE de *Google*

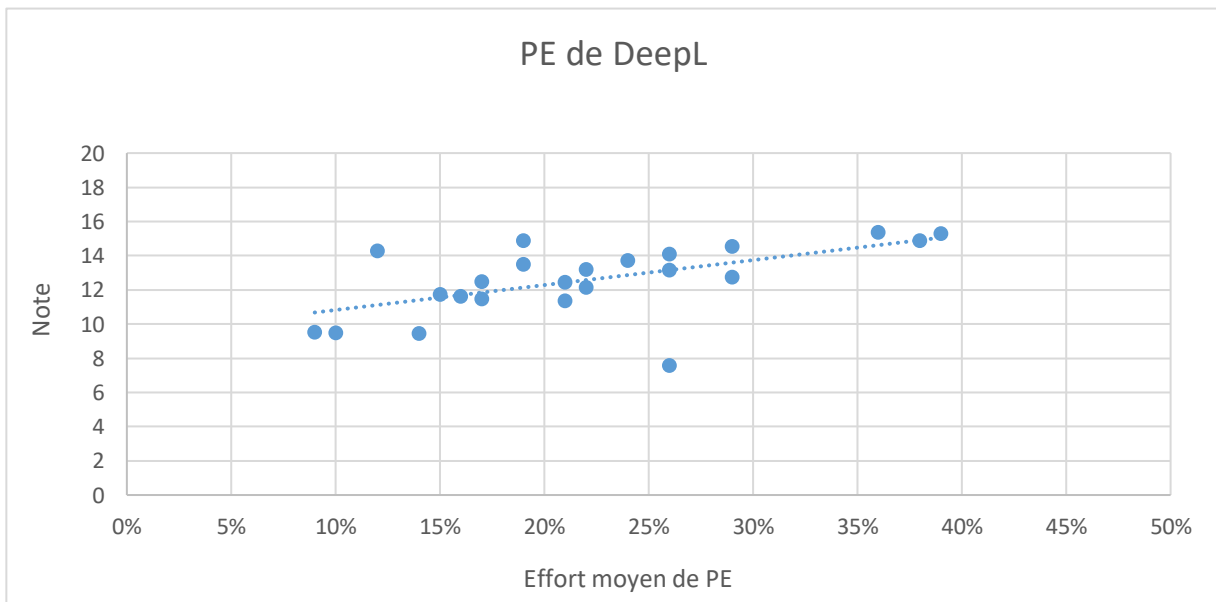


Figure 62 : Effort de PE et qualité – PE de *DeepL*

Pour répondre à cette question, nous avons à nouveau calculé le coefficient de corrélation de Pearson. Les résultats de ce test ne révèlent pas de corrélation entre l'effort moyen de PE avec *Google* et la note obtenue ($r = 0,22$ [IC 95 % : -0.214 – 0.578]). En revanche, les

résultats confirment l'existence d'une corrélation positive modérée entre l'effort de PE avec *DeepL* et la qualité des productions ($r = 0,58$ [IC 95 % : 0,222 – 0,801]). Cette corrélation indique que plus l'étudiante a apporté des modifications à la TA brute de *DeepL*, plus la qualité de sa production est élevée, et ce, quel que soit le texte source.

Soulignons que la majeure partie des participantes était des post-éditrices ayant peu d'expérience en la matière, il se peut donc que les modifications qu'elles ont apportées à la TA brute ne soient pas toujours pertinentes (voire nécessaires).

À nouveau, il serait pertinent d'approfondir toutes ces observations à l'aide d'un modèle de régression mixte.

5.3.3.3 Recueil d'impressions : post-édition ou traduction ?

Alors qu'il semblerait que la lectrice finale ne perçoive généralement pas de différences entre langue originale et langue traduite, Daems, De Clercq et al. (2017) ont cherché à savoir si l'humain était capable de distinguer un texte post-édité d'une TH. Pour ce faire, elles ont demandé à 195 étudiantes en traduction de l'Université de Gand de spécifier la nature des textes qui leur ont été présentés, en leur soumettant, soit une combinaison de deux textes, soit un seul texte. D'après les estimations qu'elles ont pu recueillir, les chercheuses ont conclu que l'humain n'avait pas été capable de faire une distinction : « [t]he user perception study showed that the participants were unable to distinguish between HT and PE texts of publishable quality. If anything, they more often incorrectly labelled PE texts as HTs than the other way around » (p. 101).

Screen (2019) s'est également intéressé à la réception de textes post-édités comparativement à des traductions pour la paire de langues anglais-gallois. Les résultats de son étude indiquent que la PE n'aurait pas d'influence négative sur le produit final d'un point de vue de la réceptrice finale, que ce soit en matière de lisibilité et d'intelligibilité : « post-edited texts, given these results, are perceived by end users to be just as readable and comprehensible as translated ones. [...] it [...] appears that post-edited texts are received just as well as translated ones » (*ibid.*, p. 147).

Aussi, dans notre cas, il nous a semblé judicieux de recueillir les impressions et intuitions des évaluatrices en leur demandant de compléter un tableau (Annexe F) en toute fin d'évaluation afin qu'elles puissent juger la nature de chaque production : s'agit-il d'une TH ou d'une PE ? De plus, pour aller un peu plus loin et pour tenter de comprendre les suppositions, nous leur avons demandé également de noter si certains éléments avaient permis d'orienter leur choix.

Le Tableau 45 reprend le taux d'estimation correcte par évaluatrice, ainsi que le taux global pour l'ensemble des cas à estimer. Notons que nous n'avons malheureusement pas recueilli les impressions des évaluatrices lors du prétest 2018.

	Texte 1	Texte 2	Texte 3	Taux global
--	---------	---------	---------	-------------

	Eval 1	Eval 2	Eval 1	Eval 2	Eval 1	Eval 2	
TH	6/9	7/9	7/8	8/8	3/7	6/7	37/48
PE	8/17	9/17	9/18	11/18	10/19	14/19	61/108
Total	14/26 54 %	16/26 61,5 %	16/26 61,5 %	19/26 73 %	13/26 50 %	20/26 77 %	98/156 63 %

Tableau 45 : Recueil impressions TH/PE

Dans l'ensemble, les évaluateurices nous ont avoué avoir rencontré des difficultés à différencier clairement les PE des TH. Pour certaines productions, leur hésitation entre les deux modes de traduction était tellement grande qu'elles sont difficilement parvenues à trancher. Les résultats chiffrés de ce recueil d'impressions montrent toutefois que toutes les évaluateurices ont estimé correctement la nature d'au moins 50 % des productions. Ainsi, sur l'ensemble des estimations – soit 156 –, 98 se sont révélées exactes, soit un taux d'estimation correcte de 63 %. Néanmoins, ces résultats signifient que, dans bon nombre de cas, tout de même, les évaluateurices ont jugé qu'il s'agissait d'une PE alors qu'il s'agissait en réalité d'une TH et vice-versa. Contrairement aux résultats de Daems, De Clercq et al. (2017), dans notre expérience, les évaluateurices sont mieux parvenues à repérer les TH (37/48) que les PE (61/108) ; beaucoup de PE ont donc été prises pour des TH. Aussi, tout comme Daems, De Clercq et al. (2017), nous estimons que ces résultats ne permettent pas de confirmer l'existence d'un *post-editeuse* puisque les évaluateurices ne sont manifestement pas parvenues à déceler des caractéristiques propres à une langue de PE qui leur auraient permis de distinguer nettement une PE d'une TH.

En dépit de cette conclusion et à toutes fins utiles, nous avons tenu à recenser et à classer les principaux éléments qui, selon elles, ont permis d'orienter leurs choix.

5.3.3.3.1 Facteurs positifs, neutres ou négatifs

Le Tableau 46 reprend les différents facteurs positifs, neutres ou négatifs ayant été rattachés, à tort ou à raison, à la TH ou à la PE par les évaluateurices.

Traduction humaine			Post-édition	
Positif	Neutre	Négatif	Positif	Négatif
Reformulations/remaniements en nombre ou pertinentes (preuves de réflexion) / efforts importants de déverbalisation et de reformulation				Traductions littérales, proches du TS ; forte présence de mots et de tournures calquées (qui "sentent" la traduction)
		Présence d'erreurs de syntaxe, grammaire et d'orthographe (ex. texte 2 : « en terme de » ; « Bangladesh »)		
	La présence de régionalismes (en l'occurrence de belgicisms)			
Traitement efficace des segments plus complexes				
	Tendance au foisonnement			
		Tendance à interpréter certains segments ; tendance à l'extrapolation		

Choix lexicaux faisant preuve d'une belle inventivité				Peu d'inventivité/ solutions peu inventives
		Présence d'erreurs typiquement humaines	Peu d'erreurs de sens/sur le plan sémantique	
		Présence d'erreurs sur le plan sémantique		
		Qualité inégale du texte cible	Homogénéité de la qualité du texte cible	
Bon travail sur l'idiomaticité/cooccurrences				Choix lexicaux bancals ; répétitions et style peu soigné

Tableau 46 : Facteurs positifs, neutres ou négatifs – TH vs PE

5.4 Synthèses et réponses aux hypothèses

Dans cette deuxième étude comparative de TH et de PE qui s'inscrit dans la continuité du prétest mené en 2018, notre objectif était d'explorer les effets de la PE de TA neuronale en contexte d'apprentissage sur la qualité du texte cible pour la paire de langues anglais-français.

5.4.1 Statistiques descriptives et inférentielles

5.4.1.1 Hypothèse HA4

Les résultats statistiques viennent confirmer l'hypothèse HA4 selon laquelle nous postulions que la PE de TA neuronale permet à des étudiantes de produire des textes cibles de qualité comparable à une TH, voire de meilleure qualité. Dans ce test, nous pouvons affirmer que, comparativement à la TH, la PE de TA neuronale n'a pas eu d'influence négative sur la qualité des productions des étudiantes. En effet, il ressort de l'évaluation humaine que les textes post-édités n'ont pas été jugés de moindre qualité que les TH. Nous avons même relevé une moyenne générale des notes en TH systématiquement plus basse par rapport aux deux autres modes de traduction ; cette observation vaut pour les trois TS (Figure 47). Globalement, il ressort de notre expérience que les textes post-édités ont été considérés comme de meilleure qualité que les TH. Preuve en est que si l'on considère l'ensemble des productions (tout TS confondu), la moyenne des notes et la valeur médiane obtenues en PE de TA (que ce soit *Google Traduction* ou *DeepL*) sont supérieures aux valeurs mesurées en TH (Figure 50). De plus, l'évaluation humaine de la qualité a mis en évidence de très hauts taux de réussite en PE de TA neuronale, particulièrement avec le moteur *DeepL* (Figure 48). Nous pouvons ainsi affirmer que, dans cette étude, la probabilité qu'une production soit jugée de bonne qualité et que celle-ci obtienne une note de réussite (supérieure ou égale à 10/20), tout TS confondu, est plus élevée lorsque l'étudiante post-édite de la TA neuronale générée par *DeepL* que lorsqu'elle traduit humainement, ou encore que lorsqu'elle post-édite de la TA neuronale générée par *Google Traduction*. Ce constat corrobore les conclusions du prétest.

Dans ce test, les statistiques inférentielles valident également l'hypothèse HA4, puisque les tests ont révélé que la PE de *DeepL* a eu une influence significative sur la qualité des productions au niveau de la distribution des erreurs, en générant moins d'erreurs d'acceptabilité (erreurs de langue, problèmes de fluidité et de cohérence en langue cible) par rapport à la TH. En revanche, ces statistiques ne nous permettent pas d'inférer une quelconque influence du mode de traduction sur les erreurs d'adéquation, ou encore sur les notes générales combinant la note en phase d'acceptabilité et la note en phase d'adéquation.

5.4.1.2 Hypothèse HA5

L'hypothèse HA5 qui postule qu'en PE de TA neuronale la distribution des erreurs n'est pas comparable à la distribution des erreurs en TH se trouve validée. En effet, nous avons mis au jour que le mode de traduction a eu des effets significatifs sur les catégories d'erreur suivantes : grammaire et syntaxe ; typographie et ponctuation ; ajout ; calque ; vocabulaire.

Nos résultats permettent de vérifier les trois sous-hypothèses de l'hypothèse HA5 (SHA5.1 ; SHA5.2 et SHA5.3).

5.4.1.2.1 Sous-hypothèse SHA5.1

Premièrement, nous pouvons affirmer que, dans cette expérience, nous avons relevé moins d'erreurs d'acceptabilité (erreurs de langue, problèmes de fluidité et de cohérence en langue cible) dans les PE (*DeepL*) par rapport aux productions humaines, comme nous l'avions postulé dans la sous-hypothèse SHA5.1. Logiquement, nous avons également constaté que les productions humaines ont obtenu, en moyenne, des notes moins élevées en acceptabilité par rapport aux PE (*DeepL*). En résumé, la PE de *DeepL* a permis aux étudiantes de commettre moins d'erreurs et d'obtenir de meilleures notes en phase d'acceptabilité comparativement à la TH.

5.4.1.2.2 Sous-hypothèse SHA5.2

Deuxièmement, nous confirmons également que la PE de *DeepL* a généré moins d'erreurs de grammaire et de syntaxe que la TH, ce qui confirme en partie la sous-hypothèse SHA5.2. Nous ne pouvons pas tirer de conclusion pour les fautes d'orthographe étant donné que le test ANOVA a abouti à des résultats non significatifs. En outre, nous avons pu mettre au jour les tendances suivantes : les PE contenaient en moyenne moins d'erreurs de vocabulaire, de même que moins de problèmes de typographie et de ponctuation que les produits de la TH.

5.4.1.2.3 Sous-hypothèse SHA5.3

Troisièmement, comparativement aux productions humaines, les PE (*Google* comme *DeepL*) contiennent bel et bien davantage de marques de littéralité et plus spécifiquement de calques fautifs, ce qui valide notre sous-hypothèse SHA5.3.

5.4.1.2.4 Sous-hypothèse SHA5.4

Penchons-nous enfin sur la sous-hypothèse SHA5.4. Par rapport aux résultats de notre prétest, la variable « texte source » n'a pas eu d'influence significative sur autant de catégories d'erreurs. Nous avons observé des écarts significatifs pour les catégories suivantes (cf. Tableau 39) : acceptabilité (nombre d'erreurs et notes) ; grammaire et syntaxe ; typographie et ponctuation ; lexicale ; incompréhension ; cohérence ; calque et vocabulaire. Nous tenons une fois de plus à préciser qu'il est relativement difficile de pouvoir attribuer ces écarts significatifs au seul facteur « thématique du TS ». En effet,

d'autres éléments peuvent également expliquer ces divergences au niveau des distributions d'erreurs, tels que la sévérité et la sensibilité propres à chaque binôme d'évaluatrices.

5.4.1.3 Hypothèse HA6

Il ressort de notre analyse que la qualité d'un texte post-édité dépend bel et bien du moteur de TA neuronale employé (hypothèse HA6). Pour rappel, nous nous attendions à obtenir des productions de meilleure qualité avec le moteur *DeepL*, ce qui se confirme puisque *DeepL* a été plus performant que *Google Traduction* dans cette expérience. Effectivement, l'analyse descriptive du corpus a révélé qu'avec *DeepL*, les étudiantes ont produit de meilleurs résultats, puisque leurs productions contiennent, en moyenne, moins d'erreurs que les PE de TA générée avec *Google Traduction*. C'est le cas pour les deux phases d'évaluation (acceptabilité et adéquation) et quel que soit le texte source (Figure 51). Les résultats de l'analyse inférentielle ont également permis de vérifier l'hypothèse HA6 puisque nous avons relevé moins d'erreurs d'acceptabilité dans les PE de *DeepL* comparativement aux PE de *Google Traduction*. En outre, nous avons également relevé significativement moins de problèmes de typographie et de ponctuation, ainsi que moins d'erreurs de vocabulaire dans les PE de *DeepL*. Cette observation confirme le discours dominant véhiculé par les développeurs et les médias, selon lequel *DeepL* serait le plus performant des moteurs génériques de TA grand public.

5.4.1.4 Hypothèse HA7

Nos résultats permettent de valider l'existence d'un effet nivelant en PE de TA neuronale sur la qualité des textes cibles (hypothèse HA7). En effet, nous avons constaté un resserrement des notes autour de la moyenne en PE de TA neuronale (*Google Traduction* et *DeepL*) (Figure 49), quel que soit le TS, ainsi qu'un resserrement des notes autour de la médiane (Figure 50). Selon nous, cette homogénéisation des notes traduit un effet nivelant en PE sur la qualité des textes cibles en comparaison à la TH. D'ailleurs, l'analyse des notes par étudiante (Figure 56 et Figure 57) vient confirmer l'existence d'un tel effet. La PE de TA neuronale — en particulier *Google Traduction* — a donné lieu à une relation inverse entre le niveau d'une étudiante en TH et la qualité de sa PE (Figure 56). Cela signifie que plus l'étudiante est faible en TH, plus elle tire avantage de la PE et, à l'inverse, plus elle est compétente en TH, plus la PE entraîne une baisse de qualité de sa production. Nous estimons que ce mode de traduction a des effets ambivalents puisque d'un côté, il permet aux étudiantes faibles de s'en sortir en réduisant le nombre moyen d'erreurs et en leur évitant donc d'obtenir de (très) mauvaises notes, et que d'un autre côté, il entraîne une diminution des performances des bons éléments. Ceci corrobore non seulement d'autres études ayant déjà mis au jour cet effet nivelant en PE chez les étudiantes (voir notamment García, 2011 et Killman, 2018), mais aussi les résultats obtenus dans le prétest.

5.4.2 Étude du processus

Concernant le processus même de post-édition, nous pouvons conclure à un gain de productivité en PE par rapport à la TH. Il ressort de notre analyse que la PE de TA neuronale permet de traduire plus de mots, plus rapidement qu'en TH, comme cela a été démontré dans diverses recherches récentes (Jia et al., 2019 ; Martikainen et Mestivier, 2020 ; Samman, 2022 ; Screen, 2019) et comme nous l'avions déjà constaté dans le prétest. Nos résultats ne nous permettent cependant pas d'établir de corrélation entre la durée de traduction/post-édition et la qualité de la production finale. Autrement dit, nous ne pouvons pas affirmer que plus l'étudiante a passé du temps sur sa traduction/post-édition, meilleure est la note qu'elle a obtenue.

En ce qui concerne l'effort moyen de PE, nos observations ne révèlent pas de grandes divergences entre la PE de *Google* et le PE de *DeepL* pour les textes 1 et 2. Toutefois, pour le texte 3 (vulgarisation scientifique), l'effort moyen de PE est moindre avec le moteur *DeepL*.

Enfin, alors que nous ne pouvons pas conclure à une corrélation entre l'effort moyen en PE de *Google* et la qualité du TC, nous pouvons conclure à une corrélation positive modérée entre l'effort en PE de *DeepL* et la qualité du TC indiquant que plus l'étudiante a modifié la TA de *DeepL*, plus la qualité de sa production est élevée, et ce, quel que soit le texte source.

5.5 Conclusion

Nous pouvons conclure de cette deuxième recherche expérimentale que la PE de TA neuronale effectuée par des étudiantes de 2^e master en traduction n'a pas eu d'influence négative sur la qualité de leurs productions. En effet, tout comme dans notre prétest, il apparaît qu'en matière de qualité globale, les textes post-édités sont de qualité comparable, voire de meilleure qualité que les textes traduits humainement. Comparativement aux notes obtenues en TH, nous avons observé en PE de TA neuronale une moyenne générale, ainsi qu'une valeur médiane systématiquement plus élevée et de hauts taux de réussite, particulièrement en PE de *DeepL*. De plus, en matière de qualité globale, l'analyse statistique a indiqué une seule influence significative entre la PE de *DeepL* et la distribution des erreurs d'acceptabilité : les étudiantes ont commis significativement moins d'erreurs d'acceptabilité lorsqu'elles ont post-édité à partir de *DeepL* par rapport à la PE de *Google* et à la TH. Nous n'avons pas pu conclure à une relation de dépendance, ni entre le mode de traduction et les erreurs d'adéquation, ni entre le mode de traduction et les notes attribuées aux productions des étudiantes.

Par ailleurs, il ressort de l'analyse détaillée des catégories d'erreurs que les productions humaines contiennent d'une part, davantage d'erreurs de grammaire et de syntaxe et d'autre part, moins de calques que les textes post-édités (*DeepL* et *Google Traduction*). De

plus, nous pouvons affirmer que les textes post-édités par les étudiantes à partir de la TA de *DeepL* contiennent significativement moins de problèmes de typographie et de ponctuation, ainsi que moins d'erreurs de vocabulaire par rapport aux deux autres modes de traduction observés.

Encore une fois, nous pensons que ces constats amènent à promouvoir l'intégration d'une formation en TA et en PE dans les cursus des apprenantes en traduction ; nous abordons cette réflexion au chapitre 8.

Concernant le processus même de post-édition, nos résultats attestent un gain de productivité en PE par rapport à la TH, mais montrent une absence de corrélation entre la durée de traduction/post-édition et la qualité de la production de l'étudiante. Enfin, nous avons pu observer une corrélation entre l'effort moyen en PE de *DeepL* et la qualité du texte cible, indiquant que plus l'étudiante a modifié la TA de *DeepL*, plus la qualité de sa production est élevée, et ce, quel que soit le texte source. Néanmoins, ces conclusions nécessiteraient d'être validées par un test statistique.

5.5.1 Hypothèses explicatives (Prétest & Test)

Dans nos deux expériences (prétest et test), nous avons relevé une moyenne des notes plus élevée, ainsi qu'un meilleur taux de réussite en PE de TA neuronale par rapport aux résultats obtenus en traduction humaine. Mais comment expliquer de telles performances ? Comment expliquer également la plus grande occurrence de calques en PE ? Nous avons tenté de dégager des hypothèses explicatives.

5.5.1.1 Accès au sens du texte source

Nous émettons une première hypothèse selon laquelle la TA neuronale permettrait aux étudiantes d'avoir, en quelque sorte, un meilleur « accès » au sens du texte source. Tout comme Guerberof Arenas (2009), Lee and Liao (2011) et Pym (2003)¹⁰⁸ l'ont déjà constaté en TA statistique, nous pensons que la TA neuronale peut pallier la compréhension lacunaire ou partielle des étudiantes en clarifiant certains passages complexes du TS, ce qui par conséquent permet d'améliorer la compréhension générale qu'elles ont du TS. Dans le prétest 2018, ce phénomène est particulièrement flagrant dans le texte 2 pour lequel le taux de réussite est extrêmement faible, tant en PE de TA statistique (13 pour cent) qu'en TH (22 pour cent), alors que le taux de réussite en PE de TA neuronale atteint, lui, 63 pour cent (Figure 26). Nous pourrions également supposer que certaines étudiantes ont des faiblesses en langue cible (fautes de grammaire, de syntaxe, tournures non idiomatiques...), et que la TA permet de combler, en tout ou en partie, ces lacunes. De plus, corroborant les résultats de Daems (2016) et de Guerberof Arenas (2009), les statistiques inférentielles du prétest révèlent un plus faible nombre d'occurrences de glissements de sens en PE de TA

¹⁰⁸ « the MT suggestions replace deficiencies in knowledge of the foreign language, which is another way of saying that the MT acts as a surrogate (and hopefully mistrusted) instructor » (Pym, 2013, p. 500).

neuronale qu'en TH. Tout comme nous venons de le signaler, nous pensons que cela peut s'expliquer par le fait que la TA permet aux étudiantes de mieux comprendre le TS. Par conséquent, les étudiantes seraient moins susceptibles de produire des glissements de sens en langue cible : « The fact that there are so few mistranslation errors in [post-edited] MT segments might indicate that using MT helps translators clarify possibly difficult aspects of the source texts thus improving general comprehension of the text » (*ibid.*, p. 18).

Cette hypothèse explicative sera, qui plus est, étayée grâce aux témoignages des étudiantes recueillis dans nos sondages (chapitre 7). Cet usage particulier de la TA a déjà été mis en évidence dans plusieurs études récentes : dans le sondage mené par Loock et Léchauguette (2021) auprès d'étudiantes en LEA, mais aussi dans l'enquête effectuée par Dirand et Rossi (2019) pour laquelle la majeure partie des étudiantes interrogées (82 %) a déclaré avoir recours à la TA afin de comprendre la langue étrangère, et enfin dans les avis recueillis par González Pastor (2021) : « A large number of respondents (16) reported to resort to MT as a problem-solving tool when faced with text fragments containing especially difficult sentences or complex syntactic structures » (p. 56).

5.5.1.2 Effet nivelant en PE de TA neuronale

En outre, nous avons observé des écarts-types réduits en PE par rapport à la TH qui reflètent un resserrement des notes autour de la moyenne, ainsi qu'une relation inverse entre le niveau des étudiantes en TH et la qualité de leur PE de TA neuronale. Selon nous, ces observations traduisent une homogénéisation des notes reflétant un effet nivelant en PE de TA neuronale. D'un côté, ces résultats nous amènent à considérer que la PE de TA neuronale permet aux étudiantes les plus faibles d'augmenter la qualité de leur production finale et d'obtenir une meilleure note que lorsqu'*elles* traduisent humainement. D'un autre côté, nous pensons que ce mode de traduction aurait également un effet limitant, car il entraînerait, chez les bons éléments, une baisse de qualité du texte cible. Les étudiantes faibles tireraient donc davantage parti de la PE de TA neuronale que les autres étudiantes. En revanche, il est interpellant de constater que les *bonnes* étudiantes seraient, *elles*, désavantagées par la PE de TA neuronale qui ne leur permet pas de produire des traductions d'aussi bonne qualité que lorsqu'*elles* traduisent humainement. Ces résultats vont dans le même sens que les observations de García (2011), Lee and Liao (2011) et Killman (2018)¹⁰⁹ qui ont mis au jour cet effet nivelant de la PE (*leveling effect*) chez les étudiantes. Au vu d'un tel constat, nous concluons que, l'un des effets pervers de la PE de TA neuronale dans ces deux expériences est la diminution des performances chez les bons

¹⁰⁹ « [T]he results of this study show that post-editing allowed students below, at, and just above the established median to improve the quality of their translation products, especially in cases of students with particularly high HT error ratios and with performances below the median. The HT quality and performances of the translators well above the median are superior to their PET quality and performances. Nevertheless, in terms of post-editing performances, these translators still managed to remain on top of the others, but to a lesser extent. Both they and the translators at and just above the median all experienced a reduction in performance levels when post-editing, while those below it all experienced a rise, giving way to a leveling effect » (Killman, 2018, p. 137).

éléments. Nous formulons l’hypothèse que ce mode de traduction entraîne une réduction des intuitions de traduction et de la créativité traductionnelle, notamment en venant perturber les processus de déverbalisation et de (re)verbalisation¹¹⁰ (Seleskovitch et Lederer, 2014). Nous pourrions dès lors légitimement nous demander s’il est pertinent d’envisager une formation en PE pour tous les profils d’étudiantes en traduction.

5.5.1.3 Maîtrise de l’orthographe et de la grammaire

Concernant les fautes d’orthographe, nos observations rejoignent celles faites par Martikainen (2019c) dans son étude comparative de PE de TA (*DeepL*) et de TH de textes spécialisés menée auprès d’apprenantes. Force est de constater que nous avons relevé très peu d’erreurs d’orthographe dans les TA brutes et à fortiori dans les textes post-édités, ce qui constitue un avantage certain, surtout pour les étudiantes qui ont des faiblesses dans ce domaine. Dans notre cas, il en va de même pour les erreurs de grammaire et de syntaxe. Ces résultats, qui ont permis de confirmer nos sous-hypothèses SHA2.4. et SHA5.2., laissent à penser que la TA – particulièrement la nouvelle génération de TA – a intégré l’ensemble des règles d’orthographe d’usage et de grammaire de la langue française en l’occurrence, ce qui débouche sur des traductions qui contiennent relativement peu de fautes de cet ordre. Nous ne pouvons néanmoins pas en dire autant des productions humaines. Cette hypothèse se confirme dans l’étude de Loock et Léchaugette (2021) : « the neural version of MT no longer "makes" blatant grammar and spelling mistakes as SMT did » (p. 214).

5.5.1.4 Effet fantôme de la TA

Comparativement à la TH, l’apparition plus fréquente de marques de littéralité et de calques constitue une tendance marquée en TA ; cette forte proximité linguistique avec le TS est d’ailleurs étayée par T. Poibeau (2019) dans son livre *Babel 2.0 Où va la traduction automatique* :

[L]es techniques utilisées [en TA] restent très proches du texte, c’est-à-dire d’une traduction mot à mot, et ne peuvent donc concevoir des traductions de plus haut niveau, où les équivalences entre langues sont trouvées grâce à des informations dépassant l’empan de la phrase simple. (p. 168)

Confirmant nos sous-hypothèses SHA2.2. et SHA5.3., cette tendance se répercute alors dans les textes post-édités, entre autres, car il y aurait un fort effet d’interférence avec le texte source induit par le processus même de PE (Čulo et al., 2014 ; Toral, 2019). Cela peut paraître évident, mais il est toujours bon de le rappeler, le texte cible généré par la TA reste en effet plus proche linguistiquement du TS (Depraetere, 2010 ; Toral, 2019) et est encore

¹¹⁰ « Si l’on s’en tient aux postulats de la théorie interprétative de la traduction (Seleskovitch et Lederer 2014 [1984]) [...], ces modifications [de sens] ne peuvent intervenir qu’au cours des deux opérations qui constituent, selon les auteures, l’acte même de la traduction, à savoir la déverbalisation du sens (compréhension comme transformation du contenu en une représentation mentale) et sa (re)verbalisation (expression dans la langue cible) » (Vogeleer et Béghin, 2020, p. 7).

considéré comme trop littéral par rapport à une TH, même dans le cas de textes généralistes à dominante informative :

[D]espite these recent advances in the field, machine translations remain quite literal [...] despite powerful architectures based on transformers and accessing the whole sentence in one go, translations remain literal and based on local equivalencies. The consequence is an accurate but very literal way of translating, that is not always convenient [...] even for articles in newspapers, NMT is often judged too literal. (Poibeau, 2022, p. 6020)

En dépit des progrès notamment en matière de fluidité, la machine continuerait donc à traduire littéralement en procédant essentiellement par calque (Martikainen, 2020). Aussi, nous avons pu constater, qu'en post-éditant, les étudiantes avaient tendance à conserver, consciemment ou inconsciemment, cette proximité avec le texte source et à y laisser, entre autres, toute une série de calques fautifs : « [students] seem to be heavily primed by the MT output (even in the case of very awkward direct calques) » (Loock et Léchaugnette, 2021, p. 214). Une influence que nous avons traduite librement en « effet fantôme » de la TA dès lors que les propositions de la machine flottent intempestivement à l'esprit de la post-éditrice et que la TA laisse sa marque dans le texte post-édité. Plusieurs facteurs peuvent être à l'origine de ce phénomène : nous sommes d'avis qu'en PE, la présence de la TA brute vient parasiter le processus de déverbalisation et par conséquent celui de reverbération, l'étudiante éprouvant ainsi des difficultés à se détacher des propositions de la TA. Il est fort probable également qu'il y ait un excès de confiance de la part des étudiantes à l'égard des propositions de la machine (Daems, 2016 ; Deneufbourg, 2019 ; Depraetere, 2010 ; Loock et Léchaugnette, 2021 ; Martikainen, 2020), ou encore que les étudiantes se soient tout simplement habituées à la TA (Guerberof Arenas et Moorkens, 2019 ; Martikainen et Kübler, 2016). Martikainen et Mestivier (2020) ont, elles aussi, observé une plus forte présence de calques en PE ; selon elles, le calque serait d'ailleurs l'une des caractéristiques majeures du *post-editeuse* (voir nbp 32), concept auquel nous consacrons le chapitre suivant.

6 Vous avez dit *post-editeuse* ?

Nous avons voulu approfondir notre étude comparative entre les textes traduits humainement et les textes post-édités, et vérifier la présence d'un *post-editeuse* dans nos corpus. Ce chapitre est divisé en deux grandes parties.

Dans la première partie, nous avons soumis nos données à différentes mesures linguistiques automatiques comme cela est fréquemment le cas en traitement automatique des langues et dans les études qui s'attachent à démontrer l'existence d'un *translationese* ou *post-editeuse*. Dans nos deux corpus de 2018 et de 2021, nous avons eu recours à trois mesures linguistiques automatiques : la richesse lexicale (densité et diversité lexicales), la longueur moyenne des phrases et l'équivalence syntaxique (analyse statistique de distance textuelle).

Dans la deuxième partie, nous complétons l'analyse quantitative par une analyse qualitative comparative des erreurs en TH et en PE relevées dans nos corpus. Nous avons souhaité mener une analyse qualitative sans nous appuyer cette fois sur la puissance de tests statistiques, car nous partageons l'avis de Gile (2011) selon lequel les conditions particulières de la recherche en traductologie requièrent parfois de s'écarter de protocoles d'expérience standard. Gile est convaincu que dans le cas d'une accumulation de « résultats sur un ensemble varié de conditions ; une éventuelle convergence des tendances observées peut être bien plus convaincante qu'un résultat statistiquement significatif lors d'une expérience contrôlée, mais unique » (*ibid.*, p. 57).

6.1 Mesures linguistiques automatiques

Dans cette première partie, nous allons explorer trois mesures linguistiques automatiques qui ont servi à analyser nos corpus. Nous commençons par la richesse lexicale à travers deux dimensions : la densité lexicale et la diversité lexicale. Nous nous pencherons ensuite sur la longueur moyenne des phrases qui est corrélée au coefficient de foisonnement et nous terminerons par l'indice d'équivalence syntaxique qui permet d'étudier la proximité syntaxique d'un texte cible par rapport à un texte source.

En nous appuyant sur la littérature scientifique dont nous proposons un aperçu ci-dessous, nous supposons que les étudiantes sont influencées 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-editeuse*), 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'étudiantes, 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

étudiantes ont tendance à se satisfaire davantage de traductions littérales et à commettre davantage de calques par rapport à la TH. Enfin, nous pensons que cette plus grande proximité linguistique sera observable tant dans le corpus 2018 que dans le corpus 2021, c'est pourquoi nous ne nous attendons pas à observer de grandes divergences entre les résultats de 2018 et les résultats de 2021.

6.1.1 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 et Lambelet, 2019, p. 66), nous avons souhaité reprendre la définition singulière et subtile qu'en donne E. Brunet¹¹¹ (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. (p. 5)

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 (1969), 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. (p. 36)

Comment mesurer cette richesse lexicale ?

Dans cette étude, nous analysons la richesse lexicale grâce à deux indices : la densité lexicale et la diversité lexicale.

¹¹¹ Étienne Brunet est un linguiste français, spécialiste en informatique et figure de proue de la statistique textuelle, il est le concepteur et développeur du logiciel Hyperbase consacré au traitement documentaire et statistique de corpus textuels.

6.1.1.1 La densité lexicale et la diversité lexicale

6.1.1.1.1 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 et al., 2013, p. 61). Alors que cet indice est fréquemment employé dans les études sur la PE (Volkart et 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 et al., 2019 et Castilho et Resende, 2022).

Selon Scarpa (2006), plus cette densité lexicale est élevée, plus difficile sera la compréhension du texte par le lecteur. Dans son étude de corpus pour la traduction de l'anglais en italien, Scarpa (2006) 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 (2022) 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. (p. 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 » (Castilho et Resende, 2022, p. 9).

En 2019, dans son étude comparative de trois systèmes de TA menée pour cinq combinaisons de langues¹¹², 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

¹¹² Allemand-anglais ; anglais-allemand ; espagnol-allemand ; anglais-français et chinois-anglais.

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.

Hypothèse

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.

6.1.1.1.2 La diversité lexicale ou variation lexicale

Bonvin et Lambelet (2019) 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). (p. 66)

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 et Way, 2018). Précisons qu'à l'instar de Toral (2019)¹¹³ entre autres, nous avons supposé une corrélation entre richesse lexicale et qualité du texte. En d'autres termes, nous jugeons que de manière générale plus une traduction est variée et dense sur le plan lexical, plus elle est de bonne qualité, tout en reconnaissant que pour certains types de textes, il peut être utile de tendre davantage vers une uniformité lexicale¹¹⁴.

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 et Resende, 2022).

Sans surprise, cette perte de richesse lexicale est également à déplorer en TA par rapport au *translationese*, lui-même moins riche que la langue originale, tel qu'en attestent

¹¹³ « [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, p. 280).

¹¹⁴ « Variety and inventiveness are not always desirable features in every kind of text. For example, excessive lexical variation might make a smartphone user's guide more difficult to follow. Nevertheless, there are various other kinds where lexical uniformity would make the text less interesting to read and less intellectually stimulating (marketing, advertising, literature, journalism, education, entertainment, and creative writing in general) » (Farrell, 2018, p. 58).

plusieurs études récentes (voir De Clercq et al., 2021 ; Hansen et Esperança-Rodier, 2023 ; Martikainen, 2019c ; Toral, 2019 ; Vanmassenhove et al., 2019 et 2021). Effectivement, Martikainen (2019c) évoque un lissage terminologique en TA statistique : « [M]ême si le moteur de traduction automatique est modelé sur la traduction humaine, la TA statistique a tendance à lisser cette variation [terminologique], en gommant les solutions minoritaires et en privilégiant les solutions de traduction majoritaires » (p. 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 et al. (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, 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’une part, il y aurait moins de variation terminologique en TA qu’en TH et que d’autre part, l’usage terminologique qui est fait en PE se rapprocherait fortement de l’usage terminologique en TA : « MT shows variation for fewer terms than HT [...] Our findings reveal levels of variation on the terminological level in the post-edited texts close, but not identical, to those of the machine translation outcomes » (p. 110).

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 » (p. 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 all cases, that of MT is lower than that of PE » (p. 276). Toujours selon lui, le travail de la post-éditrice permettrait certes d’ajouter de la diversité lexicale par rapport à la TA brute, mais ne permettrait pas d’aboutir à une production aussi variée qu’en TH en raison du haut degré d’interférence de la TA en PE :

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., p. 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.

6.1.1.2 Hypothèse

En nous fondant sur les résultats des travaux résumés ci-dessus, 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 nous attendons à observer une perte de richesse lexicale en PE par rapport à la TH. Enfin, en nous appuyant sur les observations de Vanmassenhove et al. (2019) qui concernent uniquement la TA, ainsi que sur celles de Toral (2019), nous supposons que la PE de TA neuronale se révélera moins riche lexicalement que la PE de TA statistique et, en ce sens, plus proches des TS.

6.1.1.3 Densité lexicale : Calculs et résultats

6.1.1.3.1 Calculs

Pour mesurer la densité lexicale (Figure 63), il s’agit d’établir le rapport mathématique entre le nombre de vocables¹¹⁵ lexicaux (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).

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

Figure 63 : Calcul de densité lexicale

Dans cette étude, nous avons calculé la densité lexicale grâce au logiciel *UAM corpus tool* (O’Donnell, 2016). Afin d’obtenir une comparaison pertinente des données, nous avons retenu huit productions par méthode de traduction en 2018 et sept productions par méthode en 2021. Ceci explique que nous n’obtenons pas des valeurs identiques dans les Tableau 47 et Tableau 48 pour les TS alors qu’il s’agit bel et bien des mêmes TS en 2018 et en 2021. Les résultats sont repris par expérience (2018 et 2021) dans le Tableau 47 et le Tableau 48.

¹¹⁵ Un vocable désigne un « mot, terme d’une langue considéré surtout sous le rapport de sa signification, de son individualité lexicale (*petit, petite, petits* sont trois formes d’un même vocable) » (Larousse en ligne, s.d.).

6.1.1.3.2 Résultats

Prétest 2018	Vocables lexicaux/phrased	Densité lexicale (%)
TS	10,6	52,12
TH	14,4	67
TAS	15	66,21
TAN (DeepL)	14,3	66,19
PE de TAS	15	66
PE de TAN (DeepL)	14,6	66,59

Tableau 47 : Densité lexicale – Prétest

Test 2021	Vocables lexicaux/phrased	Densité lexicale (%)
TS	10,6	51,79
TH	13,7	67,85
TA (Google)	14,6	65,49
TA (DeepL)	14,9	65,18
PE (Google)	13,8	67
PE (DeepL)	14	67,39

Tableau 48 : Densité lexicale – Test

Il apparaît très nettement dans le Tableau 47 et le Tableau 48 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 travaux sur lesquels nous nous sommes appuyée (Toral, 2019 ; Volkart et Bouillon, 2022). Malgré des valeurs très proches, nous pouvons affirmer qu'en 2018, tout comme en 2021, c'est en TH que le taux de densité lexicale est très légèrement plus élevé (67 % et 67,85 %). De plus, les résultats de 2021 révèlent que les TA brutes étaient légèrement moins denses lexicalement (65,49 % et 65,15 %) que les TH et que les PE. Ces résultats confirment l'hypothèse de travail qu'avaient formulée Castilho et Resende (2022).

6.1.1.4 Diversité lexicale : calculs et résultats

6.1.1.4.1 Calculs

La mesure la plus fréquemment utilisée pour mesurer la diversité lexicale (Figure 64) est le *Type-Token Ratio* (TTR) (Templin, 1957) ; il s'agit de calculer le rapport mathématique entre le nombre de vocables d'un texte (*types*) et le nombre total d'occurrences (*token*).

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

Figure 64 : Calcul de diversité lexicale

Toutefois, il a été démontré dans diverses études (Brezina, 2018 ; Tezcan et al., 2019) que le calcul de cette mesure est fortement influencé par la taille de l'échantillon, puisque comme l'expliquent Bonvin et Lambelet (2019), « [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 » (p. 66). 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)¹¹⁶ (Johnson, 1944), ainsi que le *Moving Average Type-Token Ratio* (MATTR)¹¹⁷ (Covington et 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*¹¹⁸.

6.1.1.4.2 Résultats

<i>Diversité lexicale (2018)</i>	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 (DeepL)	9712	658	0,07	0,36	0,36

Tableau 49 : Diversité lexicale – Prétest

<i>Diversité lexicale (2021)</i>	Occurrences	Vocables	TTR	MSTTR	MATTR
TH	9838	1045	0,11	0,39	0,39
PE (Google)	9405	800	0,09	0,37	0,37
PE (DeepL)	9603	742	0,08	0,37	0,37

Tableau 50 : Diversité lexicale – Test

¹¹⁶ 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 et Capsada, 2013, p. 449).

¹¹⁷ « 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 et McFall, 2010, p. 96).

¹¹⁸ Disponible en accès libre sur <https://github.com/LSYS/LexicalRichness>.

Les résultats qui figurent dans les Tableau 49 et Tableau 50¹¹⁹ viennent confirmer nos suppositions. Tout d’abord, nous pouvons constater 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 (2022b), de Toral (2019), ainsi que celle de Volkart et Bouillon (2022).

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) : « 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 » (p. 276). Ceci pourrait découler du fait que les post-éditrices apportent davantage de modifications en TA statistique qu’en TA neuronale. De plus, ce constat est interpellant, d’autant plus au vu de l’usage de plus en plus répandu de l’outil *DeepL* par le grand public et de l’intégration de la TA neuronale dans le flux de travail des traductrices professionnelles, 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 avons 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, nous prouve que l’indice de diversité lexicale est indépendant de l’indice de densité lexicale et qu’il était dès lors pertinent de ne pas se limiter au calcul de la densité lexicale. 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. Les causes probables de cette perte de diversité lexicale en PE sont développées dans la conclusion (6.3).

¹¹⁹ Les tableaux complets de résultats figurent en annexe (Tableau E. 1 et Tableau E. 2)

6.1.2 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.

Penchons-nous d'abord sur l'indice de longueur moyenne des phrases.

Dans son étude portant sur l'existence d'un *post-editeuse*, 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. Toral 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. (p. 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) comme « l'augmentation de volume du texte d'arrivée par rapport au texte de départ » (p. 55). 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, p. 187)

Cependant, si ce foisonnement est excessif, cela peut nuire à la qualité et à la lisibilité des traductions (Delisle, 2013). Comme le fait remarquer Cochrane (*ibid.*), 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 ci-dessous un aperçu.

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 » (p. 5). Elles ont également constaté un foisonnement entre la TA et la PE (3,97 %) qu'elles attribuent principalement aux modifications syntaxiques apportées par les post-éditrices.

La même tendance se dégage des travaux de Castilho et al. (2019). Elles ont voulu explorer l'existence d'un *post-editeuse* 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 traductrices professionnelles qu'à des étudiantes 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-editeuse* :

Post-editeuse 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., pp. 25-26)

En matière de longueur (longueur moyenne des phrases et foisonnement), elles ont abouti à la conclusion que les textes post-édités se rapprochaient 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.*, p. 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 %).

6.1.2.1 Hypothèse

De notre côté, en nous fondant sur les études de Castilho et al. (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 étudiantes présenteront une longueur moyenne de phrases et un coefficient de foisonnement moins élevés que les TH et donc plus proches des valeurs du TS.

6.1.2.2 Calculs

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 65).

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

Figure 65 : Calcul de longueur moy. des phrases

À nouveau, afin d'obtenir une comparaison pertinente des données, seules huit productions ont été retenues par méthode de traduction en 2018 avec un total de 336 phrases, et sept productions par méthode en 2021 avec un total de 294 phrases. C'est pourquoi nous n'obtenons pas deux valeurs identiques dans le Tableau 51 et le Tableau 52 pour la longueur moyenne des phrases des TS alors qu'il s'agit bel et bien des mêmes TS en 2018 et en 2021. Pour le calcul du coefficient de foisonnement, nous avons employé la formule suivante :

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

Figure 66 : Calcul du foisonnement

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

6.1.2.3 Résultats

Prétest 2018	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 (DeepL)	29,64	+19,48 %

Tableau 51 : Longueur des phrases et foisonnement – Prétest

Test 2021	Longueur moy. des phrases	Coefficient de foisonnement
TS	24,74	–
TH	30,26	+22,32 %
PE (Google)	29,98	+21,17 %
PE (DeepL)	30,12	+21,75 %

Tableau 52 : Longueur des phrases et foisonnement – Test

Confirmant ce qui a déjà été mis au jour à plusieurs reprises dans la littérature, les résultats repris dans ces deux tableaux¹²⁰ 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 car contrairement aux échantillons de ces études, nos échantillons se composent exclusivement de productions d'étudiantes ?

En matière de foisonnement, les taux oscillent entre 17 % et 22 % ce qui correspond au foisonnement communément attendu pour la traduction de l'anglais en français. Pour la comparaison entre la PE de TA statistique et la TH en 2018, 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 (2018 et 2021) sont comparables aux taux de foisonnement obtenus 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 le Tableau 51 et le Tableau 52 ne nous permettent pas de révéler l'existence de caractéristiques propres à une langue de PE (*post-editeuse*).

¹²⁰ Les tableaux complets de résultats figurent en annexe (Tableau E. 3).

Ajoutons finalement qu'il est intéressant de constater que les valeurs obtenues en 2021 sont toutes légèrement plus élevées que les valeurs de 2018. Serait-ce dû à la différence d'expérience des participantes ? Par rapport aux étudiantes de 3^e bachelier (2018), les étudiantes de 2^e master (2021) se sont-elles davantage éloignées du TS et de la TA brute, auraient-elles eu davantage recours à l'explicitation ?

6.1.3 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 pu être 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, p. 259).
- 2) *Part-of-Speech* (POS) *changes* : Il s'agit de mesurer les changements en matière d'étiquetage morphosyntaxique¹²¹ entre les phrases sources et cibles.
- 3) *Aligned Syntactic Tree Edit Distance* (ASTrED) : Ce troisième indice va mesurer 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.* p. 291). Vanroy et al. (2021) nous fournissent un exemple d'arbre de dépendance pour la phrase « He eats the cookies » :

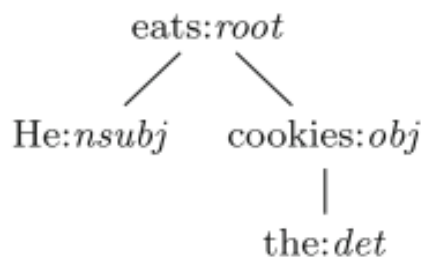


Figure 67 : Exemple d'arbre de dépendance (Vanroy et al., 2021, p. 264)

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

¹²¹ 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 » (DataFranca, 2021).

changes et ASTR_{ED}) : 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 était plus proche de l'étiquetage morphosyntaxique du TS comparativement à l'étiquetage en TH.

6.1.3.1 Hypothèse

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.

6.1.3.2 Calculs

Cette proposition de mesure d'équivalence syntaxique (Vanroy et al., 2021) a été calculée grâce au module Python ASTR_{ED}¹²² (*Aligned Syntactic Tree Edit Distance*). Ce module utilise la librairie¹²³ Stanza Python NLP (Qi et al., 2020) développée par le *Stanford NLP Group* et fréquemment utilisée en traitement automatique des langues. De plus, 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).

6.1.3.3 Résultats

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

Tableau 53 : Équivalence syntaxique – Prétest

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

Tableau 54 : Équivalence syntaxique – Test

¹²² Disponible en accès libre sur <https://github.com/BramVanroy/astred>

¹²³ « En informatique, une librairie, aussi appelée "bibliothèque de code" ou "package", est un ensemble de code prêt à l'usage qui peut être facilement importé et réutilisé par un utilisateur pour que celui-ci n'ait pas besoin de réécrire ces portions de code » (Etablab, 2021, p. 22).

Dans le Tableau 53 et le Tableau 54¹²⁴ 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 (en 2018) et par rapport à la PE de *Google* (en 2021).

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 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 (en 2018) et que les PE de *Google* (en 2021).

6.1.4 Synthèse et discussion

À l'instar de Toral (2019) et de Volkart et Bouillon (2022), mais contrairement à Daems, De Clercq et al. (2017)¹²⁵, nous pouvons dire qu'une partie de nos résultats confirme l'existence d'un *post-editeuse*. En effet, les comparaisons en matière de diversité lexicale et d'équivalence syntaxique ont permis de mettre 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 de départ, les résultats du calcul des 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 participantes aux expériences a pu jouer un rôle. En effet, le fait que les textes post-édités se sont révélés presque tout aussi denses et foisonnants que les TH peut provenir du fait que nos expériences ont été menées auprès d'apprenantes en traduction et non auprès de traductrices professionnelles, telle est en tout cas notre principale hypothèse explicative.

Si l'on suit le raisonnement de Toral (2019), ces caractéristiques propres au *post-editeuse* sont en réalité déjà présentes en TA ; « [T]hese characteristics are already present in the MT outputs that are the starting point of the PEs » (p. 280). Selon nous, ces marques caractéristiques sont dues à un double phénomène d'interférence : à la fois à l'interférence indirecte avec la langue source en PE puisque la production passe par le filtre de la TA qui

¹²⁴ Les tableaux complets de résultats figurent en annexe (Tableau E. 4).

¹²⁵ « We did not find proof of the existence of *Post-editeuse*, either perceived or measurable » (Daems, DeClercq et al., 2017, p. 101).

présente un haut degré d'interférence avec le TS et 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, p. 25).

Dernièrement, Poibeau (2022) mentionne que le problème vient du fait que la majorité des systèmes neuronaux continuent de fonctionner au niveau de la phrase, sans une prise en compte plus large du contexte.

NMT is still mainly based on knowledge inferred from large collections of parallel data, where one sentence in the source language corresponds to one sentence in the target language. From this point of view, NMT is a direct continuation of the previous segment-based approaches. Equivalences between languages are found inside the sentence at a more or less local level, despite recent attempts to integrate a wider context in the process. (pp. 6019-6020)

Selon Vanmassenhove et al. (2019), la perte de diversité lexicale en TA par rapport à la TH découlerait d'un type de biais algorithmique lié à la nature même des systèmes de TA qui ont tendance à la « surgénéralisation » (*overgeneralization*) à partir des données d'entraînement :

[T]he inherent nature of data-driven MT systems to generalise over the training data has a quantitatively distinguishable negative impact on the word choice, expressed by favouring more frequent words and disregarding less frequent ones [...] MT paradigms indeed increase/decrease the frequencies of more/less frequent words to such extent that a very large amount of words are completely 'lost in translation'. We believe, this demonstrates indeed that current systems overgeneralize and thus, we deem it appropriate to speak of a form of algorithmic bias. (p. 224 et p. 230)

Toujours selon Vanmassenhove et al. (*ibid.*), la conséquence la plus visible de ce biais algorithmique concerne les occurrences de mots ou de tournures rares :

We hypothesize that the most visible effect of such bias is to be found in the word frequencies and the disappearance (or 'non-appearance') of scarce words [...] the more frequent words observed in the input occur even more in the output, negatively affecting the frequency of less seen events or words by causing them to become even rarer events or causing them to disappear altogether. (ibid., p. 224)

Martikainen (2019c) partage également ce constat dans sa thèse de doctorat qui porte sur les sources de distorsion dans les résumés traduits de revues systématiques :

[L]a variété caractéristique de la traduction humaine, qui s'exprime à travers des solutions de traduction uniques ou des hapax, est progressivement éliminée par l'intervention de la traduction automatique, qui tend à privilégier les solutions statistiquement plus fréquentes et le calque mot-à-mot. (p. 277)

Tout comme Vanmassenhove et al. (2019), Poibeau (2022) attribue ce manque de diversité lexicale à la nature statistique de la TA neuronale : « The statistical nature of NMT also favours standard translations over more original ones. Moreover, NMT does not include reformulation or paraphrase modules that would allow the system to produce more varied or more global decisions » (pp. 6019-6020).

Déviations par rapport à l'usage

En outre, grâce à son étude de corpus portant sur la TA anglais-français, Loock (2018) a pu mettre en lumière une déviation en langue traduite automatiquement par rapport à l'usage linguistique courant en langue cible. Il a réalisé une analyse linguistique d'un double corpus de français original et de français traduit automatiquement au moyen de deux systèmes de TA : TA neuronale avec *DeepL* et TA statistique avec MT@EC/eTranslation¹²⁶. Il s'est penché sur la fréquence d'occurrence de certains phénomènes linguistiques et a découvert que les deux outils de TA engendraient une surreprésentation des phénomènes lexicaux et grammaticaux analysés en français traduit automatiquement de l'anglais par rapport au français original et « dans des proportions plus importantes par l'outil de TA neuronale [*DeepL*] » (*ibid.*, p. 798). Il peut s'agir de phénomènes lexicaux tels qu'une surreprésentation en français traduit automatiquement du lemme « chose » ou du lemme « dire » ; mais aussi de phénomènes grammaticaux, tels qu'une surreprésentation des adverbes dérivés en *-ment*, de la préposition « avec » ou encore une surreprésentation de la structure existentielle « il y a ». Toujours selon Loock (2020c), l'interférence avec la langue source seule ne suffit pas à justifier cette déviation par rapport à l'usage de la langue cible. Il pointe également du doigt les données d'entraînement : « [T]he translated data used to train the MT systems might show certain *translationese* features that get reproduced in the MT outputs » (p. 160). Aussi, il apparaît que le processus de PE ne permet manifestement pas de pallier les biais présents en TA. La TA laisserait sa trace en PE comme le faisaient remarquer Lee et Liao (2011) dans leur étude comparative de TH et de PE :

The imperfection lies mainly in the MT text constraining the students to translate freely as they, though not all, are often confined by the MT text, resulting in translations with rigid sentence patterns or unsatisfactory terminology when the MT text is incorrect to begin with. (p. 138)

Čulo et Nitzke (2016) renvoient au concept de *shining through* chez Teich (2003)¹²⁷ qu'elles appliquent à la TA pour expliquer la faible variation lexicale en PE :

¹²⁶ Il s'agit de l'outil de la Direction générale de la traduction (DGT) de la Commission européenne. Au moment de l'étude menée par R. Loock, le service était en transition entre la technologie statistique et la technologie neuronale, d'où l'appellation MT@EC/eTranslation. Depuis le 15 novembre 2017, la DGT utilise uniquement la traduction automatique neuronale, appelée eTranslation.

¹²⁷ Teich (2003) définit la notion de *shining through* comme suit : « In a translation into a given target language (TL), the translation may be oriented more towards the source language (SL), i.e. the SL shines through. This contributes to making a translation different from a comparable text in the same language as the TL, i.e. a text that is originally produced in the TL » (p. 145).

Our findings reveal levels of variation on the terminological level in the post-edited texts close, but not identical, to those of the machine translation outcomes. MT shows variation for fewer terms than HT, but in cases of variation, usually exhibits stronger variation. In PE, the variation patterns of MT were carried over even though the task description explicitly stated to correct inconsistent terminology. [...] This [...] indicates a shining through (Teich, 2003) of the MT in the post-editing products on the terminological level. (p. 110)

Ce constat est partagé par Castilho et Resende (2022) dont les travaux en TA littéraire confirment que la TA influe sur les choix lexicaux et syntaxiques posés par la post-éditrice :

[T]ranslators should be aware of the priming effect of the raw MT output on their lexical and syntactic choices [...] [B]ased on our results, we assumed that literary creativity in PE texts may be compromised [...], due to the influence of the MT lexical and syntactic choices on the translators' choices. (p. 20)

Simplification et standardisation

Tout d'abord, qu'entend-on par simplification ? La simplification lexicale peut se définir comme le « process and/or result of making do with less words » (Blum-Kulka et Levenston, 1983, p. 119). En traduction, le phénomène de simplification a été défini par Baker (1996) comme suit : « tendency to simplify the language used in translation [...] it involves making things easier for the reader » (pp. 181-182), comparativement à des textes rédigés en langue originale. Cette simplification se caractérise principalement par une limitation de la variation du vocabulaire employé, mais elle peut également se manifester par un réagencement syntaxique et un découpage des phrases : « *[S]implification can also be reflected by number of sentences and sentence length [a]s translators tend to split long sentences into smaller ones to facilitate text comprehension » (Castilho et al., 2019, p. 21).*

Tandis que la normalisation, ci-après également « standardisation », est la « tendency to exaggerate features of the target language and to conform to its typical patterns » (Baker, 1996, p. 183). Loock (2016) précise qu'il s'agit de la « tendance qu'auraient les traducteurs à surexploiter les caractéristiques de la langue cible, en se conformant à des choix linguistiques typiques, standard, et donc à être par conséquent moins créatifs » (p. 183).

Et force est de constater que ces « universaux de traduction »¹²⁸ (Baker, 1993) valent également en TA¹²⁹. Pour Yvon (2019), il n'est ainsi pas surprenant que les systèmes de TA reproduisent les biais du *translationese* en les amplifiant dès lors qu'ils sont « entraînés, au moins en partie sur des textes qui sont des traductions » (p. 65) et que les algorithmes répètent ce qu'ils ont appris. À cet égard, Toral (2019) nous invite à considérer le *post-*

¹²⁸ « [T]ranslation universals are hypotheses of linguistic features common to all translated texts regardless of the source and target languages » (Castilho et Resende, 2022, p. 2).

¹²⁹ « [B]ecause MT systems are trained on huge amounts of human-translated-parallel data this is also in line with the normalization translation universal » (De Clercq et al., 2021, p. 40).

editeuse comme une version exacerbée du *translationese* : « We find [...] evidence of post-editeuse, which can be thought of as an exacerbated translationese » (p. 280).

Concrètement, la présence de tels phénomènes en TA signifie que la majeure partie des textes traduits à l'aide de la TA sont plus simples d'un point de vue lexical et plus standardisés qu'en TH ; il y a moins d'inventivité et de créativité, ainsi qu'une plus grande uniformité dans les tournures, expressions usitées et choix des mots comparativement à la TH (Farrell, 2018). Loock (2019) s'interroge sur l'usage de plus en plus systématique d'outils de TA et de TAO :

Si les nouvelles traductions s'inspirent d'anciennes, quelle part reste-t-il pour la créativité ? [...] les traductions fournies par la TA, ainsi que les mémoires de traduction peuvent venir brider [la créativité du traducteur] et in fine standardiser la langue des textes traduits. (pp. 62-63)

L'exemple donné par Deneufbourg (2021) nous permet de mieux comprendre le phénomène de standardisation en TA :

[D]ans un texte comportant des expressions idiomatiques colorées [...] la machine [les] remplacera par des équivalents "explicatifs" – certes corrects, mais plus terre-à-terre. La machine ne cherchera pas à "faire du beau", à chatouiller la poésie du verbe, mais se contentera de rendre le sens. (s.p.)

Plusieurs chercheuses (Castilho et Resende, 2022 ; De Clercq et al., 2021 ; Farrell, 2018 ; Loock, 2019 ; Toral, 2019) ont pu démontrer que ces phénomènes persistaient après PE. C'est ce qui ressort de l'étude menée par Toral (2019) : « In a nutshell, we have found that PEs tend to be simpler and more normalised and to have a higher degree of interference from the source text than HTs » (p. 280). De la même manière, les travaux de Farrell (2018) révèlent un phénomène de standardisation et d'homogénéisation des choix opérés en PE : « What the findings of the primary experiment show [...] is an apparent normalization and homogenization of the choices made by post-editors » (p. 58).

Vers un appauvrissement linguistique

Outre leurs conséquences sur le plan lexical, que nous venons de voir, de tels biais algorithmiques peuvent entraîner des conséquences négatives sur le plan morphologique, comme nous l'expliquent Vanmassenhove et al. (2019) :

Apart from a general effect on lexical diversity, such behavior might also lead to the disappearance or amplified use of certain morphological variants of the same word, accounting, for example, for the already observed over-use of male forms in ambiguous sentences, the preference for certain verb forms over other less frequent ones (3rd person > 1st person), or the difficulties of MT systems to appropriately handle morphologically richer target languages in general. (p. 224)

Dès lors, la persistance de tels écueils en PE combinée à un haut degré d'interférence avec la langue source (Toral, 2019), tant en TA qu'en PE¹³⁰, n'est pas anodine et risquerait même, selon certaines, d'engendrer un appauvrissement de la langue.

C'est d'ailleurs de cette manière que Vanmassenhove et al. (2021) qualifient la langue propre à la TA (*machine translationese*) : « We hypothesize that the “algorithmic bias” [...] could also lead to an artificially impoverished language: “machine translationese” » (p. 2203).

À cet égard, Yvon (2019) compare même les caractéristiques du *machine translationese* – qu'il nomme « e-traductionnais » – à une piètre traduction littérale :

Pour l'essentiel, le "e-traductionnais" présente les mêmes caractéristiques qui le rapprochent d'une traduction littérale : moindre variété lexicale, surreprésentation des vocables très fréquents, tendance à reproduire l'ordre des mots des textes sources, auxquelles s'ajoutent les effets des erreurs propres aux traductions automatiques, relatifs, par exemple, à la faible cohésion lexicale ou à la méconnaissance de termes ou d'idiomes. (p. 65)

Un tel appauvrissement de la langue cible fait partie des craintes ressenties également par Toral (2019) dès lors que le *post-editeuse* ne serait pas épargné par ce phénomène :

Because PEs are simpler and have a higher interference from the source language than HTs, the extensive use of PE rather than HT may have serious implications for the target language in the long term, for example that it becomes impoverished (simplification) and overly influenced by the source language (interference). (p. 280)

Farrell (2018) est aussi de cet avis. Il craint un appauvrissement de la langue cible, tout particulièrement lorsque l'anglais devient la langue de travail privilégiée dans une culture donnée : « Failure to remedy this homogenization may eventually lead to lexical impoverishment of the target language, particularly in cultures where English has become the primary working language in which new written material is created » (*ibid.*, p. 58). G. Deneufbourg (2021) voit également d'un très mauvais œil la standardisation toujours plus grande du langage dans la mesure où cela mènerait à une réduction de notre capacité d'expression, et donc de notre pensée ; il admet l'existence d'un cercle vicieux qui renforcerait cet appauvrissement linguistique :

¹³⁰ « [W]e have shown that these issues cannot be attributed to PE *per se* but that they originate in the MT systems used as the starting point for PE ». (Toral, 2019, p. 280)

Tout indique qu'une utilisation exponentielle de la traduction automatique initie un cercle vicieux qui génère, à terme, un appauvrissement de la langue : la machine produit [...] des textes toujours plus standardisés, qui servent ensuite eux-mêmes de matière première pour alimenter d'autres moteurs, qui raboteront à leur tour les textes, et ainsi de suite [...] Les nouvelles traductions s'inspirent continuellement de ce qui existe déjà, cette mécanique peut, à terme, nuire à l'inventivité, à la créativité, à l'originalité. (*ibid.*, s.p.)

Certes, la post-éditrice pourra chercher à limiter les effets de ces biais en augmentant la richesse lexicale et en atténuant le manque d'originalité et de créativité, mais comme nous le rappelle avec justesse Farrell (2018), ces stratégies vont à l'encontre de la visée première de la PE :

Obviously it would be possible to train post-editors to add originality and inventiveness to their work by purposely editing parts where there are no formal errors, but this clearly defeats the object of post-editing [...] It is evident that the additional post-editing effort required to eliminate what are effectively MT markers is likely to nullify a great deal, if not all, of the time and cost-saving advantages of PEMT. (p. 50 et p. 58)

6.2 Analyse qualitative des erreurs

Dans cette partie, nous présentons les résultats de l'analyse qualitative comparative des erreurs contenues dans les TH et dans les PE. Indépendamment de l'analyse quantitative, nous avons procédé à une analyse linguistique de notre corpus dans le but de parvenir à déceler les effets de la (PE de) TA neuronale sur la qualité finale en observant simplement les productions des étudiantes et en comparant les propositions de traduction pour un même passage. Nous formulons tout d'abord quelques observations générales avant de détailler et d'exemplifier les effets observés. Dans un souci de clarté, nous avons décidé de répartir ces effets en deux grandes catégories : effets positifs et effets négatifs. Nos observations sont classées en sous-catégories, chacune étant illustrée par un ou plusieurs exemples tirés de nos corpus d'expérience 2018 et 2021. Bien que chaque expérience ait été faite pour trois modes de traduction, nous avons décidé de ne pas faire figurer d'exemples tirés des PE de TA dite statistique. Nous nous concentrons uniquement sur la comparaison de ce qui nous paraît le plus pertinent dans le contexte actuel : la TH et la PE de TA neuronale (*Google Traduction 2020* et *DeepL 2018 & 2020*).

6.2.1 Observations générales

Des erreurs imputables à la TA

Dans nos expériences, nous sommes arrivées au même constat que Daems (2016) dans son étude comparative entre la TH et la PE de TA statistique : « most student PE errors do find

their origin in MT errors » (p. 64)¹³¹. En effet, nous pouvons affirmer que la grande majorité des erreurs commises en PE (tant celles communes avec la TH que celles spécifiques à ce mode) était déjà présente dans la TA brute. Cette constatation nous incite à penser que les étudiantes auraient pu post-éditer davantage la sortie de TA et qu'il serait dès lors bénéfique de les former à repérer les erreurs récurrentes en TA (voir chapitre 8).

Des erreurs communes

Même si l'analyse statistique des productions (chapitres 3 et 4) a révélé des divergences significatives dans la distribution des erreurs entre la TH et la PE, nous avons toutefois repéré des erreurs communes aux deux modes de traduction, notamment des calques fautifs en PE comme en TH pour les trois TS.

Des erreurs variées

De manière générale, nous avons observé une plus grande diversité d'erreurs en TH, ce qui corrobore les résultats de l'étude comparative (TH et PE) de Martikainen (2019c) menée avec des apprenantes novices en traduction médicale : « dans ces deux types de textes produits par des apprenants novices en traduction médicale, [...] la traduction humaine se caractérise par la plus grande variété des erreurs et des sources de distorsion potentielles » (p. 301).

6.2.2 Effets de la (PE de) TA neuronale sur la qualité finale

6.2.2.1 Effets positifs

Commençons par parcourir ce que nous avons décelé comme effets positifs de la (PE de) TA neuronale sur la qualité linguistique des productions recueillies. Nous pouvons affirmer que dans cette étude, la (PE de) TA neuronale a permis aux étudiantes d'éviter un certain nombre d'erreurs sur le plan non seulement formel, mais aussi sémantique. Précisons qu'il s'agit d'une sélection subjective et non exhaustive d'éléments jugés pertinents.

6.2.2.1.1 Calques fautifs évités

Même si l'analyse statistique a révélé qu'il y avait proportionnellement davantage de calques en PE qu'en TH (cf. chapitres 4 et 5) dans notre corpus, les propositions de la TA

¹³¹ « The high number of wrong collocation errors, word sense errors, article and compound errors (the four categories where post-editing scored worse than human translation) for the first experiment can all be explained by a high number of errors in the MT output for the same passages. The same holds true for the article and compound errors found in the second experiment. It would seem that student post-editors for general texts would benefit most from extra training in spotting and solving wrong collocations as well as word sense errors » (Daems, 2016, p. 64).

neuronale ont tout de même permis aux étudiantes d'éviter certains calques fautifs commis en TH (Tableau 55, Tableau 56 et Tableau 57).

Texte source 1	<i>It recognises that the referendum decision must be implemented, but also that to implement it with gung-ho gusto, of the kind that Boris Johnson could provide, would be perilously divisive.</i>			
DeepL 2018 & 2020 Google 2020	... doit être mise en œuvre.			
Traduction correcte	... doit être mise en œuvre.			
Traduction fautive	... doit être implémentée.			
Occurrence de l'erreur	2018	TH : 25 % (2 sur 8) PE de DeepL : 0 % (0 sur 8)	2021	TH : 22 % (2 sur 9) PE de Google : 0 % (0 sur 8) PE de DeepL : 0 % (0 sur 7)

Tableau 55 : Calques fautifs évités (1)

Ce premier exemple est un calque fautif du verbe *to implement*. Effectivement, l'emploi dans ce contexte du terme « implémenter » dans le sens de « mettre en œuvre » constitue en français une impropriété lexicale ; « implémenter » et « implémentation » étant des termes réservés au domaine de l'informatique¹³². Deux TH du corpus 2018 et deux TH du corpus 2021 contenaient cette erreur.

Texte source 1	<i>With two resignations in one week, four Westminster watchers discuss whether the Conservative leader and her party can cling on to power.</i>		
DeepL 2018 & 2020 Google 2020	Avec deux démissions en une semaine...		
Traduction correcte	Avec deux démissions ...		
Traduction fautive	Avec deux résignations en une semaine...		
Occurrence de l'erreur	2018	TH : 25 % (2 sur 8) PE de DeepL : 0 % (0 sur 8)	

Tableau 56 : Calques fautifs évités (2)

¹³² Implémenter : « réaliser la phase finale d'élaboration d'un système informatique, afin de le rendre fonctionnel » (Office québécois de la langue française, 2003, s.p.).

L'exemple suivant illustre une erreur due au calque du terme anglais *resignation*. Ce substantif a été traduit de manière littérale par « résignation » en français, ce qui entraîne un glissement de sens dans le TC. Cette erreur se retrouve pourtant dans deux productions humaines du corpus 2018. Ce calque est en revanche totalement absent du corpus 2021.

<i>Texte source</i> 3	<i>Of the various manifestations of synesthesia, the most common involves seeing monochromatic letters, digits and words in unique colors.</i>			
<i>TA brute</i>	<i>DeepL</i> 2018 & 2020	...des lettres, des chiffres et des mots monochromes ...		
	<i>Google</i> 2020	...des lettres, des chiffres et des mots monochromatiques ...		
<i>Traduction correcte</i>	...des lettres, des chiffres et des mots monochromes ...			
<i>Traduction fautive</i>	... monochromatiques ...			
<i>Occurrence de l'erreur</i>	2018	TH : 50 % (4 sur 8) PE de <i>DeepL</i> : 0 % (0 sur 8)	2021	TH : 29 % (2 sur 7) PE de <i>Google</i> : 78 % (7 sur 9) PE de <i>DeepL</i> : 12,5 % (1 sur 8)

Tableau 57 : Calques fautifs évités (3)

Dans ce dernier exemple, l'adjectif *monochromatic* a été traduit de manière calquée dans plusieurs productions humaines, ainsi que dans sept PE de *Google*, tandis que la TA brute de *DeepL* en proposait une traduction adéquate en 2018 comme en 2020. L'adjectif « monochromatique » existe bel et bien, mais il s'agit d'un terme vieilli (Le Petit Robert, s.d.) auquel on préférera l'emploi de l'adjectif « monochrome ». Remarquons toutefois qu'en PE de *Google*, deux étudiantes ont modifié cet adjectif et ont évité le calque.

6.2.2.1.2 Régionalismes évités

Le recours à la TA neuronale a également permis à certaines étudiantes de ne pas faire usage de régionalismes (Tableau 58).

<i>Texte source</i> 3	<i>Does the note B taste like horseradish?</i>			
<i>DeepL</i> 2018 & 2020 <i>Google</i> 2020	... a un/le goût de raifort ?			
<i>Traduction</i> <i>correcte</i>	... a le/un goût du/de raifort ?			
<i>Traduction</i> <i>fautive</i>	... goûte le raifort ?			
<i>Occurrence</i> <i>de l'erreur</i>	2018	TH : 37,5 % (3 sur 8) PE de <i>DeepL</i> : 0 % (0 sur 9)	2021	TH : 29 % (2 sur 7) PE de <i>Google</i> : 11 % (1 sur 9) PE de <i>DeepL</i> : 0 % (0 sur 8)

Tableau 58 : Régionalismes évités

Dans cet exemple, l'usage transitif du verbe « goûter » dans le sens « avoir le goût de », « avoir une saveur de » constitue un belgicisme (Druide, 2022) qu'il convenait d'éviter. Sur nos deux corpus, six productions présentaient ce régionalisme, il s'agit de cinq TH et d'une PE de *Google*.

6.2.2.1.3 Maîtrise des règles d'usage du français (grammaire et orthographe)

Comme nous l'avions formulé dans la sous-hypothèse SHA2.4, il apparaît que la nouvelle génération de TA applique la grande majorité des règles grammaticales et orthographiques (de la langue française en l'occurrence), ce qui débouche sur des traductions qui contiennent relativement peu d'erreurs de grammaire et d'orthographe, voire pas du tout. Nous ne pouvons néanmoins pas en dire autant des productions humaines récoltées dans notre corpus (Tableau 59 et Tableau 60).

<i>Texte source 1</i>	<i>The time may well come when in order to achieve Brexit, a Gordian knot needs to be cut. But neither the country nor the party is ready for that.</i>			
<i>DeepL 2018 & 2020 Google 2020</i>	Mais ni le pays ni le parti ne sont prêts pour cela.			
<i>Traduction correcte</i>	Mais ni le pays ni le parti ne sont prêts ...			
<i>Traduction fautive</i>	Mais ni le pays ni le parti n'est prêt ...			
<i>Occurrence de l'erreur</i>	2018	TH : 50 % (4 sur 8) PE de <i>DeepL</i> : 0 % (0 sur 8)	2021	TH : 33 % (3 sur 9) PE de <i>Google</i> : 12,5 % (1 sur 8) PE de <i>DeepL</i> : 0 % (0 sur 7)

Tableau 59 : Maîtrise des règles d'usage du français (1)

Dans ce premier exemple, l'accord du verbe avec des sujets unis par la conjonction « ni » suit la règle générale d'accord avec des sujets coordonnés, qui est d'accorder avec l'ensemble des sujets, c'est-à-dire de mettre le verbe au pluriel même si les sujets sont au singulier (Office québécois de la langue française, 2023). Dans nos deux corpus, sept productions humaines présentent cette erreur de grammaire et une seule PE de *Google*.

<i>Texte source 3</i>	<i>What makes synesthesia different from drug-induced hallucinations is that synesthetic sensations are highly consistent</i>			
<i>DeepL 2018 & 2020 Google 2020</i>	Ce qui différencie la synesthésie...			
<i>Traduction correcte</i>	Ce qui différencie la synesthésie...			
<i>Traduction fautive</i>	Ce qui différentie la synesthésie...			
<i>Occurrence de l'erreur</i>	2018	TH : 0 % (0 sur 8) PE de <i>DeepL</i> : 0 % (0 sur 8)	2021	TH : 14 % (1 sur 7) PE de <i>Google</i> : 0 % (0 sur 9) PE de <i>DeepL</i> : 0 % (0 sur 8)

Tableau 60 : Maîtrise des règles d'usage du français (2)

Dans ce deuxième exemple, le terme « différencie » a été mal orthographié dans une TH en 2021, alors que cette faute d'orthographe n'apparaît dans aucune PE.

6.2.2.2 Effets négatifs

L'influence de la TA neuronale a également engendré divers types d'erreurs, plus ou moins graves, qui se retrouvent pour la majeure partie dans les productions finales des étudiantes, soit après PE. Nous avons sélectionné les failles qui nous paraissent les plus symptomatiques.

6.2.2.2.1 Marques de littéralité

On note une forte tendance de la technologie neuronale à traduire littéralement : « despite the [...] recent advances in the field, machine translations remain quite literal » (Poibeau, 2022, p. 6019). Ces traductions littérales de l'anglais en l'occurrence ont pour effet de mener à des problèmes de terminologie et d'idiomaticité entravant la fluidité générale du texte final, voire à des erreurs plus graves en traduction que sont les calques fautifs, les glissements de sens¹³³ ou encore les contresens¹³⁴. En effet, nous avons constaté qu'en post-éditant, les étudiantes n'ont généralement pas modifié les propositions de la machine qui leur semblaient d'emblée satisfaisantes. *Elles* ont été enclines à accepter des tournures maladroites, non idiomatiques ou fautives qui sont très souvent le fruit d'une traduction littérale. Ce constat est corroboré par les observations de Loock et Léchaugnette (2021) : « [students] seem to be heavily primed by the MT output (even in the case of very awkward direct calques) » (p. 214). Depraetere (2010, p. 6) avait également déjà remarqué dans son expérience qu'en post-éditant, les étudiantes voyaient généralement peu d'inconvénients à conserver des traductions qui ne sont pas adéquates alors que d'autres possibilités plus idiomatiques existent. De même, les conclusions de l'étude comparative menée par Martikainen et Kübler (2016) rendent compte d'une « tendance à la traduction par équivalence formelle et littérale » (p. 8) en PE par rapport à la TH, ce qu'elles considèrent comme révélateur de l'interférence de la TA dans le processus de traduction. Il y a quelques années, l'étude menée par Loock (2018) sur un double corpus de français original et de français traduit automatiquement à partir de l'anglais, a montré que les deux moteurs de TA sélectionnés comme outils (en l'occurrence *DeepL* et *MT@EC/eTranslation*) ne tenaient pas compte de « l'usage linguistique, au-delà des règles et dont le respect permet d'atteindre la fluidité et l'idiomaticité de la langue cible attendues sur le marché » (p. 792). Tout comme la langue source exerce une influence sur la langue cible en TH (cf. concept de *shining through* chez Teich (2003)), la TA va influencer la post-éditrice et laisser son empreinte dans la langue post-éditée (voir section 8.3.5).

¹³³ Pour rappel, par « glissement de sens », nous entendons une ambiguïté sémantique, une perte partielle, voire totale, de sens d'un segment du TS sans pour autant que cela mène à un contresens ou à un non-sens.

¹³⁴ Delisle définit le contresens comme une « faute de traduction qui consiste à attribuer à un segment du texte de départ un sens contraire à celui qu'a voulu exprimer l'auteur » (p. 649).

6.2.2.1.1 Marques de littéralité et calques fautifs

Bien que nous venions de montrer dans la section précédente que certains calques fautifs ont été évités grâce à la présence de la TA brute, nous tenons à souligner que ce mode de traduction aura davantage eu l'effet inverse dans nos expériences. De plus, cette observation a été démontrée dans l'analyse statistique : les étudiantes ont commis significativement plus de calques fautifs en PE de TA neuronale qu'elles n'en ont commis en TH. Nous jugeons que ce constat confirme que l'un des défauts majeurs de la TA est de procéder principalement par calque (Martikainen, 2020).

<i>Texte source 1</i>	<i>This summer she offered the British people the chance to turn her into an elected dictator, and they decided in their wisdom they would rather keep her as a prime minister who must operate with a degree of tact.</i>			
<i>DeepL 2018 & 2020 Google 2020</i>	Cet été, elle a offert au peuple britannique la chance de la transformer en dictateur élu...			
<i>Traduction fautive</i>	... offert au peuple britannique la chance ...			
<i>Traduction correcte</i>	... elle a offert au peuple britannique l'opportunité de elle a donné au peuple britannique la possibilité de ...			
<i>Occurrence de l'erreur</i>	2018	PE de <i>DeepL</i> : 75 % (6 sur 8) TH : 12,5 % (1 sur 8)	2021	PE de <i>Google</i> : 87,5 % (7 sur 8) PE de <i>DeepL</i> : 0 % (0 sur 7) TH : 11 % (1 sur 9)

Tableau 61 : Marques de littéralité et calques fautifs

Cet exemple concerne la traduction de la tournure *to offer someone the chance to do something*. Il convenait d'éviter une traduction calquée de type « offrir à quelqu'un la chance de faire quelque chose » puisque l'acception courante du mot « chance » au singulier en français a une connotation positive favorable et conserve une notion de hasard contrairement à l'acception anglaise dans notre contexte : possibilité, occasion. Il aurait fallu choisir des tournures qui ne soient pas calquées, telles que celles qui ont été adoptées dans la plupart des TH ou celles qui ont été correctement post-éditées : « donner la possibilité de » ou encore « offrir l'opportunité de ». Notons toutefois que l'emploi du terme « opportunité » dans le sens d'« occasion (favorable) » ou de « circonstance opportune » reste critiqué même s'il est aujourd'hui bien implanté dans l'usage (Druide, 2022). Dans nos deux corpus, la proposition donnée en TA brute aurait dû être modifiée par l'étudiante, mais elle ne l'a pas été dans 13 PE sur 23 ; on retrouve également cette tournure calquée dans deux TH. Alors qu'en 2018, 75 % des PE de *DeepL* présentaient ce

calque, ce n'est plus du tout le cas en 2021. En 2021 toujours, les étudiantes de 2^e master ont manifestement mieux post-édité ce passage en TA de *DeepL* qu'en TA de *Google*, sans que nous puissions avancer d'explications.

