

Post-édition et traduction humaine en contexte académique : une étude empirique

Perrine Schumacher
Université de Liège et Université de Genève

Received: 30/04/2020

Accepted: 20/07/2020

Résumé

Plus que jamais, le traducteur doit acquérir de nouvelles compétences s'il veut être en phase avec les évolutions technologiques récentes et à venir. Depuis 2017, la demande en post-édition (PE) de traduction automatique (TA) ne cesse de croître et cette tâche se voit confiée essentiellement à des traducteurs. Si les progrès accomplis par la TA neuronale par rapport à la TA statistique sont indéniables, qu'en est-il de la qualité finale d'une PE par rapport à celle d'une traduction humaine (TH) ? Pour le savoir, nous avons mené une expérience contrôlée avec des étudiants en traduction pour la combinaison de langues anglais-français. Dans cet article, nous présentons les résultats de l'analyse comparative du corpus après qu'il a été soumis à une évaluation humaine. Nos résultats révèlent notamment un meilleur taux de réussite en PE de TA neuronale qu'avec les deux autres méthodes (TH et PE de TA statistique), ainsi qu'un effet nivelant de la PE de TA neuronale.

Mots-clés

Post-édition, traduction automatique neuronale, expérience contrôlée, évaluation de la qualité en traduction, étudiants en traduction.



Introduction

Alors que notre société est entrée de plain-pied dans l'ère de l'intelligence artificielle, pour le meilleur ou pour le pire, la tendance semble être à l'automatisation progressive de tâches complexes ; citons la conduite

autonome, la reconnaissance faciale et l'opération qui nous intéresse ici : la traduction automatique (TA).

Plus que jamais, le traducteur professionnel doit s'adapter et acquérir de nouvelles compétences s'il veut être en phase avec les évolutions technologiques récentes et à venir. Le paysage traductionnel ne cesse, en effet, de se redéfinir, favorisant l'émergence de nouvelles pratiques, de nouvelles tâches (localisation de jeux vidéo, pré- et post-édition, etc.) en lien avec le développement de nouveaux outils technologiques. La TA neuronale est aujourd'hui considérée par beaucoup comme une technologie très prometteuse, voire capable de concurrencer la traduction humaine¹ (TH) (Shoshan, 2018). Ainsi, nous proposons d'étudier une forme de collaboration humain-machine en plein regain d'intérêt dans le secteur de la traduction : la post-édition (PE). Il convient d'emblée de préciser que nous entendons le terme « post-édition » au sens de la définition qui en est donnée dans le rapport de la *Translation Automation User Society* (TAUS) (Massardo et al., 2016) : il s'agit du « 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]. »

L'arrivée de la TA neuronale (*Neural Machine Translation*) fait aujourd'hui grand bruit. Cette nouvelle génération de TA est le fruit des récents progrès accomplis dans le domaine de l'intelligence artificielle et plus précisément en *deep learning* ou « apprentissage profond ». Il s'agit d'une technologie qui est désormais utilisée par la majorité des moteurs de TA, supplantant de ce fait la TA statistique.

Contrairement à la TA statistique, ce nouveau modèle de TA se fonde sur une approche neuronale qui utilise un ensemble d'algorithmes d'apprentissage (réseaux de neurones artificiels) pour apprendre à représenter la langue source en caractéristiques complexes (*high-level features*). Ces représentations sont obtenues grâce à la méthode de « plongement lexical » (*word embedding*) et prennent la forme de vecteurs de nombres réels. Les réseaux de neurones, dont le fonctionnement nécessite une importante puissance de calcul informatique, sont entraînés à partir de gigantesques corpus composés de segments en langue source et de leurs traductions qui représentent des centaines de milliers voire

¹ Nous considérons une « traduction humaine » comme une traduction conventionnelle, réalisée sans recours à un moteur de TA.

des millions d'unités de traduction (Forcada, 2017). Plus le corpus d'entraînement est grand, meilleure sera la qualité des résultats générés par le moteur de TA et plus grande sera la capacité du système à s'autoaméliorer. Pour reprendre les termes de Barbara Cassin dans son *Éloge de la traduction* : « la qualité est, pour de bon et sans ironie, une propriété émergente de la quantité » (Cassin, 2016 : 143).

La présente étude s'inscrit dans le cadre d'une thèse doctorale menée en cotutelle entre l'université de Liège (Belgique) et l'université de Genève (Suisse). Les objectifs principaux de notre recherche sont d'explorer les potentiels effets de la PE sur la qualité de la production finale lorsque cette tâche est pratiquée dans un contexte académique, autrement dit par des étudiants en traduction, et de tirer les enseignements des résultats obtenus pour contribuer au perfectionnement de la formation initiale en traduction.

Afin de répondre à ces objectifs, nous avons décidé de mener une recherche expérimentale contrôlée auprès d'étudiants en traduction pour la combinaison de langues anglais-français. Dans cet article, nous présentons tout d'abord une brève revue de la littérature empirique avant de détailler notre méthodologie de recherche. Enfin, nous analysons les résultats des statistiques descriptives obtenus dans cette étude comparative de TH et de PE.

1. Revue de la littérature

Depuis quelques années déjà, le domaine de la TA fait l'objet d'un regain d'intérêt non seulement sur le marché de la traduction, mais aussi dans le monde de la recherche scientifique.

Bien que le pan de la recherche en traductologie consacré aux nouvelles technologies et plus particulièrement à la PE soit relativement jeune, nous avons pu recenser des dizaines d'études sur le sujet. Toutefois, 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 recherche à l'autre, ce qui rend les résultats obtenus souvent difficilement comparables. Malgré cette base de comparaison peu favorable, plusieurs études ont permis de nourrir et d'orienter notre réflexion.

1.1. Post-édition et traduction humaine

De la lecture de la littérature scientifique semble se dégager un 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 (Plitt et Masselot, 2010 ; Depraetere et al., 2014 ; Daems, 2016 ; Jia et al., 2019 ; Martikainen et Mestivier, 2020).

Outre ce gain de temps, il est légitime de s'interroger sur la qualité du produit fini. Plusieurs chercheurs se sont intéressés à cette question et se sont attelés à comparer le produit de la TH au produit de la PE à l'aide de diverses méthodes et dans divers contextes.

Certaines de ces études tendent à montrer qu'il n'y aurait pas de différence significative, en termes de qualité, entre des textes ou des phrases post-éditées et des traductions humaines (Garcia, 2011 ; Depraetere et al., 2014 ; Daems, 2016 ; Daems et al., 2017 ; Jia et al., 2019). Autrement dit, selon les conclusions des différentes études susmentionnées, le processus de PE n'aurait pas d'influence négative sur la qualité globale d'un texte cible et permettrait d'obtenir une production de qualité comparable à la qualité d'une TH.

C'est ce qui ressort, entre autres, de l'expérience contrôlée menée par Joke Daems (2016) auprès de plusieurs traducteurs professionnels et d'étudiants en traduction. Dans son étude comparative de TH et de PE pour la combinaison de langues anglais-néerlandais, Daems conclut que la PE permet non seulement de traduire plus rapidement qu'en TH, mais aussi d'aboutir à des productions de qualité globale comparable aux productions humaines, car les résultats ne montrent pas de différences statistiquement significatives entre les erreurs commises en PE et celles commises en TH.

