



**UNIVERSITÉ
DE GENÈVE**

Archive ouverte UNIGE

<https://archive-ouverte.unige.ch>

Master

2018

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

L'intégration de la traduction automatique avec post-édition aux systèmes
de mémoire de traduction : la traduction de textes techniques dans SDL
Trados Studio 2017

Imbres, Margherita

How to cite

IMBRES, Margherita. L'intégration de la traduction automatique avec post-édition aux systèmes de mémoire de traduction : la traduction de textes techniques dans SDL Trados Studio 2017. Master, 2018.

This publication URL: <https://archive-ouverte.unige.ch/unige:112067>

© This document is protected by copyright. Please refer to copyright holder(s) for terms of use.



**UNIVERSITÉ
DE GENÈVE**

**FACULTÉ DE TRADUCTION
ET D'INTERPRÉTATION**

MARGHERITA IMBRES

L'INTÉGRATION DE LA TRADUCTION AUTOMATIQUE AVEC POST-ÉDITION AUX SYSTÈMES DE MÉMOIRE DE TRADUCTION

La traduction de textes techniques dans SDL Trados Studio 2017

Directrice de mémoire : Marianne Starlander

Jurée : Pierrette Bouillon

Mémoire présenté à la Faculté de traduction et d'interprétation (Département TIM, Unité d'italien) pour
l'obtention de la Maîtrise universitaire en traduction, mention technologies de la traduction

Université de Genève

Août 2018

Table de matières

1. Introduction.....	3
2. La traduction automatique.....	7
2.1 Introduction et bref historique	7
2.2 Les systèmes statistiques	11
2.3 Les systèmes neuronaux	13
2.4 Vers l'intégration de la TA aux outils d'aide à la traduction	17
2.5 Vers l'intégration de la TA dans le flux de traduction des entreprises	18
3. L'évaluation de la traduction automatique.....	21
3.1 L'évaluation humaine	22
3.2 L'évaluation automatique	24
3.2.1 Bilingual Evaluation Understudy (BLEU).....	26
3.2.2 Translation Edit Rate (TER).....	26
3.2.3 Human-targeted Translation Edit Rate (HTER).....	27
4. La post-édition.....	30
4.1 Définition, utilité et champs d'application	31
4.2 Évolution de la post-édition	37
4.3 Degrés de post-édition	40
4.4 Avantages et inconvénients	42
4.5 Approches de la post-édition : le profil du post-éditeur	43
5. L'expérience	46
5.1. L'intérêt et la finalité	46
5.2 L'hypothèse et les questions de recherche.....	48
5.3 Le milieu expérimental	50
5.4 Les données.....	51
5.5 Profil des participants.....	52
5.6 Les logiciels.....	53
5.7 La méthodologie.....	56
6. Les résultats.....	60
6.1 Présentation et analyse des résultats	60
6.1.1 Première partie : traduction et post-édition.....	60

6.1.2 Deuxième partie : évaluation automatique	77
6.1.3 Troisième partie : récapitulatif des réponses des participants au questionnaire.....	81
7. Conclusion	86
7.1 Conclusions de l'étude	86
7.2 Limites, améliorations et perspectives futures	88
Bibliographie.....	90
Sitographie	93
Index des figures	94
8. Annexes	95
I Protocole de travail	95
II Instructions.....	97
III Questionnaire	99
Remerciements	104

1. Introduction

Ce mémoire naît de l'intérêt toujours croissant pour la traduction automatique (TA) dans le domaine des technologies de la traduction. En raison de la place toujours plus importante qui lui est accordée dans le travail d'un traducteur, nous assistons depuis longtemps à une époque de transition et d'évolution qui voit l'apparition et le consolidation du « post-éditeur » comme une figure professionnelle à part, distincte de celle du traducteur.

Malgré la résistance opposée à la post-édition surtout de la part les traducteurs qui ont le plus d'expérience, cette activité est indéniablement en train de gagner une place toujours prépondérante dans le marché de la traduction. D'ailleurs, en ce qui concerne l'aspect de l'évolution des logiciels, il est intéressant de noter une progressive intégration de la traduction automatique aux outils d'aide à la traduction, et plus précisément aux systèmes de mémoire de traduction (Zaretskaya et al., 2015). Nous analyserons cet aspect à la section 2.4.

Le présent mémoire vise à évaluer l'intérêt de l'intégration du flux de travail « traduction automatique avec post-édition » lors de la traduction de textes techniques avec un système de mémoire de traduction par un traducteur indépendant. Nous avons conduit notre étude en testant la traduction automatique fournie par *SDL Language Cloud*¹, soit sous la forme du *Baseline Engine (module de TA de base fourni par SDL)*, soit sous forme de *l'Adaptive Engine*² (*module de TA qui s'adapte aux corrections de l'utilisateur*), dans le cadre d'un projet de traduction effectué sur le système de mémoire de traduction *SDL Trados Studio 2017*³ (Freelance version). Nous présenterons dans les détails les logiciels utilisés dans l'expérience dans la section 5.6.

Notre hypothèse est que l'intégration d'un module de traduction automatique avec post-édition dans le flux de travail d'un traducteur indépendant qui travaille avec un système de mémoire de traduction est rentable pour la traduction de textes techniques. Afin de tester notre hypothèse, nous avons conduit une étude qui a demandé la participation de 8 traducteurs (y compris nous-même). Notre expérience comporte trois parties principales : premièrement, les phases de traduction et de post-édition (section 6.1.1) ; deuxièmement, l'évaluation de la traduction à l'aide de mesures automatiques TER, HTER et BLEU

¹ <https://www.sdl.com/it/languagecloud/> (dernière consultation : 06/03/2018)

² <https://www.sdltrados.com/products/trados-studio/adaptivemt/> (dernière consultation : 06/03/2018)

³ <https://www.sdltrados.com/products/trados-studio/> (dernière consultation : 06/03/2018)

(section 6.1.2) et troisièmement, un questionnaire rempli par tous les participants (6.1.3). Nous décrivons le déroulement des différentes phases à la section 5.7, dédiée à la méthodologie.

D'après notre hypothèse, le travail du traducteur serait réduit à une simple post-édition bilingue de la traduction cible, une tâche idéalement plus rentable, plus efficace et moins fatigante par rapport à une retraduction complète de la source. Plus spécifiquement, la fonctionnalité *Adaptive Engine* garantirait une amélioration progressive de la qualité de la traduction et un gain de productivité au traducteur, en réduisant au fur et à mesure l'effort de post-édition. Par conséquent, l'intégration de la traduction automatique à un outil d'aide à la traduction dans le flux de travail d'un traducteur indépendant serait rentable.

La première partie de l'expérience comporte plusieurs phases intermédiaires (section 6.1.1). Dans un premier temps, nous avons testé l'*Adaptive Engine*, une nouvelle fonctionnalité proposée par le système de mémoire de traduction *SDL Trados Studio 2017*. Il s'agit d'un module de traduction automatique personnel, intégrable aux projets de traduction comme toute autre ressource, capable de mémoriser les choix traductifs de l'utilisateur. D'après (SDL, 2017b, non énuméré), cette fonctionnalité serait capable de s'adapter et de reproduire le style, le ton, ainsi que la terminologie de l'utilisateur. Pour cela, nous avons demandé au traducteur (en l'occurrence, nous-même) de retraduire ou post-éditer des segments issus de textes techniques. L'objectif était, d'un côté, d'évaluer le degré d'acceptation de la proposition de la traduction automatique de la part du traducteur, et de l'autre, d'analyser le type et le taux de changements apportés dans la phase de post-édition pour tester la capacité effective d'apprentissage de la fonctionnalité. Afin de garantir une majeure objectivité à l'étude, nous avons ensuite conduit deux tests de retraduction pour évaluer l'efficacité de l'*Adaptive Engine* en termes de mémorisation des corrections du traducteur et adaptation aux choix traductifs.

Ensuite, nous avons testé spécifiquement le *Baseline Engine*. Cette fois, nous avons demandé à 7 participants de retraduire ou post-éditer des segments issus de textes techniques avec l'objectif d'évaluer le degré d'acceptation de la suggestion de la traduction automatique, ainsi que le comportement du traducteur dans la post-édition.

Finalement, nous avons comparé le temps mis pour effectuer la même traduction par le participant qui a utilisé la fonctionnalité *Adaptive Engine* avec le temps mis en moyenne par les participants qui ont utilisé le *Baseline Engine*. Nous avons ainsi vérifié l'impact de la fonctionnalité *Adaptive Engine* sur la productivité temporelle par rapport au *Baseline Engine*.

Dans la deuxième partie (section 6.1.2), nous avons évalué la qualité de la traduction automatique par rapport à une traduction de référence à l'aide de mesures automatiques. Notre objectif était, d'un côté, de déterminer préalablement la qualité de la traduction « brute », afin d'obtenir un premier indicateur de l'intérêt de l'intégration de la TA et de l'autre côté, de vérifier l'effort de post-édition et la distance d'édition entre les traductions post-éditées et la référence. Enfin, notre but était d'étudier s'il y a une corrélation entre la perception de la qualité de la traduction automatique de la part des traducteurs et les résultats des mesures automatiques. Ce dernier aspect en particulier a été approfondi à l'aide du questionnaire final (Annexe III).

Dans l'étude, une place prépondérante a aussi été accordée à l'évaluation de la perception des tâches, des logiciels et de l'environnement de travail de la part des participants. Nous avons prévu un questionnaire final à cet effet dans la troisième partie de l'étude (6.1.3). L'objectif de ce questionnaire était d'une part, comme nous l'avons vu, d'établir une corrélation entre la perception de la qualité de la traduction automatique de la part des participants et les résultats des mesures automatiques ; de l'autre, de connaître la perception du traducteur par rapport au flux de travail et aux technologies utilisées, afin d'obtenir un indicateur supplémentaire de la rentabilité de l'intégration.

La structure du présent mémoire est la suivante : au chapitre 2, nous allons traiter la traduction automatique. Après un bref survol de son évolution et de son histoire, nous décrivons les principaux systèmes de traduction automatique à l'heure actuelle et ensuite nous analyserons leur architecture. Nous allons conclure la section avec l'état de l'art, ainsi que les perspectives de recherche actuelles en la matière.

Au chapitre 3, nous passerons à l'évaluation de la qualité de la traduction automatique ; nous allons ainsi présenter les méthodes adoptées dans notre étude.

Ensuite, au chapitre 4, nous allons aborder la post-édition ; après une description de cette activité, de ses buts et finalités, nous passerons à l'analyse des différents degrés de post-édition existants en relation à leur utilité. Après, nous allons décrire les avantages et les inconvénients liés à son exploitation. Nous allons ensuite tracer le profil du post-éditeur, une figure professionnelle en cours d'affirmation sur le marché du travail ; finalement, nous allons exposer les principales pistes de recherche actuelles en matière de post-édition.

Le chapitre 5 sera dédié à la description de l'expérience : d'abord, nous allons présenter l'intérêt et la finalité de l'étude ; ensuite, nous passerons à l'énonciation de notre hypothèse et des questions de

recherche et nous finirons avec la description du milieu expérimental. Dans cette section, nous allons décrire l'environnement et les étapes de travail suivies en relation à nos questions préliminaires. Ensuite, nous allons présenter les données expérimentales. Après, nous allons tracer le profil des participants à l'expérience. Puis, nous passerons à la description des logiciels utilisés dans l'étude. Finalement, nous allons conclure le chapitre avec l'exposition de la méthodologie de travail.

Le chapitre 6 sera consacré aux résultats obtenus. Nous allons présenter et analyser les données recueillies dans l'optique de la recherche. Nous allons conclure le chapitre avec un récapitulatif des réponses au questionnaire.

Finalement, le chapitre 7 sera consacré aux conclusions. Après avoir reparcouru brièvement les étapes de notre expérience, nous passerons à l'interprétation des résultats de l'étude, en cherchant à répondre aux questions et sous-questions de recherche. Finalement, nous allons exposer les limites et les possibles améliorations du travail, ainsi que les perspectives futures.

2. La traduction automatique

Après avoir donné quelques définitions préliminaires, nous allons effectuer un bref survol de l'origine de la traduction automatique (TA) et son développement au fil du temps. Nous allons nous concentrer sur les systèmes de TA statistiques dont le *Baseline Engine* est un exemple. Nous présenterons également les systèmes neuronaux qui sont au centre de l'évolution actuelle de la traduction automatique. Finalement, nous allons conclure le chapitre avec un panorama de l'état de l'art relativement à l'intégration de la TA aux outils d'aide à la traduction et au flux de travail des entreprises.

2.1 Introduction et bref historique

La traduction automatique (TA) se définit comme « l'application de l'informatique à la traduction des textes d'une langue naturelle de départ (ou langue source LS) dans une langue d'arrivée (ou langue cible LC) », et « représente l'une des premières applications informatiques non numériques » (Bouillon, 1993, p. 15). Elle appartient à la discipline de la traductique, qui « recouvre l'ensemble des disciplines relatives à l'informatisation de la traduction » (Ibidem). Mais la traduction automatique est aussi un vaste domaine qui comprend différentes sous-disciplines ; selon une classification proposée par (William John Hutchins & Somers, 1997, p. 3), elle comprend aux deux extrêmes, d'un côté, la traduction automatique assistée par l'humain (TAAH), aussi appelée traduction automatique assistée (TAA), et de l'autre, la traduction entièrement automatique de haute qualité (TEAHQ). La disposition de ces deux sous-disciplines aux pôles s'explique à partir du degré d'automatisation des tâches ou, inversement, du degré d'intervention du traducteur humain : si la TAAH accorde une place majeure au traducteur, bien qu'il soit parfois relégué au simple rôle de « réviseur » de la traduction sortante du logiciel, la TEAHQ, au contraire, vise à automatiser complètement la tâche de la traduction et par conséquent à mettre totalement à l'écart l'être humain. D'après (William John Hutchins & Somers, 1997, p. 3), « [t]he central core of MT itself is the automation of the full translation process ». En effet, dès le départ, la TEAHQ était le premier objectif de traduction automatique. Finalement, l'idée d'une automatisation complète du processus de traduction dans son entièreté, sans l'intervention humaine, est restée pendant longtemps « un horizon inaccessible » (Jacqmin, 1993, p. 43) et n'a donné des résultats satisfaisants que dans des cas très rares, dont le plus connu est l'expérience *TAUM-METEO*. Il s'agit d'un projet développé et implémenté en 1978 à l'Université de Montréal, au Canada, dont la finalité est la traduction des bulletins météo entre l'anglais et le français. Dans cette classification, il faut préciser que la TA n'inclut pas les outils informatiques d'aide à la traduction, qui font partie du domaine de la traduction humaine assistée par ordinateur (THAO), en raison de la primauté accordée à l'être humain dans l'activité de traduction. La relation entre traduction

automatique et outils d'aide à la traduction (ou « Computer-Aided Translation » en anglais) est résumée dans la Fig. 1.

