La main de Thôt
ISSN : 2272-2653
Éditeur : Carole Filière

9 | 2021
La traduction littéraire et SHS à la rencontre des nouvelles technologies de la traduction : enjeux, perspectives et défis

Collaboration homme-machine dans la traduction des métadonnées en SHS : expérience de traduction automatique post-éditée pour quatre revues françaises

Katell Hernandez Morin et Franck Barbin

http://interfas.univ-tlse2.fr/lamaindethot/986

Référence électronique
Katell Hernandez Morin et Franck Barbin, « Collaboration homme-machine dans la traduction des métadonnées en SHS : expérience de traduction automatique post-éditée pour quatre revues françaises », *La main de Thôt* [En ligne], 9 | 2021, mis en ligne le 20 mars 2023, consulté le 21 mai 2023. URL : http://interfas.univ-tlse2.fr/lamaindethot/986

Collaboration homme-machine dans la traduction des métadonnées en SHS : expérience de traduction automatique post-éditée pour quatre revues françaises

Katell Hernandez Morin et Franck Barbin

PLAN

TEXTE

Introduction

1 La question de la collaboration homme-machine dans le domaine de la traduction a suscité depuis plusieurs décennies de nombreuses recherches, *a fortiori* avec l'amélioration spectaculaire des moteurs de traduction automatique (CADWELL et al., 2016 ; KENNY, 2017 ; ROSSI et al., 2019). Cet article présente un exemple de collaboration homme-machine dans la traduction des métadonnées en sciences humaines et sociales (SHS). Il s'appuie plus précisément sur un projet expérimental de traduction automatique post-éditée pour quatre revues françaises publiées aux Presses universitaires de Rennes (PUR). Il repose ainsi sur deux pans de la littérature : les travaux portant sur

la post-édition de la traduction automatique (MARTIKAINEN et al., 2016 ; CASTILHO et al., 2017 ; MACKEN et al., 2020) et ceux abordant l'évaluation de la qualité de la traduction automatique neuronale, en comparant les métriques humaines et automatisées (LOOCK, 2018 ; CHATZIKOUMI, 2020).

2 Le projet que nous présentons ici part du constat global suivant : la relative déficience linguistique en anglais des métadonnées des revues scientifiques françaises, qui nuit notamment à la bonne visibilité et notoriété internationales des travaux de recherche desdites revues. Ce constat de faiblesse avait été établi après un travail d'expertise sur les métadonnées des revues publiées par les PUR en vue d'un hébergement sur le portail Cairn.info. Cet état des lieux a donc logiquement trouvé un écho dans l'appel à projets lancé en décembre 2018 par le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (MESRI) en direction des éditeurs pour des services de traduction. Cet appel cherchait à « stimuler l'amélioration qualitative des traductions scientifiques des revues scientifiques », « soucieuses d'améliorer leur audience internationale ». Il se situe donc en droite ligne de notre projet qui vise à concevoir une méthode, transférable à d'autres revues et domaines disciplinaires, combinant la traduction automatique neuronale (TAN) et la post-édition humaine pour améliorer la qualité des métadonnées des articles (titres, résumés, mots-clés, table des matières, etc.) du français vers l'anglais dans le processus éditorial des revues.

3 Il faut préciser dès lors qu'il ne s'agit pas d'une recherche fondamentale portant sur la traduction automatique (TA), mais d'une recherche-action destinée à être directement applicable par les enseignants-chercheurs et par les revues scientifiques. L'objectif est de créer à terme une méthodologie adaptée et conviviale pour traduire les métadonnées d'articles de recherche.

4 Nous décrivons tout d'abord le projet qui sous-tend cet article, en mettant notamment en évidence la finalité dudit projet, le rôle joué par les différents partenaires, les étapes du projet, la constitution du corpus et le choix de l'outil de TA. Nous examinons ensuite la méthodologie d'évaluation de la qualité des traductions employée, en présentant le choix de la grille d'évaluation et les résultats préliminaires de l'évaluation comparative (entre traduction humaine existante et

TA). Nous abordons les perspectives d'optimisation du processus de post-édition, qui passent par la création de glossaires thématiques ou de listes de mots-clés, par l'alimentation de mémoires de traduction et par la production de recommandations aux auteurs des revues scientifiques. Cela nous amène à recenser les apports et les limites envisagés de l'étude, aussi bien pour les revues que pour la traductologie.

1. Description du projet

5 Ce projet, financé par le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (MESRI), a pu voir le jour grâce à la collaboration de trois partenaires situés à Rennes : LIDILE (unité de recherche en Linguistique, Ingénierie et Didactique des Langues), et plus précisément l'axe de recherche TRASILT (Traduction spécialisée, ingénierie de la langue et terminologie), la MSHB (Maison des Sciences de l'Homme en Bretagne) et les PUR (Presses universitaires de Rennes). Il ambitionne de mettre au point une méthode économe et efficace de traduction qui pourra potentiellement bénéficier à l'ensemble des chercheurs et des professionnels au service des revues scientifiques. Il s'agit de concevoir une méthode qui combine la traduction automatique neuronale (TAN) et la post-édition humaine pour améliorer la qualité des métadonnées des articles du français vers l'anglais au sein du processus éditorial des revues scientifiques. L'objectif est de développer une méthodologie de traduction qui puisse être reproduite et transférée à d'autres revues et domaines disciplinaires.