Contrairement aux opinions qui semblent prévaloir pour l'heure dans le monde de la traduction, d'autres études ont été jusqu'à révéler que le recours à la PE de TA permettait d'aboutir à des textes cibles de meilleure qualité qu'en TH, c'est-à-dire à des productions soit qui contiennent moins d'erreurs que les traductions réalisées sans l'intervention d'une machine, soit qui ont été jugées comme étant de meilleure qualité. (Garcia, 2011 ; Martikainen et Mestivier, 2020 ; Plitt et Masselot, 2010).

Toutefois, dans la littérature empirique récente, certains auteurs tirent aussi des conclusions qui vont en sens inverse en démontrant que la PE entraîne des répercussions négatives sur la qualité des textes cibles. En effet, plusieurs publications ont mis en évidence que les textes post-édités que ce soit par des

étudiants en traduction ou par des professionnels de la traduction sont toujours jugés comme étant de moins bonne qualité que les TH (Čulo et al., 2014 ; Martikainen et Kübler, 2016 ; Sycz-Opoń et Galuskina, 2017).

Dans leur étude comparative réalisée avec 12 traducteurs professionnels et 12 étudiants pour la combinaison de langues anglais-allemand, Čulo et al. (2014) se sont intéressés aux divergences entre la PE et la TH sur les plans lexical et grammatical. Ils ont notamment observé l'apparition plus fréquente d'expressions non idiomatiques et de constructions non grammaticales dans les textes post-édités. Quant à Sycz-Opoń et Galuskina (2017), leur étude comparative (polonais – anglais/français) leur a permis de constater que, même après PE, les textes cibles contenaient encore beaucoup d'erreurs dues à la TA, elles ont également noté que les étudiants n'étaient pas parvenus à gommer l'étrangeté stylistique imputable à la TA.

Nous pensons que la divergence de conclusions que nous venons de résumer ici tient principalement à la diversité des systèmes d'évaluation employés et des types de textes évalués ainsi qu'aux différentes attentes en termes de qualité. Il convient également de signaler que parmi les études susmentionnées, très peu ont été menées pour la combinaison de langues anglais-français. Ce constat nous a poussée à conduire notre propre recherche expérimentale dont nous détaillons la méthodologie au point suivant.

2. Méthodologie

2.1. Participants

Les participants à cette expérience étaient 28 étudiants (21 filles et 7 garçons) inscrits en 3^e année de Bachelier traduction–interprétation de l'Université de Liège (ULiège). La langue maternelle de ces étudiants est le français à l'exception d'un étudiant germanophone.

Leur participation s'est faite sur une base volontaire. Avant de prendre part à cette expérience, chaque participant a été invité à signer un formulaire de consentement libre et éclairé reprenant les 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).

Pour mieux connaître l'opinion des étudiants sur la TA et sur la PE ainsi que leur expérience en la matière et enfin pour avoir un retour sur le test, nous avons conçu deux courts questionnaires à remplir en ligne via LimeSurvey. Le premier questionnaire était à remplir en début de session et le deuxième en fin de session. L'analyse détaillée des réponses récoltées grâce à ces deux questionnaires n'est pas reprise dans cet article.

Bien que ces étudiants aient reconnu avoir recours à la TA que ce soit quotidiennement pour certains ou bien occasionnellement pour d'autres, ils ont déclaré avoir peu d'expérience en PE, voire pas du tout. En outre, aucun étudiant n'a suivi de formation en TA et PE avant de prendre part à cette recherche expérimentale, l'une des raisons étant certainement que l'ULiège n'offre actuellement aucune formation de ce type. Il nous a néanmoins semblé opportun d'examiner les stratégies intuitives de PE de ces étudiants, car parmi les traducteurs professionnels à qui l'on confie aujourd'hui des tâches de PE, très peu ont reçu une formation spécifique en la matière.

2.2. *Sélections des textes sources et traduction*

Étant donné que nous voulons à explorer les effets de la PE dans un contexte d'enseignement, nous avons cherché, tant que faire se peut, à reproduire l'environnement des cours de traduction dispensés à l'ULiège.

À cet effet et afin de limiter la subjectivité relative à la sélection des textes sources dans notre expérience, 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. Nous avons tenu à nous concentrer sur trois types de textes différents : un article de presse, un article économique et 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 qui est un trouble de la perception sensorielle.

L'évaluation de la complexité textuelle en vue d'une traduction semble dépendre de tellement 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 dispensent aux étudiants de la filière traduction-interprétation.

Pour la tâche de TH, l'étudiant pouvait disposer de toutes les ressources qu'il jugeait utiles et nécessaires (dictionnaires explicatifs et traductifs, glossaires, Internet, etc.). Il lui était toutefois interdit d'avoir recours à un moteur de TA.

Pour les tâches de PE, les deux moteurs de TA que nous avons utilisés sont *Bing Translator* comme système de TA statistique et *DeepL Translator* comme système de TA neuronale. Les traductions automatiques ont été générées le 23 janvier 2018. À titre informatif, le tableau ci-dessous présente les scores obtenus grâce à la métrique automatique BLEU (Papineni et al., 2002) pour chaque texte et pour chaque système de TA². Pour diverses raisons que nous évoquons au point 2.4.1, ces résultats sont difficilement interprétables et nous disent bien peu de choses sur la qualité réelle de la TA brute. Nous avons toutefois tenu à les indiquer à toutes fins utiles.

| Score BLEU | Texte 1 | Texte 2 | Texte 3 |
|-----------------------------|---------|---------|---------|
| TA statistique brute | 20,24 | 24,59 | 22,20 |
| TA neuronale brute | 20,30 | 36,14 | 29,41 |

Tableau 1. Scores BLEU par texte et par système de TA.

Robert (2010) distingue deux grands types de PE : la PE brute et la PE évoluée professionnelle. Le premier type consiste en la révision, modification, correction, remaniement et relecture directe d'un texte traduit par un moteur de TA, tandis que le deuxième type est produit par un processus qui associe plusieurs technologies de TA avec des outils de traduction assistée par ordinateur (TAO), ce deuxième type étant très répandu dans le domaine privé.

² Ces scores BLEU ont été calculés sur Tilde (<https://www.tilde.com>) en utilisant comme traduction de référence soit la traduction humaine ayant reçu la meilleure note dans notre corpus, soit la traduction humaine collective réalisée par les étudiants en classe bien après l'expérience.

Dans un contexte pédagogique, nous avons décidé de nous concentrer sur la PE dite « brute » à partir de moteurs de TA disponibles gratuitement en ligne.

Toujours pour les deux tâches de PE, il a été demandé aux étudiants de réaliser une PE complète (*full post-editing*) dans le but d'aboutir à une production de qualité comparable ou égale à la qualité d'une TH. Ils ont reçu, pour ce faire, une liste de consignes de PE que nous avons reprises de la TAUS et du *Centre for Next Generation Localisation* (CNGL) (2010) et légèrement adaptées aux fins de notre étude.