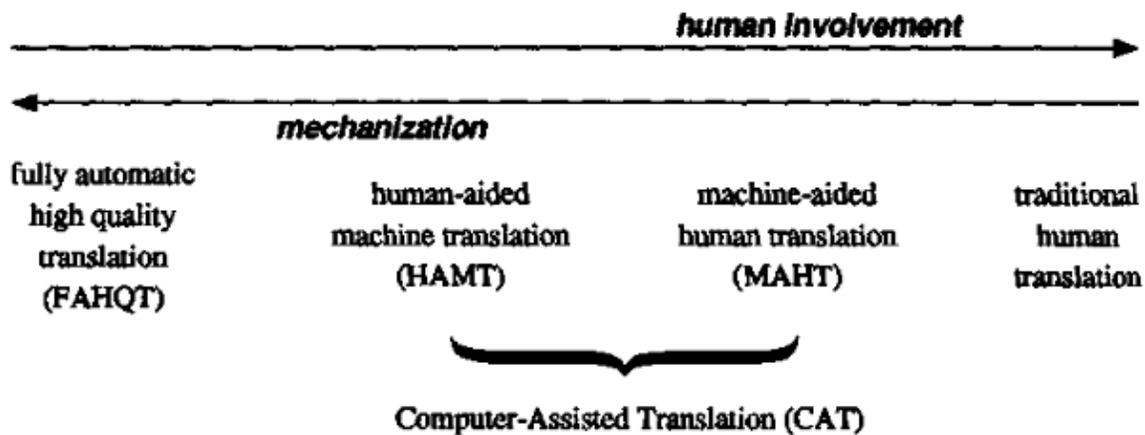


Figure 1 : Classification des types de traduction, Hutchins & Somers, 1997, p. 148

Actuellement, la majorité des systèmes de traduction automatique nécessite un certain degré d'intervention de la part du traducteur humain. Celle-ci peut intervenir soit avant la soumission du texte à la traduction, soit directement sur la traduction sortante du logiciel ; dans le premier cas, on parlera de pré-édition, et dans le deuxième de post-édition. La pré-édition est une activité *a priori* qui comprend l'ensemble des tâches effectuées sur la source (traduction entrante), finalisées à la préparation du texte au traitement par le logiciel de traduction automatique, en anticipant les potentielles « sources d'erreur » et les constructions susceptibles de créer des ambiguïtés ; elles sont typiquement réalisées à l'aide d'un ensemble de règles, qui visent à limiter le langage du texte source pour faciliter son traitement de la part du logiciel (Nyberg et al., 2003, p. 245). Cette étape s'applique surtout à la TA linguistique.

La post-édition, au contraire, prévoit l'intervention du traducteur humain *a posteriori*, pour « réviser » ou « corriger » la cible de la traduction automatique (Allen, 2003, p. 297). Le degré d'intervention sur la traduction sortante varie en fonction du type de texte, de la qualité demandée, ainsi que de la finalité à laquelle il est destiné. Un texte destiné à la publication, comme par exemple un rapport ou un manuel d'instructions, requerra une post-édition « complète », alors que le texte d'un courriel électronique ou d'un blog ou encore d'un forum demandera une simple traduction pour l'assimilation, c'est-à-dire, une traduction minimale pour parvenir à une compréhension approximative du contenu (Koehn, 2010, p. 217) ; par conséquent, l'intervention humaine sera mise de côté, ou limitée à la seule correction des erreurs « grossières ». Nous allons détailler l'activité de post-édition au chapitre 4.

Maintenant, nous passons au panorama des événements les plus marquants dans l'histoire des recherches en traduction automatique.

L'histoire des études en traduction automatique se développe autour d'un événement d'importance capitale, le rapport ALPAC de 1966, qui constitue le grand tournant dans les études dans le domaine. En effet, cet événement établit deux pistes de recherche : tout ce qui a été créé avant 1966 constitue la première génération des systèmes de traduction automatique, alors que ce qui suit est généralement indiqué comme la deuxième génération (Quah, 2006, p. 30).

Le développement de la discipline coïncide avec la conception et les premières utilisations des ordinateurs. En effet, les premiers espoirs de la recherche en traduction automatique commencent dans un contexte d'espionnage, en pleine Guerre Froide (débutée en 1947), lorsque les ordinateurs commençaient à être utilisés pour déchiffrer les codes entre l'anglais et le russe. On attribue au mathématicien Warren Weaver de la Fondation Rockefeller en 1949 l'idée d'utiliser les ordinateurs pour traduire des langues naturelles, une conception qui a permis de poser les bases de recherche dans le domaine, et qui a consacré le début de la première génération de logiciels de traduction automatique (William John Hutchins & Somers, 1997, p. 5). Sur la base des succès obtenus dans le déchiffrement des langages codés à l'aide des ordinateurs, dans son *Memorandum*, Weaver avait envisagé la possibilité d'exploiter les technologies naissantes pour parvenir à une traduction entièrement automatique de haute qualité. En particulier, il était le premier à avoir établi une analogie entre déchiffrement et traduction et donc à avoir effectué « un lien entre la traduction et la cryptographie » (H. Somers, 2003, p. 2). Dans les années suivantes, il a encouragé la recherche aux Etats-Unis afin d'utiliser les premiers ordinateurs pour la traduction. Les langues analysées à cette époque étaient l'anglais et le russe.

Finalement, en 1964, les autorités qui avaient financé le projet souhaitèrent examiner les progrès réalisés dans le domaine et les fruits de leurs onéreux investissements. Pour cela, elles nommèrent le comité ALPAC (*Automatic Language Processing Advisory Committee*), une commission chargée de l'évaluation des résultats obtenus en matière de TEAHQ en termes de qualité, vitesse et coûts.

En novembre 1966, le comité parvient à la rédaction du document qui met fin pour une longue période aux financements et aux recherches dans le domaine de la traduction automatique. Dans le document, la commission conclut que les systèmes évalués étaient très lents, moins précis par rapport à une traduction humaine, ainsi que très coûteux. Pour la première fois, le rapport mit en lumière l'inadéquation de la TEAHQ, tout en soulignant son échec total. Aussi, à partir de ce moment, les recherches dans le domaine

des technologies subissent un brusque virage pour commencer à s'intéresser aux outils d'aide à la traduction. À partir des années 90, nous assisterons à l'essor des premières mémoires de traduction (William John Hutchins, 1996, p. 6).

L'un des problèmes de fond dans l'évaluation reposait dans la surestimation des capacités de la traduction automatique. Les conclusions très négatives auxquelles parvient le rapport étaient biaisées par les attentes excessivement élevées de la communauté scientifique et du grand public envers la traduction automatique. En effet, à l'époque, une traduction entièrement automatique de haute qualité, sans intervention humaine et sans aucune restriction sur le type de textes, était encore loin d'être la réalité (Jacqmin, 1993, p. 44).

Heureusement, en Europe et au Canada, les études purent continuer à l'aide de subsides privés : c'est grâce à ces groupes de recherche que nous assisterons à la reprise de l'intérêt vers la traduction automatique à partir de la fin des années 1970.

En 1975, dans une période qui voit une multiplication des acteurs (et des langues) sur la scène internationale, nous assistons à la reprise des recherches en raison de l'intérêt renouvelé dans les possibilités offertes par la traduction automatique ; c'est le début de la deuxième génération des systèmes de traduction automatique. Les recherches se focalisent surtout autour des systèmes linguistiques qui utilisent une approche linguistique - *rule-based*, c'est-à-dire, basés sur l'application de règles encodées par le traducteur humain. L'un des plus grands succès de cette époque, ainsi que l'un de plus efficaces systèmes de traduction entièrement automatique de bonne qualité, est représenté par le projet déjà cité *TAUM-METEO* pour la traduction des bulletins météorologiques de l'anglais vers le français.

À partir des années 80, les chercheurs abandonnent l'étude des systèmes *rule-based* pour se tourner vers les approches basées sur corpus (en anglais *corpus-based*) ; maintenant, nous assistons à une nouvelle phase de transition : la performance d'un système de traduction automatique n'est plus réalisée à travers l'application d'une série de règles, encodées dans le logiciel manuellement par un développeur humain ; nous envisageons la possibilité d'exploiter les ressources lexicales et syntaxiques par extraction d'un corpus de textes alignés. Cette troisième génération de systèmes constituera la base pour le développement de l'approche statistique, développée à partir des années 2000.

À partir de 2016, en outre, nous observons une nouvelle tendance dans les systèmes de traduction automatique : le nouveau défi est constitué par la traduction neuronale (en anglais « neural machine translation »). À l'heure actuelle, un nombre toujours croissant de sociétés a effectué la transition de la

traduction statistique (section 2.2) vers l'approche neuronale. Nous examinerons cette nouvelle approche à la section 2.3 et finirons ce chapitre en abordant l'intégration de la TA aux systèmes de mémoire de traduction.

2.2 Les systèmes statistiques

La « grande famille » des systèmes statistiques présente plusieurs sous-catégories, chacune basée sur une approche particulière. Principalement, nous y retrouvons les approches *word-based* (basées sur les mots) et celles *phrase-based* (basées sur les phrases). Chaque système fonde son fonctionnement sur un calcul de probabilité effectué à partir d'un corpus de textes bilingues plus ou moins étendu, qui constitue son répertoire de connaissances ainsi que sa base d'entraînement. Plus précisément, ces systèmes reposent sur des modèles statistiques et probabilistes entraînés à partir d'un corpus d'entraînement. Ces modèles n'incluent aucune information linguistique. En revanche, ils contiennent des informations relatives aux propriétés distributionnelles des mots et des phrases, à partir desquelles ils parviennent à choisir la traduction la plus probable. Pour cela, le logiciel tient compte de la nature statistique du langage : à partir d'un corpus bilingue, il recherche les traits de langage récurrents, facilement mesurables, et vérifie comment les informations extraites de ces traits peuvent être exploitées pour prédire les traductions futures (Trujillo, 1999, p. 20). Nous verrons dans les détails en quoi consiste le processus de traduction.

L'approche statistique basée sur un corpus bilingue aligné fut l'une des premières techniques proposées pour la TA (H. L. Somers, 1993, p. 149). Concrètement, le processus de traduction consiste en deux étapes fondamentales : l'entraînement (en anglais *training process*) et le décodage (ou *processing phase*). Lors de la première étape, le système utilise un grand corpus de textes alignés par phrases pour extraire automatiquement les informations relatives à la traduction des mots et aux possibles alignements des mots au niveau des phrases. Dans la deuxième, il se sert de ces informations pour traduire les nouvelles phrases qui lui sont soumises (Poibeau, 2017, p. 45).

L'idée de base de cette approche est la suivante : étant donné un corpus parallèle assez vaste, toute phrase du corpus de la langue A est une traduction possible de toute phrase du corpus de la langue B, avec une probabilité plus ou moins grande. Par conséquent, d'après (H. L. Somers, 1993, p. 157), « la tâche de la traduction consiste à trouver pour une phrase donnée S la phrase C dans la langue cible dont la probabilité $\Pr(S|C)$ est maximale ».

Cette probabilité, à son tour, dépend de la probabilité que les mots d'une phrase donnée C correspondent à ceux de la phrase S ; par conséquent, une variable importante dont il faut tenir compte est représenté par l'ordre des mots.

Comme nous l'avons vu, l'approche statistique se fonde sur un calcul probabiliste. Pour déterminer la traduction d'une phrase du texte source, le système utilise des probabilités mesurées à partir d'un corpus. Ces probabilités sont encodées dans des modèles, qui correspondent aux règles des systèmes linguistiques (ou *rule-based*). La combinaison de ces modèles permet l'extraction de la meilleure traduction pour une phrase donnée. Voyons dans les détails de quoi il s'agit et comment ces systèmes fonctionnent concrètement.

Il existe deux modèles différents : le modèle du langage, extrait du corpus monolingue, qui détermine si la phrase C est fluide – c'est-à-dire, si elle « est ressentie comme naturelle » en langue source - et le modèle de traduction, extrait du corpus bilingue, qui détermine si la traduction est fidèle à la source.

Comme nous l'avons mentionné plus haut, à partir de la combinaison de ces deux modèles, le système est capable de déterminer la meilleure traduction pour une phrase donnée S ; celle-ci correspond à celle qui maximise l'équation suivante : fluidité (C) * fidélité (C,S)

Comme nous l'avons déjà expliqué, le modèle du langage se base sur un corpus monolingue : son objectif est donc d'évaluer la fluidité de la traduction. Il se base sur un modèle n-gramme. Un n-gramme correspond à une séquence de n mots. Il existe plusieurs modèles de n-grammes possibles : les unigrammes calculent les probabilités de mots individuels dans le corpus, les bigrammes de toutes les séquences de deux mots, et ainsi de suite. À travers le modèle du langage, le système segmente chaque phrase du texte et mesure la probabilité d'apparition de chaque séquence de n mots contenues dans le corpus d'entraînement. Finalement, la phrase qui obtient la plus grande probabilité pour l'ensemble de ses n-grammes sera sélectionnée comme traduction candidate.

Le modèle de traduction se base sur un corpus bilingue aligné. Comme nous l'avons vu, son objectif est d'évaluer la fidélité de la traduction. À cet effet, le système calcule d'abord le nombre de fois où le segment source et le segment cible sont extraits comme bi-segment, c'est-à-dire, comme deux segments en relation de traduction mutuelle. Ensuite, ce résultat est divisé par le nombre de fois où le segment est extrait avec tous les segments. Finalement, la phrase qui obtient le meilleur score est choisie comme traduction.

Il existe aussi un troisième modèle dont le système dispose : l'ordonnement. Ce modèle permet de calculer la probabilité qu'un segment change de place, ce qui nous permet aussi de traiter les changements d'ordre des mots.

Comme nous l'avons vu, dans les systèmes statistiques, les modèles de probabilités assignent un « score » à toute traduction « candidate » pour chaque phrase à traduire. Maintenant, le système doit opérer un choix pour sélectionner la « meilleure » traduction. Nous parvenons à la dernière étape du processus, effectuée par le modèle de « décodage », dont l'objectif est de déterminer et de sélectionner la traduction qui a obtenu le meilleur score. (Bouillon, 2016, non énuméré).

Comme nous l'avons déjà mentionné, les systèmes statistiques représentent l'un de plus grands succès en traduction automatique depuis les logiciels de deuxième génération, et constituent la base des développements futurs dans la recherche dans le domaine des outils d'aide à la traduction.

