

ANALYSE COMPARATIVE DE POST-ÉDITION ET DE TRADUCTION HUMAINE EN CONTEXTE ACADÉMIQUE

Perrine Schumacher
Université de Liège et de Genève

Antonio Sutura
Université de Liège

1. INTRODUCTION

Fruit des récents progrès accomplis dans le domaine de l'intelligence artificielle et plus précisément de l'apprentissage profond (*Deep Learning*), la traduction automatique (TA) neuronale a désormais largement supplanté l'ancienne approche dite statistique.

Dans ce contexte d'avancées technologiques, le paysage traductionnel ne cesse de se transformer et d'ouvrir la voie à de nouvelles pratiques dont la post-édition (PE), à savoir le « processus de révision humaine de traductions qui ont été générées par un moteur de TA dans le but d'obtenir un produit final répondant aux exigences qui ont été convenues au préalable (ex. : niveau de qualité négocié entre le client et le prestataire de services) [Traduction libre]. » (Massardo, van der Meer et Khalilov 2016)

La PE est, en effet, aujourd'hui en plein boom dans le secteur des services langagiers et se voit confiée principalement à des traducteurs/trices. Les progrès indéniables réalisés en TA et cet essor de la PE conduisent non seulement les professionnels du secteur à s'adapter (Way 2018), mais aussi les universités à s'interroger sur le contenu des formations en traduction (Loock 2019).

Ainsi, cette étude a pour objet d'explorer les potentielles différences, en matière de qualité de texte cible, entre la TH et la PE effectuées par des étudiants en traduction et de pouvoir en tirer les implications pour perfectionner l'enseignement de la traduction.

En nous appuyant notamment sur les études menées par Čulo et al. (2014), Depraetere (2010), ainsi que Sycz-Opoń et Gałuskińska (2017), nous formulons l'hypothèse que la PE influe sur la qualité du produit fini et ne permet pas aux étudiants d'obtenir des productions de qualité comparable à des TH. Pour tester cette hypothèse, une première recherche expérimentale a été

menée en février 2018 avec vingt-huit étudiants en traduction de l'Université de Liège (Belgique) pour la combinaison de langues anglais-français. Cette expérience nous a permis de comparer la qualité des produits de la TH à celle de PE de TA statistique et de PE de TA neuronale.

Le présent article est structuré de la manière suivante : nous faisons tout d'abord un rapide état de l'art des études comparatives récentes de PE et de TH avant de présenter notre méthodologie de recherche ; ensuite, nous parcourons les résultats de l'analyse statistique de notre corpus après qu'il a été soumis à évaluation humaine et enfin, nous tirons les conclusions de cette étude.

2. ÉTAT DE L'ART

Alors que le gain de temps et, par conséquent, de productivité conféré par la PE par rapport à la TH a été largement démontré dans de nombreuses études récentes (Daems 2016 ; Jia, Carl et Wang 2019 ; Martikainen et Mestivier 2020 ; Screen 2019), qu'en est-il des effets de cette méthode de traduction sur la qualité du produit fini chez les étudiants ? Il apparaît que plusieurs études ont déjà été consacrées à cette question, mais les résultats sont souvent divergents. Nous présentons ci-dessous un aperçu de plusieurs études qui se sont attachées à explorer le produit de la PE d'un point de vue qualitatif.

En effet, si certaines études tendent à montrer que la PE a des effets négatifs sur la qualité des textes cibles (Čulo et al. 2014 ; Depraetere 2010 ; Sycz-Opoń et Gałuska 2017), d'autres travaux ont démontré que la PE n'entraînait pas de baisse de qualité globale en comparaison à la TH (Daems, De Clercq et Macken 2017 ; Garcia 2011 ; Jia, Carl et Wang 2019 ; Martikainen et Mestivier 2020).

Čulo et al. (2014) ont étudié les divergences entre la PE et la TH sur les plans lexical et grammatical dans le cadre d'une étude réalisée avec des traducteurs professionnels et des étudiants pour la combinaison de langues anglais-allemand. Sur l'ensemble des productions, ils ont constaté que le phénomène d'interférence avec le TS était plus marqué en PE qu'en TH et cela s'est traduit, en autres, par la présence de tournures non idiomatiques et de constructions non grammaticales dans les textes post-édités.

Dans son étude de cas sur la PE (anglais-français) effectuée avec des étudiants de Master 2, Depraetere (2010) a conclu que les étudiants ne semblent pas avoir de problèmes à se satisfaire de propositions imparfaites qui ne sont pas idiomatiques à 100 % et qu'ils ont ainsi tendance à

apporter peu de modifications stylistiques et phraséologiques. De plus, les résultats de son étude indiquent que la plupart des erreurs recensées dans le corpus étaient des calques ou des pertes d'informations (omissions).

Enfin, dans leur étude empirique sur la TA du polonais en anglais et du polonais en français, Sycz-Opoń et Gałuskina (2017) notent que les textes post-édités n'ont pas atteint une qualité satisfaisante et que les étudiants ne sont notamment pas parvenus à éliminer l'étrangeté stylistique (*stylistic awkwardness*) imputable à la TA.

Parmi les diverses études dont les résultats vont dans le sens opposé, citons celle, plus ancienne, menée par Garcia (2011) avec des apprenants traducteurs. Les résultats statistiques de cette étude font état d'une amélioration de la qualité en PE de TA par rapport à la traduction conventionnelle, indépendamment du sens de traduction (anglais-chinois/chinois-anglais), de la difficulté du texte source et du niveau des apprenants. En outre, cette recherche a permis de mettre en lumière que la PE était particulièrement bénéfique aux apprenants faibles.

L'étude de Daems, De Clercq et Macken (2017) a pour objet de mesurer l'existence de caractéristiques propres à une langue de PE, ce que les autrices nomment *Post-editese*. Leur corpus est composé de productions réalisées à la fois par des traducteurs professionnels et par des étudiants de Master en traduction. Les conclusions de leurs recherches ne permettent pas de confirmer l'existence d'un *Post-editese* ; il serait, en effet, impossible de faire la distinction entre une PE complète¹ et une TH (anglais-néerlandais), que ce soit en termes de qualité, au niveau lexical, syntaxique ou sémantique.

Concernant l'étude de Jia, Carl et Wang (2019) menée avec 30 étudiants de Master en traduction, les résultats indiquent que la PE de TA neuronale permettrait d'obtenir des productions (anglais-chinois) tout aussi fluides (*fluency*) et fidèles (*accuracy*) que les TH.