6 Il est important à ce stade de rappeler ce que l'on entend par métadonnées d'articles de recherche et par post-édition. Le terme de métadonnées, ici dans une acception restrictive, comprend les seuls éléments destinés à être traduits, à savoir la titraille (titre, sous-titre, sommaire, etc.), les résumés, les mots-clés, le texte descriptif de l'auteur (bionote), les remerciements et les notes de la rédaction. Le terme de post-édition désigne l'activité au cours de laquelle des traducteurs professionnels relisent un texte produit par la traduction automatique (TA) et le corrigent pour supprimer les erreurs sémantiques et linguistiques en vue de le rendre intelligible, exact et grammaticalement correct (ALLEN, 2003 ; ROBERT, 2010).

1.1 Partenaires

7 Chaque partenaire de ce projet apporte son active contribution, autant en matière d'expertise qu'au niveau institutionnel. La MSHB apporte sa connaissance des normes et des standards de l'information scientifique et technique. Elle coordonne le suivi du projet et sert de relais avec les réseaux de compétences (réseau Médici, réseau REPERES, etc.) et les infrastructures de recherche nationales (OpenEdition, RnMSH). De leur côté, les PUR apportent leur connaissance de l'édition et des chaînes éditoriales de publication des revues. Elles assurent la mutualisation et l'implémentation des bonnes pratiques. L'axe TRASILT de LIDILE, quant à lui, s'appuie sur son expérience du monde de la traduction professionnelle et sur les recherches qu'il a déjà réalisées dans ce champ d'étude (traduction en sciences humaines et sociales, post-édition, évaluation de la qualité, influence des technologies sur la qualité des traductions scientifiques). Il contribue ainsi à l'évaluation de la qualité de la traduction des métadonnées et à l'élaboration d'une nouvelle méthode de traduction pour les revues scientifiques. Ce projet s'inscrit dans la coopération engagée entre les PUR et la MSHB visant à améliorer les pratiques des revues.

1.2 Étapes

8 Notre projet se décompose en trois étapes principales. Il s'agit tout d'abord de comparer la traduction en anglais de métadonnées d'articles précédemment publiée (et jugée perfectible) et la traduction des mêmes données générée par la TAN. Cette première étape est réalisée par une équipe d'enseignants-chercheurs de TRASILT qui exercent également une activité professionnelle de traduction et/ou de révision. Cette première évaluation de la qualité des traductions par une grille dédiée est complétée par une phase de post-édition humaine de la TAN : la traduction automatique des métadonnées d'autres articles est post-éditée et améliorée par des traducteurs professionnels. L'objectif de ces deux premières étapes est de déterminer les éléments qualitatifs et les limites de chaque production (humaine vs. TAN) et de proposer une première ébauche de méthode. Cette méthode sera ensuite testée sur l'un des numéros des quatre revues à paraître en 2020 afin d'offrir la méthode de traduction la

plus adaptée à des enseignants-chercheurs et aux éditeurs de revues. Dans la méthode proposée, l'outil de TAN sera intégré aux méthodologies de gestion de la qualité des traductions (mémoires de traduction et glossaires).

1.3 Corpus

9 Notre projet s'appuie sur un corpus d'articles de recherche publiés dans quatre revues des PUR, Annales de Bretagne et des Pays de l'Ouest, ArcheoSciences, Éducation & Didactique et Norois, qui traitent respectivement d'histoire, d'archéologie, de didactique et de géographie. Il faut noter que les revues sélectionnées sont toutes en libre accès sur OpenEdition Journals et Persée, et sont diffusées via le portail Cairn.info. Le choix des articles a reposé sur les critères suivants : ils devaient avoir été publiés en 2017, appartenir à un même numéro de la revue, être rédigés en français, présenter des métadonnées en anglais et traiter de sujets assez diversifiés (pour offrir le spectre sémantique le plus large possible). Nous avons ainsi choisi seize articles publiés en 2017 pour la phase 1 d'analyse et seize autres articles publiés en 2017 pour la phase 2 de post-édition. Pour une bonne représentativité des domaines de spécialité, nous avons logiquement sélectionné quatre articles tirés du même numéro publié en 2017 de chacune des quatre revues susnommées. Dans la phase 3 de test de la méthode, nous choisirons un numéro de chaque revue qui sera publié en 2020 (certainement le dernier de l'année pour laisser le temps de mettre en place cette nouvelle procédure de traduction).