L'une des pistes de recherche moderne qui dérive de l'approche statistique est représentée par les systèmes neuronaux. Le débat est encore ouvert sur l'effective utilité de cette nouvelle approche. La question que se pose la communauté scientifique est la suivante : est-ce qu'ils reflètent vraiment l'état de l'art en traduction automatique ?

Dans la section suivante, nous présentons les systèmes neuronaux ; nous allons décrire leur architecture et le déroulement du processus de traduction. Ensuite, nous présentons les résultats d'une étude effectuée par (Castilho et al., 2017), qui compare la qualité de la traduction des systèmes statistiques et neuronaux à travers des méthodes d'évaluation humaines et automatiques. Ce travail démontre qu'il est encore prématuré de considérer l'approche neuronale comme étant l'état de l'art en traduction automatique, et souligne la nécessité d'effectuer des études supplémentaires dans le domaine. Nous concluons la section avec quelques réflexions élaborées à partir des résultats de l'expérience.

2.3 Les systèmes neuronaux

Conçus et développés sur la base des architectures statistiques, ces systèmes fondent leur fonctionnement sur le *deep learning*, un concept lié à l'intelligence artificielle, réalisé grâce à un réseau neuronal artificiel. Le point de départ de cette nouvelle approche est la volonté d'imiter le cerveau humain : les neurones, ce sont les cellules préposées à la transmission et au traitement des informations élémentaires, à partir desquelles le cerveau est capable de construire des concepts et des idées plus complexes. De même, les réseaux de neurones artificiels sont conçus pour élaborer des concepts complexes à partir de petits

morceaux d'informations élémentaires, assemblées de manière hiérarchique. Dans la pratique, le procédé de traduction se configure comme une phase d'extrapolation et d'élaboration de l'information simple, à partir de laquelle le système dérive un concept complexe. Grâce au *deep learning*, à partir des données d'entraînement, le système serait capable d'extraire et d'apprendre les informations contextuelles basiques de manière autonome et graduelle, pour ensuite produire une traduction (Poibeau, 2017, p. 25).

Les systèmes neuronaux naissent de l'intégration d'un module de systèmes statistiques. D'après (Koehn, 2017, p. 6),

[a] neural network is a machine learning technique that takes a number of inputs and predicts outputs. In many ways, they are not very different from other machine learning methods but have distinct strengths.

Les modèles neuronaux supposent le développement d'un réseau neuronal *end-to-end*, c'est-à-dire, capable de traiter la phrase à traduire dans son entièreté et d'effectuer le processus de traduction en une seule étape. Le point de départ est toujours constitué par des corpus de textes bilingues alignés. D'abord, le réseau neuronal extrait et analyse l'information contenue dans les corpus et parvient à l'organiser de façon structurée. Ensuite, étant donné une phrase X à traduire, le système est entraîné pour maximiser la probabilité d'une séquence cible Y sans avoir recours à des informations linguistiques externes supplémentaires (Castilho et al., 2017, p. 109). L'idée à la base de l'approche neuronale est donc celle d'exploiter des modèles de réseaux neuronaux artificiels pour apprendre un modèle statistique pour la traduction automatique. Cependant, à la différence des approches statistiques « classiques », qui basaient le processus de traduction sur l'analyse d'une grande quantité de données (*big data*) et l'application d'algorithmes pour l'extraction de la meilleure phrase cible parmi les potentielles candidates, les architectures neuronales ont été conçues pour travailler de façon autonome, tout comme un cerveau humain. À partir d'une série de données, le système est capable d'extraire et d'apprendre en autonomie les informations linguistiques et sémantiques, pour construire une sorte de « modèle de qualité ». Ensuite, il parvient à appliquer ses connaissances et à résoudre les situations traductives problématiques. Grâce au *deep learning*, le système parvient à construire un réseau de connaissances plus profond, qui lui permet de généraliser les problèmes et les aborder « d'en haut », tout comme un cerveau humain. En définitive, si les approches statistiques se focalisaient sur la quantité de données, les systèmes neuronaux, en revanche, mettent plutôt l'accent sur la qualité (Systran, 2017, non énuméré).

Comme nous l'avons vu plus haut, à la différence des systèmes linguistiques (*rule-based*) et statistiques, l'approche neuronale traite la phrase à traduire dans son entièreté dans une seule étape, sans effectuer aucune phase intermédiaire entre la source et la cible. Voyons maintenant le processus de ce type de traduction.

Le système neuronal modèle le processus de traduction à l'aide d'un réseau neuronal artificiel (Ibidem). Ce réseau est composé d'une série de sous-réseaux (ou modèles), chacun d'eux préposé à une tâche spécifique : le processus de traduction se fonde sur l'interaction de ces modèles.

Le premier modèle est préposé à l'analyse de la phrase source et à l'extraction du sens. Le deuxième modèle, spécialisé dans l'analyse syntaxique et sémantique, extrait des informations relatives à la grammaire et au contenu sémantique. Le troisième modèle se focalise sur les mots-clés de la phrase. Le quatrième modèle apporte des informations contextuelles pour assurer la fluidité de la traduction. Finalement, le cinquième modèle adapte la traduction à un domaine spécifique. La principale caractéristique de ces modèles est le fait d'être « multidimensionnels ». En effet, chaque composant, que ce soit une phrase ou un mot, est placé dans un contexte sémantique vaste. De ce fait, la supposition à la base de l'approche neuronale est que les composants qui apparaissent dans le même contexte ont une chance plus élevée d'avoir le même signifié. À partir de ces informations, le système est donc capable d'identifier et de grouper les mots dans des plongements (ou *words embeddings*) qui appartiennent au même contexte traductionnel – pour ensuite déduire le sens et générer des traductions (Poibeau, 2017, p. 20).

Une autre caractéristique importante de ces modèles est qu'ils sont continus. Étant donné que chaque composant (mot, phrase) est inséré dans un contexte très riche, il peut être comparé dans un espace sémantique continu – ce qui rend l'approche neuronale très flexible.

Concrètement, dans un système neuronal, la traduction se configure comme un processus en deux étapes : l'encodage et le décodage. Pour simplifier, l'encodeur extrait l'information linguistique et sémantique de la source et l'organise de façon intégrée dans des « contextes », qui constituent la représentation numérique de la phrase. À partir de ces contextes, le décodeur génère la traduction en produisant les mots-cibles.

L'approche neuronale représente une étape importante dans les recherches en traduction automatique : elle a été conçue pour surmonter les difficultés posées par la traduction automatique statistique, notamment le besoin de vastes corpus parallèles pour entraîner les systèmes (Levin et al., 2017, p. 80) ;

(Shterionov et al., 2017, p. 74). Beaucoup d'études récemment conduites montrent l'efficacité de cette nouvelle approche, notamment celle effectuée par (Levin et al., 2017), qui déclare même la supériorité des systèmes neuronaux par rapport à celle de l'approche statistique au niveau de la qualité. Grâce à ces résultats prometteurs, cette approche est en train de gagner en popularité autant au niveau académique que dans l'industrie de la traduction. Pour cette raison, un certain nombre de sociétés fournisseuses de logiciels de traduction automatique, telles que *Google*, *Microsoft* et *Systran*, ont inclus la traduction neuronale dans leur offre. En particulier, après le lancement du nouveau système, *Google* a démontré que la qualité de la traduction automatique de l'approche neuronale surpasse de manière substantielle celle de l'approche linguistique *phrase-based*, en réduisant le taux d'erreurs de traduction d'environ 60% (Wu et al., 2016, p. 1). Le 18 Mars 2018, une équipe de chercheurs de *Microsoft*⁴ a annoncé avoir créé le premier système de traduction automatique neuronale capable de traduire des phrases des articles de presse du chinois vers l'anglais (et vice-versa) avec la même précision que la traduction humaine, ce qui leur a permis d'atteindre la « parité avec l'être humain » (l'objectif *achieving human parity*, tel qu'ils le définissent). D'après les résultats de l'étude, la qualité de la traduction serait à la parité de celle produite par l'être humain, ce qui rendrait l'approche neuronale le nouvel état de l'art en traduction automatique (Hassan et al., 2018, p. 1).

Cependant, il est important de préciser que cette approche, jusqu'à récemment, a fait l'objet d'études et d'améliorations. D'après (Poibeau, 2017, p. 30), l'une des difficultés principales résidait dans le traitement des phrases plus élaborées. Ce problème était lié à un certain nombre de raisons : premièrement, à la complexité de la phase d'entraînement ; deuxièmement, aux difficultés rencontrées dans le traitement des mots inconnus, ce qui résultait parfois dans des traductions bizarres et imprécises. Pour éviter ces obstacles, les chercheurs ont élaboré des techniques d'optimisation dont la finalité est surtout de faciliter le processus d'apprentissage de l'encodeur et du décodeur. L'une des solutions qui a été proposée est l'implémentation d'un *attention model*. Ce modèle permettrait au système de se focaliser sur les mots considérés importants aux fins de la traduction à un stade donné du procédé, afin d'établir des connexions entre l'encodeur et le décodeur. Par conséquent, l'*attention model* faciliterait l'entraînement du système et, finalement, il optimiserait la traduction (Systran, 2017, non énuméré).

⁴ <https://blogs.microsoft.com/ai/machine-translation-news-test-set-human-parity/> (dernière consultation : 02/05/18)

Pour rendre compte de l'état actuel de la recherche en traduction neuronale, nous résumons les résultats d'une étude effectuée par (Castilho et al., 2017). Le but du travail est de comparer la qualité de la traduction effectuée par des systèmes neuronaux et des systèmes statistiques en utilisant des méthodes d'évaluation humaines et automatiques. L'expérience est organisée en trois études ; chaque étude analyse les performances des systèmes sur des textes appartenant à un domaine spécifique. À la suite de chaque étude, comme nous l'avons déjà mentionné, la qualité de la traduction sortante est évaluée à travers des méthodes humaines et automatiques. D'après les résultats, bien que les scores obtenus par l'approche neuronale avec l'évaluation automatique soient prometteurs, l'évaluation humaine montre des résultats mélangés. L'étude en conclut que, actuellement, il n'est pas possible d'affirmer que l'approche neuronale constitue l'état de l'art dans la recherche en traduction automatique, et démontre la nécessité d'effectuer des études supplémentaires pour mesurer l'efficacité réelle de cette nouvelle approche par rapport aux systèmes statistiques.

Comme nous pouvons le déduire, au sein de la communauté scientifique, les opinions demeurent divergentes relativement à la place à accorder à l'approche neuronale dans la recherche en traduction automatique. D'après (Shterionov et al., 2017, p. 74),

(...) it is still uncertain to which extent NMT can replace SMT as core technology for large-scale translation projects. The main reasons are the computational (and financial) cost of NMT and the uncertainty in the actual quality: while NMT output is often very fluent, sometimes it lacks adequacy or is even completely wrong.

En définitive, avant de déclarer l'approche neuronale comme étant l'état de l'art en traduction automatique, il serait nécessaire d'effectuer plus de recherches, en particulier au niveau de la qualité de la traduction sortante, ainsi que d'évaluer le rapport coûts-bénéfices lié à implémentation de l'approche dans le flux de travail de projets à grande échelle.

Avec cette section, nous concluons la description des systèmes de traduction automatique statistiques et neuronaux. Dans la section suivante, nous allons exposer un panorama des pistes d'intégration de la traduction automatique aux outils d'aide à la traduction à l'heure actuelle.

2.4 Vers l'intégration de la TA aux outils d'aide à la traduction

D'après (Zaretskaya et al., 2015, non énuméré), à l'heure actuelle, il existe principalement deux méthodes pour intégrer la traduction automatique aux outils d'aide à la traduction : l'une dite « interne » et l'autre

dite « externe ». La première consiste en l'exploitation d'un logiciel de traduction automatique et d'une mémoire de traduction pour améliorer la qualité de la traduction sortante, augmenter la productivité et réduire l'effort de post-édition. La deuxième prévoit l'inclusion des suggestions du logiciel de traduction automatique avec les propositions d'autres ressources, telles que des mémoires de traduction, des bases de données terminologiques, ainsi que des glossaires. Plus spécifiquement, il existe deux méthodes de combinaison externe : l'une dite *batch processing* (ou *offline method*) et l'autre dite *interactive processing* (ou *online method*). La première consiste en l'application de la traduction automatique avant la phase de traduction pour les segments pour lesquels aucune correspondance n'est extraite de la mémoire de traduction. En revanche, la deuxième permet de soumettre les segments pendant la phase de traduction grâce à un plugin ou à un *API* (en anglais *Application Programming Interface*). Dans cette optique, l'inclusion d'un module de traduction automatique au système de mémoire de traduction *SDL Trados Studio* se configure plutôt comme une intégration externe interactive. Cependant, elle présente des innovations par rapport aux méthodes précédentes, notamment en ce qui concerne la fonctionnalité *Adaptive Engine*.

2.5 Vers l'intégration de la TA dans le flux de traduction des entreprises

La traduction automatique représente souvent une méthode plus rentable par rapport à la traduction humaine ; un certain nombre d'études démontrent son efficacité, particulièrement en ce qui concerne la traduction des textes techniques rédigés en langage contrôlé, tels que ceux utilisés dans notre expérience ((O'Brien, 2004) ; (Fiederer & O'Brien, 2009)). Dans les premières années, les recherches étaient focalisées dans l'application de la TA à des domaines restreints ; l'un des exemples typiques était le domaine touristique (Vilar et al., 2012, p. 73). Successivement, avec l'amélioration de la qualité de la traduction automatique, les chercheurs se sont lancés le nouveau défi d'appliquer la technologie à des domaines plus riches et variés du point de vue terminologique et syntaxique ; par conséquent, ils ont manifesté la tendance à surestimer les capacités de la TA dans les domaines techniques et à faire abstraction des recherches dans ces champs « de niche ». D'après (Vilar et al., 2012, p. 73), en particulier, la communauté scientifique a négligé l'étude de la véritable faisabilité de l'exploitation de techniques de TA dans des situations de travail réelles :

[i]n real-world translation environments efficiency, both in terms of cost and time, is of critical importance. Even more when the volume of texts to translate is large. Machine translation (MT) seems to be a good candidate for achieving these goals,

but somehow surprisingly the economic feasibility of MT and the fitness for real-world needs of professional translators and Language Service Providers (LSPs) have been hardly analysed so far.