6.2.2.2.1.2 Marques de littéralité et glissements de sens ou contresens

Comme nous l'avons mentionné plus haut, une traduction littérale peut également mener à un glissement de sens, voire à un contresens, en langue cible, que ce soit en PE ou en TH.

<i>Texte source 2</i>	<i>A number of other emerging markets will also take centre stage – Mexico could be larger than the UK and Germany by 2050 in PPP terms and six of the seven largest economies in the world could be emerging markets by that time.</i>			
<i>DeepL 2018 & Google 2020</i>	...six des sept plus grandes économies du monde pourraient être des marchés émergents d'ici là.			
<i>DeepL 2020</i>	...six des sept plus grandes économies du monde pourraient être des marchés émergents à cette date.			
<i>Traduction fautive</i>	...six des sept plus grandes économies du monde pourraient être des marchés émergents d'ici là/à cette date.			
<i>Traduction correcte</i>	Il se pourrait que six des sept plus grosses économies au monde soient des marchés actuellement émergents.			
<i>Occurrence de l'erreur</i>	2018	PE de <i>DeepL</i> : 87,5 % (7 sur 8) TH : 22 % (2 sur 9)	2021	PE de <i>Google</i> : 57 % (4 sur 7) PE de <i>DeepL</i> : 44 % (4 sur 9) TH : 0 % (0 sur 8)

Tableau 62 : Marques de littéralité et glissements de sens et contresens

Dans cet exemple tiré du TS économique, la traduction littérale que proposent les moteurs de TA engendre un contresens en français qui crée une rupture de cohérence. Il était absolument nécessaire de déverbaliser ce segment pour en dégager le sens et pour garantir la cohérence discursive. Dans la PE commune réalisée en classe après l'expérience, nous sommes arrivés à la traduction suivante : « le top sept des économies pourrait d'ici là compter six marchés émergents actuels ».

Notons que sur les deux expériences, la proposition littérale de la machine n'a pas du tout été modifiée par les étudiantes dans 15 PE sur 22. En outre, les étudiantes de 2^e master (2021) ont davantage post-édité la TAN brute par rapport aux étudiantes de 3^e bachelier (2018), sans que cela ne signifie pour autant que ces modifications aient permis d'aboutir à une traduction satisfaisante. En effet, la traduction de ce segment a manifestement posé

énormément de difficultés aux étudiantes puisque la seule proposition jugée sémantiquement correcte par les évaluatrices provient d'une TH en 2021.

6.2.2.2.1.3 Marques de littéralité et compléments circonstanciels

En TA, les compléments circonstanciels (temps, lieu, etc.) sont pratiquement systématiquement traduits à la même position que dans le TS, ce qui crée très souvent des problèmes de fluidité et peut aller à l'encontre de l'usage linguistique en français. En effet, alors que l'anglais préfère conserver l'ordre grammatical, l'usage linguistique en français veut que ce type de compléments soient en général logiquement placés en tête de phrase (Bonnard, 1981, p. 125). Il appartient alors à la *post-éditrice* de tenir compte de cette tendance à l'antéposition des compléments circonstanciels en français puisqu'elle participe de la meilleure fluidité du texte cible.

Texte source 2	<i>But this will require patience to ride out the storms we have seen recently in economies like, for example, Brazil, Nigeria and Turkey, all of which still have considerable long-term economic potential based on our analysis.</i>			
TA brute	DeepL 2018	... qui ont toutes encore un potentiel économique considérable à long terme sur la base de notre analyse .		
	Google 2020	... qui ont toutes encore un potentiel économique considérable à long terme basé sur notre analyse .		
	DeepL 2020	... qui, d'après notre analyse , ont encore un potentiel économique considérable à long terme.		
Traduction à éviter	...un potentiel économique considérable à long terme sur la base de notre analyse .			
Traduction à privilégier	Des marchés qui, selon notre analyse , conservent... D'après nos analyses, ces dernières...			
Occurrence	2018	PE de DeepL : 87,5 % (7 sur 8) TH : 56 % (5 sur 9)	2021	PE de Google : 33 % (3 sur 9) PE de DeepL : 12,5 % (1 sur 8) TH : 43 % (3 sur 7)

Tableau 63 : La place des compléments circonstanciels

Dans l'exemple repris dans le Tableau 63, le complément « based on our analysis » a été laissé en toute fin de phrase dans plusieurs TH et dans la majorité des PE de DeepL en 2018. En 2021, les étudiantes de 2^e master ont davantage veillé à respecter cette tendance à l'antéposition des compléments circonstanciels puisque seules trois PE de Google sur neuf auraient dû faire l'objet d'une modification à cet endroit et ne l'ont pas été. Dans toutes les autres productions, les étudiantes ont naturellement placé l'élément circonstanciel soit

en début de phrase, soit en incise, rendant la phrase entière plus fluide. Remarquons enfin que *DeepL* s'est amélioré puisque dans la TA proposée en 2020, le complément est mis en incise ; l'intervention de *la post-éditrice* n'était ainsi plus nécessaire.

6.2.2.2.1.4 Ponctuation calquée

Outre des calques lexicaux et syntaxiques, il est arrivé dans nos deux expériences que les étudiantes conservent les signes de ponctuation ayant été calqués de la langue source par la machine, en ne respectant pas forcément les règles d'usage du français. Il est connu que l'un des écueils des moteurs de TA génériques est de recopier en grande partie la ponctuation du TS sans tenir compte de l'ensemble des règles de ponctuation en vigueur dans la langue cible. Un exemple classique pour la combinaison de langues que nous étudions est la non-conversion des guillemets anglais (doubles apostrophes) en guillemets français (doubles chevrons). Et il ressort de l'étude menée par Castilho et Resende (2022) que ce défaut est souvent encore présent après PE : « we noted that there were differences in punctuation counts between translated versions [TH] and source text and that the PE versions tended to follow the punctuation of the MT output » (p. 12).

Texte source 2	<i>A number of other emerging markets will also take centre stage – Mexico could be larger than the UK and Germany by 2050 in PPP terms and six of the seven largest economies in the world could be emerging markets by that time.</i>			
TA brute	<i>DeepL</i> 2018	Un certain nombre d'autres marchés émergents occuperont également une place centrale - le Mexique pourrait être		
	<i>Google</i> 2020	Un certain nombre d'autres marchés émergents occuperont également le devant de la scène - le Mexique pourrait être		
	<i>DeepL</i> 2020	Un certain nombre d'autres marchés émergents occuperont également le devant de la scène : le Mexique pourrait être		
Emploi fautif	Un certain nombre d'autres marchés émergents occuperont également le devant de la scène – d'ici 2050, le Mexique pourrait être...			
Emploi correct	Un certain nombre d'autres marchés émergents se retrouvera également sur le devant de la scène : d'ici 2050, le Mexique pourrait dépasser...			
Occurrence de l'erreur	2018	PE de <i>DeepL</i> : 62,5 % (5 sur 8) TH : 33 % (3 sur 9)	2021	PE de <i>Google</i> : 57 % (4 sur 7) PE de <i>DeepL</i> : 0 % (0 sur 9) TH : 25 % (2 sur 8)

Tableau 64 : Ponctuation calquée

Penchons-nous, avec cet exemple, sur l'usage du tiret cadratin (Tableau 64). En anglais, comme en français, ce signe de ponctuation peut avoir plusieurs fonctions. Son emploi dans

l'exemple qui nous intéresse permet notamment de signaler « other kinds of break in a sentence where a comma, semicolon, or colon would be traditionally used » (Oxford University Press, 2021). Or, cet usage ne correspond à aucun usage de ce signe de ponctuation en français (voir outils d'aide à la rédaction fiche explicative sur TERMIUM Plus^{®135}) puisque contrairement à l'usage courant en anglais, le tiret cadratin en français « capte l'attention du lecteur. Il est le signe tout désigné pour souligner une opposition ou une conclusion inattendue, pour créer un effet de chute ou d'insistance » (Portail linguistique du Canada, 2022). Il convenait dès lors de le rendre d'une autre manière, par exemple par un autre signe de ponctuation adéquat ou encore en créant une incise. Dans nos deux corpus, le tiret cadratin du TS a été reproduit dans plus de la moitié des PE (à l'exception des PE de *DeepL* en 2021), ainsi que dans quelques TH. Il apparaît que le moteur *DeepL* ne propose pas la même ponctuation en septembre 2020 qu'en 2018 ; le tiret cadratin est en effet remplacé par le deux-points. Par conséquent, on ne retrouve pas du tout cet emploi inadéquat du tiret cadratin dans les PE de *DeepL* en 2021.

6.2.2.2 Effacement d'informations

Il n'est pas rare d'observer des effacements d'informations en TA (Loock, 2019). En effet, il arrive que pour diverses raisons qui nous sont inconnues, un terme, un syntagme, une partie de phrase voire toute une phrase du TS, n'apparaisse pas dans la traduction générée par le moteur. Et souvent, cet effacement d'informations est également relevé dans le texte post-édité par l'apprenante, car cette dernière ne le repère très probablement pas (Depraetere, 2010¹³⁶) et ne retourne pas au TS lorsqu'elle post-édite.

<i>Texte source 2</i>	<i>To realise this growth potential, emerging market governments need to implement structural reforms to improve macroeconomic stability, diversify their economies away from undue reliance on natural resources (where this is currently the case), and develop more effective political and legal institutions.</i>
<i>DeepL 2018</i>	... diversifier leurs économies et *** de développer des institutions politiques et juridiques plus efficaces.
<i>Traduction fautive</i>	... diversifier leurs économies et *** de développer...
<i>Occurrence de l'erreur</i>	PE de <i>DeepL</i> : 50 % (4 sur 8) TH : 11 % (1 sur 9)

Tableau 65 : Effacement d'informations

¹³⁵<https://www.btb.termiumpius.gc.ca/>

¹³⁶ « Sometimes, the machine fails to translate part of the source text. On a number of occasions, the majority of students failed to notice information loss » (Depraetere, 2010, p. 5).

Nous pouvons constater dans le Tableau 65 qu'en 2018, la TA n'a pas fourni de proposition pour tout un segment de phrase et que seules 50 % des étudiantes en PE ont pris la peine de corriger cet effacement en retournant au TS et en proposant une traduction du segment manquant. À nouveau, le moteur de TA semble avoir progressé puisqu'en 2020, il n'y a plus d'effacement d'informations ; l'élément de phrase a cette fois été traduit par *DeepL* et par *Google Traduction*.

6.2.2.2.3 Rupture de cohésion textuelle et biais de genre

La résolution des ruptures de cohésion textuelle et la reproduction des biais de genre constituent encore aujourd'hui des limites majeures en TA. Nous traitons ces deux aspects dans le même point, car nous jugeons qu'ils sont interdépendants : une rupture de cohésion textuelle peut entraîner l'apparition d'un biais de genre, et vice versa. Effectivement, « [l]es limitations de la traduction automatique ne se réduisent pas à leur incapacité à prendre en charge certains phénomènes linguistiques ; un autre problème important est l'existence de biais systématiques, en particulier de genre » (Wisniewski et al., 2021, pp. 11-12). Il a été démontré que puisque le moteur de TA recycle les solutions les plus fréquentes, il reproduit de ce fait les biais de genre présents dans les corpus d'entraînement (Martikainen, 2020). Pour illustrer cette problématique, prenons l'un des exemples les plus connus, celui de la traduction du substantif neutre anglais *a nurse* auquel un moteur de TA va presque systématiquement assigner le genre féminin (Wisniewski et al., 2021). Ainsi, sans indication explicite de genre, une phrase telle que « *the nurse is cleaning the room* » sera rendue, dans la plupart des cas et par la majorité des moteurs par « l'infirmière nettoie la chambre », reproduisant, voire exacerbant, de ce fait les stéréotypes (de genre et autres) dont nos sociétés ont déjà tant de mal à s'affranchir. En témoigne cet exemple (Figure 68) donné par Vanmassenhove (2020) généré avec *Google Traduction* en mai 2019 :

EN N.: I am smart.

FR M.: Je suis **intelligent**.

EN N.: I am beautiful.

FR M.: Je suis **beau**.

EN N.: I am beautiful but not smart.

ES F.: Je suis **belle** mais pas **intelligente**.

Figure 68 : Biais de genre (Vanmassenhove, 2020, p. 132)

Comme l'expliquent Wisniewski et al. (2021), ce problème de transfert d'informations d'une langue à une autre est dû en partie à certaines spécificités linguistiques propres à chacune de ces langues : « the transfer of gender information from French, where

grammatical gender is a property of all nouns, and agreement rules exist within the noun phrase, to English, where gender is only overtly used in rare constructs involving human agents and pronoun coreference » (p. 311). Alors qu'il est relativement aisé de résoudre ce type de problèmes en TH – de l'anglais vers le français en l'occurrence – puisque la traductrice humaine va généralement pouvoir s'appuyer sur le contexte pour déduire le genre de ce qu'il doit traduire et déterminer l'accord morphologique adéquat, ce n'est pas le cas en TA. Étant donné que les systèmes neuronaux se basent sur des probabilités d'occurrences (Forcada, 2017)¹³⁷, il est impossible pour ces algorithmes non conscients d'en faire autant, puisqu'ils ne font essentiellement que reproduire les structures récurrentes apprises à partir de larges corpus d'entraînement (Moorkens et al., 2018b, p. 256).

Human translators rely on contextual information to infer the gender of the speaker in order to make the correct morphological agreement. However, most current MT systems do not; they simply exploit statistical dependencies on the sentence level that have been learned from large amounts of parallel data. (Vanmassenhove, 2020, p. 130)

Cette perte d'informations contextuelles est due également au principe de fonctionnement même des systèmes puisqu' « en TA, l'unité de base reste la phrase. Le rapport qu'elle entretient avec celles qui la précèdent et celles qui la suivent se limite à une recherche de l'entourage lexical le plus probable » (Grass, 2022, p. 10).

[S]entences are translated in isolation. As a consequence, pieces of information necessary to determine the gender of the speakers might be lost. In such cases the MT system will opt for the statistically most likely variant, which depending on the training data, will either be the male or the female form. (Vanmassenhove, 2020, p. 130)

La TA est donc, logiquement, relativement mauvaise en résolution d'anaphores, notamment pronominales (Poibeau, 2022¹³⁸ ; Sennrich, 2018¹³⁹) :

¹³⁷ « In both NMT and SMT, a target sentence is a translation of a source sentence with a certain probability of likelihood; in principle, all target sentences can be a translation of a source sentence, but we are interested in the most likely one » (Forcada, 2017, p. 300).

¹³⁸ « It is also clear that most current systems work at the sentence level and are, for instance, very bad at dealing with pronouns, when the source and the target language differ in this respect [...] Note however some preliminary attempts to address the problem, like in (Luong and Popescu-Belis, 2016) or (Fu et al., 2019) » (Poibeau, 2022, p. 6020).

¹³⁹ « Most MT systems at the sentence-level do not have access to adequate context that may be required for the translation of pronouns » (Sennrich, 2018, cité dans Jwalapuram et al., 2020, p. 2270).

For example, when the source language only has one gender and the target language different pronouns for masculine and feminine, one gender has to be chosen, often at random, by the translation system, which is thus often wrong. This kind of mistake is frequent for any decision the system has to make that involves a context larger than a single sentence. (Poibeau, 2022., p. 6020)

Aussi, la correction de ce type de défaillances algorithmiques est aussi complexe qu'indispensable, d'autant plus que Bender et al. (2021), Vanmassenhove et al. (2018) et Vanmassenhove (2020) ont montré dans leurs travaux que les modèles de langue ne font pas que reproduire les biais de genre présents dans les corpus d'entraînement, ils renforcent aussi leur apparition. Si l'on peut dire que ces problèmes sont aujourd'hui partiellement résolus dans le cas de constructions peu complexes, il n'en reste pas moins que certains écueils persistent, malgré les innombrables tentatives d'améliorations, comme le montre l'exemple tiré de notre corpus et repris dans le Tableau 66.

Texte source 1	<i>Theresa May is in a stronger position than the press is willing to admit. Before the election, her frailties were ignored. Since that time, a narrative of extreme vulnerability has taken hold. Yet she remains in Downing Street, and the loss of two ministers who were found in different ways to have misbehaved does not change the powerful reasons for keeping her there. This summer she offered the British people the chance to turn her into an elected dictator.</i>			
DeepL 2018 & 2020 Google 2020	Theresa May [...] elle a offert au peuple britannique la chance de la transformer en dictateur élu.			
Traduction à éviter	... de la transformer en (un) dictateur élu.			
Traduction à privilégier	... d'en faire une dictatrice élue.			
Occurrence	2018	PE de DeepL : 87,5 % (7 sur 8) TH : 50 % (4 sur 8)	2021	PE de Google : 50 % (4 sur 8) PE de DeepL : 57 % (4 sur 7) TH : 56 % (5 sur 9)

Tableau 66 : Ruptures de cohésion textuelle et biais de genre

Étant donné que le substantif anglais *dictator* ne contient pas de marqueur de genre, le moteur de TA a assigné le genre masculin à l'ensemble du syntagme nominal *an elected dictator*, faisant fi des nombreux référents anaphoriques (surlignés en gras dans la phrase en anglais) : *Theresa May* ; *she* ; *turn her into*. Il apparaît qu'en 2018, quatre étudiantes sur huit en TH ont judicieusement accordé ce syntagme au féminin contre seulement un sur huit en PE de DeepL, les sept autres étudiantes ayant décidé de conserver telle quelle la sortie de la TA. En 2021, les étudiantes de 2^e master ont davantage post-édité la TA brute

puisque le syntagme a été mis au féminin dans sept PE sur quinze. En revanche, plus d'une étudiante sur deux en TH a tout de même choisi de traduire ce syntagme à la forme masculine. Cet usage dit générique de la forme masculine, alors qu'une forme féminine existe, n'a cependant pas été jugé fautif par nos évaluatrices.

6.2.2.2.4 Rupture de cohérence - Irrationalité

Nous avons observé certaines ruptures de cohérence textuelle imputables à la TA. En effet, tout comme la TA statistique, la TA neuronale continue à produire, parfois, des erreurs typiquement « machine », c.-à-d. des erreurs qu'un traducteur humain doué de raison ne commettrait normalement pas. Nous qualifions ces traductions d'« irrationnelles » dans le sens où elles s'opposent à la raison, elles n'obéissent pas et ne sont pas conformes au bon sens, à la logique (CNRTL, s.d.). Ces traductions irrationnelles ont pour effet de trahir la présence de la TA en révélant à la lectrice finale que le texte qu'elle a sous les yeux n'a pas été traduit par un humain. Ce phénomène entre ainsi en complète contradiction avec les exigences actuelles du marché de la traduction qui, sauf exception, promeut l'invisibilité de la traductrice en exigeant une fluidité et une idiomatité maximales (Loock, 2018), de telle sorte que la lectrice ne puisse pas se rendre compte qu'il s'agit d'une traduction. Ces traductions entraînant des problèmes évidents d'intelligibilité ou de compréhensibilité¹⁴⁰ du texte, la post-éditrice doit impérativement y remédier en tant qu'unique garante de la cohérence textuelle (Martikainen, 2019d), ce qui fut le cas dans l'exemple repris au Tableau 67.

<i>Texte source 1</i>	<i>Andrew Gimson: There's no replacement for May nor appetite for election.</i>	
<i>TA brute</i>	<i>DeepL</i> 2018 & 2020	Andrew Gimson: Il n'y a pas de remplaçant pour le mois de mai ni d'appétit pour les élections.
	<i>Google</i> 2020	Andrew Gimson: Il n'y a pas de remplaçant pour May ni d'appétit pour les élections
<i>PE correcte</i>	Il n'y a pas de remplaçant potentiel pour T. May ...	
<i>TH correcte</i>	Il n'y a ni de remplaçant prévu pour Theresa May ...	
<i>Occurrence de l'erreur</i>	/	

Tableau 67 : Irrationalité

¹⁴⁰ « Il m'apparaît que les notions d'intelligibilité et de compréhensibilité sont très proches, mais que la première semble conçue en examinant prioritairement la relation texte-lecteur tandis que la deuxième traduit davantage le point de vue lecteur-texte. Cette divergence de points de vue vient sans doute de ce que les chercheurs en intelligibilité proviennent des sciences du langage et les chercheurs en compréhensibilité, des sciences de l'éducation et de la psychologie cognitive » (Beaudet, 2001, p. 11).

Cet exemple illustre un cas de rupture de cohérence textuelle qui a été corrigé en PE. Le nom propre « May » a été traduit automatiquement par « le mois de mai », ce qui est totalement irrationnel dans ce contexte textuel, d'autant plus que dans le segment qui précède, le nom complet – « Theresa May » – apparaît. Il est surprenant de constater que *DeepL* propose toujours cette traduction fautive en septembre 2020 alors que ce n'est pas le cas de *Google Traduction*. Dans nos deux corpus 2018 et 2021, aucune production finale ne contient cette traduction irrationnelle, ce qui signifie que toutes les TA brutes ont été correctement modifiées par les étudiantes. Précisons qu'en date du 29 mars 2023, *DeepL* ne commettait plus cette rupture de cohérence textuelle.

6.2.2.2.5 Incohérence temporelle

Nous avons remarqué à plusieurs reprises que la TA neuronale a introduit des problèmes sur le plan de la cohérence temporelle (pertinence des temps verbaux) qui, très souvent, n'ont pas été corrigés durant la PE. Nous supposons que la TA n'analyse pas encore suffisamment le texte dans son ensemble ou qu'elle ne parvient pas à restituer adéquatement la cohérence temporelle de l'énoncé du TS. Ces problèmes de pertinence des temps verbaux ne sont pas sans conséquence ; en effet, les évaluatrices ont estimé qu'ils introduisaient des glissements de sens dans le TC.

<i>Texte source 3</i>	<i>But synesthetes also report making computational errors because 6 and 8 have the same color.</i>			
<i>TA brute</i>	<i>DeepL 2018 & Google 2020</i>	Mais les synesthetes rapportent aussi avoir fait des erreurs de calcul...		
	<i>DeepL 2020</i>	Mais les synesthésiens déclarent également faire des erreurs de calcul...		
<i>Traduction fautive</i>	... avoir fait des erreurs de calcul...			
<i>Traduction correcte</i>	... déclarent qu' ils font des erreurs de calcul/... déclarent faire des erreurs de calcul...			
<i>Occurrence de l'erreur</i>	2018	PE de <i>DeepL</i> : 78 % (7 sur 9) TH : 0 % (0 sur 8)	2021	PE de <i>Google</i> : 44 % (4 sur 9) PE de <i>DeepL</i> : 0 % (0 sur 8) TH : 0 % (0 sur 7)

Tableau 68 : Incohérence temporelle (1)

Dans le premier exemple (Tableau 68), le gérondif *making* a été traduit, dans six PE de TA neuronale sur neuf, par une construction infinitive au passé, « avoir fait », marquant une relation temporelle d'antériorité, alors que la temporalité qui aurait logiquement dû être privilégiée est celle du présent, ce qui a été le cas dans toutes les productions humaines. En effet, étant donné que la synesthésie est un trouble de la perception sensorielle permanent, les manifestations de ce phénomène biologique ne disparaissent pas du jour au lendemain. Cette rupture de cohérence temporelle s'est retrouvée dans de nombreuses PE, bien qu'elle ait totalement disparu des PE de *DeepL* effectuées en 2021, parce qu'une fois de plus, le moteur ne propose plus la TA brute qu'il proposait en 2018.

Texte source 1	<i>This summer she offered the British people the chance to turn her into an elected dictator, and they decided in their wisdom they would rather keep her as a prime minister</i>			
TA brute	DeepL 2018 & Google 2020	Cet été [...] et ils ont décidé, dans leur sagesse, qu'ils préfèreraient la garder comme/en tant que ...		
	DeepL 2020	Cet été [...] et ils ont décidé, dans leur sagesse, qu'ils préfèraient la garder comme...		
Traduction fautive	...ils ont décidé qu'ils préfèreraient la garder...			
Traduction correcte	... ils ont décidé qu'ils préfèraient ...			
	... ils ont préféré ...			
	... ils ont sagement décidé qu'il était préférable ...			
Occurrence de l'erreur	2018	PE de <i>DeepL</i> : 75 % (6 sur 8) TH : 0 % (0 sur 8)	2021	PE de <i>Google</i> : 50 % (4 sur 8) PE de <i>DeepL</i> : 0 % (0 sur 7) TH : 0 % (0 sur 9)

Tableau 69 : Incohérence temporelle (2)

Dans ce deuxième exemple (Tableau 69), c'est la valeur temporelle de la forme verbale *they would rather keep her* qui n'a pas été correctement analysée. Celle-ci a été rendue par un conditionnel ayant valeur de futur dans le passé « préféreraient », alors que pour des raisons de cohérence temporelle (simultanéité d'actions), elle aurait dû être traduite par une forme verbale au passé (passé composé ou imparfait) ou encore par une tournure impersonnelle, ce qui fut le cas dans toutes les TH 2018 et 2021, ainsi que dans les PE de *DeepL* en 2021, puisque le moteur a fourni une meilleure version brute de ce segment que la version 2018.

6.2.2.2.6 Irrégularité terminologique

Une autre limite bien connue en TA est le manque de cohérence dans l'utilisation des termes (Lee et Liao, 2011 ; Martikainen, 2019c ; O'Brien, 2012b ; Schumacher, 2020b). Martikainen (2019c) note plus spécifiquement un manque de régularité dans l'usage des acronymes, sigles et abréviations au sein d'un même texte qui serait une caractéristique propre en TA statistique (p. 237).

Pour notre part, l'analyse qualitative des erreurs dans nos deux corpus a montré une plus grande irrégularité terminologique en PE qu'en TH.

<i>Texte source 2</i>	<i>China could be the largest economy in the world, accounting for around 20 % of world GDP in 2050, with India in second place and Indonesia in fourth place (based on GDP at PPPs).</i> <i>A number of other emerging markets will also take centre stage – Mexico could be larger than the UK and Germany by 2050 in PPP terms</i>
<i>DeepL 2018</i>	... avec l'Inde en deuxième position et l'Indonésie en quatrième position (sur la base du PIB aux PPA). ... le Mexique pourrait être plus grand que le Royaume-Uni et l'Allemagne d'ici 2050 en termes de PPP ...
<i>Traduction fautive</i>	(... PPA) ... en termes de PPP ...
<i>Traduction correcte</i>	(... PPA) ... en matière de PPA ...
<i>Occurrence</i>	PE de <i>DeepL</i> : 50 % (4 sur 8) TH : 22 % (2 sur 9)

Tableau 70 : Irrégularité terminologique

Dans le Tableau 70, alors que l'acronyme PPP (*Purchasing Power Parity*) apparaît deux fois dans le TS et ce, de manière rapprochée, les résultats de la TA neuronale 2018 sont incohérents. La machine propose tout d'abord de le traduire par « PPA », ensuite par « PPP ». La traduction attendue de cet acronyme en français est PPA (Parité du Pouvoir d'Achat)¹⁴¹. Or, en post-éditant, la majorité des étudiantes ont conservé la proposition

¹⁴¹ « La parité de pouvoir d'achat (PPA) est un taux de conversion monétaire qui permet d'exprimer dans une unité commune les pouvoirs d'achat des différentes monnaies. Ce taux exprime le rapport entre la quantité d'unités monétaires nécessaire dans des pays différents pour se procurer le même panier de biens et de services » (TERMIUM Plus®2023, s.p.).

erronée « en termes de PPP » alors que l’acronyme avait été correctement traduit par la TA à la ligne supérieure (« sur la base du PIB aux PPA »). Cette irrégularité a été recensée dans quatre PE de *DeepL* sur huit contre deux TH sur neuf ; elle n’a toutefois plus été recensée dans les productions 2021. Encore une fois, la tendance semble avoir été de faire confiance à ce que la machine propose sans se référer au TS. Nous pensons que ce type d’irrégularités entraîne une rupture de la cohésion textuelle et a pour effets d’entraver la compréhension du texte source, et de dérouter sinon de perdre le lecteur.

6.2.2.2.7 Perte de richesse linguistique

Comme nous l’avons détaillé dans la première partie du présent chapitre, l’une des conséquences néfastes du recours de plus en plus systématique aux outils de TA serait une perte de richesse lexicale non seulement en TA brute, mais aussi en PE (Farrell, 2018 ; Martikainen, 2022b ; Toral, 2019). Certaines affirment même que cet effet pourrait conduire à un appauvrissement de la langue cible (Deneufbourg, 2021 ; Toral, 2019 ; Vanmassenhove et al., 2021). Cette perte de richesse lexicale est reflétée par des phénomènes de simplification et de standardisation en TA. Autrement dit, il a été démontré que les textes traduits automatiquement – même après post-édition – sont généralement plus simples d’un point de vue lexical et plus standardisés qu’en TH ; il y a moins d’inventivité et de créativité, ils présentent une plus grande uniformité dans les tournures usitées et dans le choix des mots comparativement à la TH (Farrell, 2018).

Dans nos corpus, étant donné que nous disposons de plusieurs productions d’un même TS, il est aisé de comparer les différentes propositions de traduction d’un même segment. Les deux exemples repris dans le Tableau 71 et le Tableau 72 permettent de se rendre compte de la « variabilité constitutive des traductions [humaines] » (Braunstein, 2016, p. 63) et d’illustrer ce que nous considérons comme une perte de richesse linguistique en TA et en PE.

<i>Texte source 1</i>	<i>Her duty is to carry on with this unenviable task. And since she is a dutiful woman, that is probably what she will do.</i>	
<i>TA brute</i>	<i>DeepL 2018</i>	Son devoir est d'accomplir cette tâche peu enviable. Et puisque c'est une femme dévouée, c'est probablement ce qu'elle fera.
	<i>Google 2020</i>	Son devoir est de poursuivre cette tâche peu enviable. Et comme elle est une femme respectueuse, c'est probablement ce qu'elle fera.
	<i>DeepL 2020</i>	Son devoir est de poursuivre cette tâche peu enviable. Et comme elle est une femme dévouée, c'est probablement ce qu'elle fera.

Exemple PE	Son devoir est de poursuivre cette tâche peu enviable. Et comme elle est une femme dévouée, c'est probablement ce qu'elle fera.
Variabilité en TH	<ul style="list-style-type: none"> ○ Son devoir est de poursuivre cette besogne peu enviable. Puisque c'est une femme dévouée, elle l'accomplira sans doute. ○ Il en va de son devoir de continuer cette tâche peu enviable. Et vu qu'elle est une femme dévouée, c'est probablement ce qu'elle fera. ○ Son devoir est de continuer d'assumer cette tâche ingrate. Et en tant que femme dévouée, c'est probablement ce qu'elle fera. ○ Il en va de son devoir, de poursuivre cette tâche, bien peu enviable. Femme dévouée, elle ne se défilera certainement pas. ○ Il est de son devoir de poursuivre cette tâche que personne ne lui envie. Et puisqu'elle a le sens du devoir, c'est probablement la voie qu'elle choisira.

Tableau 71 : Perte de richesse linguistique en PE (1)

Texte source 1	<i>It recognises that the referendum decision must be implemented, but also that to implement it with gung-ho gusto, of the kind that Boris Johnson could provide, would be perilously divisive.</i>	
TA brute	<i>DeepL 2018</i>	Il reconnaît que la décision référendaire doit être mise en œuvre, mais aussi que sa mise en œuvre avec une hargne débridée, comme celle que Boris Johnson pourrait fournir, serait dangereusement discordante.
	<i>Google 2020</i>	Il reconnaît que la décision référendaire doit être mise en œuvre, mais aussi que la mettre en œuvre avec enthousiasme, comme celui que Boris Johnson pourrait fournir, serait dangereusement source de division.
	<i>DeepL 2020</i>	Il reconnaît que la décision du référendum doit être mise en œuvre, mais aussi que la mettre en œuvre avec un enthousiasme débordant, du type de celui que Boris Johnson pourrait apporter, serait dangereusement source de division.

<i>Exemple PE</i>	Il reconnaît que la décision du référendum doit être mise en œuvre, mais affirme aussi que mettre en œuvre cette décision avec un enthousiasme débordant, à la manière de Boris Johnson, pourrait être une dangereuse source de division.
<i>Variabilité en TH</i>	<ul style="list-style-type: none"> ○ Ils admettent qu'il faut mettre en place le référendum, mais ils reconnaissent également que, s'ils le font à la manière cavalière de Boris Johnson, cela pourrait dangereusement diviser le pays. ○ Alors que certains reconnaissent que la décision du référendum doit être mise en œuvre, d'autres pensent que l'appliquer dans un élan enthousiaste et naïf, tel que pourrait le faire Boris Johnson, engendrerait de dangereux clivages. ○ S'il reconnaît la nécessité d'observer les résultats du référendum, il craint que l'attitude va-t-en-guerre, affichée à cet égard par Boris Johnson notamment, ne sème de dangereuses divisions. ○ Il reconnaît que la décision du référendum doit être implémentée, mais également que son [sic] mise en œuvre zélée, tel que l'envisagerait Boris Johnson, pourrait également se révéler clivante au point de devenir dangereuse. ○ Le parti reconnaît que la décision à l'issue du référendum doit être implémentée mais aussi que la mettre en application avec trop d'enthousiasme, tel que Boris Johnson pourrait le faire, sèmerait la discorde. ○ Les conservateurs admettent que la décision émanant du référendum doit être respectée. Mais accueillir cette décision avec enthousiasme, à l'instar d'un Boris Johnson, pourrait porter atteinte à l'unité du pays.

Tableau 72 : Perte de richesse linguistique en PE (2)

Remarquons tout d'abord que pour chaque exemple, les trois TA brutes se ressemblent fortement. Alors que dans la majorité des PE, les étudiantes sont restées très proches de la TA brute pour traduire ces deux segments, les traductions proposées en TH sont linguistiquement bien plus riches et variées. Précisons que parmi les propositions qui figurent dans le Tableau 72, plusieurs contiennent notamment des glissements de sens et des calques fautifs ; nous avons toutefois tenu à reprendre ces exemples pour montrer la variété des formulations en TH. Nous jugeons que ces deux exemples attestent l'existence du « *priming effect* » ou « effet fantôme » de la TA que nous explicitons au chapitre 8.

6.2.3 Synthèse et discussion

L'analyse qualitative des erreurs présentes dans les productions des étudiantes en traduction nous a permis de relever certains des effets bénéfiques de la (PE de) TA neuronale sur la qualité, mais également les limites, toujours existantes, de cette technologie. Nos résultats ont montré que la grande majorité des erreurs repérées dans les textes post-édités étaient déjà présentes en TA neuronale. En effet, la plupart de ces erreurs n'ont manifestement pas fait l'objet d'une modification de la part des étudiantes. C'est également l'une des conclusions de l'étude faite par Killman (2018) ; il a pu remarquer que les erreurs les plus fréquentes observées dans les PE sont en réalité des sorties de TA brute qui n'ont pas été modifiées par les étudiantes :

In some cases, [students] failed to edit incorrect MT output in more instances than engage in making necessary changes [...], showing a negative impact on performance levels. In all cases, unchanged incorrect MT output is the most abundant post-editing error category. (p. 137)

En outre, nous pensons que de manière générale, nos observations viennent renforcer l'hypothèse selon laquelle, puisque les textes post-édités passent par le filtre de la TA, ceux-ci présentent un plus haut degré d'interférence avec la langue source que les textes traduits humainement (Toral, 2019). Nous en discutons plus longuement au chapitre 8.

6.3 Conclusion

6.3.1 Mesures linguistiques automatiques

6.3.1.1 Richesse lexicale

Les résultats de nos deux expériences ont montré que les textes traduits humainement ont une densité lexicale très légèrement supérieure à celle des textes post-édités et de la TA brute, bien que les valeurs soient très proches. De plus, nous avons observé que les productions humaines présentent le degré de diversité lexicale le plus élevé. Nous pouvons en conclure que la richesse lexicale en TH est supérieure à celle en TA brute, ainsi qu'en PE de TA statistique comme de TA neuronale. Enfin, les résultats indiquent que les textes post-édités à partir de *DeepL* ont une richesse lexicale inférieure à celle des TH, ainsi qu'à celle des PE de *Google* et des PE de TA statistique.

6.3.1.2 Longueur moyenne des phrases

Les résultats indiquent une longueur moyenne de phrases supérieure en traduction par rapport aux textes sources. Les mesures obtenues en post-édition sont similaires à celles

en traduction humaine, suggérant que les versions post-éditées se rapprochent davantage des traductions humaines que des textes sources. Les PE de TA neuronale ont des taux de foisonnement comparables à ceux des traductions humaines, tandis que les PE de TA statistique ont des taux de foisonnement inférieurs à ceux des traductions humaines.

6.3.1.3 Équivalence syntaxique

D'après les résultats des trois indices d'équivalence syntaxique, les PE sont plus proches syntaxiquement des textes sources que les TH. De plus, les PE de *DeepL* présentent une moindre proximité syntaxique avec les TS que les PE de *Google* et que les PE de TA statistique. La proximité syntaxique peut être schématisée comme suit :

Expérience 2018 : TS < PE de TAS < PE de TAN (*DeepL*) < TH

Expérience 2021 : TS < PE (*Google*) < PE (*DeepL*) < TH

En résumé, les résultats du calcul des trois indices d'équivalence syntaxique repris dans les Tableau 53 et Tableau 54 révèlent la même tendance, à savoir que les PE présentent une plus grande proximité syntaxique avec les textes sources que les TH. Nous estimons que ces observations dévoilent une fois de plus l'effet d'interférence de la langue source en PE et viennent étayer l'existence d'un *priming effect* ou « effet fantôme » de la TA en PE que nous développons au chapitre 8.

6.3.2 Analyse qualitative des erreurs

Nous pouvons dire que notre analyse linguistique a mis au jour des traits caractéristiques du *post-editeuse* par rapport au *translationese*, telles que la maîtrise des règles d'usage du français (grammaire et orthographe), la tendance à la traduction littérale, à l'effacement d'informations, à la rupture de la cohésion textuelle et de la cohérence, à l'irrégularité terminologique et à la perte de richesse linguistique.

Aussi, nous considérons que nos observations confirment la tendance selon laquelle la post-éditrice serait encline à faire démesurément confiance aux sorties du moteur de TA et à s'en satisfaire trop rapidement, sans doute en raison de l'apparente fluidité des traductions produites (Guerberof Arenas, 2009 ; Deneufbourg, 2019). Selon Deneufbourg, la TA induirait « une sorte d'excès de confiance dans le chef des traducteurs, toutes expériences confondues, et favorise[rait] donc la présence résiduelle d'erreurs de sens moins visibles » (*ibid.*, s.p.), mais cette tendance serait particulièrement marquée chez les étudiantes (Daems, 2016 ; Depraetere, 2010 ; Dirand et Rossi, 2019 ; Loock et Léchauguette, 2021 ; Martikainen, 2020). Toutefois, nous tenons à préciser que, dans notre cas, il se peut que cet excès de confiance découle en partie des consignes de PE que nous avons données aux étudiantes avant l'expérience, et plus spécifiquement de la consigne suivante : « Exploiter au maximum le résultat brut de la traduction automatique ». Il nous est cependant impossible d'estimer l'influence de cette consigne sur nos résultats. Dans

tous les cas, il apparaît que les étudiantes ne prennent pas la peine de retourner au TS, ou pas suffisamment (Elming et al., 2014¹⁴² ; Larssonneur, 2022 ; Loock et Léchaugnette, 2021 ; Volkart et al., 2022), que ce soit par souci d'efficacité, par manque de vigilance ou par paresse intellectuelle, laissant des erreurs parfois graves dans leur PE.

Pour conclure, en dépit des nombreux progrès réalisés par la nouvelle génération de TA, nous pensons que les résultats des mesures automatiques en matière de diversité lexicale et d'équivalence syntaxique, ainsi que l'examen des erreurs relevées dans les textes post-édités, mettent en lumière la nécessité de former les futures traductrices à la pratique de la PE. En effet, nous sommes persuadée qu'une fois formées, les étudiantes seraient davantage en mesure de repérer et de lisser ces marques laissées par la TA en PE, et de déceler et corriger les erreurs imputables à la TA. Nous développons cette nécessité de formation plus en détail dans le dernier chapitre de cette thèse.

¹⁴² « The observation that translator had more gaze activity on the current source segment when translating than when post-editing may be explained by the fact that a translation suggestion is already presented for post-editing, so less inspiration from looking at the source is needed » (Elming et al., 2014, p. 161).

7 Connaissances et perceptions des étudiantes

7.1 Introduction

Outre les études du processus et du produit de la PE, nous sommes convaincue que les connaissances et les perceptions sur la TA et la PE qu'ont les traductrices et étudiantes en traduction, constituent un objet d'étude utile et précieux. Ainsi, l'objectif du présent chapitre est d'avoir un état des lieux des connaissances et des perceptions qu'ont les étudiantes en traduction de l'Université de Liège sur la TA et sur la PE. Précisons d'emblée que nous entendons analyser ces connaissances et perceptions au sens de Rossi (2019a), dont les recherches s'inscrivent dans le champ de la traductologie pragmatique et plus précisément en ergonomie cognitive, dont l'objet porte sur les « processus mentaux, tels que la perception, la mémoire, le raisonnement et les réponses motrices, dans leurs effets sur les interactions entre les personnes et d'autres composantes d'un système »¹⁴³. Pour ce faire, nous avons interrogé les étudiantes au moyen de sondages en ligne qui ont été soumis et récoltés à deux reprises auprès d'échantillons différents : une fournée en 2018 et une fournée en 2021. Dans notre analyse, nous nous attachons à comparer les données récoltées en 2018 à celles récoltées en 2021.

Dans le présent chapitre, après avoir formulé nos hypothèses, nous passons en revue la littérature scientifique sur le sujet qui nous occupe. Nous présentons ensuite notre méthodologie de collecte des données et décrivons la composition de nos sondages et des deux fournées de participantes. Nous donnons quelques précisions concernant le traitement des données avant de présenter et d'interpréter les résultats de ces sondages par question. Nous proposons également une synthèse des résultats par sondage afin de répondre aux différentes hypothèses, et terminons par la partie conclusion et discussion.

¹⁴³ Définition adoptée par l'Association Internationale d'Ergonomie (IEA) voir site du CNAM : <https://ergonomie.cnam.fr/ergonomie/index.html>

7.2 Hypothèses

Dans ce chapitre, nous tentons de répondre à notre deuxième question de recherche :

Question B – Quelles sont les connaissances et les perceptions des étudiantes en traduction sur la TA et sur la PE ?

Afin de répondre à cette question, nous formulons les neuf sous-questions suivantes :

B1. Quel usage font les étudiantes de la TA ?

Hypothèse HB1

Concernant l'usage de la TA, nous pensons que nous allons récolter des résultats semblables à ceux obtenus par Dirand et Rossi (2019), à savoir que la majorité des étudiantes interrogées se servent régulièrement de la TA, et la plupart du temps afin de mieux comprendre le texte source, mais aussi afin de gagner du temps.

B2. Quel est le rapport des étudiantes à la PE ?

Hypothèse HB2

Concernant la PE, nous croyons que parmi la fournée 2018, peu d'étudiantes sauront ce que recouvre ce terme et très peu auront une expérience en la matière, notamment car les étudiantes interrogées en 2018 n'ont pas eu l'occasion d'expérimenter ces outils dans le cadre de leur formation. En revanche, en ce qui concerne la fournée 2021, nous nous attendons à ce que l'ensemble des étudiantes soient familières de cette pratique étant donné l'intégration systématique de la PE dans leur cursus dès la 3^e année de bachelier à partir de l'année académique 2019-2020.

B3. Comment les étudiantes considèrent-elles les performances actuelles en TA et comment imaginent-elles ces performances à l'avenir ?

Hypothèse HB3

Il ressort de plusieurs études que, d'une manière générale, les traductrices ne sont pas réfractaires à la technologie (Koskinen et Ruokonen, 2017 ; LeBlanc, 2017). Nous pensons que les étudiantes croient encore davantage en la technologie et en son potentiel. En outre, nous partageons pleinement l'avis de Dirand et Rossi (2019), selon lequel les étudiantes seraient « sensibles aux promesses que font les fabricants [au sujet des nouvelles technologies de la traduction dont la TA], et qui sont souvent relayées dans la presse » (p. 70) ; des promesses qui découlent en général d'une surévaluation des performances des moteurs de TA (Loock, 2019). Aussi, nous pensons que la plupart des étudiantes se montreront satisfaites par la qualité d'une TA brute. De plus, nous croyons que les étudiantes vont se montrer relativement optimistes quant à l'avenir des outils de TA, et qu'elles seront convaincues d'une amélioration des performances de cette technologie dans un avenir plus ou moins proche. Néanmoins, en nous appuyant sur ce que rapportait la TAUS en 2017 à l'égard des attentes sur le marché de la traduction, nous ne pensons pas

que les étudiantes considèrent que la qualité de la TA brute pourra un jour égaler la qualité en TH : « La plupart des acteurs du marché de la traduction ne s’attendent toujours pas à ce que la qualité de production de la TA atteigne de sitôt les mêmes niveaux que la traduction humaine » (Massardo et van der Meer, 2017).

B4. Quelle est la perception des étudiantes par rapport au processus de PE et à la qualité qui en résulte ?

Hypothèse HB4

Concernant la perception du processus de PE et de la qualité, on trouve parmi les avis recueillis par Moorkens (2020) celui d’une traductrice indépendante pour l’irlandais : « [Post-editing] slows my workflow down by about 10-15 % and also results in a poorer, more stunted translation ». Ces propos rejoignent en partie les constatations de Daems (2016) : la majeure partie des traductrices professionnelles et des étudiantes qu’elle a interrogées ont estimé que la qualité obtenue en TH était meilleure ou comparable à la qualité obtenue en PE. Pour notre part, nous nous attendons à ce que la plupart des étudiantes soient d’avis que la qualité en PE est comparable à la qualité en TH, et ce, d’autant plus en 2021 qu’en 2018 en raison de l’insertion d’exercices de PE dans les cours en traduction à l’Université de Liège en 2019-2020.

B5. Les étudiantes jugent-elles l’intégration d’une formation à la (PE de) TA dans leur cursus nécessaire ?

Hypothèse HB5

En nous appuyant sur les travaux de Şahin (2013)¹⁴⁴, de Breyel et Grass (2020)¹⁴⁵ et de Jia et al. (2019)¹⁴⁶, nous nous attendons à ce que les étudiantes se montrent très favorables à l’intégration d’une formation à la (PE de) TA dans leur cursus universitaire. En effet, nous partons du principe que la plupart de ces futures traductrices sont conscientes des changements profonds que connaît notre société, liés notamment à la transition numérique, et par conséquent de la nécessité d’acquérir « un certain nombre de compétences qui [les] mettent [...] en adéquation avec les besoins d’un marché en pleine évolution » (Froeliger, 2014, pp. 59-60).

B6. Les étudiantes considèrent-elles l’un des modes de traduction comme moins exigeant ?

Hypothèse HB6

Nous supposons qu’elles vont trouver la PE moins exigeante que la TH (Moorkens, 2018 ; Yamada, 2015) notamment, car elles vont considérer que ce mode de traduction est à la

¹⁴⁴ « [N]ovice translators are positive about the place of this new component [MT post-editing] in the curriculum » (Şahin, 2013, p. 1).

¹⁴⁵ « [T]ranslators are aware of the need to adapt to the evolution of their profession, which means focusing on new tasks with higher added value » (Breyel et Grass, 2020, p. 21).

¹⁴⁶ « [P]rofessional translators are found to be reluctant to take post-editing jobs due to their negative perceptions of MT quality and post-editing work, while student translators seem to show greater potential to become future post-editors to fill this gap » (Jia et al., 2019, pp. 60-61).

fois plus rapide (Killman, 2018 ; Martikainen et Mestivier, 2020 ; Schumacher, 2020b) et moins intense sur le plan cognitif que la TH (Daems, 2016 ; Jia et al., 2019 ; O'Brien, 2006 ; Carl et al., 2015 ; Screen, 2016).

B7. Les étudiantes montrent-elles une préférence pour l'un des modes de traduction ?

Hypothèse HB7

D'une manière générale, nous pensons que les jeunes générations ne sont a priori pas réfractaires aux changements et aux nouvelles technologies. Selon nous, les étudiantes vont ainsi se montrer, pour la plupart, positives à l'égard de la PE, comme cela a déjà été mis en avant dans d'autres travaux (García, 2010 ; Jia et al., 2019 ; Moorkens et O'Brien, 2015 ; Moorkens, 2018 ; Scansani et al., 2019). Aussi, contrairement aux résultats de Daems (2016), nous estimons qu'elles vont préférer la PE à la TH, comme Yamada (2015) a pu l'observer.

B8. Quel est le degré de satisfaction des étudiantes en PE et en TH ?

Hypothèse HB8

La PE étant souvent considérée comme une pratique moins gratifiante (Moorkens, 2020¹⁴⁷ ; Scheepers et Schulz, 2017) et moins satisfaisante que la TH (Daems, 2016 ; Fulford, 2002 ; Koehn, 2009), nous nous attendons à ce que les étudiantes se montrent davantage satisfaites par le processus de TH que par celui de PE.

B9. L'opinion des étudiantes a-t-elle changé avec l'expérience ?

Hypothèse HB9

Il nous est difficile d'estimer la proportion d'étudiantes dont la participation à cette expérience est susceptible de changer leurs perceptions sur la TA et sur la PE. Nous estimons que plusieurs étudiantes pourraient effectivement se montrer plus positives à l'égard de la TA et de la PE après leur participation, puisque des résultats comparables ont été rapportés par Daems, Vandepitte et al. (2017), par García (2010), par Gaspari (2001) et par Scansani et al. (2019). Toutefois, nous croyons que, pour la majeure partie des participantes, cette courte expérience ne va pas impulser de changement d'opinion. Prenons l'étude comparative menée par Daems (2016), s'il y a, certes, davantage d'étudiantes qui se sont montrées positives à l'égard de la PE que l'inverse, l'opinion de la majorité d'entre elles n'a pas changé en cours d'expérience : « Most participants' attitude towards PE did not change after participating in the experiment [...]. Participants who changed their mind usually felt more positive about PE after the experiment [...] than more negative » (p. 70). De même, l'analyse des deux questionnaires proposés par Rossi (2019a) à des étudiantes de master de traduction spécialisée, avant et après le suivi d'un cours de 12h sur la TA et sur la PE, ne fait pas non plus état d'une évolution majeure des perceptions

¹⁴⁷ « Some participants complained that post-editing is "tedious" and "a thankless task, in every way » (Moorkens, 2020, p. 66).

des étudiantes. Si certaines étudiantes ont témoigné d'une perception plus favorable à la suite du cours, d'autres ont, en revanche, vu leurs craintes renforcées et ont manifesté des perceptions plus défavorables.

7.3 Revue de la littérature

Même si les travaux qui tentent de recueillir et d'analyser les connaissances et les perceptions des (futurs) traductrices sont peu nombreux (Rossi, 2019a) et en dépit de leurs résultats contradictoires, nous avons tenté d'en donner un aperçu général.

D'après Koskinen et Ruokonen (2017), étudier les connaissances et les perceptions des traductrices permettrait, entre autres, d'améliorer la pratique de la traduction et de la PE en contexte professionnel : « The better we understand all of the factors that affect translators' willingness and ability to adapt to changing technology, the better we understand professional translation practice » (p. 23). Il y aurait, en outre, un lien direct entre perceptions de la TA et performances en PE : « [A] positive attitude to MT has been found to be a factor in PE performance (de Almeida 2013 ; Mitchell 2015) » (cités dans Stasimioti et Sosoni, 2019, p. 126). De plus, selon de Faria Pires (2020), de telles recherches seraient également bénéfiques pour l'intégration de la PE dans les cursus des étudiantes.

À travers les différentes études que nous présentons dans cette partie, nous verrons que les perceptions des traductrices à l'égard de la TA et de la PE sont partagées (Scansani et al., 2019) entre rejet et engouement, entre crainte et optimisme, comme l'évoquaient déjà Carl et al. en 2015 :

PEMT [post-editing of machine translation], like every technological shift in the translational world, provokes a wide range of mixed feelings reaching from skepticism to enthusiasm (similar to the introduction of translation memory systems in the 1990s) re-opening the discussion about the growing influence of machines (i.e., computers and CAT tools) on the translational workflow and the translator as such. (p. 150)

Plusieurs études montrent que la traductrice ne serait pas réfractaire à la technologie de manière générale (voir notamment Koskinen et Ruokonen, 2017 ; LeBlanc, 2017). Selon les conclusions de Koskinen et Ruokonen (2017) en l'occurrence, les traductrices seraient même tout à fait prêtes à accepter un nouvel outil technologique pour autant qu'il leur permette d'augmenter leur efficacité : « [T]ranslators are probably quite willing to adopt new technology as long as it makes their work more efficient » (p. 22). Il semblerait toutefois que cela ne soit plus le cas lorsque l'usage de technologies entraîne une certaine perte d'autonomie : « when translators perceive a loss of control over their work because of technologies, they do not tend to perceive technologies in a positive light » (*ibid.*, p. 59).

Or, dans le cas des outils de TA, c'est plutôt une certaine réticence et une attitude négative de la part de la traductrice qui ressortent d'un large pan de la littérature spécialisée (Breyel et Grass, 2021 ; Fulford, 2002 ; Dirand et Rossi, 2019 ; Läubli et Orrego-

Carmona, 2017¹⁴⁸; O'Brien, 2002 et 2012b ; Peraldi, 2016 ; Moorkens et O'Brien, 2015 ; Moorkens, 2017 ; Saint-André, 2015 ; Sakamoto, 2019¹⁴⁹; Vieira, 2020). Dirand et Rossi (2019) reconnaissent effectivement que : « [...] face aux progrès de la traduction automatique, les traducteurs humains ressentent le besoin de résister, souvent en s'opposant » (p. 75). En outre, étant donné que la nouvelle génération de TA est très souvent vantée comme une technologie révolutionnaire, capable de concurrencer, voire d'égaliser, la traduction humaine (Hassan et al., 2018 ; Shoshan, 2018), bon nombre de traductrices ont tendance à considérer cette technologie et la pratique exponentielle de la PE comme une réelle menace pour leur profession :

[T]here is a fear for translators of losing their jobs or seeing the price of their work drop. There is also a reluctance of professional translators towards post-editing. The two main reasons are that post-editing is considered an alienating activity devoid of creativity. (Breyel et Grass, 2020, p. 20)

Cette crainte est également partagée par les étudiantes : « Findings [...] revealed widespread preoccupation among students about how technology and MT affects professional work » (González Pastor, 2021, p. 57). Ainsi, la post-édition est souvent vue d'un mauvais œil et est considérée de surcroît comme une pratique peu satisfaisante (Daems, 2016 ; Fulford, 2002 ; Koehn, 2009 ; Rossi, 2019a), rébarbative (Peraldi, 2016)¹⁵⁰, ingrate (Grass, 2022 ; Moorkens, 2020 ; Scheepers et Schulz, 2017¹⁵¹), déshumanisante (O'Brien, 2012b)¹⁵², voire dégradante (SFT, 2022).

¹⁴⁸ « Our study provides evidence that MT is often portrayed negatively among translators on social media outlets. The suspicions about a negative attitude towards MT that stemmed from our qualitative analysis of Facebook and LinkedIn posts (Section 3) were supported by the results of the sentiment analysis carried out on Twitter data » (Läubli et Orrego-Carmona, 2017, p. 65).

¹⁴⁹ « In *Why do many translators resist post-editing? A sociological analysis using Bourdieu's concepts*, Akiko Sakamoto draws on Bourdieu's concepts of capital, field and habitus to present a sociological analysis of translators' attitudes to post-editing. Her study is based on post-editing training manuals, focus-group interviews conducted with sixteen UK project managers and a survey of 155 company websites. She proposes that current perceptions of post-editing place post-editors in a different position from translators across axes representing the volume of capital (i.e. remuneration) and the capital type (cultural or economic). [...] Sakamoto presents a hypothetical model where post-editors can be on the same level as translators on the capital volume axis. She concludes by saying that this would represent a healthier model for introducing post-editing services to translators » (Vieira et al., 2019, p. 9).

¹⁵⁰ « L'activité de post-édition est souvent perçue comme étant particulièrement répétitive et rébarbative » (Peraldi, 2016, p. 69).

¹⁵¹ « CAT tools mostly use post-editing features, and translators in general find post-editing an extremely boring, tedious and unrewarding chore » (Scheepers et Schulz, 2017, p. 20).

¹⁵² « Some translators feel dehumanised by the technology they are required to use. Having to fix the errors created by a machine translation system (or created by a human translator and propagated by a TM system) understandably irks some translators to such a degree that they refuse to interact with the technology » (O'Brien, 2012b, p. 108).

Common arguments against MT include a dislike for correcting repetitive errors that a human translator would never make, a fear of losing language proficiency by working with poor MT output and a dislike of having one's freedom of expression limited (Wagner, 1985:213). (O'Brien, 2002, p. 100)

Enfin, notons que les traductrices chevronnées se montreraient davantage négatives à l'égard de la TA par rapport aux traductrices novices : « their attitudes towards the technology were considerably more negative than that of the novice group » (Moorkens et O'Brien, 2015, p. 80). En effet, comparativement aux traductrices chevronnées, les étudiantes se montreraient, elles, plus positives à l'égard de la TA et plus disposées à faire de la PE (Scansani et al., 2019 ; Yamada, 2015).

Pour illustrer ces perceptions négatives, nous avons repris un des commentaires reçus par Saint-André (2015) lorsqu'elle a tenté de recruter des participantes pour un sondage sur la PE dans le cadre de sa thèse de doctorat à l'Université d'Ottawa : « nous avons pu constater de la part de quelques intervenants une certaine méfiance, voire de l'animosité, envers la TA et, par conséquent, envers la PÉ. Par exemple, selon une intervenante, "une personne qui accepte de réviser des "textes" issus de la traduction automatique contribue au naufrage de la profession de traducteur." "Que l'on puisse se former à cette chose que vous appelez postédition me laisse sans voix", ajoute-t-elle » (p. 151).

Yamada (2015) a relevé qu'au Japon, il y avait certains a priori sur la PE : « it is also believed (in Japan, at least) that PE requires less skill than HT, and continuing to perform PE may cause one's translation skills to deteriorate over time » (p. 50).

Läubli and Orrego-Carmona (2017) se sont penchés sur les perceptions des traductrices, partagées sur les réseaux sociaux. Ils ont voulu compléter les résultats d'une précédente analyse qualitative de posts *Facebook* and *LinkedIn* en passant au crible la teneur de plus de 13 000 tweets en anglais publiés entre 2015 et 2017 et contenant les termes « *machine translation* » ou « *machine translated* ». Avec un rapport de 3:1, les résultats de leur analyse quantitative ont montré que les perceptions négatives de la TA l'emportaient largement sur les perceptions positives.

Afin de mieux appréhender ce phénomène, Vieira (2020) s'interroge sur le véritable objet de cette résistance ; les traductrices se montrent-elles défavorables à la technologie, ou aux conséquences de l'introduction de cette technologie sur le marché ?

This lack of clarity applies particularly to whether it is technology itself or its market effects that are predominantly deemed problematic. The difference between having a negative attitude to technology and having a negative attitude to the perceived repercussions of technology is a small but important one. (p. 8)

Si l'objet de cette résistance n'est pas clair, nous pouvons néanmoins tenter d'en expliquer l'origine. Alors d'où vient cette réticence, voire cette résistance ? Les raisons sont nécessairement multiples.

Il se peut que cette attitude découle du fait que les traductrices professionnelles sont de plus en plus fréquemment amenées à produire des traductions ou post-éditions qui soient juste de qualité satisfaisante, tel qu'il ressort des entretiens menés par LeBlanc (2016) : « Mais avant tout, ce sont les remarques sur la *qualité* des traductions qui ressortent le plus souvent des entretiens [...] les traducteurs se rendent compte que les nouvelles conditions dans lesquelles ils sont appelés à travailler rendent difficile "l'adhésion aux normes et aux conventions de rédaction dans la langue-culture d'arrivée" (Scarpa, 2010 : 97) » (p. 90).

En 2003, Allen faisait déjà remarquer que l'intégration de la TA posait une problématique nouvelle en traduction : « the acceptance and use of half- or semi-finished texts » (p. 297). En effet, revoir ses exigences de qualité et se satisfaire d'une qualité moindre n'est pas une mince affaire : « Translators constantly concerned about having to work to lower quality standards, to produce just 'good enough' (postedited) translation » (Kenny, 2015, diapositive 36). O'Brien (2010) souligne qu'une certaine pratique de la post-édition dans le domaine privé exige de la traductrice de désapprendre une bonne partie de ce qu'elle a appris et de réduire ses exigences en matière d'excellence : « Essentially, translators are asked to unlearn much of what they are taught regarding quality and professionalism: Ignore style, fluency, cohesion, coherence, text function, context, end user – Do more, of lower quality, for much less pay » (diapositive 23).

Peut-être est-ce dû à l'influence du buzz entourant la TA neuronale : « The recent improvements in MT quality, along with the hype generated by reports of NMT quality (Castilho et al., 2017b), have been unnerving for some translators » (Moorkens et al., 2018b, p. 256). Véhiculant un discours dithyrambique, la presse, les médias, ainsi que les développeuses de systèmes exposent leur « conception agonistique, qui oppose l'homme à la machine et envisage leurs relations sur le modèle du combat expansionniste, parfois même de la lutte à mort, lorsqu'il est envisagé que le traducteur devienne 'une espèce en voie de disparition' (Marr, 2018) » (Dirand et Rossi, 2019, p. 64). Plusieurs autrices dont Deneufbourg (2021) et Moorkens (2018) constatent et déplorent une certaine exagération médiatique du phénomène qui serait liée, entre autres, à une surévaluation très (trop) fréquente des performances des moteurs de TA neuronale dans la presse généraliste, mais aussi spécialisée (Loock, 2019, p. 59). En effet, journalistes et conceptrices ont une forte tendance à surévaluer les performances de cette technologie promettant une véritable révolution pour le monde de la traduction et allant parfois jusqu'à prédire l'extinction pure et simple de la traductrice humaine : « The rise of NMT has been accompanied by a good deal of media hyperbole about neural networks and machine learning, some of which has suggested that several professions, including translation, may be under threat » (Moorkens, 2018, p. 375). Cette construction discursive nous conditionne à accueillir ces progrès technologiques comme un véritable miracle « neuronal » (Rossi, 2019b), comme en témoignent les titres de presse repris sur les Figure 69 et Figure 70.

LA TA DANS LA PRESSE ET LES MÉDIAS

Should Human Translators Fear the Rise of Machine Translation?
www.adaptcentre.ie

Le traducteur, une espèce en voie de disparition?
www.7sur7.be/tech/

AI Translates News Just as Well as a Human Would
The AI was able to translate Chinese to English with "human parity"
LOU DEL BELLO | MARCH 16TH 2019

Un premier livre entièrement traduit par une intelligence artificielle
Victor De Sepausy - 0210.2018 | Edition - Les maisons - intelligence artificielle traduction - livre traduire machine - robot traduire livre

24 FEB 2016
The Future Does Not Need Translators
https://blog.taus.net

Figure 69 : La TA dans la presse et les médias (Schumacher, 2020a)

LE MIRACLE DE LA TA NEURONALE

Miracle
Fait extraordinaire qui porte à l'étonnement et à l'admiration, réussite exceptionnelle (Etym.) Empr. au lat. *miraculum* «prodige, merveille, chose extraordinaire»
Source : <https://www.cnrtl.fr/definition/miracle>

Newest GNMT: time to witness the miracle of Google Translate
Published on October 4, 2016

Google's AI translation system is approaching human-level accuracy
But there's still significant work to be done

DeepL Translator – The New Wondrous Translating Machine
Jan 26, 2019

What Exactly is DeepL?
The DeepL Translator has taken off like a rocket and is going strong. It's the new miracle in artificial intelligence.

the guardian
Tech is removing language barriers - but will jobs be lost in translation?
Could Microsoft's Star Trek-inspired translation service ever replace professional human translators?

Google Translate update makes it pretty much as good as a human translator
By Emma Wiggin, September 20, 2016 | Computing
Getting chatty around the globe just got easier

TAUS Blog
The Future Does Not Need Translators

Electric Translator Is Latest 'Miracle'
The age of miracles is only now dawning, it appeared to-jest, then typed out:
day, with the introduction of "Tysocretina sua opyze, Oklahoman", 25 Jan 1954; "Electrical brain can now translate foreign languages", by Alice Hughes [synthesized]
Owosso Argus-Press (Michigan), 14 Jan 1954; "Modern machines do wonders, but there's possible errors" Pittsburgh Press, 8 Jan 1954; "Translators beware! You can be replaced" Presque Isle Star Herald (Maine), 28 Jan 1954; "New IBM electronic machine changes Russian into English"
Scholarite, 20 Jan 1954; "Translating machine" Science, 122, 21 October 1955, pp.745-746; "Mechanical translation: new challenge to communication", by Jacob Ornstein [4]
Science News Letter, 23 Jan 1954; "Language translation by electronic "brain" [4] Scranton Times (Pennsylvania), 8 Jan 1954; "Device hurdles word barriers"
Fed into the machine and their magnetic drum. English translation emerge all! The "brain" also was most simultaneously, "taught" rules of grammar so it could change the order of words when necessary to make a smooth intelligent English sentence. "We transmit thoughts by (smooth job of creating the means of speech," the brain translator, said it has a vocabulary translated on a high-speed lary of only 250 words but it potentially could store up Mrs. Pella, who does not hundreds of thousands.

I. Le miracle « neuronal »
II. Les usages actuels
III. Risques et besoins

Caroline Rossi
Univ. Grenoble Alpes
Autres Digis_Hum 2019
Les documents numériques en langues
JEUDI 17 OCTOBRE 2019

Figure 70 : Le miracle de la TA neuronale (Rossi, 2019b, diapositive 3)

Ensuite, ce besoin de résister n'est pas un phénomène nouveau comme nous le rappellent Sánchez-Gijón et al. (2019) ; les outils de TAO avaient également soulevé à l'époque des réticences de la part des traductrices : « We should not forget that translators were also reluctant to use CAT tools when they first appeared, but most now feel comfortable with them (Moorkens 2017) » (*ibid.*, p. 33). Ce scepticisme naturel (Gaspari, 2001) et cette attitude d'emblée réfractaire découlent peut-être tout simplement d'une certaine peur du

changement (Koskinen et Ruokonen, 2017) et d'un sentiment d'anxiété face à l'automatisation (Rossi et Chevrot, 2019 ; Vieira, 2020) que peuvent ressentir certaines personnes. Il apparaît que ce phénomène est particulièrement observable chez les traductrices n'ayant que peu ou pas d'expérience en nouvelles technologies : « potential fears of job displacement and automation anxiety [...] may be more common among those with no knowledge of MT who might “fear the unknown” » (Vieira, 2020, p. 9). À la suite d'observations et d'entretiens menés auprès de traductrices travaillant pour la direction générale de la traduction (DGT) de la Commission européenne, Rossi et Chevrot (2019) ont également pu mettre en évidence une corrélation entre les appréhensions des traductrices et leurs connaissances en matière de TA : « the translators who perceived MT as a threat were regularly those with the lowest scores for MT knowledge » (s.p.).

Peut-être que cette réticence provient également de la pression accrue qui pèse sur la profession. Dans la conjoncture actuelle, d'aucunes affirment que le secteur de la traduction est de plus en plus mis sous pression : « Translation is currently described as a profession under pressure from automation, falling prices and globalized competition » (Vieira, 2020, p. 1). Et Saint-André (2015) nous rappelle que la post-édition est une « activité intimement liée à la productivité et à la rentabilité, [et] donc à l'aspect commercial de la traduction » (p. 53). LeBlanc (2016) fait remarquer que la traductrice professionnelle est appelée à « maintenir un rythme de productivité toujours croissant » (p. 84). Et les avis qu'il a pu recueillir témoignent de la pression qui pèse sur les épaules des traductrices professionnelles lorsque de nouveaux outils d'aide à la traduction sont intégrés à leur pratique :

Si, d'un côté, les traducteurs constatent que l'introduction de certains outils d'aide à la traduction a, dans une certaine mesure, entraîné des gains de productivité, voire une réduction du travail répétitif, d'un autre côté, ils constatent pleinement l'effet de ces outils sur les pratiques administratives et commerciales qui ont depuis lors été adoptées dans les entreprises et services au sein desquels ils travaillent. (*ibid.*, p. 90)

Des pratiques qui les obligent à accélérer leur rythme de travail avec notamment l'instauration de nouvelles manières de calculer la productivité (*ibid.*). Pour do Carmo et Moorkens (2020), il est bien compréhensible de constater une certaine résistance sur le marché de la traduction professionnelle. Ils sont, en effet, convaincus que dans un tel contexte la traductrice professionnelle est plus susceptible de ressentir davantage de pression que de satisfaction ; le message qui leur est envoyé est clair et limpide : « post-edit or perish » (litt. : post-éditer ou périr) (*ibid.*, p. 38) : « Unilateral imposition of PE from an unpredictable MT source, paid at a lower word rate, makes PE feasible only if the processes involved are much more efficient than normal translation » (*ibid.*).

Du reste, il se peut que cette résistance à l'égard de la TA soit corrélée à la (dé)valorisation de la profession. Cronin (2013) soulignait à cet égard le risque d'une déprofessionnalisation engendrée par la démonétisation des services de traduction :

As a number of high-profile online translation services are free, there can be even greater reluctance to pay for translation services supplied by professional translators. The demonetization of translation in this context makes its status even more uncertain in economies dominated by the logic of exchange value, where what is valuable is what you pay for. [...] “[It] brings with it an inevitable status anxiety around the profession of the translator. [...] demonetization in a market economy implies de-professionalization. (pp. 135-136)

Et effectivement, beaucoup de professionnels jugent la rétribution de leur travail insatisfaisante (Guerberof Arenas, 2013 ; SFT, 2022) et en baisse comme l’ont constaté Sakamoto (2019) et Alonso et Vieira (2020) : « A by-product of the adoption of MT as a professional tool is that clients expect lower translation rates » (p. 394). Selon les estimations de Lommel (2018a) : le tarif au mot en PE correspondrait à 65 % du tarif au mot en TH : « Prices tend to be 65 % of the total for HT, but do not account for effort, so many translators feel they are getting paid less for the same amount of work » (p. 6). Déjà en 2012, O’Brien¹⁵³ évoquait la problématique de la baisse de rémunération comme l’un des motifs de cette réticence à recourir à la TA ; idem chez Moorkens et al. (2016) : « Aside from its perceived lower quality and the greater effort needed when post-editing output, translators are reluctant to use MT because of falling pay rates » (p. 3). Plus récemment, De Faria Pires (2020) partage toujours ce même constat : « la faible rémunération dont cette pratique [la post-édition] fait l’objet contribuerait à l’image négative qu’ont les professionnels à son égard » (p. 81). À cet égard, les deux principales raisons du refus d’effectuer des travaux de PE avancées par 686 traductrices dans l’enquête 2022 sur les pratiques professionnelles en traduction menée par la SFT, sont les tarifs trop bas et le caractère inintéressant de la tâche (Figure 71).

¹⁵³ « The professional translator is [...] demoted to the status of a fixer (Krings 2001) of seemingly unintelligent errors. That they are paid lower rates to fix such errors than to create their own translation adds to the feelings of negativity » (O’Brien, 2012b, pp. 108-109).



Base : 686

Raisons du refus de la post-édition



Figure 71 : Raisons du refus de la post-édition (SFT, 2022, diapositive 27)

S’inspirant de Cooper (2004), O’Brien (2012) prétend que ce n’est pas uniquement la technologie qui entraîne une dévalorisation et une déshumanisation du métier, ce sont également les humains eux-mêmes : « The more relevant point here is that it is how the technology is created, or implemented, that has a dehumanising effect. Technology created without consideration for the task or end users removes those end users from the equation » (p. 109). En outre, si l’on considère les conclusions des travaux de Vieira (2020), ce n’est pas tant l’outil en soi qui constituerait une menace, mais bien les conséquences des pratiques commerciales qui en découlent dans le monde professionnel et les limites de cette technologie : « [M]ost criticism of MT concerned primarily not a fear of being outperformed by MT systems or an intrinsic aversion to the technology, but rather MT’s current limitations and some of the business practices that surround its use » (p. 16).

De plus, comme le soutient Moorkens (2022b) dans son article « Le traducteur, une espèce en voie de disparition ? » paru très récemment dans *le Courrier de l’Unesco*, il y aurait même un risque que « [c]es méthodes de travail fortement automatisées ne finissent par rendre le métier de traducteur peu attractif » (s.p.). Sur ce point, Alonso et Vieira (2020) insistent sur la nécessité d’accorder une plus grande place à de telles considérations économiques dans le débat sur l’avenir de la traduction et des technologies de la traduction : « [I]ssues of [...] economic fairness are increasingly important aspects of ensuring that translation technology’s benefits are reaped in a positive and sustainable manner » (p. 400).

Do Carmo et Moorkens (2020) remettent cependant en question ce discours qui témoigne d'une telle résistance : « The resistance narrative comes from translator perception and from an industry view of post-editing as a low-cost, low-skill revision task. [...] There are enough indications [...] that th[is] resistance narrative may not correspond to reality » (p. 38). En s'appuyant sur des données chiffrées, ils estiment que la pratique de la PE est, en effet, bien plus intégrée que ce que l'on pourrait croire : « the numbers show that [...] PE is much more widely used than the resistance narratives lead us to think » (*ibid.*, p. 39). Néanmoins, alors que l'opinion générale qui se dégage chez les traductrices au sujet de la TA et de la PE semble plutôt négative, tous les travaux sur le sujet n'abondent pas dans ce sens. Nous vous proposons de passer à présent en revue les principales études qui comprennent une analyse des perceptions et de l'attitude des (futurs) traductrices à l'égard des outils de TA.

Guerberof Arenas (2013) a soumis un sondage de perception de la post-édition (anglais-espagnol) à trois réviseuses, ainsi qu'à 24 traductrices professionnelles qui avaient une certaine expérience en post-édition même si toutes ne la pratiquent pas de manière régulière. Les avis qu'elle a récoltés sur la PE sont partagés : neuf participantes ont déclaré ne pas apprécier la post-édition contre sept qui disent apprécier cette pratique et huit autres qui se montrent indifférentes. Concernant la perception des efforts, 40 % des répondantes ont trouvé que la PE nécessitait davantage d'efforts que la TH et 20 % pensent l'inverse. 30 % estiment que ces deux méthodes sont équivalentes en termes d'efforts et enfin, 10 % ont répondu qu'elles ne savaient pas se prononcer. Tandis que toutes les traductrices, sauf une, se disent satisfaites de manière générale de leur métier, elles se disent peu satisfaites de la rémunération qu'elles perçoivent en PE.

Carl et al. (2015) ont réalisé un questionnaire qui a été soumis à 12 traductrices professionnelles et à 12 étudiantes de langue maternelle allemande afin de connaître leur avis sur la tâche de PE (anglais-allemand) qu'elles venaient d'effectuer. En s'appuyant sur le Tableau 73, elles concluent que, dans l'ensemble, les participantes ont fait montre d'une perception subjective négative à l'égard de la PE. Remarquons que les étudiantes se sont montrées davantage satisfaites que les traductrices professionnelles.

	Highly satisfied	Somewhat satisfied	Neutral	Somewhat dissatisfied	Highly dissatisfied
Professionals	9%	45.5%	9%	27.5%	9%
Students	8%	67%	8%	17%	0
Overall	9%	56%	9%	22%	4%

Tableau 73 : Degré de satisfaction par rapport à la tâche de PE (Carl et al., 2015, p. 155)

Les études qui suivent portent toutes sur les perceptions non plus de traductrices professionnelles mais bien d'étudiantes.

Dans le cadre de leurs recherches menées avec 30 étudiantes chinoises en master, Jia et al. (2019) ont soumis deux questionnaires (pré- et post-expérience) aux participantes afin de

connaître leur opinion sur la PE (anglais-chinois) et leur ressenti sur la tâche de PE qu'elles venaient d'effectuer. De manière générale, les étudiantes se sont montrées très positives concernant la PE de TA neuronale : « Most of the participants had a very positive attitude towards post-editing before the experiment, although they all had almost no former post-editing experience » (*ibid.*, pp. 77–78). Dans ces questionnaires, l'ensemble des étudiantes a déclaré que la PE leur a permis d'être plus rapides qu'en TH. De plus, la majeure partie d'entre elles pense que la PE implique moins d'efforts cognitifs que la TH. Concernant la qualité des productions finales, 26 étudiantes sur 30 considèrent que la qualité en TH reste supérieure à la qualité en PE. Enfin, la majorité des participantes se montre en faveur d'une formation en PE et de son intégration dans les cursus ; elles envisagent également la pratique professionnelle de la PE de manière positive : « [Students] agreed or strongly agreed that post-editing would provide them with new sources of work and new professional skills » (*ibid.*, p. 77). Ajoutons que ces résultats ne nous paraissent guère surprenants, compte tenu du degré d'acceptation et d'intégration des outils numériques dans la société chinoise, et particulièrement au sein de la jeune génération, férue de nouvelles technologies.

Yamada (2015) a mené une expérience avec 43 étudiantes en langues et communication issus de deux universités japonaises et leur a demandé de juger ensuite de la complexité de la tâche de PE (anglais-japonais) par rapport à la traduction humaine. En résumé, 74 % d'entre elles ont trouvé que la PE était plus aisée que la traduction humaine et les participantes ont également estimé que la PE leur avait permis de réduire d'environ 25 % les efforts de traduction.

Avant de démarrer leur expérience comparative de PE et de révision (italien-anglais), Scansani et al. (2019) ont soumis un questionnaire aux 47 étudiantes du master en traduction spécialisée de l'Université de Bologne. Il est apparu que toutes les participantes ont exprimé une opinion favorable à la TA. Dans cette expérience, les chercheuses italiennes se sont penchées sur le concept de confiance en TA sous l'angle de la productivité : « Investigating how trust towards MT influences translator trainees' behaviour towards the output [...], is [...] crucial to evaluate the likelihood that translators convincingly embrace MT » (*ibid.*, p. 73). Pour ce faire, elles ont voulu voir si l'apprenante en traduction se comportait de la même manière lorsqu'elle pensait devoir post-éditer de la TA neuronale que lorsqu'elle pensait être amenée à réviser une TH. Elles ont mesuré l'effort temporel et l'effort technique grâce aux indices HTER¹⁵⁴ et WPS¹⁵⁵. Leurs résultats ne rendent pas compte d'un manque de confiance de la part des étudiantes à l'égard de la TA. Ainsi, il semblerait que la prochaine génération de traductrices ne s'embarrasse pas de préjugés à l'égard de la PE : « after receiving training on this new technology and before

¹⁵⁴ HTER (*Human-targeted Translation Error Rate*) (Snover et al., 2009) : Alors que l'indice TER fait appel à une traduction de référence, HTER consiste à évaluer la distance d'édition entre une traduction candidate générée automatiquement et une version post-éditée par un traducteur humain (qui servira de traduction de référence), cf. section 2.3.7.3.

¹⁵⁵ Word per second

entering the translation industry, a new generation of translators does not seem to be affected by prejudice against PEMT as much as one could expect » (*ibid.*, p. 78).

Moorkens (2018) a mené deux activités pédagogiques avec 46 étudiantes de 2^e année en traduction et neuf doctorantes en traductologie de l'Université de la ville de Dublin. Ces étudiantes ont effectué une évaluation comparative entre la TA statistique et la TA neuronale pour diverses combinaisons de langues ¹⁵⁶ en utilisant trois méthodes d'évaluation (effort de PE, adéquation et annotation d'erreurs). À la suite de ces activités, *elles* ont pu exprimer leur ressenti : il apparaît que la majorité d'entre *elles* se sont montrées positives à l'égard de la technologie et seraient prêtes à travailler avec la TA neuronale à l'avenir. *Elles* ont été particulièrement étonnées par la qualité et par la fluidité en TA neuronale, mais évoquent tout de même des problèmes d'omission et des erreurs de traduction (*mistranslation*¹⁵⁷). Enfin, *elles* ont apprécié la tâche de PE qui leur a paru plus aisée que la TH.