Quant à Martikainen et Mestivier (2020), elles ont récemment conduit une étude comparative pour la traduction (anglais-français) de textes spécialisés avec des étudiants de Master 1 et 2 et ont constaté une amélioration de la qualité en PE de TA neuronale avec une réduction moyenne du nombre d'erreurs de moitié par rapport à la TH. Leurs résultats semblent, par ailleurs, confirmer 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 » (Martikainen et Mestivier 2020).

¹ PE pour laquelle la qualité attendue doit être comparable à la qualité d'une TH.

Nous pensons que ces conclusions divergent notamment en fonction des paramètres d'expérience, tels que la taille du corpus de travail, la diversité des critères d'évaluation de la qualité ou encore des combinaisons de langues étudiées.

3. METHODOLOGIE

3.1. Questions de recherche

Dans cette étude, nous cherchons à répondre aux questions de recherche suivantes :

Quels sont les effets de la PE sur la qualité d'un texte cible, lorsque cette tâche est effectuée par des étudiants en traduction ?

Sous-questions :

- Pour la traduction anglais-français en contexte académique, la PE permet-elle d'obtenir des textes de qualité comparable à celle de traductions humaines ?

Étant donné notre hypothèse principale, à savoir que la PE ne permet pas d'aboutir à des productions de qualité comparable à une TH, nous nous attendons à ce que les textes post-édités contiennent, de manière générale, un plus grand nombre d'erreurs que les TH et qu'ils obtiennent une moins bonne note globale sur vingt par rapport aux TH.

- Pour la traduction anglais-français en contexte académique, la PE présente-t-elle des erreurs comparables à celles en TH ?

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 de TA qu'en TH. Nous nous attendons, premièrement, à relever davantage d'erreurs d'adéquation en PE de TA qu'en TH, c'est-à-dire davantage d'erreurs de traduction sur le plan sémantique (contresens, glissement de sens, omission, etc.).

Toutefois, nous supposons également que les textes post-édités contiendront moins d'erreurs de langue (syntaxe, grammaire et orthographe d'usage) que les TH.

- Pour la traduction anglais-français en contexte académique, la qualité de textes post-édités varie-t-elle en fonction du type de TA utilisé (approche statistique ou approche neuronale) ?

À cet égard, nous pensons que les étudiants produiront des productions de meilleure qualité avec la TA neuronale. Nous nous attendons à ce que les PE de TA neuronale contiennent moins

d'erreurs que les PE de TA statistique et qu'elles se voient attribuer de meilleures notes que les PE de TA statistique.

3.2. Expérience

Pour répondre à ces questions de recherche, nous avons mené une recherche expérimentale comparative entre la TH et la PE avec des étudiants de traduction. Nous avons demandé à aux étudiants de traduire et de post-éditer trois textes différents de l'anglais en français ; chaque étudiant a produit une TH, une PE de TA statistique et une PE de TA neuronale. Les textes source ont été traduits selon trois modes de traduction différents. La composition du corpus est détaillée dans le graphique 1.

Graphique 1. Composition du corpus

En termes de qualité attendue en PE, nous avons demandé aux étudiants de produire une PE complète (*full post-editing*) et ainsi d'aboutir à un texte post-édité dont la qualité est comparable à celle d'une TH. Ils ont reçu une liste de consignes tirée de la TAUS/CNGL (2010) et légèrement adaptées aux fins de notre étude. Ce niveau de qualité est défini par la TAUS/CNGL (2010) 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. »

Étant donné que l'expérience se déroule en contexte pédagogique, nous avons cherché à reproduire, autant que possible, l'environnement de la formation en traduction proposée à l'Université de Liège.

3.2.1. Choix du corpus

Trois textes source (entre 300 et 340 mots par texte) ont été sélectionnés par trois traductrices et enseignantes de traduction de l'Université de Liège. Il s'agit d'un article de presse, d'un article économique et d'un article de vulgarisation scientifique. L'article de presse en date du 9 novembre 2017 s'intitule *Can Theresa May and her government survive? Our writers' verdicts* et est tiré du journal britannique *The Guardian*. Le texte économique s'intitule *How will the global economic order change by 2050?*, 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, l'article de vulgarisation scientifique, *What is synesthesia*, a été publié dans la revue américaine *Scientific American* le 11 septembre 2006 et traite de la synesthésie,

à savoir un trouble de la perception sensorielle. Bien entendu, nous avons vérifié qu'aucune traduction de ces textes n'était accessible en ligne au moment de notre expérience.

3.2.2. Outils de traduction

Pour la tâche de TH, l'étudiant était libre de disposer de tous les outils jugés nécessaires (dictionnaires explicatifs et traductifs, glossaires, Internet, etc.) à la seule condition de ne pas avoir recours à un moteur de TA.

Concernant les outils de TA, nous avons employé Bing Translator comme système de TA statistique car il était l'un des rares moteurs gratuits à toujours proposer l'approche statistique au moment de la tenue de notre expérience et DeepL Translator comme système de TA neuronale. Nous avons décidé d'utiliser DeepL car il était considéré, à tout le moins en 2018, comme le traducteur automatique grand public le plus performant, devançant notamment les solutions similaires de Google, Microsoft et Facebook (Smolentceva 2018). Les TA ont été générées le 23 janvier 2018. À titre informatif, le tableau 1 présente les scores donnés par la métrique d'évaluation automatique BLEU (Papineni et al. 2002) pour chaque texte selon chaque système de TA.

Tableau 1. Scores BLEU

3.2.3. Évaluation de la qualité

À nouveau, comme notre expérience se déroule en contexte pédagogique, nous avons cherché à reproduire, dans la mesure du possible, les conditions dans lesquelles les étudiants sont évalués et nous avons ainsi opté pour une évaluation humaine de la qualité.

Cependant, afin de limiter les biais de subjectivité, les productions des étudiants ont été évaluées par plusieurs évaluateurs ; chaque production a été évaluée deux fois. De plus, en nous inspirant de la typologie d'erreurs développée par Daems (2016), nous avons divisé l'évaluation de la qualité en deux phases : acceptabilité et adéquation.

La phase d'acceptabilité entend évaluer le respect des normes et usages de la langue cible.