La 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. »

2.3. Déroulement de l'expérience

Cette expérience contrôlée s'est déroulée en une seule session qui a duré 3h30 environ. Les 28 étudiants ont été regroupés dans un amphithéâtre, ils disposaient de leur ordinateur portable personnel.

Durant l'expérience *stricto sensu*, les étudiants ont été amenés à traduire, de l'anglais vers le français standard, les trois textes sources sélectionnés préalablement selon trois méthodes de traduction différentes.

Il a été demandé à chaque étudiant de produire :

- une traduction humaine (sans recours à un moteur de TA)
- une PE de TA brute générée par un moteur de TA statistique (en l'occurrence BingTranslator) ;
- une PE de TA brute générée par un moteur de TA neuronale (en l'occurrence DeepL Translator).

L'étudiant n'était pas en mesure d'identifier le modèle de TA utilisé (TA statistique ou TA neuronale).

Afin d'éviter tout biais, tout effet d'ordre et afin que chaque texte soit également traduit selon les trois méthodes de traduction différents, nous avons envisagé la répartition des tâches selon un plan croisé (principe du carré latin)

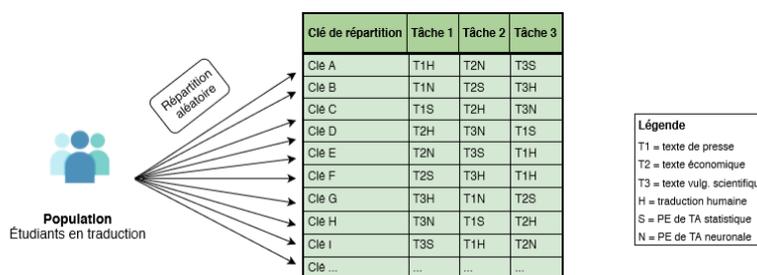
entre les facteurs suivants : textes, méthodes de traduction et ordre dans lequel les tâches devaient être accomplies (graphique 1).

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

Aucune limite de temps par tâche ne leur a été imposée. Tous les étudiants ont terminé l'expérience en moins de 3h30.

Ces tâches ont été réalisées en utilisant l'outil *MateCat*³. Nous avons choisi cet outil de TAO en ligne, car il s'agit d'un outil gratuit, *open source* qui est aussi très intuitif et personnalisable. De plus, il nous a permis ensuite de disposer d'informations précieuses telles que l'effort de PE en pourcentage et la durée de PE.

Après exclusion de plusieurs productions pour des raisons de représentativité statistique, notre corpus final se compose de 75 productions (25 TH, 25 PE de TA statistique et 25 PE de TA neuronale).



Graphique 1. Répartition des tâches selon un plan croisé (principe du carré latin). (Schumacher, 2019)

2.4. Exploitation du corpus

Étant donné que notre objectif principal est 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

³ MateCat est accessible gratuitement en ligne : <https://www.matecat.com>

manière totalement anonyme, il nous était impossible de relier la production à l'étudiant, à quelque étape que ce soit.

2.4.1. *Évaluation de la qualité en traduction et critères adoptés*

Unless 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. (Depraetere et al., 2014 : 99)

L'évaluation de la qualité en traduction est, en soi, un vaste sujet, ô combien délicat et subjectif, dont l'étude dépasse largement le cadre de cet article. C'est pourquoi nous nous limiterons à décrire le type d'évaluation que nous avons décidé d'adopter dans le cadre de cette expérience.

Nous exposons ici les raisons pour lesquelles nous avons opté pour une évaluation humaine de la qualité. Étant donné qu'il s'agit d'une étude expérimentale en contexte pédagogique, il ne nous a pas semblé opportun d'avoir recours à des systèmes d'évaluation qui ont été développés en contexte de pratique professionnelle et qui sont essentiellement utilisés sur le marché professionnel de la traduction, telles les métriques «LISA QA», «DQF-MQM», etc.

Nous n'avons pas non plus souhaité utiliser de métriques spécifiques à l'évaluation de la TA, telles que BLEU, ROUGE, NIST, METEOR, etc. Parmi les nombreuses limites de ces métriques, citons le fait que, souvent, les deux critères d'évaluation employés sont trop simples et rudimentaires par rapport aux critères généralement appliqués dans l'évaluation d'une TH ; il s'agit essentiellement de juger la fluidité du texte cible (*fluency*) et la fidélité du texte cible au texte source (*adequacy*) sans procéder à une analyse détaillée des erreurs commises. Comme le soulignent très bien Toudic et al. (2014), les critères utilisés par ces métriques d'évaluation sont insuffisants :

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. (Toudic et al., 2014 : 25)

De plus, ces métriques se fondent sur le principe de « l'étalon-or » (*gold standard*) qui se heurte à la « variabilité constitutive des traductions » (Braunstein, 2016) ; la qualité d'une traduction produite par un moteur de TA est évaluée par comparaison à une ou plusieurs traductions de référence qui ont été effectuée(s) par des traducteurs humains.

Enfin, selon Koehn (2010), il est, en effet, difficile de croire que de telles métriques soient capables de refléter de quelconques différences sémantiques entre la TA brute et la (les) traduction(s) de référence ou même le texte source.

Au vu des objectifs que nous nous sommes fixés dans cette étude, ces métriques ne se révéleront donc pas d'une grande utilité. C'est pourquoi nous nous sommes concentrés uniquement sur une évaluation humaine de la qualité. En outre, il nous a semblé pertinent que les productions recueillies soient évaluées selon une typologie d'erreurs qui soit la plus proche possible de la manière dont elles sont jugées au cours des études universitaires en traduction. Malheureusement, 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, chacun ayant développé sa propre façon de faire au-delà d'un socle commun de critères d'évaluation. Cette absence de méthode nous a conduite à concevoir notre propre typologie.

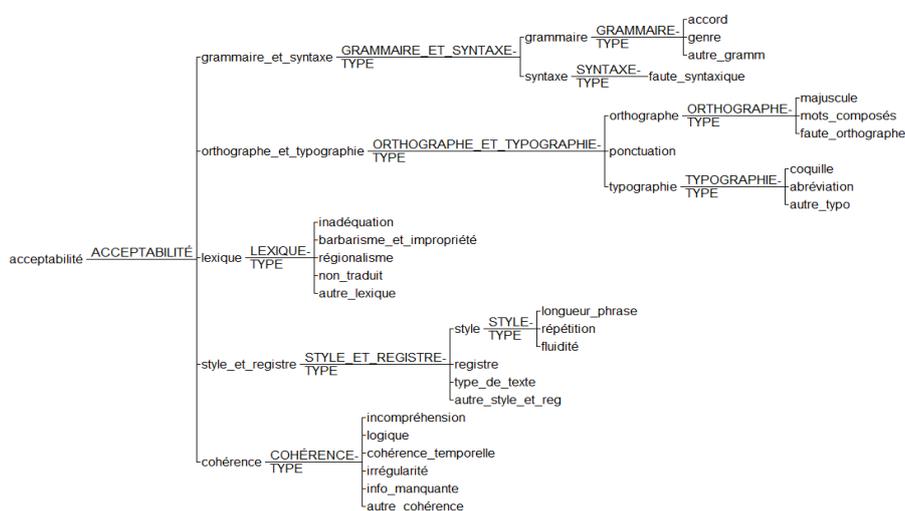
Pour ce faire, nous nous sommes inspirée, d'une part, des principaux critères communément employés au sein de la filière traduction-interprétation de l'ULiège et, d'autre part, de la typologie d'erreurs développée par Daems (2016) dans le cadre de sa thèse de doctorat intitulée *A translation robot for each translator? A comparative study of manual translation and post-editing of machine translations: process, quality and translator attitude*. Cette typologie nous a semblée particulièrement adaptée à notre contexte de recherche et à la pratique pédagogique. Nous avons trouvé son idée de procéder à une évaluation de la qualité en deux phases très judicieuse notamment, car cette division permet une analyse fine des erreurs et de leur origine. Ces deux phases font directement écho aux notions d'*acceptability* et d'*adequacy* proposées par Toury (1995).