D'après les auteurs, beaucoup plus d'efforts devraient être investis dans l'exploration des potentialités de la TA dans les domaines techniques plus restreints, plus proches aux besoins des utilisateurs, ainsi que dans l'analyse de sa rentabilité dans des environnements de travail concrets. De plus, ils reprochent parfois un certain degré de résistance de la part des potentiels clients à financer l'entraînement des logiciels de traduction automatique, cette ressource étant considérée comme « gratuite ». Dans l'article, (Vilar et al., 2012) ont présenté les résultats d'une étude pilote conduite dans un environnement de travail réel, qui portait sur l'applicabilité de techniques de traduction automatique statistique dans un projet de traduction concret. Le projet a été effectué par une agence de services linguistiques (*Language Service Provider* ou *LSP*, en anglais) qui utilisait habituellement des systèmes de mémoire de traduction dans son flux de travail. Le matériel à traduire consistait en de la documentation technique, plus précisément des diapositives, dont le contenu était très répétitif. L'un des objectifs de l'agence était de garder les coûts bas, afin de maintenir l'activité rentable. Les prérequis pour le logiciel de traduction automatique étaient, d'un côté, la facilité d'intégration de la TA au système de mémoire de traduction, et donc la possibilité pour le traducteur d'exploiter aisément les deux ressources comme d'habitude ; de l'autre, la qualité de la traduction sortante devrait être suffisamment élevée pour ne pas requérir aucune post-édition, afin de modérer les coûts. En d'autres termes, les segments traduits automatiquement devaient être présentés comme *Perfect Match*, qui équivaut à une correspondance à 100% de la mémoire de traduction. Les résultats montrent que la qualité de la traduction automatique était suffisamment élevée pour envisager d'intégrer avec succès la TA dans le flux de travail de l'agence comme complément au système préexistant de mémoire de traduction.

À ce propos, il est aussi intéressant de citer l'article de (Hofmann & Lèpan, 2017). Les auteurs présentent les résultats très satisfaisants de l'intégration et de l'implémentation d'un système de traduction automatique pour la traduction de la documentation officielle des Chemins de Fer Fédéraux de la Suisse (CFF). Depuis 2001, le département des Services Linguistiques au sein des CFF dispose de trois ressources fondamentales : un système de gestion de projets de traduction, un système de gestion de la terminologie, ainsi qu'un système de mémoire de traduction. Grâce à ces ressources, avec les années, les Services Linguistiques des CFF sont parvenu à constituer un vaste dictionnaire de terminologie officielle. Dans la perspective de l'étude, l'intégration de la traduction automatique était vue dans l'optique d'une valeur

ajoutée au travail des traducteurs. Deux approches ont été conçues : l'une prévoyait l'intégration de la traduction automatique dans le flux de travail des Services Linguistiques comme support pour les traducteurs, sous forme de suggestions de traduction, à côté des résultats offerts par les mémoires de traduction ; l'autre prévoyait l'intégration de la traduction automatique dans l'intranet des CFF, comme support pour tous les employés en fournissant des traductions *ad hoc*. D'après les résultats, grâce à l'introduction de suggestions de traduction additionnelles, générées par le système de TA, à côté des suggestions de la MT, les correspondances à 100% ont augmenté de 50%, tandis que les correspondances floues (70-99%) ont doublé. L'étude démontre l'intérêt ainsi que les avantages dérivant de l'implémentation d'un système de traduction automatique au sein d'une agence de Services Linguistiques – notamment le gain sur la productivité – et encourage son introduction au sein d'autres entreprises comme solution pour maximiser la production.

Similairement, dans notre travail, nous explorerons la rentabilité de l'intégration d'un module de traduction automatique avec post-édition dans le flux de travail d'un traducteur indépendant comme moyen pour maximiser la production. Spécifiquement, nous allons examiner l'intérêt de l'intégration en mesurant trois variables : la productivité temporelle, la qualité, ainsi que la satisfaction du traducteur.

Avec cette section, nous concluons le chapitre dédié à la traduction automatique.

Dans le chapitre suivant, nous abordons la thématique de l'évaluation de la traduction automatique, et nous décrivons les deux méthodes principales : l'évaluation humaine et l'évaluation automatique.

3. L'évaluation de la traduction automatique

L'évaluation de la traduction automatique revêt un rôle central dans l'expérience traitée dans le présent mémoire. En effet, une place importante sera accordée à l'évaluation de la qualité de la traduction sortante du *Baseline Engine* et de l'*Adaptive Engine* par rapport à une traduction de référence. À cet effet, il est primordial de dédier une section à l'explication de son utilité et de son intérêt, ainsi que présenter les méthodes adoptées dans notre étude.

L'évaluation de la traduction automatique peut être considérée comme une sous-catégorie de la discipline relative à l'évaluation des logiciels d'aide à la traduction. D'après (Kit Chunyu & Tak-Ming, 2015, p. 213),

[t]he evaluation of MT/CAT deals with the issue of quantifying their effectiveness. In a broad sense, it is intended to systematically assess the quality, success and efficacy of any aspect of an MT/CAT system that gives rise to a concern about the degree of the system's usefulness.

Cependant, l'évaluation de la traduction des systèmes reste au centre de la discipline : en effet, elle est le principal intérêt ainsi que le plus pertinent critère de jugement de la performance d'un logiciel (Ibidem).

La perspective adoptée dans le présent travail est celle de l'analyse de la qualité de la traduction sortante du logiciel. Par conséquent, nous ne tiendrons pas compte des attributs du système-même, c'est-à-dire, de ses caractéristiques intrinsèques et extrinsèques, comme par exemple son utilité, sa fiabilité, la facilité à l'usage ou le degré d'intégration avec d'autres outils dans le flux de travail.

L'évaluation de la traduction automatique est inséparable du développement des premières technologies de MT. Cependant, elle commence à s'affirmer en tant que discipline autonome dès l'apparition sur le marché des premiers logiciels de traduction automatique commerciaux de type statistique. En effet, la compétition globale accrue a fait naître la nécessité de développer des critères d'évaluation objectifs pour permettre l'encadrement de chaque logiciel au niveau de son utilité et de sa performance.

Il existe principalement deux approches à l'évaluation de la traduction automatique. La première que nous allons analyser est l'approche dite « humaine », effectuée par des êtres humains ; la deuxième, plus rentable, est l'évaluation « automatique », effectuée normalement par l'ordinateur à travers des mesures automatiques. Nous précisons que, dans le travail, nous avons adopté cette dernière approche.

Les questions que nous pouvons nous poser et qui sont à la base de l'exécution de toute évaluation sont, par exemple, les suivantes : à quel niveau de qualité correspond la traduction dite « brute » (non post-

éditée) du système de traduction automatique en examen ? Ou plus simplement, dans quelle mesure la traduction sortante du système est-elle « adéquate » pour la finalité à laquelle le texte est destiné ? Et encore, comment cette traduction peut-elle être utilisée de la façon la plus rentable ?

Avant de nous focaliser sur les différentes modalités d'évaluation, il est important de définir préalablement ce que nous attendons d'un système. En effet, les questions exposées ci-dessus nous amènent à une problématique centrale, à la base de toute évaluation : la définition de la notion de qualité. Il n'existe pas une façon univoque d'entendre ce concept : dans le cadre du présent travail, nous allons adopter la perspective de (Krauwier, 1991, p. 245) : « the main purpose of evaluation is to determine to what extent the makers of a system have succeeded in mimicking the human translator ». Nous pouvons déduire que le point de départ, ainsi que l'étalon-or pour juger la qualité d'une traduction automatique, est la traduction humaine, ainsi appelée « référence ». L'objectif de l'évaluation sera donc d'estimer le degré auquel le système est capable d'imiter le traducteur humain. Nous verrons dans les prochaines sections les différentes méthodes pour déterminer la qualité, particulièrement en relation au type et à la finalité du texte.

Dans la prochaine section, nous allons présenter les différentes approches à l'évaluation disponibles, en analysant leur degré d'utilité ainsi que les respectifs avantages et inconvénients.

3.1 L'évaluation humaine

D'après (Koehn, 2010, p. 218), l'évaluation humaine se fonde sur les jugements exprimés relativement à la qualité d'une traduction de la part des êtres humains (pas nécessairement des traducteurs) en vertu de leurs connaissances et de leur faculté de discernement. Elle prévoit la présentation d'un certain nombre de phrases traduites par un logiciel de traduction automatique à un juge humain, auquel on demande de s'exprimer sur la justesse de la cible. Ce type d'évaluation peut être effectué avec ou sans phrase source ; dans le premier cas on parlera d'évaluation bilingue, dans le deuxième d'évaluation monolingue. Cependant, les évaluateurs bilingues sont parfois très difficiles à trouver, et sont très coûteux par rapport aux monolingues (Ibidem). Pour éviter ce problème, il est aussi possible d'effectuer une évaluation en utilisant des évaluateurs monolingues qui comprennent la langue cible et de leur fournir une traduction de référence.

Normalement, il est fréquent de fournir aux évaluateurs une échelle de points (normalement, avec des valeurs de 1 à 5) afin de mesurer la qualité de chaque phrase.

Dans l'approche humaine, l'évaluation est effectuée à l'aide de deux critères : l'adéquation (en anglais *adequacy* ou *fidelity*) et la fluidité (*fluency*).

L'adéquation mesure le degré auquel la traduction sortante transmet et conserve le sens de la phrase source. Elle est effectuée en version bilingue, avec la phrase source et la cible, auxquelles parfois nous ajoutons la référence. En reprenant l'échelle de points, nous pouvons associer à chaque valeur un jugement : la valeur 1 correspond à aucun sens transmis ; 2, très peu de sens ; 3, beaucoup de sens ; 4, la plupart du sens ; 5, la totalité du sens.

En revanche, l'évaluation de la fluidité peut être effectuée en mode monolingue, c'est-à-dire, seulement avec la cible. Elle mesure le degré de « fluidité » de la traduction, c'est-à-dire, si elle semble « naturelle » à un natif de cette langue.

Si nous reprenons l'échelle de points, nous pouvons assigner une valeur de 1 à une phrase incompréhensible ; 2 à une phrase non fluide ; 3 si elle reflète la langue produite par un non-natif ; 4, si elle est d'un bon niveau ; 5, si elle est parfaite.

Ces deux critères peuvent être évalués sur une échelle absolue ou relative. Dans le premier cas, les annotateurs évalueront la traduction sortante d'un même système par rapport à une référence ; en revanche, dans le deuxième cas, les annotateurs évalueront la qualité de la traduction sortante de deux systèmes différents phrase par phrase (Koehn, 2010, p. 222). Dans ce dernier cas, la question que nous allons leur poser sera : « Est-ce que la traduction du système A est meilleure que celle du système B, pire, ou équivalente ? »

L'un des inconvénients de cette approche est le manque de précision de ces descriptifs. En effet, il est difficile (voire impossible) d'établir une échelle « globale » et parfaitement objective, capable d'exprimer exactement d'une façon universellement acceptable les signifiés associés aux scores. C'est pourquoi, il est difficile pour les annotateurs d'assurer la cohérence dans leurs jugements.

Une autre difficulté est relative aux annotateurs eux-mêmes. En effet, un nombre illimité de facteurs peut déterminer le jugement : chacun va s'exprimer selon sa perception, son degré de connaissance de la langue source ou cible, son éducation ; par conséquent, bien qu'il soit indispensable d'uniformiser le plus possible notre jury, il est impossible d'assurer l'homogénéité ou la cohérence absolues dans leurs jugements.

Ensuite, un inconvénient ultérieur est représenté par le temps : dans ce sens, l'évaluation humaine est très dispendieuse. Cela va à l'encontre des intérêts des chercheurs, qui souhaitent obtenir des résultats dans les meilleurs délais, et parfois exécuter plusieurs évaluations par jour.

Finalement, un autre inconvénient est lié aux coûts ; les annotateurs doivent percevoir une récompense pour leur travail ; de plus, cette rémunération sera plus élevée pour les bilingues.

Cette longue série d'inconvénients a encouragé les chercheurs à élaborer d'autres systèmes plus rentables pour estimer la qualité de la traduction d'un système de traduction automatique ; en effet, comme nous verrons plus bas, le grand atout lié à l'exploitation d'une mesure automatique est lié à sa rentabilité. Dans la prochaine section, nous allons présenter les mesures adoptées dans notre travail ; aussi, nous passerons à l'examen de l'utilité et des objectifs de cette approche.

3.2 L'évaluation automatique

D'après (Koehn, 2010, p. 222), l'évaluation automatique consiste en « an automatic method for assessing the quality of machine translation output ».

Avant d'expliquer le concept sous-jacent à toute évaluation automatique, il est important d'exposer ce que nous attendons d'une mesure automatique, particulièrement en comparaison avec l'évaluation humaine.

D'après (Koehn, 2010, p. 220), une mesure doit être avant tout rentable. Comme nous l'avons déjà vu, le coût représente l'un des majeurs obstacles liés à l'évaluation. L'intérêt des chercheurs est d'exploiter une mesure automatique pour exécuter plusieurs évaluations sur plusieurs domaines, possiblement dans de brefs délais. Ainsi, un autre atout exigé de toute mesure est le fait qu'elle soit adaptable, c'est-à-dire, capable de s'adapter à plusieurs champs et typologies textuelles. Un autre prérequis de toute mesure est qu'elle soit significative, c'est-à-dire, que chaque score puisse correspondre à un critère d'évaluation objectif. Ensuite, pour qu'elle soit valide, une mesure doit être cohérente et fiable ; plus précisément, elle doit permettre à plusieurs annotateurs qui utilisent les mêmes critères de parvenir aux mêmes conclusions. Cette concordance est appelée accord inter-annotateurs. Finalement, pour qu'elle soit acceptable, nous attendons qu'une mesure soit stable pendant l'évaluation, c'est-à-dire, qu'elle ne donne pas lieu à des fluctuations dans les résultats ; par exemple, lorsque nous testons plusieurs parties du même corpus de données, les scores doivent rester autant que possible homogènes.

Pour évaluer l'effective fiabilité de ces mesures, l'institut NIST (Koehn, 2010, p. 229) a conduit une étude qui évaluait la performance des systèmes de traduction automatique avec des scores automatiques et des jugements humains. Les résultats ont montré que les systèmes qui avaient reçu des scores automatiques très bas avaient aussi reçu des jugements humains très bas. Cette étude a révélé un haut niveau de corrélation entre mesures automatiques et humaines, et a démontré la fiabilité de ces méthodes.

Cependant, les mesures automatiques comportent aussi des limites. D'après les résultats de l'évaluation conduite par NIST en 2010 (*The NIST Metrics for Machine Translation 2010 Challenge - MetricsMaTr10*⁵), « [a]utomatic metrics have not yet been proved able to consistently predict the usefulness, adequacy, and reliability of MT technologies ». En outre, (Way, 2010, p. 27) reconnaît aussi que

(...) while the introduction of automatic evaluation metrics in MT...has largely been beneficial, they have to a large extent taken on too much importance, especially since real translation quality is what we should be concerned with.