Par le biais de questionnaires, Rossi (2019a) s'est penchée sur « les craintes des étudiants au regard des évolutions actuelles de la TA et de leur perception de ces phénomènes comme une menace pour le traducteur » (p. 100). Elle fait d'ailleurs remarquer que peu de travaux sont en effet consacrés aux perceptions et appréhensions des étudiantes en matière de TA. Toutefois, elle en mentionne deux (Doherty et Moorkens, 2013 ; Koskinen et Ruokonen, 2017, p. 18) dont il ressort un *a priori* défavorable chez ce groupe cible avec peu ou pas d' « expérience d'usage » en TA. Rossi (2019a) ajoute qu'il a été démontré que l'expérience d'usage pouvait avoir un rôle dans l'acceptation de nouvelles technologies informatiques (Venkatesh et Bala, 2008, cités dans *ibid.*, p. 100) et que « des travaux ont également montré que les traductrices ayant l'expérience d'environnements de traduction assistée par ordinateur étaient plus susceptibles d'accepter d'utiliser la traduction automatique (Lagoudaki, 2008, p. 266 ; Doherty et Moorkens, 2013) » (*ibid.*, pp. 99–100). Afin de tester ce lien entre expérience d'usage et perceptions, Rossi a élaboré et donné un cours d'introduction de 12h sur la TA et sur la PE (anglais – français) à des étudiantes du master de Traduction spécialisée multilingue de l'Université Grenoble Alpes, mélangeant des aspects théoriques et des aspects pratiques. Dans le but de savoir si les perceptions des 19 étudiantes avaient changé avec cette expérience, elle leur a demandé de remplir deux questionnaires avant et après le cours. Bien que ce cours ait manifestement permis aux étudiantes de se faire une opinion plus tranchée en la matière, Rossi n'a toutefois pas pu constater d'évolution des perceptions défavorables telle qu'elle l'espérait. Parmi ces 19 étudiantes, 14 ont effectivement changé d'avis : sept étudiantes ont exprimé une opinion plus favorable au terme de l'expérience tandis que les sept autres ont vu leurs craintes renforcées.

¹⁵⁶ « Language pairs chosen by cohort one were English to French (8 students), French to English (19 students), German to English (13 students), Spanish to English (4 students), and English to Spanish (2 students). Cohort 2 chose Chinese to Spanish, Arabic to English, and English to Spanish (2), Chinese (2), Turkish, and Russian » (Moorkens, 2018, p. 381).

¹⁵⁷ « Mistranslations: incorrectly translated word, wrong gender, number, or case » (Moorkens, 2018, p. 380)

Penchons-nous à présent sur la recherche de Dirand et Rossi (2019). Elles ont évalué les contacts réels d'étudiantes avec la TA par le biais d'un bref sondage en ligne contenant des questions fermées (Figure 72). La majeure partie des étudiantes *interrogées* est issue de licences de langues étrangères appliquées (LEA) en France, le reste des étudiantes provient de diverses filières en sciences humaines et sociales (droit, économie et gestion, sciences sociales, psychologie, histoire, etc.). Leur étude révèle un usage de plus en plus répandu des outils de TA chez les étudiantes ; sur les 306 étudiantes interrogées, seules quatre ont déclaré ne jamais avoir utilisé un moteur de TA ; les autrices concluent donc à une « familiarité indéniable des étudiants avec ce nouvel outil que constitue désormais la traduction automatique » (p. 70). Et sur les 283 étudiantes en langues, alors que plus de la moitié est d'avis qu'aucun outil technologique ne peut actuellement remplacer totalement une traductrice humaine pour traduire un texte de qualité publiable, 41 % pensent que cela sera possible dans 10 ans. Elles en déduisent que « les étudiants sont donc tous sensibles aux promesses que font les fabricants [au sujet des nouvelles technologies de la traduction dont la TA], et qui sont souvent relayées dans la presse » (*ibid.*, p. 70).

LES USAGES ACTUELS		
1. Etudiants de différentes filières (306)		
Dirand et Rossi (Biotraducteur et traducteur automatique : l'homme outillé et la machine-homme, <i>Des mots aux actes</i> 2020)		
Question	Réponse des étudiants en langues (219)	Réponse des autres étudiants (87)
Quand vous entendez le mot « traducteur », à quoi pensez-vous en premier ?	Une personne dont le métier consiste à traduire (76%)	Un logiciel de traduction (56%)
Pensez-vous qu'un outil technologique peut totalement remplacer un humain pour traduire un document diffusable au grand public de nos jours ?	Non, jamais (56%) Oui mais dans 10 ans (41%) Oui (3%)	Oui, mais dans 10 ans (60%) Non, jamais (31%) Oui (9%)
Avez-vous déjà utilisé un logiciel de traduction automatique comme Google Traduction ou une application mobile ?	Oui (98%) Pour comprendre une langue étrangère (82%) Pour écrire dans une langue étrangère (66%) Dans le cadre d'un travail universitaire (60%)	Oui (99%) Pour comprendre une langue étrangère (90%) Pour écrire dans une langue étrangère (29%) Dans le cadre d'un travail universitaire (19%)

Figure 72 : Les usages actuels (Rossi, 2019b, diapositive 8)

Pour terminer ce tour d'horizon, intéressons-nous de nouveau à l'expérience de Loock et Léchaugette (2021) à la suite de laquelle les chercheuses ont invité les étudiantes de 3^e année en LEA de l'Université de Lille à remplir, sur une base volontaire, un questionnaire anonyme concernant leur usage des outils de TA. Sur les 89 étudiantes ayant répondu, 83 % ont déclaré avoir recours à des outils de TA en ligne pour leurs préparations de cours, dont

10 % y auraient recours systématiquement. Les outils les plus employés sont *DeepL*, *Reverso* et *Google Traduction*. Il est intéressant de noter que seuls 5 % d'entre elles disent copier et coller l'ensemble de TS dans le moteur de TA, tandis que d'autres procèdent plutôt par phrases, syntagmes ou même par mots, ou encore que certaines traduisent d'abord le TS sans s'aider de la TA avant de comparer leur production à la sortie de la TA. Enfin, 93 % des sondées ont affirmé être capables de repérer et de corriger les erreurs de TA. Et pourtant, les résultats de cette étude montrent un vrai décalage entre leurs perceptions et leur capacité réelle à repérer et à corriger les erreurs de la TA : « While the majority of students think that MT tools cannot replace translators and that they are capable of identifying the errors in the MT output, our results show that students are actually quite far from having developed a sufficient critical use of MT » (*ibid.*, p. 214). En effet, l'analyse détaillée des erreurs révèle que les étudiantes ne sont pas parvenues à repérer la plupart des erreurs en TA ; elles ont repéré en moyenne moins de huit erreurs sur 20 et seulement deux tiers ont été correctement post-éditées.

7.4 Méthodologie

En nous appuyant sur cette littérature et en nous fiant à l'avis de Daems (2016) selon lequel « [s]urveys are the best way to gain insight into translators' experience with and attitude towards machine translation and post-editing » (p. 76), nous avons choisi d'adopter une méthode de recherche descriptive en interrogeant directement les apprenantes en traduction au moyen de sondages d'opinion. Pour chacune des deux expériences menées (en 2018 et en 2021), nous avons élaboré deux courts sondages que les répondantes devaient remplir en ligne grâce à l'outil de sondages de l'Université de Genève : LimeSurvey (*Limesurvey GmbH*). En outre, afin de voir si l'opinion des étudiantes a changé au cours de l'expérience, nous leur avons proposé de remplir un premier questionnaire avant de prendre part à l'expérience (ci-après « sondage pré-expérience ») et un deuxième questionnaire après avoir participé (ci-après « sondage post-expérience »). Nous avons également décidé de soumettre le sondage pré-expérience à d'autres étudiantes de la filière traduction-interprétation de l'Université de Liège en 2018, ainsi qu'en 2021 afin de pouvoir disposer d'un plus large échantillon et de pouvoir comparer les résultats de ces différents groupes de répondantes.

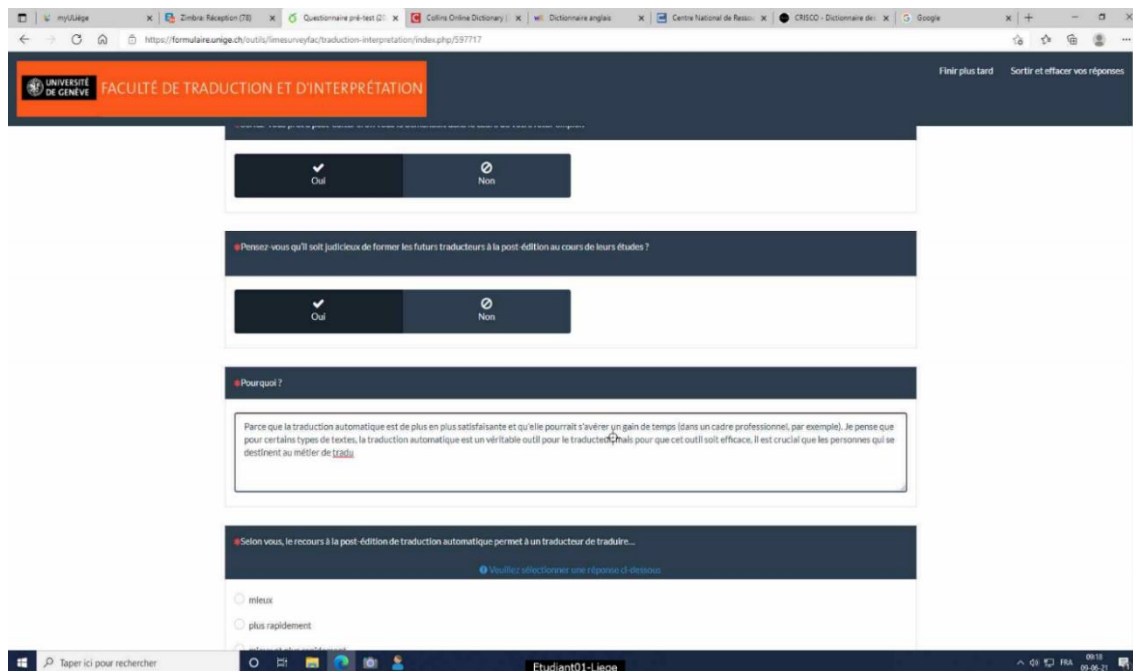


Figure 73 : Capture d'écran – sondage via LimeSurvey

7.4.1 Composition du sondage pré-expérience

Le sondage pré-expérience est composé de dix questions et quatre sous-questions fermées, ainsi que de deux sous-questions ouvertes. Avec ce sondage, nous entendons répondre aux hypothèses HB1 à HB5. La première partie de ce questionnaire est consacrée à des questions portant uniquement sur la traduction automatique. Les questions et sous-questions 1, 2, 2.1 et 2.2 vont nous permettre de répondre à l'hypothèse HB1 ; les questions et sous-questions 2.3, 3 et 4 à l'hypothèse HB3. Alors que la deuxième partie de ce sondage porte sur la post-édition, les questions et sous-questions 5, 6 et 6.1 permettent de vérifier l'hypothèse HB2. Nous répondrons à l'hypothèse HB4 grâce aux questions 7 et 10 ; enfin, les questions et sous questions 8, 8.1 et 9 nous aideront à tester l'hypothèse HB5.

7.4.2 Composition du sondage post-expérience

Le sondage post-expérience est composé de neuf questions et sous-questions fermées et de trois questions et trois sous-questions ouvertes. Grâce à ce sondage, nous entendons vérifier les hypothèses HB6 à HB9. Les questions 1 à 3, 12 et 13 vont nous permettre de répondre à l'hypothèse HB8, les questions 4 à 9 et la question 11 à l'hypothèse HB7, la question 10 à l'hypothèse HB6 et les questions 14 et 16 à l'hypothèse HB9.

La composition exhaustive des deux sondages est accessible sur <https://doi.org/10.7910/DVN/R8EJKN>

7.4.3 Participantes

7.4.3.1 Fournée 2018

Dans cette première fournée, les étudiantes en traduction avaient entre 19 et 30 ans, *elles* sont pour la grande majorité de langue maternelle française à l'exception d'une étudiante germanophone parmi les 28 participantes en BAC3 et de deux étudiantes anglophones, deux étudiantes hispanophones, d'une étudiante germanophone et d'une étudiante de langue maternelle bosniaque parmi les 77 étudiantes de M1 et M2.

7.4.3.1.1 Sondage pré-expérience

En 2018, le formulaire pré-expérience en ligne a été soumis automatiquement aux 28 étudiantes (21 filles et 7 garçons) participant à l'expérience et inscrites en BAC3 en traduction et en interprétation de l'Université de Liège au moment où nous avons mené notre expérience. L'ensemble de ces étudiantes ont rempli ce formulaire jusqu'au bout. Ce questionnaire a également été soumis à l'ensemble des étudiantes de M1 et M2 de la filière pour l'année académique 2017-2018, il a été entièrement rempli et validé par 77 d'entre *elles* (55 filles et 22 garçons).

7.4.3.1.2 Sondage post-expérience

En 2018, le sondage post-expérience a uniquement été soumis aux 28 étudiantes de BAC3 ayant participé à l'expérience comparative de TH et de PE.

7.4.3.2 Fournée 2021

Dans cette deuxième fournée, les sondées avaient entre 19 et 35 ans. *Elles* sont *toutes* francophones (langue maternelle), à l'exception d'une étudiante anglophone et d'une étudiante bilingue français-néerlandais parmi les étudiantes de BAC3 et de M1.

7.4.3.2.1 Sondage pré-expérience

En 2021, le sondage pré-expérience a été envoyé aux 24 étudiantes (19 filles et 5 garçons) participant à notre expérience et inscrites en M2 en traduction de l'Université de Liège, toutes l'ont entièrement rempli. Parallèlement, il a également été soumis aux étudiantes de BAC3 et de M1 de la filière pour l'année académique 2020-2021. Nous avons récolté 50 questionnaires entièrement remplis et validés (43 filles et 7 garçons).

7.4.3.2.2 Sondage post-expérience

En 2021, le sondage post-expérience a une fois de plus été soumis uniquement aux 24 étudiantes de M2 ayant participé à l'expérience comparative de TH et de PE.

7.5 Résultats

En additionnant les différents groupes de répondantes au sondage pré-expérience, nous disposons de 105 questionnaires remplis en 2018 et de 74 en 2021. Nous avons pu ainsi récolter un total de 179 avis d'étudiantes. Le sondage post-expérience a été entièrement complété par toutes les étudiantes auxquelles nous l'avions soumis pour un total de 52 questionnaires. Nous nous attachons à comparer les fournées 2018 et avons veillé, autant que possible, à présenter en premier lieu les réponses données par la fournée 2018 et en deuxième lieu celles données par la fournée 2021. Dans un souci de clarté, les résultats sont présentés en tableaux et par question. Dans l'analyse, ils sont exprimés essentiellement en pourcentage à des fins de comparaison.

Abréviations

Par souci de lisibilité, nous utilisons les abréviations suivantes :

BAC3 : 3^e année de bachelier

M1 : 1^{re} année de master :

M2 : 2^e année de master

HMT : tâche de traduction humaine

SMT : post-édition de traduction automatique statistique (*BingTranslator*)

NMT : post-édition de traduction automatique neuronale (*DeepL*)

GMT : post-édition de *Google Traduction*

DMT : post-édition de *DeepL*

7.5.1 Sondage pré-expérience

A) Traduction automatique

❖ Question 1 – De quel(s) moteur(s) de traduction automatique avez-vous déjà entendu parler ? (Plusieurs réponses possibles)



Figure 74 : Sondage pré-expérience (Question 1)

Question 1								
De quel(s) moteur(s) de traduction automatique avez-vous déjà entendu parler ? (Plusieurs réponses possibles)								
	2018				2021			
	BAC3		M1 et M2		M2		BAC3 et M1	
	Nbr	%	Nbr	%	Nbr	%	Nbr	%
<i>Google Traduction</i>	28	100 %	76	99 %	24	100 %	50	100 %
Systran	2	7 %	18	23 %	0	0 %	4	8 %
Bing Translator	9	32 %	26	34 %	7	29 %	10	20 %
Reverso	27	96 %	73	95 %	23	96 %	50	100 %
Microsoft Translator Hub	0	0 %	4	5 %	0	0 %	2	4 %
<i>DeepL</i>	12	43 %	51	66 %	24	100 %	50	100 %

Tableau 74 : Sondage pré-expérience (Question 1)

En 2018, les deux moteurs de TA les plus connus des étudiantes sont *Google Traduction* (100 % et 99 %) et *Reverso* (96 % et 95 %). En revanche, *DeepL*, le moteur de TA proposé par *Linguee*, ne se place qu'en 3^e position (43 % chez les BAC3 et 66 % chez les M1), très probablement car ce moteur de TA, lancé publiquement fin août 2017, commençait seulement à se faire connaître au moment où les étudiantes ont répondu à ce questionnaire en février 2018. *Bing Translator* est connu par un peu plus d'un tiers d'entre elles (32 % et 33 %). La technologie *Systran* est à première vue plus connue des étudiantes de M1 et M2 (23 %) que des étudiantes de BAC3 (7 %). Enfin, parmi les étudiantes de BAC3, personne n'a déclaré connaître *Microsoft Translator Hub*, et elles sont seulement 5 % à connaître ce moteur de TA en M1 et M2.

Pour 2021, on retrouve à peu près la même ventilation des réponses. Notons toutefois, sans grande surprise, une nette augmentation pour *DeepL* (100 %) qui est désormais connu de l'ensemble des sondées, tout comme l'est toujours *Google Traduction* (100 %). Tandis que les moteurs *Systran*, *Bing Translator* et *Reverso* sont en perte de vitesse ; il apparaît qu'avec le temps, ces trois outils sont de moins en moins connus des étudiantes.

❖ Question 2 – Avez-vous déjà utilisé un moteur de traduction automatique ?

Question 2								
Avez-vous déjà utilisé un moteur de traduction automatique ?								
	2018				2021			
	BAC3		M1 et M2		M2		BAC3 et M1	
	Nbr	%	Nbr	%	Nbr	%	Nbr	%
Oui	25	89 %	74	96 %	24	100 %	50	100 %
Non	3	11 %	3	4 %	0	0 %	0	0 %

Tableau 75 : Sondage pré-expérience (Question 2)

Les réponses figurant dans le Tableau 75 montrent qu'en 2018, la grande majorité des étudiantes (89 % chez les BAC3 et 96 % chez les M1 et M2) a déjà fait usage d'un moteur de TA. Seules six étudiantes parmi les 105 interrogées ont répondu ne jamais avoir utilisé cette technologie, dont trois figurent parmi les participantes à notre pré-test. On note une évolution dans les réponses obtenues en 2021 puisque l'ensemble des 74 étudiantes interrogées a déclaré avoir déjà eu recours à un système de TA.

Note : Les trois sous-questions qui suivent ont uniquement été posées aux étudiantes ayant répondu par l'affirmative à la question 2. Ainsi, dans les tableaux ci-dessous, la ligne « non complété ou non affiché » reprend les étudiantes qui ont indiqué n'avoir jamais fait usage de la TA.

❖ Question 2.1. – À quelle fréquence avez-vous recours à la traduction automatique ?

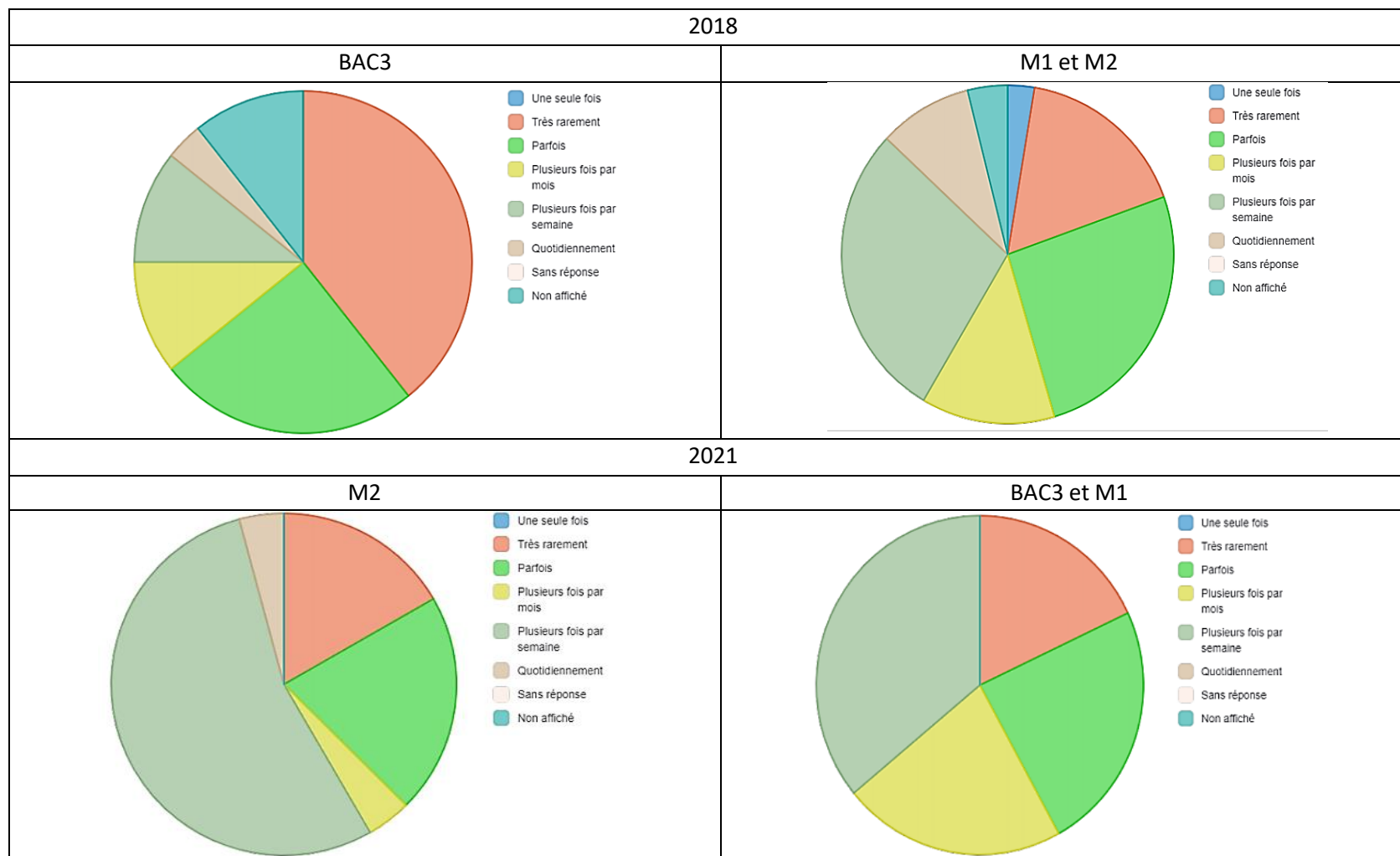


Figure 75 : Sondage pré-expérience (Question 2.1)

Question 2.1.								
À quelle fréquence avez-vous recours à la traduction automatique ?								
	2018				2021			
	BAC3		M1 et M2		M2		BAC3 et M1	
	Nbr	%	Nbr	%	Nbr	%	Nbr	%
Une seule fois	0	0 %	2	3 %	0	0,00 %	0	0 %
Très rarement	11	39 %	13	17 %	4	17 %	9	18 %
Parfois	7	25 %	20	26 %	5	21 %	12	24 %
Plusieurs fois par mois	3	11 %	10	13 %	1	4 %	11	22 %
Plusieurs fois par semaine	3	11 %	22	29 %	13	54 %	18	36 %
Quotidiennement	1	4 %	7	9 %	1	4 %	0	0 %
Non complété ou Non affiché	3	11 %	3	4 %	0	0 %	0	0 %

Tableau 76 : Sondage pré-expérience (Question 2.1)

En 2018, alors que la plupart des étudiantes de BAC3 (39 %) déclarent ne faire usage de la TA que « très rarement », contre 17 % en master, plus d'un quart des étudiantes de master (29 %) disent y avoir recours « plusieurs fois par semaine » contre 11 % en BAC3. Un quart de l'ensemble des sondées déclare utiliser « parfois » la TA. Seule une étudiante de BAC3 (4 %) et 7 étudiantes de master (9 %) ont indiqué avoir un usage quotidien de la TA. De manière générale, il apparaît que le recours à la TA est plus fréquent chez les étudiantes de master que chez les étudiantes de BAC3, même si l'usage quotidien de la TA reste marginal, c'est du moins ce que reflètent les réponses reçues.

En 2021, il apparaît que les habitudes d'utilisation ont évolué et que l'usage s'est intensifié : à présent, plus d'un tiers (36 %) des étudiantes de BAC3 et de M1, ainsi que plus de la moitié (54 %) des étudiantes de M2, estiment avoir recours à la TA plusieurs fois par semaine. Néanmoins, cette fois, seule une étudiante de M2 a admis utiliser quotidiennement la TA.

❖ Question 2.2. – Quel(s) moteur(s) de traduction avez-vous déjà utilisé(s) ?

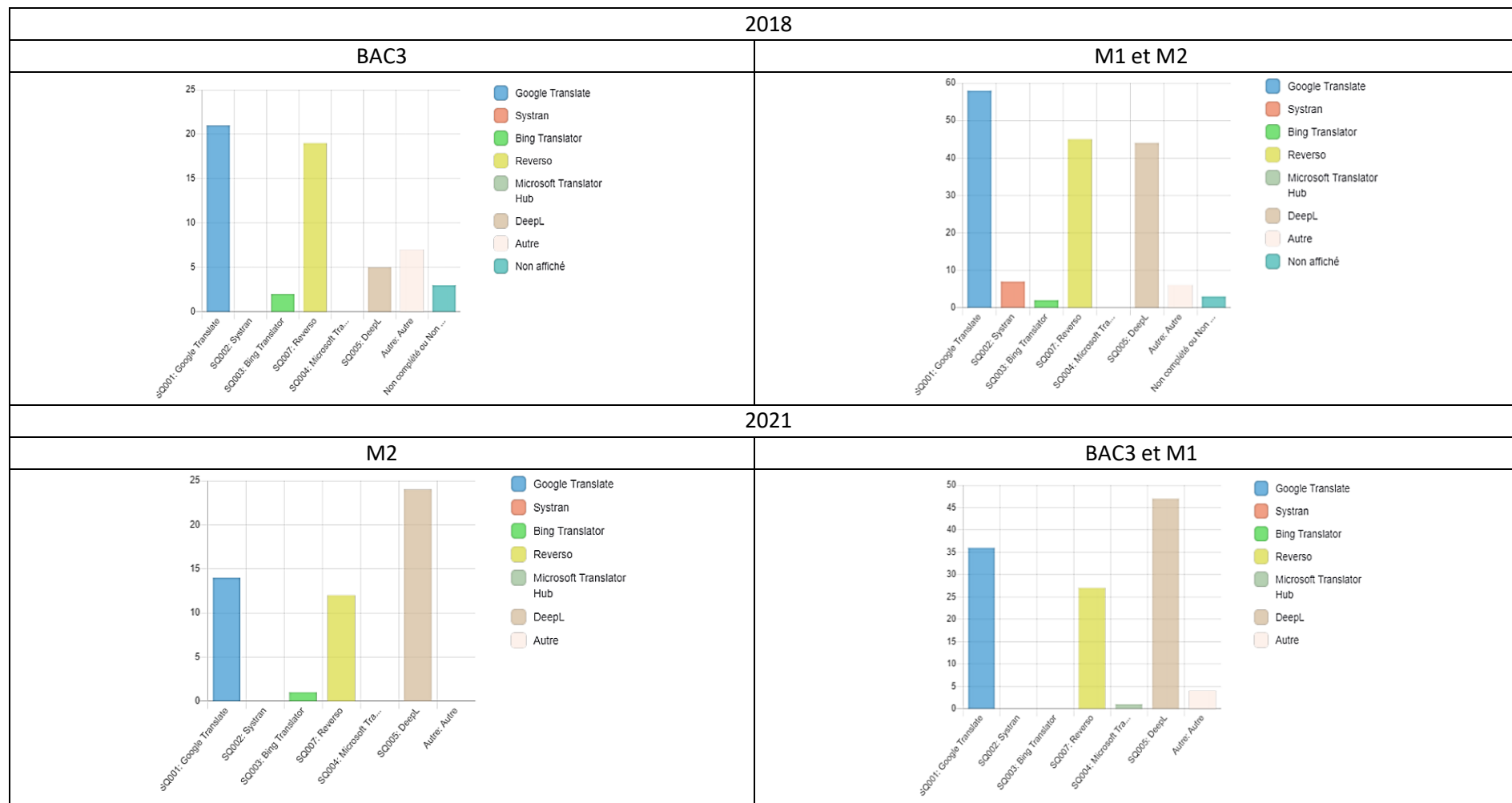


Figure 76 : Sondage pré-expérience (Question 2.2)

Question 2.2.								
Quel(s) moteur(s) de traduction avez-vous déjà utilisé(s) ? (Plusieurs réponses possibles)								
	2018				2021			
	BAC3		M1 et M2		M2		BAC3 et M1	
	Nbr	%	Nbr	%	Nbr	%	Nbr	%
<i>Google Traduction</i>	21	75 %	58	75 %	14	58 %	36	72 %
<i>Systran</i>	0	0 %	7	9 %	0	0 %	0	0 %
<i>Bing Translator</i>	2	7 %	2	3 %	1	4 %	0	0 %
<i>Reverso</i>	19	68 %	45	59 %	12	50 %	27	54 %
<i>Microsoft Translator Hub</i>	0	0 %	0	0 %	0	0 %	1	2 %
<i>DeepL</i>	5	18 %	44	57 %	24	100 %	47	94 %
Autre	7	25 %	6	8 %	0	0 %	4	8 %
Non complété ou Non affiché	3	11 %	3	4 %	0	0 %	0	0 %

Tableau 77 : Sondage pré-expérience (Question 2.2)

D'après les réponses données par les sondées en 2018, les deux moteurs de TA que le plus grand nombre d'entre elles a déjà utilisés sont *Google Traduction* (75 % et 75 %) et *Reverso* (68 % et 59 %). Seules cinq étudiantes de BAC3 (18 %) signalent avoir déjà utilisé *DeepL*. En revanche, elles sont 57 % à l'avoir déjà utilisé en master. Elles sont très peu à avoir déjà employé *BingTranslator* (7 % et 3 %). Dans la catégorie « autre », les étudiantes ont cité : *Pons*, *Wordreference*, *Linguee*, *Vertalen.nu* et *Naver*¹⁵⁸. Nous notons qu'il semble y avoir ici une confusion dans l'esprit des étudiantes entre les dictionnaires traductifs (*Wordreference*, *Vertalen.nu*, etc.), les concordanciers multilingues disponibles en ligne (*Linguee*, etc.) et les moteurs de traduction automatique (*DeepL*, etc.). Cela est certainement dû au caractère équivoque de la question qui leur a été posée. Nous aurions dû préciser : « Quel(s) moteur(s) de traduction **automatique** avez-vous déjà utilisé(s) ? » ; la question a été formulée de cette manière en 2021. En 2021, on constate l'arrivée fracassante de *DeepL* (100 % et 94 %) devant *Google Traduction* (58 % et 72 %) et *Reverso* (50 % et 54 %). Remarquons que, selon ces résultats, 6 % des sondées en BAC3 et en M1 – soit trois étudiantes – n'auraient jamais utilisé *DeepL*. Dans la catégorie « autre », les étudiantes ont cité cette fois : *Memsorce*, *SDL Trados* et *Pons*.

¹⁵⁸ « En juillet 2017, *Naver* a lancé *Papago*, un traducteur mobile basé sur l'IA qui utilise une technologie de réseau neuronal appelée N2MT (*Naver Neural Machine Translation*). Il peut traduire des textes et des phrases dans 14 langues différentes (coréen, anglais, chinois, japonais, espagnol, français, hindi, thaï, allemand, russe, portugais, italien, vietnamien, indonésien) en analysant le contexte au lieu de faire une analyse statistique » (« *Naver* (entreprise) », 2022, s.p.).

❖ Question 2.3. – En général, comment qualifieriez-vous la qualité de la (des) traduction(s) ainsi obtenue(s) ?

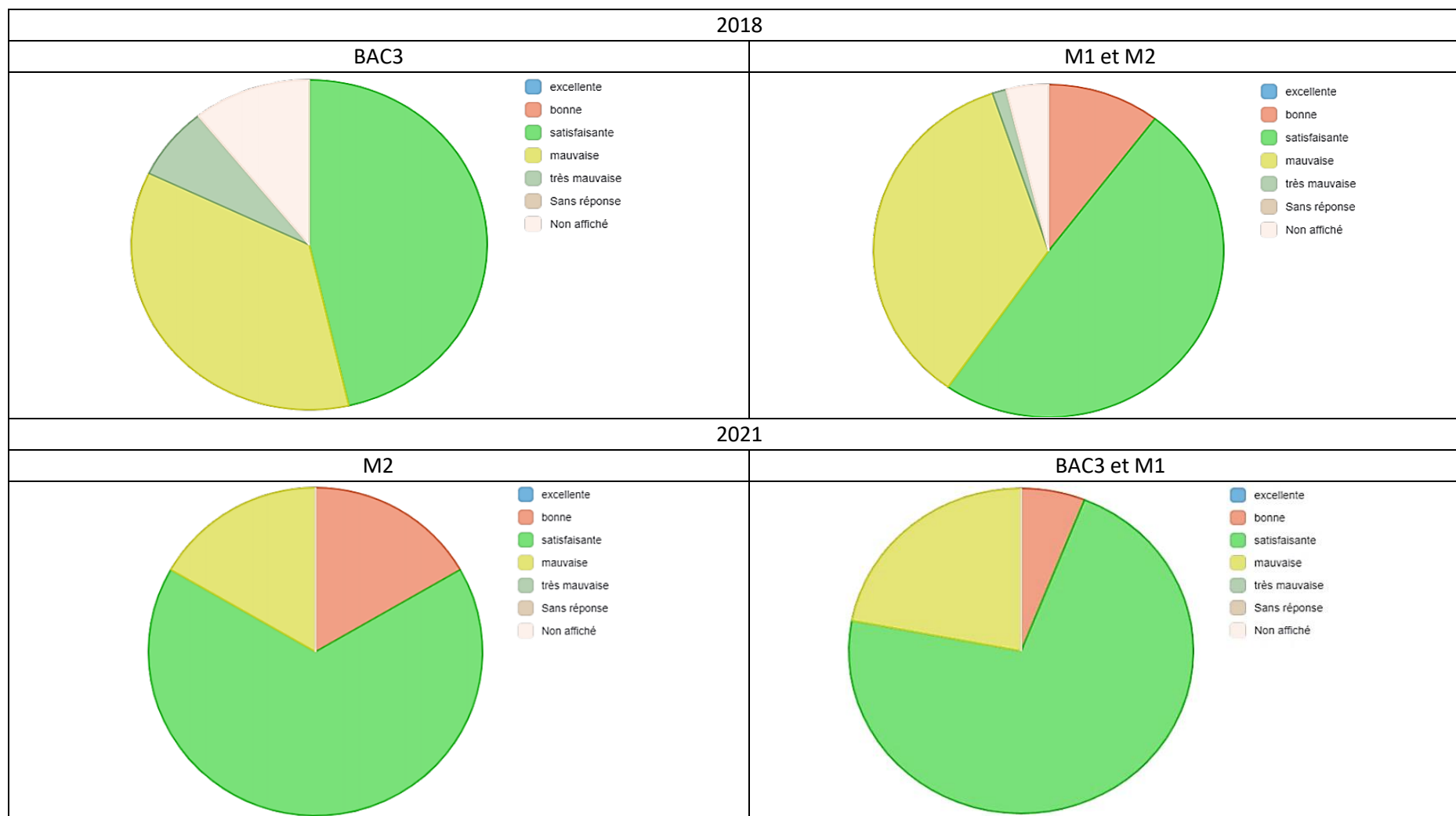


Figure 76 : Sondage pré-expérience (Question 2.3)

Question 2.3.								
En général, comment qualifieriez-vous la qualité de la (des) traduction(s) ainsi obtenue(s) ?								
	2018				2021			
	BAC3		M1 et M2		M2		BAC3 et M1	
	Nbr	%	Nbr	%	Nbr	%	Nbr	%
excellente	0	0 %	0	0 %	0	0 %	0	0 %
bonne	0	0 %	8	10 %	4	17 %	3	6 %
satisfaisante	13	46 %	38	49 %	16	67 %	36	72 %
mauvaise	10	36 %	27	35 %	4	17 %	11	22 %
très mauvaise	2	7 %	1	1 %	0	0 %	0	0 %
Non complété ou Non affiché	3	11 %	3	4 %	0	0 %	0	0 %

Tableau 78 : Sondage pré-expérience (Question 2.3)

En 2018, systématiquement un peu moins de la moitié des étudiantes interrogées (46 % et 49 %) ont indiqué qu'elles considéraient la qualité d'une traduction générée par un moteur de TA comme « satisfaisante ». 10 % des étudiantes de M1 et M2 la considèrent même comme « bonne », tandis que les autres participantes la considèrent comme « mauvaise » (36 % et 35 %). Très peu d'étudiantes jugent cette qualité « très mauvaise » (7 % et 1 %). Enfin, remarquons qu'aucune étudiante de BAC3 ne considère la TA capable de produire des traductions de bonne qualité, tandis qu'elles sont tout de même 10 % à le penser en master.

Elles sont plus nombreuses en 2021 à juger la qualité de la TA « satisfaisante » (67 % et 72 %), voire « bonne » (17 % et 6 %), et elles sont forcément moins nombreuses à la juger « mauvaise » (17 % et 22 %). Enfin, plus aucune étudiante ne considère la qualité d'une traduction générée par un moteur de TA comme « très mauvaise ».

- ❖ Question 3 – Cochez l'affirmation qui vous semble exacte. Pour vous, la qualité d'une traduction dite « conventionnelle » ou « humaine » (sans recours à la traduction automatique) est...

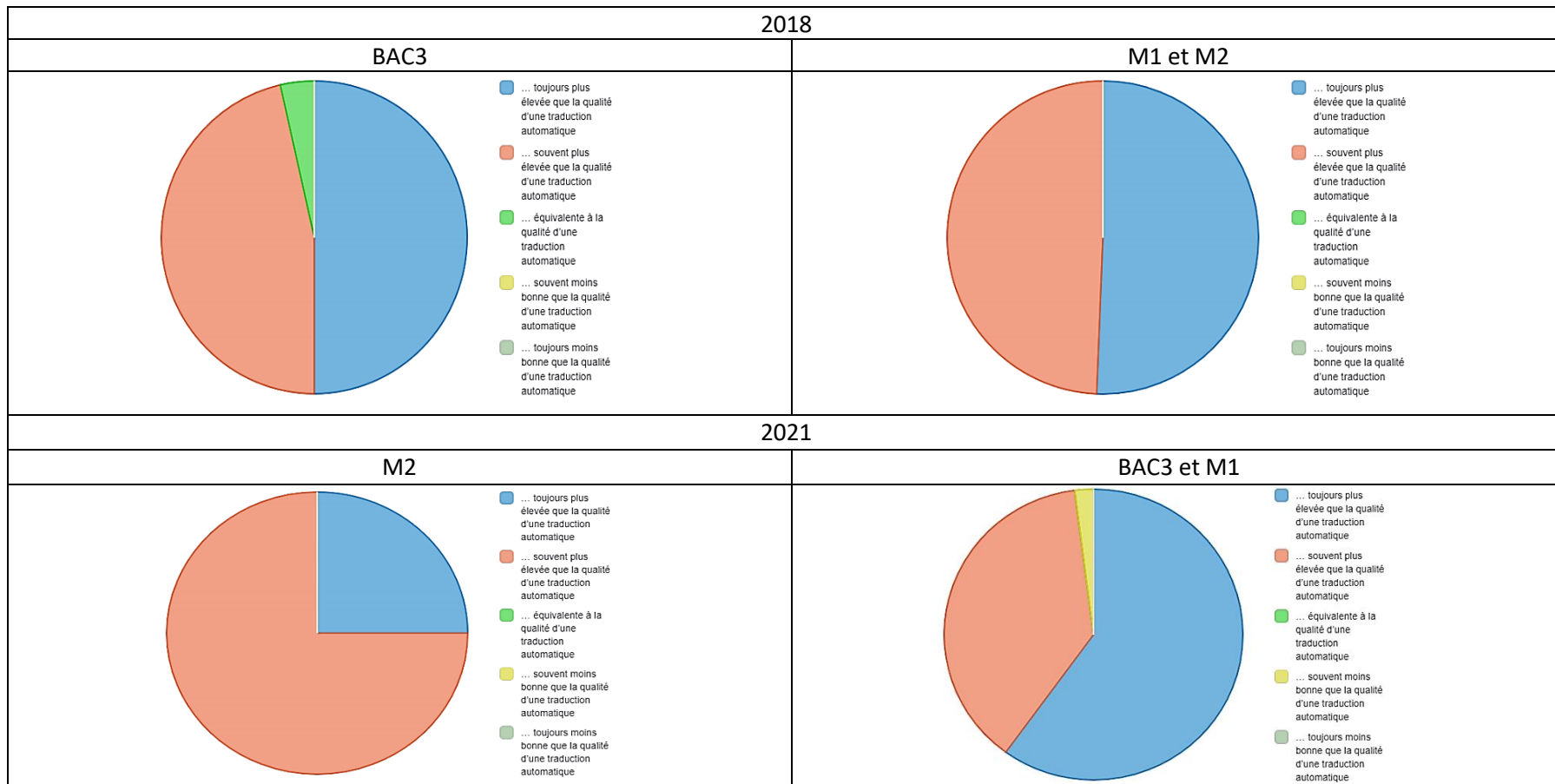


Figure 77 : Sondage pré-expérience (Question 3)

Question 3								
Cochez l'affirmation qui vous semble exacte. Pour vous, la qualité d'une traduction dite "conventionnelle" ou "humaine" (sans recours à la traduction automatique) est...								
	2018				2021			
	BAC3		M1 et M2		M2		BAC3 et M1	
	Nbr	%	Nbr	%	Nbr	%	Nbr	%
... toujours plus élevée que la qualité d'une traduction automatique	14	50 %	39	51 %	6	25 %	30	60 %
... souvent plus élevée que la qualité d'une traduction automatique	13	46 %	38	49 %	18	75 %	19	38 %
... équivalente à la qualité d'une traduction automatique	1	4 %	0	0 %	0	0 %	0	0 %
... souvent moins bonne que la qualité d'une traduction automatique	0	0 %	0	0 %	0	0 %	1	2 %
... toujours moins bonne que la qualité d'une traduction automatique	0	0 %	0	0 %	0	0 %	0	0 %

Tableau 79 : Sondage pré-expérience (Question 3)

Les réponses à cette question sont sans équivoque : pour la très grande majorité des étudiantes interrogées, la qualité de la TH reste, de manière générale, plus élevée que la qualité d'une TA. En 2018, pour 50 % d'entre elles, c'est toujours le cas. Pour un peu moins de la moitié d'entre elles (46 % et 49 %), c'est souvent le cas. Autrement dit, selon ces dernières, il arrive que, parfois, la TA soit de meilleure qualité que la TH. Enfin, une étudiante de BAC3 estime que la TH est de qualité équivalente à la qualité d'une TA.

Alors qu'en 2018 la répartition des réponses entre les affirmations 1 et 2 était équilibrée, ce n'est plus le cas en 2021 : trois quarts des étudiantes de M2 estiment que la qualité de la TH est souvent supérieure à la qualité en TA, et seul le dernier quart estime que c'est toujours le cas. Alors que chez les étudiantes de BAC3 et de M1, les résultats montrent que celles-ci ont une meilleure estime de la TH que de la TA puisqu'elles sont 60 % à trouver que la qualité en TH est toujours plus élevée qu'en TA, et 38 % à penser que c'est souvent le cas. Remarquons tout de même qu'une étudiante de ce groupe juge que la qualité en TH est souvent moins bonne que la qualité en TA.

- ❖ Question 4 – Cochez l'affirmation qui vous semble exacte. Pensez-vous que dans un avenir, proche ou lointain, la qualité d'une traduction automatique sera...

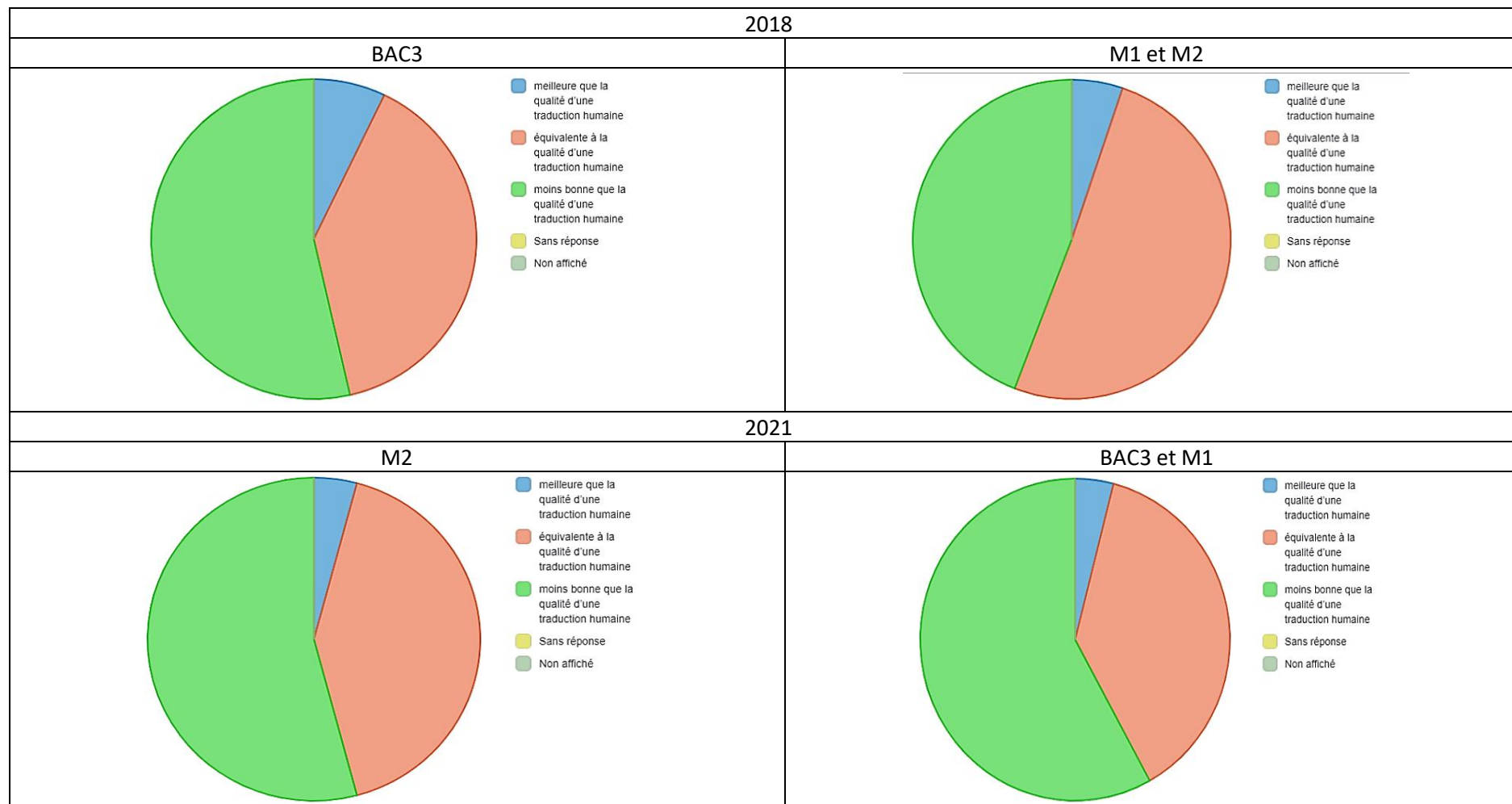


Figure 78 : Sondage pré-expérience (Question 4)

Question 4									
Cochez l'affirmation qui vous semble exacte. Pensez-vous que dans un avenir, proche ou lointain, la qualité d'une traduction automatique sera...									
	2018				2021				
	BAC3		M1 et M2		M2		BAC3 et M1		
	Nbr	%	Nbr	%	Nbr	%	Nbr	%	
meilleure que la qualité d'une traduction humaine	2	7 %	4	5 %	1	4 %	2	4 %	
équivalente à la qualité d'une traduction humaine	11	39 %	39	51 %	10	42 %	19	38 %	
moins bonne que la qualité d'une traduction humaine	15	54 %	34	44 %	13	54 %	29	58 %	
Non complété ou Non affiché	0	0 %	0	0 %	1	4 %	2	4 %	

Tableau 80 : Sondage pré-expérience (Question 4)

Dans cette question, la qualité est entendue de manière générale, sans précisions concernant le genre textuel ou la thématique du TS.

En 2018, alors que les étudiantes restent d'avis que dans le futur, la qualité d'une TA restera inférieure à celle d'une TH (54 % et 44 %), 39 % des étudiantes de BAC3 et 51 % des étudiantes de master pensent que la qualité de la TA sera équivalente à la qualité d'une TH dans un avenir proche ou lointain. Deux étudiantes de BAC3 (7 %) et quatre étudiantes de master (5 %) témoignent d'une foi sans bornes en la technologie et jugent que dans un futur, proche ou lointain, la qualité de la TA dépassera la qualité de la TH.

Nous avons obtenu des réponses similaires en 2021 : plus de la moitié des sondées jugent que la qualité d'une TA restera inférieure à celle d'une TH (54 % et 58 %). 42 % des M2 et 38 % des BAC3 et M1 présumant qu'à l'avenir la qualité en TA sera équivalente à la qualité en TH. Et seules trois étudiantes témoignent cette fois d'un grand optimisme technologique.

Si nous comparons les réponses obtenues pour cette question à celles obtenues à la question 3 (Tableau 79), il apparaît que de nombreuses étudiantes s'attendent à une (nette) amélioration de la qualité des traductions générées par un moteur de TA dans l'avenir, qu'il soit proche ou lointain. Néanmoins, force est de constater que, contrairement à ce que l'on aurait pu croire, et malgré les progrès accomplis en TA entre 2018 et 2021, il y a tout aussi peu d'étudiantes en 2021 qui estiment que la qualité en TA brute devancera un jour la qualité en TH.

B) Post-édition

- ❖ Question 5 – Avez-vous déjà entendu parler de « post-édition » ?

Question 5								
Avez-vous déjà entendu parler de « post-édition » ?								
	2018				2021			
	BAC3		M1 et M2		M2		BAC3 et M1	
	Nbr	%	Nbr	%	Nbr	%	Nbr	%
Oui	11	39 %	50	65 %	22	92 %	44	88 %
Non	17	61 %	27	35 %	2	8 %	6	12 %

Tableau 81 : Sondage pré-expérience (Question 5)

Au moment où nous avons soumis ce sondage, début 2018, deux tiers des étudiantes de BAC3 (61 %) n'avaient jamais entendu le terme « post-édition » contre seulement 35 % des étudiantes de master, soit en tout 44 étudiantes sur 105.

En 2021, ce ne sont plus que 12 % des BAC3 et M1 et 8 % des M2, soit dans l'ensemble huit étudiantes sur 74 qui ne savent pas ce que recouvre ce terme. Ces résultats témoignent certes du fait que la TA et la PE prennent de plus en plus de place sur le marché des services langagiers, mais surtout de l'intégration de la PE dans les cours destinés aux étudiantes en traduction à l'Université de Liège, notamment dans le cours de master « théorie et pratique de la traduction anglais-français ». Il est tout de même interpellant de constater que huit étudiantes ont déclaré ne jamais avoir entendu parler de « post-édition », peut-être s'agit-il d'étudiantes n'ayant pas l'anglais dans leur combinaison de langues. Il est fort à parier, qu'à l'heure actuelle, la totalité des étudiantes en traduction savent ce qu'est la PE, quel que soit leur niveau d'études.

- ❖ Question 6 – Avez-vous déjà post-édité (c.-à-d. révisé) un texte traduit par un moteur de traduction automatique ?

Question 6								
Avez-vous déjà post-édité (c.-à-d. révisé) un texte traduit par un moteur de traduction automatique ?								
	2018				2021			
	BAC3		M1 et M2		M2		BAC3 et M1	
	Nbr	%	Nbr	%	Nbr	%	Nbr	%
Oui	5	18 %	39	51 %	24	100 %	47	94 %
Non	23	82 %	38	49 %	0	0 %	3	6 %

Tableau 82 : Sondage pré-expérience (Question 6)

En 2018, 18 % des étudiantes interrogées en BAC3 – soit cinq étudiantes – ont déclaré avoir déjà post-édité de la TA au moment où ce sondage leur a été soumis, tandis que ce pourcentage s'élève à 51 % chez les étudiantes de master.

En 2021, ce sont 94 % des BAC3 et M1 – soit 47 étudiantes – et les 24 participantes à l'expérience qui ont déclaré avoir déjà post-édité. Entre 2018 et 2021, la tendance s'est ainsi clairement inversée puisque la majeure partie des étudiantes a désormais déjà expérimenté la PE.

❖ Question 6.1. – Dans quel contexte avez-vous post-édité ? Précisez

Question 6.1.								
Dans quel contexte avez-vous post-édité ? Précisez. (Plusieurs réponses possibles)								
	2018				2021			
	BAC3		M1 et M2		M2		BAC3 et M1	
	Nbr	%	Nbr	%	Nbr	%	Nbr	%
contexte académique	5	18 %	34	44 %	24	100 %	46	92 %
contexte professionnel	0	0 %	4	5 %	3	13 %	0	0 %
par curiosité	0	0 %	6	8 %	1	4 %	3	6 %
Autre	0	0 %	2	3 %	0	0 %	2	4 %
Non complété ou Non affiché	23	82 %	38	49 %	0	0 %	3	6 %

Tableau 83 : Sondage pré-expérience (Question 6.1)

Pour le sondage 2018, les cinq étudiantes de BAC3 ayant déjà une expérience en PE ont déclaré avoir post-édité dans un contexte académique. Quant aux étudiantes de master, 44 % ont post-édité dans un contexte académique, seules 5 % l'ont fait en contexte professionnel (stage), 8 % par simple curiosité et enfin, 3 % – soit deux étudiantes – ont déclaré avoir post-édité dans un autre contexte¹⁵⁹.

Pour le sondage 2021, il apparaît que l'ensemble des M2 et 92 % des BAC3 et M1 ont déjà post-édité dans un contexte académique. Trois étudiantes de M2 ont déjà post-édité en contexte professionnel et une seule par curiosité. À nouveau, deux étudiantes ont déclaré

¹⁵⁹ Étudiant 16 : « j'avais corrigé une traduction faite par Google Traduction d'une lettre d'une amie à sa famille d'accueil (dans le cadre d'un Erasmus) » ; Étudiant 73 : « Lorsque je ne comprends vraiment pas une phrase, le logiciel peut m'éclairer parfois ».

avoir post-édité dans un autre contexte¹⁶⁰. Ces usages rappellent les réponses recueillies par González Pastor (2021) : « As to MT adoption, most students reported to have used MT on their own motion at some time for their classroom translation assignments » (p. 57). Par rapport à 2018, nous savons qu'en 2021, la PE était intégrée à différents cours suivis par les étudiantes interrogées, ce qui explique les très hauts taux obtenus en 2021 pour l'usage en contexte académique.

Nous leur avons demandé d'apporter des précisions concernant leur usage de la PE dans une question ouverte. Dans la grande majorité des réponses, il ressort que les étudiantes ont ainsi déjà été exposées à la PE dans le cadre de leur formation universitaire, que ce soit lors d'un cours ou lors de préparations à domicile (González Pastor, 2021 ; Looock et Léchaugnette, 2021). Certains commentaires témoignaient toutefois d'un autre usage ; nous les avons regroupés dans les tableaux ci-dessous. Il s'agit de réponses données par les étudiantes de BAC3 et par les étudiantes de master en 2018 et en 2021 sans distinction.

Gain de temps

Bon nombre d'étudiantes ont indiqué avoir recours à la PE en raison du gain de temps que ce mode de traduction confère, notamment lorsqu'elles manquent de temps pour effectuer les préparations attendues dans le cadre de leur formation.

Identifiant	Commentaire
16	[...] lorsque j'ai beaucoup de préparations en retard. Cela m'évite ainsi de devoir faire des recherches de vocabulaire.
24	J'ai utilisé la post-édition notamment durant mon stage de master II, afin de gagner du temps lorsque la charge de travail était élevée. J'ai également dû réviser un document dans ce même contexte, ce qui s'est avéré en fait être de la post-édition car la cliente qui voulait s'assurer de la qualité de son texte avait clairement eu recours pour sa traduction personnelle à un moteur de traduction automatique.
36	Lorsque je n'ai pas le temps de préparer un texte pour un cours, je le traduis dans un moteur de traduction automatique et ensuite je le révise
49	Quand de multiples traductions sont à faire pour une même journée, un peu de post-édition permet de s'en sortir très bien et de varier l'exercice
68	Quand on me donne plusieurs gros textes à traduire à rendre dans un délais [sic] très court.
73	Lorsque j'avais beaucoup trop de traductions à réaliser pour le même jour.

Tableau 83 : Sondage pré-expérience (Gain de temps)

¹⁶⁰ Étudiant 30 : « Pour traduire un document pour un ami » ; Étudiant 39 : « Post-édition d'un texte anglais pour un ami (qui avait mal utilisé les outils de traduction automatique pour la rédaction de son texte) ».

Accès au sens

Plusieurs commentaires témoignent du fait que les étudiantes ont recours à la PE lorsqu'elles sont confrontées à un problème de compréhension du TS ou lorsqu'elles ne sont pas certaines du sens d'un passage du TS. La TA leur permettrait-elle ainsi d'avoir accès au sens du TS ? Nous avons dans les tableaux ci-dessous quelques commentaires qui illustrent cet usage.

Identifiant	Commentaire
4	Lorsque je n'étais pas certaine du sens de la phrase
6	Lorsque je ne comprenais pas le sens d'une phrase et que la traduction automatique m'a permis de comprendre le sens, sans que la phrase ne soit parfaite.
8	Lorsqu'un paragraphe d'un texte me pose un soucis [sic] durant une traduction, j'utilise souvent un moteur de traduction automatique et je le post-édite ensuite.
73	Lorque [sic] je ne comprends vraiment pas une phrase, le logiciel peut m'éclairer parfois

Tableau 84 : Sondage pré-expérience (Accès au sens)

Comparaison

Enfin, quelques étudiantes ont déclaré qu'il leur arrivait de post-éditer dans une volonté de comparer cette tâche à une TH, ce qui rejoint les usages mis en avant par Loock et Léchaugnette (2021).

Identifiant	Commentaire
2	Post-édition de traductions automatiques pour comparer le temps que cette tâche prend par rapport à une traduction humaine "directe"
36	J'ai déjà post-édité certaines traductions pour comparer le temps de travail entre cette méthode et une traduction "humaine".

Tableau 85 : Sondage pré-expérience (Comparaison)

- ❖ Question 7 – Cochez l'affirmation qui vous semble exacte. D'après vous, en post-éditant, la qualité de la traduction finale est...

Question 7								
Cochez l'affirmation qui vous semble exacte. D'après vous, en post-éditant, la qualité de la traduction finale est...								
	2018				2021			
	BAC3		M1 et M2		M2		BAC3 et M1	
	Nbr	%	Nbr	%	Nbr	%	Nbr	%
...meilleure que si le texte avait été traduit de manière conventionnelle	5	18 %	14	18 %	2	8 %	4	8 %
...tout aussi satisfaisante que si le texte avait été traduit de manière conventionnelle	21	75 %	43	56 %	17	71 %	37	74 %
...moins bonne que si le texte avait été traduit de manière conventionnelle	2	7 %	20	26 %	5	21 %	9	18 %

Tableau 86 : Sondage pré-expérience (Question 7)

En 2018, la majeure partie des étudiantes de BAC3 (75 %) est d'avis que la PE permet d'obtenir un texte cible d'une qualité tout aussi satisfaisante que la qualité d'une TH, tandis qu'*elles* sont un peu plus de la moitié (55 %) à le penser en master. Tandis qu'en 2021, *elles* sont 71 % en M2 et 74 % en BAC3. Les étudiantes sont ainsi davantage satisfaites de la qualité d'une PE en 2021.

En outre, en 2018, plusieurs étudiantes interrogées (18 % et 18 %) sont même convaincues qu'il est possible, en post-éditant, d'obtenir une qualité supérieure à la qualité d'une TH. En 2021, les étudiantes sont un peu moins confiantes avec 8 % en M2 et 8 % en BAC3 et M1. Enfin, deux étudiantes de BAC3 (7 %) et un quart des étudiantes de master (26 %) – soit 20 étudiantes – ont affirmé en 2018 que la qualité d'une PE reste inférieure à celle d'une TH. Alors qu'en 2021, ce taux ne s'élève qu'à 14 étudiantes sur 74 (19 %).

- ❖ Question 8 – Seriez-vous prête à post-éditer si on vous le demandait dans le cadre de votre futur emploi ?

Question 8								
Seriez-vous prête à post-éditer si on vous le demandait dans le cadre de votre futur emploi ?								
	2018				2021			
	BAC3		M1 et M2		M2		BAC3 et M1	
	Nbr	%	Nbr	%	Nbr	%	Nbr	%
Oui	25	89 %	72	94 %	22	92 %	49	98 %
Non	3	11 %	5	6 %	2	8 %	1	2 %

Tableau 87 : Sondage pré-expérience (Question 8)

Que ce soit en 2018 ou en 2021, la majorité des étudiantes interrogées se disent prêtes à post-éditer si on le leur demandait dans le cadre de leur futur emploi. Sur les 179 avis recueillis au total, seules 11 étudiantes ont répondu ne pas être prêtes à le faire. Il apparaît que les étudiantes sont conscientes des changements technologiques que connaissent actuellement le secteur de la traduction et la profession de traductrice, puisque la majorité – voire la quasi-unanimité (en 2021) – d’entre *elles* se disent prêtes à accomplir des tâches de PE dans le cadre de leur futur emploi. Cette tendance confirme les impressions recueillies dans l’étude de Moorkens (2018) et de Jia et al. (2019) : « [Most students] agreed or strongly agreed that post-editing would provide them with new sources of work and new professional skills » (*ibid.*, p. 77).

- ❖ Question 8. 1. – Expliquez brièvement pourquoi.

Cette sous-question n’a été posée qu’aux étudiantes ayant répondu par la négative à la question 8. En raison d’un oubli de notre part dans le questionnaire pré-expérience soumis en 2018, nous ne disposons pas des réponses pour les étudiantes de BAC3. L’ensemble des réponses 2018 et 2021 est repris dans le Tableau 88.

Identifiant (ID)	Réponse
3	Travailler en collaboration avec la machine me semble judicieux, mais je ne gagne rien en post-éditant. Soit parce que je ne fais pas assez attention à la relecture et donc que la traduction reste de moins bonne qualité, soit parce que, au contraire, je fais tellement attention à tous les petits détails que finalement, cela me prend autant de temps qu'une traduction humaine qui pourrait directement être meilleure.
13	Je préfère faire mes recherches de mon côté, et trouver les tournures qui conviennent le mieux à mon style de traduction. On est fort influencé par la proposition de la traduction automatique.
38	5 ans d'études pour vérifier ce qu'un ordinateur écrit.
47	J'ai effectué 5 ans d'études, dont 3 ans où aucun ordinateur n'était permis lors des examens. Cela nous a permis de nous détacher du texte source et d'avoir recours à diverses stratégies. Si nous commençons à revenir en arrière et à ne faire que de la post-edition, soit corriger que le texte traduit, alors je pense notre profession ne devrait plus s'appeler traducteur, mais correcteur. Certes, nous sommes amenés à réviser certaines traduction [sic], mais il ne faudrait pas la révision ne devienne l'unique activité du traducteur.
53	Pas assez stimulant de se contenter de relire et de corriger les éventuelles erreurs
58	Je préférerais faire la traduction moi-même que de relire le travail d'une machine, car je pense que le traducteur humain comprend mieux certaines subtilités.
78	Post-éditer, et donc utiliser un moteur de traduction automatique, c'est ne pas faire justice à un auteur.
81	Je préfère procéder manuellement à la totalité du texte source.

Tableau 88 : Sondage pré-expérience (Question 8.1)

Une étudiante (3) reconnaît qu'elle ne considère pas la PE comme une activité judicieuse (pas de gain de temps, moindre qualité par rapport à une TH). Plusieurs étudiantes (13, 58 et 81) invoquent simplement leur préférence pour la TH. D'autres mentionnent le caractère peu stimulant (53), voire peu gratifiant (38), de la relecture et de la correction de textes produits par une machine par rapport à la TH. Une étudiante (47) considère la PE comme un retour en arrière par rapport à la TH et estime que le rôle de la traductrice ne peut être cantonné à de la révision. Enfin, une étudiante (78) estime que la PE ne permet pas de « faire justice à un auteur » sans donner d'explication complémentaire.

- ❖ Question 9 – Pensez-vous qu’il soit judicieux de former les futures traductrices à la post-édition au cours de leurs études ?

Question 9								
Pensez-vous qu’il soit judicieux de former les futures traductrices à la post-édition au cours de leurs études ?								
	2018				2021			
	BAC3		M1 et M2		M2		BAC3 et M1	
	Nbr	%	Nbr	%	Nbr	%	Nbr	%
Oui	25	89 %	62	81 %	24	100 %	45	90 %
Non	3	11 %	15	19 %	0	0 %	5	10 %

Tableau 89 : Sondage pré-expérience (Question 9)

La vaste majorité des étudiantes interrogées sont convaincues qu’il est judicieux de former les futures traductrices à la PE au cours de leurs études. Remarquons toutefois qu’en 2018, 15 étudiantes de master (19 %) et trois étudiantes de BAC3 (10 %) n’en sont pas convaincues, de même que cinq étudiantes de BAC3 et M1 en 2021. Ainsi, nous pouvons conclure qu’il existe un large consensus parmi les étudiantes ayant été interrogées par nos soins sur la nécessité de former les futures traductrices à la PE et d’intégrer une formation spécifique dans leur cursus universitaire.

- ❖ Question 9.1. – Pourquoi ?

Nous avons demandé à toutes les étudiantes, qu’elles aient répondu positivement ou négativement à la question 9, d’argumenter leur position. Par souci de lisibilité, nous présentons uniquement une sélection, la plus représentative possible, des arguments avancés par les étudiantes. Ces arguments sont répartis en deux grandes catégories (favorables et défavorables), ainsi qu’en sous-catégories. Nous n’avons pas jugé opportun de faire ici la distinction entre les réponses 2018 et 2021 ni entre les catégories d’étudiantes.

Arguments favorables

Quand il s'agit d'expliquer les raisons pour lesquelles les étudiantes se montrent en faveur d'une formation à la PE, *elles* ont évoqué :

A. La nécessité de s'adapter aux évolutions technologiques du métier de traductrice et de répondre aux exigences du marché

Identifiant (ID)	Réponse
7	Nous vivons dans un monde où il est nécessaire de travailler rapidement pour répondre aux délais et aux attentes de nos employeurs, et ce surtout dans le domaine de la traduction. La post-édition est donc à présent primordiale dans la formation d'un traducteur, tout comme la maîtrise de la TAO.
10	Je crois que dans un avenir proche ou lointain, les traducteurs seront engagés essentiellement pour réaliser des tâches de post-édition. Cependant, je pense que la traduction humaine devrait rester au centre de la formation d'un traducteur.
13	Parce que c'est une réalité avec laquelle il va falloir composer. Le métier de traducteur est amené à évoluer avec les nouvelles technologies. Il faut donc former les futurs traducteurs en conséquence
15	Si le métier de traducteur en lui-même disparaît, il faudra bien s'adapter.
16	Car la post-édition ne peut être évitée à l'heure actuelle et que les traducteurs doivent pouvoir s'adapter aux évolutions. Si des clients demandent à des traducteurs de post-éditer, il faut y être préparé un minimum.
17	Car la traduction automatique va continuer à se développer et être formé à la post-édition ouvrira donc plus de portes aux traducteurs
18	Car les nouvelles technologies font partie intégrante de la profession de traducteur. Il est donc important de se tenir à jour et d'avancer avec notre époque.
19	Les traducteurs automatiques sont de plus en plus fréquents, il vaut donc mieux préparer les traducteurs à « corriger » ce que leurs clients auront essayé de traduire eux-même [sic].
22(b)	Je pense que cela représente le futur de la profession. Les traducteurs ne doivent pas chercher à lutter contre la technologie, mais bien l'adapter aux exigences de la profession. Selon moi, une machine ne pourra jamais traduire comme un être humain, mais elle peut grandement faciliter le travail.
23	Parce que les étudiants doivent être prêts à affronter la réalité du marché, qui évolue vers davantage de post-édition.
24	La post-édition est une pratique de plus fréquente et de plus en plus demandée. Il y a peu de chances qu'un (futur) traducteur n'y soit pas confronté. Il est important que nous puissions maîtriser tous les aspects du métier de traducteur afin d'exercer notre travail correctement. La post édition est donc, selon moi, une compétence importante à avoir.
27	L'utilisation de la traduction automatique massive va transformer le métier de traducteur. La post-édition risque de devenir un des derniers débouchés pour les étudiants de traduction.

36	La post-édition va très certainement devenir incontournable dans le domaine de la traduction dans les prochaines années. Déjà aujourd'hui elle est largement utilisée par les entreprises de traduction. Ainsi, former les étudiants à la post-édition leur permettra d'avoir les outils nécessaires à leur vie professionnelle.
46	Les moteurs de traduction permettent d'accélérer le rythme de travail, ils feront tôt ou tard partie du quotidien de tous les traducteurs. La formation doit, selon moi, correspondre [sic] à la réalité pour être professionnalisante. La traduction manuelle ne sera réservée qu'à des domaines particuliers, comme la littérature, auxquels peu de traducteurs auront concrètement accès.
49	La post-édition est, pour moi, tout aussi importante que la traduction vu l'ampleur que prennent les outils de traduction automatiques.
53	Car je pense que ça fait partie de l'avenir de la traduction. Je pense que le métier de traducteur va beaucoup évoluer ces prochaines années. Je pense également que de plus en plus d'employeurs vont exiger des traducteurs qu'ils maîtrisent les outils de post-édition. Je pense aussi qu'il est dans l'intérêt du traducteur de savoir maîtriser le plus d'outils de traduction possible. Si un employeur hésite entre deux traducteurs, il choisira sûrement celui qui est le plus polyvalent.
64	Car c'est le métier de demain.
79	Les techniques vont évoluer et les programmes seront sûrement de mieux en mieux faits. Autant apprendre à évoluer avec son temps et apprendre à travailler avec la technologie.
82	Il faut bien se rendre compte des changements qui se préparent. Refuser l'avènement des traducteurs automatiques et leurs performances toujours plus rapides, dont la qualité ne fait qu'augmenter, serait vouer le métier de traducteur à disparaître complètement. Il est temps d'embrasser le monde moderne dans lequel nous vivons (et serons amenés à vivre). Le traducteur automatique parfait n'est pas encore au point, mais cela ne saurait tarder d'ici une dizaine d'années.

Tableau 90 : Sondage pré-expérience (Question 9.1.) Arguments favorables (A)

À la lecture de ces commentaires, il se dégage chez certaines une vision très pessimiste de l'avenir de la traductrice humaine puisqu'elles évoquent une disparition partielle ou totale des tâches de TH. D'autres encore se montrent très optimistes concernant l'évolution des performances de la TA dans un avenir plus ou moins proche.

B. Le gain de temps actuel ou projeté

Identifiant (ID)	Réponse
2	Les traducteurs automatiques sont de plus en plus performants et post-éditer permet parfois de gagner énormément de temps.
5	La traduction automatique prend de plus en plus d'importance. Elle peut faire gagner beaucoup de temps au traducteur, même si ce n'est pas toujours le cas.
16	Quand les logiciels de traduction automatique seront plus performants, cela permettra de gagner du temps. Pour le moment, on en perd justement trop, car il y a trop de fautes et ça va parfois plus vite de traduire directement.
20	La post-édition pourrait faire gagner du temps aux traducteurs
24	Lorsque je suis vraiment pressée, la post-édition me permet de traduire plus rapidement. Il est possible que l'on nous demande plus tard de traduire dans des délais très ou trop courts et je trouve qu'apprendre à post-éditer pourrait se révéler utile. Cela nous permettrait de traduire peut-être plus rapidement sans pour autant rendre une traduction de mauvaise qualité.

Tableau 91 : Sondage pré-expérience (Question 9.1.) Arguments favorables (B)

C. L'apprentissage de nouvelles compétences

Identifiant (ID)	Réponse
6	Parce que je pense qu'il est intéressant d'être formé dans un maximum de domaines.
11	Il est toujours intéressant de se former à diverses manières de travailler son domaine. Il faut également noter que l'informatique est omniprésente et qu'il est donc important de s'y retrouver.
19	Il est important de se familiariser avec tous les outils mis à notre disposition
39	Il est judicieux de former les futurs traducteurs à la post-édition au cours de leurs études, car la post-édition est une compétence qui sera peut-être demandée au traducteur dans le cadre professionnel. De plus, la post-édition est l'analyse d'une [sic] texte sous un autre angle que la traduction. Recourir à la post-édition en cours offre la possibilité au futur traducteur d'élargir sa vision.
71	Car c'est une autre tâche que de traduire, je pense donc qu'il est utile que les futurs traducteurs soient aptes à post-éditer des textes rapidement et avec une bonne technique.
75	Il est toujours intéressant d'être formé à de nouvelles méthodes de traduction.

Tableau 92 : Sondage pré-expérience (Question 9.1.) Arguments favorables (C)

D. Le développement d'une approche raisonnée des outils de TA

Identifiant (ID)	Réponse
4	Acquérir un regard critique par rapport au résultat généré par un moteur de traduction automatique et acquérir la compétence de "reformulation", car c'est ce dont nous avons le plus besoin en traduction, je pense. On comprend l'idée, mais la transcrire n'est pas le plus aisé.
19	Il s'agit d'un outil qui, bien utilisé, peut se révéler un véritable gain de temps et une aide précieuse. Toutefois, si l'on n'a pas appris à s'en servir correctement, il peut vite devenir "dangereux" et donner de médiocres résultats. Il est donc important d'apprendre à en faire un usage correct pour pouvoir profiter pleinement de ses avantages.
21	[...] Je pense que pour certains types de textes, la traduction automatique est un véritable outil pour le traducteur, mais pour que cet outil soit efficace, il est crucial que les personnes qui se destinent à la traduction soient conscientes des "failles" de cet outil, d'où l'intérêt d'inclure la post-édition à notre formation.
22(a)	La traduction automatique est de plus en plus performante et représente une grande partie de notre avenir, il faut avancer avec les nouvelles technologies et les employer à bon escient.
27	Cela nous montrerait la valeur de la traduction manuelle et les limites de la traduction automatique et cela nous aiderait également à développer un esprit critique par rapport à nos traductions.
35	C'est une pratique de plus en plus courante, elle peut aussi nous être utile lors de traduction [sic] plus "traditionnelles" en fonction du type de texte à traiter (pour être plus rapide ou éviter certains types de fautes). C'est important de pouvoir identifier les problèmes récurrents et de s'approprier une méthode de travail lors de la post-édition.
42	Oui, car la traduction automatique est de plus en plus présente dans le monde de la traduction, mais elle ne fournit pas toujours des traductions "parfaites". C'est là que la post-édition intervient et se montre d'ailleurs très utile. Malheureusement, elle comprend certaines difficultés, à savoir quand faut-il s'arrêter de réviser ? On a tendance à vouloir modifier toujours plus, ne jamais être satisfait, comme si on voulait s'approprier la traduction. C'est pourquoi, une formation à la post-édition serait judicieuse : apprendre à être efficace, à ne pas trop vouloir en faire, à savoir s'arrêter, tout en étant capable de fournir une post-édition de qualité (ne pas faire l'impasse sur des calques, glissements de sens, etc.)
50	Si la post-édition devient plus commune, il faut former les futurs traducteurs à utiliser correctement cet outil. Il ne faut pas voir les traducteurs automatiques comme l'ennemi des traducteurs, mais justement comme une aide à la traduction.
54	Car la post-édition requiert une certaine technique et une capacité de critiquer ce que le logiciel a traduit. Sans expérience ou apprentissage, cela s'avérera difficile et la PE ne serait pas satisfaisante

Tableau 93 : Sondage pré-expérience (Question 9.1.) Arguments favorables (D)

E. L'importance d'apprendre à faire de la révision

Identifiant (ID)	Réponse
21	Parce que tout traducteur doit être capable de réviser une traduction qui n'est pas la sienne sans pour autant recommencer à zéro ou remanier complètement le travail de l'autre traducteur
25	Un traducteur indépendant peut être amené à faire une révision de texte à n'importe quel moment durant sa carrière. Un ami qui a terminé ses études de traduction en juin 2016 a déjà été amené à réviser des textes quelques mois après qu'il ait lancé son activité d'indépendant. C'est pourquoi il serait judicieux de former les futurs traducteurs à la révision des textes.
32	Savoir repérer des fautes dans un texte à réviser est un outil important à apprendre, que la traduction ait été faite par une machine ou une personne.
44	Je pense que réviser une traduction s'apprend autant que traduire. Il est difficile de bien réviser un texte si on n'a jamais appris à le faire comme il est difficile de bien traduire un texte si on a [sic] pas appris les techniques de traduction.

Tableau 94 : Sondage pré-expérience (Question 9.1.) Arguments favorables (E)

Notons que certaines étudiantes ne semblent faire aucune distinction entre la révision de textes et la post-édition.

F. Autres

Identifiant (ID)	Réponse
2	Meilleurs résultats : Le traducteur automatique permet d'éviter certains contresens et d'éviter les oublis. Même s'il est fréquent que le traducteur automatique commette des erreurs, ces erreurs sont facilement repérables lors de la révision.
23	Pour les rassurer. Les traducteurs automatiques sont de plus en plus performants et offrent même parfois une petite banque linguistique qui permet de pallier les imprécisions rencontrées. De ce fait, il est logique de s'inquiéter pour l'avenir de la traduction, or en se penchant sur cette facette de la traduction, le constat est évident : la traduction humaine demeure essentielle à une traduction qui se veut idiomatique et riche.
24	Oui, car les expressions idiomatiques, collocations, cooccurrences, etc. viendront plus naturellement lorsqu'ils seront déjà munis d'une base traduite. Il faudra également se remettre en question.
30	Post-éditer n'est pas aussi simple qu'il y paraît

Tableau 95 : Sondage pré-expérience (Question 9.1.) Arguments favorables (F)

Arguments défavorables

Quant aux arguments avancés en défaveur d'une formation en PE, nous avons relevé dans leurs commentaires :

A. La diminution de la qualité par rapport à la TH et le manque de confiance en la TA

Identifiant (ID)	Réponse
5	Je pense que si on forme des futurs traducteurs à la post-édition, ils auront tendance à y avoir recours plus souvent et cela pourrait diminuer la qualité de leur travail.
52	Peu confiance en la traduction automatique
59	Car la post-édition n'est pas aussi belle qu'une traduction totalement humaine.
81	Car ils ne développent pas les mêmes automatismes et les mêmes capacités qu'un étudiant qui traduit l'entièreté d'un texte par lui-même.

Tableau 96 : Sondage pré-expérience (Question 9.1.) Arguments défavorables (A)

B. La dévalorisation de l'être humain par rapport à la machine (considérations éthiques)

Identifiant (ID)	Réponse
10(a)	Je pense qu'il est important que l'étudiant en traduction soit conscient que son travail sera toujours plus élaboré et plus correct que le travail fourni par un programme de traduction. Plus important encore, il doit sentir son travail valorisé. Or, faire croire à l'étudiant que son travail est valorisé lorsqu'il passe après un programme n'est pas très correct. Dans ce cas-ci, pourquoi ne pas passer directement par l'humain.
10(b)	Parce qu'il est plus intéressant d'être formé à traduire "comme des grands" plutôt que de nous apprendre à corriger un texte écrit par une machine. Il n'y aura plus de traducteurs si on n'apprend plus aux étudiants à traduire...
20	C'est de la traduction normale, il ne faut pas non [sic] former les étudiants à travailler en "collaboration" avec des machines dont le but premier était de remplacer les traducteurs, faut pas rêver non plus.

Tableau 97 : Sondage pré-expérience (Question 9.1.) Arguments défavorables (B)

C. L'importance de former d'abord les étudiantes à traduire avant de les former à post-éditer

Identifiant (ID)	Réponse
7	Je dirais oui et non. Il faudrait d'abord passer par la traduction manuelle pour former les étudiants et ensuite apprendre la post-édition.
38	Il vaut mieux qu'ils s'habituent à traduire manuellement pendant les premières années d'études.
78	Ça les encourageraient [sic] à utiliser cette méthode trop souvent. Ou alors, en Master 1 ou 2, quand les élèves ont déjà bien appris les réflexes de bases et ont découvert leur style de traduction. Je pense qu'utiliser un moteur de traduction automatique ne doit pas être un réflexe, et encore moins dès le bachelier. Cela doit rester une aide, à prendre en dernier recours.