Nous avons donc repris cette idée et avons prévu que l'évaluation se fasse en deux temps :

- Une 1^{ère} phase qui permet d'évaluer « l'acceptabilité » du texte cible (fluidité)

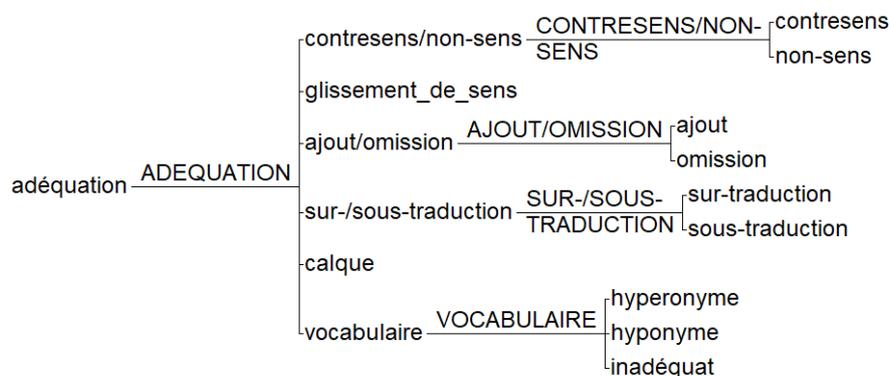
- Une 2^e phase qui sert à évaluer « l'adéquation » du texte cible au texte source (fidélité).

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 (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). Pour cette étape, l'évaluateur humain dispose uniquement du texte cible en français afin de pouvoir se placer dans la peau du récepteur.



Graphique 2. Phase d'acceptabilité

La deuxième phase permet de juger de « l'adéquation » du texte cible par rapport au texte source. Il s'agit ici de repérer les fautes de traduction, c'est-à-dire les fautes de transfert de sens (contresens, glissement de sens, calque, ajout, omission, etc.) (graphique 3). Pour cette étape, l'évaluateur dispose à la fois du texte cible et du texte source.



Graphique 3. Phase d'adéquation

Autrement dit et à l'instar de la typologie conçue par Daems (2016), si l'évaluateur repère l'erreur en ayant uniquement le texte cible sous les yeux, celle-ci sera notée durant la phase d'acceptabilité tandis que si l'erreur n'apparaît que lorsque l'évaluateur a le texte cible et le texte source sous les yeux, elle sera prise en compte dans la phase d'adéquation.

Les productions recueillies durant l'expérience ont ainsi été évaluées et annotées selon cette typologie par plusieurs juges humains, à raison de deux enseignantes par type de texte. Il s'est agi en l'occurrence d'enseignantes en traduction de notre filière avec plus de dix ans d'expérience dans l'enseignement de la traduction (Bachelier et Master) et de nombreuses années d'expérience en tant que traductrice indépendante pour deux d'entre elles. Pour ce faire, ces enseignantes 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). Nous tenons à préciser que les deux phases d'évaluation se sont, bien entendu, faites à l'aveugle ; l'évaluatrice ne savait pas si la production qu'elle jugeait était une TH ou une PE de TA.

Il leur a été demandé de souligner chaque erreur dans le texte cible et de lui assigner une étiquette selon les différentes catégories et sous-catégories préétablies (graphiques 2 et 3), une même erreur ne pouvant être soulignée qu'une seule fois. Un extrait des consignes qui ont été distribuées aux évaluateurs et qui décrivent les différentes catégories d'erreurs avec exemples est repris au bas de cet article (annexe 1).

Nous avons également prévu une étape de consolidation entre les enseignantes afin qu'elles puissent discuter des éventuelles divergences d'évaluation et adapter au besoin certaines annotations. Les valeurs d'accord interévaluatrices sont toutes supérieures ou égales à 70 pour cent.

Après cette étape, les différentes erreurs ont ensuite été comptabilisées en calculant la moyenne des erreurs entre les annotations des différentes évaluateurs. Ensuite, une note globale sur 20 a été attribuée à chaque production selon une pondération prédéfinie (tableau 2).

Toujours dans l'objectif de reproduire le contexte d'enseignement dans lequel évoluent les étudiants 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. Le tableau 2 présente en détail la pénalité qui a été attribuée à chaque catégorie d'erreur.

| Pondération | | | | | | | |
|-------------------------|-----------|---------------------|--------------------|----------------|----------------------|-------------------------|-------------|
| Note sur 20 | | | | | | | |
| Acceptabilité (40 %) | Catégorie | Grammaire /Syntaxe | Orthographe | Style/registre | Lexique | Typographie/Ponctuation | Cohérence |
| | Pénalité | 1,0 | 1,0 | 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 utilisée pour la notation sur 20.

2.5. Hypothèse

En nous appuyant sur les résultats de Čulo et al. (2014), Martikainen et Kübler (2016), ainsi que Sycz-Opoń et Gałuskina (2017), nous avons formulé l'hypothèse que, pour la paire de langues anglais-français, la PE de TA effectuée en contexte académique influe sur la qualité d'un texte cible. Nous nous attendons à ce que les textes post-édités contiennent, de manière

générale, un plus grand nombre d'erreurs et obtiennent une moins bonne note que les traductions humaines.

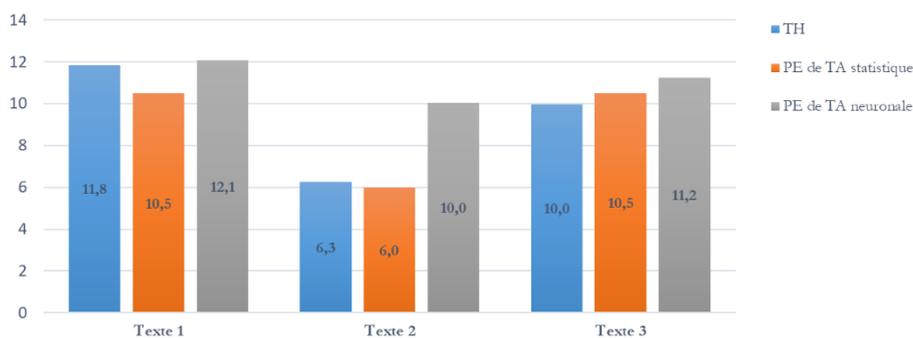
Nous tenons également à vérifier si la qualité d'un texte post-édité dépend du système de TA utilisé (TA statistique ou TA neuronale). Nous pensons, en effet, que la PE de TA neuronale permettra d'obtenir des productions de meilleure qualité qu'en PE de TA statistique.