1.4 Choix de l'outil de TAN

10 Nous allons brièvement expliquer pourquoi notre choix s'est porté sur le système de traduction automatique neuronale *DeepL Pro* et non sur la création de notre propre moteur de TA entraîné sur notre corpus. Après avoir réalisé un recensement comparatif des outils de traduction automatique potentiellement utilisables, nous avons opté pour *DeepL Pro* pour les motifs suivants. Tout d'abord, la qualité de la traduction fournie par cet outil est reconnue comme supérieure par rapport à ses concurrents directs (principalement *Google Translate*) dans plusieurs études scientifiques et professionnelles (BURBAT et al., 2018 ; MACKETANZ et al., 2018 ; SMOLENTCEVA, 2018 ; LÖBERT,

2019). Même si nous rejoignons Rudy LOOCK (2019) lorsqu'il met en garde contre les performances régulièrement surévaluées de ces moteurs de TA dans la presse, nous avons constaté que la qualité de traduction produite par *DeepL Pro* est globalement supérieure à celle de ses concurrents. Sans entrer dans les détails ici, nous avons effectué un test comparatif entre *DeepL* et *Google Translate* en soumettant à ces deux outils de TA plusieurs résumés d'articles de recherche des PUR et *DeepL* a systématiquement produit des traductions de meilleure qualité et plus fluides que *Google Translate* (pour le couple français/anglais dans notre cas). Ce système bénéficie d'une capacité d'entraînement de ses propres réseaux neuronaux sur des milliards de segments de traduction de haute qualité. Cet outil présente aussi l'avantage d'offrir une version gratuite adaptée aux besoins des enseignants-chercheurs : la simplicité d'utilisation du moteur est essentielle si l'on veut que la méthodologie de traduction proposée soit réellement appliquée de manière pérenne au sein des revues françaises. La version pro offre également une protection des données communiquées : la confidentialité des articles de recherche soumis est assurée par le cryptage des données et par l'absence de stockage des documents sur le serveur *DeepL*. Elle permet en outre la génération de documents entièrement éditables et mis en forme à l'identique, facilitant ainsi le travail de post-édition. Ce moteur de TA présente également l'avantage de pouvoir être intégré à des logiciels de traduction assistée par ordinateur (TAO) générant des mémoires de traduction. Cela permet ainsi de réutiliser et d'harmoniser la terminologie et la phraséologie au sein des revues.

11 Nous sommes toutefois conscients que cet outil présente certaines limites. L'entraînement qu'offre *DeepL* se limite à l'alimentation de l'outil par les traductions effectuées et par les éventuelles modifications réalisées en ligne : toutes les corrections apportées aux traductions sur l'interface *DeepL* sont enregistrées et servent à entraîner les algorithmes. Ce moteur de TA ne permet en aucun cas de mettre en place un entraînement spécifique adapté aux quatre domaines abordés dans les revues sélectionnées. Il ne permet pas non plus d'adosser un glossaire ou une base terminologique propre à un champ disciplinaire (sa toute nouvelle fonctionnalité d'ajout d'un glossaire permet tout au plus d'en créer un au fil de l'eau pour l'appliquer au document

en cours de traduction). Il n'offre enfin aucun accès au code source en vue d'une personnalisation de l'outil.

12 Cependant, la durée du projet (18 mois) ne permet pas de développer un outil libre, nécessitant le recrutement d'un personnel dédié et l'entraînement d'un corpus spécifique très volumineux composé de données bilingues alignées. La méthode de gestion des traductions qui sera mise en place permettra une utilisation économe et autonome du logiciel par les scientifiques et les éditeurs de revue.

2. Méthodologie d'évaluation de la qualité des traductions

13 Une fois l'outil de TAN décrit, avec ses atouts et ses limitations, revenons sur les objectifs de l'étude et sur les étapes fixées pour les atteindre. L'objet de cette étude est de développer une méthodologie de traduction encadrée reposant sur la post-édition de traduction automatique. Une qualité linguistique et scientifique optimale des métadonnées d'articles en anglais est recherchée.

2.1 La grille d'évaluation de la qualité TRASILT

14 La première étape dans cette optique est l'évaluation des traductions existantes et leur comparaison la plus objective possible avec une traduction automatique utilisant *DeepL*.

15 À cet effet et profitant de l'expertise du groupe de recherche TRASILT, le choix d'une grille d'évaluation quantitative de la qualité a été effectué : la grille TRASILT, développée et perfectionnée depuis plusieurs années au sein du groupe. Cette grille d'évaluation multicritères a vu le jour en 2011 alors que le groupe cherchait à évaluer de façon fine et dynamique les productions d'apprentis-traducteurs utilisant plusieurs méthodes (traitement de texte, TAO, reconnaissance vocale, traduction automatique avec post-édition) lors d'une expérimentation de traduction comparée (HERNANDEZ MORIN et al., 2017). L'étude de la littérature existante avait alors révélé des lacunes en matière d'outils d'évaluation objective reposant sur des critères professionnels, pouvant s'adapter indifféremment à ces différentes mé-

thodologies de traduction (O'BRIEN, 2012 ; WISNIEWSKI et al., 2013). Une grille reposant sur notre expérience de traducteurs professionnels et d'enseignants en traduction a donc été développée pour essayer de regrouper les critères d'évaluation les plus complets possibles en traduction spécialisée (non générale et non littéraire). Elle visait à limiter au maximum les biais subjectifs par l'établissement de catégories d'erreurs claires et définies, et par la différenciation des erreurs de leurs « effets » sur la traduction (impact de chaque erreur sur la qualité qui en est attendue). La recherche effectuée sur la grille a mené à l'établissement de facteurs de « gravité » de ces effets, sanctionnée par des points différenciant le degré de gravité de chaque effet. Des règles d'application de ces points de pénalité ont été affinées au cours des expériences d'évaluation des traductions comparées. En outre, des pondérations plus ou moins fortes pouvaient être appliquées à certains types d'erreurs ou d'effets de l'erreur, en fonction des attendus qui pouvaient être identifiés selon le type de document traduit ou la spécialité dont relevait la traduction.