Pour cette première phase, l'évaluatrice dispose uniquement du texte cible et repère les fautes de langue (faute de syntaxe, de grammaire, d'orthographe, barbarisme, impropriété, etc.), fautes de style (répétition, longueur de phrase, fluidité, etc.) et fautes de cohérence (logique interne, irrégularité dans l'emploi de termes, concordance des temps, etc.) (graphique 2).

Graphique 2 — Phase d'acceptabilité

La phase d'adéquation permet, dans un deuxième temps, le repérage des fautes de traduction, c'est-à-dire des fautes de transfert de sens (contresens, glissement de sens, calque, ajout, omission, etc.) (graphique 3). L'évaluatrice dispose cette fois du texte cible et du texte source.

Le détail de la typologie d'erreurs est repris à l'annexe 1.

Graphique 3 — Phase d'adéquation

3.2.4. Participants

28 étudiants (21 filles et 7 garçons) de 3e année de Bachelier traduction–interprétation de l'Université de Liège (ULiège) ont pris part à cette expérience. Ces étudiants étaient âgés entre 20 ans et 30 ans, ils sont pour la majorité de langue maternelle française à l'exception d'un étudiant germanophone.

Pour mieux connaître leur opinion sur la TA et sur la PE, ainsi que leur expérience en la matière, nous avons élaboré deux courts sondages qu'ils ont dû compléter en ligne via LimeSurvey (Schmitz 2015). Afin de voir si leur opinion a changé au cours de l'expérience, nous leur avons proposé de remplir un premier questionnaire avant de prendre part à l'expérience et un deuxième questionnaire après avoir participé. Alors que la majorité des étudiants interrogés (89 %) indique avoir déjà fait usage d'un moteur de TA, seuls 17 % ont déclaré avoir déjà post-édité de la TA avant de prendre part à notre expérience contrôlée. En outre, aucun étudiant n'a suivi de formation spécifique en TA ou en PE.

3.2.5. Résultats collectés et analyse

Après exclusion de plusieurs productions pour des raisons de représentativité, notre corpus final est composé de 75 productions (25 TH, 25 PE de TA statistique et 25 PE de TA neuronale). Pour répondre aux différentes questions de recherche, nous avons procédé à l'évaluation humaine de la qualité de ces productions en classant et en comptabilisant les erreurs. Dans le but d'explorer les effets de nos variables indépendantes (méthode de traduction et texte source) sur notre variable dépendante (la qualité du texte cible), nous nous pencherons sur la distribution des erreurs, ainsi que sur la distribution des notes attribuées aux productions des étudiants.

4. RESULTATS

Dans un souci de concision, nous nous limitons dans cet article à présenter les résultats de l'analyse statistique des données issues de notre corpus à la suite de cette évaluation humaine.

4.1. Résultats des statistiques descriptives

4.1.1. Distribution des erreurs par catégorie

Dans un premier temps, nous avons observé la distribution des erreurs relevées par catégorie indépendamment de la méthode de traduction et du texte source (voir graphiques 4 et 5).

Graphique 4 — Distribution des erreurs par catégorie pour la phase d'acceptabilité, indépendamment de la méthode de traduction

Pour la première phase de l'évaluation (graphique 4) — la phase d'acceptabilité (erreurs de langue) — la catégorie d'erreur 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 étudiants. 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 %).

Graphique 5 — Distribution des erreurs par catégorie pour la phase d'adéquation, indépendamment de la méthode de traduction

Pour la deuxième phase de l'évaluation (graphique 5) — la phase d'adéquation (erreurs de sens) — la catégorie d'erreur la plus représentée est celle des « calques ». Cette catégorie représente 50 % de l'ensemble des erreurs de traduction commises par les étudiants. 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 représente seulement 3 % et enfin la catégorie « sur-/sous-traduction » avec 0 % qui n'est pas représentée.

Les distributions des erreurs selon la méthode de traduction et selon le type de texte sont reprises dans les graphiques 6 et 7. Afin de pouvoir tirer des conclusions sur les différences observées sur ces représentations graphiques, nous avons eu recours à la statistique inférentielle.

Graphique 6 — Distribution des erreurs selon la méthode de traduction et le type de texte (phase d'acceptabilité)

4.2. Résultats des statistiques inférentielles

Afin de tester si la qualité des productions recueillies a été influencée par la méthode de traduction et/ou par le texte source (graphiques 6 et 7), nous avons cherché à démontrer l'existence de différences statistiquement significatives au niveau du nombre moyen d'erreurs relevées par catégorie en fonction de nos deux variables : méthode de traduction (TH, TAS+PE ou TAN+PE) et texte source (Texte 1 : article de presse ; Texte 2 : article économique ; Texte 3 : article de vulgarisation scientifique).

Pour ce faire, nous avons eu recours à un test d'analyse de la variance (ANOVA). Il s'agit d'un modèle statistique utilisé pour comparer les moyennes d'échantillons dans le but de tester l'influence d'une (ou plusieurs) variable(s) qualitative(s) sur une variable quantitative. Étant donné que, dans cette expérience, nous avons deux facteurs de variabilité (la méthode de traduction et le type de texte source) et plus de deux échantillons appariés, nous avons dû envisager l'application d'un test ANOVA à deux facteurs, en mesures répétées (*two-way repeated measures ANOVA*) afin de traiter les répétitions de mesures sur les mêmes sujets (les étudiants).

Ainsi, les différentes valeurs récoltées à la suite de l'évaluation humaine des productions ont été soumises au test ANOVA avec un modèle linéaire mixte. Nous tentons de déterminer si nos variables indépendantes, à savoir la méthode 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 étudiants, tantôt par les notes attribuées aux productions des étudiants. Le cas échéant, nous analysons individuellement les effets de chaque méthode 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 le traitement des données, nous avons eu recours au logiciel R 3.6.3 (Core Team 2020). 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.

Les résultats se présentent comme suit : nous analysons les effets des variables indépendantes sur :

- la qualité globale des productions (section 4.2.1) selon trois niveaux :
 - 1) les erreurs en phase d'acceptabilité (4.2.1.1)
 - 2) les erreurs en phase d'adéquation (4.2.1.2)
 - 3) les notes globales attribuées aux productions des étudiants (4.2.1.3).
- la qualité des productions par catégories d'erreurs (section 4.2.2)

4.2.1. Qualité globale

4.2.1.1. Acceptabilité

Tableau 2. ANOVA — Distribution des erreurs en phase d'acceptabilité

Le premier test ANOVA que nous avons effectué a pour objet d'analyser les effets de la méthode de traduction et du texte source sur le nombre d'erreurs comptabilisées en phase d'acceptabilité (erreurs de langue).