Tableau 98 : Sondage pré-expérience (Question 9.1.) Arguments défavorables (C)

D. La futilité d'une telle formation à la PE

Identifiant (ID)	Réponse
16	Parce qu'à partir du moment où on sait traduire, on sait aussi post-éditer.
22	Je pense que si on peut traduire, on peut réviser. En effet, selon moi, une révision demande moins d'effort qu'une traduction
33	Je pense qu'une formation en Post-Édition n'est pas nécessaire. En revanche, une formation en traduction l'est grandement. À mon sens, quand on sait traduire, on sait post-éditer, pas l'inverse.
37	Si l'on forme des étudiants à réviser des textes de traducteurs, pas besoin de les former à post-éditer.
41	Si un étudiant sait traduire et réviser les textes de ses camarades, il sait aussi post-éditer
65	Pas besoin de formation, la post-édition n'est qu'un outil pour les gens paresseux, une formation classique à la traduction permet de fournir un résultat largement satisfaisant.
74	Cela ne stimule pas la réflexion. Ce ne serait plus de la traduction, mais juste de la correction, je ne vois pas ce que l'étudiant pourrait en retirer.
85	Il me semble que si on sait traduire, on sait post-éditer. Donc les cours de traduction suffisent.

Tableau 99 : Sondage pré-expérience (Question 9.1.) Arguments défavorables (D)

Il est intéressant de constater que plusieurs étudiantes sont convaincues qu'à partir du moment où l'on est capable de traduire ou de réviser, l'on est aussi capable de post-éditer et donc qu'il est inutile de former à la PE. Pour rappel, en 2018, seuls 17 % des étudiantes de BAC3 ont déclaré avoir déjà post-édité de la TA, tandis qu'elles étaient 50 % en master. Ces commentaires laissent à penser que les étudiantes étaient peu familières de ce genre de pratique – du moins en 2018 – et qu'elles n'en connaissaient pas les spécificités.

- ❖ Question 10 – Selon vous, le recours à la post-édition de traduction automatique permet à une traductrice de traduire...

Question 10								
Selon vous, le recours à la post-édition de traduction automatique permet à une traductrice de traduire								
	2018				2021			
	BAC3		M1 et M2		M2		BAC3 et M1	
	Nbr	%	Nbr	%	Nbr	%	Nbr	%
mieux (A1)	0	0 %	2	3 %	0	0 %	4	8 %
plus rapidement (A2)	25	89 %	55	71 %	16	67 %	36	72 %
mieux et plus rapidement (A3)	0	0 %	8	10 %	5	21 %	7	14 %
aucune de ces propositions (A4)	3	11 %	12	16 %	3	12 %	3	6 %

Tableau 100 : Sondage pré-expérience (Question 10)

En 2018 et en 2021, la majorité des sondées a estimé que la PE générait un gain de temps pour la traductrice en lui permettant de traduire plus rapidement, ce qui corrobore les résultats du sondage mené par Daems (2016) auprès d'étudiantes de master en traduction. En revanche, plusieurs répondantes de BAC3 et de master sont d'avis que la PE ne permet ni de traduire plus rapidement ni de mieux traduire (21 étudiantes sur 179). Très peu de répondantes jugent que la PE permet de mieux traduire, sans pour autant générer un gain de temps : deux étudiantes de master en 2018 et quatre répondantes de BAC3 et M1 en 2021.

Et enfin en 2021, 21 % des M2 et 14 % des BAC3 et M1 pensent que la PE engendre un gain de temps, ainsi qu'un gain de qualité puisqu'elles jugent que cela permet à la fois de traduire plus rapidement et de mieux traduire, alors qu'elles étaient seulement 10 % des étudiantes de M1 et M2 à le penser en 2018.

7.5.2 Sondage post-expérience

❖ Question 1 - Quel est votre niveau général de satisfaction pour la tâche de traduction humaine (HMT) ?

Nous avons demandé aux participantes de nous faire part de leur niveau de satisfaction sur une échelle de 1 à 5 (de « pas satisfaite du tout » à « très satisfaite ») pour chacune des trois tâches accomplies durant l'expérience.

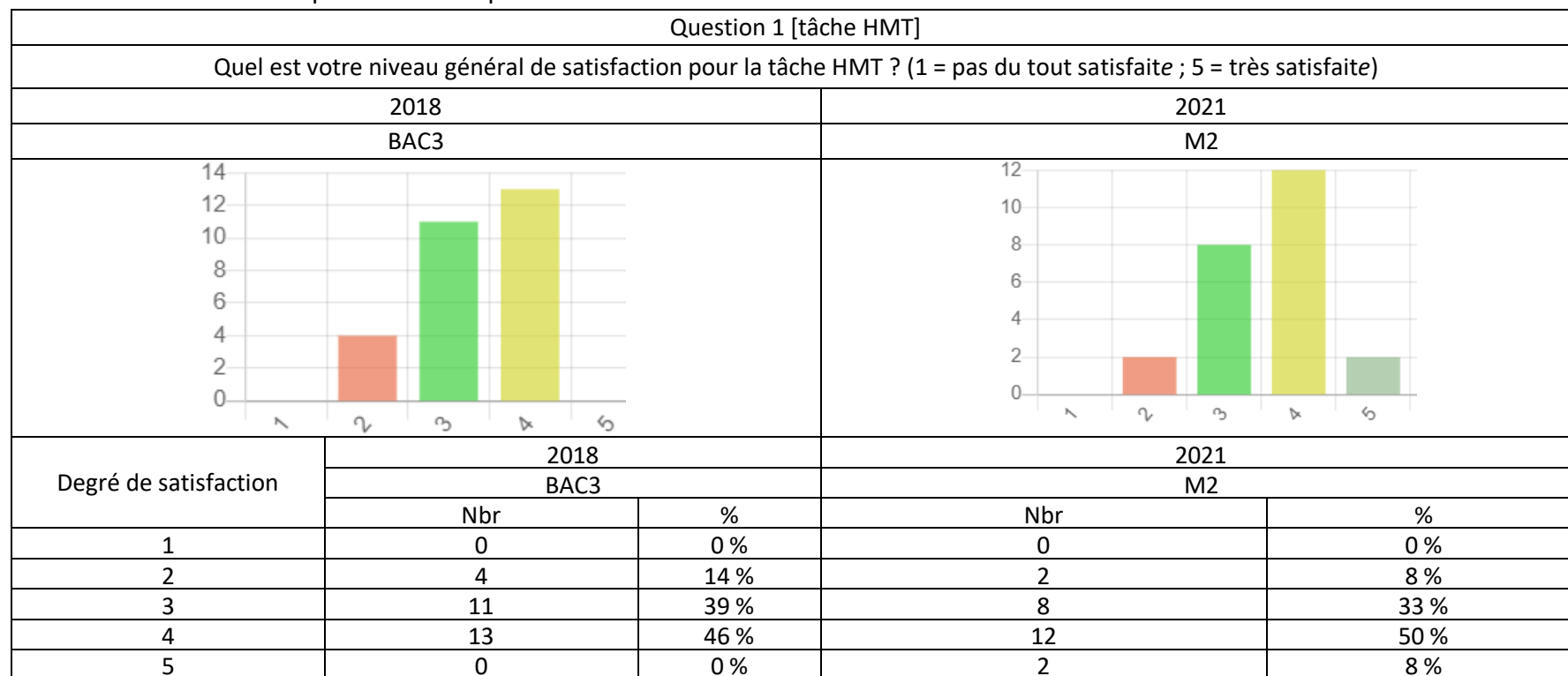


Figure 79 : Sondage post-expérience (Question 1)

Concernant la TH, en 2018, près de la moitié des étudiantes (46 %) se disent « plus que satisfaites », 11 étudiantes (39 %) se disent « satisfaites », contre quatre étudiantes (14 %) qui se déclarent « peu satisfaites » par cette tâche. Aucune étudiante ne s'avoue « insatisfaite », ni « très satisfaite ». On constate une légère augmentation de la satisfaction procurée par la tâche de TH en 2021 puisque 50 % des M2 se sont déclarées « plus que satisfaites » et deux étudiantes (8 %) se disent même « très satisfaites ».

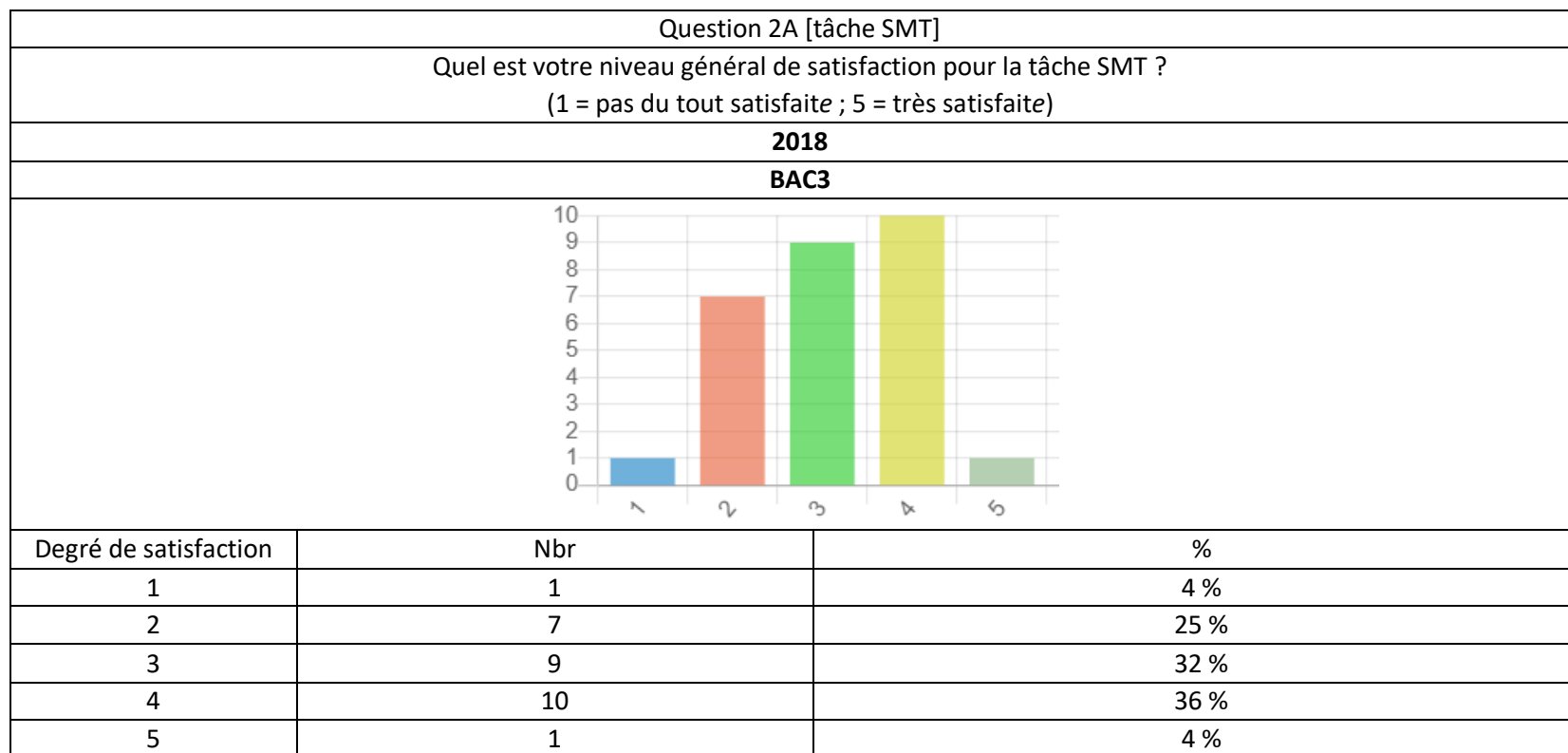
❖ **Question 2A – Quel est votre niveau général de satisfaction pour la tâche de PE de TA statistique (SMT) ?**

Figure 80 : Sondage post-expérience (Question 2A)

En 2018, pour la PE de TA statistique, environ un tiers des étudiantes (32 %) se déclarent « satisfaites » et un peu plus d'un tiers se disent « plus que satisfaites » par cette tâche. Une étudiante se dit même « très satisfaite ». En revanche, il apparaît que la PE de TA statistique a très peu satisfait sept étudiantes (25 %) et n'a pas satisfait du tout une étudiante.

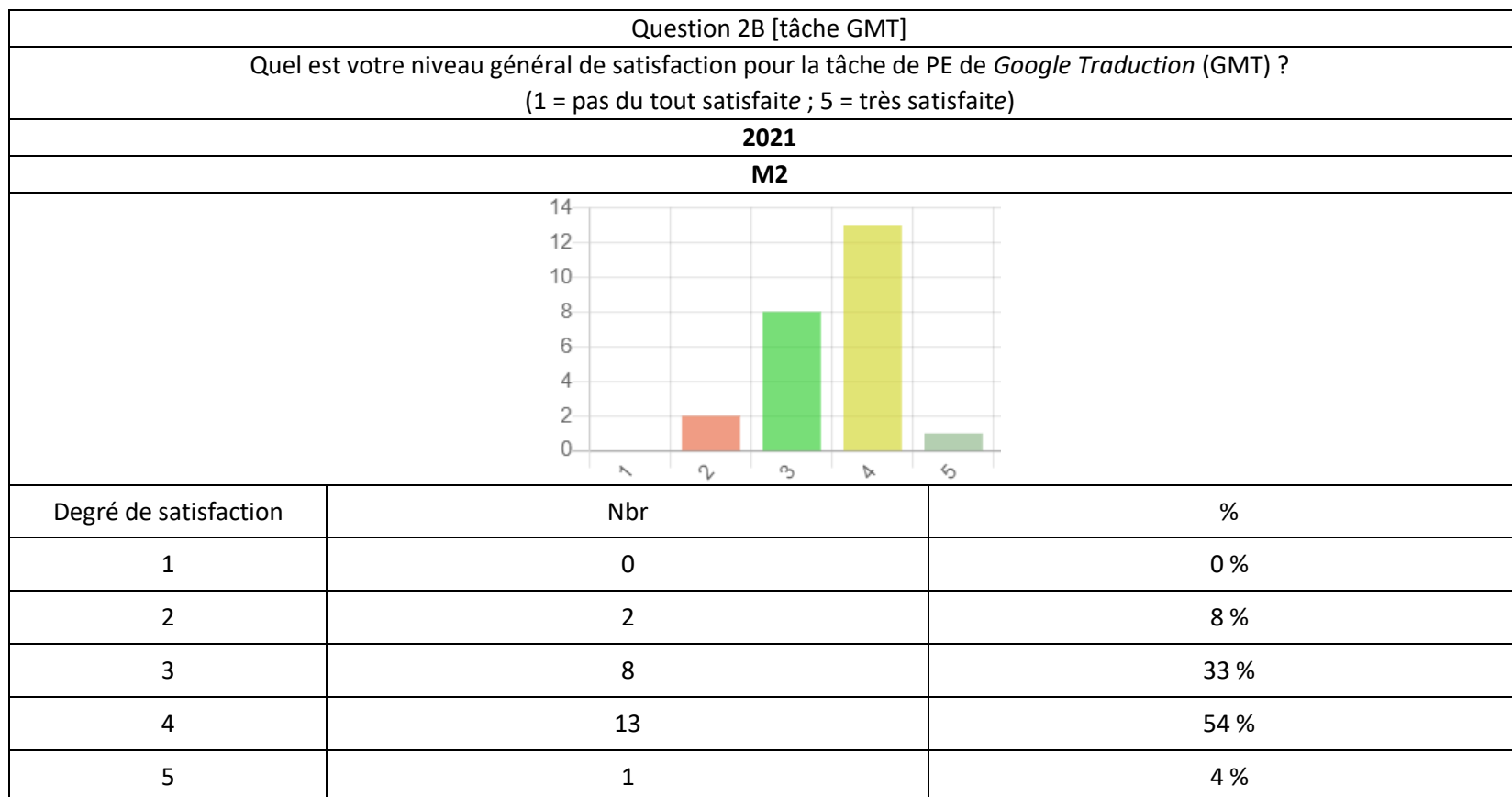
❖ Question 2B – Quel est votre niveau général de satisfaction pour la tâche de PE de *Google Traduction* (GMT) ?

Figure 81 : Sondage post-expérience (Question 2B)

En 2021, pour la PE de TA générée par *Google Traduction* (GMT), la majorité des sondées semble satisfaite, voire davantage : un tiers des étudiantes se sont dit « satisfaites » ; plus de la moitié (54 %) sont « plus que satisfaites » et une étudiante de M2 a déclaré être « très satisfaite » par cette tâche. Il nous est impossible de comparer ces résultats à ceux obtenus en 2018 étant donné qu'il ne s'agit pas du même moteur de TA (*BingTranslator* et *Google Traduction*).

❖ Question 3 – Quel est votre niveau général de satisfaction pour la tâche de PE de TA neuronale (NMT) ?

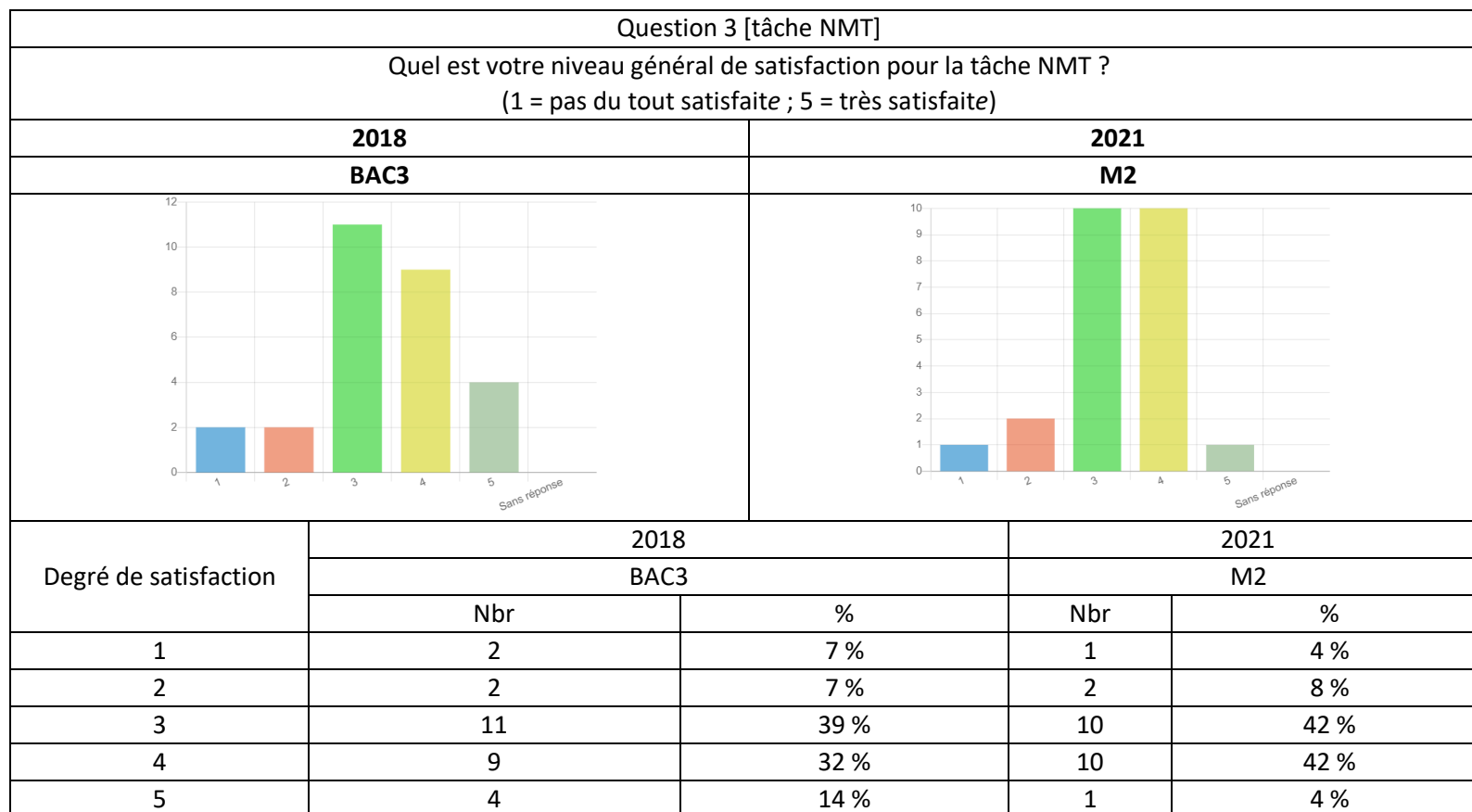


Figure 82 : Sondage post-expérience (Question 3)

En 2018, pour la PE de TA générée par *DeepL*, plus d'un tiers des étudiantes (39 %) se déclarent satisfaites et un peu moins d'un tiers (32 %) se disent « plus que satisfaites » par cette tâche. De plus, quatre étudiantes se disent même « très satisfaites ». Seules deux étudiantes sont « peu satisfaites » et deux étudiantes déclarent ne pas avoir été du tout satisfaites par cette tâche. Les taux de satisfaction obtenus en 2021 pour cette tâche sont très semblables à ceux de 2018. Il n'y a toutefois plus qu'une seule étudiante de M2 qui se dit « très insatisfaite », et une seule étudiante qui se dit « très satisfaite ».

7.5.2.1 Conclusion (questions 1 – 2A – 2B – 3)

Nous pouvons tirer comme conclusion que la tâche de PE de TA neuronale (*DeepL* et *Google*) a donné, dans l'ensemble, davantage satisfaction aux étudiantes que la PE de TA statistique (*BingTranslator*). En effet, nous avons constaté des taux de satisfaction plus élevés : Pour la PE de *DeepL* en 2018, 85 % des interrogées ont déclaré être satisfaites, voire plus, et ce taux s'élève à 88 % en 2021. Tandis que pour la PE de *Google* en 2021, ce sont 91 % des participantes qui se disent satisfaites, voire plus, contre 72 % en PE de TA statistique.

Selon nous, ces observations résultent en partie des nombreuses améliorations apportées par la nouvelle génération de TA par rapport à l'ancienne approche. Avec 85 % en 2018 et 91 % en 2021, il apparaît que le taux global de satisfaction en TH est comparable au taux de satisfaction en PE de TA neuronale. Il existe toutefois des divergences d'opinion, car la PE de *DeepL* a récolté des avis plus tranchés qu'en TH : alors qu'aucune participante n'a déclaré être insatisfaite en TH, plusieurs le sont en PE de *DeepL* (7 % en 2018 et 4 % en 2021).

Note

Les tableaux dans lesquels figurent les réponses aux questions ouvertes 4 – 5 – 6 – 7 – 8 – 9 – 17, ainsi qu'aux sous-questions 10.1 – 13.1 – 16.1 contiennent à la fois les commentaires des étudiantes de BAC3 en 2018 et ceux des étudiantes de M2 en 2021 sans distinction. Pour les questions 4 à 9, nous avons retenu uniquement les réponses des étudiantes qui nous paraissent renvoyer au mode de traduction plutôt qu'à des caractéristiques propres au texte source (complexité textuelle, etc.). Enfin, précisons que comme les étudiantes n'ont pas eu à traduire le même TS, la teneur de ces commentaires peut varier.

❖ **Question 4 – Qu'avez-vous le moins aimé dans la tâche de traduction humaine ?**

Nous avons d'abord demandé aux participantes de nous dire ce qu'elles avaient le moins apprécié dans la tâche de TH.

Question 4 [tâche HMT]	
Qu'avez-vous le moins aimé dans la tâche HMT ?	
ID	Réponse
8	La répartition en segments. Elle a du sens pour l'expérience, mais, pour mon confort personnel, j'aurais préféré pouvoir avoir accès au texte en entier (pas sous forme de segment), afin de voir s'il s'agissait d'un texte journalistique ou d'un billet. J'avais parfois l'impression que certaines phrases sortaient de nulle part, mais je suppose que c'est dû à la mise en page originale (sous-titres éventuels, commentaires, etc.)
9	Ne pas avoir le temps de laisser reposer la traduction et d'y revenir plus tard.
10	Beaucoup moins rapide que les autres tâches.
11	Cette tâche est celle qui m'a pris le plus de temps.
12	beaucoup de recherches à faire pour les termes économiques
13	Le fait de ne plus avoir de texte de base après avoir réalisé les deux autres était perturbant
15	Traduire dans un temps donné
16	le manque de contexte, j'aurais bien aimé avoir lu l'article dans son intégralité pour mieux comprendre l'enchaînement
17	L'interdiction d'utiliser la traduction automatique, qui aurait pu être utile pour traduire certains mots dans leur contexte.
19	Je passais plus de temps à traduire certaines phrases que pour les autres textes.
21	Les segments étaient fort courts.
24	Les recherches à effectuer pour bien comprendre le texte, et donc bien traduire.
25	La segmentation du texte
26	La recherche du vocabulaire
27	C'est le plus long et plus stressant, car il faut tout produire

Tableau 101 : Sondage post-expérience (Question 4)

Pour la TH (2018 et 2021), les éléments qui n'ont pas été appréciés par certaines étudiantes sont la thématique, la complexité du TS, la terminologie, la longueur des phrases ou encore les abréviations. En outre, plusieurs commentaires, repris dans le tableau ci-dessous, portent sur le mode de traduction. Certains rendent compte du caractère chronophage de la TH (10, 11, 19 et 27) par rapport aux tâches de PE, et une étudiante (27) a également trouvé cette tâche plus stressante. D'autres ont évoqué la lourdeur de la recherche documentaire (12, 17, 24 et 26). Certains commentaires portent davantage sur les conditions d'expérience que sur le mode de traduction à proprement parler. C'est le cas des commentaires 9, 13 et 15. Il est intéressant de noter que plusieurs étudiantes (8, 21 et 25) ont été dérangées par l'architecture segmentée et par le manque de contexte (16). Les étudiantes ont, en effet, peu l'habitude de traduire humainement avec un logiciel, en l'occurrence *MateCat*, qui présente le TS et la traduction en deux colonnes segmentées.

On peut légitimement se demander si cette segmentation a eu des effets cognitifs sur le processus de traduction.

❖ **Question 5A – Qu'avez-vous le moins aimé dans la tâche de PE de TA statistique (SMT) ?**

Question 5A [tâche SMT]	
Qu'avez-vous le moins aimé dans la tâche SMT ?	
ID	Réponse
1	ne pas avoir pu choisir mes mots
4	Réviser
10	Le niveau assez pauvre des mots
14	Il n'est pas évidant [sic] de corriger la phrase sans la refaire entièrement
17	Je ne voyais pas toujours directement les tournures à modifier, car la traduction était déjà assez fluide
19	Certaines tournures de phrases n'étaient pas bien traduites [sic] de base.
20	Vocabulaire bateau
23	Mauvaise traduction automatique à beaucoup d'endroits
24	Ne pas savoir exactement s'il était nécessaire de changer les mots/phrases ou pas. Il y avait des phrases qui ne me plaisaient pas, mais qui n'étaient cependant pas fausses.
26	Les fautes des termes traduits automatiquement
28	Le fait de ne pas réussir à se détacher de la traduction automatique par moment

Tableau 102 : Sondage post-expérience (Question 5A)

En 2018, pour la PE de TA statistique, outre des difficultés propres aux TS (terminologie, structure, thématique, longueur de phrases, etc.), plusieurs participantes ont formulé des commentaires qui se rapportent au mode de traduction (voir Tableau 102). A été mis en avant la difficulté de réviser (4), de repérer ou de corriger des unités de traduction mal traduites par la TA (19, 23, 26), surtout lorsque celles-ci sont déjà fluides (17). Une étudiante (24) évoque la difficulté de déterminer s'il convient de modifier certaines unités de traduction ou de les laisser telles quelles. Une autre étudiante (28) note la difficulté de parvenir à se détacher de la TA qu'elle a sous les yeux. Ne pas pouvoir choisir ses mots (1) et ne pas pouvoir retraduire entièrement la phrase (14) ont également été perçus comme des difficultés. Enfin, deux réponses évoquent la pauvreté du lexique (10) ou encore son caractère « bateau » (20).

❖ **Question 5B – Qu'avez-vous le moins aimé dans la tâche de PE de *Google Traduction* (GMT) ?**

Question 5B [tâche GMT]	
Qu'avez-vous le moins aimé dans la tâche GMT ?	
ID	Réponse
2	Pas assez de libertés [sic]
7	Texte traduit trop littéralement
8	Beaucoup de phrases étaient très peu françaises. J'ai l'impression que ça m'a pris plus de temps de les corriger que si j'avais traduit directement l'anglais.
9	Je n'étais pas satisfaite de certaines de mes tournures de phrases [sic].
11	La présence de segments déjà traduits influence un peu notre compréhension du texte source.
15	Traduire dans un temps donné
23	Fautes répétitives qui font douter

Tableau 103 : Sondage post-expérience (Question 5B)

En 2021, pour la PE de *Google*, plusieurs commentaires évoquent la qualité médiocre de la TA brute : une étudiante mentionne le manque de liberté (2), un autre la littéralité de la TA (7) ou encore le caractère fautif (23) et peu fluide des propositions de la machine (8), ce qui, toujours selon l'étudiante 8, a rendu cette tâche davantage chronophage par rapport à une TH. L'étudiante 11 souligne l'influence potentielle de la TA brute sur la compréhension du TS. A été mentionnée également, la non-satisfaction de la qualité de la PE finale (9). Enfin, on retrouve le même commentaire que pour la tâche de TH qui rend compte de la complexité de traduire dans un temps imparti (15).

❖ Question 6 – Qu'avez-vous le moins aimé dans la tâche de PE de *DeepL* (NMT) ?

Question 6 [tâche NMT]	
Qu'avez-vous le moins aimé dans la tâche NMT ?	
ID	Réponse
1	ne pas avoir pu choisir mes mots
2	Pas assez de libertés
4	réviser
6	Le manque de contexte m'a poussé à faire certains choix de traductions dont je ne peux garantir la pertinence totale
10B	Je ne sais pas dire pourquoi, mais c'est cette traduction que j'ai le moins apprécié [sic] au niveau de la post-édition. Je ne trouve pas que cette traduction était bien traduite au niveau de la syntaxe et l'ordre des mots.
11	La transformation de segments mal traduits s'avéraient [sic] parfois plus difficiles que si on avait traduit ce texte nous-mêmes.
12	nécessite de faire des recherches par notion inconnue
14	C'est parfois difficile d'accepter une phrase qui est grammaticalement correcte alors que la tournure ne plaît pas du tout.
15	Traduire dans un temps donné
17	Le temps des verbes, les tournures de phrases. La difficulté de me détacher de la traduction et en même temps de trouver un équilibre entre post-édition et traduction.
19	Certaines expressions étaient plus difficiles à traduire, la traduction automatique n'était pas satisfaisante pour ces phrases.
20	Puisqu'il s'agit d'une post édition [sic], il fallait aller chercher tous les termes spécifiques afin de vérifier leur exactitude
21	La traduction automatique n'était pas mauvaise et il était donc difficile de s'en éloigner pour vérifier que les phrases étaient correctement traduites.
22	Les tournures de phrases [sic] créaient parfois un sens erroné
23A	Quelques termes précis étaient mal traduits
23B	Parfois envie de réécrire toute la phrase
24	Ne pas savoir exactement s'il était nécessaire de changer les mots/phrases ou pas. Il y avait des phrases qui ne me plaisent pas, mais qui n'étaient cependant pas fausses.
27	Les prépositions traduites automatiquement sont sujettes à erreur et c'est pas évident à rattrapper [sic]
28	La traduction était déjà relativement bonne donc je ne savais pas si cela servait à quelque chose que j'y effectue de grands changements

Tableau 104 : Sondage post-expérience (Question 6)

Pour la tâche de PE de *DeepL* (2018 et 2021), les principales difficultés soulignées par les étudiantes sont à nouveau du même ordre que celles rencontrées en PE de TA statistique ou en PE de *Google Traduction* : la réduction de la liberté de traduction (1, 2 et 14), la recherche terminologique (12 et 20), le délai imparti (15), le manque de contexte (6), la nécessité de modifier ou de corriger les propositions erronées ou insatisfaisantes de la TA (11, 19, 22 et 23A), la difficulté de pallier la pauvreté linguistique (10A) et syntaxique (10B) des propositions émanant de la TA, la difficulté d'estimer le degré de modification à

apporter et ainsi d'éviter la sur-édition (14, 17, 23B, 24 et 28) ou encore de se détacher de la proposition machine (17, 21). Enfin, une étudiante a relevé la difficulté de déjouer l'illusion de fluidité en TA neuronale (21).

Il apparaît que plusieurs des raisons évoquées par nos participantes en PE font écho aux réponses récoltées par Moorkens et O'Brien (2015) auprès d'étudiantes novices :

Responses to the question "Did you like the task of post-editing? Why/why not?" were divided into positive, neutral and negative. [...] Those with a neutral attitude suggested that postediting limited the use of "imagination" or that it was "uncreative". Reasons given for negative responses can be grouped into four main categories to do with time, quality, tool functionality, and lack of context. Some participants complained about the raw MT quality saying it would be "easier to start from scratch". There was a perception among a few that the task took more time (than translation), was exhausting because it was repetitive, and made more difficult due to the lack of context for the segments. (p. 79)

❖ Question 7 – Qu'avez-vous le plus aimé dans la tâche de traduction humaine ?

Q7 [tâche HMT]	
Qu'avez-vous le plus aimé dans la tâche HMT ?	
Identifiant (ID)	Réponse
1	pouvoir traduire moi-même
2	Libertés laissées, car texte pas encore traduit de base [sic]
3	La traduction entièrement réalisée sans aide de <i>DeepL</i> etc
4A	traduire
4B	Le fait de ne pas avoir à post-éditer le texte m'a donné une plus grande marge de manœuvre. Je pense ainsi avoir eu une meilleure compréhension du texte source.
5	Le fait de devoir traduire tout de A à Z.
6	La liberté de traduction
7	La liberté de proposer ses propres choix
8	Les recherches à effectuer
10A	Le fait de pouvoir choisir ses propres mots.
10B	Gain de temps
11A	la traduction humaine est plus simple
11B	Pouvoir faire la traduction nous-même [sic] permet d'avoir plus de liberté d'interprétation.
12	liberté de traduire selon ses idées
13	La liberté de traduire à partir de rien
14A	Beaucoup de place pour le style personnel, c'est beaucoup plus naturel.
14B	Il faut faire des recherches pour s'informer sur le sujet, toujours intéressant pour la culture générale
16	défis
17	la production était plus facile et libre
20	J'étais libre de tourner les phrases à ma sauce, sans être influencée par la traduction produite par un logiciel.
22	Traduire moi-même, pouvoir être plus créative
23A	La liberté de traduction
23B	Plus facile de se détacher du texte
24A	Être libre de choisir sa traduction.
24B	Le thème, la liberté de traduire comme bon me semble (pas de post-édition).
25	La liberté de traduction
27	C'est gai de traduire
28	Le fait de pouvoir traduire librement

Tableau 105 : Sondage post-expérience (Question 7)

Pour la tâche de TH, ce qui a été apprécié par de nombreuses étudiantes (2, 6, 7, 10A, 12, 13, 14A, 17, 20, 22, 23A, 24A et B, 25 et 28) est la liberté de traduction que ce mode de traduction confère (Daems, 2016), le fait de pouvoir exprimer librement ses préférences

stylistiques (idiolecte), ainsi que la liberté d'interprétation (11B) avec parfois une meilleure compréhension du TS (4B). Plusieurs étudiantes (1, 3, 4A, 5 et 27) ont indiqué tout simplement avoir aimé traduire plutôt que post-éditer. Par ailleurs, deux étudiantes (11A et 17) ont perçu la TH comme plus facile que les tâches de PE, et une étudiante estime même que la TH était plus rapide (10B) (Daems, 2016). Il nous est difficile de commenter la réponse 16. Enfin, les étudiantes ont évoqué la nécessité de la recherche documentaire (8 et 14B) et la plus grande facilité à se détacher du texte pour proposer une traduction (23B). Certains de ces commentaires rejoignent les avis recueillis dans les travaux de Daems (2016) : « Those who preferred regular translation mentioned creativity and freedom as an important factor, and they did not believe that post-editing would necessarily save time » (p. 107).

❖ **Question 8A – Qu'avez-vous le plus aimé dans la tâche de PE de TA statistique ?**

Q8A [tâche SMT]	
Qu'avez-vous le plus aimé dans la tâche SMT ?	
Identifiant (ID)	Réponse
4	traduire par moi-même si je voyais que c'était totalement incorrect
5	rapide à corriger
8	Le fait d'avoir déjà une traduction et de seulement la corriger est plus agréable
9	La première piste d'une traduction possible
10	L'assez bon niveau de la traduction
13	Les phrases traduites correctement par le traducteur automatique
14	C'est amusant de voir ce que peut produire un ordinateur
16	défis
17	la rapidité avec laquelle la révision se faisait [sic]
18	Retravailler les phrases
22	rapidité
24	Plus rapide, car une première traduction était proposée
26	Avoir une base de traduction

Tableau 106 : Sondage post-expérience (Question 8A)

Pour la PE de TA statistique, certaines étudiantes (9 et 26) ont apprécié avoir une base de traduction et ne pas devoir partir d'une « page blanche ». Deux étudiantes (10, 13) ont souligné le bon niveau général ou ponctuel de la TA. D'autres ont évoqué la rapidité de cette tâche (5, 17, 22 et 24). Deux étudiantes (8 et 18) ont particulièrement aimé le travail de reformulation, de révision. L'étudiante 4 montre, *elle*, une nette préférence pour la TH par rapport à la PE. Enfin, les réponses 14 et 16 nous paraissent peu explicites et difficilement interprétables.

❖ **Question 8B – Qu'avez-vous le plus aimé dans la tâche de PE de *Google Traduction* (GMT) ?**

Q8B [tâche GMT]	
Qu'avez-vous le plus aimé dans la tâche GMT ?	
Identifiant (ID)	Réponse
1	Sujet assez facile, la traduction automatique relativement bien faite également. Pas trop de choses à modifier.
2	Texte économique donc top que certains termes soient déjà traduits par l'ordinateur
3	L'aide précieuse du traducteur automatique pour certaines phrases compliquées
4	Le thème du texte était original. La qualité de la traduction automatique était d'ailleurs assez bonne pour un thème si spécifique.
7	La relative rapidité de la tâche
11	Le sujet du texte fait que le travail de post-édition n'était pas trop exigeant.
15	Le contenu/sujet. Travail de post-édition intéressant
20	La traduction automatique était assez bonne, je n'ai donc fait quasi que des modifications de style et de fluidité
22	Garder certaines expressions correctes, mais que je n'appréciais pas
24	Le thème, le texte était déjà assez bien traduit, peu de changements à effectuer lors de la post-édition.

Tableau 107 : Sondage post-expérience (Question 8B)

En 2021, pour la PE de la TA générée par *Google Traduction*, les participantes de M2 ont fait ressortir la qualité de la TA brute et la relative aisance à post-éditer (1, 4, 11, 20 et 24), l'aide précieuse que représentent les propositions émanant de la TA (2, 3), mais aussi le gain de temps perçu en PE (7). Enfin, l'étudiante 22 n'a pas répondu pertinemment à la question posée. Une fois encore, les réponses données par nos étudiantes rejoignent les conclusions du sondage mené par Moorkens et O'Brien (2015) : « Of those who gave positive responses, the reasons they used were that the translation was already done for them and they just needed to "improve a few things" » (p. 79).

❖ Question 9 – Qu'avez-vous le plus aimé dans la tâche de PE de *DeepL* ?

Q9 [tâche NMT]	
Qu'avez-vous le plus aimé dans la tâche NMT ?	
ID	Réponse
1	Sujet intéressant. Pas trop de grosses modifications de sens à apporter. La traduction automatique aidait beaucoup
2	Grosses erreurs commises par le système, agréable de les réparer (eg: May --> mai)
3	Erreurs facilement identifiables du traducteur automatique
4	traduire par moi-même si je voyais que c'était totalement incorrect
5A	déjà bien traduit
5B	Le sujet et le fait que la traduction automatique n'était pas trop chaotique
6	La traduction automatique était plutôt bonne
8	Même si j'ai effectué beaucoup de modifications, je trouve que la plupart des phrases étaient déjà bien traduites, ce qui permet de gagner du temps
9A	La première piste d'une traduction possible
9B	Il me semble que le texte cible était déjà très bien traduit
10	Tout de même plus rapide
11	La présence des segments traduits aident [sic] parfois à comprendre certains passages plus compliqués en anglais
13	Les phrases traduites correctement par le traducteur automatique
14	Pas grand-chose à corriger et le thème était assez intéressant
16	voir que la post-édition fonctionne pas trop mal sur un texte économique
18A	Retravailler les phrases
18B	j'ai trouvé cela amusant de pouvoir voir directement où la machine a des difficultés (nom propre de Theresa May)
19	Le texte n'était pas trop mal traduit à la base [sic]
21	Découverte de nouvelles expressions sur le vocabulaire économique
22	rapidité
23	La traduction était presque parfaite
24A	Plus rapide, car une première traduction était proposée
24B	La terminologie, assez facile à trouver, les recherches assez brèves
26	Avoir une base de traduction
27	la traduction automatique de "May"...

Tableau 108 : Sondage post-expérience (Question 9)

En PE de *DeepL*, plusieurs étudiantes ont, une fois de plus, apprécié la qualité relativement bonne de la sortie en TA neuronale ; en comparaison avec la PE de TA statistique, davantage d'étudiantes (1, 5A et B, 6, 8, 9B, 13, 14, 16, 19 et 23) ont estimé que la qualité de la sortie brute de la TA était satisfaisante, bonne, voire « presque parfaite » (23). Certaines étudiantes (9A, 24A et 26) ont à nouveau apprécié avoir une base de traduction. Une étudiante (18A) a particulièrement apprécié le travail de reformulation. Ont aussi été mis en avant, l'aide que la TA représente en matière de compréhension du TS (11), ainsi que le peu de besoins en recherche terminologique (24B) et la découverte de nouvelles

expressions (21). Plusieurs étudiantes ont, une fois de plus, fait part d'une impression de gain de temps (8, 10, 22 et 24A). En outre, certaines ont trouvé qu'il était aisé de repérer et de modifier les erreurs grossières de la machine (2, 3), tandis que l'étudiante 4 confirme sa préférence pour la TH. Enfin, les étudiantes 18B et 27 se sont délecté des propositions cocasses de la machine. Une fois encore, plusieurs des arguments avancés en PE font écho aux impressions recueillies par Daems (2016) : « [T]hose who enjoyed it [post-editing] mostly enjoyed not having to start from scratch, and noted that it could save them some time, provided the output was of sufficient quality » (p. 107).

❖ **Question 10 – Quelle(s) tâche(s) étai(en)t la(les) plus exigeante(s)/complexe(s) ?**

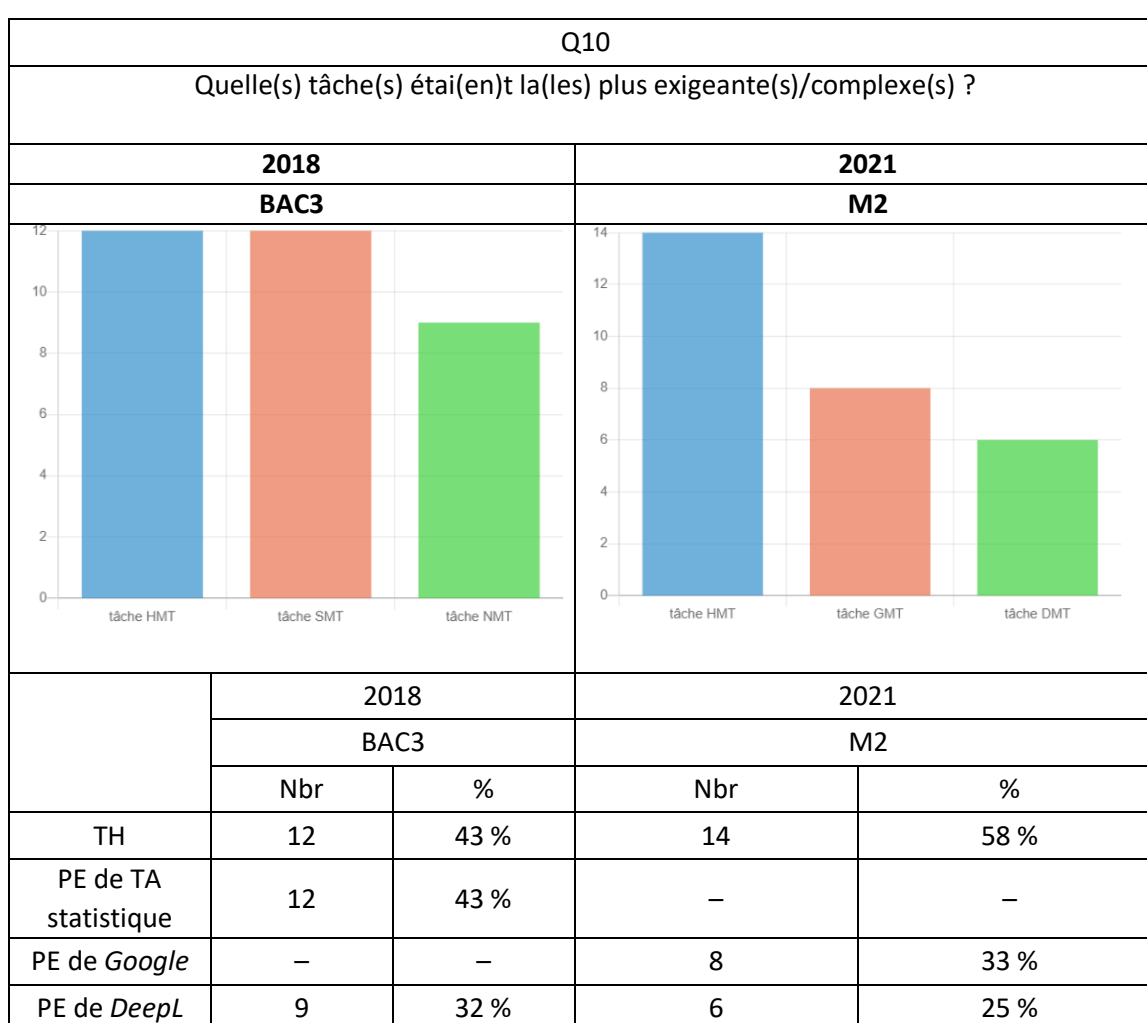


Figure 83 : Sondage post-expérience (Question 10)

Pour cette question, l'étudiante pouvait sélectionner plusieurs tâches. Selon les taux de 2018 figurant dans le Figure 83, les étudiantes ont considéré la tâche de TH comme tout aussi exigeante que la PE de TA statistique. En 2021, c'est la TH (58 %) qui a été jugée plus complexe que les deux tâches de PE. De plus, moins d'un tiers des participantes en 2018 (32 %) et seulement un quart (25 %) des M2 en 2021 ont trouvé qu'il s'agissait de la PE de

DeepL. Remarquons qu'en 2021, la PE de *DeepL* a été perçue comme moins exigeante que la PE de *Google Traduction*. Sur les deux expériences, nous pouvons dire que la TH ressort comme une tâche plus exigeante que la PE aux yeux des étudiantes (43 % en 2018 et 58 % en 2021). Nous pouvons également en déduire que sur les deux expériences, la PE de *DeepL* a été la tâche jugée comme la moins exigeante par les étudiantes, quel que soit le TS.

❖ Question 10.1. – Pourquoi ?

Nous avons à nouveau opéré une sélection représentative des justifications avancées par les étudiantes. Pour bon nombre d'étudiantes (BAC3 et M2 confondus), la TH s'est révélée une tâche plus exigeante que la PE pour diverses raisons : la TH leur a paru être une tâche davantage chronophage (12, 20, 22B) qui nécessite d'effectuer davantage de recherches (1B, 8, 9, 17B, 20, 21B). Certaines ont simplement trouvé qu'il était plus simple de « corriger/réviser » que de traduire par soi-même (3A, 4). Enfin, en PE, certaines étudiantes se sentent aidées au niveau de la compréhension du TS par les propositions de la machine, ce qui n'est pas le cas en TH (8, 11B, 21A, 21B, 22A).

1B	1. C'était le texte le plus difficile des trois (vocabulaire, thème, tournures...) 2. Il a dû être traduit 100 % "humainement", ce qui a demandé beaucoup de temps, de travail et de recherches.
3A	C'est plus difficile de traduire que de réviser.
4	Il est toujours plus complexe de traduire par soi-même, car on ne démarre pas d'une structure de base qu'on doit corriger. On doit tout faire par soi-même.
8	La tâche HMT est, selon moi, la plus exigeante, car les traducteurs doivent faire le travail de A à Z : comprendre, rechercher le vocabulaire et les concepts en anglais ET en français, puis enfin reformuler le sens dans un français correct et idiomatique. Mais c'est aussi pour cela que c'est la plus intéressante !
9	On part de rien et donc il faut faire toutes les recherches de vocabulaire et de contexte.
11B	Cette tâche était de la traduction : on ne pouvait donc se baser que sur notre compréhension personnelle du texte source, sans avoir de segments déjà traduits pouvant nous aiguiller sur le sens d'un passage donné par exemple.
12	La HMT m'a pris beaucoup plus de temps
17B	La première, car il s'agissait d'une traduction entièrement humaine, avec toutes les recherches et les difficultés qui en découlent.
20	Je trouve qu'il était plus complexe de traduire le texte sans aide de la traduction automatique, car cela prend beaucoup de temps d'aller chercher chaque information sur internet étant donné que je ne connais pas grand-chose à l'économie.
21A	Il fallait réfléchir à toutes les tournures de phrase par soi-même et trouver le mot exact alors que lors de la post-édition, on avait déjà des exemples et une idée initiale et si on doutait, on pouvait juste vérifier.
21B	Je ne m'y connais pas très bien en économie, il fallait donc faire plus de recherches que dans le premier texte. En outre, le texte n'était pas pré-traduit, ce qui peut parfois apporter une aide pour déchiffrer le texte.
22A	Il faut trouver les tournures adéquates par soi-même et il n'y a pas d'aide à la compréhension du texte en anglais, contrairement à la post-édition.
22B	Elle prend plus de temps, car il faut rechercher plus de termes à mon sens, et parfois réfléchir plus longtemps pour produire de belles phrases qui auraient pu être données par la traduction automatique en quelques secondes.

Tableau 109 : Sondage post-expérience (Question 10.1-A)

En 2018, 43 % des étudiantes de BAC3 ont trouvé qu'il s'agissait de la PE de TA statistique. Elles ont évoqué la difficulté de repérer les erreurs de la TA (17A et 25), de juger s'il convient de modifier le segment brut (11A et 25), la complexité de la tâche par rapport à une TH (15A) ou encore, le fait que c'était la première fois qu'on leur demandait de post-éditer (1 et 28).

1A	C'est un exercice qu'il ne nous avait jamais été demandé de réaliser.
11A	C'est difficile de juger si quelque chose est correct ou non quand on fait une révision d'une traduction automatique.
15A	C'était parfois très compliqué de réviser ce qui avait déjà été fait par un ordinateur que de produire sa propre traduction.
17A	Il était plus difficile de voir les fautes de tournures de phrase, il fallait se concentrer davantage sur les calques.
25	On se base sur une traduction déjà "faite". On doit se contenir [sic] de changer ce qui est correct, mais que nous nous n'exprimerions pas comme ça. Les fautes de traductions commises par quelqu'un d'autre que soi sont parfois plus difficiles à repérer puisqu'on remet tout en doute.
28	Vu que c'était la toute première fois que j'effectuais une post-édition, j'ai pris beaucoup de temps avant d'oser émettre d'autres suggestions et je décortiquais quasiment chaque phrase (même celles qui me paraissaient les plus évidentes).

Tableau 110 : Sondage post-expérience (Question 10.1-B)

En 2021, 33 % des participantes de M2 ont considéré qu'il s'agissait de la tâche de PE de *Google*. Les justifications avancées par les étudiantes se rapportent à la nécessité de devoir vérifier ce que la TA produit (2B), à la complexité de la tâche de PE par rapport à une TH (3B), ainsi qu'à la difficulté de se détacher des propositions de la machine (8 et 17B).

2B	Il s'agissait d'un domaine plus pointu et il fallait donc vérifier davantage de termes précis, car je ne me fie pas du tout à la traduction automatique.
3B	Le sujet était complexe et devoir le post-éditer rendait parfois la tâche encore plus compliquée pour certains segments.
8	En revanche, les autres tâches sont aussi compliquées, pour différentes raisons. La tâche GMT, par exemple, prend du temps parce qu'il faut presque tout modifier et qu'il est compliqué de se détacher des mots français déjà écrits sous nos yeux. Toutefois, si l'on a le temps d'aller jusqu'au bout, je pense qu'on peut produire une traduction de bonne qualité.
17B	La deuxième, car j'ai eu énormément de mal à me détacher de la traduction automatique pour faire des phrases qui me semblaient correctes et plus adaptées. La troisième [PE de <i>DeepL</i>] l'était moins, car je trouvais que la traduction automatique était plutôt bonne au départ.

Tableau 111 : Sondage post-expérience (Question 10.1-C)

Enfin, seules quelques participantes (BAC3 et M2 confondus) ont perçu la PE de *DeepL* comme exigeante. Les arguments donnés sont semblables à ceux avancés pour la PE de TA statistique et pour la PE de *Google* : complexité de la tâche de PE (4, 10) et caractère chronophage de la tâche (10) par rapport à une TH, difficulté de se détacher des propositions de la machine (15B), difficulté de repérer les erreurs de la TA (25), de juger s'il convient de modifier la proposition de la TA (11A et 25). En outre, plusieurs étudiantes ont également fait référence à la complexité du TS (2A, 7 et 19).

2A	Les sujets sont complexes voire même [sic] un peu inconnus pour ma part.
4	La traduction automatique a produit des structures de phrases assez bancales, ce qui a rendu la tâche peut-être plus complexe qu'une traduction traditionnelle.
7	Les phrases à traduire n'avaient ni queue ni tête.
10	Je n'ai pas réussi à produire une traduction de qualité, je pense que j'ai perdu du temps plus qu'autre chose à corriger les erreurs de la machine au lieu de rédiger moi-même.
11 A	C'est difficile de juger si quelque chose est correct ou non quand on fait une révision d'une traduction automatique.
15B	J'ai eu du mal à m'éloigner de la TA car je ne maîtrisais pas vraiment la terminologie économique demandée.
19	Il y avait dans le texte des tournures de phrases et des expressions difficiles à traduire.
25	On se base sur une traduction déjà "faite". On doit se contenir de changer ce qui est correct, mais que nous nous n'exprimerions pas comme ça. Les fautes de traductions commises par quelqu'un d'autre que soi sont parfois plus difficiles à repérer puisqu'on remet tout en doute.

Tableau 112 : Sondage post-expérience (Question 10.1-D)

❖ **Question 11 – Parmi les trois productions finales, laquelle/lesquelles estimez-vous être de meilleure qualité ?**

Q11				
Parmi les trois productions finales que vous venez de réaliser, laquelle/lesquelles estimez-vous être de meilleure qualité ?				
	2018		2021	
	BAC3		M2	
	Nbr	%	Nbr	%
TH	15	54 %	12	50 %
PE de TA statistique	3	11 %	–	–
PE de <i>Google</i>	–	–	7	29 %
PE de <i>DeepL</i>	10	36 %	5	21 %

Tableau 113 : Sondage post-expérience (Question 11)

Une fois de plus, les étudiantes tiennent une relative haute estime de leur production humaine, car plus d'une participante sur deux a estimé que le mode de traduction qui lui a permis d'aboutir à la meilleure qualité est la TH (54 % en 2018 et 50 % en 2021). Parmi les quatre modes de traduction, celui qui recueille le taux le plus faible (11 %) est la PE de TA statistique. Concernant la PE de *DeepL*, ce taux s'élève à 36 % des étudiantes en 2018 et à 21 % en 2021. Les taux obtenus nous amènent à nous interroger : Pourquoi les étudiantes de M2 plus expérimentées se sont-elles montrées, dans une certaine mesure, plus enclines à se satisfaire de la qualité en PE de TA par rapport à des étudiantes de BAC3 ? Nous présumons qu'une des raisons pourrait être l'influence de l'intégration d'exercices de PE

dans la formation en traduction à l'Université de Liège entre 2018 et 2021. En effet, contrairement aux étudiantes de BAC3, les étudiantes de M2 ont été sensibilisées aux avantages et aux limites de la (PE de) TA neuronale. Or, comme cela a déjà été constaté, notamment par Stasimioti et Sosoni (2019), il est courant que les étudiantes se montrent davantage favorables après avoir suivi une formation spécifique en TA et en PE.

- ❖ **Question 12 – Les tâches de post-édition vous ont-elles permis de vous concentrer davantage sur la fluidité du texte cible (style, phraséologie, tournures idiomatiques, etc.) par rapport à une traduction humaine ?**

Q12				
Les tâches de post-édition vous ont-elles permis de vous concentrer davantage sur la fluidité du texte cible (style, phraséologie, tournures idiomatiques, etc.) par rapport à une traduction humaine ?				
	2018		2021	
	BAC3		M2	
	Nbr	%	Nbr	%
Oui	14	50 %	11	46 %
Non	14	50 %	13	54 %

Tableau 114 : Sondage post-expérience (Question 12)

En 2018, on observe une parfaite répartition pour cette question : 50 % des étudiantes jugent que par rapport à la TH, les tâches des PE leur ont permis de se concentrer davantage sur la fluidité du texte cible et 50 % estiment que ce n'est pas le cas. En 2021, seuls 46 % des étudiantes de M2 ont, cette fois, répondu positivement à cette question.

- ❖ **Question 13 – Les tâches de post-édition vous ont-elles permis d'utiliser des termes, tournures ou expressions auxquels vous n'auriez pas pensé lors d'une traduction humaine ?**

Q13				
Les tâches de post-édition vous ont-elles permis d'utiliser des termes, tournures ou expressions auxquels vous n'auriez pas pensé lors d'une traduction humaine ?				
	2018		2021	
	BAC3		M2	
	Nbr	%	Nbr	%
Oui	10	36 %	11	46 %
Non	18	64 %	13	54 %

Tableau 115 : Sondage post-expérience (Question 13)

MT output can be very good and serve as inspiration for students, meaning MT should be considered a translation tool to help them in the same way translation memories and specialised electronic corpora do. (Loock, 2020c, p. 164)

Une fois encore, les avis sont partagés, même si la majorité des participantes (2018 et 2021) ne semble pas avoir tiré profit de la TA brute en ce qui concerne l'emploi de termes, tournures et expressions auxquels *elles* n'auraient pas pensé spontanément en TH.

- ❖ Question 13.1 – Si vous vous en souvenez, pourriez-vous citer un ou plusieurs exemples de ces termes, tournures ou expressions.

Q13.1	
Si vous vous en souvenez, pourriez-vous citer un ou plusieurs exemples de ces termes, tournures ou expressions	
ID	Réponse
1	je ne connaissais pas « le nœud gordien ». En temps normal, j'aurais essayé de trouver une paraphrase, mais en réalité ce terme existe aussi en français.
3	Je trouvais la phrase : « un récit de vulnérabilité extrême s'est installé », très littérale, mais finalement bien tournée et elle rendait bien le sens également. On pense à "s'installer" comme si ce récit était partout, dans nos conversations, dans les journaux, à la télé, malgré madame May.
7	« couper le nœud gordien »
8	Je ne me souviens plus exactement. Mais comme le texte est déjà produit, il est évident de retrouver des tournure [sic] différentes de celles que nous aurions produites.
9	« Occuper le devant de la scène ».
10	« sens antihoraire »
18	Je pense que j'ai moins fait de recherches pour le texte économique, car les termes étaient déjà « traduits pour moi ». J'ai fait confiance à la machine même si je pense que ce n'était pas toujours le bon choix.
19	Il n'y a rien qui me viennent [sic] comme ça, mais certaines tournure [sic] semblaient très naturelles dans la traduction automatique.
20	« au pied de la lettre »
21	« des hallucinations induites par la drogue » => j'aurais sûrement paraphrasé
22	« Au pied de la lettre » à la place de littéralement.
28	« trancher le nœud gordien »

Tableau 116 : Sondage post-expérience (Question 13.1)

La lecture de ces commentaires (notamment 18) permet d'entrevoir les risques d'un excès de confiance chez les étudiantes à l'égard des propositions de la machine. Si la TA constitue par moments une aide, il arrive également que les étudiantes aient une impression trompeuse et se reposent à tort sur la technologie. D'ailleurs, précisons que deux des tournures citées dans le Tableau 116 ont été pénalisées comme calques fautifs par les évaluatrices humaines ; il s'agit des tournures « un récit de vulnérabilité extrême s'est installé » (3) – *Since that time, a narrative of extreme vulnerability has taken hold* et « des hallucinations induites par la drogue » (21) – *What makes synesthesia different from drug-induced hallucinations*.

❖ Question 13.2 – Quelle est la tâche qui vous a le plus aidée en ce sens ?

Q13.2				
Quelle est la tâche qui vous a le plus aidée en ce sens ?				
	2018		2021	
	BAC3		M2	
	Nbr	%	Nbr	%
PE de TA statistique	2	7 %	–	–
PE de Google	–	–	3	13 %
PE de DeepL	6	21 %	7	29 %
Sans réponse	0	0 %	1	4 %

Tableau 117 : Sondage post-expérience (Question 13.2)

Nous avons demandé aux étudiantes concernées de nous dire quelle était la tâche de PE qui les avait le plus aidées à faire usage de termes, tournures ou expressions auxquels *elles* n'auraient pas pensé en TH. Au vu des résultats repris dans ce tableau, il apparaît que les étudiantes considèrent que la TA générée par *DeepL* les a davantage aidés par rapport à la TA générée par la TA statistique (7 %) et par *Google Traduction* (13 %). Peut-être serait-ce dû au fait que le moteur *DeepL* met particulièrement l'accent sur la fluidité en langue cible (Loock, 2018) ?

❖ Question 14 – Après cette expérience, seriez-vous prête à post-éditer si on vous le demandait dans le cadre de votre futur emploi ?

Q14				
Après cette expérience, seriez-vous prête à post-éditer si on vous le demandait dans le cadre de votre futur emploi ?				
	2018		2021	
	BAC3		M2	
	Nbr	%	Nbr	%
Oui	26	93 %	22	92 %
Non	2	7 %	2	8 %

Tableau 118 : Sondage post-expérience (Question 14)

Nous leur avons posé la question que nous leur avons posée avant de prendre part à l'expérience (voir Tableau 87). Aucun changement pour les participantes de M2, nous obtenons toujours 92 % de réponses positives. En revanche, il apparaît que 93 %, soit 26 étudiantes de BAC3 sur 28, se montrent prêtes à post-éditer si on le leur demandait dans le cadre de leur futur emploi (contre 25 avant l'expérience), ce qui signifie qu'une étudiante a manifestement changé d'opinion à la suite de cette expérience.

❖ Question 14.1. – Expliquez brièvement pourquoi.

Nous avons voulu connaître les raisons pour lesquelles quatre étudiantes ont répondu par la négative à la question 14. Leurs réponses figurent dans le Tableau 119. Trois de ces étudiantes expriment clairement leur préférence pour la TH : l'étudiante 1 estime que la TH reste un mode de traduction « plus efficace » que la PE sans préciser pourquoi, l'étudiante 2 réaffirme sa préférence pour la TH en arguant de la qualité du produit fini, tandis que l'étudiante 14 affirme ne pas apprécier d'être influencée par la machine dans ses choix stylistiques tout en reconnaissant l'utilité que peut avoir cet outil. Enfin, l'étudiante 12 trouve que la PE est une activité très chronophage.

Q14.1	
Expliquez brièvement pourquoi vous n'êtes pas prête à post-éditer si on vous le demandait dans le cadre de votre futur emploi ?	
ID	Réponse
1	Je reste convaincu que la traduction humaine est plus efficace.
2	Mêmes raisons que celles expliquées dans le premier questionnaire. En outre, le logiciel reste généralement proche du texte et j'ai moins tendance à modifier les tournures dans le cas de la post-édition car souvent, ce qui est proposé n'est pas faux, c'est juste "moins bon" qu'une traduction humaine.
12	Cela prend beaucoup de temps à corriger.
14	Comme dans le premier questionnaire, je préfère chercher mes propres tournures au lieu d'être influencée par la traduction automatique, même si par moment, elle peut s'avérer utile.

Tableau 119 : Sondage post-expérience (Question 14.1)

❖ **Question 15 – Auriez-vous des améliorations à suggérer pour ce test concernant les tâches de post-édition ? (problèmes rencontrés, ergonomie de l'interface, sortie brute de la traduction automatique...)**

Q15	
Auriez-vous des améliorations à suggérer pour ce test concernant les tâches de post-édition ? (problèmes rencontrés, ergonomie de l'interface, sortie brute de la traduction automatique...)	
ID	Réponse
15	le découpage du texte n'aide pas à avoir une vue d'ensemble du texte et à voir si tout est cohérent (lien entre les phrases, les idées)
18	Lors de la traduction "conventionnelle", j'aurais préféré avoir une vue totale du texte.
20	Changer l'ordre des tâches, puisque la traduction humaine est plus longue et compliquée, la mettre au début pourrait améliorer le test.
2	A la fin de la post-édition sur Theresa May, lorsque je devais post-éditer, le curseur sélectionnait d'office les mots suivants et les supprimait au fur et à mesure que je modifiais le texte. Je devais donc écrire plus loin et faire des copier-coller à replacer. De plus, je n'étais pas sur mon clavier normal (Mac) donc je ne sais pas comment faire un accent sur le A majuscule ni comment lier le oe dans le cas de noeud par exemple.
7	Une vue d'ensemble du texte final permettrait probablement de mieux se rendre compte de la fluidité et de la cohérence globale.
14	Pas accès à tous mes outils. J'utilise beaucoup antidote pour les cooccurrences et les synonymes. Je n'ai pas pu utiliser ce logiciel, ce qui a probablement diminué la qualité de mes traductions.
16	Pouvoir modifier le texte aisément d'un point de vue typographique et avoir accès au texte source complet pour mieux comprendre le contexte, mais aussi la mise en page (pourquoi certains paragraphes ne sont pas suivis d'un point, etc.).
17	Par rapport à l'interface, lorsqu'on traduit un segment, et qu'on valide la traduction, le programme ouvre automatiquement le suivant. Une fois que le texte est traduit dans son entièreté, il serait pratique de pouvoir ouvrir un segment en particulier et de le modifier (pour faire une rapide correction lors de la relecture par exemple), puis de le fermer sans que le suivant ne s'ouvre directement, car cela coupe la lecture du texte.
19	"Truck" (mot et non concept), avoir les guillemets français disponibles directement en appuyant sur la touche (+ espaces après certains éléments de ponctuation), est-il normal que certaines phrases n'aient pas de point ?

Tableau 120 : Sondage post-expérience (Question 15)

La question 15 nous a permis d'avoir le feedback des étudiantes. Trois étudiantes nous ont fait part de leurs suggestions d'amélioration. Deux commentaires concernent le découpage du texte en segments qui est perçu comme négatif, car cela ne permet pas d'avoir une vue d'ensemble du texte et porte atteinte à la cohérence (ID15). L'étudiante 18 regrette *elle* aussi de ne pas avoir eu de « vue totale du texte », mais en TH. Enfin, l'étudiante 20 juge qu'il serait préférable de proposer la tâche de TH en premier lieu pour des raisons de complexité. Sur ce point, nous avons tenu à ce que l'ordre d'exécution des tâches varie afin d'éviter tout biais d'ordre dans ces expériences.

❖ **Question 16 – Compléter l'affirmation suivante : Après avoir effectué cette expérience...**

Q16				
Compléter l'affirmation suivante : Après avoir effectué cette expérience...				
	2018		2021	
	BAC3		M2	
	Nbr	%	Nbr	%
je me sens davantage enclin.e à recourir à la traduction automatique/post-édition (A1)	12	43 %	4	17 %
je me sens moins enclin.e à recourir à la traduction automatique/post-édition (A2)	3	11 %	1	4 %
mon opinion sur la traduction automatique/post-édition n'a pas changé (A3)	13	46 %	19	79 %

Tableau 121 : Sondage post-expérience (Question 16)

En 2018, la tenue de l'expérience a manifestement permis à plusieurs étudiantes de BAC3 de se forger un avis favorable sur la TA et sur la PE puisque parmi les 28 participantes, 12 ont déclaré être davantage enclines à recourir à la TA/PE à la suite de cette expérience. Trois étudiantes se disent, en revanche, moins enclines à le faire, et les 13 participantes restantes estiment que l'expérience n'a pas eu d'influence sur leur opinion. Cependant, nous n'observons pas la même tendance en 2021. En effet, la majorité des étudiantes de M2 (79 %) ont déclaré que leur opinion n'avait pas changé avec l'expérience. Seules quatre participantes (17 %) se disent davantage disposées à faire de la TA/PE. Peut-être avaient-elles tout simplement moins de préjugés défavorables avant de participer à l'expérience par rapport aux étudiantes de BAC3 en 2018. Ces divergences entre 2018 et 2021 nous amènent à nous poser la question suivante : Les perceptions des étudiantes à l'égard de la TA et de la PE s'amélioreraient-elles au fur et à mesure des années ?

7.6 Synthèse et vérification des hypothèses

7.6.1 Sondages pré-expérience

En synthèse, l'analyse des réponses recueillies dans ces deux sondages pré-expérience 2018 et 2021 a permis de faire ressortir que :

- les moteurs de TA les plus connus des étudiantes sont *Google Traduction*, *Reverso* et *DeepL* (Tableau 74). Ce dernier n'arrivait qu'en 3^e position en 2018, tandis qu'en 2021, *DeepL* se partage la 1^{re} place ex aequo avec *Google Traduction*, tous deux connus de l'ensemble des sondées ;
- les étudiantes de master ont fait part d'une plus grande familiarité avec la TA et avec la PE que les étudiantes de BAC3. Il apparaît, en effet, que le recours à la TA est plus fréquent chez les étudiantes de master que chez les étudiantes de BAC3, même si l'usage quotidien de la TA reste marginal parmi l'ensemble des sondées (Tableau 76). Notons qu'en 2021, les habitudes d'utilisation ont toutefois évolué ; l'usage d'outils de TA et la pratique de la PE se sont intensifiés aussi bien chez les M2 que chez les BAC3 et M1. La même tendance se dégage des travaux de Dirand et Rossi (2019), Loock et Léchaugette (2021), ainsi que Resende et Way (2021). Le recours à l'outil *DeepL* est en nette progression entre 2018 et 2021 puisqu'il s'agit du premier moteur dont les étudiantes déclarent faire usage en 2021 (100 % en M2 et 94 % en BAC3 et M1) devant *Google Traduction* et *Reverso* (Tableau 77). Toujours en 2021, remarquons que 6 % des répondantes en BAC3 et en M1 (soit trois étudiantes) ont déclaré ne jamais avoir utilisé *DeepL* ;
- beaucoup d'étudiantes ont évoqué, dans diverses réponses, le gain de temps conféré par la TA et par la PE, estimant que ce mode de traduction permet de faire gagner du temps dans le processus de traduction par rapport à la TH. Cette perception de gain de temps ressort également du sondage mené auprès d'étudiantes par Daems (2016) dans le cadre de son étude comparative, ainsi que des résultats du questionnaire post-expérience élaboré par Jia et al. (2019). Plusieurs étudiantes ont également admis avoir recours à la TA dans le but d'avoir accès au sens de certains passages du TS. Notre hypothèse HB1¹⁶¹ se voit ainsi vérifiée ;

¹⁶¹ Concernant l'usage de la TA, nous avons formulé l'hypothèse selon laquelle la majorité des étudiantes interrogées se sert de la TA la plupart du temps comme aide à la compréhension du texte source, mais aussi comme moyen d'augmenter leur productivité.

- la majorité des étudiantes interrogées considèrent la qualité d'une traduction générée par un moteur de TA comme « satisfaisante », voire « bonne » (Tableau 78). Ces résultats valident l'hypothèse HB3¹⁶² et corroborent les résultats du sondage mené par Loock et Léchaugnette (2021)¹⁶³. De plus, plus le temps passe, plus la satisfaction des étudiantes concernant la qualité en TA est à la hausse. Nous pensons que cette augmentation substantielle, entre 2018 et 2021, du taux d'étudiantes qui considèrent la qualité de la TA comme satisfaisante est corrélée avec les nombreuses avancées technologiques en matière de TA et au « bond en avant » (Moorkens, 2022b) qu'a fait la qualité mesurée entre l'ancien paradigme statistique et le nouveau paradigme neuronal. Mais cette amélioration des perceptions est probablement due également à l'intégration progressive de la TA dans l'enseignement de la traduction à l'Université de Liège dès l'année académique 2019-2020 ;

- la majorité des étudiantes tiennent la TH en haute estime par rapport à la TA. En effet, pour la très grande majorité des étudiantes interrogées, la qualité en TH reste, de manière générale, plus élevée que la qualité en TA (Tableau 79). De nombreuses étudiantes s'attendent, certes, à une amélioration de la qualité des traductions générées par un moteur de TA à l'avenir, qu'il soit proche ou lointain (Tableau 80). Néanmoins, force est de constater que, contrairement à ce que l'on aurait pu croire, et malgré les progrès accomplis en TA entre 2018 et 2021, il y a toujours aussi peu d'étudiantes en 2021 qui estiment que la qualité en TA (sans intervention humaine) devancera un jour la qualité en TH. Notre hypothèse HB3¹⁶⁴ se voit à nouveau confirmée. Ces observations vont dans le même sens que les réponses récoltées par Dirand et Rossi (2019) : 56 % des étudiantes en langues interrogées ne pensent pas qu'un outil technologique soit actuellement capable de remplacer totalement une traductrice humaine lorsqu'il s'agit de « traduire un document diffusable au grand public » (*ibid.*, p. 70) et elles sont tout de même 41 % à imaginer que cela sera possible dans 10 ans ;

¹⁶² Nous avons supposé que la plupart des étudiantes se montreraient satisfaites par la qualité d'une TA brute.

¹⁶³ Il ressort du sondage de Loock et Léchaugnette (2021) que plus de la moitié des étudiantes interrogées ont déclaré être souvent satisfaites de la qualité des sorties de TA, tandis que 38 % ne le sont que parfois.

¹⁶⁴ Nous avons supposé que les étudiantes allaient se montrer relativement optimistes quant à l'avenir des outils de TA et qu'elles seraient convaincues d'une amélioration des performances de cette technologie dans un avenir plus ou moins proche. Et nous nous attendions à ce que les étudiantes ne considèrent pas que la qualité de la TA puisse égaler la TH dans un avenir plus ou moins proche.

- validant partiellement notre hypothèse HB2¹⁶⁵, les étudiantes sont bien plus familières de la PE en 2021 qu'en 2018 : seules huit étudiantes sur 74 ont déclaré ne jamais avoir entendu le terme « post-édition » contre 44 étudiantes sur 105 en 2018 (Tableau 82). Ces résultats témoignent d'une part, de l'essor de la PE dans le secteur de la traduction, et d'autre part, de l'intégration de la PE dès 2019-2020 dans les cours destinés aux étudiantes en traduction à l'Université de Liège à partir de la 3^e année de bachelier. Il est toutefois interpellant de constater qu'en 2021, huit étudiantes déclarent toujours ne jamais avoir entendu parler de « post-édition » ;
- concernant la qualité perçue en PE, la majeure partie des répondantes est convaincue qu'il est possible d'atteindre en PE une qualité comparable à la TH (Tableau 86), ce qui confirme notre hypothèse HB4¹⁶⁶, mais va à l'encontre des avis recueillis par Daems (2016). Et manifestement, la qualité perçue en PE par les étudiantes est davantage satisfaisante en 2021 qu'en 2018, ce qui est certainement lié à l'intégration d'exercices de PE dans leur formation dès l'année académique 2019-2020 ;
- notre hypothèse HB5 qui porte sur la nécessité d'intégrer une formation à la (PE de) TA au cursus des étudiantes a pu être vérifiée. En effet, de la même manière que cela a déjà été mis en avant par Guerberof Arenas et Moorkens (2019), ainsi que par Jia et al. (2019) grâce à leurs expériences menées en contexte pédagogique, il existe un large consensus parmi les étudiantes ayant été interrogées par nos soins sur la nécessité d'être formées à la PE et sur l'utilité d'intégrer une formation spécifique dans leur cursus universitaire (Tableau 89). Ces constatations sont également semblables aux observations récentes faites par González Pastor (2021) auprès d'étudiantes en traduction : « As to MT adoption, most students [...] express their willingness to receive further training in MT and other technology translation tools » (p. 57). Parmi les raisons pour lesquelles *elles* se montrent en faveur d'une telle formation, citons : la nécessité de s'adapter aux évolutions technologiques du métier de traductrice et de répondre aux exigences du marché ; le gain de temps actuel ou projeté grâce à la PE ; l'apprentissage de nouvelles compétences ; le développement d'une approche raisonnée des outils de TA (*MT Literacy*) ; l'importance d'apprendre à faire de la révision. Quant aux quelques étudiantes ayant exprimé un avis défavorable, *elles* ont justifié leur position en évoquant la diminution de la qualité par rapport à la TH et le manque de confiance en la TA ; la dévalorisation de l'être humain par rapport à la machine ; l'importance

¹⁶⁵ Pour le test en 2021, nous nous attendions à ce que l'ensemble des étudiantes soient familières de cette pratique étant donné l'intégration systématique de la PE dans leur cursus dès la 3^e année de bachelier à partir de l'année académique 2019-2020.

¹⁶⁶ Nous avons supposé que plupart des étudiantes seraient d'avis que la qualité en PE est comparable à la qualité en TH et ce, d'autant plus en 2021 qu'en 2018.

de former d'abord les étudiantes à traduire avant de les former à post-éditer et la futilité d'une telle formation à la PE ;

- enfin, nous supposons que les étudiantes sont conscientes des changements profonds que connaît le secteur de la traduction et des enjeux qui en découlent pour la profession puisque la majorité, voire la quasi-unanimité (en 2021), d'entre elles se dit prête à accomplir des tâches de PE dans le cadre de leur futur emploi (Tableau 87). Cette disposition confirme une fois de plus les impressions qui ressortent de l'étude de Moorkens (2018) et de Jia et al. (2019).

7.6.2 Sondages post-expérience

Après avoir analysé en détail les réponses obtenues dans les sondages post-expérience 2018 et 2021, nous pouvons conclure que :

- de manière générale, comme nos hypothèses l'attestent, les étudiantes participant aux tests ont fait part de perceptions positives à l'égard de la TA et de la PE, et ce malgré le peu d'expérience en la matière. C'est également le constat qui ressort des travaux de González Pastor (2021)¹⁶⁷, de Jia et al. (2019)¹⁶⁸, de Moorkens (2018), de Şahin (2013)¹⁶⁹ et de Scansani et al. (2019). En outre, comme nous l'avons vu, les apprenties auraient tendance à avoir une opinion plus favorable sur la PE que les traductrices expérimentées (Moorkens et O'Brien, 2015) ;
- en 2018 comme en 2021, la tâche de TH a été désignée comme étant plus exigeante aux yeux des étudiantes que les tâches de PE (Figure 83), ce qui corrobore notre hypothèse HB6¹⁷⁰, ainsi que les avis recueillis par Daems (2016), Moorkens (2018) et Yamada (2015) dans leurs travaux respectifs. Nous avons également pu déduire que sur les deux expériences, la PE de *DeepL* a été la tâche jugée comme la moins exigeante par les étudiantes, quel que soit le TS. Il est intéressant de voir que, selon ce que Daems (2016) a pu démontrer dans son expérience, ces perceptions reflèteraient la charge cognitive réelle : « [S]tudents' perceptions correspond to the fixation analysis which showed that post-editing was cognitively less demanding than human translation » ;

¹⁶⁷ « Undergraduate students show general positive attitude towards MT after the training and the completion of activities » (González Pastor, 2021, p. 57).

¹⁶⁸ « [T]he students generally demonstrated a positive attitude towards NMT post-editing, although they were not totally accustomed to this new way of translating » (Jia et al., 2019, p. 79).

¹⁶⁹ « [N]ovice translators are positive about the place of this new component [MT post-editing] in the curriculum » (Şahin, 2013, p. 1).

¹⁷⁰ Nous nous attendions à ce que les étudiantes considèrent la PE moins exigeante que la TH.