16 La grille TRASILT (TOUDIC et al., 2014) introduite ici (voir Tableau 1) distingue donc aujourd'hui neuf types d'erreurs possibles en traduction spécialisée : sept erreurs reposant sur des catégories plutôt traditionnelles (Sens, Grammaire / syntaxe, Orthographe / typographie, Terminologie, Phraséologie, Style, Omission / ajout), auxquelles s'ajoutent deux catégories issues de l'évaluation professionnelle : les erreurs de Localisation (définies comme l'absence d'adaptation à un public cible ou à une culture donnée) et les erreurs de PAO (défauts de mise en page et de formatage).

17 À chacune de ces erreurs identifiées, peuvent s'appliquer potentiellement quatre effets sur la qualité de la traduction évaluée : la Précision de l'information transmise, la Fonctionnalité du document traduit, la Lisibilité des contenus et la Conformité de la traduction aux différentes normes et conventions linguistiques ou professionnelles applicables.

Tableau 1 : grille TRASILT : Typologie des erreurs et des effets sur la qualité

18 Types d'erreur :

Sens	**Omission / ajout**	**Terminologie**	**Phraséologie**	**Grammaire / syntaxe**
Ambiguïté	Non-traduction d'un élément de sens du document source	Variante inappropriée (variété de langue / usage professionnel / usage « interne »)	Variante inappropriée (variété de langue / usage professionnel / usage « interne »)	Erreur morpho-syntaxique
Erreur de sens partielle	Ajout injustifié d'informations ayant un impact mineur sur le texte cible	Terme inapproprié (appartenant à un autre domaine)	Phraséologie inappropriée (appartenant à un autre domaine)	Ordre des mots
Erreur de sens complète	Ajout injustifié d'informations ayant un impact majeur sur le texte cible	Incohérence terminologique (dans le document / par rapport aux documents de référence)	Incohérence phraséologique (dans le document / par rapport aux documents de référence)	Structure de la phrase
Non-correction d'un défaut du texte source				

Orthographe	**Style**	**Localisation**	**PAO**
Mauvaise orthographe	Traduction littérale	Non-adaptation à la culture cible	Mise en page
Typographie	Longueur de la phrase	Non-adaptation au public cible	Mise en forme
Erreur de ponctuation	Manque de fluidité	Défaut de localisation des éléments et données chiffrées	Graphiques
	Registre inapproprié (langage formel/informel)		Balises
	Variété inappropriée (orthographe ou choix de mot spécifique à un pays)		Références croisées

Effets sur la qualité :

Précision	Fonctionnalité	Lisibilité	Conformité
L'erreur empêche le transfert correct des informations du document source.	L'erreur empêche l'utilisation appropriée du produit, processus ou document.	L'erreur nuit à la fluidité et à la clarté du document cible.	Le document cible n'est pas conforme aux normes, conventions ou recommandations de la langue, du pays, de la culture ou du client.

19 Les niveaux de gravité des effets de l'erreur vont de 0 (pas d'effet / effet non comptabilisé) à 3 (effet critique) en passant par 1 (effet mineur) et 2 (effet majeur). Pour éviter la dispersion des types d'effets, le nombre d'effets pour une erreur donnée est limité à 2, pour un total maximum de 5 points de pénalité. Des coefficients de pondération (minorations ou majorations) peuvent être appliqués sur certains types d'erreurs ou d'effets à minorer ou à proscrire selon la visée du document traduit ou sa spécialité. Les recherches effectuées sur la grille ont également amené le groupe à introduire l'attribution de points de bonus lorsque des traductions présentent un caractère particulièrement novateur ou enrichissent fortement le document traduit. Cet ensemble de mesures quantitatives permet de parvenir à un score global traduisant le niveau et la typologie de qualité de la traduction (grâce à un récapitulatif des totaux par type d'erreurs et par effet sur la qualité). Le score obtenu est complété par une cellule de commentaire rédigé sur la qualité globale de la traduction.

2.2 Résultats préliminaires de l'évaluation comparative de la qualité des traductions

20 La finesse de la grille d'évaluation choisie pour le projet d'optimisation des traductions des métadonnées d'articles de revues en sciences humaines et sociales nous permet donc d'adapter notre évaluation aux exigences de publication en anglais de métadonnées d'articles disponibles sur des plateformes telles que Cairn.info (2005) ou OpenEdition (1999). Ces plateformes requièrent des métadonnées en anglais – et potentiellement à l'avenir dans d'autres langues – facilement identifiables, afin d'améliorer la visibilité des articles des revues francophones auprès d'un public non francophone. Pour ce faire, les métadonnées produites doivent être exactes (correspondre aux désignations de la discipline) et harmonisées sur le plan terminologique

(termes des résumés et mots-clés cohérents au sein de la discipline) et la qualité linguistique des métadonnées doit être suffisante pour qu'elles soient publiées et consultées largement par des scientifiques non francophones. C'est donc l'objectif qui a été fixé également pour notre recherche, et qui a guidé la première étape de notre travail : l'évaluation des traductions des métadonnées existantes (sur l'année 2017) et leur comparaison avec l'évaluation de la qualité produite par une traduction automatique utilisant *DeepL* sur les mêmes métadonnées.