Méthode de traduction

Il apparaît tout d'abord que l'effet « méthode de traduction » (« methode » dans le tableau 2) est significatif car $P < 0,001$ ($F = 8,22$; $dfN = 2$; $dfD = 44$; $P = 0,0009216$), ce qui signifie qu'il y a une différence significative entre les distributions de valeurs pour les erreurs d'acceptabilité correspondant à chaque méthode de traduction.

L'analyse statistique indique que la distribution des erreurs d'acceptabilité (erreurs de langue, de style et de cohérence) a été influencée par le méthode de traduction employé, et ce, indépendamment du texte source. Quant aux résultats du test de Tukey, ils révèlent un écart statistiquement significatif entre la TA statistique et la TA neuronale ($P < 0,001$), ainsi qu'entre la TH et la TA neuronale ($P < 0,05$). Nous pouvons en conclure que les étudiants ont commis significativement moins d'erreurs de langue lorsqu'ils ont post-édité de la TA neuronale qu'ils n'en ont commises en PE de TA statistique et en TH.

Texte source

L'effet « texte source » (« texte » dans le tableau) est, lui aussi, significatif avec $P < 0,001$ ($F = 62,91$; $dfN = 2$; $dfD = 44$; $P < 0,001$), ce qui signifie qu'il y a une différence significative entre les distributions de valeurs pour les erreurs d'acceptabilité correspondant à chaque texte.

Les résultats du test de Tukey montrent l'existence de différences significatives entre tous les textes source ($P < 0,001$). Ces observations nous amènent à affirmer que, pour la phase d'acceptabilité, le type de texte source constitue bel et bien une variable dans cette expérience, au même titre que la méthode de traduction. Remarquons que le texte source pour lequel les étudiants ont commis significativement le plus grand nombre d'erreurs d'acceptabilité est le texte 2 (économique), et ce, indépendamment de la méthode de traduction. Tandis que le texte qui a engendré le moins d'erreurs dans le chef des étudiants est le texte 3 (vulgarisation scientifique).

Interaction méthode–texte

En revanche, ce test statistique n'a pas révélé d'interaction significative entre la méthode de traduction et le texte (« méthode:texte » dans le tableau 2) ($F = 1,04$; $dfN = 4$; $dfD = 37,32$; $P = 0,398$). En d'autres termes, la qualité des productions ne dépend pas significativement de l'interaction entre nos deux variables indépendantes.

4.2.1.2. Adéquation

Tableau 3. ANOVA — Distribution des erreurs en phase d'adéquation

Le deuxième test ANOVA nous permet d'analyser les effets de la méthode de traduction et du texte source sur le nombre d'erreurs comptabilisées cette fois en phase d'adéquation (erreurs de traduction).

Texte source

Il apparaît que l'effet « texte source » (texte) est significatif ($F = 20,74$; $dfN = 2$; $dfD = 44$; $P < 0,001$). Les résultats du test de Tukey montrent que toutes les moyennes entre les textes sont significativement différentes ($P < 0,01$). Nous pouvons ainsi conclure que le type de texte source a influencé la distribution des erreurs d'adéquation dans notre corpus. Ainsi, le texte source pour lequel les étudiants ont commis significativement le plus grand nombre d'erreurs d'adéquation est à nouveau le texte 2 (économique) et ce, indépendamment de la méthode de traduction. Tandis que le texte qui a engendré le moins d'erreurs d'adéquation dans le chef des étudiants est, cette fois, le texte 1 (article de presse).

Méthode de traduction

Contrairement aux résultats obtenus pour la phase d'acceptabilité, l'ANOVA effectuée pour la phase d'adéquation indique qu'il n'y aurait pas de dépendance entre la méthode de traduction

et la distribution des erreurs d'adéquation (erreurs de traduction) : l'effet « méthode de traduction » (methode) est non significatif ($F= 1,09$; $dfN= 2$; $dfD= 44$; $P=0,342$). Il n'est donc pas nécessaire de passer par un test de Tukey.

Interaction méthode–texte

Enfin, le test statistique ne révèle pas d'interaction significative entre la méthode de traduction et le texte (methode:texte) ($F= 2,10$; $dfN= 4$; $dfD= 38,45$; $P=0,098$). Par conséquent, il ne nous a pas été possible d'inférer de dépendance entre la méthode de traduction et le nombre moyen d'erreurs commises en phase d'adéquation.

4.2.1.3. Notes

Nous avons également cherché à savoir si la méthode de traduction ou le texte source ont eu une influence significative sur la note globale attribuée aux productions des étudiants (tableau 4). La note globale sur 20 comprend la note obtenue en phase d'acceptabilité (40 %) et la note obtenue en phase d'adéquation (60 %).

Tableau 4. ANOVA — Distribution des notes

Méthode de traduction

Les résultats montrent que l'effet « méthode de traduction » est significatif ($F= 6,69$; $dfN= 2$; $dfD= 44$; $P= 0,002895$). Il apparaît donc que la méthode de traduction a eu une influence significative sur la qualité des textes cibles, mesurée ici en notes globales attribuées aux productions des étudiants.

Les résultats du test de Tukey 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 étudiants sont en moyenne meilleures en PE de TA neuronale par rapport aux deux autres méthodes de traduction, et ce, indépendamment du texte source.

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.

Le test de Tukey 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 ($P<0,001$). 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 étudiants, car il a généré des notes globales significativement plus basses que les deux autres textes source.

Interaction méthode–texte

Les résultats de l'ANOVA repris dans le tableau 4 ne révèlent pas d'interaction significative entre les méthodes de traduction et les textes source ($F= 1,57$; $dfN= 4$; $dfD= 38,16$; $P= 0,201$).

4.2.2. Qualité par catégories d'erreurs

De manière plus spécifique, les tests ANOVA et Tukey (voir détails en annexe 2) nous ont également permis de mettre en évidence que la méthode de traduction a eu une influence significative sur les catégories d'erreurs suivantes :

- Calque

Le test ANOVA révèle une influence significative entre la méthode de traduction et le nombre de calques recensés dans les productions ($F= 9,29$; $dfN= 2$; $dfD= 44$; $P= 0,0004$). Les résultats du test de Tukey 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 1,888 et à celle de la PE de TA statistique de 3,275.