- les étudiantes tiennent en haute estime leur production humaine, car plus d'une participante sur deux a estimé que le mode de traduction qui lui a permis d'obtenir la meilleure qualité est la TH (Tableau 113). Cette tendance, qui infirme en partie notre hypothèse HB7¹⁷¹, fait écho aux observations recueillies par Daems (2016)¹⁷², Gaspari et al. (2014)¹⁷³ et Samman (2022)¹⁷⁴, ainsi qu'aux conclusions de Moorkens et al. (2018b) : « All participants were faster when post-editing NMT, but they all still stated a preference for translation from scratch, as they felt less constrained and could be more creative » (p. 255). De plus, parmi les quatre modes de traduction, celui qui a recueilli le moins bon score est la PE de TA statistique. Nous en déduisons que les étudiantes considèrent que ce mode ne leur a pas permis de produire des traductions finales de qualité comparable ni à la TH, ni à la PE de TA neuronale. Force est de constater que cette impression se vérifie dans l'analyse statistique mesurant le taux de réussite, et ce, quel que soit le texte source, à l'exception du texte source 3 pour lequel nous avons relevé un plus haut taux de réussite en PE de TA statistique qu'en TH. Par ailleurs, les étudiantes misent un peu plus sur la TH que sur la PE de TA neuronale en matière de résultats qualitatifs. Ici en revanche, leur perception ne peut être étayée par les résultats de l'analyse statistique puisque, dans les deux expériences, les PE de *DeepL* ont été jugées de qualité comparable aux TH, voire de meilleure qualité globale, lors de l'évaluation humaine ;
- la tâche de PE de TA neuronale (*DeepL* et *Google*) a donné, dans l'ensemble, davantage satisfaction aux étudiantes que la PE de TA statistique (Figure 81 et Figure 82). Cet écart découle probablement des nombreux progrès qualitatifs apportés par le paradigme neuronal par rapport au paradigme statistique (Martikainen, 2019b ; Moorkens, 2022b). Nos résultats révèlent également un taux global de satisfaction en TH comparable au taux de satisfaction en PE de TA neuronale (Figure 80), ce qui infirme notre hypothèse HB8 selon laquelle nous nous attendions à ce que les étudiantes se montrent davantage satisfaites en TH qu'en PE. Il existe toutefois des divergences d'opinions et la PE de *DeepL* a récolté des avis plus tranchés qu'en TH : alors qu'aucune participante n'a déclaré être insatisfaite en TH, plusieurs le sont en PE de *DeepL* (7 % en 2018 et 4 % en 2021) ;

¹⁷¹ Nous nous attendions à ce que les étudiantes se montrent, pour la plupart, positives à l'égard de la PE et avons supposé qu'elles préféreraient la PE à la TH.

¹⁷² « Most participants, students and professionals alike, preferred human translation over post-editing » (Daems, 2016, p. 109).

¹⁷³ « [A]ll the user judgments tend to express consistent preference for manual translation from scratch over PE of MT output » (Gaspari et al., 2014, p. 68).

¹⁷⁴ « [T]he vast majority of students preferred HT. Findings from the retrospective pre-test survey validate these results as the before-survey responses revealed that over 85 % of students preferred HT [over PE] » (Samman, 2022, p. 168).

- en 2018, la participation à cette expérience a permis à plusieurs étudiantes de se forger un avis (davantage) favorable sur la TA et sur la PE (Tableau 121). Dans notre cas, ce sont 43 % des participantes qui se sont déclarées être davantage enclines à recourir à la TA/PE à la suite de cette expérience. Une tendance que l'on retrouve chez Gaspari (2001), chez Şahin (2013)¹⁷⁵, chez Daems (2016)¹⁷⁶, ainsi que chez Daems, Vandepitte et al. (2017)¹⁷⁷ après que les étudiantes ont reçu une formation à la TA et à la PE, et qui se confirmera plus tard dans les travaux de Guerberof Arenas et Moorkens (2019), de Samman (2022)¹⁷⁸ et de Stasimioti et Sosoni (2019)¹⁷⁹. Cependant, nous ne pouvons pas en dire autant en 2021, ce qui confirme notre hypothèse HB9¹⁸⁰. Effectivement, nous n'avons pas constaté d'évolution des perceptions chez 79 % des étudiantes de M2. Ces étudiantes avaient-elles tout simplement moins de préjugés défavorables avant de participer à l'expérience par rapport aux étudiantes de BAC3 en 2018 ?

7.7 Conclusion et discussion

Dans ce chapitre, nous avons présenté les résultats d'un sondage pré-expérience et d'un sondage post-expérience menés en 2018 et à nouveau en 2021 auprès d'étudiantes en traduction de BAC3, de M1 et de M2 à l'Université de Liège. Nous avons analysé un total de 179 questionnaires remplis : 105 en 2018 et 74 en 2021. Nous espérons que cette analyse a pu rendre compte de l'état des connaissances et des perceptions de ces étudiantes en traduction à l'égard de la TA et de la PE. Il va sans dire que les informations récoltées sur les connaissances et sur les perceptions des (futurs) traductrices dépendent inévitablement de l'ancrage culturel des personnes interrogées et de la place faite aux nouvelles technologies dans la société dans laquelle elles évoluent. Il convient également

¹⁷⁵ « [A]s the participants got used to the practice of post-editing, the level of negativity towards this activity decreased » (Şahin, 2013, p. 13).

¹⁷⁶ « We found that if participants changed their minds after the experiment, it was usually in favour of post-editing, indicating that understanding indeed leads to acceptance » (Daems, 2016, p. 160).

¹⁷⁷ « Since translators' attitude became somewhat more positive towards post-editing after participating in the experiment, we believe that post-editing should be included in translator training » (Daems, Vandepitte et al., 2017, p. 265).

¹⁷⁸ « [P]ost-intervention responses revealed a statistically significant shift towards more acceptance of MTPE training and use, indicating that the more students learned about the features of MT and MTPE skills and practiced them, the more positive their opinions became » (Samman, 2022, résumé).

¹⁷⁹ « [T]he findings of the questionnaire analysis reveal a change in the students' attitudes and perceptions after their training in MT and PE. In particular, the students reported a more positive attitude towards MT and PE » (Stasimioti et Sosoni, 2019, p. 134).

¹⁸⁰ Nous avons estimé que plusieurs étudiantes allaient effectivement se montrer plus favorables à l'égard de la TA et de la PE après leur participation, mais nous avons avancé que, pour la majeure partie des participantes, cette courte expérience n'allait pas impulser de changement d'opinion.

de souligner l'influence de la qualité de la TA (statistique et neuronale) en fonction des combinaisons de langues concernées dans ces différentes études, les performances de la TA étant toujours inégales selon les langues et selon le sens de traduction (Koehn et Knowles, 2017¹⁸¹ ; Poibeau, 2019¹⁸²). C'est pourquoi nous tenons une fois de plus à rappeler que la validité externe des données recueillies lors des sondages se limite à notre contexte d'expérience. En effet, nos analyse et interprétation des réponses obtenues ne valent que pour le profil de post-éditrices novices que constituent les étudiantes de BAC3, de M1 et de M2 en traduction de notre filière et pour la combinaison de langues étudiées (anglais-français).

En fin de compte, nous pouvons conclure que de manière générale, nous avons observé dans les commentaires moins de réticence envers la TA et la PE en 2021 qu'en 2018. Cela nous amène à nous poser une dernière question : Pourquoi les perceptions des étudiantes à l'égard de la TA et de la PE semblent-elles s'améliorer avec le temps ? Bien évidemment, ces divergences pourraient tenir uniquement à des variables propres aux étudiantes, mais nous croyons que d'autres facteurs ne doivent pas être sous-estimés. Comme cela a déjà été constaté dans d'autres études (Daems, Vandepitte et al., 2017 ; Guerberof Arenas et Moorkens, 2019 ; Şahin, 2013 ; Stasimioti et Sосoni, 2019), il est fréquent d'observer une évolution positive des perceptions des étudiantes après que celles-ci ont reçu une formation spécifique en TA et en PE. Ce qui est bel et bien le cas des étudiantes interrogées en 2021 puisque des exercices de PE ont été intégrés dans la formation des étudiantes en 2019-2020 dès la 3^e année de bachelier. En outre, nous sommes également convaincue que plus les années passent depuis l'émergence de la TA neuronale, plus les étudiantes sont, de manière générale, ouvertes à cette technologie et à la PE, plus *elles* se montrent désireuses d'y être formées.

¹⁸¹ « NMT systems have a steeper learning curve with respect to the amount of training data, resulting in worse quality in low-resource settings, but better performance in high-resource settings » (Koehn et Knowles, 2017, p. 28).

¹⁸² « Le repérage d'équivalents traductionnels au niveau de mots ou de segments fonctionne d'autant mieux que l'on a affaire à des langues proches, qui autorisent un découpage en mots relativement similaire entre langue source et langue cible. Cela se traduit très directement dans les performances des différents systèmes [...] La traduction vers le japonais ou, plus récemment, le chinois ou l'arabe, a suscité un nombre très important de recherches. Les performances, comparées à celles obtenues avec des langues indo-européennes comme le français ou l'espagnol, restent moindres, dans la mesure où ces langues ont une structure très éloignée de l'anglais, même si de grands progrès ont été enregistrés ces dernières années, et si les réseaux de neurones ont permis de réduire l'écart de performances » (Poibeau, 2019, pp. 137-138).

8 Intégration de la post-édition aux cursus de formation en traduction : enjeux, atouts et défis

Pour répondre au dernier objectif de recherche que nous nous sommes fixé, nous souhaitons contribuer, dans ce dernier chapitre, au débat sur l'actualisation des programmes de formation en traduction et tenter de mieux cerner les enjeux, atouts et défis de l'apprentissage de la PE en nous appuyant sur les implications de nos travaux empiriques, ainsi que sur notre expérience pédagogique. En premier lieu, grâce à la littérature scientifique, nous exposons les arguments qui étayent notre position en faveur de l'intégration de la TA et de la PE à la formation initiale en traduction. En deuxième lieu, nous menons une réflexion sur l'adoption d'une approche raisonnée des outils de TA. En troisième lieu, nous présentons une liste de sept défis majeurs que pose la PE de TA neuronale aux (futures) traductrices et aux formatrices, en l'état actuel de cette technologie, et sur lesquels nous pensons qu'il convient de mettre l'accent dans l'apprentissage de la PE. Pour terminer, nous essayons de répondre à certaines interrogations sur l'avenir de la profession.

8.1 Enjeux d'une formation à la TA et à la PE

Dans cette section, nous souhaitons mettre en avant la nécessité d'intégrer une formation à la TA et à la PE dans le cursus des apprenantes en traduction afin, entre autres, qu'elles puissent s'engager plus sereinement dans leur carrière professionnelle. De plus, nous explorons des pistes de réflexion sur le contenu et sur les atouts d'une telle formation, ainsi que sur le moment opportun pour envisager son introduction.

Parmi tous les avis de scientifiques du domaine et de pédagogues que nous avons été amenée à lire et à entendre dans le cadre de ce travail de thèse, il se dégage un large consensus sur l'importance de former les étudiantes en traduction à la TA et à la PE. En effet, le besoin de repenser le cursus des futures traductrices ressort avec évidence de la littérature scientifique, qu'elle soit récente ou plus ancienne : Belam, 2003 ; Brunette et O'Brien, 2011 ; Cholewska, 2021 ; Čulo et al., 2014 ; De Faria Pires, 2018 ; Daems, 2016 ; Depraetere, 2010 ; González Pastor, 2021 ; Guerberof Arenas et Moorkens, 2019 ; Jia et al., 2019 ; Kenny, 2018 ; Kenny et Doherty, 2014 ; Killman, 2018 ; Loock, 2019 ; Loock et Léchaugnette, 2021 ; Martikainen et

Kübler, 2016 ; Nitzke et Hansen-Schirra, 2021 ; O'Brien, 2002 ; Pym, 2013 ; Robert, 2010 ; Saint-André, 2015 ; Stasimioti et Sosoni, 2019 ; Volkart et al., 2022.

Il ressort effectivement de tous ces travaux la nécessité de revoir et d'adapter les programmes de formation initiale proposés aux apprenantes en y intégrant notamment des compétences en post-édition. Cela se fait d'ailleurs déjà depuis plusieurs années, entre autres dans le milieu universitaire : « Universities are increasingly including PE training in the final years of their translation degrees or as part of technical translation or localisation Master's programmes » (Guerberof Arenas et Moorkens, 2019, p. 232). En outre, la maîtrise de la PE est désormais attendue dans le monde professionnel et requise pour l'obtention du label EMT :

In translator training, PE has already become a standard practice. Many translation programmes have, in one way or another, incorporated PE into their curricula [...] and teaching of PE is now expected for all courses included in the European Master's in Translation Network (2017) [...] We may thus say that there is a whole generation of graduate translators for whom PE is expected as part of their jobs. (do Carmo et Moorkens, 2020, p. 39)

Il y a plus de 20 ans déjà, les conclusions de la recherche empirique menée par Gaspari (2001) plaident en faveur de l'intégration de la technologie et plus précisément de cours consacrés à la TA dans les cursus des étudiantes.

Brunette et O'Brien (2011) ont souligné, elles aussi, le besoin de voir émerger une formation universitaire en post-édition : « Que ce soit par nos expériences pédagogiques ou par les réponses issues des analyses contextuelles, nous sommes convaincues de la nécessité de former des agents de postédition à l'université » (p. 6).

Même s'il semble que les compétences en post-édition puissent s'acquérir avec le temps et l'expérience (Depraetere, 2010 ; O'Brien, 2002 ; Vasconcellos, 1986¹⁸³), Saint-André (2015) défend, elle aussi, la nécessité de former les étudiantes à la post-édition plutôt que de s'en remettre uniquement à la pratique.

Les conclusions de Daems (2016) vont aussi dans ce sens. Rappelons qu'elle a mené à l'époque une recherche empirique avec d'une part, des traductrices professionnelles et d'autre part, des étudiantes en traduction, ce qui lui a permis de comparer la PE de TA statistique et la TH pour la combinaison de langues anglais-néerlandais. Elle s'est intéressée au processus même de traduction, à la qualité des textes cibles et à l'attitude des traductrices envers la post-édition. Dans ses conclusions, elle soutient l'idée d'ajouter une formation à la post-édition aux cursus universitaires afin d'armer au mieux la future génération de traductrices :

¹⁸³ « Les compétences en post-édition s'acquièrent au fil du temps et de la pratique, le niveau d'aise augmente substantiellement après un mois de post-édition à temps plein (environ 100 000 mots) » (Vasconcellos, 1986 citée par Saint André, 2015, p. 53).

Our findings imply that post-editing is a viable alternative for human translation [...] Looking at the benefits of post-editing and the fact that most participants weren't opposed to post-editing after participating, perhaps specific post-editing training could be added to the translation curriculum to make for an even better future generation of translators. (ibid., p. 114)

Pareillement, dans son article intitulé « Intégration de la traduction automatique neuronale à la formation universitaire des futurs traducteurs », De Faria Pires (2018) reconnaît que si l'on entend préparer les futures traductrices/post-éditrices aux réalités de la pratique professionnelle,

il convient plus que jamais de remettre en question les différentes formations universitaires disponibles en traduction, notamment au sein de l'Union européenne, dans le but d'harmoniser ces dernières et de permettre aux étudiants suivant ce type de cursus d'avoir, lors de l'obtention de leur diplôme, toutes les cartes en main pour entrer le plus sereinement possible sur le marché professionnel de la traduction, marché ô combien rude et compétitif. (p. 50)

Il appelle à l'intégration systématique d'un cours de PE aux cursus universitaires afin que les étudiantes diplômées puissent entrer sereinement sur le marché professionnel de la traduction :

[L]es formations en traduction auraient tout à gagner d'une inclusion systématique d'un cours de post-édition harmonisé et adapté aux évolutions technologiques en matière de traduction automatique. [...] Grâce à un tel cours, les traducteurs en formation pourraient disposer des armes nécessaires pour affronter le marché de la traduction tel qu'il est réellement, un marché compétitif où la notion de productivité occupe le devant de la scène. (ibid., p. 60)

Enfin, d'après les résultats de leur récente recherche axée sur l'étude de deux stratégies en PE, Volkart et al. (2022) ont réaffirmé l'intérêt d'une telle formation : « the fact that students spotted just half of the MT mistakes is in line with the results of other studies and emphasises the importance of MT and PE teaching in the translation curriculum » (p. 6).

En outre, cette volonté de formation n'émane pas uniquement des scientifiques et des enseignantes, elle vient aussi du corps étudiant, bien conscient des bouleversements que connaît le secteur des services langagiers comme l'attestent les réponses à nos sondages (cf. chapitre 7), ainsi qu'aux sondages récoltés par González Pastor (2021) : « [A]s to MT adoption, most students reported to have used MT on their own motion at some time for their classroom translation assignments and express their willingness to receive further training in MT and other technology translation tools » (p. 57).

8.1.1 Quels atouts ?

Dans cette section, nous entendons faire valoir les atouts majeurs d'une formation spécifique en PE à l'intention des étudiantes.

Tout d'abord, O'Brien (2002) prétendait il y a plus de 20 ans déjà qu'avoir des compétences en PE est synonyme d'accroissement des opportunités d'emploi : « Companies wishing to implement machine translation technology would therefore benefit if translation graduates were already "comfortable" with post-editing. Additionally, post-editing skills would give translators an extra boost when it comes to finding employment opportunities » (pp. 99-100). Dix ans plus tard, elle réaffirme son point de vue : la technologie ne menace pas la profession, bien au contraire, « the increasing technologisation of the profession is not a threat, but an opportunity to expand skill sets and take on new roles » (O'Brien, 2012b, p. 118). Manifestement, cet argument est d'autant plus fort de nos jours étant donné que le marché mondial de la traduction ne cesse de croître (Hickey, 2023) et que les besoins en services de post-édition ne font qu'augmenter (Guerberof Arenas et Moorkens, 2019). Plus récemment encore, Cholewska (2021)¹⁸⁴ et González Pastor (2021)¹⁸⁵ ont également confirmé cette amélioration de l'employabilité des jeunes diplômées.

En outre, rappelons que les travaux de Belam (2003) lui ont permis de mettre en avant les bénéfices d'une formation à la PE pour les étudiantes en matière d'apprentissage des langues et d'enrichissement des compétences en traduction. D'ailleurs, elle n'est pas la seule à laisser entendre que s'exercer à la tâche de PE serait bénéfique à la fois pour *la future post-éditrice* et pour *la future traductrice*, même si *cette dernière* n'est pas amenée à pratiquer la PE. C'est en tout cas ce que défendent également Robert (2013) et Saint-André (2015)¹⁸⁶. Ainsi, Robert (*ibid.*) soutient que la tâche de post-édition permettrait d'accroître la performance de *la traductrice* (dans ses fonctions habituelles). Selon elle, en acquérant des réflexes de post-édition, *la traductrice* est capable d'appréhender différemment la traduction conventionnelle et d'être plus efficace car sa vigilance est plus aiguisée encore :

¹⁸⁴ « Overall, this study confirms the validity of integrating MTPE into the educational programme for future translators. More broadly, this means adapting teaching to the pace of technological development. In order to provide the best possible education aligned with the needs of the translation market, while at the same time increasing the employability of translation graduates in the future, an MTPE course should be included » (Cholewska, 2021, p. 209).

¹⁸⁵ « Overall, findings promote the incorporation of technology and MT in particular into the translation classroom in order to enhance students' skills and maximize their employability opportunities in the current high-speed changing translation market » (González Pastor, 2021, p. 57).

¹⁸⁶ « Puisque nous partons du principe que les traducteurs sont les plus aptes à mener à bien le travail de PÉ [...], il n'est nullement question ici de remplacer la formation en traduction, mais plutôt de la bonifier [...] À notre avis, les améliorations proposées profiteraient donc non seulement aux traducteurs qui feront de la PÉ, mais aussi à ceux qui n'en feront pas » (Saint-André, 2015, p. 136).

En acquérant de nouvelles facultés au niveau du décortilage, de l'analyse, du pointage, du repérage, de l'évaluation et de la modification, le post-éditeur devient un jongleur professionnel, très réactif, vif d'esprit et capable d'alléger et de clarifier toute phrase avec une efficacité redoutable. Grâce à cette agilité intellectuelle, le post-éditeur appréhende différemment la traduction traditionnelle lorsqu'il revient à ce type d'exercice et y est encore plus performant, car sa vigilance a décuplé. (p. 33)

Parmi les autres arguments avancés en faveur d'une formation à la PE, il semble que l'expérience et la formation en PE aient un impact sur la qualité des productions. C'est ce qui ressort en tout cas chez Guerberof Arenas (2014) :

[I]t was also observed that the cluster with the least or no experience in post-editing performs better with Fuzzy matches in terms of errors than with MT matches, and this seems to indicate that experience and training on postediting might have a pay-off in terms of quality, although this might not be the only factor. (p. 73)

De plus, nous savons pertinemment que formation et expérience d'usage jouent inévitablement un rôle dans l'acceptation d'une technologie¹⁸⁷, comme le souligne Dillon (2001) :

To be accepted, a technology must satisfy basic usability requirements and be perceived as useful by its intended user community. User experience and training will impact acceptance levels as will the manner in which the technology is implemented to contribute to organizational goals and working practice. (s.p.)

On retrouve, en outre, cette idée à maintes reprises chez O'Brien en ce qui concerne l'acceptation de la TA et la PE par la traductrice :

Translators who do not have post-editing skills are frequently hostile to machine translation technology. [...] Post-editing skills should be taught because [...] it could improve the uptake of machine translation technology by improving translators' perceptions of MT and its capabilities (O'Brien, 2002, p. 100).

[T]he more the professional translator is involved in the testing, implementation, and execution of translation technology, the more ownership she feels over the technology, and the more likely it is to be seen as an aid rather than a dehumanising threat. (O'Brien, 2012b, p. 109)

¹⁸⁷ « Des travaux récents sur l'acceptation des technologies montrent d'ailleurs l'importance de cette étape de prise en main : plus un système est perçu comme facile à utiliser, plus il sera adopté facilement (Vankatesh & Bala, 2008, p. 281) » (cités dans Rossi, 2019a, p. 95).

[B]eing a "good" post-editor is closely linked with an individual's attitude towards MT as a technology [...] If a translator dislikes MT as a technology, he or she will possibly be tempted to delete or ignore every MT suggestion. (O'Brien, 2022, p. 118)

Dans leurs travaux, Rossi et Chevrot (2019) ont observé chez les traductrices de la DGT interrogées un lien direct entre niveau de connaissances en TA et perceptions négatives : « [T]he translators who perceived MT as a threat were regularly those with the lowest scores for MT knowledge » (Rossi et Chevrot, 2019, p. 189). De la même manière, nous jugeons qu'une formation en TA et en PE est d'autant plus nécessaire qu'elle permet manifestement de désamorcer « les a priori défavorables et relativement tenaces des étudiants » (Rossi, 2019a, p. 102), ainsi que leur scepticisme à l'égard de cette technologie et de cette nouvelle pratique¹⁸⁸, ce dont Gaspari (2001) était déjà bien conscient : « University education and training [...] can successfully bridge the gap between trainee translators' natural scepticism towards MT and the impending need to equip tomorrow's professionals with necessary knowledge and skills » (p. 42). Dans leur étude, Scansani et al. (2019) ont supposé que l'attitude positive à l'égard de la TA qu'elles ont observée chez l'ensemble des étudiantes de master participant était le fruit de la formation spécifique reçue dans le cadre de leurs études :

No evidence of a lack of trust towards MT emerged. This behaviour confirms the positive opinion on MT stated in the pre-experiment questionnaire [...]. This constructive attitude and the ability to interact with technology may be the result of greater awareness of the limits and strengths of MT and PE practice, acquired as part of their academic education. (ibid., pp. 77-78)

Stasimioti et Sosoni (2019), deux chercheuses de l'Université ionienne (Corfou), se prononcent clairement en faveur d'une formation puisque, comme nous l'avons mentionné au chapitre 7, elles ont constaté un changement d'opinion après que les étudiantes de 4^e année ont suivi un module de formation à la PE dans le cadre du cours « outils de traduction » :

In line with previous studies, the findings of the questionnaire analysis reveal a change in the students' attitudes and perceptions after their training in MT and PE. In particular, the students reported a more positive attitude towards MT and PE; they were also more confident and faster and they avoided over-editing. It thus appears that training is crucial if MT is to be accepted and adopted by translators and if PE is to produce target texts with human-like quality. (p. 134)

¹⁸⁸ « [T]he role of training is particularly important, as it may affect MT and PE perception and translators' performance. Several studies, in fact, have shown that the attitudes of students changed positively after training in MT and PE (Alotaibi, 2014; Koponen, 2015; Çetiner and İşısağ, 2019; Guerberof Arenas and Moorkens, 2019) » (Stasimioti et Sosoni, 2019, p. 126).

Enfin, outre le fait que la formation pousse les étudiantes à dépasser leurs idées reçues, Guerberof Arenas et Moorkens (2019) estiment que cela leur permet également d'envisager leur insertion professionnelle en toute connaissance de cause :

Translation is, above all, a practical task. Therefore, universities offering translation and interpreting as a degree need to constantly adapt to the industry and modify their syllabi so that students are well-prepared to face the professional world [...] Such training helps improve students' flexibility regarding this task and change misconceptions about MT. At the same time, and more importantly, such training provides students with the tools necessary to make their own informed decisions about the type of work they can and want to do in the future, and to negotiate rates or deadlines with possible clients or employers. (p. 232)

Pour notre part, nous avons dégagé plusieurs observations tirées de nos données empiriques qui confirment la pertinence d'une formation à la TA et à la PE. Ainsi, tout en reconnaissant les nombreuses limites, toujours existantes, de la TA neuronale, nous estimons qu'il est légitime d'encourager l'apprentissage de la PE dans l'enseignement de la traduction, notamment car nos résultats ont révélé que ce processus permet aux étudiantes d'obtenir une qualité finale comparable, voire supérieure, aux productions humaines. De plus, étant donné que, dans nos corpus, la majeure partie des erreurs contenues dans les textes post-édités sont constituées de propositions de TA non post-éditées, nous sommes convaincue que le suivi d'une formation permettrait, entre autres, aux étudiantes de mieux repérer les erreurs fréquemment commises par la machine (telles que les calques) et d'apprendre à sentir le besoin de post-éditer tel ou tel élément en fonction du niveau de PE qui est attendu. La présence de caractéristiques propres à une langue post-éditée (*post-editeuse*) dans nos corpus renforce également, selon nous, le besoin de former les apprenantes afin de leur apprendre à atténuer ces traits caractéristiques en PE, si cela est jugé nécessaire. Enfin, il ressort de nos sondages que la majorité des participantes se montre en faveur d'une formation en PE et de son intégration dans les cursus. Rappelons que parmi les principales raisons avancées, les étudiantes ont cité :

- la nécessité de s'adapter aux évolutions technologiques du métier de traductrice et de répondre aux exigences du marché ;
- le gain de temps actuel ou projeté grâce à la PE ;
- l'apprentissage de nouvelles compétences ;
- le développement d'une approche raisonnée des outils de TA ;
- l'importance d'apprendre à faire de la révision.

8.1.2 Quand prévoir cette formation ?

Au vu des différents arguments présentés en faveur de l'intégration d'une formation à la PE dans les cursus des étudiantes en traduction, il est légitime de s'interroger sur le moment idéal

pour former ces étudiantes à la TA et à la PE. S'agit-il du début, du milieu ou de la fin de parcours ? Nous ne sommes d'ailleurs pas seule à nous poser cette question et nous nous joignons à l'avis quasiment unanime exprimé sur ce point par les traductologues (voir O'Brien, 2002¹⁸⁹ ; De Faria Pires, 2018 ; Guerberof Arenas et Moorkens, 2019¹⁹⁰ ; Nitzke et Hansen-Schirra, 2021) qui considèrent que le moment idéal pour envisager une formation à la PE est en fin de parcours, c'est à dire en master, voire fin du bachelier : « We recommend integrating PE in late B.A. translation programmes or even only in M.A. studies so that a general translation competence has already been developed to a certain degree » (Nitzke et Hansen-Schirra, 2021, p. 74). Selon De Faria Pires (2018), il est en effet évident qu'un cours de PE n'est pas à prévoir durant les premières années de formation. Au vu des compétences requises, il affirme que la formation devrait idéalement avoir lieu en fin de parcours, il propose de la mettre en place sur les deux dernières années de formation, c'est-à-dire de la réserver aux étudiantes de master. Aussi, nous avançons que la principale raison pour laquelle les étudiantes ne devraient pas suivre des cours de PE durant les premières années de leur cursus est la nécessité d'acquisition de compétences préalables.

8.1.2.1 Prérequis de formation

Beaucoup s'accordent sur le fait que la post-éditrice doit avoir avant tout de solides compétences en traduction : « It is essential that post-editors are skilled translators, because they need the same basic skill set » (Nitzke et Hansen-Schirra, 2021, pp. 69-70) ; « les résultats fournis [par la TA] doivent être systématiquement contrôlés, contrôle qui n'est possible que par un acteur disposant de compétences traductionnelles solides » (Loock, 2019, p. 63).

C'est pourquoi, tout comme Kenny et Doherty (2014), ainsi que Guerberof Arenas et Moorkens (2019), nous sommes convaincue qu'une formation de qualité en post-édition ne dispense en aucun cas l'apprenante d'une formation solide en traduction : « there is a growing demand for post-editing services, but [...] it may not be wise for those who are about to graduate to focus on post-editing at the expense of other "traditional" translation skills » (Kenny et Doherty, 2014, p. 286) ; « As MT evolves, syllabi need to adapt to innovations without disregarding

¹⁸⁹ De Faria Pires (2018) résume le point de vue de O'Brien (2002) : « Comme le souligne l'auteure, il apparaît crucial de choisir le moment approprié du cursus pour y insérer un cours approfondi de post-édition de traduction automatique. Ainsi, elle considère que les étudiants doivent au préalable disposer de compétences fondamentales telles qu'une excellente maîtrise linguistique, une aptitude à la gestion terminologique de base, des compétences en informatique et en traduction spécialisée. En outre, ils devraient idéalement avoir suivi une introduction aux technologies du langage. En ce sens, il semble évident qu'un tel cours ne devrait pas être intégré aux premières années de la formation » (De Faria Pires, 2018, p. 54).

¹⁹⁰ « The faster that universities implement these changes the better, preferably at a later stage of translator training » (Guerberof Arenas et Moorkens, 2019, p. 232).

attention to the core skills that any translator needs to succeed in a language-related activity » (Guerberof Arenas et Moorkens, 2019, pp. 232-233).

Idéalement, les apprenantes doivent disposer de certaines compétences et connaissances élémentaires avant d'acquérir des compétences plus pointues, et donc avant d'être formées à la PE : « [S]tudents need to acquire a set of core skills prior to acquiring new or more "sophisticated" skills [...] [T]ranslator training should ensure that students have sufficient core skills, knowledge, and self-confidence to be involved in MT activities and PE cycles » (Guerberof Arenas et Moorkens, 2019, pp. 232-233). Parmi ces compétences élémentaires, De Faria Pires (2018) relève la maîtrise de la langue source et de la langue cible, la maîtrise des aspects relatifs à la qualité en traduction (fluidité, respect de la fonction des textes, prise en compte du public cible, etc.), l'acquisition de solides compétences en traduction spécialisée « dans la mesure où les textes qu'elles post-éditeront par la suite pourront relever de domaines divers et variés » (p. 57), ainsi qu'une connaissance minimale des outils de TAO existants. Pareillement, le modèle de compétences en PE élaboré par Nitzke et Hansen-Schirra (2021) repose avant tout sur des compétences traductionnelles :

If we consider the competence model as a house of PE competences, the architecture of the house is grounded on the basic competences we also expect from professional translators: translation competences, including bilingual, extralinguistic and research competence. This is also the basis for a skilled posteditor. (p. 69)

Pour notre part, nous estimons également qu'une bonne maîtrise des langues source et cible, ainsi que de solides compétences en traduction sont indispensables pour pouvoir déceler et modifier les problèmes de transfert de sens et les problèmes de langue en TA. S'il est vrai que la connaissance préalable de la TAO peut constituer un plus pour apprendre directement à post-éditer dans un environnement de TAO se rapprochant d'un contexte professionnel par exemple, nous ne pensons pas qu'il s'agisse d'un prérequis incontournable pour suivre une formation en PE.

8.1.2.2 En début de parcours ?

À notre connaissance, le seul avis divergent sur cette question est celui exprimé par Rudy Look. Look (2019) plaide effectivement en faveur de l'introduction de cours en TA et en PE dès le début de la formation des étudiantes en traduction afin qu'elles puissent s'approprier « un outil d'aide à la traduction qui n'a pas vocation à les remplacer » (p. 55) et qu'elles puissent développer rapidement leur esprit critique vis-à-vis des sorties machine :

En mettant dès le départ les étudiants en garde contre les risques encourus face à des résultats favorisant la fluidité de la langue cible au détriment de la fidélité au texte source [...] le formateur développe l'esprit critique des étudiants vis-à-vis d'une technologie souvent caricaturée par ailleurs. (ibid., p. 62)

Il justifie son point de vue en expliquant que les étudiantes ont désormais très facilement accès à des moteurs de TA gratuits en ligne et ce, dès le début de leurs études, voire avant. Or, il est aujourd'hui indubitable que bon nombre d'étudiantes ont recours à ce type d'outils pour leur usage personnel ou pour faire leurs préparations de cours (voir notamment González Pastor, 2021), d'où la nécessité, selon lui, d'encadrer leur usage dès le départ et ainsi d'éviter toute « pratique buissonnière » ou « pratique clandestine » (voir thèse de Bourdais (2021) sur la consultation des moteurs de TA en ligne par les lycéens français).

8.1.3 Quel contenu ?

En dépit du fait que la PE est définie dans une norme ISO (ISO 18587), il n'existe pas, à notre connaissance, de protocole précis et détaillé de PE. De plus, il y a relativement peu d'études qui proposent des canevas de contenus de cours sur la PE. C'est également le constat que fait Nitzke (2019) :

Although the field is thriving, little has been published solely on PE training yet. While many empirical studies conclude with implications on what needs to be integrated in PE training or that PE training is necessary to educate professional post-editors, only few publications focused on how to design PE training. (p. 45)

Citons tout de même les travaux pionniers de O'Brien (2002), de Belam (2003) et de Depraetere (2010), ainsi que les pistes d'exploration données par De Faria Pires (2018) en matière d'enseignement de la PE.

En 2002, O'Brien proposait un module sur la PE comportant un volet théorique et un volet pratique. Pour elle, ce volet théorique doit couvrir les aspects suivants : introduction à la post-édition, introduction à la traduction automatique, introduction à la reformulation en langue contrôlée (LC)¹⁹¹, acquisition de compétences en gestion de terminologie et en linguistique

¹⁹¹ En pré-édition, il est courant d'avoir recours à une langue contrôlée : « Pre-editing often goes through the process of converting a natural language into a controlled language. A source text is revised using a simplified form of language based on some restrictions on lexicon, grammar and syntactic structures of the source text. The controlled language, as opposed to the natural language, refers to an artificial language that does not permit an erroneous expressive variation in sentence structures and semantic features » (Shih, 2021, p. 1555). Notons que l'utilisation d'une telle langue est controversée et « pas toujours possible (activité chronophage, requérant une certaine expertise et pouvant mener à une simplification excessive) et difficilement applicable à la communication scientifique » (Loock, 2020a, p. 271).

textuelle, ainsi que de compétences de base en programmation¹⁹². Concernant le volet pratique, elle recommande que les étudiantes puissent s'exercer à la PE non seulement en classe, mais aussi à domicile afin d'acquérir un maximum d'expérience pratique, qu'elles puissent s'entraîner sur différents types de textes, avec différents moteurs de TA, et adapter leur production en fonction des différents niveaux de post-édition (PE légère / PE complète). Elle suggère également que les étudiantes apprennent à utiliser des outils de gestion terminologique et des outils d'aide à la rédaction en LC¹⁹³, qu'elles développent des compétences en analyse de corpus et en création de macros¹⁹⁴, et qu'elles conçoivent éventuellement une application basique de PE.

Belam (2003) a organisé un atelier d'introduction à la PE dans le cadre d'un cours de traduction assistée par ordinateur pour des apprenantes en langue. Loin de perturber les étudiantes ou de leur donner de mauvaises habitudes, ce court séminaire a eu des effets positifs relatifs à l'apprentissage des langues (*ibid.*, p. 4). Belam soutient, en effet, que cet atelier s'est révélé bénéfique pour les étudiantes sur plusieurs points :

It can be shown to help their language learning and their translation skills at an appropriate level, and it also helps their awareness of the communicative aspects of language, and gives them a perspective on the use of foreign languages in the workplace. (ibid., p. 1)¹⁹⁵

Quelques années plus tard, dans une étude de cas portant sur l'enseignement de la PE à des étudiantes en traduction, Depraetere (2010) tente de déterminer des stratégies de post-édition à mettre en œuvre dans un contexte de formation. La chercheuse conclut qu'il n'est souvent pas nécessaire de demander aux étudiantes de réduire les changements stylistiques et les reformulations au strict minimum. En effet, conserver des traductions peu adéquates ne semble pas leur poser de problèmes : « [W]hen teaching post-editing to translation trainees, there is no need to (over-)emphasize the lack of need for stylistic changes or rephrasing.

¹⁹² « If equipped with programming skills, the post-editor could develop his or her own programme for automatically correcting consistent errors for specific language pairs, text types and MT systems » (O'Brien, 2002, p. 103).

¹⁹³ « Students are asked to check and edit a text in the source language using a CL tool and to submit the controlled and uncontrolled texts to a number of MT systems. Post-editing of both versions would then reveal the pros and cons of controlled authoring for machine translation » (O'Brien, 2002, p. 105).

¹⁹⁴ « students would acquire practical programming skills by writing macros to automatically apply common changes in target texts » (O'Brien, 2002, p. 105).

¹⁹⁵ Daems (2016) résume ces effets bénéfiques : « Belam (2003) described the benefits of introducing students to post-editing. Firstly, the post-editing task was beneficial to language learning, as students had to study the text, learn new vocabulary, and understand the text. Secondly, the task improved students' translation skills, as they had to discuss errors as well as translation strategies, and they learned that MT sometimes provides useful suggestions. Finally, discussing the degrees of post-editing in relationship to text function, content and style trained students' communication skills » (p. 33).

Students do not seem to experience problems keeping in translations that are less than perfect » (*ibid.*, s.p.).

Par la suite, Guerberof Arenas et Moorkens (2019) ont résumé l'opinion qui ressort des principales études portant sur l'apprentissage de la TA (statistique) et de la PE, et voici les points que cette formation devrait couvrir :

[R]esearchers [...] indicate that this training should be eminently practical and include: information on MT history, analysis of several types of engines with special emphasis on SMT (the dominant MT paradigm at the time the abovementioned works were published), pre-editing and controlled language, understanding levels of PE, MT output evaluation, output error identification, and a considerable amount of PE practice. (p. 221)

8.1.3.1 Acquisition de compétences spécifiques

Les spécificités de la post-édition par rapport à la traduction et à la révision classique (O'Brien, 2002), ainsi que les compétences particulières à acquérir, sont un argument souvent invoqué en faveur d'une formation pour les (futurs) post-éditeurs. C'est ce qui ressort notamment de l'étude comparative de TH et de PE menée par Čulo et al. (2014) :

[P]ost-editing should be taught as an additional competence for translators in order to minimize interferences. In fact, translation revision (of human-made translations) usually is part of the curriculum, but post-editing MT output has its own challenges. This PEMT workflow requires additional skills. (p. 203)

De même, les conclusions de l'étude menée par Sycz-Opoń et Gałuskina (2017) ayant montré que la tâche de post-édition requiert des compétences particulières par rapport à la tâche de traduction conventionnelle, les chercheuses préconisent l'intégration d'une formation spécifique à la PE :

The results of this study suggest that the post-editing of the MT raw output is as demanding for translation students as traditional translation, however, it requires a different set of skills, such as critical thinking and perceptiveness. Therefore, a special kind of training related to the effective use of MT technology should be implemented during translation classes. (p. 195)

Martikainen et Kübler (2016) insistent, elles aussi, sur la nécessité de reconnaître la nature spécifique de la PE et les caractéristiques des textes post-édités « pour sensibiliser les traducteurs post-éditeurs à l'influence de la TA dans le processus ; à l'éloignement du texte source qu'elle produit ; à cette tendance à la traduction directe » (p. 15).

De notre côté, tout en reconnaissant les lacunes toujours existantes de la nouvelle génération d'outils de TA (Alonso et Vieira, 2020), nous plaidons en faveur de l'intégration systématique d'un cours de PE dans le cursus des étudiantes en traduction. En effet, nous estimons qu'une formation spécifique à la PE répondant aux besoins des apprenantes permettrait d'éviter, en tout ou en partie, les types récurrents d'erreurs que nous avons mises au jour dans nos corpus et qui sont attribuables, en grande partie, à la présence de la TA. Tout comme de Faria Pires (2018), nous sommes convaincue que cette formation doit être axée sur la technologie neuronale qui est désormais le paradigme dominant.

8.1.3.2 Trois piliers de la formation

Poursuivant nos lectures de la littérature scientifique, nous avons dégagé trois grands piliers sur lesquels pourrait reposer cette formation.

8.1.3.2.1 Apprécier les avantages et les limites de la TA

Just as we need to be educated about social media nowadays, we also need to be educated about the pros, cons and serious pitfalls of MT.

O'BRIEN ET EHRENSBERGER-DOW (2020)

Beaucoup préconisent qu'une telle formation mette l'accent sur une sensibilisation aux avantages et aux limites, toujours existantes, de cette technologie.

C'était déjà le cas de O'Brien (2002). Selon elle, il est important que la post-éditrice acquiert des connaissances sur cette technologie afin de percer les mystères de cette « boîte noire » :

Knowledge of MT technology in general would go a long way towards helping the post-editor understand what is going on in the so-called "black-box" and why certain errors occur consistently. Understanding the history of MT development, its current status and future prospects would ensure that the post-editor had an appreciation for the technology, its limitations and how it might improve in the future. (p. 102)

Şahin (2013) souligne, lui aussi, ce besoin d'exposer les apprenantes aux avantages et aux inconvénients liés à l'utilisation de la TA :

MT post-editing activities can and should be part of the translation curriculum so that novice translators get more familiar with this new type of work demand and they become more aware of the potential benefits and drawbacks of using machine translation in the translation process. (p. 13)

Dans sa thèse de doctorat, Saint-André (2015, p. 135) relève deux principales compétences en informatique parmi les compétences à intégrer en priorité aux programmes de formation destinée aux traductrices/post-éditrices : la maîtrise des outils et la connaissance de leurs forces et faiblesses.

Pour Daems (2016) également, il serait judicieux d'offrir une telle formation pour permettre aux étudiantes de prendre conscience des limites de cette pratique : « The most straightforward reason for teaching post-editing is making people aware of its existence and limitations. [...] [W]e believe the process can be made more efficient, especially for students » (p. 160).

Selon Massey et Ehrensberger-Dow (2017), la mission des formatrices est d'amener les étudiantes à prendre du recul par rapport aux technologies langagières, en leur montrant ce que la machine est capable de faire ou pas : « [S]tudents should be encouraged to develop the metacognitive capacity to reflect on the deployment of language technologies, by learning about the capabilities and limitations of the machines and tools with which they are and will be working » (p. 307).

Quant à Moorkens (2018), il insiste sur la démystification de la TA neuronale auprès des étudiantes en traduction en passant à nouveau par l'identification des forces et des faiblesses de ces outils : « This makes it all the more important for translation students to familiarise themselves with and demystify NMT output, and to become aware that, despite the hype about machine learning, that NMT output has many weaknesses as well as strengths » (p. 376). À l'instar de Moorkens (*ibid.*), Loock (2019) encourage, lui aussi, les enseignantes à démystifier la TA neuronale : « [I]l incombe aux formateurs de faire prendre conscience aux futurs professionnels de ce que la machine peut faire et ne pas faire afin de démystifier (terme emprunté à Moorkens, 2018) cette technologie » (p. 55).

8.1.3.2.2 Sensibiliser aux erreurs récurrentes

L'intérêt d'exposer les apprenantes aux erreurs de la TA ressort de l'étude menée par Guerberof Arenas (2009) dans la mesure où les résultats montrent qu'à force d'en voir, le post-

éditrice chevronnée parvient à repérer les erreurs caractéristiques de la TA plus rapidement et plus facilement comparativement à une post-éditrice peu expérimentée.

Ce besoin d'attirer l'attention des étudiantes sur les erreurs récurrentes, et donc prévisibles pour la plupart en TA, a été exprimé à maintes reprises par les scientifiques ; Guerra Martínez (2003)¹⁹⁶, Daems (2016)¹⁹⁷; Depraetere (2010), Killman (2018)¹⁹⁸, mais aussi par O'Brien (2010), par Peraldi (2016)¹⁹⁹, par Rossi et Carré (2022), par Saint-André (2013)²⁰⁰ ou encore par Tatsumi (2010)²⁰¹.

Depraetere (2010) compte parmi les premières à promouvoir cette sensibilisation aux erreurs typiques en TA :

[T]here is a distinct need to raise the students' awareness of typical MT errors [...] It therefore seems very useful to carry out MT error analyses, if possible with some attention for the differences in typical errors of rule-based systems and statistical MT systems. (s.p.)

Être habituée aux erreurs fréquemment commises par la machine permettrait à la post-éditrice, d'une part, de réduire le temps passé à post-éditer, comme l'avance Schäffer (2003) : « familiarity with the pattern of errors produced by a particular MT system is an important factor in reducing post-editing time » (p. 136), et d'autre part, de diminuer la charge cognitive par rapport à la TH :

¹⁹⁶ « Knowing the errors produced by an MT system can both speed up the PE process due to the post-editor's familiarity with commonly repeated MT mistakes, as well as allowing for a better knowledge of this technology and its weaknesses » (Guerra Martínez, 2003, p. 49).

¹⁹⁷ « Post-editor training should focus on helping students spot typical machine translation errors that currently often go unnoticed (such as meaning shifts, wrong collocations, logical problems, and word sense issues), and ways to solve them » (Daems, 2016, p. 160).

¹⁹⁸ « [M]ore work needs go into educating students about MT error types that can be expected and should be corrected » (Killman, 2018, p. 137).

¹⁹⁹ Peraldi (2016) explique que « les rares spécialistes de la question insistent sur la nécessité de former les traducteurs afin de faciliter le travail de repérage et de correction des erreurs, notamment en inculquant un ensemble de stratégies permettant d'identifier des patrons types d'erreurs (O'Brien 2004, 4), mais aussi dans le but de développer des "schémas mentaux inédits" (Robert 2016, 3) afin d'adapter au plus juste sa pratique de traducteur aux besoins de la traduction automatique et ainsi parvenir à faire de la post-édition une activité hautement satisfaisante sur un plan intellectuel » (p. 8).

²⁰⁰ Saint-André (2013) avance l'idée d'une formation qui inclurait « un volet sur l'identification d'erreurs récurrentes et la compréhension de leurs mécanismes, puis sur une communication efficace et constructive de ces renseignements au client » (p. 19).

²⁰¹ « What they found most useful about the PE training included that they learned common problems of MT output and how to efficiently correct errors » (Tatsumi, 2010, p. 181).

[I]t is highly recommended that post-editing training programmes should be added to universities' programmes for translator training. It is reasonable to expect that, through systematic training, the student translators will benefit more from post-editing, in terms of saving processing time and reducing cognitive effort, as compared to from-scratch translation. (Jia et al., 2019, p. 80)

Les erreurs seraient effectivement repérées et corrigées plus facilement par la post-éditrice : « Getting used to the recurrent problems found in NMT outputs for a given language pair (and domain) will help you detect them and fix them more efficiently » (Rossi et Carré, 2022, p. 55).

8.1.3.2.3 Faire ressortir les spécificités humaines

Massey et Ehrensberger-Dow (2017) recommandent d'initier les étudiantes aux différents systèmes de TA et de mettre l'accent durant la formation sur les dimensions qui sont propres à l'humain :

[A] basic introduction to MT in its various forms, from rule-based MT through phrase-based MT to NMT, is indispensable. [...] Training programmes should encourage and foster the development of the human dimensions of intuition, creativity, and ethical judgement in order to ensure that graduates can deliver high-quality human translation when needed. (p. 308)

Guerberof Arenas et Moorkens (2019) préconisent elles aussi de renforcer les compétences qui permettent de démarquer la traductrice humaine de la machine, en se concentrant entre autres sur les aspects créatifs. Elles encouragent également les enseignantes à se mettre quelque peu en retrait pendant les cours et à endosser davantage un rôle d'accompagnatrice :

[A]s technology improves, as NMT has in the last couple of years, it is [...] important to focus on those skills that differentiate humans from machines: to increase the creative aspects of the training, not only with creative writing, which should be compulsory in translation training, now more than ever, but also with project-focused activities that help students to think innovatively rather than follow a set of given instructions to perform a set of given tasks [...] It cannot be expected that trainers can learn all of the existing technology at a fast pace and at a proficient level, therefore it is advisable to replace a more central role for the trainer with a role that seeks to guide and to learn with and from the students while also sharing their experience with them. To this end, the trainer will spend more time preparing and creating the course initially, so that during the class, they can take a step back and help students to resolve unexpected issues, organise teams, and cope with unfamiliar situations. (p. 233)

En outre, la nécessité de faire valoir les aptitudes transversales de la traductrice humaine sur un marché hautement concurrentiel et technologique ressort explicitement de la préface du référentiel de compétences EMT mis jour en 2022 :

Plus que jamais, l'intelligence, le savoir et les aptitudes humaines restent donc les facteurs clés pour fournir des traductions de qualité [...] Il ne s'agit d'adopter une approche plus étroite, mais plutôt de prendre conscience du fait que les aptitudes humaines constituent un élément de différenciation sur un marché du travail technologique, où les compétences linguistiques, critiques et éthiques peuvent se combiner pour produire un ensemble d'aptitudes transversales dont doivent se doter les diplômés de demain. (Groupe d'experts EMT, 2022, p. 2)

Plus récemment, González Pastor (2021) conseille d'attirer l'attention des étudiantes sur les répercussions que cette technologie a sur la profession : « Findings [...] revealed widespread preoccupation among students about how technology and MT affects professional work, so putting the focus on how technology affects working conditions, pay and professional self-image should be one key issue when introducing MT in the translation classroom » (p. 57).

En fin de compte, quels que soient l'approche et le dispositif pédagogiques adoptés, on ne peut que se réjouir à l'idée que les choses bougent depuis plusieurs années au sein des universités et des établissements de formation, et que de plus en plus de programmes en traduction sont revus et adaptés en tenant compte de l'évolution de la profession comme le confirme O'Brien (2022) :

Over time, MT and PE have been incorporated into translator training programmes in universities. There are many different approaches to this training. Some universities offer entire stand-alone courses on post-editing, some incorporate it into revision courses, and others weave it into translation technology courses. (p. 118)

Ajoutons qu'il peut également être envisagé d'intégrer des exercices de post-édition à des cours de traduction, comme le font par exemple plusieurs enseignantes de notre filière traduction-interprétation à l'Université de Liège depuis quelques années déjà.

8.2 Pour une approche raisonnée de la TA neuronale

Compte tenu du rôle croissant des outils de TA dans le monde de la traduction, il nous paraît essentiel d'en promouvoir un usage raisonné, tant auprès des professionnel/les qu'auprès des étudiantes, en référence au concept anglophone de *MT Literacy* (Bowker et Ciro, 2019 ; Loock, 2019 et 2020d ; Loock et Léchaugnette, 2021 ; O'Brien et Ehrensberger-Dow, 2020 ; Rossi, 2019a). Bowker et Ciro (2019), les deux chercheuses à l'origine du concept de *MT literacy* ont envisagé l'encadrement de l'utilisation de la TA uniquement dans le cadre d'activités de recherche scientifique et de publication par des non-professionnel/les. Cependant, De Clercq et al. (2021) proposent d'en adopter une définition plus large : « [Translators] should develop their MT literacy [...] that is the need to know what the machine can(not) do, what the

difference is between human translations and MT output, and what to focus on during the post-editing (PE) process » (p. 22). Quant à O'Brien et Ehrensberger-Dow (2020), elles insistent sur la nécessité de comprendre les rouages de la TA et de réfléchir au contexte d'utilisation de cette technologie : « MT literacy means knowing how MT works, how it can be useful in a particular context, and what the implications are of using MT for specific communicative needs » (p. 146). En ce qui concerne le fonctionnement de ces systèmes de TA, Loock (2019, p. 56) précise que, selon lui, il n'est pas nécessaire de le comprendre dans les moindres détails. En revanche, il insiste sur le besoin de faire comprendre aux étudiantes que la performance de ces outils dépend directement de la qualité des données du corpus d'entraînement. Loock (2020a) rejoint l'approche pragmatique défendue par Bowker et Ciro (2019) qu'il soutient pour la formation des futures traductrices : « les outils de traduction automatique sont là, [...] ils sont très fréquemment utilisés par de nombreux types d'utilisateurs, [...] et il importe donc d'être formé à leur utilisation raisonnée » (Loock, 2020a, p. 271). Il existe toute une série d'interrogations auxquelles ce type de formation doit tenter de répondre :

Au-delà de la connaissance, c'est une approche raisonnée prenant en compte les enjeux qui nous semble particulièrement importante : que sait faire la machine ? Que ne sait-elle pas faire ? Quels sont les risques ? Quand avoir recours à ces outils ? Quels problèmes déontologiques ? Quelle rentabilité ? (Loock, 2019, p. 56)

D'ailleurs, les objectifs et enjeux d'une telle formation sont nécessairement multiples :

A core focus of this training is ensuring that translation students understand the most recent approaches to MT, its strengths and its limitations, how to evaluate it and how to post-edit. Importantly, understanding when and how MT ought to be used has become central to training, both for translation students and for those who are not trained in translation. (O'Brien, 2022, pp. 118-119)

Rossi (2019a) plaide pour une réflexion collective afin « d'introduire un usage raisonné et bien encadré de la TA » (p. 101) dans les programmes de formation en traduction ; elle préconise d'enseigner aux apprenantes des stratégies d'adaptation pour leur permettre de « dépasser leurs craintes » (p. 102) et pour les encourager « à faire une lecture critique des évolutions actuelles et à réfléchir au rôle qu'elles souhaitent jouer dans un environnement qui fait la part belle au progrès technologique et aux intelligences artificielles » (*ibid.*).

En nous fondant sur les opinions des chercheuses que nous venons de citer, nous jugeons qu'il incombe aux formatrices d'orienter les étudiantes vers une approche raisonnée des outils de TA, avec l'objectif d'en faire des utilisatrices averties et autonomes pour une pratique responsable de la TA.

8.2.1 Responsabilisation

Élargissons encore davantage le concept de *MT literacy* en faisant référence à Kadiu (2016) qui met l'accent sur le développement d'un usage averti et d'une pratique responsable de la TA en renforçant la responsabilisation de *la traductrice* par rapport au recours à la TA, *celle-ci* ne pouvant se déresponsabiliser de sa production :

Grand nombre d'utilisateurs ignorent les incertitudes qui régissent le fonctionnement des outils de traduction automatique en ligne ; ils négligent le fait que, sous couvert de certitude et de mécanique, ces outils sont en mutation constante, qu'ils dépendent de décisions humaines réelles, qu'ils requièrent une évaluation attentive et vigilante. (p. 218)

C'est pourquoi il est impératif, selon elle, de responsabiliser les *traductrices* à l'usage de tels outils en les amenant à « reconnaître la part d'incertitude qui sous-tend ces systèmes et exige leur participation active » (*ibid.*, p. 218).

Dans la même idée, Loock et Léchaugnette (2021) sont persuadés qu'une approche raisonnée et responsable de la TA passe nécessairement par le développement de l'esprit critique des apprenantes : « As future professionals, students need to optimise their use of information provided by computers ; placing them in the role of critical decision makers is a way to raise their awareness of the limitations of MT » (p. 205). Partant du principe que les outils de TA sont là et à la portée de toutes, *elles* reconnaissent le besoin de démystifier la TA et d'armer efficacement les apprenantes :

Our aim should however remain the same: demystify MT technology so that students understand that human intervention remains necessary, even for translation purposes outside the translation industry. Since today MT tools are unavoidable in language classes, and not only for translation tasks, we need to find ways to empower students with the necessary skills to use what machines have to offer critically and professionally. (Loock et Léchaugnette, 2021, p. 217)

Comme l'avancent O'Brien et Ehrensberger-Dow (2020), on ne peut supposer que parce que la nouvelle génération a grandi avec le numérique et est hyperconnectée, qu'elle est forcément compétente en technologies numériques. De la même manière, on ne peut supposer que, parce que les moteurs de TA sont désormais accessibles à tout le monde, les *utilisatrices* sont des *utilisatrices* averties pour autant (p. 160)²⁰². Il est donc primordial d'être

²⁰² « Being able to use technology does not imply understanding it or being able to transfer one technology-specific set of skills to another: many people consider themselves digitally literate based on their use of social media. As digital natives become adults and join professional life, it is important to consider potential disparities in their use and understanding of digital technologies » (O'Brien et Ehrensberger-Dow, 2020, p. 147).

formé adéquatement en la matière afin de déterminer si l'utilisation de la technologie est pertinente :

MT literacy training will empower people to make better decisions and potentially lead to more demand for professional translation, or at least fewer bad decisions about deployment of MT [...] Increasing MT literacy will assist with the decision as to when it is sensible—from a process, economic and cognitive perspective—to use the technology and when it is not. (ibid., p. 158 et p. 160)

O'Brien et Ehrensberger-Dow introduisent un argument supplémentaire pour justifier l'adoption d'une telle approche ; en plus d'éclairer les futures post-éditrices, celles-ci sont ainsi en mesure d'accompagner les sociétés qui désirent intégrer la TA à leurs services :

An increasing number of translator training programmes are including the topic of MT, and competence in this technology is now seen as a core component of translation competence models. This training empowers professional translators not only to be knowledgeable end users of MT, but also to act as expert advisors to companies and organisations who wish to implement MT in their workflows. (ibid., p. 157)

8.2.2 Autonomisation

Il y a quelques années, Kenny et Doherty (2014) préconisaient l'adoption d'une approche holistique de la TA statistique visant l'autonomisation des traductrices :

[T]he adoption of such tools should not force translators into roles that they may find so limiting as to no longer be interesting or professionally rewarding. This is why we advocate a pro-active, holistic approach to the teaching and learning of SMT, one in which translators have ownership, critical understanding and a good deal of control. (p. 290)

En s'appuyant sur ce qu'avaient suggéré Kenny et Doherty (*ibid.*) pour l'ancienne génération de TA, Moorkens (2018)²⁰³ préconise que la traductrice humaine soit le plus autonome possible afin d'être capable d'intervenir à chaque étape liée à l'utilisation de la TA neuronale : phase de préparation, phase d'entraînement et phase de post-édition, ce que Guerberof Arenas et Moorkens (2019) considèrent par ailleurs comme un plus sur le marché du travail :

²⁰³ « The points of potential intervention in the SMT preparation, training, and post-editing processes that may benefit from the skills of the translator, as highlighted by Kenny and Doherty (2014), still hold true for NMT. With this point in mind, I contend that helping students to learn about new technologies, including NMT, is a positive and empowering intervention » (Moorkens, 2018, p. 376).

[R]esearchers suggest that a focus on MTPE enhances overall translation training with a view to joining the workforce as professionals, and to sharing the knowledge necessary for making decisions with regards to MT by being involved with the technology at all possible stages. (p. 221)

Dans la même idée, Rico (2017) suggère d’envisager la TA comme un processus, et non comme un produit, dont la traductrice devrait être en mesure de maîtriser toutes les étapes contrairement à la position défendue par Loock (2019, cf. section 8.2) :

An interesting approach to the incorporation of MT into translator training takes as a starting point the idea that MT is not a product but a process (Rico 2017). Every step of the process (data compilation, training and evaluation of MT engines, post-editing, recycling of TM systems) requires human expertise. According to Rico (2017), if properly trained, translators would occupy the very heart of the process and would be able to control and master it in all its facets. (cité dans Alonso et Vieira, 2020, p. 400)

À cet égard, depuis quelques années, plusieurs programmes en traduction misent sur une formation qui favorise l’autonomisation des apprenantes :

[P]lutôt que de tourner le dos aux évolutions actuelles, un certain nombre de formations ont pris le parti de donner aux étudiants un accès aussi complet que possible aux nouveaux usages, et de favoriser le développement de l’autonomie du futur traducteur dans l’interaction avec ses outils. (Rossi, 2019a, p. 94)

D’après Rossi (*ibid.*), il s’agit là de la meilleure manière pour permettre aux étudiantes de considérer la TA comme un outil à leur service plutôt que comme une menace :

Impliquer le traducteur pleinement et dans tout le processus, y compris lorsqu’il travaille avec une traduction qu’il n’a pas produite, est à notre sens la meilleure façon de passer d’une conception de la TA comme machine, qui pourrait sembler toute puissante et menacer de le remplacer, à la prise en main d’un outil, qu’il s’agit de mettre à son service. (p. 95)

8.2.3 Conscientisation

Il y a un aspect qu’il nous tient à cœur d’évoquer bien qu’il dépasse le cadre de cette thèse, il s’agit de la conscientisation des enjeux que la TA et, à fortiori, la PE soulèvent et plus spécifiquement de la problématisation éthique et idéologique de cette technologie. Nous jugeons que l’intégration de la TA au flux de travail de la traductrice et l’adoption de la PE ne peuvent être considérées de manière isolée du contexte dans lequel elles interviennent. Il s’agit donc d’amener les apprenantes, en tant que futures professionnelles de la traduction, à réfléchir de manière critique aux avantages et aux limites de ces outils et de les sensibiliser aux

enjeux éthiques et idéologiques de leur utilisation. C'est ce que revendique Loock (2020a) : « [U]tiliser ces outils sans en connaître les limites et les enjeux qu'ils posent, y compris éthiques, serait risqué, voire contre-productif » (p. 271). De même, Kadiu (2016) fait valoir l'importance de conscientiser les futures traductrices sur les problèmes d'ordre juridique et éthique relatifs à l'utilisation de moteurs de TA en ligne à des fins professionnelles. En outre, Kenny et Doherty (2014) avaient relevé à l'époque plusieurs problèmes en matière de protection des données personnelles, de confidentialité et de déontologie qui sont toujours d'actualité :

[Translators] may not be able to ensure client confidentiality, for example, and non-disclosure agreements may explicitly prohibit translators from uploading client material to web-based translation services. They may also have qualms about reusing the work of other unacknowledged translators, who may not have given permission for their translations to be used in this manner. (p. 288)

Ainsi, il incombe, selon nous, aux traductrices et formatrices de s'interroger, au-delà même de considérations éthiques, sur l'idéologie sous-jacente à l'adoption de nouvelles pratiques professionnelles (dont les outils de TA font clairement partie) dans une société libérale axée sur la rentabilité des marchés et dont le modèle économique repose sur la condition impérative d'une croissance infinie dans un monde aux ressources finies. À cet égard, nous faisons référence à la conférence de M. Cronin (2017b) « Translation and Sustainability in the Age of the Anthropocene » dans laquelle il soulève cette question dérangeante et pourtant essentielle :

To what extent can the translation or the translators do we have to ask ourselves a number of ethical questions and ethical issues about our activity becoming complicit in particular forms of resource extractivism, of economic development that, in all likelihood, will lead to climatic catastrophe and destruction? (s.p.)

8.3 Les défis en post-édition

En nous appuyant sur les résultats issus de nos recherches expérimentales, sur notre expérience en tant qu'assistante dans plusieurs cours d'exercices de traduction et de post-édition, ainsi que sur les échanges que nous avons eus avec les étudiantes en traduction dans divers contextes, nous avons tenté de cerner et de définir les défis majeurs que pose aujourd'hui la PE de TA aux (futures) traductrices. Il convient de préciser que cette liste de défis n'est pas exhaustive et n'est pas organisée selon un ordre hiérarchique.

8.3.1 Développer la capacité de discernement

Comme nous l'avons montré, la PE requiert la maîtrise de compétences spécifiques (Nitzke et Hansen-Schirra, 2021). Parmi ces compétences, il est essentiel que l'apprenante développe ce que nous appelons la « capacité de discernement » et que nous considérons comme une compétence clé en post-édition. Nous sommes d'avis qu'à force d'entraînement, *la post-éditrice* doit être capable de juger en peu de temps la qualité de la sortie de la TA, et de percevoir si le texte, le paragraphe, le segment vaut la peine d'être post-édité ou bien s'il est préférable de le retraduire à partir de zéro. Cette capacité figurait déjà dans la liste des compétences à intégrer en priorité à la formation des futures post-éditrices élaborée par Saint-André (2015) : « Les deux compétences en correction à intégrer en priorité concernent le repérage des segments à modifier [...] et la rapidité avec laquelle le postéditeur décide s'il doit intervenir et comment » (p. 132). À cet égard, Way (2018) a suggéré que les développeuses d'outils de TA définissent un système de cotation qui permette à *la post-éditrice* de trancher rapidement :

Accordingly, it behoves the entire MT developer community to deliver MT output with a score that is meaningful to human post-editors, so that they can immediately decide whether it is either quicker to post-edit the MT suggestion, or to translate the source string from scratch by hand. (p. 174)

Cette capacité de discernement est primordiale, principalement car elle permet un gain de temps et donc de rentabilité pour *la post-éditrice*, mais aussi une réduction des efforts cognitifs à fournir (Jia et al., 2019, p. 80). Il semblerait que cette capacité d'analyse des phrases et de repérage des éléments à modifier, et donc d'estimation de l'effort général de PE, s'affine avec l'expérience (Nitzke et Hansen-Schirra, 2021²⁰⁴ ; Robert, 2013²⁰⁵). Aussi, nous sommes convaincue que plus *la post-éditrice* post-édite, plus *elle* améliore cette capacité de discernement. Il est donc primordial que les (futures) traductrices apprennent à développer cette compétence dès le départ.

²⁰⁴ « [I]t should become easier to predict the PE effort with increasing experience » (Nitzke et Hansen-Schirra, 2021, p. 72).

²⁰⁵ « La capacité d'analyse des phrases et de repérage des points à modifier s'affine avec l'expérience, et le post-éditeur développe, au fil des années, un haut degré d'acuité et de précision dans le processus post-éditionnel. En acquérant de nouvelles facultés au niveau du décorticage, de l'analyse, du pointage, du repérage, de l'évaluation et de la modification, le post-éditeur devient un jongleur professionnel, très réactif, vif d'esprit et capable d'alléger et de clarifier toute phrase avec une efficacité redoutable » (Robert, 2013, p. 33).

8.3.2 Trouver le juste équilibre entre sur-édition et sous-édition

The notion of "trust" in MT output will be a key factor in users' capacity to assess the appropriateness of what the machine provides.

RUDY LOOCK ET SOPHIE LÉCHAUGUETTE (2021)

Une autre compétence clé de la post-éditrice est de savoir à quel point *elle* peut se fier à la proposition de la machine : « the post-editor needs to learn to what extent he or she can trust the MT output, e.g. concerning the correct translation of terminology, and when MT decisions have to be changed » (Nitzke et Hansen-Schirra, 2021, p. 71). Ainsi, nous considérons que l'un des principaux défis de toute post-éditrice, quelle que soit son expérience, consiste à trouver le juste équilibre entre sur-édition et sous-édition : « post-editors either over-edit the MT output making preferential choices or they under-edit it leaving errors uncorrected, while sometimes they also introduce new errors » (Stasimioti et Sosoni, 2019, p. 129). D'ailleurs, les étudiantes *elles-mêmes* prennent rapidement conscience de cette difficulté majeure, « students mentioned that they found it difficult to decide which MT translations were to be retained and which ones were to be corrected during the post-editing process » (Jia et al., 2019, pp. 78-79).

Sur-édition

Plusieurs études ont permis de faire ressortir un certain excès de zèle en PE, à savoir lorsque la post-éditrice modifie la TA brute plus que nécessaire. C'est le cas de l'étude menée par De Almeida (2013) dans laquelle les changements dits « préférentiels » opérés par la post-éditrice de TA statistique (anglais-français et anglais-espagnol) représentent entre 16 % et 25 % de l'ensemble des changements effectués. Selon la définition donnée par De Almeida (*ibid.*), un changement est considéré comme préférentiel « if the sentence from the raw MT output would still be grammatically correct, intelligible and accurate in relation to the source text, even if the change in question was not implemented » (p. 100). De Almeida et O'Brien (2010) considèrent qu'une post-éditrice compétente doit avoir entre autres « [t]he ability to adhere to the guidelines, so as to minimise the number of "preferential" changes, or changes that are not strictly speaking necessary, and which are normally outside the scope of PE » (s.p.). Quelques années plus tard, Way (2018) confirme cette tendance à la sur-édition en PE : « when translators are asked to post-edit MT, they often make unnecessary changes, as while the MT output might have been acceptable to some, "it's not quite how they would have said it themselves" » (p. 166). À cet égard, l'étude de Mutal et al. (2019) a montré qu'il y avait une proportion plus importante de changements jugés non pertinents en PE de TA neuronale par rapport à la PE de TA statistique. Selon Depraetere (2010) et De Almeida et O'Brien (2010), cette tendance à la sur-édition serait particulièrement observable chez les traductrices expérimentées. Inversement, selon Koponen (2015) et Imbres (2018), ce sont les étudiantes et

les post-éditrices novices qui feraient davantage preuve d'un excès de zèle par rapport aux post-éditrices expérimentées : « les post-éditeurs débutants sont parfois excessivement zélés dans leur activité et ont tendance à hyper corriger » (*ibid.*, p. 32). Dans tous les cas, nous avons pu constater dans nos expériences, ainsi que dans les cours incluant des exercices de PE, que certaines étudiantes vont systématiquement sur-éditer la TA brute, remettant trop fréquemment les propositions de la machine en question, opérant des changements non indispensables, notamment préférentiels, et se fourvoyant dans de longues recherches terminologiques, ce qui entraîne à fortiori une baisse d'efficacité et donc de productivité. Ce manque de confiance à l'égard de la TA (Cadwell et al., 2018) entraînerait également une augmentation inutile de la charge cognitive : « [T]his mistrust can also lead to unnecessary cognitive effort. If revisers know that MT was used, they may not necessarily trust the terminology produced by the MT system and may spend considerable time checking and double-checking terms » (O'Brien et Ehrensberger-Dow, 2020, p. 150). Cette méfiance est également évoquée par Rossi (2019a) : « En grande majorité [...], les étudiants ne jugent pas les sorties de TA suffisamment fiables pour que le traducteur puisse envisager de les utiliser » (pp. 98-99) et plus récemment encore par Zimina et Gledhill (2021) : « nos apprentis-traducteurs apprennent très vite qu'il faut se méfier des premières solutions proposées par la machine » (p. 79). Plusieurs facteurs expliquent ce manque de confiance, tels le manque de régularité des résultats et l'imprévisibilité de la qualité particulièrement en TA neuronale : « MT output errors are unpredictable, inconsistent and foster distrust » (Cadwell et al., 2018, p. 314) ; « Translators' lack of trust toward MT might therefore be influenced by different factors, including inconsistency/unpredictability of its output (especially true of NMT), or misconceived expectations about its functioning » (Scansani et al., 2019, p. 74). Ensuite, De Almeida (2013) laisse entendre qu'il ne serait pas rare d'observer une tendance à la sur-édition chez les post-éditrices/traductrices ayant une opinion défavorable sur la TA et sur la PE. Enfin, selon Martikainen (2021), les étudiantes seraient particulièrement enclines à sur-éditer lorsqu'elles ont une bonne compréhension des segments qu'elles sont en train de post-éditer.

Sous-édition

Paradoxalement, nous constatons également le phénomène inverse chez les étudiantes, c'est-à-dire une tendance à sous-éditer. De manière générale, certaines sont en effet enclines à laisser des traductions peu satisfaisantes sur le plan de l'idiomaticité (Depraetere, 2010)²⁰⁶, voire fautives dans certains cas. Nous pensons que cette tendance provient très probablement d'un excès de confiance à l'égard des propositions de la TA, ce qui pose évidemment problème : « Or, une foi absolue en l'outil peut se révéler problématique car elle peut non seulement empêcher une prise de décision véritable, et déresponsabiliser l'utilisateur, mais

²⁰⁶ « [T]he literal MT translation is accepted, even in cases where a more elegant or target language-oriented alternative is possible (as is clear from the human translation). The phrasing suggested by the machine is accepted provided it is informationally equivalent even though it may not be 100 % idiomatic » (Depraetere, 2010, p. 4).

aussi conduire à des malentendus et nier la visée communicationnelle de l'acte traductif » (Kadiu, 2016, p. 215). Kadiu (*ibid.*) dénonce un manque de transparence au sein même des systèmes de TA et incite *la post-éditrice* à cultiver l'incertitude et à développer un esprit de vigilance par rapport aux propositions de la machine :

La majorité des outils actuels n'incitent pas les utilisateurs à remettre en question la valeur des résultats fournis par la machine. Le manque de transparence de leur fonctionnement et l'apparente univocité des traductions qu'ils proposent n'encouragent pas les usagers les moins avertis à adopter une attitude critique relativement aux résultats obtenus. [...] Le problème majeur des outils de traduction automatique statistique aujourd'hui ne réside pas tellement dans leur fonctionnement mécanique, mais dans la dissimulation du caractère incertain du résultat fourni par l'ordinateur, c'est-à-dire dans la non-reconnaissance de cette part d'indécidabilité qui peut et doit survenir dans toute décision traductive responsable. (pp. 214-215)

Et cette confiance, parfois aveugle, en ce que la machine propose serait particulièrement observée chez les étudiantes (Daems, 2016 ; Deneufbourg, 2019 ; Depraetere, 2010 ; Dirand et Rossi, 2019 ; Loock et Léchaugnette, 2021²⁰⁷ ; Martikainen, 2020). Déjà en 2010, Depraetere faisait remarquer dans les conclusions de son étude de cas portant sur l'enseignement de la PE, que le risque avec les étudiantes n'est pas qu'*elles* ont tendance à effectuer trop de changements ou de reformulations en post-éditant, mais plutôt qu'*elles* ont souvent trop confiance en ce que la machine leur propose²⁰⁸.

Daems (2016) en arrive au même constat à la suite de l'analyse du recours à des ressources externes (dictionnaires, concordanciers, moteurs de recherche, etc.) durant la PE : « students might be too trusting of MT quality (as evidenced by the fact that a lower number of external resources is consulted when post-editing) » (p. 113).

Dans leur étude de cas, Loock et Léchaugnette (2021) déplorent également cet excès de confiance qui les amène à s'interroger : « The question of trust boils down to the more general question of human-machine interaction: how to make sure that the human is *assisted* and not *misled* by the machine because of excessive trust in the information generated by online tools? » (p. 205). *Elles* insistent alors sur le besoin d'apprendre aux étudiantes à ne jamais se

²⁰⁷ « We interpret these results as being the reflection of excessive trust from our students: they expect the information shown on their screens to be as reliable and truthful as the results of a calculator, for instance. As a result, they seem to be heavily primed by the MT output (even in the case of very awkward direct calques) » (Loock et Léchaugnette, 2021, p. 214).

²⁰⁸ « The danger involved in post-editing at this level is not that too much will be changed, but rather that there is at times too much trust in the translation engine » (Depraetere, 2010, p. 7).

fier aveuglément aux propositions de la machine et à ne jamais faire l'économie du retour au TS.

À nouveau, selon Martikainen (2021), les étudiantes auraient tendance à sous-éditer lorsqu'elles sont confrontées à des segments complexes dont le contenu leur échappe.

Les conseils formulés par Pym (2013) s'avèrent particulièrement à propos ; pour être une post-éditrice compétente, il recommande entre autres d'apprendre à faire confiance aux données, mais aussi à s'en méfier (*learn to trust and mistrust data*). O'Brien et Ehrensberger-Dow (2020) sont en outre convaincues que la notion de confiance en TA est étroitement liée à une utilisation raisonnée des outils de TA (*MT literacy*) et soulignent la nécessité de développer un scepticisme sain à cet égard : « Healthy scepticism is an important component of MT literacy [...] Those who know about MT systems and how they operate tend to have a healthy level of mistrust in the output [...] » (p. 152 et p. 150). C'est d'ailleurs l'un des objectifs que s'était fixés Killman (2018) dans son expérience effectuée en contexte d'apprentissage :

The pedagogical goal of having the students retranslate the same text using MT was to develop a better understanding of how MT may be beneficial to the translation task and to enhance their ability to determine what aspects of the output they may trust or distrust. (p. 116)

Nous pensons dès lors que la mission de la post-éditrice consiste à trouver un juste équilibre entre méfiance et confiance, entre sur- et sous-édition et qu'une formation adéquate peut les y aider.

8.3.3 Assurer la cohésion et la cohérence textuelles

[S]'il n'existe aucune raison de principe interdisant de penser qu'une machine soit apte à produire une traduction intelligente, force est de constater que certains mécanismes cognitifs échappent à la rationalité calculatoire.

MARGARETE FLÖTER-DURR (2022)

Nous partageons pleinement l'avis de Martikainen (2019d) selon lequel l'un des rôles importants de la post-éditrice est d'assurer à la fois la cohésion textuelle et la cohérence discursive du TC, en évitant toute rupture. La cohésion textuelle est une dimension interne au texte définie comme « le maillage linguistique du texte qui assure son unité » (Nita et al., 2021, résumé). Tandis que la cohérence textuelle est « l'adéquation de ce maillage textuel à la

situation de communication » (*ibid.*) qui permet à *la lectrice* de créer des corrélations entre le texte et le monde (Kerremans, 2017, p. 53) : « Human translation involves understanding the text on two levels, i.e. establishing relations between elements of the text themselves (text internal cohesion) and correspondences between elements of text and the real world (discourse external coherence) » (Poibeau, 2022, p. 6020). Garanties de la lisibilité et de l'intelligibilité d'un texte, la cohésion et la cohérence sont « les deux normes essentielles de la textualité, c'est-à-dire les conditions qui définissent le texte » (El Hajjari, 2020, p. 53) ; « Cohesion and coherence contribute to the general texture within a text. In other words, they are a set of characteristics that allows the text to function as a whole » (Kerremans, 2017, p. 53). Et pourtant, la TA neuronale demeure une mauvaise élève en la matière puisqu'elle est incapable d'assurer la cohérence externe : « While the machine can reasonably be expected to handle issues of text cohesion at some point, discursive coherence is out of its reach, since it requires knowledge of the world » (Poibeau, 2022, p. 6020), faisant de *la traductrice* humaine la garante de la cohérence puisqu'*elle* est la seule à pouvoir comprendre les textes source et cible et créer des liens avec le monde (Martikainen, 2019d). D'autre part, la TA génère aussi régulièrement des ruptures de cohésion notamment car la majeure partie des systèmes traitent encore les données au niveau de la phrase sans prendre en compte plus largement le contexte (*ibid.*), et forcément, sans qu'il n'y ait de prise en compte du texte dans son ensemble. Quelle que soit la taille des corpus d'entraînement, il manque à la machine toute une série d'informations extralinguistiques, comme le reconnaît Poibeau (2016),

le langage est affaire d'usage et par définition, le texte est un objet figé qui exclut tout le contexte communicationnel extralinguistique. Il est douteux que l'ensemble des connaissances nécessaires au traitement automatique de la langue puisse être inféré automatiquement et directement à partir de corpus, même si ceux-ci se composent de milliards de mots. (p. 11)

C'est pourquoi l'un des rôles essentiels de *la post-éditrice* est d'assurer cohésion et cohérence textuelles puisque

les systèmes assemblent des mots, des bouts de phrases en produisant des phrases grammaticalement justes, mais parfois dépourvues de sens ou hors contexte par rapport à l'ensemble du texte, ou formulées dans un registre de langue inadéquat, ou encore dépourvues de cohérence l'une par rapport à l'autre. (Grass, 2022, p. 14)

Par ailleurs, ce défi nous amène à soulever une problématique ergonomique en PE : la segmentation du texte, c'est-à-dire « the process of splitting source texts into chunks or bits of information (like a sentence or shorter) » (Alonso et Vieira, 2020, p. 398). Il est d'autant plus complexe d'assurer la cohésion et la cohérence textuelles que, depuis l'émergence des outils de TAO, *la traductrice* se voit de plus en plus contrainte de travailler un texte dans une architecture non naturelle et très fragmentée (segment après segment) (Dragsted, 2004 ;

Lavault-Olléon, 2019 ; LeBlanc, 2013 ; Martikainen, 2020). Les traductrices professionnelles s'en plaignent d'ailleurs fréquemment (Moorkens et al., 2018b²⁰⁹). De fait, cette architecture fragmentée est problématique à bien des égards puisqu'elle les empêche notamment d'avoir une vue globale du texte : « While this [segmentation] increases productivity, it also prevents translators from having an overall grasp of the text » (Alonso et Vieira, 2020, p. 398).