Après avoir exposé les principaux prérequis de toute mesure automatique, nous passons maintenant à la description de leur fonctionnement de base. En particulier, nous allons examiner deux concepts-clés : la « précision » et le « rappel ».

Pour simplifier, lors d'une évaluation automatique, l'ordinateur compare la traduction sortante de chaque système examiné avec la traduction humaine de la même phrase, que nous avons définie comme « traduction de référence ». Dans l'optique d'une mesure basée sur les mots, la précision d'un système est calculée à partir du nombre de mots traduits correctement par le système, divisé par le nombre total des mots de la cible (Koehn, 2010, p. 223).

Le « rappel », en revanche, est calculé à partir du nombre de mots traduits correctement par le système, divisé par le nombre de mots de la référence. Dans cette optique, le rappel quantifie le nombre de mots que le système devrait traduire correctement (Ibidem).

Précision et rappel peuvent être combinés dans la *f-measure*, qui calcule la moyenne entre ces deux mesures (Koehn, 2010, p. 224) :

⁵ <https://www.nist.gov/sites/default/files/documents/itl/iad/mig/NISTMetricsMaTr10EvalPlan.pdf> (dernière consultation : 27/06/18)

$$f\text{-measure} = \frac{\text{précision} * \text{rappel}}{(\text{précision} + \text{rappel})/2}$$

Après avoir exposé les objectifs ainsi que le fonctionnement de base de l'évaluation automatique, nous passons à la présentation des mesures adoptées dans notre étude.

3.2.1 Bilingual Evaluation Understudy (BLEU)

D'après (Snover et al., 2006, p. 224), « BLEU calculates the score of a translation by measuring the number of n-grams, of varying length, of the system output that occur within the set of references. »

D'après (Kit Chunyu & Tak-Ming, 2015, p. 226), la logique derrière le calcul du score BLEU est « the closer a machine translation is to a professional human translation, the better it is ». Cette mesure calcule le nombre de n-grammes (qui correspondent, comme nous l'avons vu, à des séquences de n mots) qui occurrent dans la cible et dans une ou plusieurs traductions de référence.

Logiquement, nous pouvons déduire que le plus une phrase sera brève, d'autant plus élevé sera le score qu'elle recevra, ce qui risque de donner lieu à une évaluation non objective. Afin de ne pas biaiser l'évaluation, les chercheurs ont introduit une pénalité par la brièveté (*brevity penalty*), qui va réduire le score dans le cas où la traduction cible est trop courte (Koehn, 2010, p. 226).

Cette méthode présente un double aspect innovant ; en effet, si d'un côté elle permet de prendre en compte plusieurs références au même temps, de l'autre, elle considère aussi l'ordre des mots, ce qui assure une majeure objectivité dans le calcul de la précision.

3.2.2 Translation Edit Rate (TER)

D'après (Snover et al., 2006, p. 225), « Translation Edit Rate (TER) measures the amount of editing that a human would have to perform to change a system output so it exactly matches a reference translation ».

Concrètement, TER est calculé comme suit (Ibidem) :

$$TER = \frac{\# \text{ d'éditations}}{\# \text{ moyen de mots dans la référence}}$$

La mesure TER est fondée sur la quantification de la distance d'édition entre deux séquences de mots, ce qui correspond au nombre minimal d'opérations nécessaires pour transformer une séquence de mots dans une autre (Kit Chunyu & Tak-Ming, 2015, p. 229).

La grande différence par rapport aux méthodes précédentes réside dans un changement de perspective. Avant, la supposition à la base de toute mesure était qu'une traduction automatique contient inévitablement des défauts ; par conséquent, l'objectif primaire était de réduire le taux d'erreur le plus possible afin de parvenir à une traduction impeccable. En effet, la question sous-jacente à la création des mesures était *How good is machine translation ?* (Koehn, 2010, p. 237).

Mais, d'après (Ibidem), au lieu de juger la qualité de la traduction sortante en termes absolus, il faudrait plutôt le considérer en relation à la finalité à laquelle la traduction est destinée. La question qu'il faudrait se poser serait donc *Is machine translation good enough ?* : est-ce que la traduction automatique remplit le but auquel elle est destinée ?

Cette nouvelle perspective considère la qualité de la traduction dans une optique plus restreinte et premièrement pragmatique ; par conséquent, « en baissant la barre » des attentes, cet aspect innovant nous permet de valoriser encore plus la traduction automatique d'un système.

3.2.3 Human-targeted Translation Edit Rate (HTER)

L'une des variantes de la mesure TER est HTER (Human-targeted Translation Edit Rate). D'après (Snover et al., 2006, p. 226),

[f]inding only the minimum number of edits, without generating a new reference, is the measure defined as TER; finding the minimum of edits to a new *targeted reference* is defined as human targeted TER (HTER).

À la différence des méthodes précédentes, HTER évalue la distance d'édition entre l'hypothèse (un segment traduit automatiquement) et un segment post-édité par le traducteur humain (qui devient la nouvelle traduction de référence), en calculant le nombre minimal d'insertions, substitutions, suppressions et déplacements pour transformer la traduction automatique dans une phrase acceptable (Koehn, 2010, p. 238). Aussi, d'après les résultats de l'étude conduite par (Snover et al., 2006), HTER montre un niveau de corrélation avec les jugements humains relativement à la qualité de la traduction automatique meilleur par rapport à BLEU, même dans le cas où ce dernier est calculé sur des références générées par l'être humain (*human-targeted*). De plus, « HTER may represent a method of capturing human judgments about translation quality without the need for noisy subjective judgments » (Snover et al., 2006, p. 231). Dans le cadre de notre étude, nous allons utiliser cette mesure pour évaluer la qualité de la traduction post-éditée, afin de déterminer la rentabilité du flux de travail TA + PE.

Dans la même optique, il est intéressant de citer l'article de (Callison-Burch et al., 2007). Dans l'étude, les chercheurs ont évalué la qualité de la traduction de différents systèmes de traduction automatique pour huit paires de langues (français, allemand, tchèque et espagnol vers l'anglais et vice versa), à l'aide de mesures humaines et automatiques. Le matériel exploité pour la traduction et l'évaluation était un corpus de textes issus de Europarl, le corpus des débats parlementaires européens. L'objectif de l'étude était double ; d'un côté, d'exploiter les jugements humain – considérés comme la référence, en raison de leur fiabilité – pour évaluer les mesures automatiques, et de l'autre, d'examiner la corrélation entre ces deux typologies de mesures. Un autre objectif important était d'examiner l'impact du type de mesures adoptées dans l'évaluation des systèmes de traduction automatique. En effet, avant, BLEU était la mesure la plus utilisée, en raison de son haut niveau de corrélation avec les jugements humains ; cependant, d'après les auteurs, les résultats des travaux les plus récentes suggèrent que le niveau corrélation de BLEU avec les jugements humains n'est pas si haut comme nous l'avions cru auparavant. En particulier, d'après (Koehn & Monz, 2006, p. 104) cette mesure sous-estimerait systématiquement les systèmes de traduction automatique linguistiques.

Les mesures automatiques examinées dans l'études étaient les suivantes : *Semantic role overlap*, *Dependency overlap*, *ParaEval-Recall*, *Para-Eval-Precision*, *METEOR*, *BLEU*, *TER*, *GTM*, *WER*, *Maximum correlation training sur l'adéquation* et *Maximum correlation training sur la fluidité*.

L'une des questions fondamentales de l'étude était : quelle mesure d'évaluation automatique montre le plus haut niveau de corrélation avec les jugements humains relativement à la qualité de la traduction automatique ? Pour répondre à cette question, d'abord, les chercheurs ont effectué une vaste évaluation humaine : celle-ci leur a permis, d'un côté, de classer les différents systèmes selon leur performance, et de l'autre, de réaliser une analyse approfondie du processus d'évaluation. Ensuite, ils ont examiné le minutage, ainsi que l'accord inter- et intra-annotateurs pour trois méthodes d'évaluation humaine : le classement des phrases selon les critères de « fluidité » et « adéquation », le classement des phrases traduites entre elles, et finalement, le classement des « constituants », c'est-à-dire, des morceaux de phrases. Finalement, les chercheurs ont examiné la corrélation entre les mesures automatiques et les jugements humains. Pour cela, ils ont utilisé le coefficient de corrélation de Spearman ρ . D'après les résultats, les trois mesures moins fréquemment exploitées (*Semantic role overlap*, *ParaEval-recall* et *METEOR*) montrent le niveau de corrélation avec les jugements humains le plus élevé (à la différence de BLEU, avant considéré le meilleur indicateur). En conclusion, les résultats suggèrent que le choix des mesures à adopter a indéniablement un impact sur la comparaison des systèmes. Similairement, dans

notre étude, nous découvrirons que les mesures adoptées considérées individuellement ne reflètent pas la qualité réelle (et donc, l'exploitabilité) de la traduction automatique.

Nous avons terminé la section avec la description des mesures d'évaluation automatique adoptées dans notre étude. Maintenant, nous allons aborder la thématique de la post-édition.

Après une brève introduction relative à son origine et son développement, nous passerons à la description détaillée de cette méthode, avec une attention particulière à son utilité et à sa finalité en relation à notre expérience.

4. La post-édition

Actuellement, nous observons un intérêt toujours croissant envers la post-édition dans l'industrie de la traduction et de la localisation : d'un côté, elle est en train de s'affirmer et de se consolider comme activité intégrée à la traduction automatique – à tel point qu'elles sont devenues presque indissociables dans le flux de travail ; de l'autre, comme nous le verrons dans les sections suivantes, elle représente une méthode pour évaluer la qualité de la traduction automatique en termes de distance de post-édition entre la traduction post-éditée et la traduction de référence, à l'aide de mesures automatiques. Dans ce chapitre, nous allons examiner ces deux aspects, et nous chercherons à justifier l'intérêt de son adoption.

D'après (Koponen, 2015, p. 2),

[a]dvances of machine translation technology have in recent years increased its use in various contexts. In the translation industry, work processes involving the use of machine translated texts as a raw translation to be post-edited by a translator are becoming increasingly common.

Grâce au progrès incessant dans le domaine de la traduction automatique et à l'explosion du marché de la traduction, un nombre toujours croissant d'agences et de sociétés de l'industrie de la traduction et de la localisation se tournent vers les possibilités offertes par l'intégration de la post-édition au flux de travail. Plusieurs études montrent l'impact positif de cette combinaison sur la productivité. Dans les prochaines sections, nous allons explorer les motivations à l'origine de ce tournant.

La description de l'activité sera effectuée en relation avec l'objectif de notre travail : démontrer l'efficacité et la rentabilité de l'intégration de la traduction automatique avec post-édition dans le flux de travail d'un traducteur indépendant qui utilise un système de mémoire de traduction. La tâche occupe une place prépondérante dans notre expérience : comme nous l'avons déjà mentionné plus haut, nous allons demander à nos participants de post-éditer (ou retraduire) des segments pré-traduits par un logiciel de traduction automatique. Nous allons ensuite comparer la traduction finale (brute et post-éditée) avec une traduction de référence pour calculer l'effort de post-édition et le niveau d'exploitation de la suggestion de traduction automatique par nos participants. Finalement, la post-édition nous aidera à évaluer l'intérêt de l'intégration de la traduction automatique dans le flux de travail du traducteur.

Dans les prochaines sections, nous allons retracer l'évolution et l'intérêt de cette activité, ainsi que la perception de la part des utilisateurs, en relation avec l'expérience du présent mémoire. Après avoir donné

quelques définitions préliminaires, nous passerons à l'analyse de son utilité et de son champ d'application ; pour cela, nous examinerons l'état de l'art, ainsi que les pistes de recherche actuelles dans le domaine. Ensuite, nous présenterons l'évolution de la post-édition : nous nous pencherons sur son développement et plus spécifiquement sur la relation avec le progrès de la traduction automatique. En outre, nous discuterons chaque aspect de la discipline ayant un lien avec notre travail. À cet effet, nous ferons appel soit à la littérature pertinente, soit aux résultats des études préalables en la matière. Nous passerons ensuite à la description des degrés de post-édition existants, par rapport à la finalité et à l'usage auxquels le texte cible est destiné. Nous effectuerons également un bref panorama des principaux avantages et inconvénients liés à son usage. Finalement, nous conclurons ce chapitre avec une réflexion sur les approches et la perception de la tâche de post-édition de la part des traducteurs, et nous essayerons de tracer le profil du post-éditeur.

4.1 Définition, utilité et champs d'application

Le terme « post-édition » désigne la correction effectuée par un être humain d'un texte traduit automatiquement (Allen, 2003, p. 297). D'après (Veale et Way, 1997, dans Allen, 2003, p. 297), la post-édition est « the term used for correction of machine translation output by human linguists/editors ». (Wagner, 1985, dans Allen, 2003, p. 297) donne aussi une définition intéressante, qui rend compte de la différence entre cette tâche et la traduction humaine « traditionnelle » : « post-editing entails correction of a pre-translated text rather than translation 'from scratch' ». Deux éléments clés concourent à la définition de la tâche, ainsi qu'à son exécution : d'un côté, la traduction automatique, qui constitue l'élément sur lequel l'activité est exercée, et de l'autre, l'exécuteur de cette tâche, c'est-à-dire, l'être humain. En effet, contrairement à la révision, qui est la correction d'un texte traduit par un être humain, la post-édition est toujours effectuée à partir d'une traduction automatique. Par conséquent, le type d'erreurs à corriger sera inévitablement différent. Comme nous le verrons, cet aspect aura un impact sur la perception de la tâche de la part du post-éditeur.

Concernant la différence entre post-édition et révision, il est aussi intéressant de citer la définition de (Vasconcellos, 1987a, dans Krings & Koby, 2001, p. 7) :

[w]ith MT postediting, the focus is on adjusting the machine output so that it reflects as accurately as possible the meaning of the original text. In other words, the emphasis is on an ongoing exercise of adjusting relatively predictable difficulties, rather than on the discovery of any inadvertent lapse or error. The

passages that clearly require corrections, though many of them are minor and local, are more frequent than in traditional revision.

La différence entre les deux activités réside essentiellement dans la fréquence, le type et la répétitivité des erreurs. Alors qu'un traducteur humain, par exemple, peut mal traduire un mot une seule fois, et garder la cohérence dans le reste du texte, le logiciel de traduction automatique répétera systématiquement l'erreur, en particulier dans le cas des architectures linguistiques. Par conséquent, souvent le post-éditeur se retrouvera à corriger le même type d'erreurs encore et encore dans le texte, ce qui peut provoquer un certain degré d'irritation. Inévitablement, la tâche de la post-édition est parfois perçue comme une activité fatigante et frustrante. D'après (Koponen, 2012, p. 181),

[w]hile high quality MT may increase productivity, post-editing poor translations can be a frustrating task which requires more effort than translating from scratch. For this reason, estimating whether machine translations are of sufficient quality to be used for post-editing and finding means to reduce post-editing effort are an important field of study.