21 À ce stade du travail, nous disposons de l'analyse préliminaire de 10 des 16 articles de 2017 prévus pour la comparaison de la phase 1. Cette analyse sera complétée par le travail de post-édition de 16 autres articles de 2017 réalisé par des traducteurs professionnels spécialistes des disciplines concernées. Elle sera suivie d'une évaluation scientifique par les chercheurs du projet à l'aide de la grille TRA-SILT.

22 Le tableau 2 présenté ci-après fournit les scores de qualité globaux obtenus dans notre première évaluation (de 10 articles sur 16) sur la traduction des métadonnées de chaque article (quatre par revue), pour chaque revue (quatre revues) et chaque méthode de traduction (traduction humaine par les auteurs des articles – *a priori*[1], dite « traduction publiée » – et « traduction automatique » par *DeepL*). Ces scores reflètent le total des points de pénalité appliqués aux effets de chaque erreur sur la qualité des traductions.

Tableau 2 : Score global par méthode et par revue

Revue	Score Traduction publiée	Score Traduction automatique
ABPO		
Article 1	23	23
Article 2	42	31
ARCHEOSCIENCES		
Article 1	16	7
Article 2	76	21
Article 3	38	35
ÉDUCATION & DIDACTIQUE		
Article 1	21	11

Article 2	26	20
NOROIS		
Article 1	46	20
Article 2	18	16
Article 3	50	18
Score moyen pour les 4 revues	**34,33**	**21,75**

23 La première observation au regard de ces chiffres, est la grande variabilité des valeurs obtenues selon l'article, la revue ou la méthode de traduction concernée : si l'article 1 de la revue *ABPO* présente un score qualitatif identique de 23 (points de pénalité) quelle que soit la méthode utilisée, l'article 2 de la revue *ArcheoSciences* affiche des écarts de qualité beaucoup plus importants (76 pour la traduction publiée contre 21 points seulement pour la traduction automatique). Un schéma analogue apparaît pour l'article 3 de la revue *Norois* (50 points de pénalité pour la traduction publiée contre 18 points seulement pour la traduction automatique). Il est à noter également qu'aucune revue ne se détache par des scores particulièrement homogènes d'un article à l'autre ou d'une méthode de traduction à l'autre. Une observation plus générale permet de se rendre compte qu'à une exception près (l'article 1 d'*ABPO*), les scores qualitatifs offerts par la traduction automatique sont meilleurs (plus faibles) que ceux de la traduction publiée. Les différences sont parfois faibles (comme pour l'article 3 d'*ArcheoSciences* ou l'article 2 de *Norois*), mais le plus souvent, elles sont relativement importantes à importantes. Lorsque l'on calcule le score qualitatif moyen obtenu sur l'ensemble des revues selon la méthode de traduction, l'on obtient le chiffre de 34,33 pour la traduction publiée, contre 21,75 pour la traduction automatique. Ce ne sont que des tendances préliminaires, mais qui laissent entrevoir deux principales directions dans les interprétations possibles de l'évaluation : les niveaux de qualité des traductions en anglais par les auteurs des articles sont très hétérogènes ; et la qualité obtenue lors de l'examen de la traduction automatique par *DeepL* est généralement meilleure (ce qui ne signifie pas qu'elle soit jugée « bonne » pour une publication).

24 Si nous entrons dans le détail des erreurs les plus couramment relevées selon les revues et les méthodes de traduction appliquées, d'autres enseignements intéressants se font jour. Le tableau 3 détaille

ces types d'erreurs et le comptage, dans chaque article (entre parenthèses), des occurrences de ces erreurs par type d'erreur.

Tableau 3 : Détail des erreurs les plus couramment relevées par méthode et par revue

Revue	**Traduction publiée**	**Traduction automatique**
ABPO		
Article 1	grammaire/syntaxe (4) orthographe/typographie (4) omissions/ajouts (3) style (3)	orthographe/typographie (5) sens (3) style (3)
Article 2	grammaire/syntaxe (6) terminologie (6) omissions/ajouts (5)	sens (5) terminologie (4) grammaire/syntaxe (3)
ARCHEOSCIENCES		
Article 1	grammaire/syntaxe (4) terminologie (3)	terminologie (2)
Article 2	omissions/ajouts (15) grammaire/syntaxe (9)	terminologie (4) sens (2)
Article 3	grammaire/syntaxe (8) localisation (4)	terminologie (11) sens (5)
ÉDUCATION & DIDACTIQUE		
Article 1	grammaire/syntaxe (5) sens (3)	terminologie (3)
Article 2	grammaire/syntaxe (4) style (4)	orthographe/typographie (7) terminologie (3)
NOROIS		
Article 1	omissions/ajouts (8) sens (6) phraséologie (6)	sens (3) terminologie (3)
Article 2	omissions/ajouts (4) terminologie (4)	orthographe/typographie (3)
Article 3	sens (7) orthographe/typographie (7)	orthographe/typographie (3) localisation (3)
Types d'erreurs les plus fréquents dans les 4 revues	**grammaire/syntaxe (40)** **omissions/ajouts (35)** **sens (16)**	**terminologie (30)** **orthographe/typographie (18)** **sens (18)**