8.3.4 Adapter la qualité aux modalités de PE

Revenons à ce que nous avons évoqué au chapitre 2 concernant les niveaux de qualité attendus en PE. Il n'est pas aisé pour la post-éditrice débutante de parvenir à adapter la qualité de sa production en fonction des modalités attendues (Hynninen, 2018, p. 15), surtout lorsqu'il s'agit de faire le deuil de l'excellence : « [P]ost-editors [should] accept that sometimes the quality of the target text does not have to be 100 % to still fulfil the purpose of the target text » (Nitzke et Hansen-Schirra, 2021, p. 74). Dans certains cas, la post-éditrice doit ainsi apprendre à se contenter d'une qualité « satisfaisante » (*a good enough quality*, voir Koehn 2020²¹⁰) pour privilégier une dimension utilitaire de la TA et produire ce que Bowker (2020b) qualifie de « fit-for-purpose post-edited texts » : « Though the technology still has room for improvement, there is nevertheless ample evidence that many users remain happy to accept less-than-perfect MT output as long as it adequately serves their purpose » (p. 456). Dans les rapports rédigés par des étudiantes en traduction de l'Université d'Helsinki à la suite d'un cours de TA et de PE, Koponen (2015)²¹¹ a pu observer que la majeure partie des répondantes a déclaré avoir rencontré des difficultés pour s'adapter aux modalités de la PE légère, principalement en raison des attentes élevées en matière de qualité linguistique auxquelles *elles* sont habituées dans leur formation en traduction. Certaines étudiantes ont ainsi manifestement du mal à se

²⁰⁹ « They complained that the MT systems "conditioned" them to produce a literal translation, and found the limitation of sentential segmentation awkward » (Moorkens et al., 2018b, pp. 255-256).

²¹⁰ « Among vendors of machine translation technology, the phrase good enough is commonly used. It is an admission that machine translation technology today and in the foreseeable future is not perfect and will make mistakes. However, the technology, as it is today, is still useful for a wide range of tasks where perfect quality is not required. What constitutes good enough depends on the task. Another phrase overheard from discussions by vendors of machine translation technology is: *Quality is what the customer says it is*. Ultimately, the success of machine translation is determined if it helps accomplish the broader task, and the people working on the task are the best judge » (Koehn, 2020, pp. 41-42).

²¹¹ « Particularly the idea of "good enough" quality was new to them [students], and most (9 out of 13) explicitly commented that they had a hard time adjusting to a task where such quality was required. Thinking of the translation as a text that needed to be simply sufficient to convey the meaning needed some adjustment. However, most indicated that after the course, they now had a better understanding of differing situations where a lesser quality of language might suffice. Most also explicitly mentioned that they considered it important to learn to adapt to different requirements. Most of the students (7) who raised this theme [adjusting to PE quality requirements] also explicitly connected it to their prior translator training, where the need for "perfect" language is generally emphasized. They found it difficult to accept a translation that was "less than perfect" » (Koponen, 2015, p. 15).

satisfaire d'une qualité moindre en PE, ce dont nous avons pu nous rendre compte durant les différents cours pratiques donnés à l'Université de Liège. Tout en continuant à maintenir des exigences élevées en matière de qualité en traduction comme en PE puisque selon l'adage « qui peut le plus, peut le moins », nous sommes d'avis qu'il serait bon de travailler ponctuellement la faculté d'adaptation des étudiantes en leur demandant de réaliser des tâches de PE légère pour les contraindre à n'effectuer que les changements nécessaires en vue d'obtenir une qualité satisfaisante.

8.3.5 Atténuer l'effet fantôme

Se référant à ce que nous avons mis au jour dans le chapitre 6 et faisant écho à la notion de *shining through* définie par Teich (2003), nous sommes convaincue que tout comme la langue source interfère avec la langue cible en TH, la TA va avoir une influence sur la langue post-éditée pouvant « entraîner des "déviances" vis-à-vis de la norme attendue en langue originale » (Loock, 2016, p. 208) et limiter la créativité de la post-éditrice : « PE, as noted previously, has been shown to prime the translator and limit creativity » (Moorkens et al., 2018b, p. 255). Selon nous, la présence de la TA induit un biais cognitif dit « biais d'ancrage » selon lequel il est difficile de se détacher de sa première impression. Nous pensons, en effet, qu'il n'est pas toujours évident pour la (future) traductrice de parvenir à s'éloigner de la traduction brute affichée sur un écran une fois lue (Loock, 2019), autrement dit, de réussir à « lever le nez » de la TA brute et à se détacher, parfois radicalement si nécessaire, des propositions de traduction qui sont directement sous ses yeux. Les post-éditrices seraient ainsi influencées malgré elles par la machine ; il s'agirait du *priming effect* de la TA (Carl et Schaeffer, 2017²¹² ; Green et al., 2013 ; Loock et Léchaugnette, 2021 ; Toral, 2019) que Loock et Léchaugnette (2021) définissent comme : « the influence of the MT output preventing them [translators] from considering alternative solutions (priming) » (p. 217). Comme nous l'avons vu au chapitre 6, cet effet fantôme de la TA a pour conséquence d'aboutir à une PE qui reste très proche de la traduction machine :

[T]he MT output performs a translation that tends to be as equivalent as possible to their source texts. It is possible that when post-editing the raw MT output translators are primed by the MT choices even though they were instructed to change the text to achieve a high-quality translation for publication standards, thus resulting in a PE text similar to the MT output. (Castilho et Resende, 2022, p. 20)

De plus, nous considérons cet effet fantôme comme un facteur ayant une influence néfaste sur la qualité du texte cible. Martikainen (2019a) souligne le fait que la traduction humaine

²¹² « Post-editors are heavily primed during this process so that they more easily accept sub-optimal translations which human translators, working from scratch, would otherwise not produce » (Carl et Schaeffer, 2017, p. 44).

exige un plus grand degré d'interprétation du TS qu'en PE : « [L]'interprétation du texte source joue une part importante [...], et dans la post-édition, cette étape est absente, puisque c'est la traduction automatique qui se substitue à l'interprétation première du texte source par le traducteur » (p. 265). Tout comme l'affirmait déjà LeBlanc (2016) par rapport aux mémoires de traduction, nous croyons que de la même manière, la présence de la TA tend à la fois à restreindre les intuitions de traduction, et de façon générale à limiter la créativité traductionnelle de la post-éditrice (Cadwell et al., 2018²¹³ ; Grass, 2022²¹⁴ ; Guerberof Arenas et Toral, 2021²¹⁵ ; Jia et al., 2019²¹⁶ ; Moorkens et al., 2018b²¹⁷), tout en encourageant la passivité, en entraînant une certaine dépendance aux outils et en nuisant aux réflexes naturels en traduction (*ibid.*, p. 83). Et cela se ressent logiquement sur la qualité de la production finale. C'est également l'avis partagé par O'Brien et Ehrensberger-Dow (2020) : « [T]he presence of the MT output might cause translators to fixate on it with too little regard for the source text (see Elming et al., 2014), which could be detrimental to the creative process » (p. 152). Étant donné que la PE requiert certaines compétences spécifiques par rapport au processus de TH, nous supposons que les facultés cognitives à l'œuvre en PE sont en partie distinctes des facultés cognitives en TH. Aussi, nous croyons que cet effet fantôme entrave notamment le processus de déverbalisation (cf. nbp 110) et entrave toute « approche somatique de la traduction » (*The Somatics of Translation*)²¹⁸ ; approche scientifique proposée par Robinson dès 1991 dans son ouvrage *The Translator's Turn* se fondant essentiellement et avant tout sur l'intuition en traduction.

Nous trouvons à cet égard que les études cognitives en PE sont d'une importance capitale pour approfondir l'exploration des effets de la PE sur les stratégies cognitives déployées par la traductrice.

²¹³ « Reasons not to use MT identified during coding and analysis of the focus group data: Because using MT diminishes creativity » (Cadwell et al., 2018, p. 311).

²¹⁴ « Dans les domaines où le sens et le style sont prépondérants, c'est-à-dire dans l'intégralité des sciences humaines, une ébauche automatique peut être contreproductive en ce qu'elle vient brider l'élan créatif du traducteur, voire l'induire en erreur » (Grass, 2022, p. 13).

²¹⁵ « The method applied to this study, by quantifying errors and creative shifts, gives us an interesting perspective on the differences between translation modalities, and reveals that HT scores higher for creativity than MTPE, all other things being equal » (Guerberof Arenas et Toral, 2021, pp. 22-23).

²¹⁶ « [T]he students found from-scratch translation and post-editing very different in many aspects. Eleven students mentioned that there was much less room for them to show their creativity and linguistic skills and also less freedom during post-editing, especially for the general language texts » (Jia et al., 2019, p.78).

²¹⁷ « PE is a task disliked by many translators, who have complained that it constrains their work, allows limited opportunities for creativity » (Moorkens et al., 2018b, p. 240).

²¹⁸ « Translator should just translate any way s/he feels, based on "pure intuition" – which is to say, out of pure solipsistic subjectivity, without the slightest thought or other concern for accuracy or equivalence [...] I was actually saying that we *feel* accuracy, we *feel* equivalence and we *feel* the norms that governs those things » (Robinson, 2015, p. 299).

Afin d'atténuer au maximum cet effet fantôme, nous sommes convaincue qu'il est indispensable, non seulement d'apprendre aux étudiantes à développer un regard critique sur les informations fournies par la machine (Loock et Léchauguette, 2021), mais aussi de leur faire prendre conscience de leur valeur ajoutée par rapport à la machine (Loock, 2019).

8.3.6 Déjouer l'illusion de fluidité

La TA neuronale, parce qu'elle produit de résultats en apparence meilleurs, requiert une vigilance accrue du traducteur afin de déceler des problèmes plus subtils.

BAPTISTE DIRAND ET CAROLINE ROSSI (2019a)

Il est indubitable que la nouvelle génération de TA a fait d'énormes progrès par rapport à la TA statistique (Loock et Léchauguette, 2021 ; Poibeau, 2019 ; Moorkens, 2022b), notamment en matière de fluidité en langue cible (Loock, 2018 ; Macken et al., 2019 ; Rossi et Carré, 2022). En effet, la TA neuronale, et plus spécifiquement le cas de l'outil *DeepL* (Loock 2018), est connue pour mettre l'accent sur la fluidité et sur une certaine élégance linguistique en langue d'arrivée. Selon Koehn (2018), cette technologie produirait un « bel emballage stylistique et syntaxique [susceptible] de mener le post-éditeur à passer à côté d'erreurs de sens évidentes et potentiellement graves, qui existent toujours en dépit du caractère sophistiqué des approches neuronales » (cité dans De Faria Pires, 2018, p. 56). Bon nombre d'études ont ainsi démontré que l'un des grands torts actuels de cette technologie était de produire ponctuellement une « illusion de fluidité » (Deneufbourg, 2019 ; Guerberof Arenas et Moorkens, 2019 ; Martikainen, 2019b ; Rossi et Carré, 2022 ; Way, 2019²¹⁹), et ce, parfois même au détriment de la fidélité au sens du texte source (Loock, 2019, Macken et al., 2019 ; Nitzke et Hansen-Schirra, 2021) : « the quality of NMT output is often "human"-like: smooth, elegant and fluent, even in cases when it contains translation errors » (*ibid.*, p. 74). Dès lors, il arrive que les propositions de la machine soient grammaticalement et syntaxiquement correctes et « sonnent bien », mais qu'elles masquent en réalité des erreurs sémantiques plus ou moins graves (Guerberof Arenas et Moorkens, 2019) de type glissement de sens, contresens, non-sens, ajout/omission fautive, etc. qui sont souvent difficiles à repérer par l'étudiante :

²¹⁹ « [Neural] MT output can be deceptively fluent; sometimes perfect target-language sentences are output, and less thorough translators and proofreaders may be seduced into accepting such translations, despite the fact that such translations may not be an actual translation of the source sentence at hand at all » (Way, 2019, p. 320).

Even if fluency is better in NMT output when compared with SMT, this fluency can be deceiving because even if sentences are grammatically correct, the content might be wrong. This is a known weakness in NMT systems. The output is far from perfect and still presents errors such as omissions, additions, terminology mismatches and mistranslations, not to mention the usual difficulties in translating long sentences (over 25 words). (p. 231)

Cette illusion de fluidité en TA est particulièrement pernicieuse, car elle a pour conséquence de renforcer la confiance dont certaines étudiantes font preuve symptomatiquement à l'égard des propositions de la machine (Deneufbourg, 2019), ce qui peut les amener à sous-éditer : « student post-editors mostly tend to spontaneously accept MT output as such, and rarely edit these suggestions that convincingly mimic fluent human translation » (Martikainen, 2019b, p. 13). Cette étude nous aura effectivement permis de nous rendre compte que l'élégance linguistique n'était en fait, par moments, que superficielle. En effet, après avoir passé les différentes productions des étudiantes au peigne fin, nous remarquons qu'à certains égards, les propositions de la TA neuronale nous apparaissent davantage dangereuses et trompeuses pour les étudiantes que celles de la TA statistique. De prime abord, ces propositions de traduction leur paraissent satisfaisantes, tandis que les mauvaises propositions de la TA statistique sont, en général, tellement inadéquates ou saugrenues que l'étudiante n'a d'autres choix que de post-éditer. Ce ressenti trouve notamment écho dans les travaux de Dirand et Rossi (2019), Moorkens (2018), Mutal et al. (2019), Nitzke et Hansen-Schirra (2021) et Yamada (2019). Ce dernier a pu observer dans une étude comparative que les étudiantes avaient moins bien post-édité en TA neuronale qu'en TA statistique. Ce phénomène serait dû, selon lui, au fait que la TA neuronale a tendance à générer des erreurs qui ressemblent davantage à des erreurs qu'un humain commettrait, ce qui rendrait les erreurs plus difficilement repérables par les étudiantes²²⁰. Cette difficulté est également soulevée par Rossi et Carré (2022) : « even though NMT quality has undeniably been getting better, it is probably not easier to deal with than other types of MT, and MT is never a simple recipe for success. Instead, you will need to pay attention to small mistakes hidden in a fluent MT output » (p. 55). Néanmoins, précisons que tout comme le formule De Faria Pires (2020) dans l'un de ses postulats : les problèmes de qualité présents dans les textes post-édités ne sont pas systématiquement imputables à la TA brute. Quant à Mutal et al. (2019), leurs résultats confirment que la TA neuronale est capable de paraphraser et d'utiliser correctement des synonymes. *Elles* ont également remarqué que par rapport à la TA statistique, les traductrices ont plus de difficultés en PE de TA neuronale à déterminer s'il est nécessaire de modifier une phrase : « We also observe there to be less

²²⁰ « The results of this study suggest that the more advanced, human-like translation abilities of NMT make it even more challenging for student translators to meet a professional standard of post-editing quality [...] [T]he advent of NMT [...] resolves typical SMT errors (in its raw output) before human post-editing, generating a similar error distribution to humans – NMT translates like humans, and produces errors like humans » (Yamada, 2019, p. 87 et p. 97).

agreement on differences to be corrected in NMT than SMT, suggesting that errors are easier to identify in SMT » (Mutal et al., 2019, p. 2). Enfin, *la* (future) post-éditrice se doit d'être davantage vigilante en TA neuronale pour déceler des problèmes subtils, car ce mode de traduction produit des résultats qui sont meilleurs en apparence (Dirand et Rossi, 2019) ; d'ailleurs, plus la TA brute paraît fluide, plus il convient d'être prudent (Rossi et Carré, 2022). C'est pourquoi nous sommes persuadée que l'un des grands défis en PE consiste à ne pas se laisser leurrer par cette apparente fluidité et ainsi parvenir à déjouer l'un des grands pièges actuels en TA neuronale.

8.3.7 Revaloriser le rôle de l'humain

Avec la révolution numérique et les incontestables progrès de la traduction automatique/machine (TA) depuis l'arrivée de la TA neuronale, il semble incontournable d'enseigner aux futurs professionnels la façon dont ils peuvent et doivent dégager leur plus-value de traducteur humain.

RUDY LOOCK (2019)

Loock (2020b) est convaincu que pour parvenir à maîtriser les outils de TA et à les utiliser correctement, il ne faut pas les rejeter, mais il ne faut pas les aduler non plus, il faut plutôt s'attacher à dégager la valeur ajoutée de *la* traductrice humaine. Nous souscrivons de toute évidence à la nécessité de revaloriser le rôle de *la* traductrice humaine par rapport à une machine qui traduit automatiquement, d'autant plus auprès des traductrices de demain. Dans une étude de cas, Loock (2018) a relevé une déviance par rapport à l'usage linguistique en français et ce, malgré les progrès qualitatifs en TA neuronale. Ce non-respect de l'usage linguistique est problématique, affirme-t-il, car il révèle à *la* lectrice le statut traduit du texte, ce qui va à l'encontre des attentes du marché de la traduction en matière d'invisibilité de *la* traductrice. D'où cette nécessité de faire ressortir la valeur ajoutée de l'apprenante par rapport à une machine :

In an educational setting, by focusing on specific linguistic features, translators-to-be can be sensitised to the performance and limits of MT systems and therefore define their added value over the machine, since MT output, in spite of indisputable progress, does not seem to take into account language norms such as frequencies of use. Students can then become aware of the gap that exists between original and machine-translated language, while in order to reach the invisibility required by the industry this gap should be at least reduced. (Loock, 2020c, p. 165)

Cette prise de conscience du rôle essentiel de l'humain est primordiale car il est un fait qu'en PE, beaucoup de traductrices ont perdu le plaisir de traduire et se sentent fréquemment dévalorisées dans cette tâche « d'autant plus ingrate qu'elle n'est guère créative et substituée à la démarche valorisante du transfert de sens celle plus ardue de reconstitution du sens et de la cohésion » (Grass, 2022, p. 12).

Elles se sentent ainsi reléguées à un rôle de simple « réparatrice » des erreurs de la TA pour lequel elles reçoivent de surcroît une moindre rétribution :

In the context of MT, not only can translators feel replaced by the machine, but the machine generates fundamental linguistic errors that a trained human translator would rarely generate. The professional translator is then demoted to the status of a fixer (Krings 2001) of seemingly unintelligent errors. That they are paid lower rates to fix such errors than to create their own translation adds to the feelings of negativity. (O'Brien, 2012b, pp. 108-109)

Martikainen (2022a) déplore, elle aussi, le caractère déshumanisant de ces outils de TA : « these tools [...] impose a workflow that tends to reduce the human translator's role to simply correcting mistakes made by the machine, in a one-way process with no real interaction » (s.p.).

DePalma et Lommel (2017) reconnaissent aussi que bon nombre de traductrices ont perdu le goût pour leur profession, le métier ayant perdu de ses qualités attractives :

Until now language technology developers have focused their work on speeding up the process and lowering costs. Those drivers have left many translators feeling alienated from the very aspects of their work that attracted them to the job in the first place – the creativity of language, the challenge of solving difficult problems, and the ability to work on stimulating texts and topics. Translators often find that they spend as much time managing the technology as they do translating, and that their rates are always under pressure. (s.p.)

Pourtant, comme le fait justement remarquer O'Brien (2022), la TA est une technologie imparfaite (p. 105) et les utilisatrices ne sont bien souvent pas conscientes « des limites de ces outils, de leurs lacunes, des imperfections qui rendent l'intervention humaine indispensable à tous les stades de leur utilisation » (Kadiu, 2016, p. 218). Est-il besoin de rappeler qu'à ce jour, aucun outil technologique ne peut remplacer totalement une traductrice humaine et produire un texte de qualité publiable (Deneufbourg, 2020 ; Loock et Léchaugette, 2021 ; Vieira et al., 2019 ; Zimina et Gledhill, 2021²²¹) ? Pour ces raisons, nous estimons, tout comme Kenny

²²¹ « [S]i les plateformes de TA actuellement disponibles fournissent des solutions de traduction adéquates, elles ne sont pas en mesure de prendre en charge la traduction de discours "rédactionnels" ou "promotionnels" sans intervention humaine » (Zimina et Gledhill, 2021, p. 79).

(2022)²²², que la TA neuronale devrait être vue comme un outil d'aide parmi d'autres, comme une technologie qui vient s'ajouter à toute la palette d'outils d'aide à la traduction qui permet à la traductrice d'être plus rapide et efficace, et ainsi de faire face à la masse toujours plus grande et urgente de documents à traduire : « [T]he productivity gains brought about by machine translation (MT) can help translators meet ever-tighter deadlines and respond to pressing demands for publishing content simultaneously in different languages » (Martikainen, 2022a, s.p.). Nos propos trouvent écho chez Loock (2020c) : « MT output can be very good and serve as inspiration for students, meaning MT should be considered a translation tool to help them in the same way translation memories and specialised electronic corpora do » (p. 164), ainsi que chez Alonso et Vieira (2020) :

[D]espite its potentially negative consequences, technology can also have a positive impact on translation as a practice and profession. [...] While MT can be perceived to threaten translators' standing and professional status, it is in a relationship of complementarity with translators and not in one of exclusion. (p. 400)

En dépit des progrès technologiques et des allégations infondées de parité humaine en TA (Grass, 2022²²³ ; Loock et Léchauguette, 2021²²⁴ ; Poibeau, 2022²²⁵), l'humain reste incontestablement un maillon essentiel de la chaîne de traduction et nous devons non seulement en prendre conscience, mais aussi en faire prendre conscience. Ce dernier défi s'adresse donc tout particulièrement au corps enseignant : « [I]t is incumbent on translation trainers to ensure that students are made aware of their usefulness in order to maximise their agency as translators, and to fulfil industry employment needs » (Moorkens, 2018, p. 375).

²²² « NMT is just the latest in a line of technologies designed to automate translation, albeit one that has risen to prominence remarkably quickly » (Kenny, 2022, p. 46).

²²³ « La question de la parité entre la traduction automatique (TA) neuronale et la traduction humaine (TH) a fait couler beaucoup d'encre virtuelle. Si elle peut être atteinte dans certains cas, sur des textes ne présentant que peu d'ambiguïtés, elle est pour l'instant totalement impossible pour des textes complexes » (Grass, 2022, p. 11).

²²⁴ « Contrary to popular (and student) belief, most of the time MT output is never ready-for-use and often requires post-editing (PE), with both accuracy and fluency errors [...] MT output is rarely perfect and systematically requires checking » (Loock et Léchauguette, 2021, p. 205 et p. 207).

²²⁵ « [I]t is clear that performance varies a lot from one domain to the other, or from one type of text to the other. MT is not a solved task and has not reached human parity between any language pair (although some results obtained in evaluation conferences on specific data from specific domains may suggest the opposite). [...] A quick look at current systems clearly shows that we are still far from any "human parity" » (Poibeau, 2022, p. 6018 et p. 6019).

8.4 Quel avenir pour la traduction professionnelle ?

À tous de savoir s'adapter, évoluer, muter vers ce que seront les nouveaux enjeux de la communication multilingue et multiculturelle à l'avenir.

ÉLISABETH LAVAUT-OLLEON (2018)

Sommes-nous en train de scier la branche sur laquelle nous sommes assises ?

Cette question, O'Brien la posait déjà en 2012, quoique sous une autre forme : « By repairing MT segments and letting the machine learn about those repairs, are we progressively making ourselves redundant? Or, again, are we freeing ourselves up for those harder, subtler, more complex problems? » (p. 117). Et cette question fait inévitablement place à toute une série d'autres interrogations : Quelle est aujourd'hui la place de *la traductrice* humaine, « se voit-elle libéré[e] ou au contraire aliéné[e] par la machine ? » (Poncharal et Birkan-Berz, 2022, s.p.). La TA neuronale fait-elle de *la traductrice* une traductrice augmentée ou une traductrice diminuée ? (*ibid.*). Les outils de TA représentent-ils une réelle menace pour la profession ? Comment repenser le rôle de *la traductrice* ? Quelle est la valeur ajoutée de l'humain par rapport à la machine ? Quel rôle pour l'humain à l'avenir ? Comme le recommande Poibeau (2022), « [t]he minimum is at least to consider these questions seriously and not just put forward the replacement of humans as an unquestionable desirable advantage of technology (p. 6022) ». Et c'est ce que nous allons nous atteler à faire dans la dernière partie de ce chapitre.

8.4.1 Les outils de TA : une réelle menace pour la profession ?

[M]algré le bouleversement des fondements épistémologiques de la traduction, des conditions de l'exercice de la profession de traducteur et de son statut social dû à l'essor des outils de l'intelligence artificielle, la traduction humaine n'est pas pour autant vouée à disparaître.

MARGARET FLÖTER-DURR (2022)

Parmi les différentes protagonistes du monde de la traduction, certaines se montrent relativement optimistes concernant le futur de la profession. *Elles* considèrent, en effet, que

l'avenir des traductrices n'est pas si sombre dès lors que les progrès en matière de TA par apprentissage profond permettent à la traduction professionnelle de répondre aux besoins de notre société libérale et capitaliste en contribuant à l'augmentation de la productivité des traductrices et en permettant de satisfaire à la demande en traduction. En outre, le nombre de personnes travaillant dans le secteur de la traduction n'a jamais été aussi élevé qu'aujourd'hui (Moorkens, 2022b, p. 24). Le marché des services langagiers se porte très bien et les prévisions pour 2023-2027 sont tout aussi bonnes comme le montre la Figure 84. Ce secteur continue ainsi à se développer sur le plan international. D'après le rapport Nimdzi 2023, le marché mondial était estimé en 2022 à 64,7 milliards de dollars et devrait atteindre 69,3 milliards de dollars en 2023 (Hickey, 2023).

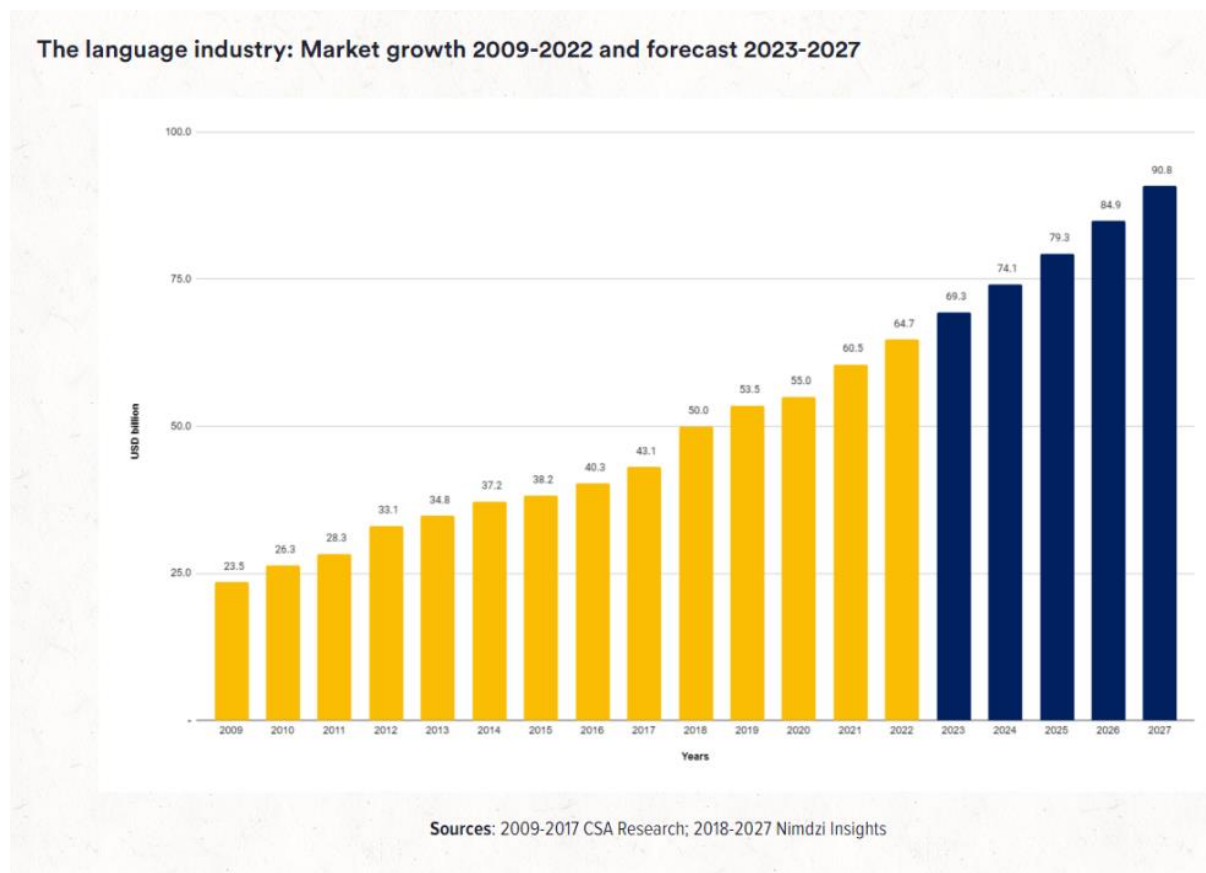


Figure 84 : Rapport Nimdzi 2023 (Hickey, 2023)

O'Brien (2012), par exemple, prévoyait déjà que l'augmentation des interactions humain-machine, en l'occurrence avec la TA statistique à l'époque, aurait des conséquences positives : « By interacting more with computer-supported translation tools, translators will enable more translation to occur, into more languages and, hopefully, in language directions that do not

normally gain attention through lack of resources » (p. 119). Et c'est toujours le cas à la suite du changement de paradigme de TA : « Evidence suggests that only a small portion of content produced globally is translated (Language Industry Survey 2019) but as MT becomes more successful, translation volumes are increasing globally » (O'Brien et Ehrensberger-Dow, 2020, p. 160). Mion (2020) estime à cet égard que « les nouvelles technologies n'ont pas volé le travail des humains. Dans la plupart des cas, elles l'ont modifié », idem pour Moorkens (2022b, p. 24) qui reconnaît que les systèmes de TA neuronale en ligne, devenus très performants et accessibles à toutes, n'ont pas tué les métiers de la traduction, ils les ont transformés. Ainsi, Dirand et Rossi (2019) estiment que les progrès réalisés en TA « ne condamnent pas le traducteur humain, [...] ils permettent plutôt d'envisager des interactions homme-machine plus nombreuses et plus fluides (sorties de TA intégrées à un outil de TAO, TA interactive s'ajustant à mesure que le traducteur écrit) » (p. 76).

8.4.1.1 L'humain, l'humain et encore l'humain

Fondée à l'origine sur le fonctionnement du cerveau biologique (Poibeau, 2019), la TA basée sur les réseaux de neurones artificiels est paradoxalement tributaire de l'activité du cerveau humain à plusieurs niveaux. Prenons par exemple les énormes bases de données d'entraînement sur lesquelles reposent ces systèmes et qui « ne sont autres que les travaux de traducteurs humains » (Dirand et Rossi, 2019, p. 72) : « An important bottleneck to neural MT is the availability and quality of multilingual data [and] it should be noted that high-quality data require human effort and expertise » (Alonso et Vieira, 2020, p. 395) ; « NMT still relies on human translations or at least translations validated by humans as training data » (Kenny, 2022, p. 46). Ainsi, un moteur de TA « ne peut fonctionner sans exploiter des données humaines » (Poncharal et Birkan-Berz, 2022, s.p.).

[T]he success of some forms of translation technology, such as statistical MT and neural MT, is in fact heavily dependent on their having access to large volumes of training data made up of high-quality human translations. So, far from replacing translators, current state-of-the-art technologies are actually dependent on them. (Bowker, 2019, p. 465)

Dirand et Rossi (2019, p. 76) regrettent que l'on mette bien trop souvent en avant les progrès de la TA plutôt que les nombreuses interactions humain-machine dont découlent d'ailleurs ces progrès. Après tout, c'est l'humain qui sélectionne les corpus spécialisés utilisés pour l'entraînement des outils, ainsi que les ressources linguistiques et terminologiques à intégrer. Et pourtant tout effort humain est invisibilisé, les outils de TA donnant l'impression que le processus de traduction est complètement automatisé et ne requiert aucune intervention humaine.

[N]ot only does free online machine translation imply that translation is "an agentless, automatic function that can be realized in no time at all" (Cronin 2012, 47) thus obscuring the human labour that produces the translated and other data on which SMT is based; systems like Google Translate also obscure the labour of the computer scientists who build SMT systems, and can give the impression that there is nothing to SMT. (Kenny et Doherty 2014, p. 288)

En outre, les moteurs de TA sont appelés « des traducteurs », ce qui est loin d'être neutre ; Rossi (2019b) parle d'une personnification de l'outil ayant un effet subliminal sur les utilisatrices.

Pourtant, l'humain est la pierre angulaire de la technologie neuronale :

Humans play an important role for MT, especially in operational contexts, for example by providing an expertise in the preparation of the data, establishing the terminology of a domain and correcting MT output during the post-edition phase. All this should be valued and the goal of NMT should not be to replace humans in the first place. (Poibeau, 2022, p. 6022)

À bien y réfléchir, l'humain est effectivement omniprésent et indispensable à chaque stade du processus : de la conception à l'évaluation des performances de l'outil, en passant par la sélection et la préparation des corpus d'entraînement, la prédiction et bien évidemment la post-édition : « MT is developed, fed, evaluated and trained by humans, who, in turn, can use MT technology to increase productivity » (Alonso et Vieira, 2020, p. 394).

Ainsi, comment la TA pourrait-elle se substituer à l'humain alors qu'elle en dépend tant ?

8.4.2 Traductrice augmentée ou traductrice diminuée ?

Dans une volonté d'optimiser le travail de la traductrice, les frontières entre TH et TA deviennent floues, la TA étant de plus en plus intégrée aux outils de TAO (Castilho et al., 2018)²²⁶, ce qui laisse « présager une tout autre manière d'envisager la traduction automatique, plus proche de la traduction humaine assistée par ordinateur que de la post-édition pure » (Hansen, 2022, p. 22). Vieira (2019) évoque même un changement de paradigmes qui permet de replacer l'humain au cœur du processus de traduction en passant

²²⁶ « We contend that the boundaries between HT and MT (and, to some extent, also those between translation as understood in academia versus in the industry) are increasingly blurring; this is apparent, in particular, in software and web localisation as well as in a wide range of technical and specialised domains, where MT (often supported by PE) is becoming widely used alongside the now commonplace computer-assisted translation (CAT) software such as translation memories, especially for projects involving major languages as source or target, or for language pairs with substantial commercial interest » (Castilho et al., 2018, p. 11).

de la TA assistée par l'humain (*human-assisted MT*) à la traduction humaine assistée par ordinateur (*machine-assisted human translation*) :

[M]uch has been done lately to improve the use of MT in professional translation. MT is now available in most computer-assisted translation (CAT) tools. As implied by their name, these tools changed the focus from human-assisted MT to machine- (or computer-) assisted human translation. Unlike the paradigm from early MT research, machine-assisted human translation puts humans at the centre of translation production. The incorporation of MT into CAT tools and, more recently, a higher degree of integration of these tools' different features – such as translation memories, terminological resources and MT itself – all contribute to an understanding of MT as a resource that helps human translators improve their productivity. (pp. 319-320)

Pour Martikainen (2022a), cette intégration de la TA à l'interface TAO constitue de nos jours le scénario le plus confortable pour la traductrice : « The ideal scenario for translator well-being would seem to be integration of MT in the usual translation workflow, wherein MT is just another potentially useful tool available for the professional translator through the CAT interface » (s.p.). Elle montre l'importance d'exercer les étudiantes à ce type de tâches de TH assistée par ordinateur :

Students are [...] trained to put in practice a workflow inspired by best practice, i.e. integrating different MT engines into the CAT environment of their choice so as to maintain the benefits offered by CAT tools, particularly for ensuring coherence, and limit MT interference. (ibid., s.p.).

Aujourd'hui, grâce à toute une série de nouveaux outils dont les outils de TA, la traductrice peut évoluer « dans un environnement de travail révolutionnaire » (Mion, 2020, s.p.) qui lui permet de pratiquer un nouveau type de traduction, la « traduction augmentée » (*augmented translation*, voir DePalma et Lommel, 2017) :

Dans cet environnement, il·elles ont accès à des outils qui améliorent leur efficacité et leur productivité et accroissent leurs compétences. À l'aide de dictionnaires numériques, d'Internet, de bases terminologiques, de corpus, de la recherche vocale, de logiciels de dictée vocale, de mémoires de traductions, de correcteurs automatique [sic] et de bien d'autres outils et fonctions, les professionnel·le·s des services linguistiques sont amenés à effectuer un nouveau genre de traduction : la traduction augmentée. (Mion, 2020, s.p.)

Avec ce concept de traduction augmentée, l'humain se voit réattribuer une place centrale, comme l'illustre la Figure 85, lui permettant de reprendre goût à la traduction. Un tel environnement hyper technologique lui permet entre autres d'échapper aux tâches répétitives

et épuisantes de son métier, et de se concentrer davantage sur d'autres aspects, notamment sur son potentiel créatif, ce que DePalma et Lommel (2017) avaient déjà envisagé :

They work in a technology-rich environment that automatically processes many of the low value tasks that consume an inordinate amount of their time and energy. It brings relevant information to their attention when needed. This computing power will help language professionals be more consistent, more responsive, and more productive, all the while allowing them to focus on the interesting parts of their jobs rather than on "translating like machines". (s.p.)

Augmented Translation Ties Humans and Machines Together

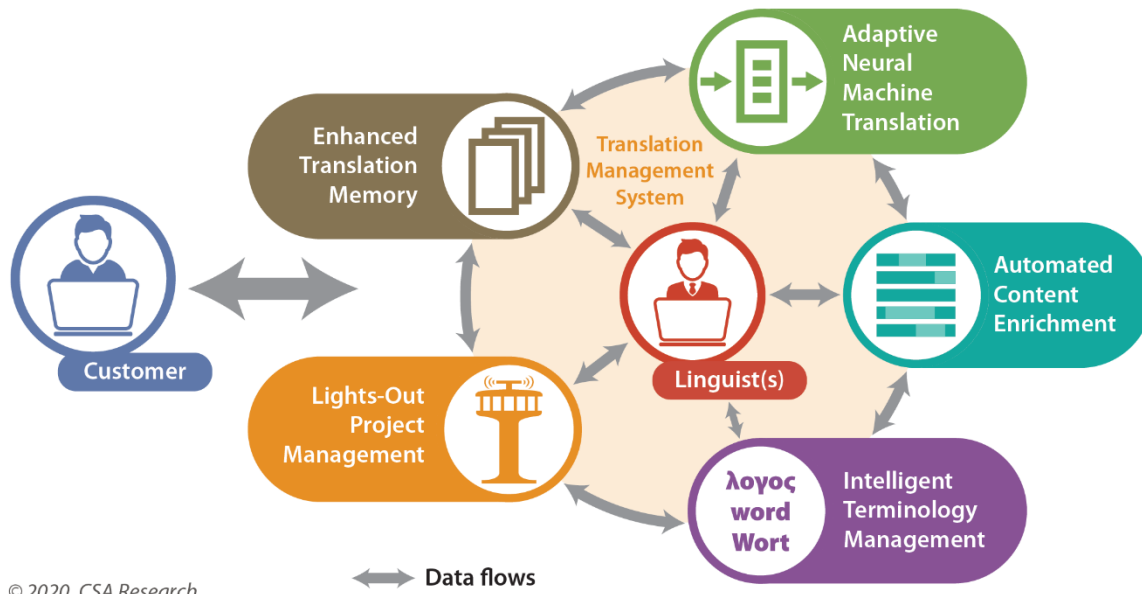


Figure 85 : La traduction humaine augmentée (Lommel, 2020)

8.4.2.1 TA interactive et TA adaptative

Dans ce nouvel environnement, parmi les dernières innovations mises à disposition de la traductrice figurent des outils combinant PE de TA interactive (Peris et al., 2017)²²⁷ et de TA

²²⁷ « Peris et al. (2017) have proposed what they call interactive neural machine translation, a special kind of interactive machine translation or interactive translation prediction, which had customarily been performed so far using SMT instead of NMT. In these prediction–completion translation workflows, the system suggests possible continuations which may be accepted [...] or ignored by the professional translators as they type the target text » (Forcada, 2017, p. 296).

adaptative (par ex. le logiciel Lilt²²⁸). Ces deux concepts sont parfois utilisés de manière synonyme alors qu'il y a bel et bien une distinction à faire :

An interactive system tries to autocomplete the text the user is going to type; it either predicts the text the user is going to type or changes the MT suggestion on the basis of what is typed, whereas an adaptive system is an MT system that learns from corrections on the fly and is continuously trained. (Daems et Macken, 2019, p. 118)

Sur le continuum de l'intervention humaine en PE (Figure 86) proposé par Vieira (2019), la PE de TA interactive et de TA adaptative se situe logiquement à l'extrême opposé de la PE automatique.

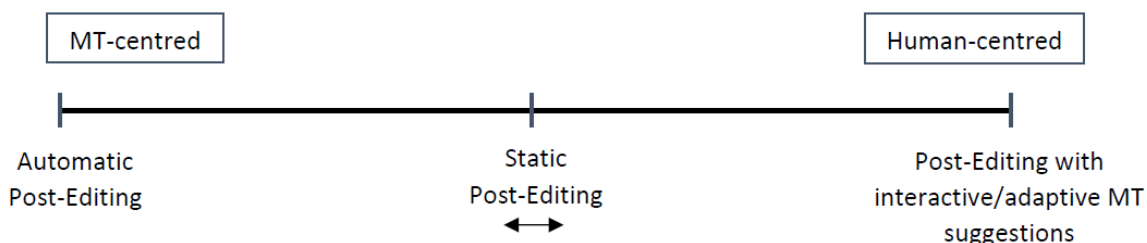


Figure 86 : Continuum de l'intervention humaine en PE (Vieira, 2019, p. 328)

Comment fonctionnent concrètement ces outils ? Pour y voir plus clair, référons-nous à l'explication détaillée donnée par Martikainen (2022a, s.p.) concernant l'outil Lilt : ce logiciel s'adapte à deux stades différents ; comme la plupart des moteurs de TA, Lilt s'adapte à l'utilisatrice et s'améliore au fil du temps. Toutefois, contrairement aux moteurs de TA statique, les systèmes de TA interactive (ou « TA adaptative en temps réel ») s'adaptent également pendant le processus de traduction. Le logiciel tient compte immédiatement de chaque mot traduit afin d'adapter la sortie proposée et de prédire la formulation recherchée par l'utilisatrice. De cette manière, si la proposition initiale de la machine n'est pas bonne, une nouvelle sortie est proposée après la saisie manuelle de quelques mots de la traduction souhaitée.

²²⁸ <https://lilt.com>

« This technology – currently found in Lilt and SDL BeGlobal – learns from translators on the fly. It adapts to the content they work on, automatically learning terminology and style. It remembers what linguists have previously translated at the sub-segment level, and goes beyond translation memory to help translate text it has never seen before in a way that is consistent with how the individual professional works. Rather than post-editing MT output of dubious quality, linguists see the results as suggestions they can choose to use or not. The more they use the system, the better these suggestions will become » (DePalma et Lommel, 2017, s.p.).

Présenté comme une solution alternative à la « post-édition de TA statique », ce type d'outils est vu comme un moyen de pallier certains des écueils en TA statique tels que le peu de marge de manœuvre laissée à *la traductrice*, comme le souligne Martikainen (2022a) : « Adaptive MT has been advertised as "Machine Translation for Human Translators" [Denkowski et al., 2015], a way to recenter the translation process on the human and foster more genuine interaction with the machine » (s.p.), mais aussi comme une possibilité de tirer plus intelligemment parti des sorties de TA : « Adaptive and interactive MT can provide translators with more intelligent ways of leveraging automatic suggestions and of avoiding or speeding up repetitive work » (Alonso et Vieira, 2020, p. 400).

En revalorisant le rôle de l'humain dans le processus, ces nouvelles fonctionnalités peuvent, selon nous, permettre à *la traductrice* de retrouver un certain plaisir à traduire, entre autres car l'effet fantôme de la TA est atténué, comme le font remarquer Rossi and Chevrot (2019) dans leur étude menée à la DGT : « Within this intricate interweaving of technology and human intervention, machine translation, which appears only in the form of easily dismissed segments, may well be of minor importance and its impact limited » (p. 180). Effectivement, *la traductrice* ne se voit plus imposer une proposition de traduction :

Contrarily to most existing systems, MT output is not presented to the user in the space devoted to the target text. Instead, for each segment, the system shows the source text, followed by a blank space where the user inserts or types the translation, and only under this blank space, the suggested MT output that adapts to the target text being produced. The user interface is simple and intuitive, and allows for easy insertion of translation memory matches, MT suggestions, and glossary terms. (Martikainen, 2022a, s.p.)

De plus, il ressort de l'expérience menée par Martikainen (*ibid.*) que les étudiantes de master en traduction expriment une nette préférence pour la PE de TA neuronale interactive et adaptative par rapport à la PE de TA neuronale statique : « The experiment conducted with master's students suggests that interactive/adaptive MT indeed does have the potential to offer a better user experience than post-editing static MT » (s.p.).

8.4.3 Quel rôle pour l'humain à l'avenir ?

Selon Lavault-Olléon (2018), la survie de la profession va fortement dépendre de l'appropriation de la TA et c'est en mettant en exergue le rôle déterminant des spécificités humaines, telles que la médiation interculturelle, la sensibilité linguistique, mais aussi les compétences en gestion de projets multilingues, que les *traductrices* feront valoir leur valeur ajoutée. En outre, on voit apparaître aujourd'hui des nouveaux services tels que l'hyper-localisation ou la transcréation.

Vieira (2019) souligne, à cet égard, l'influence positive que peut avoir le champ de la traductologie :

Given commercial MT developers' decades- old tendency to oversell the results of their systems [...] translation studies have an important role to play in ensuring a sound understanding of MT's place in professional translation. This is likely to become even more critical as MT technology continues to evolve and gain wider public visibility. (p. 326)

Pour ne pas céder à la panique, Vieira (2020) propose que l'ensemble des protagonistes du secteur s'asseyent autour de la table pour repenser le rôle de la traductrice :

Open dialogue among translation industry stakeholders and the exploration of business models that integrate rather than fragment the role of translators across domains are considered here to be more productive responses to advances in technology than giving in to automation anxiety. (p. 17)

Bernardini et al. (2020, p. 302) estiment que la mission de la traductrice ne peut se limiter à une intervention en bout de course (en PE ou pour les tâches d'évaluation par exemple). Forte de ses compétences linguistiques, la traductrice doit être partie prenante d'une collaboration interdisciplinaire avec les conceptrices et prestataires de services, en vue d'améliorer les systèmes d'intelligence artificielle : « New roles involve the appreciation, understanding and critical assessment of AI technologies and the ability to post-edit, approve and vouchsafe the correctness of translations, all firmly rooted in human expertise in languages, cultures and the science of translation » (*ibid.*, p. 302).

Breyel et Grass (2021) sont, elles, convaincues que l'enjeu dans le domaine de l'intelligence artificielle est de remplacer des humains conscients par des algorithmes non conscients. Aussi, elles estiment que pour pouvoir continuer à exercer, la traductrice doit être capable d' « anticiper, de favoriser la créativité, la qualité de communication et le travail en équipe. Les formations des traducteurs doivent aussi se diversifier en incluant les techniques d'écriture, la rédaction technique » (p. 105), etc.

Pour notre part, notre opinion concorde en tout point avec celle de Rossi (2019a) et de Kenny et Doherty (2014)²²⁹ qui jugent que les formatrices en traduction ont en quelque sorte le devoir moral d'assurer la pérennité de la profession : « Il appartient aux traductrices et traducteurs

²²⁹ « [A]s academics who train translators we have a vested interest in, but also an ethical commitment to ensuring the sustainability of the profession. Otherwise we are educating students in our own interest but with little regard for theirs. [...] Such sustainability must partly reside in translators' ability to evolve and to adopt whatever tools are useful and currently available to them so as to remain relevant and competitive » (Kenny et Doherty, 2014, p. 290).

de demain de négociier et de construire leur rôle dans ce paysage mouvant, et aux formatrices et formateurs de les y encourager » (Rossi, 2019a, p. 103). L'une des clés, selon nous, pour y parvenir est de développer les précieuses facultés de renouvellement et d'adaptation des étudiantes, quelle que soit la technologie ou l'outil (*DeepL Traduction, Google Traduction, ChatGPT d'Open AI, DeepL Write, etc.*) qu'il conviendra de prendre en main pour parvenir à trouver sa place sur un marché hautement concurrentiel.

Understanding translation technologies and the appropriate evaluation techniques is critical to the successful integration of these technologies in the language services industry of today, where the lines between human and machine have become increasingly blurred and adaptability to change has become a key asset that can ultimately mean success or failure in a competitive landscape. (Moorkens et al., 2018a, p. 1)

9 Conclusions et perspectives

We will have to negotiate the perpetual challenge of what it means to be human in the translation age, knowing that the only certainty in our digital world is that there is none.

MICHAEL CRONIN (2013)

Alors que le secteur de la traduction professionnelle est en pleine mutation technologique et que l'actualisation des programmes de formation en traduction constitue un enjeu déterminant tant pour l'avenir de l'enseignement de la traduction que pour le futur de la profession, nous nous sommes intéressée dans cette thèse à la post-édition de TA en contexte d'apprentissage.

Notre principale intention a été d'explorer les potentielles différences, en matière de qualité, entre les produits de la post-édition et de la traduction humaine. Pour ce faire, nous avons confronté ces deux processus de traduction dans des conditions expérimentales impliquant des étudiantes en traduction. Nous avons eu également pour objectif d'évaluer les connaissances et les perceptions des étudiantes en traduction à l'égard de la TA et de la PE. Nous avons enfin souhaité nourrir une réflexion plus large sur l'intégration de la TA et de la PE aux cursus de formation en traduction, ainsi que sur les défis de l'apprentissage de la PE.

Dans cette dernière partie, nous faisons la synthèse de nos travaux et présentons le récapitulatif des résultats en commençant par les résultats empiriques du prétest et du test (étude du produit et étude du processus) avant de passer aux conclusions tirées des mesures linguistiques automatiques et de l'analyse qualitative des erreurs contenues dans nos corpus. Puis nous examinons les principales tendances qui se sont dégagées des différents sondages soumis aux étudiantes en matière de connaissances et de perceptions sur la TA et sur la PE. Nous résumons par la suite notre discussion sur l'introduction de la TA et de la PE dans les programmes de formation en traduction, ainsi que notre point de vue concernant l'avenir de la traduction professionnelle. Nous soulignons enfin les principales limites de cette étude et proposons pour terminer des pistes de recherches futures.

9.1 Récapitulatif des résultats

9.1.1 Résultats empiriques

9.1.1.1 Étude du produit

Question A – Quels sont les effets de la post-édition en contexte d'apprentissage sur la qualité d'un texte cible ?

Afin de répondre à notre première question de recherche, nous avons mené deux expériences contrôlées en contexte d'apprentissage pour la combinaison de langues anglais-français. Dans la première expérience (prétest) menée en 2018 avec les étudiantes en traduction de 3^e bachelier, nous avons comparé les produits de la TH, de la PE de TA statistique (*BingTranslator*) et de la PE de TA neuronale (*DeepL*). La deuxième expérience (test) a été menée en 2021 avec les étudiantes en traduction de 2^e master et nous avons comparé cette fois les produits de la TH et de la PE de TA neuronale générée par deux moteurs différents : *Google Traduction* et *DeepL*. Rappelons que la cohorte d'étudiantes ayant pris part au test n'est pas la même que celle qui a participé au prétest.

Grâce à l'évaluation humaine des productions et aux résultats de l'analyse quantitative, nous pouvons conclure que, comparativement à la TH, la PE de TA, qu'elle soit statistique ou neuronale, n'a pas eu d'influence négative sur la qualité finale des productions. En d'autres termes, la PE n'a pas généré de productions de moindre qualité globale par rapport à la traduction humaine.

Cependant, les résultats des deux expériences ont mis en évidence une relation inverse entre le niveau d'une étudiante en TH et la qualité de sa production en PE de TA neuronale, ce que nous désignons par « effet nivelant ». Nous pouvons conclure à un effet nivelant de la PE de TA neuronale sur la qualité finale. Plus l'étudiante a de faibles compétences en TH, plus *elle* tire avantage de la PE et, à l'inverse, plus *elle* a des compétences élevées en TH, plus la PE entraîne une baisse de qualité de sa production finale.

9.1.1.1.1 Prétest

Analyse globale

Dans le prétest, la PE de TA neuronale a permis d'obtenir une qualité globale supérieure à celle des PE de TA statistique et des TH. L'analyse statistique a démontré qu'en PE de TA neuronale, les étudiantes ont commis significativement moins d'erreurs d'acceptabilité par rapport à la PE de TA statistique et à la TH et *elles* ont obtenu, en moyenne, de meilleures notes par rapport aux deux autres modes de traduction, quel que soit le TS. Nous n'avons cependant pas pu

inférer de dépendance entre le mode de traduction et la distribution des erreurs d'adéquation. Ces résultats confirment donc que la qualité des textes post-édités dépend, entre autres, de la technologie de traduction automatique employée (paradigme statistique ou paradigme neuronal).

Analyse détaillée

L'analyse détaillée des catégories d'erreurs du prétest a indiqué que la distribution d'erreurs en PE n'est pas comparable à celle observée en TH. Les textes post-édités (TA statistique et TA neuronale) contiennent plus de calques et moins d'erreurs d'orthographe que les TH. Les PE de TA neuronale contiennent moins de glissements de sens par rapport à la TH et moins d'erreurs de grammaire et de syntaxe par rapport aux PE de TA statistique et aux TH. Enfin, il y avait davantage de problèmes de cohérence dans les PE de TA statistique que dans les PE de TA neuronale et dans les TH.

9.1.1.1.2 Test

Analyse globale

Concernant le test, il apparaît que les textes post-édités (*Google Traduction* et *DeepL*) sont de qualité globale comparable, voire de meilleure qualité, que les textes traduits humainement, ce qui est particulièrement le cas en PE de *DeepL*. Les résultats statistiques ont indiqué que les étudiantes ont commis significativement moins d'erreurs d'acceptabilité lorsqu'elles ont post-édité à partir de *DeepL* par rapport à la PE de *Google Traduction* et à la TH. Nous n'avons cependant pas pu conclure à une relation de dépendance, ni entre le mode de traduction et les erreurs d'adéquation, ni entre le mode de traduction et les notes attribuées aux productions des étudiantes.

Analyse détaillée

L'analyse détaillée des catégories d'erreurs du test a montré que la distribution d'erreurs en PE de TA neuronale n'est pas comparable à celle observée en TH. Les productions humaines contiennent davantage d'erreurs de grammaire et de syntaxe et moins de calques que les textes post-édités (*DeepL* et *Google Traduction*). De plus, les problèmes de typographie, de ponctuation et d'erreurs de vocabulaire sont moins présents dans les PE de *DeepL* par rapport aux PE de *Google Traduction* et aux TH. Ces résultats ont confirmé que la qualité d'un texte post-édité en contexte d'apprentissage dépend, entre autres choses, du moteur de TA neuronale employé (*Google Traduction* ou *DeepL*).

9.1.1.2 Étude du processus

Dans les deux expériences, les résultats sommaires de l'étude du processus ont révélé un gain de productivité en post-édition par rapport à la traduction humaine, mais une absence de

corrélation entre la durée de traduction/post-édition et la qualité des productions. Nous avons toutefois constaté que plus l'étudiante a passé du temps à post-éditer la TA statistique, meilleure est la qualité finale de sa production (faible corrélation positive).

9.1.1.2.1 Prétest

Nous avons noté une réduction notable de l'effort de PE en TA neuronale par rapport à la PE de TA statistique, ainsi qu'une faible corrélation positive entre le taux d'effort de PE et la qualité des textes post-édités, qui est davantage marquée en PE de TA statistique.

9.1.1.2.2 Test

Nous avons observé une corrélation entre l'effort moyen en PE de *DeepL* et la qualité du texte cible, autrement dit, plus l'étudiante a modifié la TA brute de *DeepL*, meilleure est la qualité finale de sa production.

9.1.2 Vous avez dit *post-editeuse* ?

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

9.1.2.1 Mesures linguistiques automatiques

9.1.2.1.1 Richesse lexicale

Malgré des valeurs très proches, nous avons constaté que les TH sont très légèrement plus denses lexicalement que les TA brutes et que les PE. Nous avons constaté également que les productions humaines présentaient le degré de diversité lexicale le plus élevé, ce qui nous a amenée à conclure à une plus grande richesse lexicale en TH par rapport à la TA brute et à la PE de TA, qu'elle soit statistique ou neuronale. De plus, parmi l'ensemble des productions, les PE de *DeepL* se révèlent les moins riches lexicalement comparativement aux textes traduits humainement, aux PE de *Google Traduction*, ainsi qu'aux PE de TA statistique. Notons que cette richesse lexicale moindre en PE de *DeepL* n'a manifestement pas nui à la qualité finale puisque ces productions ont obtenu des résultats comparables, voire de meilleurs résultats, que les autres productions.

9.1.2.1.2 Longueur moyenne des phrases

Les résultats rendent compte d'une longueur moyenne des phrases plus élevée en langue traduite par rapport aux textes sources. Les valeurs calculées en PE sont très proches de celles

obtenues en TH, ce qui laisse entendre que les versions post-éditées se rapprochent davantage des TH que des TS. Concernant le foisonnement, les PE de TA neuronale présentent des taux de foisonnement comparables à ceux des TH, tandis que les PE de TA statistique foisonnent moins que les TH.

9.1.2.1.3 Équivalence syntaxique

Les résultats des calculs d'équivalence syntaxique révèlent la même tendance : les PE présentent une plus grande proximité syntaxique avec les textes sources que les TH. Par ailleurs, les textes post-édités qui s'éloignent le plus des textes sources sont les PE de *DeepL*.

En conclusion, une partie des mesures linguistiques automatiques (richesse lexicale et équivalence syntaxique) a confirmé l'existence d'un *post-editeuse* puisque les textes post-édités se sont révélés moins riches lexicalement que les TH (tout particulièrement en PE de *DeepL*) et plus proches syntaxiquement de la langue source que les TH.

9.1.2.2 Analyse qualitative des erreurs

L'analyse qualitative des erreurs contenues dans les textes post-édités et dans les productions humaines nous a permis d'illustrer, à l'aide d'exemples tirés de nos corpus, certains des effets tant positifs que négatifs du recours à la (PE de) TA neuronale par rapport à la TH. Il s'avère, en outre, que la majorité des erreurs relevées en PE étaient déjà présentes dans la TA brute et n'ont pas été corrigées par les étudiantes.

Ainsi, l'analyse qualitative des erreurs dans nos corpus a également permis de mettre au jour des caractéristiques propres au *post-editeuse* : maîtrise des règles d'usage du français (grammaire et orthographe), marques de littéralité, effacement d'informations, rupture de cohésion textuelle, rupture de cohérence, irrégularité terminologique, perte de richesse linguistique.

9.1.3 Étude des connaissances et perceptions

Question B – Quelles sont les connaissances et les perceptions des étudiantes en traduction sur la TA et sur la PE ?

L'analyse des réponses aux sondages soumis aux étudiantes de 3^e bachelier, de 1^{er} et 2^e master en 2018 et en 2021, nous a permis de dégager les tendances suivantes :

- le recours à la TA et à la PE s'est intensifié entre 2018 et 2021 aussi bien chez les étudiantes de 1^{er} et 2^e master que chez les étudiantes de 3^e bachelier. Leur outil de TA en ligne de prédilection est *DeepL*, suivi de *Google Traduction* et de *Reverso* ;

- la majorité des étudiantes a perçu un gain de temps en PE par rapport à la TH. De plus, certaines étudiantes attestent l'utilisation de la TA comme moyen d'accéder au sens de certains passages du texte source ;
- la plupart des étudiantes interrogées considèrent la qualité d'une traduction générée par un moteur de TA comme « satisfaisante » ou « bonne ». De plus, nous avons observé une amélioration de la satisfaction des étudiantes à l'égard de la qualité de la TA avec le temps. Selon nous, cette amélioration des perceptions est due aux progrès réalisés en TA neuronale, ainsi qu'à l'intégration progressive de cette technologie dans le cursus de formation en traduction de l'Université de Liège ;
- la majeure partie des étudiantes estiment que la qualité en traduction humaine reste plus élevée que la qualité d'une TA brute. Ces étudiantes s'attendent à une évolution de la qualité en TA à l'avenir, même si peu d'entre elles estiment que la qualité d'une TA sans intervention humaine dépassera un jour la qualité d'une TH ;
- les apprenantes en traduction sont bien plus familières de la post-édition en 2021 qu'en 2018, ce qui résulte indéniablement de l'introduction d'exercices de PE dans les cours de traduction de l'Université de Liège à partir de la 3^e année d'étude, conjuguée à l'essor de cette pratique dans le monde professionnel ;
- de nombreuses étudiantes sont convaincues qu'il est possible d'obtenir en PE une qualité comparable à la TH. En outre, les étudiantes se sont montrées davantage satisfaites de la qualité de la PE en 2021 qu'en 2018, ce qui porte à croire que l'intégration de la PE dans leur formation initiale a eu un effet positif sur leurs perceptions ;
- les résultats des sondages ont montré que la très grande majorité des étudiantes est favorable à l'introduction d'une formation spécifique aux outils de TA et à la PE dans leur cursus. Les étudiantes semblent d'ailleurs conscientes du bouleversement technologique que traverse le secteur de la traduction et des enjeux qui en découlent pour la profession, et se disent prêtes à pratiquer la PE dans leur future carrière professionnelle ;
- de manière générale, les étudiantes ayant participé au prétest et au test ont fait part de perceptions positives à l'égard de la TA et de la PE, malgré leur expérience limitée dans ce domaine ;

- en 2018 comme en 2021, les étudiantes ont considéré la tâche de TH comme étant plus exigeante que les tâches de PE. En revanche, la PE de *DeepL* a été jugée comme la tâche la moins exigeante pour les deux expériences ;
- en matière de qualité du produit fini, les étudiantes participantes tiennent en relative haute estime leur production humaine, car plus d'une sur deux (54 % en 2018 et 50 % en 2021) a estimé que c'est la TH qui leur a permis d'obtenir la meilleure qualité. La PE de TA statistique a obtenu le moins bon score parmi les différents modes de traduction, les étudiantes ne considérant pas que ce mode de traduction leur a permis de produire des traductions de qualité comparable, ni à la TH ni à la PE de TA neuronale ;
- les étudiantes se sont montrées plus satisfaites par la tâche de PE de TA neuronale (*DeepL* et *Google*) que par la tâche de PE de TA statistique, probablement en raison des améliorations apportées par le paradigme neuronal. Le taux global de satisfaction en TH était comparable à celui en PE de TA neuronale ;
- enfin, en 2018, la participation au prétest a permis à 43 % des participantes (3^e bachelier) de se montrer plus enclines à utiliser la TA/PE et de développer une attitude plus favorable à l'égard de ces outils. En revanche, lors de l'expérience 2021, seules quatre participantes de 2^e master (17 %) se disent davantage enclines à faire de la TA/PE. Nous n'avons donc pas constaté d'évolution notable des perceptions en 2021.

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

Dans la dernière partie de cette thèse, nous avons souligné la pertinence de repenser les cursus de formation en traduction en fonction des évolutions technologiques de la profession. Nous avons proposé à cet égard une réflexion sur le contenu et sur les avantages d'une formation à la TA et à la PE, ainsi que sur le moment approprié pour introduire une telle formation dans les programmes d'études en traduction. Dans la lignée de Bowker et Ciro (2019), de Loock (2019 et 2020d), de O'Brien et Ehrensberger-Dow (2020) et de Rossi (2019b), nous plaidons pour une approche raisonnée des outils de TA qui cherche à faire des étudiantes des utilisatrices averties, autonomes et responsables. De plus, en nous appuyant sur notre expérience pédagogique et sur les résultats de notre étude, nous avons dressé une liste de sept grands défis auxquels, selon nous, sont confrontées les traductrices, aussi bien novices que chevronnées, en PE de TA neuronale et sur lesquels il convient de porter un point d'attention particulier dans l'apprentissage de la PE. Il s'agit de pouvoir développer la capacité de discernement, de trouver le juste équilibre entre sur-édition et sous-édition, d'assurer la cohésion et la cohérence textuelles, d'adapter la qualité aux modalités de PE, d'atténuer l'effet

fantôme de la TA, de déjouer l'illusion de fluidité en TA et de revaloriser le rôle de l'humain dans le processus de traduction.

En outre, nous avons avancé que les améliorations dans le domaine de la TA, ainsi que l'essor de la PE, et l'accroissement de la demande en traduction sont interdépendants. Nous en avons profité pour réaffirmer que, malgré les progrès indéniables de la TA neuronale, l'humain continue de jouer un rôle prépondérant et indispensable en traduction. Par ailleurs, nous avons montré qu'il existe des façons ingénieuses d'intégrer la technologie neuronale aux outils de TAO (TA interactive, TA adaptative, etc.) afin de recentrer le processus de traduction sur l'humain. Enfin, nous avons exprimé notre conviction selon laquelle l'avenir de la traduction professionnelle dépend de la capacité des traductrices à s'approprier les nouvelles technologies de la traduction et à valoriser leurs spécificités humaines. Nous partageons l'idée que les formatrices en traduction ont, à cet égard, un rôle déterminant à jouer pour assurer la pérennisation de la profession en développant les capacités d'adaptation des étudiantes, quelle que soit l'innovation technologique à appréhender (moteur de TA, chatbot, outil d'aide à la rédaction ou de reformulation), ainsi qu'en mettant l'accent sur la valeur ajoutée de la traduction humaine dans cette « civilisation de l'algorithme » (Rouvroy, 2018).

Pour conclure, nous espérons non seulement que cette étude comparative aura permis de mieux cerner les enjeux, les atouts et les défis de l'apprentissage de la PE, mais aussi que nos résultats et nos réflexions pourront contribuer à faire évoluer l'enseignement de la traduction en promouvant un usage raisonné de la TA neuronale auprès des apprenantes et formatrices en traduction.

9.2 Limites expérimentales

Nous tenons à mettre en évidence les principales limites de ces travaux. Comme dans toute recherche expérimentale en sciences humaines, nous n'avons pas pu contrôler l'ensemble des facteurs. Nous savons aussi que certains de nos choix méthodologiques ont entraîné plusieurs limites dans cette étude : le choix d'une méthode d'échantillonnage non probabiliste (échantillonnage consécutif), la non-vérification de l'authenticité des traductions humaines (uniquement dans le prétest), l'intervention de l'autrice de cette étude en tant qu'évaluatrice de la qualité des productions recueillies, et ce, même si ce fut le cas uniquement dans le cadre du prétest 2018. À cela s'ajoutent la fragilité de nos observations en matière de processus de PE (vérification statistique nécessaire), et le choix d'une évaluation humaine de la qualité en traduction qui demeure une tâche hautement complexe et empreinte de subjectivité. Une subjectivité que nous avons toutefois tenté de limiter dans la mesure de nos moyens, non seulement en travaillant avec plusieurs évaluatrices (Saldanha et O'Brien, 2014), mais aussi en employant une typologie d'erreurs commune et en prévoyant une phase de mise en commun des annotations entre évaluatrices. Nous voulons également faire remarquer qu'il est fort

probable que nos évaluateuses humaines, en tant qu'enseignantes et traductrices expérimentées, aient jugé la qualité des productions de manière plus sévère que si ces productions avaient été évaluées par d'autres profils d'évaluateuses (Fiederer et O'Brien, 2009), ce qui pourrait avoir entraîné un phénomène de « surcorrection », principalement pour la phase d'acceptabilité. Par ailleurs, bien que l'ordre dans lequel les tâches ont été effectuées variait d'une étudiante à l'autre (Figure 13), nous n'avons tenu compte à aucun moment de la potentielle influence de cet ordre sur nos résultats.

Nous tenons ensuite à rappeler que les résultats présentés dans cette thèse sont le fruit de deux recherches expérimentales menées avec des apprenantes pour la combinaison de langues anglais-français. Ces résultats sont limités aux paramètres d'expérience et ne valent que dans le cadre spécifique dans lequel ces expériences se sont tenues ; nous n'avons dès lors aucune prétention de généralisation. Nous sommes, en effet, consciente que toute modification des conditions expérimentales (population observée, choix du texte source, critères d'évaluation de la qualité, etc.) pourrait aboutir à d'autres résultats. De plus, les conditions expérimentales dans lesquelles les étudiantes ont été placées ne sont pas comparables aux conditions habituelles en traduction professionnelle. De futurs travaux de recherche sont nécessaires pour pouvoir confirmer et généraliser nos observations. Nous considérons également comme potentielle limite le fait que les trois moteurs de TA auxquels nous avons eu recours (*Bing Translator*, *Google Traduction* et *DeepL*) sont des outils génériques grand public ; on pourrait légitimement s'attendre à de meilleures performances avec des moteurs entraînés spécifiquement pour nos tâches. Enfin, soulignons une dernière limite et non des moindres : les TA brutes à partir desquelles les étudiantes ont travaillé ont été générées par ces moteurs de TA début 2018 et fin 2020. Or, puisque nous savons que les avancées technologiques dans ce domaine s'enchaînent à un rythme effréné (Breyel et Grass, 2021 ; Poibeau, 2019 ; Rossi et Carré, 2022), il est fort probable qu'à l'heure où nous écrivons ces lignes, ces moteurs ne proposent plus les mêmes sorties.

9.3 Perspectives

La présente étude ouvre plusieurs perspectives de recherches futures étant donné que tous les facettes et enjeux de la PE n'ont pas pu être traités dans les limites de cette thèse. Il serait notamment intéressant d'effectuer une analyse qualitative approfondie des changements opérés par les étudiantes dans les textes post-édités, en estimant par exemple la proportion de changements indispensables et de changements préférentiels ou non nécessaires, etc. De plus, nous n'avons pas du tout exploré, dans cette thèse, la réception des productions auprès d'un public extérieur à l'étude. L'un des objectifs pourrait être d'observer si ce public exprime

une préférence pour l'un des modes de traduction, s'il parvient à repérer la TA brute parmi les PE et les TH, etc.

D'autres études expérimentales en contextes pédagogique et professionnel pourraient également être envisagées afin de faire ressortir la pertinence d'une formation en TA et en PE. Le suivi d'une formation en TA et en PE pourrait, par exemple, constituer l'une des variables indépendantes de l'expérience. En outre, il serait nécessaire de proposer un contenu concret de formation à la PE à des fins didactiques et scientifiques. Il pourrait s'agir d'imaginer des activités pédagogiques qui facilitent l'acquisition des compétences spécifiques à la PE, telles que des scénarios d'apprentissage qui mettraient les étudiantes dans des conditions optimales pour développer de bons réflexes de PE, qui les sensibiliseraient aux erreurs récurrentes en TA et qui feraient ressortir la valeur ajoutée de l'humain.

Par ailleurs, alors que nous avons étudié essentiellement le produit de la PE, il conviendrait de se pencher davantage sur l'étude du processus de PE. Il serait pertinent, par exemple, de faire appel aux sciences cognitives pour comparer les stratégies cognitives en PE et en TH et pour évaluer les effets du processus de PE sur les processus mentaux habituels en traduction.

Enfin, nous pensons que le bouleversement que connaît le monde de la traduction ne se limite pas à l'arrivée d'outils de TA performants. En effet, il y a toute une série d'outils qui méritent également, selon nous, de faire l'objet de recherches. Nous pensons notamment aux toutes dernières innovations technologiques en matière d'IA, telles que le lancement de *ChatGPT* et autres chatbots ou de l'outil de reformulation *DeepL Write*, innovations susceptibles de venir enrichir encore la palette d'outils de la traductrice.

Bibliographie

- Albarello, L., Bourgeois, É., et Guyot, J. (2010). *Statistique descriptive: Un outil pour les praticiens chercheurs*. De Boeck Supérieur. <https://doi.org/10.3917/dbu.albar.2010.01>
- Allen, J. (2003). Post-editing. In H. Somers (Ed.), *Computers and Translation: A translator's guide* (pp. 297-317). John Benjamins Publishing Company. <https://doi.org/10.1075/btl.35.19all>
- Allignol, C. (2007). Caractérisation des fautes en fin de formation. Expérience pédagogique dans un cours de traduction spécialisée allemand-français. *The Journal of Specialised Translation*, 8, 64-84.
- Alonso, E., et Vieira, L. N. (2020). The impact of technology on the changing role of the translator in globalized translation workflows. In E. Bielsa et D. Kapsaskis (Eds.), *Routledge Handbook of Translation and Globalization* (pp. 391-405). Routledge.
- ALPAC (1966). *Language and machines. Computers in translation and linguistics*. A report by the Automatic Language Processing Advisory Committee (ALPAC), National Academy of Sciences, National Research Council.
- Álvarez-Vidal, S., Liver, A., et Badia, T. (2021). What do post-editors correct? A fine-grained analysis of SMT and NMT errors. *Revista Tradumàtica. Tecnologies de la Traducció*, 19, 131-147. <https://doi.org/10.5565/rev/tradumatica.286>
- Anceaux, F., et Sockeel, P. (2006). Mise en place d'une méthodologie expérimentale : hypothèses et variables. *Recherche en soins infirmiers*, 84, 66-83. <https://doi.org/10.3917/rsi.084.0066>
- Arrondo, F. (2012). *Fidélité inter-juges d'un instrument de mesure des symptômes pré-morbides et prodromiques de la schizophrénie. À propos d'une étude rétrospective réalisée à Nancy sur 50 dossiers*. [Thèse de doctorat en médecine, Université de Lorraine]. <https://hal.univ-lorraine.fr/hal-01733147>
- Astolfi, J., Darot, É., Ginsburger-Vogel, Y., et Toussaint, J. (2008). Chapitre 8. Erreur. In J. Astolfi, É. Darot, Y. Ginsburger-Vogel et J. Toussaint (Eds.), *Mots-clés de la didactique des sciences : Repère, définitions, bibliographies* (pp. 87-91). De Boeck Supérieur.
- Baccini, A. (2010). *Statistique Descriptive Élémentaire*. Publications de l'Institut de Mathématiques de Toulouse. <https://www.math.univ-toulouse.fr/~baccini/zpedago/asde.pdf>
- Bahdanau, D., Cho, K., et Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR) (San Diego, 7-9 mai 2015)*. <https://dx.doi.org/10.48550/arXiv.1409.0473>

- Baker, M. (1993). Corpus linguistics and translation studies : implications and applications. In M. Baker, G. Francis et E. Tognini-Bonelli (Eds.), *Text and technology. In honor of John Sinclair* (pp. 233-250). John Benjamins.
- Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and translation: Studies in language engineering in honour of Juan C. Sager* (pp. 175-186). John Benjamins.
- Balling, L. W. (2008). A brief introduction to regression designs and mixed-effects modelling by a recent convert. In S. Göpferich, A. L. Jakobsen et I. Mees (Eds.), *Looking at Eyes. Eye tracking Studies of Reading and Translation Processing* (pp. 175-192). Samfundslitteratur.
- Bally, Ch. (1951). *Traité de Stylistique française* (3^e éd.). Georg & Cie S.A./Klincksieck.
- Banerjee, S., et Lavie, A. (2005). METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization* (pp. 65-72).
- Barrault, L., Bojar, O., Costa-jussà, M. R., Federmann, C., Fishel, M., Graham, Y., Haddow, B., Huck, M., Koehn, P., Malmasi, S., Monz, C., Müller, M., Pal, S., Post, M., et Zampieri, M. (2019). Findings of the 2019 conference on machine translation (WMT19). In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers)* (pp. 1-61). Association for Computational Linguistics.
- Bayerl, P., et Paul, K. (2011). What Determines Inter-Coder Agreement in Manual Annotations? A Meta-Analytic Investigation. *Computational Linguistics*, 37, 699-725. http://dx.doi.org/10.1162/COLI_a_00074
- Beaudet, C. (2001). Clarté, lisibilité, intelligibilité des textes : Un état de la question et une proposition pédagogique. *Recherches en Rédaction Professionnelle*, 1(1), 1-19.
- Bédoire, A., Blanchais, J., Bossard, M., Coutard, M., Gilein, B., Goupil, J-L., ... Lherbier, M. (2001). *Exploiter l'évaluation C.E.2. De l'analyse des résultats à la mise en œuvre de réponses pédagogiques adaptées*. Edusarthe.
- Belam, J. (2003). Buying up to falling down: A deductive approach to teaching post-editing. In M. L. Forcada, H. Somers et A. Way (Eds.), *MT Summit IX, Third Workshop on Teaching Machine Translation* (pp. 1-10). Association for Machine Translation in the Americas.
- Bénard, M., Bordet, G., et Kübler, N. (2022). Réflexions sur la traduction automatique et l'apprentissage en langues de spécialité. *ASp*, 82, 81-98. <https://doi.org/10.4000/asp.8113>

- Bender, E. M., Gebru, T., McMillan-Major, A., et Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)* (pp. 610-623). Association for Computing Machinery.
- Bentivogli, L., Bisazza A., Cettolo M., et Federico M. (2016). Neural versus phrase-based machine translation quality: a case study. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (Austin, Texas, 1-5 novembre 2016)*. <https://doi.org/10.48550/arXiv.1608.04631>
- Bernardini, S., Bouillon, P., Ciobanu, D., van Genabith, J., Hansen-Schirra, S., O'Brien, S., Steiner, E., et Teich, E. (2020). Language service provision in the 21st century: challenges, opportunities and educational perspectives for translation studies. In S. Noorda, P. Scott et M. Vukasovic (Eds.), *Bologna Process beyond 2020: Fundamental values of the EHEA* (pp. 297-303). Bononia University Press. <https://archive-ouverte.unige.ch/unige:138544>
- Blain, F. (2013). Modèles de traduction évolutifs. Ordinateur et société [cs.CY] [Thèse de doctorat en Informatique, Le Mans Université]. <https://tel.archives-ouvertes.fr/tel-01142926/document>
- Blanchon, H., et Boitet, C. (2007). Pour l'évaluation des systèmes de TA par des méthodes externes fondées sur la tâche. *Traitement Automatique des Langues (TAL)*, 48(1), 33-65.
- Blum-Kulka, S., et Levenston, E.A. (1983). Universals of Lexical Simplification. In C. Faerch et G. Casper (Eds.), *Strategies in Inter-language Communication* (pp.119-139). Longman.
- Boers, M., et Liu, H. (2013). *Compte rendu des visites effectuées par Marion Boers et Henry Liu aux institutions européennes et aux organisations sœurs à Bruxelles, Luxembourg et Amsterdam en février et mars 2013* (Rapport – février, mars 2013). Fédération Internationale des Traducteurs (FIT). http://wa1.fit-ift.org/wp-content/uploads/2014/02/4.1-CO_rapportvisite_partenaire_FRA.doc
- Bonnard, H. (1981). *Code du français courant, Grammaire seconde, première, terminale*. Magnard.
- Bonvin, A., et Lambelet, A. (2019). Exploration empirique de la richesse lexicale : la perception humaine. *Linguistik Online*, 100(7), 65-94. <https://doi.org/10.13092/lo.100.6018>
- Bouillon, P. (2012). Analysis of existing metrics and proposal for a task-oriented metric. In *Proceedings of the 16th Annual conference of the European Association for Machine Translation. ACCEPT-Automated Community Content Editing PorTal (Trente, Italie, 28-30 mai 2012)*. European Association for Machine Translation. https://www.accept.unige.ch/Products/D9.1_Analysis_of_existing_metrics_and_proposal_of_a_task-oriented_metric.pdf
- Bouillon, P., et Clas, A. (1993). *La traductique : études et recherches de traduction par ordinateur*. Les Presses de l'Université de Montréal.

- Bourdais, A. (2021). *Traducteurs en ligne et enseignement-apprentissage de l'anglais : pratiques de litt ratie num rique en lyc e*. [Th se de doctorat en sociologie, Universit  de Lyon]. <https://theses.hal.science/tel-03610877>
- Bowker, L. (2000a). A Corpus-based Approach to Evaluating Student Translations. *The Translator*, 6(2), 183-210. Routledge.
- Bowker, L., (2020b). Fit-for-purpose translation. In M. O'Hagan (Ed.), *The Routledge Handbook of Translation and Technology* (pp. 453-468), Routledge.
- Bowker, L., et Ciro, J. B. (2019). *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*. Emerald Group Publishing.
- Braunstein, N. (2016). La traduction de l'intraduisible en psychanalyse. In N. Braunstein (Ed.), *Traduire la psychanalyse : Interpr tation, sens et transfert* (pp. 17-63).  r s.
- Breyel, C., et Grass, T. (2020). Machine translation and "biotranslation": an unlikely marriage. *SKASE Journal of Theoretical Linguistics*, 17(4), 9-22.
- Breyel, C., et Grass, T. (2021). Traduction automatique et biotraduction : le mariage for c . *Traduire*, 244, 94-106. <https://doi.org/10.4000/traduire.2350>
- Brezina, V. (2018). *Statistics in corpus linguistic: A practical guide*. Cambridge University Press.
- Brunet, E. (2015). La r p tition dans la phrase. Etude statistique. *Semen*, 38, <https://doi.org/10.4000/semen.10287>
- Brunette, L., et O'Brien, S. (2011). Quelle ergonomie pour la pratique post ditrice des textes traduits? *ILCEA*, 14. <https://doi.org/10.4000/ilcea.1081>
- Bussy, S. (2014). *Deep Learning –  tude de l'article : Layer-wise training of deep generative models* [Projet de fin d' tude, T l com SudParis]. https://simonbussy.com/doc/report_PFE.pdf
- Cadwell, P., O'Brien, S., et Teixeira, C.S.C. (2018). Resistance and accommodation: factors for the (non-) adoption of machine translation among professional translators. *Perspectives*, 26(3), 301-321. <https://doi.org/10.1080/0907676X.2017.1337210>
- Carl, M., et Schaeffer, M. (2017). Why translation is difficult: A corpus-based study of nonliterality in post-editing and from-scratch translation. *HERMES – Journal of Language and Communication in Business*, 5(6), 43-57.
- Carl, M., Dragsted, B., Elming, J., Hardt, D., et Jakobsen, A. L. (2011). The process of post-editing: A pilot study. In B. Sharp, M. Zock, M. Carl et A. L. Jakobsen (Eds.), *Proceedings of the 8th International NLPCS Workshop* (pp. 131-142). Samfundslitteratur.