Nous allons introduire ici un concept important : la relation entre la qualité de la traduction automatique et l'effort de post-édition. Ce concept nous servira de base pour nos prochaines réflexions, notamment pour retracer l'évolution de la post-édition en rapport avec le progrès dans la TA.

Il est aussi important de préciser que, normalement, le niveau de qualité attendu par l'utilisateur final est souvent inférieur par rapport à celui attendu par le post-éditeur. D'après (W. John Hutchins, 1986, dans Krings & Koby, 2001, p. 15),

[i]nitially, post-editors tend to be overzealous in their correction of MT texts, changing not only grammatical and lexical mistakes but also altering the style of texts. They have to learn to edit without complete rewriting.

Les post-éditeurs débutants sont parfois excessivement zélés dans leur activité et ont tendance à hyper corriger. Comme nous le spécifierons plus bas, une formation adéquate ainsi qu'un bon niveau d'expérience constituent deux prérequis indispensables pour un post-éditeur.

La post-édition n'est pas une activité « standardisée ». À l'heure actuelle, il existe plusieurs niveaux de post-édition applicables. Comme nous le verrons par la suite dans le détail, le choix est déterminé par deux

raisons essentielles : d'un côté, la qualité de la traduction automatique sortante, et de l'autre, la finalité à laquelle le texte est destiné.

Le concept qui fait le lien entre la qualité souhaitée dans le texte final et le degré de post-édition est l'effort de post-édition. Concrètement, ce terme désigne l'effort investi par l'être humain dans la correction d'un texte traduit automatiquement (Krings & Koby, 2001, p. 179). La mesure de l'effort de post-édition représente un indice important de la qualité de la traduction automatique : intuitivement, plus le texte à post-éditer est bon, moins l'effort humain investi à la post-édition sera grand. D'après (Krings & Koby, 2001, p. 178), « the question of post-editing effort is the key issue in the evaluation of the practicality of machine translation systems ». L'auteur mentionne trois niveaux pour mesurer l'effort de post-édition : effort temporel, effort cognitif et effort technique. Plus précisément, l'effort temporel indique le temps requis pour post-éditer une phrase ; l'effort cognitif représente le type et l'étendue des procédés cognitifs mis en œuvre lors de la correction et finalement, l'effort technique désigne le nombre d'éditations (insertions, substitutions ou réordonnements) effectuées au cours de la tâche. Si les aspects temporels et techniques sont plus facilement calculables, il existe encore beaucoup de divergences sur une méthode d'évaluation objective et universelle de l'effort cognitif.

Comme nous l'avons déjà mentionné, il existe un lien profond entre qualité de la traduction automatique et post-édition. Plusieurs études démontrent que la qualité de la TA peut influencer la rapidité de post-édition ; cette dernière constitue, en effet, un facteur important pour déterminer l'effort de post-édition, et pourtant il n'est pas le seul. D'autres facteurs influencent l'activité, notamment, comme nous l'avons déjà mentionné, l'effort cognitif et technique. Cependant, d'après (Krings & Koby, 2001, p. 179), le facteur déterminant est représenté par l'effort cognitif, résultant de la combinaison de l'effort temporel et technique. Pour cela, d'après (Ibidem),

[t]he type and extent of cognitive processes triggered by the post-editing task must be defined qualitatively and quantitatively, and correlated to the corresponding deficiencies in machine translations as triggering factor.

Aux fins d'une estimation objective de l'effort de post-édition, il faudrait d'abord déterminer le type et l'étendue des processus cognitifs impliqués dans l'exécution de la tâche, et ensuite établir une corrélation entre ces processus et les déficiences dans la traduction automatique. Finalement, la mise au point de

techniques visant la réduction de l'effort cognitif aurait également comme conséquence de limiter la frustration associée à la tâche ressentie par le post-éditeur (Koponen, 2012, p. 181).

Concernant la relation entre effort cognitif et qualité de la traduction automatique, il est intéressant de citer l'étude conduite par (Koponen et al., 2012). Comme l'expliquent les auteurs, la post-édition représente actuellement l'une des méthodes pour évaluer la qualité de la traduction automatique ; celle-ci est calculée en termes de distance de post-édition entre une traduction post-éditée et une référence à l'aide de mesures automatiques. Cependant, comme le reprochent les auteurs, si ces mesures représentent de très bons indicateurs de l'effort technique en ce qui concerne le nombre de frappes, elles ne rendent toutefois pas compte de l'effort cognitif réel demandé au post-éditeur pour la tâche. Aussi, d'après (Koponen, 2012, non énuméré), la mesure semi-automatique HTER n'est pas toujours en corrélation avec la perception du traducteur de l'effort de post-édition. L'effort cognitif en particulier peut varier en fonction de plusieurs facteurs, notamment le type d'erreurs et le contexte. En définitive, les auteurs suggèrent que l'effort cognitif ne serait pas toujours directement proportionnel à l'effort technique : concrètement, certaines erreurs peuvent demander beaucoup d'effort mental, et pourtant elles ne comportent que très peu d'éditations au niveau technique. Pour cela, les auteurs proposent d'utiliser le temps comme mesure de l'effort cognitif, et explorent les possibles corrélations entre les deux indicateurs. L'objectif du travail était donc de déterminer certaines catégories d'erreurs dans la TA, et d'établir une corrélation entre ces dernières et les différents niveaux de difficulté dans le traitement, en ce qui concerne le temps requis pour la correction. D'après les résultats, le type d'erreurs a effectivement un impact sur le temps de post-édition : les segments qui comportaient des erreurs classées comme « faciles » à traiter (donc moins dispendieux du point de vue cognitif) étaient associés à des temps de correction plus brefs, et vice-versa.

L'une des méthodes permettant de faciliter le traitement du texte source par le logiciel de TA et par conséquent d'améliorer la qualité de la traduction est l'adoption des règles d'un langage contrôlé dans la rédaction des textes. (O'Brien, 2004) explore les possibles corrélations entre traduisibilité et effort de post-édition. Afin d'évaluer la traduisibilité d'un texte, l'auteur identifie des « indicateurs de traduisibilité » (*translatability indicators* ou simplement *TIs*), qui sont définis comme « linguistic features known to be problematic for MT ». Concrètement, les indicateurs de traduisibilité consistent en des constructions difficiles à traiter par le logiciel de traduction automatique ; quelques exemples de *TIs* sont, dans le cas de la langue anglaise, la forme passive, le gérondif, les ellipses, ou encore les longues séquences de substantifs. D'après les résultats, l'effort temporel demandé pour post-éditer les phrases qui présentent

le moins de *TIs* est en général inférieur à celui investi pour traiter celles qui présentent le plus de *TIs*. En conclusion, l'application de règles de langage contrôlé améliore la traduisibilité du texte et réduit effectivement l'effort de post-édition. Nous anticipons ici que les textes que nous traiterons dans le travail seront vraisemblablement rédigés en langage contrôlé.

Comme nous l'avons déjà mentionné à la section 3.2, la qualité de la TA peut être évaluée grâce à des mesures automatiques. Cependant, les scores résultant de ces évaluations sont difficiles à interpréter de la part des chercheurs ou des traducteurs chargés de post-éditer la traduction sortante du logiciel. De plus, comme nous l'avons déjà mentionné, ils ne rendent pas compte de l'effort cognitif effectif demandé par la tâche. (O'Brien, 2011) explore les corrélations entre deux mesures d'évaluation de la traduction automatique (GTM ou General Text Matcher et TER) et la productivité de traducteurs professionnels dans la post-édition. Dans cette étude, la productivité est déterminée par la rapidité de post-édition, ainsi que par l'effort cognitif ; ce dernier est calculé comme le temps de fixation moyen, et est mesuré à l'aide d'un outil de suivi du regard (*eye-tracking*).

Comme nous l'avons déjà expliqué, TER mesure le nombre d'édérations nécessaires pour convertir une traduction « brute » en la traduction de référence. Le choix de son adoption dans l'étude se justifie par le fait que, à la différence d'autres mesures automatiques, TER ne requiert que peu de références pour être en corrélation avec les jugements humains (Snover et al., 2006, p. 230).

GTM évalue la similarité entre la traduction « brute » et la traduction de référence à travers les mesures de précision, rappel et leur f-measure composite (ou moyenne harmonique). Le choix de son adoption dans l'étude se justifie par un certain nombre de raisons : d'après les résultats d'études précédentes, il a obtenu le plus haut niveau de corrélation avec la rapidité de post-édition par rapport à d'autres mesures (Tatsumi, 2009, non énuméré), ainsi que le plus haut niveau de corrélation avec les jugements humains relativement aux critères de fluidité et d'adéquation (Turian et al., 2003, non énuméré).

Les résultats montrent l'existence d'une corrélation entre les deux mesures, soit pour la rapidité de post-édition, soit pour le temps de fixation moyen. En ce qui concerne en particulier la rapidité de post-édition, les segments qui ont obtenu le plus haut niveau de GTM et TER sont aussi ceux qui demandent le moins d'effort ; quant au temps moyen de fixation, les segments qui ont obtenu le plus bas niveau de GTM requièrent le temps de fixation le plus élevé.

D'après les résultats, de manière générale, la post-édition des segments qui ont obtenu les plus hauts scores de GTM et TER requiert moins d'effort cognitif et temporel, par rapport aux segments qui

présentent des scores de niveau moyen ou bas. En conclusion, nous pouvons établir une corrélation entre ces mesures automatiques et la productivité de post-édition.

Concernant l'impact de la qualité de la TA sur la productivité en post-édition, il est intéressant de citer l'étude conduite par (Koehn & Germann, 2014). L'objectif était d'examiner la façon dont le comportement et la productivité des post-éditeurs sont influencés par la qualité du système de traduction automatique sous-jacent. Dans l'étude, la productivité est calculée comme le temps requis pour traiter chaque segment. Les auteurs analysent les traductions finales post-éditées de quatre logiciels de TA statistique, ainsi que les données de l'activité des post-éditeurs. D'après les résultats, le logiciel le plus performant a permis d'obtenir un gain de productivité de 20%. Les auteurs en concluent que l'amélioration de la qualité de la traduction automatique se traduirait effectivement par une hausse de la productivité :

(...) [w]e find that better machine translation leads to less time spent on editing, but more importantly, less time spent of figuring out harder translation problems
(...)(Koehn & Germann, 2014, p. 45).

Un autre résultat significatif est qu'il existe une variation plus importante entre les post-éditeurs qu'entre le gain de productivité déterminé par les quatre systèmes.

Concernant les différences dans la qualité de la traduction automatique post-éditée et la traduction humaine, il est intéressant de citer deux travaux. Le premier, qui a été conduit par (Fiederer & O'Brien, 2009), part de la question suivante : est-ce que la qualité de la traduction automatique doit être nécessairement inférieure à celle de la traduction humaine ? Pour répondre à cette interrogation, les chercheurs ont demandé à onze annotateurs qualifiés d'évaluer 30 segments source, 3 segments post-édités et 3 segments traduits en suivant des critères de clarté, de précision et de style. Le matériel de l'expérience était un texte technique (un guide de l'utilisateur) traduit de l'anglais vers l'allemand. D'après les résultats, les segments traduits automatiquement et post-édités ont obtenu le meilleur score au niveau des critères de clarté et de précision, tandis que les segments traduits par l'humain ont été jugés meilleurs du point de vue du style. De plus, de manière générale, les segments rédigés en langage contrôlé ont obtenu le meilleur score pour les critères de clarté et de précision – ce qui confirme une fois de plus l'impact positif de la rédaction en LC sur la qualité de la traduction automatique. Finalement, lorsqu'il leur a été demandé de choisir leur segment « préféré », la majorité des participants a choisi un segment traduit par l'être humain.

La deuxième étude a été conduite par (O'Curran, 2014). Sur la base des résultats obtenus par (Fiederer & O'Brien, 2009), l'auteur analyse la qualité au niveau linguistique des résultats d'un test sur la productivité de la post-édition, effectué sur des textes issus du domaine technique. Ce test contient soit des segments traduits automatiquement et ensuite post-édités, soit des segments traduits par l'être humain. D'après les résultats, de manière générale, les segments traduits par l'humain contiennent plus d'erreurs par mot que les segments traduits automatiquement. Cette étude confirme les résultats de travaux précédents, d'après lesquels moins d'erreurs avaient été repérées dans les segments post-édités que dans les traductions humaines. Par conséquent, elle démontre encore une fois l'intérêt de l'intégration de la traduction automatique avec post-édition dans le flux de travail, ainsi que l'effet positif sur la qualité, en ce qui concerne les textes techniques.

Finalement, concernant l'impact sur la qualité finale, il est intéressant de citer l'étude conduite par (Guerberof Arenas, 2009). Dans le travail, l'auteur compare la qualité finale résultant de la post-édition de segments issus d'une mémoire de traduction – des correspondances floues à 80-90% – par rapport à la qualité de segments traduits automatiquement et post-édités. D'après les résultats, la qualité du produit résultant de l'activité des traducteurs est supérieure pour les segments traduits automatiquement par rapport aux correspondances floues issues de la mémoire de traduction.

Dans cette section, nous avons décrit la tâche de la post-édition, en nous basant soit sur la littérature existante en la matière, soit sur les études et les travaux pertinents, en relation avec l'objectif de notre travail.

Dans la prochaine section, nous passerons à la description de l'évolution de l'activité de post-édition. Nous chercherons à définir son importance dans l'industrie de la traduction et de la localisation, ainsi que la place qui lui est accordée dans le flux de travail d'un traducteur. Notre objectif est de démontrer l'intérêt de l'intégration de la TA + PE dans le flux de travail d'un traducteur.

4.2 Évolution de la post-édition

D'après (Allen, 2003, p. 297), l'exploitation de la post-édition est strictement liée à l'adoption et à l'implémentation de techniques de traduction automatique dans les flux de travail. Depuis une vingtaine d'années, plusieurs facteurs ont contribué à l'intégration de la TA dans les marchés de la traduction et de la localisation, premièrement la pression exercée par la mondialisation et le multilinguisme. Avec l'expansion de leurs activités dans les quatre coins du monde, de grandes et petites entreprises ont ressenti la nécessité d'adopter une approche multilingue dans leurs méthodes de communication. Par

conséquent, elles se sont tournées vers les possibilités offertes par le domaine des technologies de la traduction, à la recherche de méthodes plus rentables et efficaces. Comme l'explique (O'Brien, 2011, p. 1),

[r]ecent advances in machine translation technology have led to an increased implementation by organisations with large translation volumes and a broad range of target language requirements. Consequently, technical translators who work as in-house translators or, more commonly, as freelance translators for these organisations are increasingly asked to post-edit, as opposed to translate or revise human translations that are recycled through translation memory systems.