25 Le premier constat évident, lorsque l'on observe la récurrence des erreurs relevées sur les traductions publiées (effectuées par les auteurs), est la part dominante des défauts de grammaire ou de syntaxe (apparaissant en gras pour tous les articles où ce type d'erreurs est le plus présent). Il s'agit du type d'erreurs le plus fréquent pour cette méthode de traduction (avec 40 occurrences). Les écarts de construction syntaxique entre le français et l'anglais et la connaissance approfondie de l'anglais que requiert une rédaction scientifique de qualité, peuvent expliquer ces défauts récurrents dans les traductions vers l'anglais de chercheurs non spécialistes de langue. Voici un exemple du type d'erreur de syntaxe fréquemment rencontré : dans un article de la revue *ArcheoSciences*, « *The two* ***potters workshops*** *[...] are therefore contemporary* » traduisait la phrase « Les deux **ateliers de potiers** [...] seraient donc sensiblement contemporains » (on s'attendrait plutôt à trouver une formule telle que « *The two* ***potters' workshops*** » ou « *The two* ***pottery workshops*** »).

26 Le deuxième constat est la forte présence, dans ces traductions également, des omissions et des ajouts (c'est-à-dire la non-traduction ou l'ajout injustifié d'un élément sémantique par rapport aux éléments figurant dans le texte source). Ces omissions et ajouts injustifiés (les seuls pénalisés, car ceux-ci peuvent aussi se justifier sur certains passages) représentent 35 occurrences dans l'ensemble des revues. Une hypothèse d'explication peut être la liberté que prennent les auteurs dans la traduction de leur propre résumé d'article (certaines informations ou précisions du résumé en français ne figurant plus dans le résumé produit en anglais). Cette adaptation peut relever d'un choix personnel, comme d'une difficulté à exprimer certaines réalités énoncées dans la langue maternelle avec autant de précision ou de nuances, dans une langue étrangère. L'exemple ci-après en est l'illustration : dans le passage suivant (issu d'un article de la revue d'histoire ABPO), « ***The relief from*** *the East lodging of Suscinio was long considered a simple decorative element of the façade* » qui traduisait l'extrait « Longtemps considéré comme un simple élément ornemental de la façade du logis Est de Suscinio, **le relief aux cerfs** surmontant l'entrée [...] », peut être considéré comme un choix de (non) traduction libre du complément « aux cerfs », ou comme une stratégie d'évitement consécutive à la difficulté de l'auteur à traduire ce complément. Quoi qu'il en soit, dans cet article, l'apparition du cerf sur les

reliefs du château était centrale et cet élément de sens devait donc être traduit dans le passage.

27 L'analyse des résultats obtenus avec la traduction automatique, quant à elle, révèle la présence majoritaire d'un autre type d'erreur : les erreurs de terminologie (30 erreurs comptabilisées). Cette incapacité des outils de traduction automatique généralistes à effectuer des choix contextuels de termes selon les disciplines, et à harmoniser ces choix au sein d'un même document, est connue (LÄUBLI, 2018 ; LOOCK, 2018). Un exemple assez révélateur (tiré d'un article de la revue de géographie *Norois*) peut être donné ici : le terme de « **paysage fluvial** » est traduit différemment par le logiciel selon qu'il apparaît dans le résumé de l'article (« ***river landscape*** ») ou dans la liste des mots-clés (« ***fluvial landscape*** »). L'on notera que cette défaillance dans la gestion terminologique est nettement moins présente dans les traductions effectuées par les auteurs des articles (13 erreurs sur l'ensemble des revues). Le deuxième type d'erreurs le plus souvent observé sur la traduction automatique des métadonnées est le type « orthographe/typographie » (18 erreurs). Le type « sens » est également relevé selon la même récurrence (18 erreurs dans toutes les revues). Les erreurs d'orthographe et de typographie peuvent s'expliquer en partie (après observation du détail des erreurs) par la gestion défaillante par *DeepL* des différences de ponctuation (espaces insécables, gestion des espacements) et de typographie (majuscules/minuscules) entre le français et l'anglais. Cette limitation du logiciel *DeepL* est également connue. Les erreurs de sens récurrentes s'expliquent par la gestion aléatoire des noms propres (les toponymes notamment) par l'outil, ainsi que par la traduction aléatoire, là aussi, des concepts (renvoyant à d'autres réalités que celles décrites en français). Deux exemples viennent illustrer ces erreurs : dans la revue *ArcheoSciences*, un article traite d'un site de poterie-tuilerie médiévale situé dans le **département de la Manche**. Dans le résumé traduit en anglais par *DeepL*, nous ne parlons plus du département de la Manche, mais de « *the Saint-Georges-de-Rouelley site* (***Channel***) » (!). Toujours dans le même article, le terme de « datation haute » apparaît dans le contexte suivant : « le réexamen des données archéomagnétiques recueillies au milieu des années 1980 conduit à une **datation plus haute d'un demi-siècle** (1260-1280 au lieu de 1325-1350) ». Le logiciel *DeepL* livre cette traduction du passage : « (…) *leads to a*

higher dating of half a century ». Cet emploi n'a pas de sens en anglais et devrait être remplacé par l'expression « *an* ***earlier dating*** (…) ».