Ainsi, nous pouvons affirmer que les étudiants ont commis, en moyenne, davantage de calques fautifs en PE, que ce soit de TA statistique ou de TA neuronale, qu'ils 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 générerait moins de calques fautifs dans le chef des étudiants que l'ancienne génération de TA.

- Glissement de sens

Le test ANOVA fait apparaître une dépendance entre la méthode de traduction et le nombre de glissements de sens comptabilisés dans le corpus ($F= 3,366$; $dfN= 2$; $dfD= 44$; $P= 0,043$). Le test de Tukey 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 étudiants ont commis, en moyenne, plus de glissements de sens lorsqu'ils ont traduits « humainement » que lorsqu'ils ont post-édité de la TA neuronale.

Il semblerait que, dans cette expérience, le recours à la PE de TA neuronale ait permis de réduire les erreurs de glissements de sens par rapport à la TH.

- Orthographe

Le test ANOVA nous permet d'inférer une relation de dépendance entre la méthode de traduction et le nombre de fautes d'orthographe relevées dans les productions des étudiants ($F=16,62$; $dfN=2$; $dfD=44$; $P<0,001$). Le test de Tukey 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 en conclure que, dans notre expérience, lorsque les étudiants ont traduit « humainement », ils ont commis davantage de fautes d'orthographe que lorsqu'ils ont post-édité.

- Grammaire et syntaxe

Les résultats du test ANOVA font apparaître une relation de dépendance entre la méthode de traduction et le nombre d'erreurs de grammaire et de syntaxe présentes dans les productions des étudiants ($F=12,826$; $dfN=2$; $dfD=44$; $P<0,001$). Le test de Tukey 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 ($P<0,01$).

Il ressort de ces résultats qu'en PE de TA neuronale, les étudiants ont commis, en moyenne, moins d'erreurs de grammaire et de syntaxe par rapport aux deux autres méthodes de traduction.

- Cohérence

Cette catégorie concerne des erreurs de type : problèmes de logique interne, irrégularité dans l'emploi de termes, concordance des temps, etc. Les résultats du test ANOVA montrent que la méthode de traduction a eu une influence sur le nombre d'erreurs de cohérence ($F=6,04$; $dfN=2$; $dfD=44$; $P=0,004$). Le test de Tukey fait apparaître deux écarts significatifs : le PE de TA statistique présente un résultat supérieur à la fois au résultat obtenu en PE de TA neuronale ($P<0,01$), ainsi qu'au résultat obtenu en TH ($P<0,05$).

Dans cette expérience, nous pouvons affirmer que la PE de TA statistique a engendré davantage d'erreurs de cohérence que les deux autres méthodes.

En revanche, nos résultats révèlent que, dans cette expérience, la méthode de traduction n'a pas eu d'influence significative sur les catégories d'erreurs suivantes : contresens et non-sens ; ajout et omission ; vocabulaire ; typographie et ponctuation ; style et registre.

Discussion et conclusion

Tout d'abord, les résultats de cette première recherche expérimentale viennent infirmer notre principale hypothèse selon laquelle la PE de TA ne permet pas d'aboutir à des productions de qualité comparable à des TH. En effet, contrairement à ce que nous supposions, les résultats de l'analyse statistique nous poussent à conclure que, dans cette expérience, la PE n'a pas eu d'effets négatifs sur la qualité globale des productions finales et que les textes post-édités par les étudiants sont de qualité globale équivalente aux TH, voire supérieure dans le cas de la PE de TA neuronale.

Les premiers tests ANOVA ont permis de révéler que la méthode 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) et au niveau des notes attribuées aux productions. Nous n'avons pas pu inférer de dépendance entre la méthode de traduction et la distribution des erreurs d'adéquation (erreurs de traduction).

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, nous pouvons en tirer deux conclusions : premièrement, les étudiants ont commis significativement moins d'erreurs de langue lorsqu'ils ont post-édité de la TA neuronale qu'ils n'en ont commises en TH et deuxièmement, les notes obtenues en PE de TA neuronale sont, en moyenne, meilleures qu'en TH, et ce, indépendamment du texte source.

Alors que notre hypothèse principale se trouve réfutée, ces résultats confirment toutefois notre deuxième hypothèse, 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 la PE de TA neuronale a, non seulement, généré moins d'erreurs de langue qu'en PE de TA statistique, mais qu'elle a aussi permis d'aboutir à des productions qui ont récolté de meilleures notes qu'en PE de TA statistique. Nous pensons que ces constatations jouent en faveur d'une intégration de la PE, et particulièrement de la PE de TA neuronale, dans la formation initiale des étudiants dès lors que cette méthode de traduction permettrait manifestement d'augmenter la qualité globale de leurs productions.

Nous pouvons conclure que, dans cette expérience, la PE de TA (statistique ou 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, De Clercq et Macken 2017 ; Depraetere et al. 2014 ; Garcia 2011 ; Jia, Carl et Wang 2019). En outre, la PE de TA neuronale a permis d'aboutir à des productions de qualité globale supérieure par rapport aux deux autres méthodes de traduction. Ce constat vient notamment confirmer les résultats de l'étude comparative menée par Martikainen et Mestivier (2020) avec des étudiants de Master 1 et 2 dans laquelle il est fait état d'une amélioration de la qualité en PE de TA neuronale (DeepL) par rapport à la TH.

Par ailleurs, de manière plus spécifique, nos résultats montrent que la méthode de traduction a eu des effets significatifs sur les catégories d'erreurs suivantes : calque ; glissement de sens ; orthographe ; grammaire et syntaxe ; cohérence. Ainsi, il ressort de l'analyse statistique que, dans cette expérience, les textes post-édités (TA statistique et TA neuronale) par les étudiants contiennent significativement plus de calques et moins de fautes d'orthographe que les TH ; la PE de TA neuronale a généré moins de glissements de sens par rapport à la TH, ainsi que moins d'erreurs de grammaire et de syntaxe par rapport aux deux autres méthodes de traduction et enfin, que la PE de TA statistique a produit davantage d'erreurs de cohérence par rapport aux deux autres méthodes.