Les objectifs de cette intégration étaient, d'un côté, de maximiser la production et d'optimiser les communications entre les différents acteurs, et de l'autre, de limiter les coûts, garder les prix avantageux, ainsi que d'assurer la diffusion de la documentation technique au même moment dans de différentes parties du monde et, par conséquent, de la traduire dans plusieurs langues (Allen, 2003, p. 298). Par ces motifs, l'intégration de la traduction automatique et la post-édition du contenu représentait une solution rentable et convenante.

Un autre facteur clé qui a contribué à ce tournant dans le flux de travail est représenté par le changement dans les attentes relatives à la qualité de la traduction. Avant, l'objectif de la traduction était presque toujours la production d'un texte de haute qualité, impeccable au niveau syntaxique, grammatical et terminologique, ainsi que potentiellement destiné à la publication. Maintenant, la mondialisation a mis l'accent sur la rapidité des échanges des informations. Par conséquent, au lieu d'une qualité « parfaite », nous recherchons plutôt une traduction adéquate au but auquel le texte est destiné. Dans ce sens, l'assimilation (*gisting*) représente un phénomène intéressant : il s'agit d'une traduction automatique sommaire dont la finalité est de comprendre l'essence d'un texte, en laissant de côté les détails et les tournures stylistiques. L'une des premières réalisations des outils pour l'assimilation était la création du portail *babelfish* en 1997. Il s'agissait d'un logiciel de traduction automatique disponible gratuitement sur Internet, capable de fournir des traductions instantanées dans plusieurs langues, et finalement, de garantir l'accès à un contenu à un utilisateur pour lequel la langue de rédaction est inconnue.

Concernant le futur de la traduction automatique et de la post-édition, d'après (Krings & Koby, 2001, p. 10), « machine translation will continue to be used, and post-editing will continue to be a need » : comme nous l'avons déjà mentionné, un concept clé à retenir est le lien profond entre l'adoption de la traduction

automatique et la diffusion de la post-édition ; ce lien se traduit inévitablement par une relation d'interdépendance dans le flux de travail.

En effet, d'après (Daems et al., 2015, p. 31),

[i]n recent years, machine translation (MT) and its subsequent post-editing have become more widely accepted in the translation industry. Especially when it comes to technical texts, machine translation has proven its worth, with companies like Autodesk reporting on productivity increases in comparison with human translation ranging from 20 to 131%, depending on the language combination and translator.

Comme nous le verrons, les avantages dérivants de l'intégration de ces deux fonctionnalités sont connus dans le marché de la traduction, spécifiquement en ce qui concerne l'impact sur la productivité. À ce propos, il est intéressant de citer l'étude conduite par (Zhechev, 2012) au sein de *Autodesk*⁶, société créatrice de logiciels basée en Suisse. Le but du travail était d'évaluer les effets de l'intégration de la traduction automatique dans le flux de travail, ainsi que de tester l'impact sur la productivité dérivant de cette combinaison. L'étude calcule les gains de productivité dérivant de la post-édition de segments traduits automatiquement versus des segments traduits par l'humain. D'après les résultats, il existe une corrélation significative entre le nombre d'éditations apportées par les traducteurs à la traduction « brute » - la distance d'édition - et le gain de productivité, par rapport à une traduction « de zéro ».

En conclusion, nous résumons les pistes de recherches actuelles en post-édition et traduction automatique :

[t]he main goal of postediting research is no longer finding out whether or not post-editing can be used, but rather finding out when it cannot be used, and how machine translation systems can be improved to better suit post-editors' needs (Daems et al., 2015, p. 31).

Comme nous pouvons le déduire, actuellement, les recherches en post-édition procèdent en parallèle avec celles en traduction automatique. En effet, nous pouvons observer un dialogue constant entre les deux disciplines, qui reflète leur lien dans le flux de travail : l'amélioration de l'une est toujours associée au progrès et à la facilitation de l'autre. Cet aspect revêt une importance centrale dans notre expérience.

⁶ <https://www.autodesk.ch/fr> (dernière consultation : 16/04/18)

Nous passons maintenant à la description de différents niveaux de post-édition existants.

4.3 Degrés de post-édition

Comme nous l'avons déjà mentionné, il existe plusieurs niveaux de post-éditions applicables. Le choix du degré à adopter dépend de plusieurs facteurs. Nous pouvons citer, entre autres, la durée de vie du texte, le volume de la documentation à traiter, le type d'informations, les attentes relatives à la qualité souhaitée, ainsi que la finalité du texte source (Allen, 2003, p. 298). Cependant, d'après (Krings & Kobayashi, 2001, p. 9), « [t]he factor that ultimately determines the quality level up to which raw machine translation output should be edited is the user's need ». Dans l'ensemble, comme confirmé par (Loffler-Laurian, 1994, p. 20), le principal critère pour déterminer le niveau de post-édition est la finalité de la traduction. Intuitivement, un texte qui servira uniquement pour le *gisting* requerra un niveau de post-édition nul, ou du moins inférieur à celui demandé par un texte destiné à la publication.

Dans la catégorisation des niveaux de post-édition existants, nous allons utiliser la classification proposée par (Allen, 2003, p. 300). Selon cette classification, le degré de post-édition est lié aux différentes approches de la TA, ainsi qu'à la finalité de l'intégration d'un système de traduction automatique.

Il existe principalement deux approches à l'utilisation des systèmes de traduction automatique : l'une est dite « traduction pour l'assimilation » (*translation for assimilation*) ; l'autre est dite « traduction pour la communication » (*translation for communication*) (Allen, 2003, p. 301).

La première se définit comme traduction pour la compréhension (*translation for understanding*) ; elle comprend les techniques finalisées à l'extraction du contenu sémantique essentiel d'un texte, et fait abstraction des détails et tournures stylistiques.

L'autre approche disponible est la post-édition dite « rapide ». Elle est surtout appliquée aux textes destinés à une circulation interne (non destinés à la publication) et prévoit l'application d'un taux de correction minimal, limité essentiellement aux erreurs « grossières », afin de produire un document compréhensible et informatif.

Passons maintenant à l'approche la plus fréquente, c'est-à-dire, la traduction finalisée à la transmission et à la diffusion d'une information à grande échelle (*translation for dissemination*). Dans la plupart des cas, ce type d'approche entraîne la production d'un contenu destiné à la publication ; par conséquent, un certain degré de post-édition sera toujours requis.

Dans le cas de la traduction pour la diffusion, il existe principalement deux niveaux de post-édition applicables en fonction de la qualité souhaitée ; d'un côté, nous avons la post-édition dite « minimale » (*minimal post-editing*, en anglais) (Allen, 2003, p. 304), et de l'autre, celle dite « complète » (en anglais *full post-editing*) (Allen, 2003, p. 306).

Une traduction destinée à la publication qui ne requiert aucun degré de post-édition était avant presque inexistante ; l'un de très rares cas était le projet déjà cité *TAUM-METEO*. Cependant, grâce au progrès de la traduction automatique, la qualité de la traduction « brute » est toujours plus élevée, ce qui entraîne un besoin de post-édition progressivement inférieur.

L'approche de la post-édition la plus commune est celle dite « complète ». Comme le nom le suggère, elle consiste en la correction intégrale d'un texte traduit automatiquement ; le résultat doit être une traduction de haute qualité, comparable à une traduction humaine, et par conséquent apte à la publication. Le problème réside ici dans la question de la rentabilité et dans la vitesse d'exécution de l'activité ; parfois, la retraduction complète d'un texte peut s'avérer plus rapide que la post-édition d'un texte traduit automatiquement. Au cours de notre expérience, nous examinerons cet aspect crucial.

La post-édition minimale prévoit l'application d'un taux de correction limité à l'essentiel pour produire un texte compréhensible et exploitable. Un exemple typique de post-édition minimale est celle applicable à un texte destiné à être transmis à un tiers, tels que des clients, des lecteurs ou des utilisateurs. Dans ce cas, puisque le texte sera destiné à la diffusion, et éventuellement à la publication, il sera soumis à un certain niveau de post-édition avant d'être envoyé.

En ce qui concerne spécifiquement cette dernière approche, l'une des difficultés réside dans la détermination objective et universellement acceptable des critères de correction minimale. En effet, comme nous le démontrerons aussi dans notre étude, le degré, le type et l'extension des corrections varient largement d'un post-éditeur à l'autre par rapport à son expérience, sa connaissance des langues source et cible ou plus simplement sa perception personnelle. Pour ces raisons, les entreprises qui ont implémenté l'activité de post-édition ont souvent mis au point des directives internes afin d'uniformiser les attentes. Cependant, à l'heure actuelle, les divergences persistent sur le niveau et la « profondeur » des corrections.

Dans le cadre de notre expérience, nous demanderons à nos traducteurs d'effectuer une post-édition complète, à l'aide des directives *TAUS*⁷. Nous présenterons le contenu de ces directives dans les instructions, que vous trouverez en annexe (Annexe II).

4.4 Avantages et inconvénients

Passons maintenant à une brève présentation des avantages et des inconvénients liés à l'adoption de la post-édition de la traduction automatique. L'un des premiers avantages est la cohérence ; d'après (McElhaney et Vasconcellos, 1988, dans Krings & Koby, 2001, p. 11),

[w]ith machine translation, the post-editor has the assurance, at the mechanical level, that nothing has been skipped or repeated, and also that there are not likely to be errors in spelling (...).

Alors que le traducteur humain peut commettre des omissions, ou des fautes d'orthographe sporadiques, la traduction automatique sera toujours cohérente dans ses traductions. De plus, si d'un côté les constructions incorrectes seront présentes soit dans la traduction humaine, soit dans la traduction automatique, le logiciel de TA sera plus enclin à commettre des fautes « locales » (Ibidem). Cet aspect s'applique surtout à l'approche linguistique.

Cependant, elle comporte aussi des inconvénients. D'après (Krings & Koby, 2001, p. 12), la constante exposition du traducteur à un langage stylistiquement et grammaticalement imparfait pourrait influencer négativement sa capacité d'identification et de correction des erreurs. En particulier, les apprenants étrangers risquent d'assimiler rapidement des tournures erronées, et de perdre la capacité de discerner les constructions standardisées de celles incorrectes. À l'heure actuelle, il est toutefois important de revoir cette réflexion à la baisse, à la lumière du progrès actuel en matière de qualité de la traduction automatique statistique et neuronale.

Après un bref panorama des avantages et des inconvénients liés à la post-édition, passons maintenant à la description de différentes approches de cette pratique.

⁷ <https://www.taus.net/academy/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines> (dernière consultation : 20/08/18)

4.5 Approches de la post-édition : le profil du post-éditeur

D'après (Krings & Koby, 2001, p. 12), le professionnel le plus apte à exercer l'activité de post-éditeur est, intuitivement, le traducteur : « only a translator can judge the accuracy of a translation ». En effet, en raison d'une connaissance approfondie des langues source et cible, le traducteur est celui qui dispose des compétences nécessaires pour détecter les erreurs de la traduction automatique et apporter les modifications nécessaires. Comme le confirment (McElhaney et Vasconcellos, 1988, dans Krings & Koby, 2001, p. 12),

[t]he translator is the one best able to pick up errors in the machine translation, (...) he has fund knowledge about the cross-language transfer of concepts, and he has technical resources at his disposal which he knows how to use in the event of doubts. Moreover, for the very reason that translators are best suited to the task, the more experienced they are, the more effective they will be.

Comme nous l'avons vu, l'expérience en post-édition revêt un rôle essentiel dans les compétences du post-éditeur : pour cette raison, depuis quelques années, plusieurs auteurs insistent sur la nécessité d'introduire une formation spécifique en post-édition dans le parcours académique du traducteur. D'après (Koponen, 2015, p. 2),

[t]he increasing use of post-editing processes also raises questions related to teaching and training of translators and post-editors, and institutions offering translator training have started to incorporate post-editing in their curricula.

De plus, plusieurs auteurs soulignent l'aspect « graduel » dans l'acquisition des compétences en post-édition ; d'après (Vasconcellos, 1987b, dans Krings & Koby, 2001, p. 16),

[p]ost-editing skills are developed gradually, and initial judgements are bound to be reversed. The level of comfort is greatly increased at the end of 100,000 words – the equivalent of a month of full-time post-editing.

Plusieurs facteurs déterminent le succès d'un post-éditeur ; certaines compétences sont également demandées par la révision « traditionnelle », notamment, des connaissances linguistiques des langues source et cible, la maîtrise du sujet/domaine de compétence, des connaissances interculturelles, ainsi que de bonnes capacité de recherche et de repérage de l'information et de la documentation (O'Brien, 2002, p. 100). En particulier, plusieurs auteurs insistent spécifiquement sur un aspect : l'attitude du traducteur-

post-éditeur envers la traduction automatique. Un certain nombre d'études confirment le rôle clé joué par l'attitude face à l'activité de la part de l'être humain, à tel point qu'elle peut influencer le temps requis pour l'exécution de la tâche. D'après (Lange, 1998, dans Krings & Koby, 2001, p. 13), à cause de l'attitude négative de la part du post-éditeur, le temps requis par la traduction automatique avec post-édition pourrait même dépasser celui demandé par la seule traduction humaine.

Comme nous l'avons déjà mentionné, acquérir des capacités techniques en post-édition demande du temps et de l'expérience. Dans l'ensemble, outre une attitude positive, d'autres prérequis sont des compétences linguistiques et techniques, ainsi qu'une bonne aptitude à la résolution de problèmes (Krings & Koby, 2001, p. 12). Comme l'expliquent (McElhaney & Vasconcellos, 1988, in Krings & Koby, 2001, p. 7),

[i]n addition to the professional experience that a translator brings to the task, it is essential that he have good keyboard skills, quick reaction to the challenges that arise in the text, and a flair for solving problems creatively.

Plus spécifiquement, les compétences techniques incluent, entre autres, une connaissance approfondie des logiciels de traitement de texte, des connaissances en rédaction en langage contrôlé, un certain degré de familiarité avec les outils d'aide à la traduction et de gestion de la terminologie, ainsi qu'une bonne maîtrise des raccourcis clavier (O'Brien, 2002, p. 100).