28 Si nous nous penchons enfin sur les sanctions quantitatives (pénalités) appliquées par les évaluateurs aux erreurs relevées dans les métadonnées d'articles traduites, rappelons que la grille d'évaluation utilisée pénalise les effets des erreurs sur la qualité du document et non les erreurs elles-mêmes. Des pénalités de 1 à 5 points peuvent s'appliquer en fonction des effets négatifs des erreurs sur la Précision, la Fonctionnalité, la Lisibilité ou la Conformité des données traduites. Le tableau 4 présenté ci-après détaille les effets les plus fortement sanctionnés dans les traductions pour chaque article et selon les deux méthodes de traduction. Les pénalités appliquées (entre parenthèses) à chaque effet sont à rapprocher du score global (total des pénalités) obtenu avec chaque méthode sur chaque article (voir tableau 2).

Tableau 4 : Détail des effets les plus pénalisés par méthode et par revue

Revue	Traduction publiée	Traduction automatique
ABPO		
Article 1	conformité (13) précision (5)	fonctionnalité (10) conformité (9)
Article 2	fonctionnalité (17) conformité (11) précision (10)	fonctionnalité (13) lisibilité (7) précision (6)
ARCHEOSCIENCES		
Article 1	conformité (7) précision (5)	fonctionnalité (3) conformité (2) lisibilité (2)
Article 2	fonctionnalité (24) précision (22) conformité (21)	fonctionnalité (8) lisibilité (8)
Article 3	fonctionnalité (14) conformité (14)	fonctionnalité (10) lisibilité (10) précision (9)
ÉDUCATION & DIDACTIQUE		
Article 1	conformité (9) fonctionnalité (7)	conformité (6) fonctionnalité (3) précision (2)
Article 2	conformité (15) lisibilité (8)	conformité (9) fonctionnalité (7)
NOROIS		

Article 1	conformité (16) précision (14) fonctionnalité (13)	conformité (7) lisibilité (7)
Article 2	fonctionnalité (6) conformité (5) précision (5)	fonctionnalité (6) conformité (6)
Article 3	fonctionnalité (20) conformité (11) précision (11)	fonctionnalité (10) conformité (5)
Types d'effets les plus pénalisés dans les 4 revues	**fonctionnalité (94)** **conformité (122)**	**fonctionnalité (70)** **conformité (44)**

29 De ce tableau de résultats, se détachent deux tendances visibles : l'atteinte à la fonctionnalité du document traduit, tout d'abord, représente un poids très important dans les sanctions qualitatives appliquées, quelle que soit la méthode employée (traduction humaine par des non-spécialistes de langue – 94 points de pénalité – ou traduction automatique – 70 points).

30 Ce résultat n'est pas une surprise, étant donné que l'atteinte à la fonctionnalité du document « empêche l'utilisation appropriée du produit, processus ou document » (selon la définition de la grille rappelée dans le tableau 1). Dans le contexte d'un usage scientifique des métadonnées des documents, l'objectif essentiel de transmission de la connaissance est ainsi « empêché » par les erreurs, et il est logique que l'évaluateur les sanctionne fortement.

31 La deuxième tendance est la suivante : les deux effets les plus pénalisés sont, pour les deux méthodes de traduction, la fonctionnalité et la conformité. Des différences importantes apparaissent, cependant : les pénalités liées à la conformité sont nettement plus présentes dans les traductions publiées (leur niveau est supérieur à celui des pénalités comptabilisées pour la fonctionnalité : 122 points contre 94). Ces points de pénalité peuvent être rapprochés des nombreuses erreurs de grammaire et de syntaxe relevées dans les traductions publiées (par les auteurs) – qui affectent la conformité des articles aux normes linguistiques et scientifiques de publication en anglais. Il n'est pas surprenant, là non plus, que l'évaluateur y ait appliqué un poids important. Dans les traductions automatiques des métadonnées, ce poids de l'atteinte à la conformité linguistique et scientifique est moindre (44 points de pénalité sur l'ensemble des revues, contre 70 points pour l'effet « fonctionnalité »). Rappelons à titre d'hypo-

thèse d'explication que la traduction automatique neuronale a démontré, dans un certain nombre d'études, sa capacité à produire des énoncés globalement fluides sur le plan syntaxique (SUTSKEVER et al., 2014 ; HASSAN et al., 2018).

32 Si ces premiers résultats nous confortent dans l'intérêt d'utiliser la traduction automatique neuronale en soutien aux auteurs, éditeurs ou responsables de plateformes de publication (en utilisant *DeepL* ou un autre outil commercial ou en accès libre), ils soulignent aussi la nécessité de compléter cet apport par tout un travail d'harmonisation linguistique, terminologique et d'encadrement de la post-édition humaine des traductions automatiques.