3. Résultats

À des fins de concision, nous avons décidé de nous limiter dans cet article à l'analyse des résultats des statistiques descriptives. Dans un souci de clarté, nous allons parcourir différents graphiques qui permettent de visualiser les résultats obtenus.

3.1. Comparaison de la qualité des productions par méthode de traduction

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).

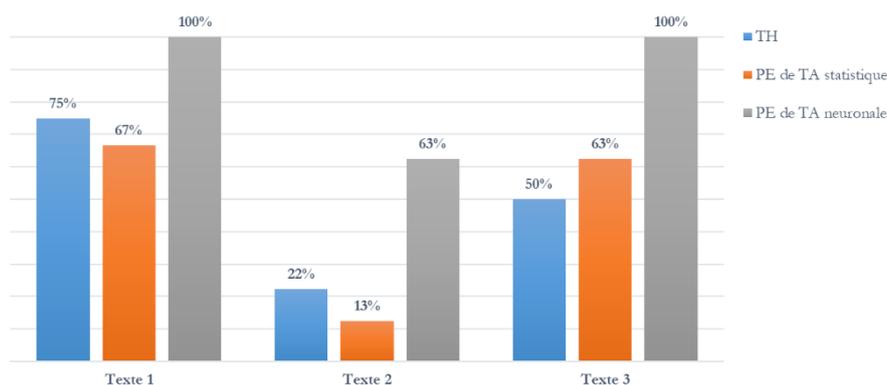


Graphique 4. Moyenne des notes obtenues selon la méthode de traduction et le type de texte source

Sur le graphique 4, nous pouvons observer 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).

3.2. Taux de réussite

Nous nous sommes également penchée sur le taux de réussite selon la méthode de traduction, la note de réussite à l'Université étant fixée à 10/20 (contre 12/20 auparavant). 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.

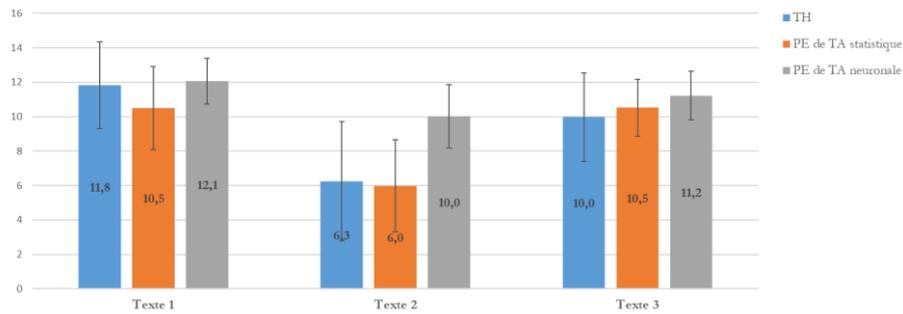


Graphique 5. Taux de réussite selon la méthode de traduction et le texte source.

Sur le graphique 5, nous pouvons constater qu'il semblerait que la méthode de traduction influe sur le taux de réussite, et ce, pour les trois textes sources. La PE de TA neuronale permet ainsi d'obtenir systématiquement un meilleur taux de réussite qu'en PE de TA statistique ou qu'en TH. Ainsi, nous pouvons affirmer que, dans cette étude, la probabilité qu'un étudiant obtienne une note de réussite (supérieure ou égale à 10/20) est plus élevée lorsque ce dernier post-édite de la TA neuronale que lorsqu'il traduit humainement ou encore que lorsqu'il post-édite de la TA statistique.

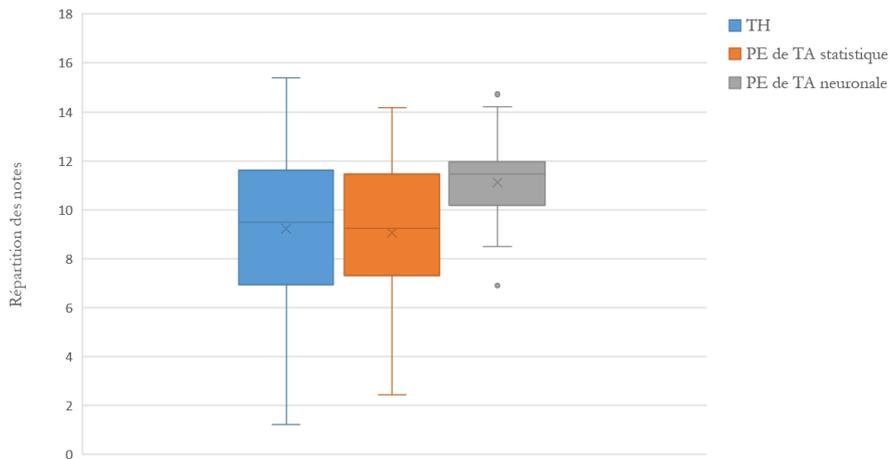
3.3. Moyenne des notes et écart-type

Les graphiques 6 et 7 nous permettent de comprendre un peu mieux les résultats obtenus en matière de taux de réussite. Nous avons repris sur ce graphique la moyenne des notes obtenues en fonction de la méthode de traduction et du texte source ainsi que l'écart-type représenté par la barrette.



Graphique 6. Moyenne des notes et écart-type.

Sur le graphique 6, 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 méthodes.

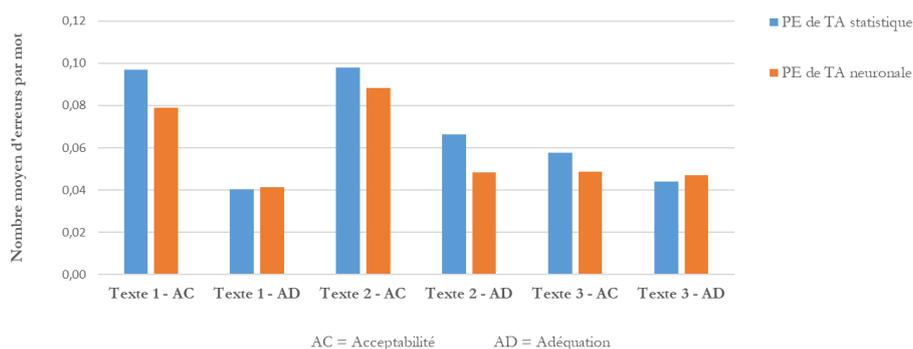


Graphique 7. Répartition des notes selon la méthode de traduction

Le graphique 7 en boîte à moustaches illustre clairement ce resserrement des notes obtenues. Nous avons repris ici l'ensemble des notes obtenues selon la méthode de traduction, indépendamment du texte source. Pour chaque méthode 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 ; le rectangle délimité par le 1er quartile (Q1) et le 3e 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. Ce graphique nous permet, en outre, 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 méthodes de traduction. Le resserrement des notes se traduit ici par une réduction de l'écart interquartile (Q3-Q1) : 5,09 en TH ; 4,30 en PE de TA statistique et 2,03 en PE de TA neuronale.