Cette apparition plus fréquente de calques constitue une tendance marquée en PE ; une des raisons avancées par Čulo et al. (2014) est qu'il y aurait un fort effet d'interférence avec le texte source induit par le processus même de PE. Il paraît évident, mais il est toujours bon de le rappeler : le moteur de TA va générer un texte qui reste plus proche linguistiquement du texte source qu'une TH ne le serait (Depraetere 2010). Il a été constaté, qu'en post-éditant le texte cible, les étudiants 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. Il se pourrait que ce phénomène provienne d'un excès de confiance de la part des étudiants à l'égard des propositions de la machine (Daems 2016 ; Deneufbourg 2019 ; Depraetere, 2010). 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-editese*. Aussi, nous sommes convaincus qu'il convient de porter un point d'attention particulier à ce type d'erreurs dans l'enseignement de la TA et de la PE.

Enfin, nous tenons à souligner que nous sommes bien conscients que ces résultats ne peuvent valoir que dans le cadre spécifique de cette expérience menée avec des apprenants et pour la

combinaison de langues anglais-français. De futurs travaux de recherche sont nécessaires pour pouvoir confirmer ces observations, nous avons l'intention de reproduire prochainement cette expérience avec des étudiants de Master en traduction.

Bibliographie

Core Team, R. (2020): *R: A language and environment for statistical computing (Computer software)*. Vienna, Austria. <https://www.R-project.org/> [Google Scholar]

Čulo, Oliver; Gutermuth, Silke; Hansen-Schirra, Silvia et Nitzke, Jean (2014): «The Influence of Post-Editing on Translation Strategies», in Sharon O'Brien, Laura Winther-Balling, Michael Carl, Michel Simard et Lucia Specia (eds.), *Post-editing of Machine Translation: Processes and Applications*. Newcastle upon Tyne: Cambridge Scholars Publishing, 200-218.

Daems, Joke (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.

Daems, Joke; De Clercq, Orphée et Macken, Lieve (2017): «Translationese and Post-edited: How Comparable Is Comparable Quality?», *Linguistica Antverpiensia New Series-themes in Translation Studies*, 16, 89-103. <http://hdl.handle.net/1854/LU-8516838>

Deneufbourg, Guillaume (2019): *Post-édition de traduction automatique : se méfier des apparences*, article inspiré d'une communication présentée au Congrès annuel de l'American Translators Association, le 25 octobre 2018. La Nouvelle-Orléans, Louisiane, États-Unis. <https://www.ata-divisions.org/FLD/index.php/tag/guillaume-deneufbourg/>

Depraetere, Ilse (2010, mai): *What counts as useful advice in a university post-editing training context? Report on a case study*, in François Yvon and Viggo Hansen (éds.) *EAMT 2010 : compte-rendu de la 14e conférence annuelle de l'Association européenne de traduction automatique, 27-28 mai 2010*. Saint-Raphaël, France. <http://www.mt-archive.info/EAMT-2010Depraetere-2.pdf>

Depraetere, Ilse, De Sutter, Nathalie et Tezcan, Arda (2014): «Post-edited quality, post-editing behaviour and human evaluation: a case study», in Sharon O'Brien, Laura Winther-Balling, Michael Carl, Michel Simard et Lucia Specia (eds.), *Post-editing of Machine Translation: Processes and Applications*. Newcastle upon Tyne: Cambridge Scholars Publishing, 78-108.

Garcia, Ignacio (2011): «Translating by post-editing: Is it the way forward?», *Machine Translation*, 25, 217-237. <https://doi.org/10.1007/s10590-011-9115-8>

Jia, Yanfang; Carl, Michael et Wang, Xiangling (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. https://www.jostrans.org/issue31/art_jia.php

Lang, Tom (2007): «Documenting Research in Scientific Articles: Guidelines for Authors, 2 – Reporting Hypothesis Tests», *Chest*, 131(1), 317-319.

Loock, Rudy (2019): «La plus-value de la biotraduction face à la machine», *Traduire*, 241, 54-65. <http://journals.openedition.org/traduire/1848>

Martikainen, Hanna et Mestivier, Alexandra (2020, janvier): *Les outils de traduction nouvelle génération : quel effet sur la qualité des textes traduits ?* Communication présentée à la journée d'études Traduction & Qualité 2020 : Biotraduction et traduction automatique, Université de Lille, France.

Massardo, Isabella; van der Meer, Jaap et Khalilov, Maxim (2016): *TAUS translation technology landscape report*, De Rijp : TAUS.

Papineni, Kishore; Roukos, Salim; Ward, Todd et Zhu, Wei-Jing (2002): «Bleu: a method for automatic evaluation of machine translation», in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. Philadelphia, 311-318.

Schmitz, Carsten (2015): *Limesurvey: An Open Source survey tool*. Hamburg, Germany: Limesurvey Project Team.

Screen, Benjamin (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.

Smolentceva, Natalia (2018, 5 décembre): «DeepL: Cologne-based startup outperforms Google Translate. *Deutsche Welle*». <https://www.dw.com/en/deep-l-cologne-based-startup-outperforms-google-translate/a-46581948>

Sycz-Opoń, Joanna et Gałuskina, Ksenia (2017): «Machine Translation in the Hands of Trainee Translators – an Empirical Study». *Studies in Logic, Grammar and Rhetoric*, 49 (1), 195-212.

TAUS/CNGL (2010): *Consignes relatives à la post-édition des traductions automatiques*. <https://www.taus.net/academy/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines-french>