3. Perspectives d'optimisation du processus de post-édition

3.1 Recommandations et guides de style

33 Les gains de qualité globale passeront en effet par l'élaboration de consignes précises et harmonisées établies en fonction de chaque revue. Celles-ci pourront prendre la forme d'un guide de style élaboré à partir des défauts et irrégularités constatés lors de l'étude des traductions humaines par les auteurs (rappels de règles syntaxiques de l'anglais, par exemple) et des problèmes récurrents posés par la traduction automatique (incohérence terminologique, style littéral de la traduction, etc.). Elles pourront également s'appuyer sur les recommandations existantes effectuées par les éditeurs de revues aux auteurs en matière de rédaction des articles, de choix des mots-clés (par exemple, éviter l'emploi de termes polysémiques ou de synonymes pour désigner un même concept, dans la mesure du possible). Les recommandations des revues concernées par cette recherche ont déjà été rassemblées et une coordination avec les responsables d'édition des PUR est prévue tout au long du projet.

3.2 Supports terminologiques et traductionnels

34 Les recommandations et guides de style devront s'accompagner de tout un travail d'harmonisation terminologique, voire phraséologique, qui passe par l'élaboration de glossaires et peut bénéficier de l'apport de mémoires de traduction (et de bases de données terminologiques) alimentées par les traductions effectuées. La linguistique de corpus et les outils d'analyse contribuent également à pouvoir établir des glossaires bilingues représentatifs, en choisissant des sources terminologiques fiables (articles de revues francophones et anglophones faisant autorité et reconnues pour leur qualité rédactionnelle) et en améliorant les sources disponibles, par un travail scientifique en lien avec les spécialistes disciplinaires. Ce travail a commencé avec la mise en place de glossaires bilingues améliorables extraits à partir des articles de 2017 des quatre revues.

3.3 Apports et limites de l'étude envisagés

35 L'objectif qui sous-tend ce projet est l'amélioration de la visibilité internationale des revues françaises en sciences humaines et sociales. Le défi principal est celui de la coopération active entre les experts en traduction, les chercheurs en sciences humaines et sociales et les responsables éditoriaux, afin de tendre vers une harmonisation des pratiques et des choix linguistiques et terminologiques2. Cette coopération est mue, d'une part, par le constat de pratiques hétérogènes et génératrices de limitations qualitatives dans les traductions (ce qui empêche leur large diffusion) ; et par l'intérêt pour le traitement traductologique de l'objet scientifique, d'autre part. Les métadonnées en sciences humaines et sociales n'ont en effet été que peu traitées et analysées à l'aide de mémoires de traduction ciblées ; et l'efficacité de la traduction automatique dans ces domaines reste largement à éprouver, tant la diversité des discours et des productions est grande et la technologie, évolutive.

36 Cette recherche n'offre qu'une perspective, un angle de vue initial sur un moteur de traduction automatique (*DeepL*) et son exploitation en-

richie ciblant spécifiquement les métadonnées d'articles de revues, avec une approche terminologique et phraséologique des disciplines visées. L'outil de traduction choisi pour l'étude en raison de son accessibilité immédiate et du niveau de qualité globale qu'il offre, présente des limites et n'est peut-être pas, à terme, l'outil idéal pour traiter ce genre de contenus : un outil en accès libre ou *ad hoc*, qui puisse être entraîné et s'adapter au contexte spécifique de chaque article, serait sans doute souhaitable dans une recherche au long cours, même si la technologie évolue vite et si *DeepL* lui-même s'est déjà doté d'une fonctionnalité de gestion d'un glossaire en trois langues (anglais, français et allemand).

37 De même, la première langue cible choisie pour cette étude, l'anglais, pour des raisons évidentes de recherche de diffusion large des articles (ou à tout le moins, de plus grande visibilité des métadonnées donnant accès à ces articles)3, ne saurait être la seule langue à envisager dans ce type de projets. L'enjeu de la traduction des métadonnées scientifiques dépasse largement cette *lingua franca* et devrait permettre de les rendre accessibles dans une plus grande diversité de langues, pour refléter les intérêts et les collaborations potentielles d'auteurs de nationalités diverses. Le projet prévoit d'élargir le travail d'adaptation à des langues telles que l'espagnol ou le breton, en fonction de la pertinence de ces langues pour chaque revue.

38 Enfin, si le défi de la reproductibilité des méthodes, d'une discipline scientifique et d'un mode de discours à l'autre, est grand, ce type de recherche offre l'occasion de transférer de bonnes pratiques de rédaction et de traduction vers la communauté scientifique en sciences humaines et sociales, en permettant aux revues d'être accompagnées dans leur démarche d'harmonisation des usages linguistiques.

NOTE DE FIN

1 Nous n'avons pas accès, par article ou par revue, à l'identité de l'auteur de la traduction des métadonnées, mais l'hypothèse la plus vraisemblable est que l'auteur traduit la plupart du temps lui-même ou elle-même les métadonnées de son article (avec ou sans l'aide d'un tiers et/ou de la traduction automatique). Nous savons, en revanche, que les traductions des résumés

ou des mots-clés peuvent être retouchées lors du travail éditorial des revues (par des équipes majoritairement francophones).

2 L'enjeu pour les revues n'est pas de parvenir à une simple harmonisation des métadonnées en anglais, mais d'appliquer cette harmonisation aux métadonnées rédigées dans la langue de départ.

3 Ce choix linguistique nous est imposé par la sélection de l'anglais pour la traduction des métadonnées par les revues elles-mêmes.

AUTEURS

Katell Hernandez Morin

Franck Barbin