3.4. Comparaison des systèmes de TA

Nous avons cherché également à vérifier notre deuxième hypothèse, à savoir si la qualité d'un texte post-édité dépend du système de TA employé (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 selon le niveau moyen d'erreurs comptabilisées par texte source.



Graphique 8. Comparaison des systèmes de TA (statistique et neuronal) d'après le nombre moyen d'erreurs par mot du texte source.

Sur le graphique 8, 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 pourquoi nous avons pris comme indicateur le nombre moyen d'erreurs par mot du texte source.

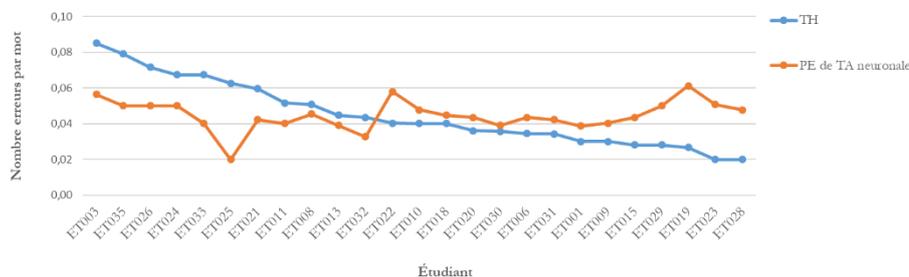
Les résultats représentés sur ce graphique semblent indiquer que la PE de TA neuronale permet d'aboutir à des productions de meilleure qualité, c'est-à-dire, contenant, en moyenne, moins d'erreurs que les PE de TA statistique, à l'exception des résultats obtenus pour le texte 1 et le texte 3 en phase d'adéquation, où les PE de TA neuronale contiennent légèrement plus d'erreurs par rapport aux PE de TA statistique.

3.5. Résultats par étudiant

Enfin, au vu des résultats auxquels nous sommes arrivée, nous avons cherché à savoir si les étudiants, pris 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.

3.5.1 Nombre moyen d'erreurs

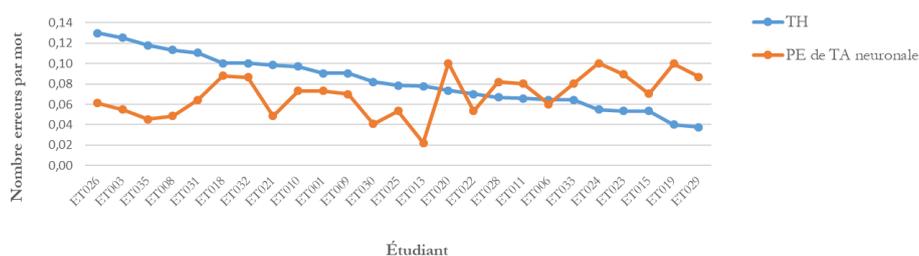
Une fois de plus, nous avons normalisé les résultats en prenant comme référence le nombre d'erreurs par mot du texte source. Sur les graphiques 9 et 10, les traductions humaines (TH) sont classées de manière croissante : de la production qui contient le plus grand nombre d'erreurs par mot à celle qui en contient le moins, autrement dit, de la moins bonne TH à la meilleure, indépendamment du texte source.



Graphique 9. Résultats par étudiant pour la phase d'adéquation.

Sur le graphique 9, la TH de l'étudiant 22 (ET022) se situe à la médiane (0,04). Nous pouvons d'emblée constater cette tendance : les étudiants 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 tous les étudiants 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, tous les étudiants 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.

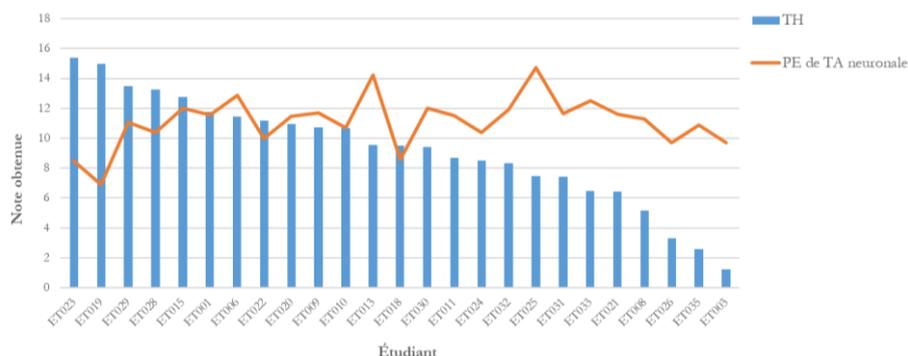
Nous observons également cette tendance, toutefois de manière moins marquée, pour les erreurs de langue (phase d'acceptabilité) sur le graphique 10. La médiane se situe, cette fois, à 0,06 (ET006 et ET033).



Graphique 10. Résultats par étudiant pour la phase d'acceptabilité.

3.5.2. Notation

Nous avons voulu vérifier si cette tendance était également observable au niveau de la notation. Sur le graphique 11, les TH sont classées à présent de manière décroissante : de la meilleure production à la moins bonne.



Graphique 11. Notes obtenues par étudiant.

Les résultats obtenus en comparant les notes obtenues en TH et en PE de TA neuronale pour un même étudiant nous permettent également d’observer cet « effet nivelant » de la PE de TA neuronale. Les étudiants que nous considérons comme les plus faibles, à savoir ceux dont la note arrondie en TH n’atteint pas le seuil de réussite de 10/20 — ET030 jusque ET004 inclus — sont ceux qui obtiennent systématiquement une meilleure note en PE de TA et donc, nous en déduisons que ce sont ceux qui bénéficient le plus de cette méthode de traduction. À l’inverse, les étudiants dont nous considérons les productions humaines comme de bonne qualité (note supérieure ou égale à 12/20) — ET023 à ET015 inclus — ont automatiquement obtenu une note plus basse en PE de TA neuronale.

4. Discussion

Les résultats partiels de cette première recherche expérimentale viennent infirmer en grande partie notre hypothèse de départ qui était que la PE effectuée par des étudiants peut avoir une influence négative sur la qualité d’un texte cible et se repercuter sur la notation.

En effet, contrairement à ce que nous pressentions, nous avons relevé une moyenne des notes plus élevée, ainsi qu’un meilleur taux de réussite en PE de TA neuronale par aux résultats obtenus en PE de TA statistique et en TH. Mais comment expliquer ce meilleur taux de réussite ? L’hypothèse interprétative que nous proposons est que la TA neuronale permettrait à certains étudiants de pallier la compréhension lacunaire ou partielle qu’ils ont

du texte source (terminologie, logique interne, etc.). La TA neuronale leur donnerait, en quelque sorte, accès au sens du texte source. Dans notre expérience, ce phénomène est particulièrement flagrant dans le cas du texte 2 pour lequel le taux de réussite est extrêmement faible aussi bien en PE de TA statistique (13 %) qu'en TH (22 %) alors que le taux de réussite en PE de TA neuronale atteint, lui, 63 % (graphique 5). Nous pourrions également supposer que ces étudiants 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.