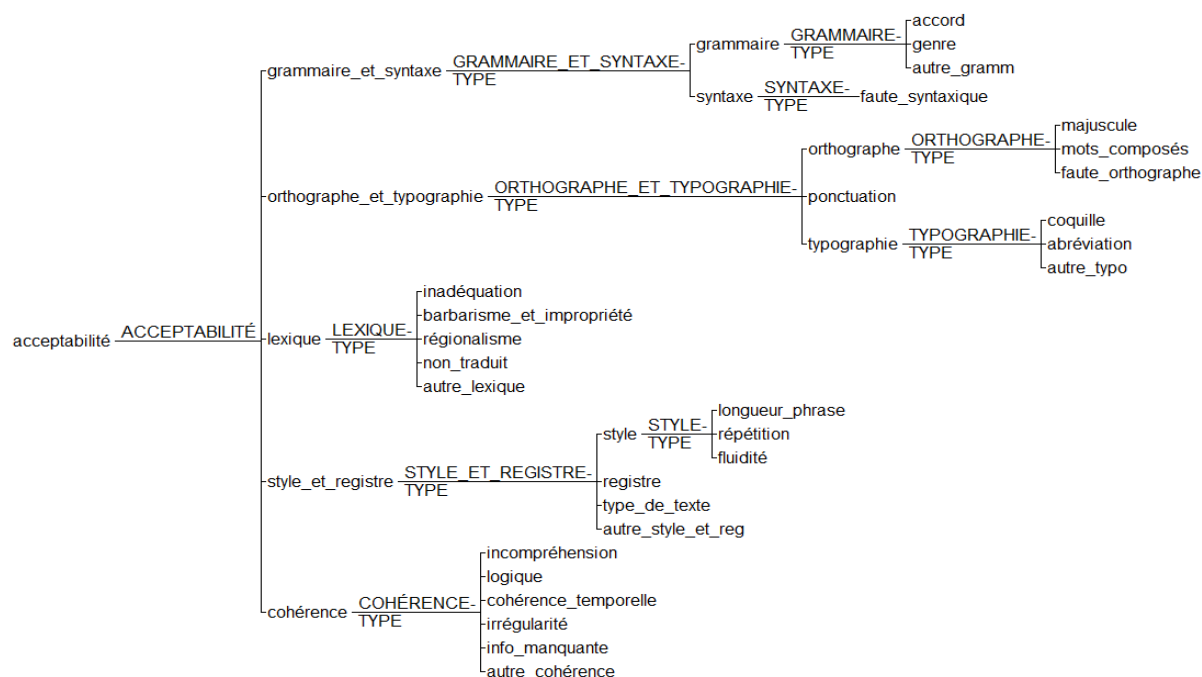
Way, Andy (2018): «Quality expectations of machine translation», in Joss Moorkens, Sheila Castilho, Federico Gaspari et Stephen Doherty (eds.), *Translation Quality Assessment: From Principles to Practice*. Berlin : Springer, 159-178.

Annexe 1

Consignes détaillées pour l'évaluation humaine

Acceptabilité (*Acceptability*)

Aperçu des 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/de syntaxe (voir ci-dessous), ou d'une coquille (voir plus loin « orthographe et typographie »).

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
- **Autre_gramm**
Merci de préciser en quoi cette erreur ne rentre dans aucune autre catégorie.

Syntaxe

- **Faute_syntactique**
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 et typographie

Non-respect des règles d'orthographe et de typographie d'usage.

► Note : Pour certains cas problématiques, c'est à vous de distinguer la faute d'orthographe de la faute de typographie. Ex. : « étique ». À vous d'estimer s'il s'agit-il d'une graphie intentionnelle (faute d'orthographe) ou non intentionnelle (faute typographique).

Orthographe

► Note : Une erreur orthographique récurrente ne doit être soulignée qu'une seule fois par texte.

- **Majuscule**
Majuscule manquante ou superflue
- **Mots composés**
Trait d'union manquant ou superflu
- **Faute d'orthographe**
Faute d'orthographe d'usage

Ponctuation

Faute de ponctuation, signe manquant ou superflu.

Typographie

► Note : Une erreur typographique récurrente (ex. : utilisation de guillemets anglais ou lieu de guillemets français) ne doit être soulignée qu'une seule fois par texte.

- **Coquille**
Omission, addition, interversion ou substitution d'un ou de plusieurs caractères typographiques. (Espace manquante ou superflue ; mauvais type de guillemets (anglais "... " ou français « ... »)), etc.
Ex. : L'infromation ; il a été promené chien ; il a été promené le **le** chien
- **Abréviation**
Abréviation incorrecte
On distingue généralement trois grandes catégories d'abréviations :
 - Les abréviations proprement dites (c.-à-d. ; Mmes) ;
 - Les sigles et les acronymes (OPEP ; BENELUX) ;
 - Les symboles (h = heure ; °C = degré ; g = gramme).
- **Autre_typo**
Merci de préciser en quoi cette erreur ne rentre dans aucune autre catégorie.

Lexique

- **Inadéquation**
Terme inadéquat

Ex. La Chine pourrait être la plus grande économie mondiale, **comptant** environ 20 % du PIB mondial.
= [...] **regroupant/comptabilisant** environ 20 % du PIB mondial.

- **Barbarisme et impropriété**

Barbarisme : Forme d'un mot qui n'existe pas en français ; déformation d'un mot, faute de conjugaison.

Ex. : « abrégier » pour « abréger » ; « il envoira » pour « il enverra »

Impropriété : 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, québécoïsme, etc. jugé fautif ou inapproprié dans ce contexte.

- **Non traduit**

Mot/élément de la phrase qui n'a pas été traduit

- **Autre lexique**

Merci de préciser en quoi cette erreur ne rentre dans aucune autre catégorie.



Style et registre

Incompréhension

Tournure/segment incompréhensible ou absurde.

Style

- **Longueur_phrase**

Phrase jugée trop longue ou trop courte

- **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.

- **Fluidité**

Construction grammaticalement correcte, mais il y a un problème de fluidité. (Mauvaise cooccurrence, non-idiomaticité, collocation malheureuse, lourdeur, déformation de proverbes...)

Registre

Registre jugé trop formel/informel

Type_de_texte

Non-respect des éventuelles contraintes/normes liées au type de texte. Veuillez préciser.

Autre_style_et_reg

Merci de préciser en quoi cette erreur ne rentre dans aucune autre catégorie.



Cohérence

- **Logique**

Problème de 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

- **Info_manquante**

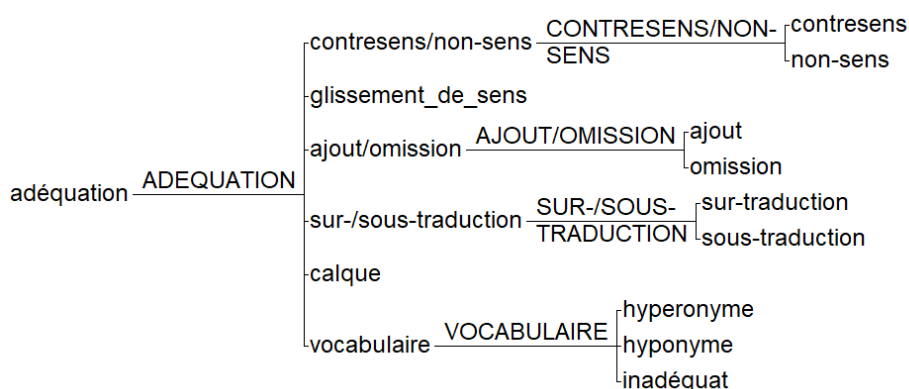
Information manquante

- **Autre_cohérence**

Merci de préciser en quoi cette erreur ne rentre dans aucune autre catégorie.