- Carl, M., Gutermuth, S., et Hansen-Schirra, S. (2015). Post-editing machine translation Efficiency, strategies, and revision processes in professional translation settings. In A. Ferreira et J. Schwieter (Eds.), *Psycholinguistic and Cognitive Inquiries into Translation and Interpreting Vol. 115* (pp. 145-174). Benjamins Translation Library.
- Cassin, B. (2016). *Éloge de la traduction. Compliquer l'universel*. Fayard.
- Castagnoli, S., Ciobanu D., Kübler N., Kunz K., et Volanschi A. (2011). Designing a Learner Translator Corpus for Training Purposes. In N. Kübler (Ed.), *Corpora, Language, Teaching, and Resources: From Theory to Practice* (pp. 221-248). Peter Lang.
- Castilho, S., et Resende, N. (2022). Post-Editese in Literary Translations. *Information*, 13(2), 1-22.
- Castilho, S., Doherty, S., Gaspari, F., et Moorkens, J. (2018). Approaches to Human and Machine Translation Quality Assessment. In Moorkens, J., Castilho, S., Gaspari, F., et Doherty, S. (Eds.), *Translation Quality Assessment. Machine Translation: Technologies and Applications* (pp. 9-38). Springer.
- Castilho, S., Moorkens, J., Gaspari, F., Sennrich, R., Sosoni, V., Georgakopoulou, Y., Lohar, P., Way, A., Miceli Barone, A., et Gialama, M. (2017). A Comparative Quality Evaluation of PBSMT and NMT using Professional Translators. In *Proceedings of MT Summit XVI, vol.1: Research Track* (pp. 116-131). http://aamt.info/app-def/S-102/mtsummit/2017/wp-content/uploads/sites/2/2017/09/MTSummitXVI_ResearchTrack.pdf
- Castilho, S., Resende, N.C.A., et Mitkov, R. (2019). What Influences Post-editese features? A preliminary study. In *Proceedings of the Second Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT 2019) (Varna, Bulgarie, 5-6 septembre 2019)*. <http://rgcl.wlv.ac.uk/wp-content/uploads/2019/11/HiT-IT2019-proceedings.pdf>
- Castillo, J. J. (2009, 17 mai). Non-Probability Sampling. *Explorable.com*. <https://explorable.com/non-probability-sampling>
- Chaudiron, S. (2004). *Évaluation des systèmes de traitement de l'information*. Lavoisier.
- Cho, K., van Merriënboer, B., Bahdanau, D., et Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation* (pp. 103-111). Association for Computational Linguistics.
- Cholewska, D. (2021). Machine Translation Post-Editing (MTPE) from the Perspective of Translation Trainees: Implications for Translation Pedagogy. In *Proceedings of Machine Translation Summit XVIII: Users and Providers Track* (pp. 200-210). Association for Machine Translation in the Americas.
- Chuquet, H., et Paillard, M. (1989). *Approche linguistique des problèmes de traduction anglais-français*. Ophrys.

- CNRTL (s.d.). Irrationnel. In *Centre national de ressources textuelles et lexicales*. En ligne. <https://www.cnrtl.fr/definition/irrationnel>
- Cochrane, G. (1995). Le foisonnement, phénomène complexe. *TTR*, 8(2), 175-193. <https://doi.org/10.7202/037222ar>
- Cochrane, G. (2000). *Le foisonnement dans les textes de spécialité, illusion d'optique ou réalité quantifiable ?* [Thèse de doctorat]. Université Laval.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.
- Cooper, A. (2004). *The Inmates are Running the Asylum: Why Hi-Tech Products Drives us Crazy and How to Restore the Sanity*. SAMS.
- Covington, Michael, A. et McFall, Joe. D. (2010). Cutting the Gordian Knot: The Moving-Average Type-Token Ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2), 94-100.
- Cronin, M. (2013). *Translation in the Digital Age*. Routledge.
- Cronin, M. (2017a). *Eco-translation: Translation and ecology in the age of the Anthropocene*. Routledge, Taylor & Francis Group.
- Cronin, M. (2017b, septembre). *Translation and Sustainability in the Age of the Anthropocene*. Communication présentée à la FIT de l'Université de Genève, Genève. <https://mediaserver.unige.ch/play/VN3-2755-2017-2018-B>
- Cumbreño, C., et Aranberri, N. (2021). What Do You Say? Comparison of Metrics for Post-editing Effort. In M. Carl (Eds.), *Explorations in Empirical Translation Process Research. Machine Translation: Technologies and Applications*. Springer. <https://doi.org/10.1007/978-3-030-69777-8>
- Čulo, O., et Nitzke, J. (2016). Patterns of terminological variation in post-editing and of cognate use in machine translation in contrast to human translation. In *Proceedings of the 19th Annual Conference of the European Association for Machine Translation, EAMT 2017* (pp. 106-114). European Association for Machine Translation.
- Čulo O., Gutermuth, S., Hansen-Schirra, S., et Nitzke, J. (2014). The Influence of Post-Editing on Translation Strategies. In S. O'Brien, L. Winther-Balling, M. Carl, M. Simard et L. Specia (Eds.), *Post-editing of Machine Translation: Processes and Applications* (pp. 200-218). Cambridge Scholars Publishing.
- 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]. Université de Gand.

- Daems, J., et Macken, L. (2019). Interactive adaptive SMT versus interactive adaptive NMT: a user experience evaluation. *Machine Translation*, 33(1/2), 117-134. <http://www.jstor.org/stable/45222464>
- Daems, J., De Clercq, O., et Macken, L. (2017). Translationese and Post-editeese: How comparable is comparable quality? *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 16, 89-103. <https://doi.org/10.52034/lanstts.v16i0.434>
- Daems, J., Vandepitte, S., Hartsuiker, R., et Macken, L. (2017). Translation Methods and Experience: A Comparative Analysis of Human Translation and Post-editing with Students and Professional Translators. *Meta*, 62(2), 245-270. <https://doi.org/10.7202/1041023ar>
- Dalügge-Momme, N. (2016, décembre). *Devons-nous craindre la traduction automatique ? Intégration de la traduction automatique et avenir de notre profession*. Communication présentée à la Faculté de Philosophie et Lettres de l'Université de Liège, Liège.
- DataFranca (2021). Étiquetage morphosyntaxique. In *le Grand Lexique français de l'intelligence artificielle*. https://datafranca.org/wiki/Étiquetage_morphosyntaxique
- DataFranca (2023). Système à base de règles. In *le Grand Lexique français de l'intelligence artificielle*. https://datafranca.org/wiki/Système_à_base_de_règles
- De Almeida, G. (2013). *Translating the post-editor: an investigation of post-editing changes and correlations with professional experience across two Romance languages* [Thèse de doctorat], Dublin City University.
- De Almeida, G., et O'Brien, S. (2010). Analysing post-editing performance: correlations with years of translation experience. In *Proceedings of the 14th Annual Conference of the European Association for Machine Translation (Saint-Raphaël, France, 27-28 mai 2010)*. European Association for Machine Translation. <https://aclanthology.org/2010.eamt-1.19>
- De Clercq, O., de Sutter, G., Loock, R., Cappelle, B., et Plevoets, K. (2021). Uncovering Machine Translationese Using Corpus Analysis Techniques to Distinguish between Original and Machine--Translated French. *Translation Quarterly*, 101, 21-45.
- De Faria Pires, L. (2018). Intégration de la traduction automatique neuronale à la formation universitaire des futurs traducteurs : pistes d'exploration. *Myriades*, 4, 50-61.
- De Faria Pires, L. (2020). *Effets de l'expérience professionnelle sur la qualité des textes post-édités à la DGT : Étude de cas à partir d'une traduction automatique statistique et neuronale* [Thèse de doctorat en Langues, Lettres et Traductologie]. Université de Mons.
- De Ketele, J.-M., et Roegiers, X. (1993). *Méthodologie du recueil d'informations*. De Boeck Université.

- DeepL GmbH (2020). *Comparaison de DeepL Traducteur avec ses concurrents*. Cologne : DeepL GmbH <https://www.deepl.com/fr/quality.html>
- DeepL GmbH (2022). *Pourquoi DeepL ?* Cologne : DeepL GmbH. <https://www.deepl.com/fr/whydeepl>
- DeepL GmbH (2023). *DeepL Write*. Cologne : DeepL GmbH. <https://www.deepl.com/en/blog/introducing-deepl-write>
- Delisle, J. (1993). *La traduction raisonnée. Manuel d'initiation à la traduction professionnelle de l'anglais vers le français*. Presses de l'Université d'Ottawa.
- Delisle, J. (2013). *La traduction raisonnée. Manuel d'initiation à la traduction professionnelle de l'anglais vers le français* (3^e éd.). Presses de l'Université d'Ottawa.
- Deneufbourg, G. (2013). *La fluence discursive en interprétation de conférence. Approche évaluative par la traduction à vue*. Specialized Master's Article – UMONS. https://www.researchgate.net/publication/330579786_La_fluence_discursive_en_interpretation_de_conference_Approche_evaluative_par_la_traduction_a_vue
- Deneufbourg, G. (2018, mai). *Post-édition de traduction automatique : se méfier des apparences*. Communication présentée à la journée d'étude ULB TRADITAL : La post édition, avenir radieux ou fin d'un monde ? Université libre de Bruxelles, Bruxelles. [Présentation PowerPoint].
- Deneufbourg, G. (2019). *Post-édition de traduction automatique : se méfier des apparences*. Article inspiré d'une communication présentée au Congrès annuel de l'American Translators Association (Nouvelle-Orléans, 25 octobre 2018). <https://www.ata-divisions.org/FLD/index.php/tag/guillaume-deneufbourg/>
- Deneufbourg, G. (2020, janvier). *La traduction, un métier en voie de disparition ?* Communication présentée à la journée d'études Traduction & Qualité 2020 : Biotraduction et traduction automatique, Université de Lille, Lille.
- Deneufbourg, G. (2021, 13 octobre). *Traduction automatique : la dangereuse « sagesse des foules »*. *The Conversation*. <https://theconversation.com/traduction-automatique-la-dangereuse-sagesse-des-foules-169376>
- DePalma, D. (2013, décembre). *Post-editing in practice*. *tcworld magazine*. <https://www.tcworld.info/e-magazine/translation-and-localization/post-editing-in-practice-459/>
- DePalma, D. A., et Lommel, A. (2017, 15 février). *Augmented translation powers up language services* [Billet de blog]. <http://www.commonseadvisory.com/Blogs.aspx?1=1&moduleID=390&Contenttype=ArticleDe tAD&Aid=37907>

- Depraetere, I. (2010, mai). *What counts as useful advice in a university postediting training context? Report on a case study*. Communication présentée à la 14^e conférence annuelle de l'Association européenne de traduction automatique (EAMT) 2010, Saint-Raphaël, France. <http://www.mtarchive.info/EAMT-2010Depraetere-2.pdf>
- Depraetere, I., De Sutter, N., et Tezcan, A. (2014). Post-edited quality, post-editing behaviour and human evaluation: a case study. In S. O'Brien, L. Winther-Balling, M. Carl, M. Simard et L. Specia (Eds.), *Post-editing of Machine Translation: Processes and Applications* (pp. 78-108). Cambridge Scholars Publishing.
- Dillon, A. (2001). User Acceptance of Information Technology. In W. Karwowski (Ed.), *Encyclopedia of Human Factors and Ergonomics* (pp. 1-10). Taylor and Francis.
- Dirand, B., et Rossi, C. (2019). Biotraducteur et traducteur automatique. L'homme outillé et la machine-homme. In É. Lavault-Olléon et M. Zimina (Eds.), *Des mots aux actes* (pp. 63-78). Classiques Garnier.
- Do Carmo, F., et Moorkens, J. (2020). Differentiating editing, post-editing, and revision. In M. Koponen, B. Mossop, I. Robert et G. Scocchera (Eds.) *Translation Revision and Post-editing: Industry Practices and Cognitive Processes* (pp. 35-49). Routledge (Taylor & Francis).
- Doddington, G. (2002). Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of the second international conference on Human Language Technology Research* (pp. 138-145). Morgan Kaufmann Publishers Inc.
- Doherty, S. (2016). The Impact of Translation Technologies on the Process and Product of Translation. *International Journal of Communication*, 10, 947-969.
- Doherty, S. (2017). Issues in human and automatic translation quality assessment. In D. Kenny (Ed.), *Human issues in translation technology* (pp. 131-148). Routledge.
- Doherty, S., et Moorkens, J. (2013). Investigating the experience of translation technology labs: pedagogical implications. *The Journal of Specialised Translation*, 19, 122-136.
- Doherty, S., et O'Brien, S. (2014). Assessing the usability of raw machine translated output: A usercentered study using eye tracking. *International Journal of Human-Computer Interaction*, 30(1), 40-51.
- Doherty, S., Moorkens, J., Gaspari, F., et Castilho, S. (2018). On Education and Training in Translation Quality Assessment. In J. Moorkens, S. Castilho, F. Gaspari, et S. Doherty (Eds.), *Translation Quality Assessment. Machine Translation: Technologies and Applications, vol 1* (pp. 95-106). Springer. https://doi.org/10.1007/978-3-319-91241-7_5
- Dragsted, B. (2004). *Segmentation in Translation and Translation Memory Systems – An Empirical Investigation of Cognitive Segmentation and Effects of Integrating a TM System into the Translation Process*. Samfundslitteratur.

- Druide (2022). *Antidote (version 10) : Le remède à tous vos mots* [Logiciel]. Druides Informatique Inc.
- Durieux, C. (1990). Le foisonnement en traduction technique d'anglais en français. *Meta*, 35(1), 55-60. <https://doi.org/10.7202/002689ar>
- Dussart, A. (2005). Faux sens, contresens, non-sens... un faux débat ? *Meta*, 50(1), 107-119. <https://doi.org/10.7202/010661ar>
- El Hajjari, S. (2020). La Cohérence Textuelle Comme Outil et Comme Finalité. *Recherches en Langue et Littérature Françaises*, 13(24), 51-64.
- Elie, C., et Colombet, I. (2011). Méthodes d'estimation de la reproductibilité. *Sang Thrombose Vaisseaux*, 23(3), 138-45.
- Elming, J., Balling, L. W., et Carl, M. (2014). Investigating User Behaviour in Post-editing and Translation using the CASMACAT Workbench. In S. O'Brien, L. W. Balling, M. Carl, M. Simard et L. Specia (Eds.), *Post-editing of Machine Translation: Processes and applications* (pp. 147-169). Cambridge Scholars Publishing.
- Etalab (2021). Lexique des termes techniques – Pseudonymiser des documents grâce à l'IA. In etalab.gouv.fr. <https://guides.etalab.gouv.fr/pseudonymisation/lexique/>
- Farnir, F., Massart L., et Rives, A. (2021). *Corrélation de Pearson* (Biostatistiques). Repéré sur http://www.biostat.ulg.ac.be/pages/Site_r/corr_pearson.html
- Farrell, M. (2018). Machine Translation Markers in Post-Edited Machine Translation Output. In *Proceedings of the 40th Conference Translating and the Computer* (pp. 50-59). Asling.
- Fédération Wallonie-Bruxelles (2021). Enseignement supérieur – conditions de réussite et voies de recours. Portail officiel de la fédération Wallonie-Bruxelles. <http://www.enseignement.be/index.php?page=23812>
- Federico, M., Bertoldi, N., Cettolo, M., Negri, M., Turchi, M., Trombetti, M., Cattelan, A., Farina, A., Lupinetti, D., Martines, A., Massidda, A., Schwenk, H., Barrault, L., Blain, F., Koehn, P., Buck, C., et Germann, U. (2014a). The MateCat Tool. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics : System Demonstrations* (pp. 129-132). Dublin City University et Association for Computational Linguistics. <https://aclanthology.org/C14-2028>
- Federico, M., Bertoldi, N., Trombetti, M., et Cattelan, A. (2014b). [Working with MateCat: user manual and installation guide](#). In *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas: Tutorials* [Présentation PowerPoint]. Association for Machine Translation in the Americas. <https://aclanthology.org/2014.amta-tutorials.3>

- Fiederer, R., et O'Brien, S. (2009). Quality and machine translation: A realistic objective? *The Journal of Specialised Translation*, 11, 52-74.
- Flanagan, M., et Christensen, T. (2014). Testing post-editing guidelines: how translation trainees interpret them and how to tailor them for translator training purposes. *The Interpreter and Translator Trainer*, 8(2), 257-275.
- Flöter-Durr, M. (2022). Les limites épistémologiques des techniques numériques actuelles de l'intelligence artificielle en traduction : Epistemological limits of current digital techniques of artificial intelligence in translation. *Lebende Sprachen*, 67(1), 4-44. <https://doi.org/10.1515/les-2022-0004>
- Fontenelle, T. (2019, mai). *Le traducteur ne sera pas remplacé par la technologie – Il sera remplacé par un traducteur qui utilise la technologie*. Communication présentée à la journée d'étude CIRTI Le traducteur à l'ère des nouvelles technologies 2019, Université de Liège, Liège. <http://www.traduction2019.uliege.be/>
- Forcada, M. L. (2017). Making sense of neural machine translation. *Translation Spaces*, 6(2), 291-309. <https://doi.org/10.1075/ts.6.2.06for>
- Fort, K., François C., et Ghribi, M. (2010). Évaluer des annotations manuelles dispersées : les coefficients sont-ils suffisants pour estimer l'accord inter-annotateurs ? In *Actes de la 17^e conférence sur le Traitement Automatique des Langues Naturelles (TALN)* (pp. 341-350). ATALA.
- Froeliger, N. (2013). *Les Noces de l'analogique et du numérique : De la traduction pragmatique*. Belles lettres.
- Froeliger, N. (2014). De la traductologie des états d'âme et « vice versa » : vers une étude des aspects psychologiques en traduction. *Vertimo studijos*, 7, 59-67.
- Fulford, H. (2002). Freelance translators and machine translation: an investigation of perceptions, uptake, experience and training needs. In *Proceedings of the 6th EAMT Workshop: Teaching Machine Translation (Manchester, 14-15 novembre 2002)*. <https://aclanthology.org/2002.eamt-1.13>
- Gana, K., et Broc, G. (2018). *Introduction à la modélisation par équations structurales : manuel pratique avec lavaan*. ISTE Editions.
- Gangloff, B. (2008). *Les fondements méthodologiques du savoir scientifique*. Presses Universitaires de Rouen et du Havre.
- García, I. (2010). Is Machine Translation Ready Yet? *Target*, 22(1), 7-21. <https://doi.org/10.1075/target.22.1.02gar>
- García, I. (2011). Translating by post-editing: Is it the way forward? *Machine Translation*, 25, 217-237. <https://doi.org/10.1007/s10590-011-9115-8>

- García, I. (2012). A brief history of postediting and of research on postediting. *Revista Anglo Saxonica*, 3, 293-310.
- Gaspari, F. (2001). Teaching Machine Translation to Trainee Translators: a Survey of Their Knowledge and Opinions. In M.L. Forcada, J.A. Pérez-Ortiz et D.R. Lewis (Eds.), *Proceedings of the Workshop "Teaching Machine Translation" (Saint-Jacques-de-Compostelle, 18-22 septembre 2001)*. <https://aclanthology.org/2001.mtsummit-teach>
- Gaspari, F., Toral, A., Naskar S. K., Groves, D., et Way, A. (2014). Perception vs. reality: measuring machine translation post-editing productivity. In S. O'Brien, M. Simard et L. Specia (Eds.), *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas* (pp. 60-72). Association for Machine Translation in the Americas.
- Geitgey, A. (2016, 22 août). Machine Learning is Fun! Part 5: Language Translation with Deep Learning and the Magic of Sequences. *Medium*. <https://medium.com/@ageitgey/machine-learning-is-fun-80ea3ec3c471>
- Gélinas, S. (2017). *Test de Zulliger : réflexions critiques et évidences de validité basées sur le construit*. [Thèse de doctorat en psychologie, Université du Québec]. <https://depot-e.uqtr.ca/id/eprint/8187>
- Gellerstam, M. (1986). Translationese in Swedish novels translated from English. In L. Wollin et H. Lindquist (Eds.), *Translation Studies in Scandinavia Vol. 4* (pp. 88-95). CWK Gleerup.
- Gérard, F.-M. (2002). L'indispensable subjectivité de l'évaluation. *Antipodes*, 156, 26-34.
- Gerlach, J. (2008). *Les Interlangues en TA : l'exemple de MedSLT* [Mémoire de Master en Traduction, Université de Genève]. <https://archive-ouverte.unige.ch/unige:5563>
- Gerlach, J. (2015). *Improving statistical machine translation of informal language: a rule-based pre-editing approach for French Forums* [Thèse de doctorat en Traitement Informatique Multilingue, Université de Genève]. <https://doi.org/10.13097/archive-ouverte/unige:73226>
- Gile, D. (2005). *La traduction. La comprendre, l'apprendre*. Presses Universitaires de France. <https://doi.org/10.3917/puf.gile.2005.01>
- Gile, D. (2011). La recherche traductologique : méthodes ou approche? *TTR*, 24(2), 41-64. <https://doi.org/10.7202/1013394ar>
- Gondard-Delcroix, C. (2006). *La combinaison des analyses qualitative et quantitative pour une étude des dynamiques de pauvreté en milieu rural malgache* [Thèse de doctorat en Sciences économiques, Université Montesquieu - Bordeaux IV]. <https://theses.hal.science/tel-00165502>

- González Pastor, D. (2021). Introducing Machine Translation in the Translation Classroom: A Survey on Students' Attitudes and Perceptions. *Revista Tradumàtica. Tecnologies de la Traducció*, 19, 47-65. <https://doi.org/10.5565/rev/tradumatica.273>
- Gouadec, D. (1981). Paramètres de l'évaluation des traductions. *Meta*, 26(2), 99-116. <https://doi.org/10.7202/002949ar>
- Gouadec, D. (1989). Comprendre, évaluer, prévenir : pratique, enseignement et recherche face à l'erreur et à la faute en traduction. *TTR*, 2(2), 35-54. <https://doi.org/10.7202/037045ar>
- Gouadec, D. (2007). *Translation as a Profession*. John Benjamins Publishing Company.
- Gouadec, D. (2010). Quality in translation. In Y. Gambier et L. van Doorslaer (Eds.), *Handbook of translation studies Vol 1* (pp. 270-275). John Benjamins Publishing Company. <https://doi.org/10.1075/hts.1.qua1>
- Grass, T. (2022). L'erreur n'est pas humaine. *Traduire*, 246, 10-23. <https://doi.org/10.4000/traduire.2763>
- Grbíć, N. (2008). Constructing interpreting quality. *Interpreting*, 10(2), 232-257.
- Green, S., Heer, J., et Manning, C.D. (2013). The efficacy of human post-editing for language translation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 439-448). Association for Computing Machinery. <https://doi.org/10.1145/2470654.2470718>
- Groupe d'experts EMT (2022). *Référentiel de compétences de l'EMT*. Commission européenne, Bruxelles. https://commission.europa.eu/system/files/2023-01/emt_competence_fw_k_2022_fr.pdf
- Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, 11, 1-21.
- Grunenwald, N. (2021). *Sur les bouts de la langue. Traduire en féministe/s*. La Contre Allée.
- Guerberof Arenas, A. (2008). *Productivity and Quality in the Post-editing of Outputs from Translation Memories and Machine Translation* [Thèse de doctorat en Traduction et études interculturelles, Universitat Rovira I Virgili]. <http://hdl.handle.net/10803/90247>
- Guerberof Arenas, A. (2009). Productivity and quality in MT post-editing, In *Proceedings of MT Summit XII-Workshop: Beyond Translation Memories: New Tools for Translators MT, AMTA 2009 (Ottawa, 26-30 août 2009)*. <https://aclanthology.org/2009.mtsummit-btm.7>
- Guerberof Arenas, A. (2013). What do professional translators think about post-editing. *The Journal of Specialised Translation*, 19, 75-95.
- Guerberof Arenas A., et Moorkens J. (2019). Machine translation and post-editing training as part of a master's programme. *The Journal of Specialised Translation*, 31, 217-238.

- Guerra Martínez, L. (2003). *Human Translation versus Machine Translation and Full Postediting of Raw Machine Translation Output* [Mémoire de master en Traductologie, Dublin City University]. <http://sceuromix.com/enlaces/>
- Hallgren, K. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor Quantitative Methods in Psychology*, 8(1), 23-34. <https://doi.org/10.20982%2Ftqmp.08.1.p023>
- Halliday, M. A. K. (1985). *Spoken and written language*. Oxford University Press.
- Hansen, D. (2021). Les lettres et la machine : un état de l'art en traduction littéraire automatique. In *Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles. Vol. 2 : 23e Rencontres jeunes Chercheurs en Informatique pour le TAL (RECITAL)* (pp. 28-45). ATALA.
- Hansen, D., et Esperança-Rodier, E. (2023). Human-Adapted MT for Literary Texts: Reality or Fantasy? In S. Castilho, R. Caro Quintana, M. Stasimioti et V. Sosoni (Eds.), *Proceedings of the International Conference New Trends in Translation and Technology NeTTT 2022* (pp. 178-190). Incoma Ltd.
- Hansen, D., Esperança-Rodier, E., Blanchon, H., et Bada, V. (2022). La traduction littéraire automatique : Adapter la machine à la traduction humaine individualisée. *Journal of Data Mining and Digital Humanities. Vers une robotique du traduire [Hors-série]*. <https://dx.doi.org/10.46298/jdmdh.9114>
- Hartley, A., et Popescu-Belis, A. (2004). Évaluation des systèmes de traduction automatique. In S. Chaudiron (Ed.), *Évaluation des systèmes de traitement de l'information. Sciences et technologies de l'information* (pp.311-335). Hermès.
- Hassan, H., Aue, A., Chen, C., Chowdhary, V., Clark, J., Federmann, C., Huang, X., Junczys-Dowmunt, M., Lewis, W., Li, M., Liu, S., Liu, T.-Y., Luo, R., Menezes, A., Qin, T., Seide, F., Tan, X., Tian, F., Wu, L., Wu, S., Xia, Y., Zhang, D., Zhang, Z., et Zhou, M. (2018). Achieving Human Parity on Automatic Chinese to English News Translation. *ArXiv preprint*. <https://arxiv.org/abs/1803.05567>
- Hickey, S. (2023). *The 2023 Nimdzi 100: The Ranking of the Top 100 Largest Language Service Providers* (Rapport – février 2023). <https://www.nimdzi.com/nimdzi-100-top-lsp/>
- Hothorn T., Bretz F., et Westfall P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical Journal*, 50(3), 346-363. <https://doi.org/10.1002/bimj.200810425>
- House, J. (1998). Quality of translation. In M. Baker (Ed.), *The Routledge Encyclopedia of Translation Studies* (pp. 197-200). Routledge.
- Hu, K., et Cadwell, P. (2016). A comparative study of post-editing guidelines. *Baltic Journal of Modern Computing*, 4(2), 346-353.
- Hutchins, W. J. (2000). *Early years in machine translation: memoirs and biographies of pioneers*. John Benjamins Publishing Company.

- Hynninen, V. (2018). *Enseigner la postédition rapide. Étude d'une méthode d'apprentissage* [Mémoire de master en Philologie française, section traduction]. Université de Helsinki.
- Idiomaticité (2020, 3 avril). In *Wiktionnaire*. <https://fr.wiktionary.org/wiki/idiomaticit%C3%A9>
- Imbres, M. (2018). *L'intégration de la traduction automatique avec post-édition aux systèmes de mémoire de traduction : la traduction de textes techniques dans SDL Trados Studio 2017* [Mémoire de Maitrise en Traduction, Université de Genève]. <https://archive-ouverte.unige.ch/unige:112067>
- IRD – Institut de recherche et de documentation pédagogique (2017). Validité écologique. <https://www.irdp.ch/institut/validite-ecologique-2182.html>
- Jia, Y., Carl, M., et Wang, X. (2019). How does the post-editing of Neural Machine Translation compare with from-scratch translation? A product and process study. *The Journal of Specialised Translation*, 31, 60-86.
- Jilinskaya, M. (2012). *Approche psychométrique et différentielle de la mesure du leadership par la méthode à 360 degrés* [Thèse de doctorat en psychologie, Université Paris-Ouest Nanterre La Defense]. <https://www.theses.fr/2012PA100112/document>
- Johnson, W. (1944). Studies in language behavior: A Program of Research. *Psychological Monographs*, 56(2), 1-15. https://www.mpi.nl/publications/item_2350946
- Jurafsky, D., et Martin J.H. (2008). *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. (2^e éd.). Prentice-Hall.
- Jwalapuram, P., Joty, S.R., et Shen, Y. (2020). Pronoun-targeted fine-tuning for NMT with hybrid losses. In B. Webber, T. Cohn, Y. He et Y. Liu (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 2267-2279). Association for Computational Linguistics.
- Kadiu, S. (2016). Des zones d'indécidabilité dans la traduction automatique et dans la traduction humaine. *Meta*, 61(1), 204-220. <https://doi.org/10.7202/1036990ar>
- Kenny, D. (2015, novembre). *Machine Translation: broadening our horizons*. Communication présentée au Translating Europe Regional Workshop on the opportunities and threats of machine translation for professional translators, Durham. <http://www.ciol.org.uk/content/machine-translation-opportunities-and-threats>
- Kenny, D. (2017, juillet). *Of Meaning, Materiality and Magic: Conceptualising Machine Translation in the 21st Century*. Communication présentée au Translation [and Interpreting] in Transition 3, Université de Gand, Gand.

- Kenny, D. (2018). Sustaining Disruption? The Transition from Statistical to Neural Machine Translation. *Revista Tradumàtica. Tecnologies de la Traducció*, 16, 59-70. <https://doi.org/10.5565/rev/tradumatica.221>
- Kenny, D. (2022). *Machine translation for everyone: Empowering users in the age of artificial intelligence* (Translation and Multilingual Natural Language Processing 18). Language Science Press.
- Kenny, D., et Doherty, S. (2014). Statistical Machine Translation in the Translation Curriculum: Overcoming Obstacles and Empowering Translators. *The Interpreter and Translator Trainer*, 8(2), 276-294. <https://doi.org/10.1080/1750399X.2014.936112>
- Killman, J. (2018). Translating the same text twice: An English-Spanish comparative product study of post-edited translations vs. human translations. *The Journal of Internationalization and Localization*, 5, 114-141. <https://doi.org/10.1075/jial.18003.kil>
- Koby, G. S., et Melby, A.K. (2013). Certification and Job Task Analysis (JTA): Establishing Validity of Translator Certification Examinations. *The International Journal of Translation and Interpreting Research*, 5(1), 174-210. <http://dx.doi.org/10.12807/ti.105201.2013.a10>
- Koehn, P. (2009). A Process Study of Computed Aided Translation. *Machine Translation*, 23(4), 241-263.
- Koehn, P. (2010). *Statistical Machine Translation*. Cambridge University Press.
- Koehn, P. (2020). *Neural Machine Translation*. Cambridge University Press.
- Koehn, P., et Monz, C. (2006). Manual and automatic evaluation of machine translation between European languages. In *Proceedings of the Workshop on Statistical Machine Translation* (pp. 102-121). Association for Computational Linguistics.
- Koehn, P., et Knowles, R. (2017). Six Challenges for Neural Machine Translation. In *Proceedings of the 1st Workshop on Neural Machine Translation* (pp. 28-39). Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/W17-3204>
- Koglin, A. (2015). An empirical investigation of cognitive effort required to post-edit machine translated metaphors compared to the translation of metaphors. *Translation and Interpreting*, 7, 126-141.
- Koo, T. K., et Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine*, 15(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Koponen, M. (2015). How to teach machine translation post-editing? Experiences from a post-editing course. In *Proceedings of the 4th Workshop on Post-editing Technology and Practice (Miami, 30-3 novembre 2015)*. <https://aclanthology.org/2015.mtsummit-wptp.1>

- Koponen, M. (2016). Is Machine Translation Post-editing Worth the Effort? A Survey of Research into Post-editing and Effort. *The Journal of Specialised Translation*, 25, 131-148.
- Koponen, M, Aziz, W., Ramos, L., et Specia, L. (2012). Post-editing time as a measure of cognitive effort. In *Workshop on Post-Editing Technology and Practice (San Diego, 28 octobre 2012)*. Association for Machine Translation in the Americas.
- Koskinen, K., et Ruokonen, M. (2017). Love letters or hate mail? Translators' technology acceptance in the light of their emotional narratives. In D. Kenny (Ed.), *Human Issues in Translation Technology* (pp. 8-24). Routledge.
- Krings, H. P. (2001). *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*. The Kent State University Press.
- Kübler, N. (2013). Traduction pragmatique, linguistique de corpus, traducteur : un ménage à trois explosif ? In *Proceedings of Tralogy 2013, Futures in technology for translation (Paris, 17-18 janvier 2013)*. <https://hal.science/hal-02497335/document>
- Kuznetsova, A., Brockhoff, P.B., et Christensen, R.H.B. (2017). LmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1-26. <https://doi.org/10.18637/jss.v082.i13>
- Lacruz, I. (2017). Cognitive Effort in Translation, Editing, and Post-editing. In J.W. Schwieter et A. Ferreira (Eds.), *The Handbook of Translation and Cognition* (pp. 386-401). John Wiley & Sons <https://doi.org/10.1002/9781119241485.ch21>
- Lang, T. (2007). Documenting research in scientific articles: Guidelines for authors: 3. Reporting multivariate analyses. *Chest*, 131(2), 628-632. <https://doi.org/10.1378/chest.06-2088>
- Language Industry Survey. (2018). *Expectations and Concerns of the European Language Industry* (Rapport – 2018). European Union of Associations of Translation Companies. https://euatc.org/wp-content/uploads/2019/10/2018_Language_Industry_Survey_Report.pdf
- Language Industry Survey. (2019). *Expectations and Concerns of the European Language Industry* (Rapport – 2019). European Union of Associations of Translation Companies. <https://euatc.org/wp-content/uploads/2019/11/2019-Language-Industry-Survey-Report.pdf>
- Larousse. (s. d.). Vocable. In *Larousse Dictionnaire en ligne*. <https://www.larousse.fr/dictionnaires/francais/vocable/82341>
- Larose, R. (1998). Méthodologie de l'évaluation des traductions. *Meta*, 43(2), 163-186. <https://doi.org/10.7202/003410ar>

- Larsonneur, C. (2022, février). *Lectures traductologiques de l'humain et de la machine*. Communication présentée au séminaire Tract 2021-2022 L'impact de l'intelligence artificielle sur la traduction littéraire : théorie, pratique, création, Université Sorbonne Nouvelle, Paris.
- Läubli, S., et Orrego-Carmona, D. (2017). When google translate is better than some human colleagues, those people are no longer colleagues. In *Proceedings of the 39th Conference Translating and the Computer* (pp. 56-69). ASLing. <https://doi.org/10.5167/uzh-147260>
- Läubli, S., Sennrich, R., et Volk, M. (2018). Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 4791-4796). Association for Computational Linguistics.
- Läubli, S., Castilho, S., Neubig, G., Sennrich, R., Shen, Q., et Toral, A. (2020). A set of recommendations for assessing human-machine parity in language translation. *Journal of Artificial Intelligence Research*, 67, 653-672. <https://doi.org/10.1613/jair.1.11371>
- Lavault-Olléon, E. (2018). Regard diachronique et prospectif sur un demi-siècle de traductologie pragmatique. *ASp*, 74, 5-26. <https://doi.org/10.4000/asp.5239>
- Lavault-Olléon, E., et Carré, A. (2012). Traduction spécialisée : l'ergonomie cognitive au service de la formation. *ASp*, 62, 67-77. <https://doi.org/10.4000/asp.3117>
- Lavault-Olléon, E., et Allignol, C. (2014). La notion d'acceptabilité en traduction professionnelle : où placer le curseur ? *ILCEA*, 19. <https://doi.org/10.4000/ilcea.2455>
- Lavie, A. (2011). Evaluating the Output of Machine Translation Systems. In *Proceedings of the 13th MT Summit, Tutorial Abstracts, (Xiamen, Chine, 19 septembre 2011)*. <https://aclanthology.org/2011.mtsummit-tutorials.3.pdf>
- Le Petit Robert (s.d.). Monochromatique. In *Dictionnaire le Petit Robert de la langue française* [En ligne]. <https://petitrobert.lerobert.com/robert.asp>.
- Le Petit Robert (s.d.). Stochastique. In *Dictionnaire le Petit Robert de la langue française* [En ligne]. <https://petitrobert.lerobert.com/robert.asp>.
- LeBlanc, M. (2013). Translators on Translation Memory (TM). Results of an ethnographic study in three translation services and agencies. *The International Journal for Translation and Interpreting Research*, 5(2), 1-13. <http://dx.doi.org/10.12807/ti.105202.2013.a01>
- LeBlanc, M. (2016). La traduction spécialisée à l'ère des nouvelles technologies : quel effet sur le texte de spécialité ? *Studia Romanica Posnaniensia*, 43(1), 77-92. <https://doi.org/10.14746/strop.2016.425.006>

- LeBlanc, M. (2017). I Can't Get No Satisfaction! Should We Blame Translation Technologies or Shifting Business Practices? In D. Kenny (Ed.), *Human Issues in Translation Technology* (pp. 45-62). Routledge.
- Lee, J., et Liao, P. (2011). A Comparative Study of Human Translation and Machine Translation with Post-editing. *Compilation and Translation Review*, 4(2), 105-149. <https://doi.org/10.29912/CTR.201109.0005>
- Lee-Jahnke, H. (2001). Aspects pédagogiques de l'évaluation des traductions. *Meta*, 46(2), 258-271.
- Légal, J.-B. (2008). *Quelques rappels concernant la méthode expérimentale*. Université Paris X-Nanterre. <http://j.b.legal.free.fr/Blog/share/Concepts/methodo.pdf>
- Legendre, P., et Borcard, D. (s.d.), *L'analyse de variance à un critère de classification (ANOVA)*. Université de Montréal. <https://fr.scribd.com/document/515625216/Anova>
- Léon, J. (2002). Le CNRS et les débuts de la traduction automatique en France. *La revue pour l'histoire du CNRS*, 6. <https://doi.org/10.4000/histoire-cnrs.3461>
- Léon, J. (2015). *Histoire de l'automatisation des sciences du langage*. ENS Éditions.
- Li, B., Wang, Z., Liu, H., Jiang, Y., Du, Q., Xiao, T., Wang, H., et Zhu, J. (2020). Shallow-to-Deep Training for Neural Machine Translation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020* (pp. 995-1005). Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/2020.emnlp-main.72>
- Limesurvey GmbH (s.d.). LimeSurvey: An Open Source survey tool [Logiciel]. LimeSurvey GmbH, Hambourg, Allemagne. <http://www.limesurvey.org>
- Lin, C.-Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Proceedings of Workshop on Text Summarization Branches Out* (pp. 74–81). Association for Computational Linguistics.
- Loffler-Laurian A.-M. (1984). Vulgarisation scientifique : formulation, reformulation, traduction. *Langue française*, 64, 109-125.
- Lommel, A. (2018a). Augmented Translation: A New Approach to Combining Human and Machine Capabilities. In *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas: Volume 2, User Track* (pp. 5-12). Association for Machine Translation in the Americas.
- Lommel, A. (2018b). Metrics for Translation Quality Assessment: A Case for Standardising Error Typologies. In J. Moorkens, S. Castilho, F. Gaspari, et S. Doherty (Eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 109-127). Springer.

- Lommel, A. (2020, 4 novembre). Augmented Translation: Are We There Yet? [Billet de blog]. *CSA Research*. <https://csa-research.com/Blogs-Events/Blog/augmented-translation-2020>
- Lommel, A., Burchardt, A., Popović, M., Harris, K., Avramidis, E., et Uszkoreit, H. (2014). Using a new analytic measure for the annotation and analysis of MT errors on real data. In *Proceedings of the 17th Annual Conference of the European Association for Machine Translation – EAMT 2014* (pp. 165-172).
- Loock, R. (2016). *La traductologie de corpus*. Presses universitaires du Septentrion.
- Loock, R. (2018). Traduction automatique et usage linguistique : une analyse de traductions anglais-français réunies en corpus. *Meta*, 63(3), 786-806. <https://doi.org/10.7202/1060173ar>
- Loock, R. (2019). La plus-value de la biotraduction face à la machine. *Traduire*, 241, 54-65.
- Loock, R. (2020a). Compte rendu de [Bowker, Lynne et Buitrago Ciro, Jairo (2019) : *Machine Translation and Global Research : Towards Improved Machine Translation Literacy in the Scholarly Community*. Bingley : Emerald Publishing, 111 p.]. *Meta*, 65(1), 270-272. <https://doi.org/10.7202/1073650ar>
- Loock, R. (2020b, janvier) : *Introduction*. Communication présentée à la journée d'études Traduction & Qualité 2020 : Biotraduction et traduction automatique, Université de Lille, Lille.
- Loock, R. (2020c). No more rage against the machine: how the corpus-based identification of machine-translationese can lead to student empowerment. *The Journal of Specialised Translation*, 34, 150-170.
- Loock, R. (2020d, mars). *Pour une approche raisonnée de la traduction automatique*. Communication présentée à la journée d'études Mots/Machines #2 : Traduction et sentiment : quel sens pour la machine ? Brest.
- Loock, R., et Léchauguette S. (2021). Machine translation literacy and undergraduate students in applied languages: report on an exploratory study. *Revista Tradumàtica. Tecnologies de la Traducció*, 19, 204-225. <https://doi.org/10.5565/rev/tradumatica.281>
- Loock, R., Mariaule, M., et Oster, C. (2013). Traductologie de corpus et qualité : étude de cas. In *Actes du colloque Tralogy II, Session 5 - Assessing Quality in MT / Mesure de la qualité en TA (Paris, 17-18 janvier 2013)*. <http://odel.irevues.inist.fr/tralogy/index.php?id=243>
- Lu, X. (2012). The Relationship of Lexical Richness to the Quality of ESL Learners' Oral Narratives. *The Modern Language Journal*, 96, 190-208.
- Lupo, L., Dinarelli, M., et Besacier, L. (2022). Focused Concatenation for Context-Aware Neural Machine Translation. In *Proceedings of the Seventh Conference on Machine Translation (WMT)* (pp. 830–842). Association for Computational Linguistics. <https://aclanthology.org/2022.wmt-1.77.pdf>

- Macken, L., Van Brussel, L., et Daems, J. (2019). NMT's wonderland where people turn into rabbits: a study on the comprehensibility of newly invented words in NMT output. *Computational Linguistics in the Netherlands Journal*, 9, 67-80.
- Mallinson, J., Sennrich, R., et Lapata, M. (2017). Paraphrasing revisited with neural machine translation. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (pp. 881-893). Association for Computational Linguistics.
- Martikainen, H. (2019a). Qualité fonctionnelle et distorsion d'interprétation en traduction médicale. In E. Lavault-Olléon et M. Zimina (Eds.), *Des mots aux actes. Traduction et technologie, regards croisés sur de nouvelles pratiques* (pp. 257-270). Classiques Garnier.
- Martikainen, H. (2019b). Post-Editing Neural MT in Medical LSP: Lexico-Grammatical Patterns and Distortion in the Communication of Specialized Knowledge. *Informatics*, 6(3), 26, 1-15 <https://doi.org/10.3390/informatics6030026>
- Martikainen, H. (2019c). *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], Université Paris-Diderot.
- Martikainen, H. (2019d, mai). *Post-éditer à l'heure neuronale : quelle valeur ajoutée pour l'humain face à la machine ?* Communication présentée à la journée d'étude CIRTI Le traducteur à l'ère des nouvelles technologies 2019, Université de Liège, Liège.
- Martikainen, H. (2020, novembre). Traduire à l'ère numérique : point de vue francophone ? Communication présentée au Webinaire Traduction et humanités numériques, Università Ca' Foscari Venezia.
- Martikainen, H. (2021, avril). *Enseigner la post-édition en langue de spécialité dans le cadre d'un projet réel*. Communication présentée au colloque AFFUMT : Former aux métiers de la traduction aujourd'hui et demain, Université Grenoble Alpes, Grenoble.
- Martikainen, H. (2022a). 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. (2022b, avril). *Investigating the usability of automatic metrics for characterizing translated vs post-edited texts*. Communication présentée lors de la conférence internationale Tralogy3 - Traduction humaine et traitement automatique des langues : vers un nouveau consensus ? Paris.
- Martikainen, H., et Kübler, N. (2016). Ergonomie cognitive de la post-édition de traduction automatique : enjeux pour la qualité des traductions. *ILCEA*, 27. <https://doi.org/10.4000/ilcea.3863>

- Martikainen, H., et Mestivier, A. (2020, janvier). *Les outils de traduction nouvelle génération : quel effet sur la qualité des textes traduits ?* Communication présentée à la journée d'étude Traduction & Qualité 2020 : Biotraduction et traduction automatique, Université de Lille, Lille.
- Martínez Melis, N. (2001). *Évaluation et didactique de la traduction : le cas de la traduction dans la langue étrangère* [Thèse de doctorat en Traduction, Universitat Autònoma de Barcelona]. <http://hdl.handle.net/10803/5251>
- Martínez Melis, N., et Hurtado Albir, A. (2001). Assessment In Translation Studies: Research Needs. *Meta*, 46(2), 272-287. <https://doi.org/10.7202/003624ar>
- Massardo, I., van der Meer, J., et Khalilov, M. (2016). *TAUS Translation Technology Landscape Report* (Rapport – septembre 2016). TAUS.
- Massardo, I., et van der Meer, J. (2017). *The translation industry in 2022. A report from the TAUS Industry Summit* (Rapport – mars 2017). TAUS.
- Massey, G., et Ehrensberger-Dow, M. (2017). Machine learning: Implications for translator education. *Lebende Sprachen*, 62(2), 300-312. <https://doi.org/10.1515/les-2017-0021>
- McGraw, K.O., et Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30-46. <https://doi.org/10.1037/1082-989X.1.1.30>
- Mesa-Lao, B. (2014). Gaze Behaviour on Source Texts: An Exploratory Study Comparing Translation and Post-Editing. In S. O'Brien, L. Winther Balling, M., M. Simard et L. Specia (Eds), *Post-Editing of Machine Translation* (pp. 219-245). Cambridge Scholars Publishing.
- Mion, E. A. (2020, 28 avril). Le concept de traduction augmentée [Billet de blog]. <https://fr.eamtranslations.com/post/le-concept-de-traduction-augment%C3%A9e>
- Mion, E. A. (2021, 30 mars). Qu'est-ce qu'un moteur de traduction automatique générique ? [Billet de blog]. <https://fr.eamtranslations.com/post/qu-est-ce-qu-un-moteur-de-traduction-automatique-g%C3%A9n%C3%A9rique>
- Moorkens, J. (2017). Under pressure: translation in times of austerity. *Perspectives*, 25(3), 464-477. <https://doi.org/10.1080/0907676X.2017.1285331>
- Moorkens, J. (2018). What to expect from Neural Machine Translation: a practical in-class translation evaluation exercise. *The Interpreter and Translator Trainer*, 12(4), 375-387. <https://doi.org/10.1080/1750399X.2018.1501639>
- Moorkens, J. (2020). Comparative satisfaction among freelance and directly-employed Irish-language translators. *Translation and Interpreting*, 12(1), 55-73. <http://dx.doi.org/10.12807/ti.112201.2020.a04>

- Moorkens, J. (2022a). Ethics and machine translation. In D. Kenny (Ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence* (pp. 121-140). Language Science Press. <http://dx.doi.org/10.5281/zenodo.6759984>
- Moorkens, J. (2022b). Le traducteur, une espèce en voie de disparition ? *Le Courrier de l'UNESCO*, 2, 24-25. <https://courrier.unesco.org/fr/articles/le-traducteur-une-espece-en-voie-de-disparition>
- Moorkens, J., et O'Brien, S. (2015). Post-editing evaluations: trade-offs between novice and professional participants. In *Proceedings of the 18th annual conference of the European Association for Machine Translation EAMT 2015* (pp. 75-81). European Association for Machine Translation. <https://aclanthology.org/W15-4910>
- Moorkens, J., O'Brien, S., Silva, I.A.L., Fonseca, N., et Alves, F. (2015). Correlations of perceived post-editing effort with measurements of actual effort. *Machine Translation*, 29(3-4), 267-284.
- Moorkens, J., Lewis, D., Reijers, W., Vanmassenhove, E., et Way, A. (2016). Translation resources and translator disempowerment. In *Proceedings of ETHICA 2016: ETHics in corpus collection, annotation and application* (pp. 49-53).
- Moorkens, J., Castilho, S., Gaspari, F. et Doherty, S. (2018a). Introduction. In J. Moorkens, S. Castilho, F. Gaspari et S. Doherty (Eds.), *Translation Quality Assessment. Machine Translation: Technologies and Applications*, 1-6. Springer.
- Moorkens, J., Toral, A., Castilho, S., et Way, A. (2018b). Translators' perceptions of literary post-editing using statistical and neural machine translation. *Translation Spaces*, 7(2), 240-262. <https://doi.org/10.1075/ts.18014.moo>
- MQM Website (2022). The MQM Error Typology. *Themqm.org*. <https://themqm.org/error-types-2/typology/>
- Müller, R., et Büttner, P. (1994). A critical discussion of intraclass correlation coefficients. *Statistics in medicine*, 13(23-24), 2465-2476. <https://doi.org/10.1002/sim.4780132310>
- Muller, Ch. (1969). La statistique lexicale. *Langue française*, 2, Le lexique, sous la direction de Louis Guilbert, 30-43. <https://doi.org/10.3406/lfr.1969.5419>
- Mutal, J. D., Volkart, L., Bouillon, P., Girletti, S., et Estrella, P. S. (2019). Differences between SMT and NMT Output - a Translators' Point of View. In *Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT 2019)* (pp. 75-81). Incoma Ltd. http://dx.doi.org/10.26615/issn.2683-0078.2019_009
- Naver (entreprise). (2022, 1^{er} août). In *Wikipédia*. https://fr.wikipedia.org/wiki/Naver_%28entreprise%29#Notes_et_r%C3%A9f%C3%A9rences

- Nita, R., Brunet, B., Caron, P., Kleiber, G., et Vergez-Couret, M. (2021). *Cohérence et cohésion textuelles*. Lambert-Lucas.
- Nitzke, J. (2016). Monolingual post-editing: An exploratory study on research behaviour and target text quality. In S. Hansen-Schirra et S. Grucza (Eds.), *Eye-tracking and Applied Linguistics* (pp. 83-109). Language Science Press.
- Nitzke, J. (2019). *Problem solving activities in post-editing and translation from scratch: A multi-method study*. Language Science Press.
- Nitzke, J., et Hansen-Schirra, S. (2021). *A short guide to post-editing* (Translation and Multilingual Natural Language Processing 16). Language Science Press.
- Nunziati, G. (1990). Pour construire un dispositif d'évaluation formatrice. *Cahiers pédagogiques*, 280, 47-64.
- Nunziatini, M., et Marg, L. (2020). Machine Translation Post-Editing Levels: Breaking Away from the Tradition and Delivering a Tailored Service. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation* (pp. 309-318). European Association for Machine Translation.
- O'Brien, S. (2002). Teaching post-editing: a proposal for course content. In *Proceedings of the 6th EAMT Workshop: Teaching Machine Translation* (pp. 99-106). European Association for Machine Translation.
- O'Brien, S. (2006). Pauses as indicators of cognitive effort in post-editing Machine Translation output. *Across Languages and Cultures*, 7(1), 1-21. <http://dx.doi.org/10.1556/Acr.7.2006.1.1>
- O'Brien, S. (2010). Introduction to Post-Editing: Who, What, How and Where to Next? In *Proceedings of the 9th Conference of the Association for Machine Translation in the Americas: Tutorials (Denver, 31 octobre-4 novembre 2010)*. Association for Machine Translation in the Americas.
- O'Brien, S. (2012a). Towards a Dynamic Quality Evaluation Model for Translation. *The Journal of Specialised Translation*, 17, 55-77.
- O'Brien, S. (2012b). Translation as human-computer interaction. *Translation Spaces*, 1, 101-122.
- O'Brien, S. (2022). How to deal with errors in machine translation: Postediting. In D. Kenny (Ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence* (pp. 105-120). Language Science Press. <https://doi.org/10.5281/zenodo.6759982>
- O'Brien, S., et Ehrensberger-Dow, M. (2020). MT Literacy: A cognitive view. *Translation, Cognition & Behavior*, 3(2), 145-164. <https://doi.org/10.1075/tcb.00038.obr>

- O'Donnell, M. (2016). *The UAM CorpusTool 3.3* [Logiciel]. <http://corpustool.com>
- Oates, B. J. (2005). *Researching information systems and computing*. Sage.
- Office québécois de la langue française (2003). Implémentation. In *Le grand dictionnaire terminologique (GDT)*. <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8361143/mise-en-oeuvre>
- Office québécois de la langue française (2009). Faute de traduction. In *Le grand dictionnaire terminologique (GDT)*. <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/26502700/faute-de-traduction>
- Office québécois de la langue française (2023). Accord du verbe avec des sujets unis par *ni*. In *Le grand dictionnaire terminologique (GDT)*. <https://vitrinelinguistique.oqlf.gouv.qc.ca/21490/la-grammaire/le-verbe/accord-du-verbe-avec-le-sujet/sujets-joints-par-une-conjonction/accord-du-verbe-avec-des-sujets-unis-par-ni>
- OpenAI (2022). *ChatGPT: Optimizing Language Models for Dialogue*. <http://chat.openai.com/>
- Organisation internationale de normalisation (2017). *ISO 18587 : Services de traduction — Post-édition d'un texte résultant d'une traduction automatique — Exigences*. <https://www.iso.org/obp/ui/#iso:std:iso:18587:ed-1:v1:fr>
- Organisation internationale de normalisation (2018). *ISO 9241-11 : Ergonomie de l'interaction homme-système — Partie 11: Utilisabilité — Définitions et concepts*. <https://www.iso.org/obp/ui/#iso:std:iso:9241:-20:ed-2:v1:fr>
- Ortiz-Martínez, D., González-Rubio, J., Alabau, V., Sanchis-Trilles, G., et Casacuberta, F. (2016). Integrating Online and Active Learning in a Computer-Assisted Translation Workbench. In M. Carl, S. Bangalore et M. Schaeffer (Eds.), *New Directions in Empirical Translation Process Research. New Frontiers in Translation Studies* (pp. 57-76). Springer.
- Oxford University Press (2021). How To Use An Em Dash. In *Oxford University Press* [En ligne]. <https://www.lexico.com/grammar/dash-and-em-dash>
- Papineni, K., Roukos, S., Ward, T., et Zhu, W. (2002). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (pp. 311-318). Association for Computational Linguistics. <http://dx.doi.org/10.3115/1073083.1073135>
- Peraldi, S. (2016). De la traduction automatique brute à la post-édition professionnelle évoluée : le cas de la traduction financière. *Revue française de linguistique appliquée*, XXI(1), 67-90.

- Pérez-Ortiz, J. A., Forcada, M.L., et Sánchez-Martínez, F. (2022). How neural machine translation works. In D.Kenny (Ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence* (pp. 141-164). Language Science Press. <https://doi.org/10.5281/zenodo.6760020>
- Poibeau, T. (2016). Traduire sans comprendre ? La place de la sémantique en traduction automatique. *Langages*, 201, 77-90. <https://doi.org/10.3917/lang.201.0077>
- Poibeau, T. (2019). *Babel 2.0 : Où va la traduction automatique ?* Odile Jacob.
- Poibeau, T. (2022). On “Human Parity” and “Super Human Performance” in Machine Translation Evaluation. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 6018-6023). European Language Resources Association.
- Poirier, E. A. (2022). Vers une évaluation empirique des textes traduits et de la qualité en traduction. *Journal of Data Mining and Digital Humanities. Vers une robotique du traduire [Hors-série]*. <https://doi.org/10.46298/jdmdh.9123>
- Poncharal, B., et Birkan-Berz, C. (2022). *Appel à contributions du colloque TRACT - Traduction littéraire et intelligence artificielle : théorie, pratique, création (Paris, 20 et 21 octobre 2022)*. <https://lit-trans-ai.sciencesconf.org/>
- Popel, M., Tomkova, M., Tomek, J., Łukasz Kaiser, Uszkoreit, J., Bojar, O., et Žabokrtský, Z. (2020). Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nature communications*, 11(1), 4381. <https://doi.org/10.1038/s41467-020-18073-9>
- Popescu-Belis, A. (2019). Context in neural machine translation: A review of models and evaluations. *ArXiv preprint*. <https://doi.org/10.48550/arXiv.1901.09115>
- Popović, M. (2011). Hjerson: an open source tool for automatic error classification of machine translation output. *Prague Bulletin of Mathematical Linguistics*, 96, 59-67. <https://doi.org/10.2478/v10108-011-0011-4>
- Popović, M. (2015). Chrf: Character n-gram f-score for automatic MT evaluation. In *Proceedings of the tenth workshop on statistical machine translation* (pp. 392-395). Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/W15-3049>
- Popović, M. (2018). Error Classification and Analysis for Machine Translation Quality Assessment. J. Moorkens, S. Castilho, F. Gaspari et S. Doherty (Eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 129-158). Springer. https://doi.org/10.1007/978-3-319-91241-7_7
- Popović, M. (2020). Informative Manual Evaluation of Machine Translation Output. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 5059-5069). International Committee on Computational Linguistics. <http://dx.doi.org/10.18653/v1/2020.coling-main.444>

- Popović, M. (2021). On nature and causes of observed MT errors. In *Proceedings of the 18th Biennial Machine Translation Summit Virtual* (pp. 163-175). Association for Machine Translation in the Americas.
- Post, M. (2018). A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers* (pp. 186-191). Association for Computational Linguistics. <https://doi.org/10.18653/v1/W18-6319>
- Prégent, R. (1990). *La préparation d'un cours*. Éditions de l'École polytechnique.
- Pym, A. (2013). Translation Skill-Sets in a Machine-Translation Age. *Meta*, 58(3), 487-503. <https://doi.org/10.7202/1025047ar>
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J., et 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* (pp. 101-108). Association for Computational Linguistics.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <http://www.r-project.org/index.html>
- Ramírez-Sánchez, G. (2022). Custom machine translation. In D. Kenny (Ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence* (pp. 165-186.). Language Science Press. <https://doi.org/10.5281/zenodo.6760022>
- Rei, R., Stewart, C., Farinha, A. C., et Lavie, A. (2020). COMET: A neural framework for MT evaluation. In B. Webber, T. Cohn, Y. He et Y. Liu (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 2685-2702). Association for Computational Linguistics. <https://dx.doi.org/10.18653/v1/2020.emnlp-main.213>
- Reiß, K., Vermeer, H. J. (2014). *Towards a general theory of translational action: skopos theory explained*. Routledge.
- Rémillard, J. (2018). *Utilité et utilisation de la traduction automatique dans l'environnement de traduction : une évaluation axée sur les traducteurs professionnels* [Thèse de doctorat en Traduction, Université d'Ottawa]. <http://dx.doi.org/10.20381/ruor-22046>
- Resende, N., et Way, A. (2021). Can Google Translate rewire your L2 English processing? *Digital*, 1, 66-85. <https://doi.org/10.3390/digital1010006>
- Robert, A.-M. (2010). La post-édition : l'avenir incontournable du traducteur ? *Traduire*, 222, 137-144. <https://doi.org/10.4000/traduire.460>

- Robert, A.-M. (2013). Vous avez dit post-éditrice ? Quelques éléments d'un parcours personnel. *The Journal of Specialised Translation*, 19, 29-40.
- Robinson, D. (2015). The somatics of tone and the tone of somatics: The Translator's Turn revisited. *Translation and Interpreting Studies. The Journal of the American Translation and Interpreting Studies Association*, 10(2). 299-319. <https://doi.org/10.1075/tis.10.2.09rob>
- Rossi, C. (2018). *Acquisition, traduction, usages : des linguistiques énonciatives et cognitives à l'automatisation* [Habilitation à diriger des recherches en Linguistique, Université Lumière Lyon 2]. <https://hal.science/tel-03659740>
- Rossi, C. (2019a). L'apprenti traducteur et la machine. Des connaissances aux perceptions de la traduction automatique. In E. Lavault-Olléon et M. Zimina (Eds.), *Des mots aux actes. Traduction et technologie, regards croisés sur de nouvelles pratiques* (pp. 93-105). Classiques Garnier.
- Rossi, C. (2019b, 17 octobre). *Les usages actuels de la traduction automatique*. Communication présentée à l'Atelier DigitHum 2019 : les humanités numériques en langues, Paris. https://dighum.humanum.fr/atelier/2019/4_rossi.php
- Rossi, C., et Carré, A. (2022). How to choose a suitable neural machine translation solution: Evaluation of MT quality. In D. Kenny (Ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence* (pp. 51-79). Language Science Press. <https://doi.org/10.5281/zenodo.6759978>
- Rossi, C., et Chevrot, J-P. (2019). Uses and perceptions of Machine Translation at the European Commission. *The Journal of Specialised Translation*, 31, 177-200.
- Rouvroy, A. (2018). Homo juridicus est-il soluble dans les données ? In *Law, norms and freedom in cyberspace = Droit, normes et libertés dans le cybermonde : liber amicorum Yves Poulet* (pp. 417-444). Larcier.
- Rumsey, D. J. (2003). *Statistics for dummies* (2^e éd.). Wiley.
- Sager, J. C. (1989). Quality and Standards – The Evaluation of Translations, In C. Picken (Ed.), *The Translator's Handbook* (pp. 91-102). Aslib.
- Şahin, M. (2013). Using MT post-editing for translator training. In *Actes du colloque Tralogy II. Session 6 - Teaching around MT (Paris, 17-18 janvier 2013)*. <https://hal.science/hal-02497609>
- Saint-André, L. (2015). *Quelle formation donner aux traducteurs-postéditeurs de demain ?* [Thèse de doctorat en traductologie, Université d'Ottawa]. <http://dx.doi.org/10.20381/ruor-6777>
- Sakamoto, A. (2019). Why do many translators resist post-editing? A sociological analysis using Bourdieu's concepts. *The Journal of Specialised Translation*, 31, 201-216.

- Saldanha, G., et O'Brien, S. (2014). *Research Methodologies in Translation Studies*. Routledge.
- Salmi, L. (2021). A good servant but a bad master: Finnish Translators' Perceptions on Translation Technology. *Revista Tradumàtica. Tecnologies de la Traducció*, 19, 112-130. <https://doi.org/10.5565/rev/tradumatica.287>
- Samman, H. M. (2022). *Evaluating machine translation post-editing training in undergraduate translation programs - an exploratory study in Saudi Arabia* [Thèse de doctorat en Langues modernes, University of Southampton]. <http://eprints.soton.ac.uk/id/eprint/469163>
- Sánchez-Gijón, P., Moorkens, J., et Way, A. (2019). Post-editing neural machine translation versus translation memory segments. *Machine Translation*, 33, 31-59. <https://doi.org/10.1007/s10590-019-09232-x>
- Savenkov, K., et Lopez, M. (2022). The State of the Machine Translation 2022. In *Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas. Volume 2: Users and Providers Track and Government Track* (pp. 32-49). Association for Machine Translation in the Americas.
- Scansani, R., Bernardini, S., Ferraresi, A., et Bentivogli, L. (2019). Do translator trainees trust machine translation? An experiment on post-editing and revision. In *Proceedings of MT Summit XVII, Volume 2: Translator, Project and User Tracks* (pp. 73-79). International Association for Machine Translation.
- Scarpa, F. (2006). Corpus-based quality assessment of specialist translation: A study using parallel and comparable corpora in English and Italian. In M. Gotti et S. Šarčević (Eds), *Insights into specialized translation-linguistics insights* (pp. 155-172). Peter Lang.
- Scarpa, F. (2010). *La traduction spécialisée. Une approche professionnelle à l'enseignement de la traduction* (traduit et adapté par M. Fiola). Presses de l'Université d'Ottawa.
- Scheepers, T., et Schulz, P. (2017). *Interactive Neural Translation Assistance for Human Translators*. <https://thijs.ai/papers/scheepers-schulz-interactive-neural-translation.pdf>
- Schumacher, P. (2019). Avantages et limites de la post-édition. *Traduire*, 241, 108-123. <https://doi.org/10.4000/traduire.1887>
- Schumacher, P. (2020a, février). *La traduction automatique neuronale : révolution ou illusion ?* Communication présentée aux Formapéros Langue maternelle, langue étrangère, influence et transferts, Haute École de la Ville de Liège, Liège.
- Schumacher, P. (2020b). La traduction automatique neuronale : technologie révolutionnaire ou poudre de perlimpinpin ? Compte-rendu d'une expérience pédagogique. *Al-Kīmiyā*, 18, 67-89. <https://hdl.handle.net/2268/256446>

- Schumacher, P. (2020c). Post-édition et traduction humaine en contexte académique : une étude empirique. *Transletters: International Journal of Translation and Interpreting*, 3, 239-274. <https://hdl.handle.net/2268/250040>
- Schumacher, P., et Sutera, A. (2022). Analyse comparative de post-édition et de traduction humaine en contexte académique. In C. Expósito Castro, M. D. M. Ogea Pozo et F. Rodríguez Rodríguez (Eds.), *Theory and practice of translation as a vehicle for knowledge transfer. Théorie et pratique de la traduction comme véhicule de transfert des connaissances* (pp. 173-208). Editorial Universidad de Sevilla. <https://hdl.handle.net/2268/289924>
- Schumacher, P. (2023 - à paraître). Traduction humaine et post-édition : contrôle qualité en contexte académique. *Meta*, 68 (3).
- Screen, B. (2016). What does translation memory do to translation? The effect of translation memory output on specific aspects of the translation process. *Translation and Interpreting*, 8(1), 1-18. <http://dx.doi.org/10.12807/ti.108201.2016.a01>
- Screen, B. (2019). What effect does post-editing have on the translation product from an end- user's perspective? *The Journal of Specialised Translation*, 31, 133-157.
- Secară, A. (2005). Translation evaluation – a state of the art survey. In *Proceedings of the eCoLoRe/MeLLANGE Workshop* (pp. 39-44). Centre for Translation Studies, University of Leeds.
- Seleskovitch, D., et Lederer, M. (2014). *Interpréter pour traduire* (5^e éd.). Les Belles Lettres. <https://doi.org/10.4000/traduire.717>
- Senez, D. (1998). Post-editing Service for Machine Translation Users at the European Commission. In *Proceedings of Translating and the Computer 20*. Aslib.
- SFT commission Statistiques (2022). *Rapport de l'enquête 2022 sur les pratiques professionnelles en traduction* (Rapport – novembre 2022) [diapositives PowerPoint]. https://www.sft.fr/sites/default/files/2022-11/2022_SFT_resultats-enquete-statistiques-metiers-de-la-traduction.pdf
- Shih, C.-L. (2021). How to Empower Machine-Translation-to-Web Pre-Editing From the Perspective of Grice's Cooperative Maxims. *Theory and Practice in Language Studies*, 11 (12), 1554-1561. <https://doi.org/10.17507/tpls.1112.07>
- Shoshan, O. (2018, 19 avril). Make Neural Machine Translation Better, Faster – a New Way to Measure NMT Quality. *Slator. Language Industry Intelligence*. <https://slator.com/sponsored-content/make-neural-machine-translation-better-faster-a-new-way-to-measure-nmt-quality/>
- Shterionov, D., Superbo, R., Nagle, P. et al. (2018). Human versus automatic quality evaluation of NMT and PBSMT. *Machine Translation*, 32, 217-235. <https://doi.org/10.1007/s10590-018-9220-z>

- Shweta Bajpai, R.C., et Chaturvedi, H. K. (2015). Evaluation of Inter-Rater Agreement and Inter-Rater Reliability for Observational Data: An Overview of Concepts and Methods. *Journal of the Indian Academy of Applied Psychology*, 41 (3), 20-27.
- Simoulin, A., et Crabbé, B. (2021). Un modèle Transformer Génératif Pré-entraîné pour le _____ français (Generative Pre-trained Transformer in _____ (French) We introduce a French adaptation from the well-known GPT model). In *Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles. Vol 1 : conférence principale* (pp. 246–255). ATALA.
- Smolentceva, N. (2018, 5 décembre). DeepL: Cologne-based startup outperforms Google Translate. *Deutsche Welle*. <https://www.dw.com/en/deepl-cologne-based-startup-outperforms-google-translate/a-46581948>
- Snover, M., Dorr, B., Schwartz, R., Micciulla, L., et Makhoul, J. (2006). *A Study of Translation Edit Rate with Targeted Human Annotation*. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers* (pp. 223–231). Association for Machine Translation in the Americas.
- Snover, M., Dorr, B., Madnani, N., et Schwartz, R. (2009). Ter-Plus: Paraphrase, Semantic, and Alignment Enhancements to Translation Edit Rate. *Machine Translation*, 23, 117-127. <https://doi.org/10.1007/s10590-009-9062-9>
- Stasimioti, M., et Sosoni, V.(2019). Undergraduate Translation Students' Performance and Attitude vis-à-vis Machine Translation and Post-editing: Does Training Play a Role. In *Proceedings of the Translating and the Computer Conference* (pp. 125-136). Asling. https://www.asling.org/tc41/wp-content/uploads/TC41-Proceedings_125-136.pdf
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation*, 9(4). <https://doi.org/10.7275/96jp-xz07>
- Strubell, E., Ganesh, A., et McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3645-3650). Association for Computational Linguistics.
- Stymne, S., et Ahrenberg, L. (2012). On the practice of error analysis for machine translation evaluation. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (pp. 1785-1790). European Language Resources Association (ELRA).
- Sutskever, I., Vinyals, O., et Le, Q. (2014). Sequence to Sequence Learning with Neural Networks. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence et K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (pp. 3104-3112). <https://doi.org/10.48550/arXiv.1409.3215>

- Sycz-Opoń, J., et Gałuskińska, K. (2017). Machine Translation in the Hands of Trainee Translators – an Empirical Study. *Studies in Logic, Grammar and Rhetoric*, 49 (1), 195-212. <https://doi.org/10.1515/slgr-2017-0012>
- Tatsumi, M. (2010). *Post-Editing Machine Translated Text in A Commercial Setting: Observation and Statistical Analysis* [Thèse de doctorat en Langues appliquées et études interculturelles, Dublin City University]. <http://doras.dcu.ie/16062/>
- TAUS/CNGL (2010). *Consignes relatives à la post-édition des traductions automatiques*. Traduction de courtoisie Lexcelera. https://o.taus.net/index.php?option=com_rfiles&layout=preview&tmpl=component&path=Articles%2Ftaus-cngl-consignes-relatives-a-la-post-edition-des-traductions-automatiques.pdf
- Teich, E. (2003). *Cross-Linguistic Variation in System and Text: A Methodology for the Investigation of Translations and Comparable Texts*. Mouton de Gruyter. <https://doi.org/10.1515/9783110896541>
- Televic Education (2018). Manuel d'utilisation v4.1 AVIDAnet®Live. <https://support.televic-education.com/hc/en-us/articles/213859829--FR-Manuel-d-utilisation-AVIDAnet-Live>
- Templin, M. C. (1957). *Certain language skills in children; their development and interrelationships*. University of Minnesota Press.
- TERMIUM Plus® (2023). *La banque de données terminologiques et linguistiques du gouvernement du Canada*. Portail linguistique du Canada, Bureau de la traduction, Services publics et Approvisionnement Canada. <http://www.termiumplus.gc.ca>.
- Tezcan, A., Daems, J., et 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* (pp. 40-49). European Association for Machine Translation.
- To, V., Fan, S., et 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-edited: an Exacerbated Translationese. In *Proceedings of Machine Translation Summit XVII: Research Track* (pp. 273-281). European Association for Machine Translation.
- Toral, A., et Way, A. (2015). Machine-assisted translation of literary text: A case study. *Translation Spaces*, 4, 240-267. <https://doi.org/10.1075/ts.4.2.04tor>
- Toral, A., et Way, A. (2018). What Level of Quality Can Neural Machine Translation Attain on Literary Text? In J. Moorkens, S. Castilho, F. Gaspari, et S. Doherty (Eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 263-287). Springer. https://doi.org/10.1007/978-3-319-91241-7_12

- Toral, A., Castilho, S., et Hu, K. et Way, A. (2018). Attaining the Unattainable? Reassessing Claims of Human Parity in Neural Machine Translation. In *Proceedings of the Third Conference on Machine Translation: Research Papers* (pp. 113-123). Association for Computational Linguistics.
- Torruella, J., et Capsada, R. (2013). Lexical Statistics and Tipological Structures: A Measure of Lexical Richness. *Procedia - Social and Behavioral Sciences*, 95, 447-454. <https://doi.org/10.1016/j.sbspro.2013.10.668>
- Toudic, D., Hernandez Morin, K., Moreau, F., Barbin, F., et Phuez, G. (2014). Du contexte didactique aux pratiques professionnelles : proposition d'une grille multicritères pour l'évaluation de la qualité en traduction spécialisée. *ILCEA*, 19. <https://doi.org/10.4000/ilcea.2517>
- Toury, G. (1995). The Nature and Role of Norms in Translation. In G. Toury (Ed.), *Descriptive Translation Studies and Beyond* (pp. 53-69). John Benjamins Publishing Company.
- Toury, G. (2012). *Descriptive Translation Studies and Beyond* (2^e éd.). John Benjamins Publishing Company.
- Troiano, F., Permentiers, J., et Springael, E. (2000). *Traduction, adaptation et éditng multilingue : Mode d'emploi des services linguistiques et multimédias* (2^e éd.). Telos Communication Group (TCG).
- Ure, J. (1971). Lexical Density and Register Differentiation. In G. Perren et J. Trim (Eds.), *Applications of Linguistics* (pp. 443-452). Cambridge University Press.
- Van Egdom, G.-W., Verplaetse, H., Schrijver, I., Kockaert, H., Segers, W., Pauwels, J., Bloemen, H., Wylin, B., Huertas-Barros, E., Vandepitte, S., et Iglesias-Fernández, E. (2018). How to Put the Translation Test to the Test? On Preselected Items Evaluation and Perturbation. In *Quality Assurance and Assessment Practices in Translation and Interpreting* (pp. 26–56). IGI GLOBAL; Hershey, PA. <https://doi.org/10.4018/978-1-5225-5225-3>
- Van Egdom, G.-W., Vieira, L. N., et Absolon, J. (2018). Towards testing post-editing performance: A future-proof diagnostic tool. *Revista Tradumàtica*, 16, 114-124. <https://doi.org/10.5565/rev/tradumatica.211>
- van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI Ethics*, 1, 213-218. <https://doi.org/10.1007/s43681-021-00043-6>
- Vanmassenhove, E. (2020). On the Integration of Linguistic Features into Statistical and Neural Machine Translation. *ArXiv preprint*. <https://doi.org/10.48550/arXiv.2003.14324>
- Vanmassenhove, E., Hardmeier, C., et Andy, W. (2018). Getting Gender Right in Neural Machine Translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing EMNLP* (pp. 3003–3008). Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/D18-1334>

- Vanmassenhove, E., Shterionov, D., et Way, A. (2019). Lost in Translation: Loss and Decay of Linguistic Richness in Machine Translation. In *Proceedings of Machine Translation Summit XVII: Research Track* (pp. 222-232). European Association for Machine Translation.
- Vanmassenhove, E., Shterionov, D. et 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* (pp. 2203-2213). Association for Computational Linguistics.
- Vanroy, B., De Clercq, O., Tezcan, A., Daems, J., et 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* (pp. 259-294). Springer. https://doi.org/10.1007/978-3-030-69777-8_10
- Vasconcellos, M. (1986). Post-editing on-screen: machine translation from Spanish into English. In *Proceedings of the 8th Conference Translating and the Computer: A profession on the move (Londres, 13-14 novembre 1986)*. Aslib.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., et Polosukhin, I. (2017). Attention is All you Need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan et R. Garnett (Eds.), *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 6000-6010). Curran Associates.
- Vieira, L. N. (2017). From process to product: links between postediting effort and post-edited quality. In A. Lykke Jakobsen et B. Mesa-Lao (Eds.), *Translation in Transition: Between Cognition, Computing and Technology* (pp. 162-186). John Benjamins Publishing Company. <https://doi.org/10.1075/btl.133>
- Vieira, L. N. (2019). Post-Editing of Machine Translation. In M. O'Hagan (Ed.), *The Routledge Handbook of Translation and Technology* (pp. 319-335). Routledge. <https://doi.org/10.4324/9781315311258>
- Vieira, L. N. (2020). Automation anxiety and translators. *Translation Studies*, 13(1), 1-21.
- Vieira, L. N, Alonso, E., et Bywood, L. (2019). Introduction: Post-Editing in Practice – Process, Product and Networks. *The Journal of Specialised Translation*, 31. http://www.jostrans.org/issue31/issue31_toc.php
- Vogeleer S., et Béghin L. (2020). *Déverbaliser – reverbaliser : La traduction comme acte de violence ou comme manipulation du sens ?* Presses de l'Université Saint-Louis. <https://doi.org/10.4000/books.pusl.26559>
- Volkart, L. (2018). *Traduction automatique statistique vs. neuronale : Comparaison de MTH et DeepL à La Poste Suisse* [Mémoire de Maîtrise en Traduction, Université de Genève]. <https://archive-ouverte.unige.ch/unige:113749>

- Volkart, L., et 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* (pp. 71-79). European Association for Machine Translation.
- Volkart, L., Girletti, S., Gerlach, J., Mutal, J. D., et Bouillon, P. (2022). Source or target first? Comparison of two post-editing strategies with translation students. *Journal of Data Mining and Digital Humanities. Vers une robotique du traduire* [Hors-série]. <https://doi.org/10.46298/jdmdh.9067>
- Wagner, E. (1985). Post-editing Sysstran – A challenge for Commission Translators. *Terminologie et Traduction*, 3, 1-7.
- Way, A. (2013). Traditional and emerging use-cases for machine translation. In *Proceedings of Translating and the Computer 35* (pp. 1-12). Aslib.
- Way, A. (2018). Quality expectations of machine translation. In J. Moorkens, S. Castilho, F. Gaspari et S. Doherty (Eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 159-178). Springer.
- Way, A. (2019). *Machine translation: Where are we at today?* In E. Angelone, G. Massey et M. Ehrensberger-Dow (Eds.), *The Bloomsbury Companion to Language Industry Studies* (pp. 311-332). Bloomsbury Publishing. <http://dx.doi.org/10.5040/9781350024960.0018>
- Wilms, F. J. (1981). Von Susy zu Susy-BSA - Forderungen an ein anwenderbezogenes MU-System. *Sprache und Datenverarbeitung*, 1(2), 38-43.
- Wisniewski, G., Zhu, L., Ballier, N., et Yvon, F. (2021). Screening Gender Transfer in Neural Machine Translation. In *Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP* (pp. 311-321). Association for computational linguistics. <http://dx.doi.org/10.18653/v1/2021.blackboxnlp-1.24>
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *ArXiv preprint*. <https://doi.org/10.48550/arXiv.1609.08144>
- Yamada, M. (2015). Can college students be post-editors? An investigation into employing language learners in machine translation plus post-editing settings. *Machine Translation*, 29, 49-67. <https://doi.org/10.1007/s10590-014-9167-7>
- Yamada, M. (2019). The impact of Google Neural Machine Translation on Post-editing by student translators. *The Journal of Specialised Translation*, 31, 87-106.
- Yvon, F. (2019). Les deux voies de la traduction automatique. *Hermès, La Revue*, 85, 62-68. <https://doi.org/10.3917/herm.085.0062>

- Yvon, F. (2020, décembre). *Comprendre la traduction neuronale. Pour l'analyser, l'évaluer, la diagnostiquer.. et mieux la contrôler ?* Communication présentée au colloque CIUTI 2020 [en ligne].
- Zaretskaya, A., Vela, M., Corpas Pastor, G., et Seghiri, M. (2016). Measuring Post-editing Time and Effort for Different Types of Machine Translation Errors. *New Voices in Translation Studies*, 15, 63-92. <http://hdl.handle.net/2436/622208>
- Zehnalová, J. (2013). Tradition and Trends in Translation Quality Assessment. In J. Zehnalová, O. Molnár et M. Kubánek (Eds.), *Tradition and Trends in Trans-Language Communication* (pp. 41-58). Univerzita Palackého.
- Zeman, D., Fishel, M., Berka, J. et Bojar, O. (2011). Addicter: what is wrong with my translations? *Prague Bulletin of Mathematical Linguistics*, 96, 79-88. <http://dx.doi.org/10.2478/v10108-011-0013-2>
- 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* (pp. 1-19). Association for Computational Linguistics. <https://doi.org/10.18653/v1/K17-3001>
- Zhang, D., Kim, J., Crego, J., et Senellart, J. (2017). Boosting Neural Machine Translation. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing. Volume 2: Short Papers* (pp. 271-276). Asian Federation of Natural Language Processing.
- Zimina, M., et Gledhill, C. (2021). L'impact de la traduction automatique sur les pratiques langagières et professionnelles des apprentis-traducteurs : entre apports en efficacité et menaces pour la diversité des discours. In J.-C. Beacco, J. C. Herreras et Ch. Tremblay (Eds.), *Traduction automatique et usages sociaux des langues. Quelles conséquences pour la diversité linguistique ?* (pp. 63-82). Observatoire européen du plurilinguisme (OEP).