En outre, nous avons observé des écarts-types réduits en PE par rapport à la TH (graphique 6) qui reflètent un resserrement des notes autour de la moyenne (c'est particulièrement le cas pour la PE de TA neuronale, voir graphique 7), ainsi qu'une relation inverse entre le niveau des étudiants en TH et la qualité de leur PE de TA neuronale (graphiques 8, 9 et 10). Selon nous, ces observations traduisent un effet nivelant de la PE de TA neuronale. D'un côté, ces résultats nous amènent à considérer que la PE de TA neuronale permet aux étudiants les plus faibles d'augmenter la qualité de leur production finale et d'obtenir une meilleure note que lorsqu'ils traduisent humainement. D'un autre côté, nous pensons que cette méthode de traduction aurait également un effet limitant, car elle entraînerait, chez les « bons » étudiants, une baisse de qualité du texte cible. Les étudiants faibles tireraient donc davantage parti de la PE de TA neuronale que les autres étudiants. En revanche, il est déconcertant de constater que les bons étudiants seraient, eux, désavantagés par la PE de TA neuronale qui ne leur permet pas d'aboutir à des productions d'aussi bonne qualité que lorsqu'ils traduisent humainement. Ces résultats vont dans le même sens que les observations de Garcia (2011) et Killman (2018) qui ont, tous deux, déjà évoqué cet effet nivelant de la PE (*leveling effect*) chez les étudiants. Au vu d'un tel constat, nous pensons qu'un des effets négatifs de la PE de TA neuronale est de brider, en quelque sorte, les bons éléments en restreignant leurs intuitions de traduction et en limitant leur créativité traductionnelle. Nous pourrions dès lors légitimement nous demander s'il est pertinent d'envisager une formation en PE pour tous les profils d'étudiants en traduction.

Les différentes interprétations que nous venons de formuler dans cette partie, en particulier concernant l'effet nivelant de la PE de TA neuronale, mériteraient d'être approfondies et de faire l'objet de futurs travaux de recherche.

Conclusion

La présente étude comparative nous aura permis d'explorer les potentiels effets de la PE en contexte académique sur la qualité du texte cible.

Il serait prématuré de tirer ici des conclusions générales, cependant les résultats obtenus dans cette recherche expérimentale nous amènent à infirmer notre hypothèse principale, car la PE ne semble pas avoir eu d'influence négative sur la qualité globale des productions des étudiants. En effet, dans notre expérience, les textes post-édités n'ont pas obtenu de moins bonnes notes que les deux autres méthodes de traduction, bien au contraire.

Tout d'abord, nous avons observé une moyenne des notes plus élevée, ainsi qu'un meilleur taux de réussite en PE de TA qu'en PE de TA statistique et qu'en TH (graphique 4 et 5). Dès lors, nous pouvons soutenir que dans cette étude, la probabilité qu'un étudiant de 3^e bachelier obtienne une note de réussite (supérieure ou égale à 10/20) est plus grande en PE de TA neuronale qu'elle ne l'est dans les deux autres méthodes de traduction.

Nos résultats indiquent également que la qualité d'un texte post-édité semble dépendre du système de TA utilisé (TA statistique ou TA neuronale). Effectivement, dans cette étude, la PE de TA neuronale a permis d'aboutir à des productions qui contiennent, de manière globale, moins d'erreurs que les PE de TA statistique (graphique 8). Ces résultats confirment d'ailleurs ceux obtenus dans l'étude comparative entre la PE de TA statistique et la PE de TA neuronale (anglais-japonais) menée par Yamada (2019) auprès d'étudiants universitaires.

Par ailleurs, nos résultats tendent à montrer l'existence d'un effet nivelant en PE de TA neuronale chez les étudiants en traduction. 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 (graphiques 6 et 7) et par l'établissement d'une relation inverse entre le niveau d'un étudiant en TH et la qualité de sa PE de TA neuronale (graphiques 8, 9 et 10). Cette relation peut s'interpréter comme suit : plus l'étudiant est faible en TH, plus la PE lui permet d'augmenter la qualité de ce qu'il produit et à l'inverse, plus l'étudiant est compétent en TH, plus la PE entraîne une baisse de qualité de sa production finale. Les étudiants faibles tireraient donc davantage profit de la PE de TA neuronale que les autres étudiants. A contrario, les bons étudiants, outre le fait de ne tirer aucun avantage de la PE de TA neuronale, seraient desservis par

cette méthode de traduction qui ne leur permet pas de produire des textes cibles d'aussi bonne qualité qu'en TH.

Enfin, nous tenons à souligner les principales limites de ce travail. En effet, nous sommes consciente que certains de nos choix méthodologiques ont entraîné plusieurs limites dans cette étude : le choix d'une méthode d'échantillonnage non probabiliste, l'intervention de l'autrice de cette étude dans l'évaluation humaine des productions recueillies et de surcroît, le fait que l'évaluation de la qualité en traduction reste une tâche empreinte de subjectivité, et ce, même si nous avons veillé à avoir recours à plusieurs évaluateurs humains ainsi qu'à une typologie d'erreurs commune. Nous tenons à souligner qu'il ne s'agit là que des résultats partiels d'une première expérience qui nécessite d'être reproduite pour confirmer et élargir ces conclusions.

Remerciements

Nous souhaitons remercier particulièrement trois enseignantes de la filière traduction-interprétation de l'ULiège, Valérie Maris, Bénédicte Klinkenberg et Valérie Bada, pour leur aide et leur soutien précieux dans ce projet.

Nous tenons également à remercier la Professeure Pierrette Bouillon de l'Université de Genève pour sa relecture minutieuse et ses conseils avisés.

Références

- Braunstein, N. A. (2016) *Traduire la psychanalyse : Interprétation, sens et transfert*. Toulouse, France, ERES.
- Carl, M., Dragsted, B., Elming, J., Hardt D. et Lykke, A. (2011) "The process of post-editing: A pilot study." *Copenhagen Studies in Language*, 41, pp. 131-142.
- Cassin, B. (2016) *Éloge de la traduction. Compliquer l'universel*. Paris, Fayard.
- Čulo, O., Gutermuth, S., Hansen-Schirra, S. et Nitzke J. (2014) "The Influence of Post-Editing on Translation Strategies". In O'Brien, S., Winther-Balling, L., Carl, M., Simard, M. et Specia, L. (eds.) *Post-editing of Machine Translation: Processes and Applications*. Newcastle upon Tyne, Cambridge Scholars Publishing, pp. 200-218.
- 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.