Adéquation (*Adequacy*)

Aperçu des 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. : *And Canavero has yet to explain why moving a head onto a new body **would produce anything other than death or dementia.***

→ M. Canavero doit cependant expliquer en quoi transplanter une tête sur un corps étranger **pourrait provoquer la mort ou la démence** du patient.

= [...] **pourrait aboutir à une autre fin que la mort ou la démence** du patient.

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

Ambiguïté sémantique, perte partielle ou totale de sens sans pour autant mener à un contresens/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/Omission

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

« 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 !



Sur-/sous-traduction

Sur-traduction

« Le fait de ne pas avoir recours à une implicite lorsqu'elle s'impose conduit à une « surtraduction », faute de traduction qui consiste à traduire explicitement des éléments du texte de départ qui devraient normalement rester implicites dans le texte d'arrivée ». (Delisle 2013)

Ex. : d) *No parking at any time.*

→ Stationnement interdit **en tout temps**.

= Stationnement interdit ». (Delisle 2013 : 206)

Sous-traduction

« Le fait de ne pas recourir à une explicitation lorsqu'elle s'impose aboutit à une sous-traduction, faute de traduction qui consiste à omettre dans le texte d'arrivée les compensations, étoffements ou explicitations qu'exige une traduction idiomatique et conforme au sens attribué au texte de départ par le traducteur » (Delisle 2013).

Ex. : *Non-biological samples may be in the raw unprocessed form or as the final product, available either in bulk quantities or in small "street amounts".* [Domaine de la drogue]

Sous-traduction : Les échantillons non organiques peuvent être des produits bruts, non transformés, ou des produits élaborés, à leur stade final, et se présenter soit en vrac soit en petites « **quantités de rue** ».

Version explicitée : Les échantillons non organiques peuvent être des produits bruts, non transformés, ou des produits élaborés, à leur stade final, et se présenter soit en vrac soit en petites « **quantités destinées à la vente dans la rue** ».

(Delisle 2013 : 213)



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_syntactique**

« 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**

« Mot ou terme qui entretient une relation hiérarchique avec un autre mot ou terme dont l'extension sémantique est plus restreinte.

Ex. : l'hyperonyme *animal* englobe les hyponymes *chat, cheval, chien*, etc.

► Note : Il arrive que l'hyperonyme de la langue de départ doive être traduit par plusieurs hyponymes dans la langue d'arrivée.

Ex. 1. Faute d'un hyperonyme français correspondant à *nuts* dans l'expression *mixed nuts*, le traducteur se voit contraint d'utiliser plusieurs homonymes : *amandes, arachides, noix d'acajou, pistaches*, etc. » (Delisle 2013 : 662)

- **Hyponyme**

« Mot ou terme qui entretient une relation hiérarchique avec un autre mot ou terme dont l'extension sémantique est plus large

Ex. : Les hyponymes *canapé, chaise, fauteuil* et *table* sont de même niveau et englobés dans l'hyperonyme *meuble*.

► Note : Le fait de traduire un hyperonyme du texte de départ par un hyponyme du texte d'arrivée constitue parfois une faute de traduction. » (Delisle 2013 : 663)

Ex. : *He bought me flowers for my birthday.* → Il m'a offert des **roses** pour mon anniversaire.

- **Inadéquat**

Traduction non adéquate dans le contexte sans pour autant mener à un glissement de sens.

Ex. : *But let's say this guy has figured out a way to **reattach** a broken spinal cord so that a brain could be moved to another body and work.*

→ Mais admettons que ce type ait effectivement trouvé un moyen de **raccorder** une moelle épinière sectionnée [...]

= [...] un moyen de **rabouter/ressouder/reconnecter** [...]

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 - accord

= 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 ne disposez pas du TS pour le vérifier, veuillez toutefois annoter cette erreur dans la première PUIS dans la deuxième phase d'évaluation lorsque vous disposerez du TS.

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

Encore merci pour votre précieuse contribution à ce projet et bonne évaluation.

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. (2013). *La Traduction raisonnée : manuel d'initiation à la traduction professionnelle de l'anglais vers le français*. Les Presses de l'Université d'Ottawa, 3e éd.

Delisle, J., Lee-Jahnke, H. et Cormier, Monique C. (1999). *Terminologie de la traduction = Translation terminology = Terminología de la traducción = Terminologie der Übersetzung*. (FIT monograph series = Collection FIT ; 1). Amsterdam-Philadelphia : John Benjamins.

Toury, G. (1995). The Nature and Role of Norms in Translation. In idem, *Descriptive Translation Studies and Beyond*. Amsterdam-Philadelphia : John Benjamins. 1995, 53-69 [En ligne]. Disponible sur : <http://spinoza.tau.ac.il/~toury/works> (Consulté le 02/03/2018)

Annexe 2

ANOVA analysis

```
print(params$title)

## [1] "DonneesANOVA_ADEQ_Calque.csv"
df <- read.table(params$file, header = TRUE, sep = ";")
```

Anova

Première version

```
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
```

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  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)
```

ANOVA analysis

```
print(params$title)

## [1] "Donnees ANOVA_ADEQ_GS.csv"
df <- read.table(params$file, header = TRUE, sep = ";")
```

Anova

Première version

```
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
```

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  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)
```

ANOVA analysis

```
print(params$title)

## [1] "Donnees ANOVA_ACC_Ortho.csv"
df <- read.table(params$file, header = TRUE, sep = ";")
```

Anova

Première version

```
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
```

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  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)
```

ANOVA analysis

```
print(params$title)

## [1] "Donnees ANOVA_ACC_Gram-synt.csv"
df <- read.table(params$file, header = TRUE, sep = ";")
```

Anova

Première version

```
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
```

Multiple Comparisons of Means: Tukey Constrasts

```
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)
```

ANOVA analysis

```
print(params$title)

## [1] "Donnees ANOVA_ACC_coherence.csv"
df <- read.table(params$file, header = TRUE, sep = ";")
```

Anova

Première version

```
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
```

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  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)
```