Annexe A

Textes sources et TA brutes

Texte source 1 – Article de presse

Can Theresa May and her government survive? Our writers' verdicts

With two resignations in one week, four Westminster watchers discuss whether the Conservative leader and her party can cling on to power

Andrew Gimson: There's no replacement for May nor appetite for election

Theresa May is in a stronger position than the press is willing to admit. Before the election, her frailties were ignored. Since that time, a narrative of extreme vulnerability has taken hold. Yet she remains in Downing Street, and the loss of two ministers who were found in different ways to have misbehaved does not change the powerful reasons for keeping her there. This summer she offered the British people the chance to turn her into an elected dictator, and they decided in their wisdom they would rather keep her as a prime minister who must operate with a degree of tact.

There is no popular demand for another election – the last one was an election too many – and no popular call for some particular individual to replace her. Nor does the Conservative party have a replacement for her in mind. On the great issue of the day, which is Brexit, the party is split. It recognises that the referendum decision must be implemented, but also that to implement it with gung-ho gusto, of the kind that Boris Johnson could provide, would be perilously divisive.

The time may well come when in order to achieve Brexit, a Gordian knot needs to be cut. But neither the country nor the party is ready for that. The clear preference is for May to continue with the difficult, perhaps impossible, task of disentangling the knot. Her duty is to carry on with this unenviable task. And since she is a dutiful woman, that is probably what she will do.

Source :<https://www.theguardian.com/commentisfree/2017/nov/09/can-theresa-may-conservative-government-survive-writers-verdict> (consulté le 23/01/2018)

Texte 1 – TA brute (Bing) générée le 23/01/2018 (prétest)

Theresa May et son gouvernement peuvent-ils survivre? Les verdicts de nos écrivains

Avec deux démissions en une semaine, quatre observateurs de Westminster discutent de la question de savoir si le chef conservateur et son parti peuvent s'accrocher au pouvoir

Andrew Gourmelon: il n'y a pas de remplacement pour le mai, ni l'appétit pour l'élection

Theresa May est dans une position plus forte que la presse est prête à admettre. Avant l'élection, ses faiblesses étaient ignorées. Depuis ce temps, un récit de vulnérabilité extrême s'est mis en attente. Pourtant, elle reste à Downing Street, et la perte de deux ministres qui ont été trouvés de différentes façons d'avoir mal agi ne change pas les raisons puissantes de la garder là. Cet été, elle a offert au peuple

britannique la chance de la transformer en un dictateur élu, et ils ont décidé dans leur sagesse qu'ils préfèrent la garder comme un premier ministre qui doit opérer avec un degré de tact.

Il n'y a pas de demande populaire pour une autre élection-la dernière était une élection de trop-et aucun appel populaire pour une personne en particulier pour la remplacer. Le parti conservateur n'a pas non plus de remplaçant pour elle à l'esprit. Sur le grand numéro de la journée, qui est Brexit, le parti est divisé. Il reconnaît que la décision référendaire doit être mise en œuvre, mais aussi que pour l'appliquer avec Gung-Ho Gusto, du genre que Boris Johnson pourrait fournir, serait dangereusement la Division.

Le temps peut bien venir quand afin d'atteindre Brexit, un noeud gordien doit être coupé. Mais ni le pays ni le parti n'est prêt pour cela. La préférence claire est pour mai de continuer avec la tâche difficile, peut-être impossible, de démêler le noeud. Son devoir est de poursuivre cette tâche peu enviable. Et comme elle est une femme dévouée, c'est probablement ce qu'elle va faire.

Texte 1 – TA brute (DeepL) générée le 23/01/2018 (prétest)

Theresa May et son gouvernement peuvent-ils survivre? Les verdicts de nos écrivains

Avec deux démissions en une semaine, quatre observateurs de Westminster se demandent si le chef du Parti conservateur et son parti peuvent s'accrocher au pouvoir...

Andrew Gimson:Il n' y a pas de remplaçant pour le mois de mai ni d'appétit pour les élections.

Theresa May est dans une position plus forte que la presse ne veut bien l'admettre. Avant les élections, ses faiblesses étaient ignorées. Depuis lors, un récit de vulnérabilité extrême s'est imposé. Pourtant, elle demeure à Downing Street, et la perte de deux ministres qui ont été retrouvés de différentes façons pour avoir mal agi ne change rien aux puissantes raisons de sa présence. Cet été, elle a offert au peuple britannique la chance de la transformer en dictateur élu, et ils ont décidé, dans leur sagesse, qu'ils préféreraient la garder comme premier ministre, qui doit agir avec tact.

Il n' y a pas de demande populaire pour une autre élection - la dernière était une élection en trop grand nombre - et il n' y a pas d'appel populaire pour qu'une personne en particulier la remplace. Le Parti conservateur n' a pas non plus de remplaçant pour elle. Sur la grande question du jour, qui est Brexit, le parti est divisé. Il reconnaît que la décision référendaire doit être mise en œuvre, mais aussi que sa mise en œuvre avec une hargne débridée, comme celle que Boris Johnson pourrait fournir, serait dangereusement discordante.

Le moment est peut-être venu de couper un nœud gordien pour réaliser Brexit. Mais ni le pays ni le parti ne sont prêts pour cela. La préférence est clairement donnée à ce que May poursuive la tâche difficile, voire impossible, de démêler le nœud. Son devoir est d'accomplir cette tâche peu enviable. Et puisque c'est une femme dévouée, c'est probablement ce qu'elle fera.

Texte 1 – TA brute (DeepL) générée le 03/09/2020 (test)

Theresa May et son gouvernement peuvent-ils survivre ? Les verdicts de nos écrivains

Avec deux démissions en une semaine, quatre observateurs de Westminster se demandent si la chef des conservateurs et son parti peuvent s'accrocher au pouvoir

Andrew Gimson : Il n'y a pas de remplacement pour le mois de mai ni d'appétit pour les élections

Theresa May est dans une position plus forte que ce que la presse est prête à admettre. Avant l'élection, ses faiblesses ont été ignorées. Depuis lors, un récit d'extrême vulnérabilité s'est installé. Pourtant, elle reste à Downing Street, et la perte de deux ministres qui se sont mal comportés de différentes manières ne change rien aux puissantes raisons qui l'ont poussée à rester là-bas. Cet été, elle a offert au peuple britannique la chance de la transformer en dictateur élu, et ils ont décidé, dans leur sagesse, qu'ils préféreraient la garder comme premier ministre qui doit agir avec un certain tact.

Il n'y a pas de demande populaire pour une autre élection - la dernière était une élection de trop - et pas d'appel populaire pour qu'une personne en particulier la remplace. Le parti conservateur n'a pas non plus de remplaçant en vue. Sur la grande question du jour, qui est Brexit, le parti est divisé. Il reconnaît que la décision du référendum doit être mise en œuvre, mais aussi que la mettre en œuvre avec un enthousiasme débordant, du type de celui que Boris Johnson pourrait apporter, serait dangereusement source de division.

Le moment pourrait bien venir où, pour réaliser le Brexit, il faudra couper un nœud gordien. Mais ni le pays ni le parti ne sont prêts pour cela. La préférence va clairement à May pour poursuivre la tâche difficile, voire impossible, de démêler le nœud. Son devoir est de poursuivre cette tâche peu enviable. Et comme elle est une femme dévouée, c'est probablement ce qu'elle fera.

Texte 1 – TA brute (Google Traduction) générée le 03/09/2020 (test)

Theresa May et son gouvernement peuvent-ils survivre? Verdicts de nos écrivains

Avec deux démissions en une semaine, quatre observateurs de Westminster discutent de la possibilité pour le chef conservateur et son parti de s'accrocher au pouvoir

Andrew Gimson: Il n'y a pas de remplaçant pour May ni d'appétit pour les élections

Theresa May est dans une position plus forte que la presse ne veut bien l'admettre. Avant les élections, ses faiblesses ont été ignorées. Depuis ce temps, un récit d'extrême vulnérabilité s'est installé. Pourtant, elle reste à Downing Street, et la perte de deux ministres qui se sont révélés avoir mal agi de différentes manières ne change pas les raisons puissantes de la garder là-bas. Cet été, elle a offert au peuple britannique la chance de la transformer en dictateur élu, et ils ont décidé dans leur sagesse qu'ils préféreraient la garder en tant que Premier ministre qui doit agir avec un certain tact.

Il n'y a pas de demande populaire pour une autre élection - la dernière était une élection de trop - et aucun appel populaire pour une personne en particulier pour la remplacer. Le Parti conservateur n'a pas non plus en tête une remplaçante pour elle. Sur la grande question du jour, qui est le Brexit, le parti est divisé. Il reconnaît que la décision référendaire doit être mise en œuvre, mais aussi que la mettre en œuvre avec enthousiasme, comme celui que Boris Johnson pourrait fournir, serait dangereusement source de division.

Le moment pourrait bien venir où, pour réaliser le Brexit, un nœud gordien doit être coupé. Mais ni le pays ni le parti ne sont prêts pour cela. Il est clairement préférable que May poursuive la tâche difficile, voire impossible, de démêler le nœud. Son devoir est de poursuivre cette tâche peu enviable. Et comme elle est une femme respectueuse, c'est probablement ce qu'elle fera.

Texte source 2 – thématique économique

How will the global economic order change by 2050?

Key findings

- We project that the world economy could more than double in size by 2050, assuming broadly growthfriendly policies (including no sustained long-term retreat into protectionism) and no major global civilisation-threatening catastrophes.
- Emerging markets will continue to be the growth engine of the global economy. By 2050, the E7 economies could have increased their share of world GDP from around 35% to almost 50%. China could be the largest economy in the world, accounting for around 20% of world GDP in 2050, with India in second place and Indonesia in fourth place (based on GDP at PPPs).
- A number of other emerging markets will also take centre stage – Mexico could be larger than the UK and Germany by 2050 in PPP terms and six of the seven largest economies in the world could be emerging markets by that time.
- Meanwhile, the EU27 share of world GDP could be down to less than 10% by 2050, smaller than India.
- We project Vietnam, India and Bangladesh to be three of the world's fastest growing economies over this period. UK growth has the potential to outpace the average rate in the EU27 after the transitional impact of Brexit has passed, although we project the fastest growing large EU economy to be Poland.
- Today's advanced economies will continue to have higher average incomes, but emerging economies should make good progress towards closing this gap by 2050. This will open up great opportunities for businesses prepared to make long-term investments in these markets. But this will require patience to ride out the storms we have seen recently in economies like, for example, Brazil, Nigeria and Turkey, all of which still have considerable long-term economic potential based on our analysis.
- To realise this growth potential, emerging market governments need to implement structural reforms to improve macroeconomic stability, diversify their economies away from undue reliance on natural resources (where this is currently the case), and develop more effective political and legal institutions.

Source: <https://www.pwc.com/gx/en/world-2050/assets/pwc-the-world-in-2050-full-report-feb-2017.pdf> (consulté le 12/01/2018)

Texte 2 – TA brute (Bing) générée le 23/01/2018 (prétest)

Comment l'ordre économique mondial changera-t-il de 2050?

Principales constatations

Nous projetons que l'économie mondiale pourrait plus que doubler dans la taille de 2050, en supposant que les politiques largement growthfriendly (y compris aucune retraite soutenue à long terme dans le protectionnisme) et aucune catastrophe majeure de civilisation mondiale menaçante.

Les marchés émergents resteront le moteur de croissance de l'économie mondiale. En 2050, les économies E7 auraient pu augmenter leur part du PIB mondial d'environ 35% à près de 50%. La Chine pourrait être la plus grande économie dans le monde, représentant environ 20% du PIB mondial en 2050, avec l'Inde en deuxième place et l'Indonésie en quatrième place (basé sur le PIB aux PPP).

Un certain nombre d'autres marchés émergents seront également en phase de centre-le Mexique pourrait être plus grand que le Royaume-Uni et l'Allemagne par 2050 en termes de PPP et six des sept plus grandes économies dans le monde pourrait être les marchés émergents à ce moment-là.

Entre-temps, la part de l'UE27 dans le PIB mondial pourrait être inférieure à 10% en 2050, plus petite que l'Inde.

Nous projetons le Vietnam, l'Inde et le Bangladesh pour être trois des économies les plus rapides au monde au cours de cette période. La croissance du Royaume-Uni pourrait dépasser le taux moyen dans l'UE27 après l'adoption de l'impact transitoire de la Brexit, bien que nous projetions la plus forte croissance de l'économie de l'UE en Pologne.

Les économies avancées actuelles continueront d'avoir des revenus moyens plus élevés, mais les économies émergentes devraient faire de bons progrès en vue de combler cet écart de 2050. Cela ouvrira de grandes possibilités aux entreprises prêtes à réaliser des investissements à long terme sur ces marchés. Mais cela nécessitera de la patience pour faire sortir les tempêtes que nous avons vu récemment dans des économies comme, par exemple, le Brésil, le Nigéria et la Turquie, qui ont tous encore un potentiel économique à long terme considérable basé sur notre analyse.

Pour réaliser ce potentiel de croissance, les gouvernements des marchés émergents doivent mettre en œuvre des réformes structurelles pour améliorer la stabilité macroéconomique, diversifier leurs économies et s'en remettre aux ressources naturelles (là où c'est le cas actuellement) et développer davantage institutions politiques et juridiques efficaces.

Texte 2 – TA brute (DeepL) générée le 23/01/2018 (prétest)

Comment l'ordre économique mondial changera-t-il d'ici 2050?

Principales conclusions

Nous prévoyons que la taille de l'économie mondiale pourrait plus que doubler d'ici 2050, en supposant des politiques globalement favorables à la croissance (y compris l'absence d'un repli durable à long terme

vers le protectionnisme) et l'absence de catastrophes mondiales majeures menaçant la civilisation mondiale.

Les marchés émergents continueront d'être le moteur de croissance de l'économie mondiale. D'ici 2050, les économies de l' E7 auraient pu faire passer leur part du PIB mondial d'environ 35 % à près de 50 %. La Chine pourrait être la plus grande économie du monde, représentant environ 20 % du PIB mondial en 2050, avec l'Inde en deuxième position et l'Indonésie en quatrième position (sur la base du PIB aux PPA).

Un certain nombre d'autres marchés émergents occuperont également une place centrale - le Mexique pourrait être plus grand que le Royaume-Uni et l'Allemagne d'ici 2050 en termes de PPP et six des sept plus grandes économies du monde pourraient être des marchés émergents d'ici là.

Dans le même temps, la part de l'UE27 dans le PIB mondial pourrait être inférieure à 10% d'ici 2050, soit un niveau inférieur à celui de l'Inde.

Nous prévoyons que le Vietnam, l'Inde et le Bangladesh deviendront au cours de cette période trois des économies les plus dynamiques du monde. La croissance du Royaume-Uni pourrait dépasser le taux moyen dans l'UE27 après la disparition de l'impact transitoire de Brexit, même si nous prévoyons que la Pologne sera la grande économie de l'UE connaissant la croissance la plus rapide.

Les économies avancées d'aujourd'hui continueront d'avoir des revenus moyens plus élevés, mais les économies émergentes devraient bien progresser vers la réduction de cet écart d'ici 2050. Cela ouvrira de grandes possibilités pour les entreprises prêtes à investir à long terme sur ces marchés. Mais cela exigera de la patience pour surmonter les tempêtes que nous avons vues récemment dans des économies comme, par exemple, le Brésil, le Nigeria et la Turquie, qui ont toutes encore un potentiel économique considérable à long terme sur la base de notre analyse.

Pour réaliser ce potentiel de croissance, les gouvernements des marchés émergents doivent mettre en œuvre des réformes structurelles afin d'améliorer la stabilité macroéconomique, de diversifier leurs économies et de développer des institutions politiques et juridiques plus efficaces.

Texte 2 – TA brute (DeepL) générée le 03/09/2020 (test)

Comment l'ordre économique mondial changera-t-il d'ici 2050 ?

Principales conclusions

Nous prévoyons que la taille de l'économie mondiale pourrait plus que doubler d'ici 2050, en supposant des politiques globalement favorables à la croissance (y compris l'absence de repli durable vers le protectionnisme) et l'absence de catastrophes majeures menaçant la civilisation mondiale.

Les marchés émergents continueront à être le moteur de la croissance de l'économie mondiale. D'ici 2050, les économies des pays de l'E7 pourraient avoir augmenté leur part du PIB mondial d'environ 35 % à près de 50 %. La Chine pourrait être la plus grande économie du monde, représentant environ 20 % du PIB mondial en 2050, avec l'Inde en deuxième position et l'Indonésie en quatrième position (sur la base du PIB aux PPA).

Un certain nombre d'autres marchés émergents occuperont également le devant de la scène : le Mexique pourrait être plus grand que le Royaume-Uni et l'Allemagne en termes de PPA d'ici 2050 et six des sept plus grandes économies du monde pourraient être des marchés émergents à cette date.

Dans le même temps, la part de l'UE-27 dans le PIB mondial pourrait être ramenée à moins de 10 % d'ici 2050, soit moins que l'Inde.

Nous prévoyons que le Viêt Nam, l'Inde et le Bangladesh seront trois des économies à la croissance la plus rapide du monde au cours de cette période. La croissance du Royaume-Uni pourrait dépasser le taux moyen de l'UE-27 une fois que l'impact transitoire de Brexit sera passé, bien que nous prévoyions que la grande économie de l'UE qui connaîtra la croissance la plus rapide sera la Pologne.

Les économies avancées d'aujourd'hui continueront à avoir des revenus moyens plus élevés, mais les économies émergentes devraient bien progresser pour combler cet écart d'ici 2050. Cela ouvrira de grandes possibilités aux entreprises prêtes à investir à long terme sur ces marchés. Mais cela exigera de la patience pour surmonter les tempêtes que nous avons vues récemment dans des économies comme, par exemple, le Brésil, le Nigeria et la Turquie, qui, d'après notre analyse, ont encore un potentiel économique considérable à long terme.

Pour réaliser ce potentiel de croissance, les gouvernements des marchés émergents doivent mettre en œuvre des réformes structurelles afin d'améliorer la stabilité macroéconomique, de diversifier leurs économies en évitant une dépendance excessive aux ressources naturelles (ce qui est le cas actuellement) et de mettre en place des institutions politiques et juridiques plus efficaces.

Texte 2 – TA brute (Google Traduction) générée le 03/09/2020 (test)

Comment l'ordre économique mondial changera-t-il d'ici 2050?

Principales conclusions

Nous prévoyons que l'économie mondiale pourrait plus que doubler de taille d'ici 2050, en supposant des politiques largement favorables à la croissance (y compris aucune retraite durable à long terme dans le protectionnisme) et aucune catastrophe mondiale majeure menaçant la civilisation.

Les marchés émergents continueront d'être le moteur de croissance de l'économie mondiale. D'ici 2050, les économies de l'E7 auraient pu augmenter leur part du PIB mondial d'environ 35% à près de 50%. La Chine pourrait être la plus grande économie du monde, représentant environ 20% du PIB mondial en 2050, l'Inde en deuxième position et l'Indonésie en quatrième place (sur la base du PIB en PPA).

Un certain nombre d'autres marchés émergents occuperont également le devant de la scène - le Mexique pourrait être plus grand que le Royaume-Uni et l'Allemagne d'ici 2050 en termes de PPA et six des sept plus grandes économies du monde pourraient être des marchés émergents d'ici là.

Dans le même temps, la part de l'UE27 dans le PIB mondial pourrait être réduite à moins de 10% d'ici 2050, soit moins que l'Inde.

Nous prévoyons que le Vietnam, l'Inde et le Bangladesh seront trois des économies les plus dynamiques du monde au cours de cette période. La croissance au Royaume-Uni a le potentiel de dépasser le taux

moyen dans l'UE27 après le passage de l'impact transitoire du Brexit, bien que nous prévoyons que la grande économie de l'UE à la croissance la plus rapide soit la Pologne.

Les économies avancées d'aujourd'hui continueront d'avoir des revenus moyens plus élevés, mais les économies émergentes devraient bien progresser pour combler cet écart d'ici 2050. Cela ouvrira de grandes opportunités aux entreprises prêtes à investir à long terme sur ces marchés. Mais cela exigera de la patience pour surmonter les tempêtes que nous avons vues récemment dans des économies comme, par exemple, le Brésil, le Nigéria et la Turquie, qui ont toutes encore un potentiel économique considérable à long terme basé sur notre analyse.

Pour réaliser ce potentiel de croissance, les gouvernements des marchés émergents doivent mettre en œuvre des réformes structurelles pour améliorer la stabilité macroéconomique, diversifier leurs économies afin d'éviter une dépendance excessive aux ressources naturelles (là où c'est actuellement le cas) et développer des institutions politiques et juridiques plus efficaces.

Texte source 3 – thématique scientifique

What is synesthesia?

When you eat chicken, does it feel pointy or round? Is a week shaped like a tipped-over D with the days arranged counterclockwise? Does the note B taste like horseradish? Do you get confused about appointments because Tuesday and Thursday have the same color? [...] When you read a newspaper or listen to someone speaking do you see a rainbow of colors? If so, you might have synesthesia.

Synesthesia is an anomalous blending of the senses in which the stimulation of one modality simultaneously produces sensation in a different modality. Synesthetes hear colors, feel sounds and taste shapes. What makes synesthesia different from drug-induced hallucinations is that synesthetic sensations are highly consistent: for particular synesthetes, the note F is always a reddish shade of rust, a 3 is always pink or truck is always blue.

The estimated occurrence of synesthesia ranges from rarer than one in 20,000 to as prevalent as one in 200. Of the various manifestations of synesthesia, the most common involves seeing monochromatic letters, digits and words in unique colors. This is called grapheme-color synesthesia.

Synesthetes report having unusually good memory for things such as phone numbers, security codes and polysyllabic anatomical terminology because digits, letters and syllables take on such a unique panoply of colors. But synesthetes also report making computational errors because 6 and 8 have the same color and claim to prejudge couples they meet because the colors of their first names clash so hideously.

For too long, synesthetes were dismissed as having overactive imaginations, confusing memories for perceptions or taking metaphorical speech far too literally. Recent research, however, has documented the reality of synesthesia and is beginning to make headway into understanding what might cause such unusual perceptions [...]

The causes of synesthesia also remain unknown. [...] Whatever its etiology, synesthesia provides cognitive neuroscientists with a unique opportunity to learn more about how the brain creates our perceptual reality.

Source : <https://www.scientificamerican.com/article/what-is-synesthesia/> (consulté le 23/01/2018)

Texte 3 – TA brute (Bing) générée le 23/01/2018 (prétest)

Qu'est-ce que synesthésie?

Quand tu manges du poulet, est-ce que ça te semble pointu ou rond? Est une semaine en forme comme un basculé sur D avec les jours disposés dans le sens inverse des aiguilles d'une montre? Est-ce que la note B a un goût de raifort? Avez-vous confus sur les rendez-vous parce que mardi et jeudi ont la même couleur? Quand vous lisez un journal ou écoutez quelqu'un parler, voyez-vous un arc-en-ciel de couleurs? Si oui, vous pourriez avoir synesthésie.

Synesthésie est un mélange anormal des sens dans lequel la stimulation d'une modalité produit simultanément la sensation dans une modalité différente. Synesthètes entendre les couleurs, sentir les sons et les formes de goût. Ce qui rend synesthésie différent des hallucinations induites par la drogue est que les sensations synesthésique sont très cohérentes: pour synesthètes particulier, la note F est toujours une nuance rougeâtre de rouille, un 3 est toujours rose ou le camion est toujours bleu.

L'occurrence estimée de synesthésie varie de plus rare qu'une en 20 000 à une valeur aussi fréquente que celle de 200. Des diverses manifestations de synesthésie, le plus commun consiste à voir des lettres monochromatiques, des chiffres et des mots dans des couleurs uniques. C'est ce qu'on appelle graphème-Color synesthésie.

Synesthètes rapport ayant une mémoire inhabituellement bonne pour des choses telles que les numéros de téléphone, les codes de sécurité et la terminologie anatomique polysyllabique parce que les chiffres, les lettres et les syllabes prendre une telle panoplie unique de couleurs. Mais synesthètes également signaler des erreurs de calcul parce que 6 et 8 ont la même couleur et de prétendre à préjuger les couples qu'ils rencontrent parce que les couleurs de leurs premiers noms se heurtent si affreusement.

Pendant trop longtemps, les synesthètes ont été rejetés comme ayant des imaginations suractives, confondant des mémoires pour des perceptions ou prenant le discours métaphorique beaucoup trop littéralement. La recherche récente, cependant, a documenté la réalité de synesthésie et commence à faire des progrès dans la compréhension de ce qui pourrait causer de telles perceptions inhabituelles [...]

Les causes de synesthésie restent également inconnues. [...] Quelle que soit son étiologie, synesthésie fournit aux neuroscientifiques cognitifs une occasion unique d'en apprendre davantage sur la façon dont le cerveau crée notre perception de la réalité.

Texte 3 – TA brute (DeepL) générée le 23/01/2018 (prétest)

Qu'est-ce que la synesthésie?

Lorsque vous mangez du poulet, vous sentez-vous pointu ou rond? Est-ce qu'une semaine est en forme de D avec les jours disposés dans le sens contraire des aiguilles d'une montre? La note B a un goût de raifort? Êtes-vous confus au sujet des rendez-vous parce que les mardis et jeudis ont la même couleur?

Quand vous lisez un journal ou écoutez quelqu'un parler, voyez-vous un arc-en-ciel de couleurs? Si c'est le cas, vous pourriez avoir une synesthésie.

La synesthésie est un mélange anormal des sens dans lequel la stimulation d'une modalité produit simultanément des sensations dans une modalité différente. Les synesthéses entendent les couleurs, ressentent les sons et goûtent les formes. Ce qui différencie la synesthésie des hallucinations médicamenteuses, c'est que les sensations synesthésiques sont très constantes: pour des synesthésies particulières, la note F est toujours une teinte rougeâtre de rouille, un 3 est toujours rose ou le camion est toujours bleu.

L'incidence estimée de la synesthésie varie de plus d'un cas sur 20 000 à un cas sur 200. Parmi les diverses manifestations de la synesthésie, la plus fréquente consiste à voir des lettres, des chiffres et des mots monochromes dans des couleurs uniques. C'est ce qu'on appelle la synesthésie de couleur du graphème.

Les synesthètes rapportent avoir une mémoire exceptionnellement bonne pour des choses comme les numéros de téléphone, les codes de sécurité et la terminologie anatomique polysyllabique parce que les chiffres, les lettres et les syllabes prennent sur une telle panoplie unique de couleurs. Mais les synesthètes rapportent aussi avoir fait des erreurs de calcul parce que 6 et 8 ont la même couleur et prétendent préjuger des couples qu'ils rencontrent parce que les couleurs de leurs prénoms s'affrontent si affreusement.

Pendant trop longtemps, les synesthètes ont été rejetés comme ayant une imagination débordante, confondant les souvenirs avec les perceptions ou prenant le discours métaphorique trop au pied de la lettre. Cependant, des recherches récentes ont documenté la réalité de la synesthésie et commencent à faire des progrès dans la compréhension de ce qui pourrait causer de telles perceptions inhabituelles [...]

Les causes de la synesthésie restent également inconnues. [...] Quelle que soit son étiologie, la synesthésie offre aux neuroscientifiques cognitifs une occasion unique d'en apprendre davantage sur la façon dont le cerveau crée notre réalité perceptuelle.

Texte 3 – TA brute (DeepL) générée le 03/09/2020 (test)

Qu'est-ce que la synesthésie?

Lorsque vous mangez du poulet, vous sentez-vous pointu ou rond? Une semaine a-t-elle la forme d'un «D» basculé avec les jours disposés dans le sens antihoraire? La note «B» a-t-elle le goût du raifort? Êtes-vous confus au sujet des rendez-vous parce que mardi et jeudi ont la même couleur? [...] Lorsque vous lisez un journal ou écoutez quelqu'un qui parle, voyez-vous un arc-en-ciel de couleurs? Si tel est le cas, vous pourriez avoir une synesthésie.

La synesthésie est un mélange anormal des sens dans lequel la stimulation d'une modalité produit simultanément une sensation dans une modalité différente. Les synesthètes entendent les couleurs, ressentent les sons et goûtent les formes. Ce qui différencie la synesthésie des hallucinations médicamenteuses, c'est que les sensations synesthésiques sont très cohérentes: pour des synesthésies particuliers, la note «F» est toujours une nuance rougeâtre de rouille, un «3» est toujours rose ou «camion» est toujours bleu.

L'occurrence estimée de la synesthésie varie de plus rare d'un sur 20 000 à aussi répandu que d'un sur 200. Parmi les diverses manifestations de la synesthésie, la plus courante consiste à voir des lettres, des chiffres et des mots monochromatiques dans des couleurs uniques. C'est ce qu'on appelle la synesthésie graphème-couleur.

Les synesthètes rapportent avoir une mémoire exceptionnellement bonne pour des choses telles que les numéros de téléphone, les codes de sécurité et la terminologie anatomique polysyllabique parce que les chiffres, les lettres et les syllabes prennent une panoplie de couleurs si unique. Mais les synesthètes rapportent aussi avoir fait des erreurs de calcul car «6» et «8» ont la même couleur et prétendent préjuger des couples qu'ils rencontrent parce que les couleurs de leurs prénoms se heurtent si affreusement.

Pendant trop longtemps, les synesthètes ont été rejetés comme ayant une imagination hyperactive, confondant les souvenirs pour les perceptions ou prenant le discours métaphorique beaucoup trop littéralement. Des recherches récentes, cependant, ont documenté la réalité de la synesthésie et commencent à progresser dans la compréhension de ce qui pourrait causer de telles perceptions inhabituelles.

[...]

Les causes de la synesthésie restent également inconnues. [...] Quelle que soit son étiologie, la synesthésie offre aux neuroscientifiques cognitifs une occasion unique d'en apprendre davantage sur la façon dont le cerveau crée notre réalité perceptive.

Texte 3 – TA brute (Google Traduction) générée le 03/09/2020 (test)

Qu'est-ce que la synesthésie ?

Lorsque vous mangez du poulet, est-ce que vous avez l'impression que le poulet est pointu ou rond ? Une semaine a-t-elle la forme d'un "D" renversé, les jours étant disposés dans le sens inverse des aiguilles d'une montre ? La note "B" a-t-elle un goût de raifort ? Vous confondez-vous avec les rendez-vous parce que le mardi et le jeudi ont la même couleur ? [...] Lorsque vous lisez un journal ou que vous écoutez quelqu'un parler, voyez-vous un arc-en-ciel de couleurs ? Si oui, vous avez peut-être une synesthésie.

La synesthésie est un mélange anormal des sens dans lequel la stimulation d'une modalité produit simultanément une sensation dans une autre modalité. Les synesthésiens entendent les couleurs, ressentent les sons et goûtent les formes. Ce qui différencie la synesthésie des hallucinations induites par la drogue est que les sensations synesthésiques sont très constantes : pour certains synesthésiens, la note "F" est toujours une nuance rougeâtre de rouille, un "3" est toujours rose ou "camion" est toujours bleu.

On estime que la fréquence des synesthésies varie de plus rare qu'une sur 20 000 à aussi fréquente qu'une sur 200. Parmi les diverses manifestations de la synesthésie, la plus courante consiste à voir des lettres, des chiffres et des mots monochromes dans des couleurs uniques. C'est ce qu'on appelle la synesthésie graphème-couleur.

Les synesthésiens déclarent avoir une mémoire exceptionnellement bonne pour des choses telles que les numéros de téléphone, les codes de sécurité et la terminologie anatomique polysyllabique, car les

chiffres, les lettres et les syllabes prennent une panoplie de couleurs uniques. Mais les synesthésiens déclarent également faire des erreurs de calcul parce que le "6" et le "8" ont la même couleur et prétendent préjuger des couples qu'ils rencontrent parce que les couleurs de leurs prénoms s'entrechoquent de façon si hideuse.

Pendant trop longtemps, les synesthésiens ont été considérés comme ayant une imagination débordante, confondant les souvenirs pour les perceptions ou prenant le discours métaphorique beaucoup trop au pied de la lettre. Cependant, des recherches récentes ont documenté la réalité de la synesthésie et commencent à faire des progrès dans la compréhension de ce qui peut causer ces perceptions inhabituelles

[...]

Les causes de la synesthésie restent également inconnues. [...] Quelle que soit son étiologie, la synesthésie offre aux neuroscientifiques cognitifs une occasion unique d'en apprendre davantage sur la façon dont le cerveau crée notre réalité perceptive.

Annexe B

Consignes

B.1 Consignes étudiantes

Consignes test

- ➔ Vous allez recevoir en tout 3 e-mails sur votre adresse «...@student.uliege.be»
- + Étape 1 : Un premier e-mail contenant 1 lien pour accéder au questionnaire « pré-expérience »
- + Étape 2 : Un deuxième e-mail confirmant votre participation au questionnaire et contenant 3 liens pour accéder aux trois tâches qui vous sont assignées
- + Étape 3 : Un troisième e-mail en cours de test contenant 1 lien pour accéder au questionnaire « post-expérience ».

Navigateur recommandé pour ce test : Google Chrome

Déroulement du test

+ Étape 1 : Questionnaire pré-expérience

Cliquez sur le lien correspondant pour ouvrir le questionnaire.
N'oubliez pas de **soumettre vos réponses** sur la dernière page du questionnaire.
Vous recevez un e-mail de confirmation vous indiquant que vos réponses ont bien été enregistrées.

+ Étape 2 : Réalisation des trois tâches dans MateCat

Démarrer une tâche

- a) Pour démarrer une des tâches qui vous est assignée, cliquez sur un des trois liens que vous avez reçus par e-mail.
- b) Vous pouvez commencer à traduire/post-éditer dans la colonne de droite.

!!! N'oubliez pas de **confirmer** chaque segment en appuyant simultanément sur **CTRL + Enter** ou en cliquant sur le bouton bleu « TRANSLATED » en bas à droite du segment avant de passer au segment suivant !!!

Pas d'inquiétude, il est encore possible de modifier un segment confirmé.

Terminer une tâche

- C) Une fois que vous avez terminé, vérifiez que **TOUS** les segments sont confirmés et que la barre de progression (en bas à gauche) affiche 100%. Si tel est le cas, fermez simplement la fenêtre et passez à la tâche suivante.

✚ Étape 3 : Questionnaire post-expérience

Cliquez sur le lien correspondant pour ouvrir le questionnaire.

N'oubliez pas de **soumettre vos réponses** sur la dernière page du questionnaire. Vous recevez un e-mail de confirmation vous indiquant que vos réponses ont bien été enregistrées. *Rem : ce questionnaire reste accessible jusqu'à 48h après le test. Si vous manquez de temps, vous ne devez donc pas impérativement le compléter lors du test.*

Un grand merci pour votre contribution et bon travail !

Consignes de post-édition complète (*full post-editing*)



Consignes pour obtenir un texte cible de qualité comparable ou égale à une traduction humaine

Ce niveau de qualité est en général défini comme compréhensible (un lecteur comprend parfaitement le contenu du message), exact (la traduction a le même sens que le texte source) et stylistiquement correct [...] La syntaxe est normale, la grammaire et la ponctuation sont correctes.

- Viser une traduction correcte au niveau grammatical, syntaxique et sémantique
- Vérifier que la terminologie importante est correctement traduite [...]
- Vérifier qu'aucune information n'a été accidentellement ajoutée ou oubliée
- Réviser le contenu injurieux, inapproprié ou culturellement inacceptable
- Exploiter au maximum le résultat brut de la traduction automatique
- Appliquer les règles fondamentales d'orthographe, de ponctuation et de coupure des mots
- Vérifier que la mise en forme est correcte

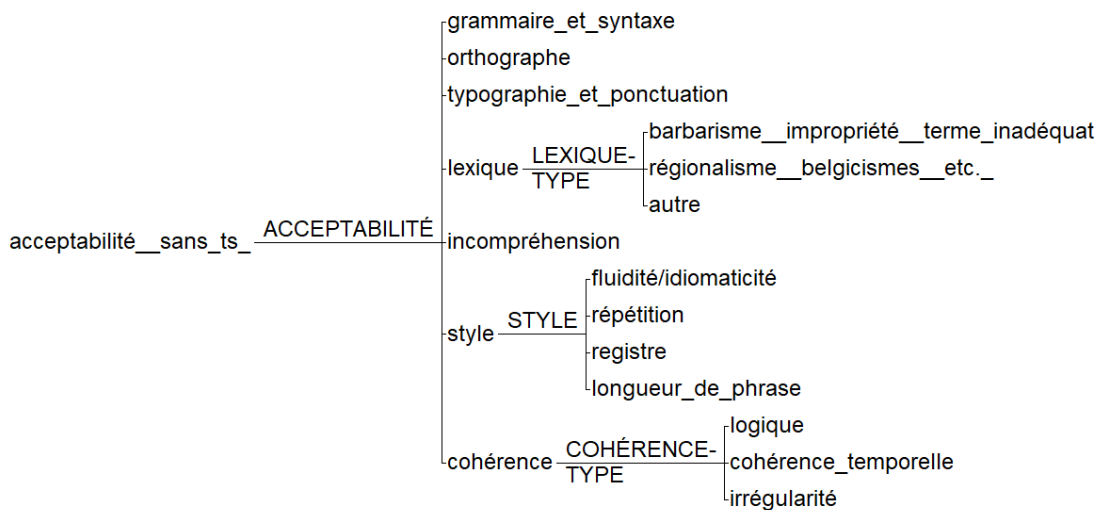
B.2 Consignes évaluatrices

Consignes détaillées pour l'évaluation humaine

TEST

Acceptabilité

Aperçu des sous-catégories



Définitions

Grammaire et syntaxe

Non-respect des règles de grammaire et de syntaxe.

► Note : À vous de juger s'il s'agit d'une faute de grammaire/orthographe/syntaxe (voir ci-dessous) ou d'une faute d'inattention, d'une coquille (voir plus loin « typographie et ponctuation »).

Grammaire

Accord

Faute d'accord (sujet, adjectif, attribut, pronom, verbe, participe passé...)

Ex. : Comme toi et moi le savent bien

Genre

Faute de genre (masculin/féminin)

Ex. : Il chercha un échappatoire

Syntaxe

Faute syntaxique : Solécisme, construction syntaxique non conforme à la grammaire ou à l'usage jugé correct (emploi incorrect d'une préposition, d'une conjonction, d'un verbe ; rupture syntaxique ; mauvais référent.).

Ex.: il est aussi grand **comme** moi ; certaines personnes ont plus de facilité en mathématiques. **Ils** s'en vantent.

Orthographe

Faute d'orthographe

Faute d'orthographe d'usage, majuscule manquante ou superflue, trait d'union manquant ou superflu, accent manquant ou superflu, etc.

► Note : Une erreur orthographique récurrente ne doit être soulignée qu'une seule fois par texte.

Typographie/ponctuation

Typographie

Non-respect des règles de typographie d'usage.

Coquille

Omission, addition, interversion ou substitution d'un ou de plusieurs caractères typographiques. Espace manquante ou superflue ; guillemets manquants ou mauvais type de guillemets (anglais "... " ou français « ... »), etc.

Ex.: L'inf~~r~~omation ; il a été promen~~é~~ chien ; il a été promen~~é~~ le **le** chien

Abréviation

Abréviation incorrecte

Etc.

Ponctuation

Faute de ponctuation, signe manquant ou superflu.

Remarque

► Note : Pour certains cas problématiques, c'est à vous de juger s'il s'agit d'une faute d'orthographe ou d'une erreur typographique. Ex. : « étique ». À vous d'estimer s'il s'agit-il d'une graphie intentionnelle (faute d'orthographe) ou non intentionnelle (faute typographique).

► Note : Une erreur typographique récurrente (ex. : utilisation de guillemets anglais au lieu de guillemets français) ne doit être soulignée qu'une seule fois par texte.

 **Lexique**

- **Barbarisme, impropiété ou terme inadéquat**

Barbarisme : Forme d'un mot qui n'existe pas en français ; déformation d'un mot.
Ex. : « abrévier » pour « abrégé »

Impropiété : Utilisation d'un mot existant mais selon une acception qu'il n'a pas.
Ex. : « recouvrir la vue » plutôt que « recouvrer la vue »

- **Régionalisme**

Belgicisme, québécoisisme, etc. jugé fautif ou inapproprié dans ce contexte.


- **Autre**

Merci de préciser en quoi cette erreur n'entre dans aucune autre catégorie.

 **Incompréhension**

Tournure/segment incompréhensible ou absurde qui entrave la compréhension du texte cible.

Ex. La Chine pourrait avoir la plus grande économie mondiale, avec environ 20% du PIB mondial en 2050, avec en deuxième place l'Inde et en quatrième place l'Indonésie (basé sur **le PIB en partenariat public-privé**).

 **Style**

- **Fluidité/Idiomatité**

Construction grammaticalement correcte, mais il y a un problème de fluidité : Mauvaise cooccurrence, tournure non idiomatique, collocation malheureuse, lourdeur, redondance, déformation de proverbes...

- **Répétition**

Un même mot/expression ou un mot/expression similaire est employé(e) trop souvent ou est trop proche de l'occurrence précédente de ce mot/expression.

► Note : Les cas de redondance sont à reprendre dans la catégorie « fluidité/idiomatité »

- **Registre**

Registre jugé trop formel/informel

- **Longueur de phrase**

Phrase trop longue ou trop courte

 **Cohérence**

- **Logique**

Problème de lien logique entre les idées, phrases, paragraphes.

- **Cohérence temporelle**

Problème de cohérence temporelle, mauvaise concordance des temps...

Ex. : Le nouveau livre de Michael Wolff accentue, au moyen d'anecdotes et de citations tapageuses, ce que l'on **a découvert** chaque jour depuis un an. →[...] ce que l'on **découvre** chaque jour depuis un an.

- **Irrégularité**

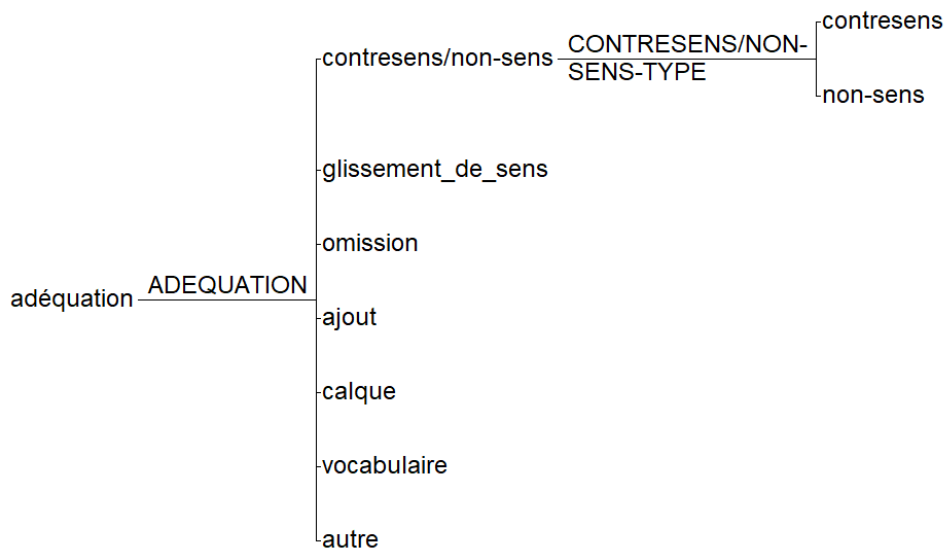
Irrégularité, manque d'uniformité dans les termes/tournures employés.

Dans le commentaire, veuillez noter le terme/tournure irrégulière et spécifier les variantes : « (terme1) (nombre d'occurrences) x ; (terme2) (nombre d'occurrences) x ».

Ex. : Commentaire : (les questionnaires d'évaluation) x 2 ; (les formulaires d'évaluation) x 1

Adéquation

Aperçu des sous-catégories



Définitions

- **Contresens/Non-sens**

- **Contresens**

« Faute de traduction qui consiste à attribuer à un segment du texte de départ un sens contraire à celui qu'a voulu exprimer l'auteur » (Delisle 2013 : 649)

Ex. : *The clear preference is for May to continue with the difficult, perhaps impossible, task of **disentangling the knot.***

→ Il y a une nette préférence pour que Mme May continue la tâche difficile, voire impossible, ~~de nouer le nœud.~~

= [...] la tâche difficile, voire impossible, **de dénouer le nœud.**

- **Non-sens**

« Faute de traduction qui consiste à attribuer à un segment du texte de départ un sens erroné qui a pour effet d'introduire dans le texte d'arrivée une formulation absurde » (Delisle 2013 : 672)

Ex. : *Trump, who does not care for government work, is more invested in his reputation as a creative writer, declaring more than once that "somebody said" that he is "the Hemingway of a hundred and forty characters".* [En référence à Twitter]

→ M. Trump, qui ne se soucie pas du travail gouvernemental, s'investit davantage dans sa réputation d'écrivain créatif, déclarant plus d'une fois que « quelqu'un a dit » qu'il est « ~~l'Hemingway des cent quarante personnages.~~ »

= [...] « l'Hemingway des cent quarante caractères ».

- **Glissement de sens**

« Faute de traduction qui consiste à ajouter ou à enlever à un segment du texte de départ une nuance de sens, ce qui altère le sens du texte » (Saint-Jacques 2011 : 39) sans pour autant mener à un contresens ou non-sens.

Ex. : *To near-universal shock, however, he won the election.*

→ Cependant, **contre toute attente**, il remporta les élections.

= Cependant, **à la surprise presque générale** [...]

- **Ajout**

Ajout

« Faute de traduction qui consiste à introduire de façon non justifiée dans le texte d'arrivée des éléments d'information superflus ou des effets stylistiques absents du texte de départ.

Ex. : *About 85% of beer sold is ale.*

→ Environ 85% de la bière vendue **dans les supermarchés** est de type ale ». (Delisle 2013 : 642)

► Note : Ne pas sanctionner l'explicitation justifiée !

- **Omission**

Omission

« Faute de traduction qui consiste à ne pas rendre dans le texte d'arrivée un élément de sens du texte de départ sans raison valable.

Ex. : *Two thirds of the customers (whose **monthly** consumption is less than 900 kWh) will receive a 0.5% increase per year.*

→ Les deux tiers des abonnés (ceux dont la consommation est inférieure à 900 kWh) subiront une hausse de 0,5 % par an.

= Les deux tiers des abonnés (ceux dont la consommation **mensuelle** est inférieure à 900 kWh) subiront une hausse de 0,5 % par an ». (Delisle 2013 : 673)

► Note : Ne pas sanctionner l'omission stratégique ni le recours justifié à l'implicite !



Calque

Calque lexical

Emprunt/traduction littérale jugée fautive ou injustifiée d'un mot ou syntagme du texte source. Le calque lexical respecte les structures syntaxiques de la langue française.

Ex. : 1. *We all know what a **sensor** is, right?* → Nous savons tous ce qu'est un **senseur**, n'est-ce pas ? = [...] un **capteur**, n'est-ce pas ?

2. *to control one's tears* → contrôler ses larmes = retenir, contenir ses larmes

Calque syntaxique

« Dans le cas d'un anglicisme syntaxique, on reproduit en français les éléments d'une structure syntaxique anglaise ». (Banque de dépannage linguistique 2002)

Ex. : 1. *to insist that* → insister que = insister sur le fait que (verbe intransitif)

2. *to fail an exam* → échouer un examen = échouer à un examen (verbe intransitif)



Vocabulaire

Terme/syntagme jugé non adéquat, trop précis ou imprécis (hyperonyme, hyponyme, vocabulaire inadéquat)



Autre

Merci de préciser en quoi cette erreur n'entre dans aucune autre catégorie.

Remarques générales

- Il est tout à fait possible qu'un même terme/segment doive être souligné au cours des deux phases d'évaluation.

Ex. : ***To near-universal shock**, however, he won the election.*

→ Cependant, **contre tout attente**, il remporta les élections.

= Acceptabilité → grammaire

= Adéquation → glissement de sens

- Si vous sentez, grâce à votre expérience, que l'étudiant.e a commis une erreur potentiellement imputable à une interférence avec le TS tel un calque fautif, mais que vous êtes seulement en phase 1 (acceptabilité) et que vous ne disposez donc pas du TS, veuillez annoter cette erreur dans la première PUIS dans la deuxième phase d'évaluation lorsque vous aurez le TS sous les yeux.

Ex. : *to fail an exam*

→ échouer un examen

= Acceptabilité → faute_syntaxique

= Adéquation → calque_syntaxique

Pour toute question, n'hésitez pas à me contacter à l'adresse : p.schumacher@uliege.be ou au 0479/406716

Encore merci pour votre précieuse contribution à ce projet et bonne évaluation !

Pondération des erreurs

Acceptabilité (erreurs de langue)

Catégorie	Grammaire/Syntaxe	Orthographe	Typographie/ponctuation	Lexique	Incompréhension	Style	Cohérence
Point de pénalité (sur 20)	1	1	0,2	0,3	1,3	0,3	0,3

Adéquation (erreurs de traduction)

Catégorie	Contresens/non-sens	Glissement de sens	Ajout	Omission	Calque	Vocabulaire	Autre
Point de pénalité (sur 20)	1,3	0,7	0,7	0,7	0,5	0,5	0,5

Note globale : note d'acceptabilité pour 40% + note d'adéquation pour 60%

Références

- Banque de dépannage linguistique. *Qu'est-ce qu'un anglicisme syntaxique ?* [En ligne]. Disponible sur : <http://bdl.oqlf.gouv.qc.ca/bdl/gabarit_bdl.asp?id=4059> (Consulté le 26/02/2018)
- 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]. Université de Gand, Faculté de Philosophie et Lettres, Gand, Belgique.
- Delisle, J., Fiola, M., Bastin, G., Farid, G., Francœur, A., Guilloton, N., Guyon, A., Le Blanc, C. et Marshman, E. (2013) *La traduction raisonnée, 3e édition : Manuel d'initiation à la traduction professionnelle de l'anglais vers le français*, University of Ottawa Press. Disponible sur : <http://www.jstor.org/stable/j.ctt6wr8bc> (Consulté le 10/07/2019)
- Delisle, J., Lee-Jahnke, H. et Cormier, M. C. (1999) *Terminologie de la traduction/Translation terminology/Terminología de la traducción/Terminologie der Übersetzung*. Amsterdam/Philadelphia, John Benjamin Publishing Company
- Saint-Jacques, F. (2011) *Étude comparative de trois traductions de Paradise Lost de l'anglais au français. Définition d'une méthodologie quantitative de l'équivalence en traduction littéraire* [Mémoire présenté dans le cadre du programme de maîtrise en traduction et terminologie pour l'obtention du grade de Maître ès arts (M.A.)], Université Laval, Faculté des Lettres, Québec, Canada.

Annexe C

Statistiques inférentielles – Prétest

Qualité globale – Acceptabilité

Multiple Comparisons of Means: Tukey Contrasts

```
tk <- multcomp::glht(mod, linfct=mcp(methode='Tukey', interaction_average=TRUE))
summary(tk)

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## TAS+PE - TAN+PE == 0 5.065 1.290 3.926 <0.001 ***
## TH - TAN+PE == 0 3.671 1.290 2.846 0.0122 *
## TH - TAS+PE == 0 -1.394 1.290 -1.080 0.5264
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 1 : Test de Tukey : Nombre erreurs_Mode (Acceptabilité)

```
summary(glht(mod, linfct=mcp(texte='Tukey', interaction_average=TRUE)))

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0 6.815 1.290 5.283 <1e-06 ***
## Texte 3 - Texte 1 == 0 -7.648 1.290 -5.929 <1e-06 ***
## Texte 3 - Texte 2 == 0 -14.463 1.290 -11.212 <1e-06 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 2 : Test de Tukey : Nombre erreurs_TS (Acceptabilité)

Qualité globale – Adéquation

```
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0 6.977 1.083 6.44 < 0.001 ***
## Texte 3 - Texte 1 == 0 3.380 1.083 3.12 0.00516 **
## Texte 3 - Texte 2 == 0 -3.597 1.083 -3.32 0.00258 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 3 : Test de Tukey : Nombre erreurs_TS (Adéquation)

Qualité globale – Notes (Acceptabilité)

Multiple Comparisons of Means: Tukey Contrasts

```
tk <- multcomp::glht(mod, linfct=mcp(methode='Tukey', interaction_average=TRUE))
summary(tk)
```

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## TAS+PE - TAN+PE == 0 -3.8319 0.6699 -5.720 <1e-04 ***
## TH - TAN+PE == 0 -3.0398 0.6699 -4.538 <1e-04 ***
## TH - TAS+PE == 0 0.7921 0.6699 1.182 0.464
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 4 : Test de Tukey : Notes_Mode (Acceptabilité)

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0 -1.7620 0.6699 -2.630 0.0232 *
## Texte 3 - Texte 1 == 0 2.7986 0.6699 4.177 <0.001 ***
## Texte 3 - Texte 2 == 0 4.5606 0.6699 6.808 <0.001 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 5 : Test de Tukey : Notes_TS (Acceptabilité)

Qualité globale – Notes (Adéquation)

```
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0 -5.5255 0.7996 -6.910 < 1e-04 ***
## Texte 3 - Texte 1 == 0 -3.2972 0.7996 -4.124 0.000104 ***
## Texte 3 - Texte 2 == 0 2.2282 0.7996 2.787 0.014635 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 6 : Test de Tukey : Notes_TS (Adéquation)

Qualité globale – Notes globales (Acceptabilité + Adéquation)

Multiple Comparisons of Means: Tukey Contrasts

```
tk <- multcomp::glht(mod, linfct=mcp(methode='Tukey', interaction_average=TRUE))
summary(tk)
```

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## TAS+PE - TAN+PE == 0 -2.1148 0.6185 -3.419 0.00179 **
## TH - TAN+PE == 0 -1.7551 0.6185 -2.838 0.01286 *
## TH - TAS+PE == 0 0.3597 0.6185 0.582 0.82995
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 7 : Test de Tukey : Notes globales_Mode

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0 -4.0403 0.6185 -6.533 <1e-04 ***
## Texte 3 - Texte 1 == 0 -0.8727 0.6185 -1.411 0.335
## Texte 3 - Texte 2 == 0 3.1676 0.6185 5.122 <1e-04 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 8 : Test de Tukey : Notes globales_TS

Grammaire et syntaxe

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## methode      78.101  39.051     2 44.000 12.8263 4.086e-05 ***
## texte        26.005  13.002     2 44.000  4.2707  0.02018 *
## methode:texte 14.273   3.568     4 37.115  1.1720  0.33890
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau C. 9 : résultats ANOVA Grammaire et syntaxe

```
tk <- multcomp::glht(mod, linfct=mcp(methode='Tukey', interaction_average=TRUE))
summary(tk)

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>|z|)
## TAS+PE - TAN+PE == 0  2.4676     0.4943  4.992 < 0.001 ***
## TH - TAN+PE == 0     1.5995     0.4943  3.236 0.00342 **
## TH - TAS+PE == 0    -0.8681     0.4943 -1.756 0.18451
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 10 : Test de Tukey – Gramm_Synt_Mode

```
summary(glht(mod, linfct=mcp(texte='Tukey', interaction_average=TRUE)))

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0 -0.9931     0.4943 -2.009  0.1100
## Texte 3 - Texte 1 == 0 -1.4051     0.4943 -2.843  0.0124 *
## Texte 3 - Texte 2 == 0 -0.4120     0.4943 -0.834  0.6821
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 11 : Test de Tukey – Gramm_Synt_TS

Orthographe

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## methode      20.8072  10.4036     2  44.000  16.6276 4.184e-06 ***
## texte        4.5476   2.2738     2  44.000   3.6341 0.03462 *
## methode:texte  1.1977   0.2994     4  37.286   0.4785 0.75123
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau C. 12 : résultats ANOVA Orthographe

```
tk <- multcomp::glht(mod, linfct=mcp(methode='Tukey', interaction_average=TRUE))
summary(tk)

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>|z|)
## TAS+PE - TAN+PE == 0  0.4583    0.2241   2.045 0.101591
## TH - TAN+PE == 0     1.2755    0.2241   5.692 < 1e-04 ***
## TH - TAS+PE == 0     0.8171    0.2241   3.647 0.000759 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 13 : Test de Tukey – Orthographe_Mode

```
summary(glht(mod, linfct=mcp(texte='Tukey', interaction_average=TRUE)))

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0  0.00463   0.22407   0.021 0.9998
## Texte 3 - Texte 1 == 0 -0.52083   0.22407  -2.324 0.0525 .
## Texte 3 - Texte 2 == 0 -0.52546   0.22407  -2.345 0.0498 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 14 : Test de Tukey – Orthographe_TS

Typographie et ponctuation

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## methode         2.670   1.335     2 44.000  0.4046   0.6697
## texte          97.334  48.667     2 44.000 14.7502 1.252e-05 ***
## methode:texte  10.049   2.512     4 36.919  0.7614   0.5571
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau C. 15 : résultats ANOVA Typographie et ponctuation

```
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0 -2.6134    0.5146 -5.079 <1e-04 ***
## Texte 3 - Texte 1 == 0 -2.1644    0.5146 -4.206 <1e-04 ***
## Texte 3 - Texte 2 == 0  0.4491    0.5146  0.873  0.658
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 16 : Test de Tukey – Typo_Ponctuation_TS

Lexique

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
## methode         7.1072   3.5536     2 44.000  1.8297 0.17245
## texte          17.6192   8.8096     2 44.000  4.5361 0.01618 *
## methode:texte  21.2888   5.3222     4 38.285  2.7404 0.04245 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau C. 17 : résultats ANOVA Lexique

Style et registre

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF  F value Pr(>F)
## methode      10.87   5.43     2 44.000   1.4202 0.2525
## texte       1314.12 657.06     2 44.000  171.6939 <2e-16 ***
## methode:texte  2.52   0.63     4 38.287   0.1646 0.9550
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau C. 18 : résultats ANOVA Style et registre

```
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0  6.5255     0.5542  11.775 <1e-09 ***
## Texte 3 - Texte 1 == 0 -3.6042     0.5542  -6.504 <1e-09 ***
## Texte 3 - Texte 2 == 0 -10.1296     0.5542 -18.279 <1e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 19 : Test de Tukey Style_registre_TS

Cohérence

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF  F value    Pr(>F)
## methode      30.318  15.159     2 44.000   6.0403 0.004809 **
## texte       288.871 144.436     2 44.000  57.5514 5.237e-13 ***
## methode:texte 10.374   2.594     4 37.174   1.0334 0.402930
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau C. 20 : résultats ANOVA Cohérence

```
tk <- multcomp::glht(mod, linfct=mcp(methode='Tukey', interaction_average=TRUE))
summary(tk)

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## TAS+PE - TAN+PE == 0 1.4884 0.4488 3.317 0.00267 **
## TH - TAN+PE == 0 0.3403 0.4488 0.758 0.72862
## TH - TAS+PE == 0 -1.1481 0.4488 -2.558 0.02827 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 21 : Test de Tukey Cohérence_Mode

```
summary(glht(mod, linfct=mcp(texte='Tukey', interaction_average=TRUE)))

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0 4.6435 0.4488 10.347 <0.001 ***
## Texte 3 - Texte 1 == 0 1.2199 0.4488 2.718 0.0179 *
## Texte 3 - Texte 2 == 0 -3.4236 0.4488 -7.629 <0.001 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 22 : Test de Tukey Cohérence_TS

Catégories en phase d'adéquation

Contresens et non-sens

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)

## boundary (singular) fit: see ?isSingular

anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## methode      7.286   3.643     2    66  2.3577    0.1026
## texte       83.974  41.987     2    66 27.1735 2.458e-09 ***
## methode:texte 7.749   1.937     4    66  1.2538    0.2971
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau C. 23 : résultats ANOVA Contresens et non-sens

```
summary(glht(mod, linfct=mcp(texte='Tukey', interaction_average=TRUE)))

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0  2.5486     0.3521  7.238 <0.001 ***
## Texte 3 - Texte 1 == 0  1.7014     0.3521  4.832 <0.001 ***
## Texte 3 - Texte 2 == 0 -0.8472     0.3521 -2.406 0.0427 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 24 : Test de Tukey CS_NS_TS

Glissement de sens

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)

anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## methode     12.538   6.2690     2 44.00  3.3669  0.04359 *
## texte       43.007  21.5034     2 44.00 11.5488 9.298e-05 ***
## methode:texte 16.519   4.1297     4 38.68  2.2179  0.08494 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau C. 25 : résultats ANOVA Glissement de sens

```
tk <- multcomp::glht(mod, linfct=mcp(methode='Tukey', interaction_average=TRUE))
summary(tk)

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## TAS+PE - TAN+PE == 0 0.4676 0.3865 1.210 0.4474
## TH - TAN+PE == 0 1.0023 0.3865 2.593 0.0257 *
## TH - TAS+PE == 0 0.5347 0.3865 1.383 0.3496
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 26 : Test de Tukey GS_Mode

```
summary(glht(mod, linfct=mcp(texte='Tukey', interaction_average=TRUE)))

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0 0.01389 0.38654 0.036 0.999
## Texte 3 - Texte 1 == 0 1.61574 0.38654 4.180 <1e-04 ***
## Texte 3 - Texte 2 == 0 1.60185 0.38654 4.144 <1e-04 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 27 : Test de Tukey GS_TS

Ajout/Omission

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)

## boundary (singular) fit: see ?isSingular

anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
## Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## methode 1.19097 0.59549 2 66 1.1762 0.3148
## texte 0.99866 0.49933 2 66 0.9863 0.3784
## methode:texte 1.36504 0.34126 4 66 0.6741 0.6123
```

Tableau C. 28 : résultats ANOVA Ajout/Omission

Calque

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## methode      134.74   67.372     2 44.000  9.2993 0.0004281 ***
## texte        465.00  232.498     2 44.000 32.0915 2.538e-09 ***
## methode:texte  41.02   10.255     4 37.783  1.4155 0.2475806
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau C. 29 : résultats ANOVA Calque

```
tk <- multcomp::glht(mod, linfct=mcp(methode='Tukey', interaction_average=TRUE))
summary(tk)

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
##   data = df)
##
## Linear Hypotheses:
##           Estimate Std. Error z value Pr(>|z|)
## TAS+PE - TAN+PE == 0  1.3866   0.7625  1.819  0.1635
## TH - TAN+PE == 0     -1.8889   0.7625 -2.477  0.0354 *
## TH - TAS+PE == 0     -3.2755   0.7625 -4.296 <0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 30 : Test de Tukey Calque_Mode

```
summary(glht(mod, linfct=mcp(texte='Tukey', interaction_average=TRUE)))

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
##   data = df)
##
## Linear Hypotheses:
##           Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0  5.5787   0.7625  7.317 <1e-04 ***
## Texte 3 - Texte 1 == 0  0.6343   0.7625  0.832  0.683
## Texte 3 - Texte 2 == 0 -4.9444   0.7625 -6.485 <1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 31 : Test de Tukey Calque_TS

Vocabulaire

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)

## boundary (singular) fit: see ?isSingular
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF DenDF F value  Pr(>F)
## methode      4.3137  2.1568     2    66  1.6286 0.203987
## texte       18.6358  9.3179     2    66  7.0359 0.001699 **
## methode:texte 25.0905  6.2726     4    66  4.7364 0.002021 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau C. 32 : résultats ANOVA Vocabulaire

```
summary(glht(mod, linfct=mcp(texte='Tukey', interaction_average=TRUE)))

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0 -1.1782    0.3260 -3.614 <0.001 ***
## Texte 3 - Texte 1 == 0 -0.3056    0.3260 -0.937  0.6167
## Texte 3 - Texte 2 == 0  0.8727    0.3260  2.677  0.0204 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau C. 33 : Test de Tukey Vocabulaire_TS

Annexe D

Statistiques inférentielles – Test

Qualité globale – Acceptabilité

```
mod2 <- lmerTest::lmer(note ~ methode + texte + (1|Etudiant), data=df)
tk <- multcomp::glht(mod2, linfct=mcp(methode='Tukey', interaction_average=TRUE))
summary(tk)

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode + texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## TAN(G) + PE - TAN(D) + PE == 0 2.2433 0.8607 2.606 0.02478 *
## TH - TAN(D) + PE == 0 2.5992 0.8607 3.020 0.00711 **
## TH - TAN(G) + PE == 0 0.3559 0.8474 0.420 0.90737
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 1 : Nombre erreurs_Mode (Acceptabilité)

```
summary(glht(mod2, linfct=mcp(texte='Tukey', interaction_average=TRUE)))

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode + texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0 5.5351 0.7734 7.157 < 0.001 ***
## Texte 3 - Texte 1 == 0 7.9356 0.7733 10.262 < 0.001 ***
```

2

```
## Texte 3 - Texte 2 == 0 2.4005 0.7733 3.104 0.00543 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 2 Nombre erreurs_TS (Acceptabilité)

Qualité globale – Adéquation

```
## Fit: lmerTest::lmer(formula = note ~ methode * texte + (1 | Etudiant),
##   data = df)
##
## Linear Hypotheses:
##               Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0   3.3966    0.8606   3.947 0.000237 ***
## Texte 3 - Texte 1 == 0   7.9677    0.8605   9.259 < 1e-04 ***
## Texte 3 - Texte 2 == 0   4.5711    0.8606   5.311 < 1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 3 : Nombre erreurs_TS (Adéquation)

Qualité globale – Notes (Acceptabilité)

```
mod2 <- lmerTest::lmer(note ~ methode + texte + (1|Etudiant), data=df)
tk <- multcomp::glht(mod2, linfct=mcp(methode='Tukey', interaction_average=TRUE))
summary(tk)
```

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode + texte + (1 | Etudiant),
##   data = df)
##
## Linear Hypotheses:
##               Estimate Std. Error z value Pr(>|z|)
## TAN(G) + PE - TAN(D) + PE == 0  -0.7996    0.5250  -1.523 0.28004
## TH - TAN(D) + PE == 0          -1.7509    0.5250  -3.335 0.00246 **
## TH - TAN(G) + PE == 0          -0.9513    0.5195  -1.831 0.15936
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 4 : Notes_Mode (Acceptabilité)

```
summary(glht(mod2, linfct=mcp(texte='Tukey', interaction_average=TRUE)))
```

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode + texte + (1 | Etudiant),
##   data = df)
##
## Linear Hypotheses:
##               Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0  -2.7938    0.4870  -5.737 < 0.001 ***
## Texte 3 - Texte 1 == 0  -4.4959    0.4869  -9.233 < 0.001 ***
```

2

```
## Texte 3 - Texte 2 == 0  -1.7022    0.4869  -3.496 0.00139 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 5 : Notes_TS (Acceptabilité)

Grammaire et syntaxe

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)

## boundary (singular) fit: see ?isSingular
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## methode      37.170 18.5851     2    63  8.4654 0.0005543 ***
## texte       56.298 28.1492     2    63 12.8217 2.13e-05 ***
## methode:texte  5.321  1.3302     4    63  0.6059 0.6598252
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau D. 6 : résultats ANOVA Grammaire et syntaxe

```
mod2 <- lmerTest::lmer(note ~ methode + texte + (1|Etudiant), data=df)

## boundary (singular) fit: see ?isSingular
tk <- multcomp::glht(mod2, linfct=mcp(methode='Tukey', interaction_average=TRUE))
summary(tk)

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode + texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>|z|)
## TAN(G) + PE - TAN(D) + PE == 0  0.0637   0.4238  0.150 0.98762
## TH - TAN(D) + PE == 0          1.5515   0.4238  3.661 0.00076 ***
## TH - TAN(G) + PE == 0          1.4878   0.4238  3.511 0.00128 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 7 : Test de Tukey – Gramm_Synt_Mode

```
## Fit: lmerTest::lmer(formula = note ~ methode + texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0  2.0881   0.4238  4.927 < 1e-04 ***
## Texte 3 - Texte 1 == 0  1.6475   0.4238  3.888 0.000279 ***
## Texte 3 - Texte 2 == 0 -0.4407   0.4238 -1.040 0.551671
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 8 : Test de Tukey – Gramm_Synt_TS

Orthographe

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)

## boundary (singular) fit: see ?isSingular

anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## methode      0.59613 0.29807     2    63  0.9637 0.3870
## texte        0.49142 0.24571     2    63  0.7944 0.4563
## methode:texte 1.93128 0.48282     4    63  1.5610 0.1957
```

Tableau D. 9 : résultats ANOVA Orthographe

Typographie et ponctuation

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## methode      24.324 12.1621     2 57.766  8.1393 0.0007688 ***
## texte        48.615 24.3077     2 40.329 16.2674 6.605e-06 ***
## methode:texte 11.984  2.9961     4 59.535  2.0050 0.1053543
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau D. 10 : résultats ANOVA Typographie et ponctuation

```
mod2 <- lmerTest::lmer(note ~ methode + texte + (1|Etudiant), data=df)
tk <- multcomp::glht(mod2, linfct=mcp(methode='Tukey', interaction_average=TRUE))
summary(tk)

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode + texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>|z|)
## TAN(G) + PE - TAN(D) + PE == 0  1.1284  0.3929  2.872  0.0114 *
## TH - TAN(D) + PE == 0          1.4672  0.3929  3.735  <0.001 ***
## TH - TAN(G) + PE == 0          0.3389  0.3892  0.871  0.6588
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 11 : Test de Tukey – Typographie&ponctuation_Mode

```
summary(glht(mod2, linfct=mcp(texte='Tukey', interaction_average=TRUE)))

##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode + texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0 1.9831 0.3671 5.402 <0.001 ***
## Texte 3 - Texte 1 == 0 0.9503 0.3671 2.588 0.0261 *
```

2

```
## Texte 3 - Texte 2 == 0 -1.0329 0.3671 -2.814 0.0135 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 12 : Test de Tukey – Typographie&ponctuation_TS

Lexique

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)

## boundary (singular) fit: see ?isSingular

anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
## Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## methode 4.152 2.076 2 63 1.1699 0.3170
## texte 110.943 55.471 2 63 31.2606 3.712e-10 ***
## methode:texte 3.577 0.894 4 63 0.5040 0.7329
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau D. 13 : résultats ANOVA Lexique

```
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode + texte + (1 | Etudiant),
##   data = df)
##
## Linear Hypotheses:
##               Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0  -1.0812    0.3798  -2.847  0.0123 *
```

2

```
## Texte 3 - Texte 1 == 0  1.9607    0.3798  5.163  <0.001 ***
## Texte 3 - Texte 2 == 0  3.0419    0.3798  8.009  <0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 14 : Test de Tukey – Lexique_TS

Incompréhension

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)

## boundary (singular) fit: see ?isSingular
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##
##          Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## methode    1.0910  0.5455     2    63  0.9012    0.4112
## texte     18.0910  9.0455     2    63 14.9440 4.881e-06 ***
## methode:texte  4.2901  1.0725     4    63  1.7719    0.1456
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau D. 15 : résultats ANOVA Incompréhension

```
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode + texte + (1 | Etudiant),
##   data = df)
##
## Linear Hypotheses:
##               Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0  0.02967    0.23031   0.129   0.991
```

2

```
## Texte 3 - Texte 1 == 0  1.11169    0.23031   4.827 3.62e-06 ***
## Texte 3 - Texte 2 == 0  1.08202    0.23031   4.698 7.23e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 16 : Test de Tukey – Incomprehension_TS

Style

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)

## boundary (singular) fit: see ?isSingular
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##               Sum Sq Mean Sq NumDF DenDF F value  Pr(>F)
## methode         14.271   7.135     2    63  2.1341 0.126826
## texte          101.483  50.741     2    63 15.1765 4.17e-06 ***
## methode:texte   50.886  12.722     4    63  3.8050 0.007812 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau D. 17 : résultats ANOVA Style

Cohérence

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)

## boundary (singular) fit: see ?isSingular
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## methode      1.725   0.863     2    63  1.4166   0.2502
## texte      66.398  33.199     2    63 54.5216 1.806e-14 ***
## methode:texte  1.481   0.370     4    63  0.6080   0.6583
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau D. 18 : résultats ANOVA Cohérence

```
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode + texte + (1 | Etudiant),
##   data = df)
##
## Linear Hypotheses:
##           Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0  1.6038     0.2232   7.186 < 1e-04 ***
```

2

```
## Texte 3 - Texte 1 == 0 -0.7112     0.2232  -3.186  0.00404 **
## Texte 3 - Texte 2 == 0 -2.3150     0.2232 -10.372 < 1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 19 : Test de Tukey – Cohérence_TS

Catégories en phase d'adéquation

Contresens et non-sens

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF  DenDF F value  Pr(>F)
## methode      10.734   5.3670     2  53.047   5.9866 0.004525 **
## texte        14.714   7.3571     2  40.138   8.2065 0.001028 **
## methode:texte 12.807   3.2018     4  54.845   3.5715 0.011668 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau D. 20 : résultats ANOVA Contresens et non-sens

Glissement de sens

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)

## boundary (singular) fit: see ?isSingular
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF  DenDF F value  Pr(>F)
## methode       8.0915   4.0458     2    63   2.2226 0.11676
## texte         4.8716   2.4358     2    63   1.3381 0.26969
## methode:texte 19.3703   4.8426     4    63   2.6603 0.04068 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau D. 21 : résultats ANOVA Glissement de sens

Omission

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF  DenDF F value  Pr(>F)
## methode      10.3212   5.1606     2  61.737 15.3778 3.808e-06 ***
## texte        11.0221   5.5111     2  42.126 16.4222 5.335e-06 ***
## methode:texte  9.4002   2.3500     4  62.531  7.0028 0.0001005 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau D. 22 : résultats ANOVA Omission

Ajout

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF  DenDF F value  Pr(>F)
## methode    2.0388  1.01942     2  61.980  4.5012 0.01495 *
## texte      1.2083  0.60416     2  41.164  2.6677 0.08140 .
## methode:texte 1.4571  0.36429     4  62.644  1.6085 0.18326
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau D. 23 : résultats ANOVA Ajout

```
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode + texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
##           Estimate Std. Error z value Pr(>|z|)
## TAN(G) + PE - TAN(D) + PE == 0 -0.1135    0.1431  -0.794  0.7070
## TH - TAN(D) + PE == 0          0.2744    0.1431   1.918  0.1336
## TH - TAN(G) + PE == 0          0.3879    0.1427   2.719  0.0179 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 24 : Test de Tukey – Ajout_Mode

Calque

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)

## boundary (singular) fit: see ?isSingular
anova(mod)

## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF  DenDF F value  Pr(>F)
## methode   108.05   54.03     2    63  9.2822 0.0002931 ***
## texte     753.33  376.67     2    63 64.7140 5.306e-16 ***
## methode:texte 44.74   11.19     4    63  1.9217 0.1177707
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau D. 25 : résultats ANOVA Calque


```
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode + texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## TAN(G) + PE - TAN(D) + PE == 0 0.1710 0.7172 0.238 0.96914
## TH - TAN(D) + PE == 0 -2.4276 0.7172 -3.385 0.00213 **
## TH - TAN(G) + PE == 0 -2.5986 0.7172 -3.623 < 0.001 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 26 : Test de Tukey – Calque_Mode

```
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
```

2

```
## Fit: lmerTest::lmer(formula = note ~ methode + texte + (1 | Etudiant),
## data = df)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0 1.6143 0.7172 2.251 0.0629 .
## Texte 3 - Texte 1 == 0 7.6239 0.7172 10.630 <1e-04 ***
## Texte 3 - Texte 2 == 0 6.0096 0.7172 8.379 <1e-04 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 27 : Test de Tukey – Calque_TS

Vocabulaire

```
mod <- lmerTest::lmer(note ~ methode*texte + (1|Etudiant), data=df)
anova(mod)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF F value  Pr(>F)
## methode      9.7335  4.8668     2 61.284  3.6755 0.031077 *
## texte       15.1067  7.5534     2 41.745  5.7046 0.006454 **
## methode:texte 5.1339  1.2835     4 62.283  0.9693 0.430767
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tableau D. 28 : résultats ANOVA Vocabulaire

```
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode + texte + (1 | Etudiant),
##   data = df)
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>|z|)
## TAN(G) + PE - TAN(D) + PE == 0  0.88495   0.34521   2.564   0.0281 *
## TH - TAN(D) + PE == 0          0.81519   0.34521   2.361   0.0479 *
## TH - TAN(G) + PE == 0         -0.06977   0.34490  -0.202   0.9777
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 29 : Test de Tukey – Vocabulaire_Mode

```
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmerTest::lmer(formula = note ~ methode + texte + (1 | Etudiant),
##   data = df)
##
## Linear Hypotheses:
##              Estimate Std. Error z value Pr(>|z|)
## Texte 2 - Texte 1 == 0  -0.4708    0.3429  -1.373  0.35512
## Texte 3 - Texte 1 == 0   0.6561    0.3429   1.913  0.13499
```

2

```
## Texte 3 - Texte 2 == 0   1.1269    0.3429   3.286  0.00297 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Tableau D. 30 : Test de Tukey – Vocabulaire_TS

Annexe E

Mesures linguistiques automatiques

```

Lexical Diversity (2018):
-----

Tokens in HT:  9778
Tokens in TAN: 9712
Tokens in TAS: 9565

Types in HT:   954
Types in TAN:  658
Types in TAS:  728

TTR in HT:     0.09756596440989977
TTR in TAN:    0.06775123558484349
TTR in TAS:    0.07611082070047047

RTTR in HT:    9.64769039962645
RTTR in TAN:   6.676849033400936
RTTR in TAS:   7.443700522585692

CTTR in HT:    6.821947304364216
CTTR in TAN:   4.721245228476647
CTTR in TAS:   5.26349111664219

MSTTR in HT:   0.37726315789473686
MSTTR in TAN:  0.35547368421052633
MSTTR in TAS:  0.3631578947368421

MATTR in HT:   0.3807580558249772
MATTR in TAN:  0.3577794420926945
MATTR in TAS:  0.3633178910214029

```

Tableau E. 1 : Diversité lexicale (module Python LexicalRichness) – Prétest

```

Lexical Diversity (2021):
-----

Tokens in HT:  9838
Tokens in GGL: 9405
Tokens in DPL: 9603

Types in HT:   1045
Types in GGL:  800
Types in DPL:  742

TTR in HT:     0.10622077658060582
TTR in GGL:    0.08506113769271664
TTR in DPL:    0.07726752056648964

RTTR in HT:    10.53568752036302
RTTR in GGL:   8.24917633186328
RTTR in DPL:   7.57182278320982

CTTR in HT:    7.449856090111173
CTTR in GGL:   5.833048523464095
CTTR in DPL:   5.3540872359504625

MSTTR in HT:   0.388421052631579
MSTTR in GGL:  0.3678888888888889
MSTTR in DPL:  0.3696842105263158

MATTR in HT:   0.38857779205482185
MATTR in GGL:  0.3698124859645102
MATTR in DPL:  0.36782293497364227

```

Tableau E. 2 : Diversité lexicale (module Python LexicalRichness) – Test

Segment length (2021):			

Average length in TS:	24.74		
Average length in TH:	30.26	expanding ratio in TH:	+22.32%
Average length in DL:	30.12	expanding ratio in DL:	+21.75%
Average length in GT:	29.98	expanding ratio in GT:	+21.17%
Segment length (2018):			

Average length in TS:	24.81		
Average length in TH:	29.59	expanding ratio in TH:	+19.27%
Average length in TAN:	29.64	expanding ratio in TAN:	+19.48%
Average length in TAS:	29.08	expanding ratio in TAS:	+17.20%

Tableau E. 3 : Longueur moyenne des phrases

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		

Tableau E. 4 : Équivalence syntaxique (module Python ASTRÉD)

Annexe F

Recueil impressions évaluatrices

	Traduction humaine ou post-édition	Commentaires	Appréciation générale (- / -- / --- / + / ++ /+++)
A			
B			
C			
D			
E			
F			
G			
H			
I			
J			
K			
L			
M			
N			
O			
P			
Q			
R			
S			
T			
U			
V			
W			
X			
Y			
Z			