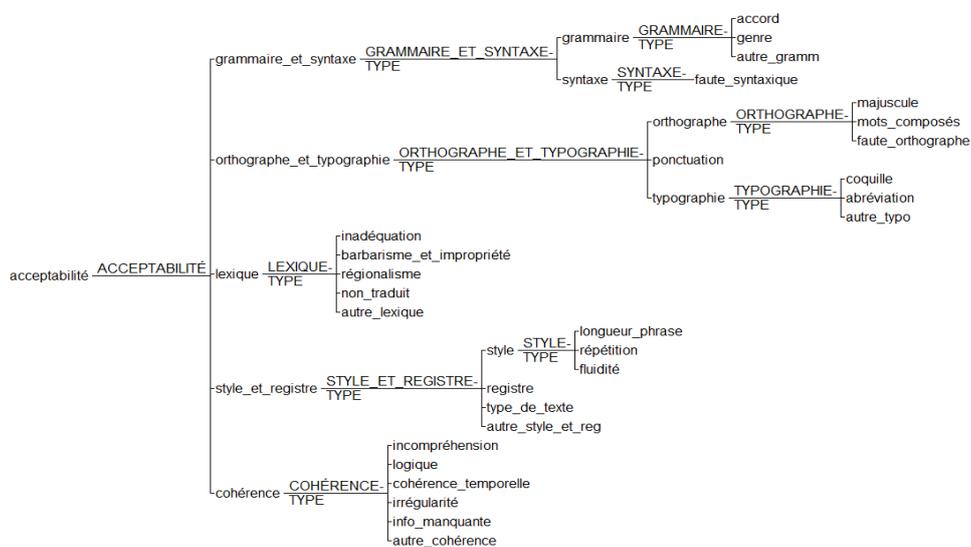
- Daems, J., De Clercq, O. et Macken, L. (2017) “Translationese and Post-editedese: How Comparable Is Comparable Quality?” *Linguistica Antverpiensia New Series-themes in Translation Studies*, 16, pp. 89-103. [En ligne] [Consulté le 3 mars 2020] <http://hdl.handle.net/1854/LU-8516838>
- Depraetere, I., De Sutter, N. et Tezcan, A. (2014) “Post-edited quality, post-editing behaviour and human evaluation: a case study.” In O’Brien, S., Winther-Balling, L., Carl, M., Simard, M. et Specia, L. (eds.) *Post-editing of Machine Translation : Processes and Applications*. Newcastle upon Tyne, Cambridge Scholars Publishing, pp. 78-108.
- Forcada, M. L. (2017) “Making sense of neural machine translation.” *Translation Spaces*, 6 (2), pp. 291-309. [En ligne] [Consulté le 2 août 2019] DOI 10.1075/ts.6.2.06for
- Garcia, I. (2011) “Translating by post-editing: Is it the way forward?” *Machine Translation*, 25, pp. 217–237. [En ligne] [Consulté le 15 février 2020] DOI 10.1007/s10590-011-9115-8
- 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, pp. 60-86. [En ligne] [Consulté le 18 juin 2019] https://www.jostrans.org/issue31/art_jia.pdf
- 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 (2), pp. 114-141. [En ligne] [Consulté le 23 mai 2019] DOI 10.1075/jial.18003.kil
- Koehn, P. (2010) *Statistical Machine Translation*. New York, Cambridge University Press.
- Martikainen H. et Kübler, N. (2016) «Ergonomie cognitive de la post-édition de traduction automatique : enjeux pour la qualité des traductions». *ILCEA*, 27. [En ligne] [Consulté le 3 mars 2020] DOI 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’études Traduction & Qualité 2020 : Biotraduction et traduction automatique, Université de Lille, France.
- Massardo, I., van der Meer, J. et Khalilov, M. (2016) *TAUS translation technology landscape report*, De Rijp : TAUS.
- O’Donnell, M. (2016) *The UAM CorpusTool 3.3*. <http://corpustool.com>
- Papineni, K., Roukos, S., Ward, T. et Zhu, W.J. (2002) “Bleu: a method for automatic evaluation of machine translation.” *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 311-318.
- Plitt, M. et Masselot, F. (2010) “A productivity test of statistical machine translation post-editing in a typical localisation context.” *The Prague Bulletin of Mathematical Linguistics*, 93, pp. 7-16.

- Robert, A.-M. (2010) «La post-édition : l'avenir incontournable du traducteur ?» *Traduire*, 222, pp. 137-144. [En ligne] [Consulté le 8 avril 2020] DOI 10.4000/traduire.460
- Schumacher, P. (2019) «Avantages et limites de la post-édition», *Traduire* 241, pp. 108-123. [En ligne] [Consulté le 18 juillet 2020] DOI 10.4000/traduire.1887
- Shoshan, O. (2018) *Make Neural Machine Translation Better, Faster – A New Way to Measure NMT Quality*. Slator [En ligne] [Consulté le 4 décembre 2019] <https://slator.com/sponsoredcontent/make-neural-machine-translation-better-faster-a-new-way-to-measure-nmtquality/>
- Sycz-Opoń, J. et Galuska, K. (2017) “Machine Translation in the Hands of Trainee Translators – an Empirical Study.” *Studies in Logic, Grammar and Rhetoric*, 49 (1), pp. 195-212.
- TAUS/CNGL (2010) *Consignes relatives à la post-édition des traductions automatiques*. [En ligne] [Consulté le 12 février 2020] <https://www.taus.net/academy/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines-french>
- 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. [En ligne] [Consulté le 3 mars 2020] <http://ilcea.revues.org/2517>
- Toury, G. (1995) “The Nature and Role of Norms in Translation”. In Toury, G. (ed.) *Descriptive Translation Studies and Beyond*. Amsterdam-Philadelphia, John Benjamins, pp. 53-69.
- Yamada, M. (2019) “The impact of Google Neural Machine Translation on Post-editing by student translators”. *The Journal of Specialised Translation*, 31, pp. 87-106. [En ligne] [Consulté le 18 juin 2019] https://www.jostrans.org/issue31/art_yamada.php

Annexe 1

Consignes détaillées pour l'évaluation humaine Acceptabilité (*Acceptability*)

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/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 n'entre dans aucune autre catégorie.

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 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 au 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 n'entre 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. : « abrévier » 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 n'entre 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 n'entre 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**

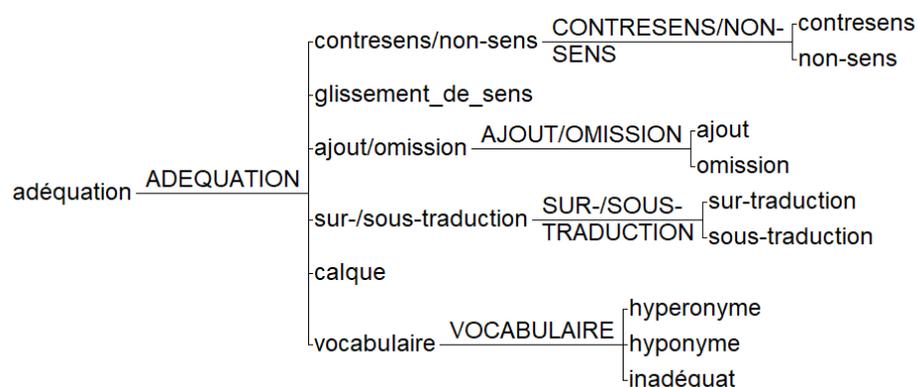
Information manquante

- **Autre_cohérence**

Merci de préciser en quoi cette erreur n'entre dans aucune autre catégorie.

Adéquation (*Adequacy*)

APERÇU DES SOUS-CATEGORIES



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 implicitation 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 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**

« 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