Dans cette perspective, il est intéressant de citer l'étude conduite par (O'Brien & de Almeida, 2010). En déplorant le manque de directives globalement adoptées en matière de post-édition, l'article explore les possibles corrélations entre le degré de performance dans la post-édition et le niveau d'expérience dans la traduction. Dans cette étude, les auteurs ont décidé de tester la combinaison du modèle d'assurance-qualité LISA QA Model et des directives de post-édition fournies par GALE (Global Autonomous Language Exploitation). Les résultats montrent que les traducteurs qui ont le plus d'expérience sont aussi les post-éditeurs les plus rapides et précis. Cependant, à la différence des post-éditeurs moins expérimentés, les traducteurs qui ont le plus d'expérience ont la tendance à apporter le plus de modifications « préférentielles », surtout au niveau stylistique – ce qui contrevient aux indications des directives. À l'inverse, les post-éditeurs « débutants » ont la tendance à apporter seulement les changements nécessaires et à se conformer davantage aux directives.

En conclusion, les auteurs mettent l'accent, encore une fois, sur l'importance d'une formation spécifique en post-édition, et encouragent son intégration dans le cadre du parcours académique du traducteur.

Concernant plus spécifiquement ce dernier aspect, il est aussi intéressant de citer l'article de (O'Brien, 2002). Comme nous l'avons déjà mentionné, depuis une dizaine d'années, l'expansion du marché de la traduction a entraîné la nécessité d'adopter de nouveaux outils d'aide à la traduction, notamment les logiciels de traduction automatique et les systèmes de mémoire de traduction. En outre, un certain nombre d'études montrent l'impact positif sur la productivité dérivant de l'intégration de la traduction automatique avec post-édition, ainsi que la viabilité de cette implémentation pour faire face à la demande toujours croissante de traductions. D'après l'auteur, l'intégration d'une formation en techniques de post-édition dans le parcours académique d'un traducteur aurait un certain nombre d'avantages : premièrement, l'attitude des traducteurs envers la traduction automatique serait plus positive. En effet,

[t]ranslators who do not have post-editing skills are frequently hostile to machine translation technology. Common arguments against MT include a dislike for correcting repetitive errors that a human translator would never make, a fear of losing language proficiency by working with poor MT output and a dislike of having one's freedom of expression limited (O'Brien, 2002, p. 100).

Une formation en post-édition rendrait également les traducteurs plus conscients des capacités de cette technologie, et leur permettrait d'être plus à l'aise avec ce type de logiciels. En définitive, une attitude moins critique envers la traduction automatique faciliterait son acceptation et son adoption dans l'environnement de travail futur.

Nous terminons la section dédiée à la post-édition. Dans ce chapitre, nous avons essayé de justifier l'importance de cette technique, ainsi que l'intérêt de son intégration dans le flux de travail d'un traducteur. Nous avons défini la tâche en nous focalisant spécifiquement sur la relation avec la traduction automatique ; nous avons présenté les différents niveaux de post-éditions disponibles, les approches à l'activité, ainsi que son utilité et son champ d'application, à l'aide des résultats d'études pertinentes. Finalement, nous avons cherché à tracer le profil du post-éditeur, en définissant les compétences requises pour cette tâche.

Maintenant, nous passons au chapitre consacré à l'expérience, dans lequel nous essayerons d'analyser la façon dont les connaissances exposées jusqu'à présent se concrétisent dans la pratique.

5. L'expérience

Après une brève introduction, nous passerons à la description détaillée de l'étude. D'abord, nous allons exposer l'intérêt et la finalité de travail inspiré de l'état de l'art dans la recherche de l'intégration de la TA avec post-édition dans la pratique professionnelle de la traduction de la section 2.5. Ensuite, nous présentons notre hypothèse, ainsi que les questions de recherche ; après, nous allons décrire le milieu expérimental, ainsi que le type et l'origine des données traitées. Nous allons ensuite tracer le profil des participants et présenter les logiciels utilisés. Finalement, nous allons conclure le chapitre avec la description de la méthodologie de travail, avec nos questions de recherche et notre hypothèse.

5.1. L'intérêt et la finalité

Comme nous l'avons déjà mentionné dans l'Introduction, au chapitre 1, la présente étude vise à évaluer l'intérêt de l'intégration d'un module de traduction automatique avec post-édition à un outil d'aide à la traduction, en l'occurrence, du système de traduction automatique fourni par *SDL Language Cloud* au système de mémoire de traduction *SDL Trados Studio*. Pour cela, nous allons évaluer la qualité de la traduction automatique (avant et après la post-édition), soit sous forme de *Baseline Engine*, soit sous forme d'*Adaptive Engine*, en prenant aussi en compte l'attitude et l'approche aux outils de la part de l'utilisateur. Notre recherche se propose d'analyser l'intérêt de l'intégration du flux de travail TA + PE relativement à la traduction de textes techniques, dans l'optique de travail d'un traducteur indépendant qui utilise un système de mémoire de traduction.

Comme nous l'avons déjà mentionné au chapitre 2, la traduction automatique fournie par *SDL Language Cloud* est de type statistique. En effet, l'un des objectifs de ce travail est d'explorer une piste alternative à l'approche neuronale. Cependant, il faut préciser que cette dernière génération de systèmes de TA est au centre des recherches actuelles dans le domaine des technologies de la traduction, tant au niveau académique, qu'au niveau scientifique : selon certains, elle représenterait actuellement l'état de l'art en matière de traduction automatique, du fait d'être parvenue à surmonter nombre d'obstacles et de difficultés traductives posés par la dernière génération de systèmes de TA (les systèmes statistiques). En effet, actuellement, un nombre toujours croissant de sociétés (y compris SDL⁸) est en train d'effectuer la transition vers cette nouvelle approche.

⁸ <https://www.sdl.com/it/about/news-media/press/2017/sdl-brings-nmt-to-its-secure-ets.html> (dernière consultation : 28/08/18)

Dans une étude pilote que nous avons effectuée préliminairement à notre expérience, nous avons testé la productivité temporelle de la fonctionnalité *Adaptive Engine* dans le cadre d'un projet de traduction afin de vérifier l'intérêt de son intégration dans le flux de travail d'un traducteur indépendant. Nous avons réalisé l'expérience à l'aide de quatre étudiants en Maîtrise en Traduction. L'étude a été organisée en deux parties : dans un premier temps, nous avons soumis aux participants un texte très répétitif à traduire sans traduction automatique ; dans un deuxième temps, environ 8 semaines après, les participants ont retraduit le même texte, cette fois à l'aide de la traduction automatique et de la fonctionnalité *Adaptive Engine*. Le texte à traduire consistait en des phrases d'une longueur entre 25 et 30 mots extraites des Rapports Annuels de la Banque Centrale Européenne de 2011 à 2016. Étant donné la complexité et la technicité du texte source, nous pouvons supposer que l'effet d'apprentissage serait théoriquement nul. De plus, les participants ont eux-mêmes déclaré n'avoir rien mémorisé de la première expérience. Nous avons enregistré le comportement des participants à l'aide du logiciel d'enregistrement d'écran *BB Flashback Express*⁹. Aussi, d'après les enregistrements vidéo, nous avons remarqué que les participants ont cherché à nouveau beaucoup de terminologie déjà traduite lors de la première expérience, ce qui justifierait encore une fois un effet d'apprentissage nul. D'après l'analyse des résultats, pour les quatre participants, nous avons pu constater un gain temporel significatif : entre 108 et 151 secondes par segment (phrase). Ce gain temporel est aussi justifié par un haut niveau d'exploitation des suggestions du logiciel de TA. Ces résultats préliminaires satisfaisants nous ont encouragé à poursuivre l'étude de cette fonctionnalité.

Dans le présent travail, nous allons évaluer l'intérêt de l'intégration d'un module de traduction automatique avec post-édition à un système de mémoire de traduction ; plus précisément, nous allons tester et comparer la performance de deux modules de traduction automatique (*Adaptive Engine* et *Baseline Engine*) en mesurant trois variables fondamentales : la qualité de la traduction (avant et après post-édition) à l'aide de mesures automatiques, le temps de post-édition, ainsi que la satisfaction de l'utilisateur. Concernant ce dernier aspect en particulier, une place importante sera accordée à l'opinion et à la perception des participants par rapport à la qualité de la traduction automatique, à la tâche de post-édition, à l'environnement de travail, ainsi qu'aux technologies testées. Le questionnaire final prévu à cet effet nous aidera à recueillir l'opinion des participants.

⁹ http://www.bbflashback.it/BBFlashBack_FreePlayer.aspx (dernière consultation : 16/03/2018)

Nous avons organisé notre travail en trois parties principales : traduction et post-édition (section 6.1.1), évaluation de la qualité de la traduction automatique (section 6.1.2), questionnaire pour les participants (section 6.1.3).

Après avoir exposé l'intérêt du travail, nous passons maintenant à la formulation de l'hypothèse et des questions de recherche.

5.2 L'hypothèse et les questions de recherche

L'hypothèse sous-jacente à notre étude est que l'intégration d'un module de traduction automatique avec post-édition dans le flux de travail d'un traducteur indépendant dans le cadre d'un système de mémoire de traduction serait rentable pour la traduction de textes techniques. À cet effet, nous allons vérifier si la traduction automatique fournie par *SDL Language Cloud* - soit sous forme de *Baseline Engine*, soit sous forme d'*Adaptive Engine* - est adéquate pour la traduction des textes techniques de l'anglais vers l'italien. Plus précisément, la fonctionnalité *Adaptive Engine* garantirait une amélioration progressive de la qualité de la traduction et augmenterait la productivité du traducteur, en réduisant au fur et à mesure l'effort de post-édition. Dans l'optique de l'étude, le travail du traducteur serait donc réduit à une simple post-édition bilingue de la traduction cible, une tâche idéalement plus rentable, plus efficace et moins fatigante par rapport à une retraduction complète de la source.

Pour vérifier notre hypothèse, nous allons mesurer un certain nombre de variables. Premièrement, nous nous focaliserons sur la performance de l'*Adaptive Engine*. Nous allons tester le degré d'apprentissage de la fonctionnalité en analysant le taux d'erreurs et le niveau d'intégration des préférences traductives de l'utilisateur de la part du système, afin de démontrer son efficacité en termes de capacité d'adaptation au style et à la terminologie du traducteur. En outre, afin de garantir une meilleure objectivité à l'étude, nous conduirons ensuite deux tests de retraduction.

Après, nous passerons au *Baseline Engine* : nous allons étudier le degré d'acceptation de la suggestion de la traduction automatique de la part de chaque participant, le taux d'éditations effectuées, ainsi que le temps de post-édition. En outre, nous allons mesurer l'impact de la fonctionnalité sur la productivité temporelle par rapport à l'*Adaptive Engine*, afin de comparer la performance des deux modules.

Ensuite, nous passerons à l'évaluation de la qualité de la TA. D'abord, nous allons analyser manuellement la traduction sortante. Puis, nous allons évaluer la qualité de la traduction « brute » du système par rapport à une référence (la traduction italienne officielle du guide utilisateur) en calculant le TER et le BLEU. Ces

données nous serviront comme point de départ pour les analyses successives. Ensuite, nous calculerons le HTER : pour chaque participant, nous comparerons la version finale post-éditée (notre nouvelle référence) avec la traduction « brute » pour déterminer l'effort de post-édition. Finalement, nous calculerons le BLEU : pour cela, nous comparerons la version « brute » avec chaque version post-éditée (la référence) afin de déterminer une possible corrélation entre perception de la qualité de la traduction automatique et effort de post-édition. Le travail se conclura avec un questionnaire.

Le présent travail est structuré autour de deux questions principales et trois sous-questions.

La première question de recherche sera : **est-ce que l'intégration de la traduction automatique avec post-édition dans le flux de travail d'un traducteur indépendant qui utilise un système de mémoire de traduction est rentable pour la traduction de textes techniques ?** Afin de répondre à cette question, premièrement, nous allons nous interroger sur la qualité de la traduction automatique sortante (la même pour le *Baseline Engine* et l'*Adaptive Engine*). Pour évaluer la rentabilité de l'intégration, nous allons étudier d'une part, la qualité de la traduction sortante, et d'autre part, le temps nécessaire à effectuer la tâche de post-édition. Nous considérons que la qualité de la traduction sortante est un indicateur de l'exploitabilité de la TA, ainsi qu'un prédicteur de la quantité de post-éditions nécessaires à produire un texte apte à diffusion. La première sous-question sera donc : **quelle est la qualité de la traduction automatique « brute » ?** À cette fin, nous allons comparer la qualité de la traduction sortante non post-éditée avec une traduction de référence pour calculer le score TER (sur lequel nous nous baserons pour déterminer la distance d'édition), ainsi que le BLEU, qui nous donnera une indication relativement à la qualité de la traduction sortante.

Ensuite, en ce qui concerne plus spécifiquement la fonctionnalité *Adaptive Engine*, nous allons nous interroger sur son effective capacité d'apprentissage et d'intégration des corrections du traducteur – ce qui théoriquement permettrait une progressive réduction de l'effort de post-édition. La deuxième question de recherche sera : **dans quelle mesure est-ce que la fonctionnalité est capable de s'adapter aux préférences stylistiques et terminologiques de l'utilisateur ?** Pour répondre à cette interrogation, nous allons examiner le nombre d'intégrations des corrections du traducteur et, inversement, le taux de répétition des erreurs. Finalement, nous allons vérifier l'intérêt de l'utilisation de cette fonctionnalité par rapport au simple *Baseline Engine*. Pour cela, nous allons comparer le temps requis en moyenne pour traduire le même texte à l'aide du *Baseline Engine* et de l'*Adaptive Engine*. Nous allons donc chercher de répondre à la deuxième sous-question : **quel module permet la meilleure productivité temporelle ?** En

répondant à cette question, nous serons en mesure d'évaluer l'efficacité de la fonctionnalité *Adaptive Engine* par rapport au *Baseline Engine*.

Ensuite, nous nous intéresserons au degré d'acceptation de la suggestion de la traduction automatique par le traducteur, ce qui représente un indicateur supplémentaire de l'intérêt de l'intégration. Pour cela, d'un côté, nous examinerons combien de fois le traducteur accepte la proposition de TA au lieu de retraduire le segment de zéro, et de l'autre, nous analyserons le nombre de segments post-édités. Finalement, nous allons comparer le texte traduit automatiquement et post-édité (notre *human-targeted reference*) avec la traduction brute et nous calculerons HTER et BLEU. La troisième sous-question sera donc : **quel est l'effort de post-édition ?** En répondant à cette question, d'un côté, nous serons en mesure de déterminer le degré de tolérance de l'effort de post-édition ; de l'autre, nous pourrions évaluer la qualité de la traduction « brute » par rapport à la traduction finale post-éditée, en prenant aussi en compte la perception et le comportement de nos participants.

Enfin, l'étude se terminera par un questionnaire construit pour nous donner une vue d'ensemble de l'attitude et de la perception des participants envers la tâche de post-édition, de leur niveau d'appréciation de la qualité de la traduction automatique et de l'environnement du travail, ainsi que de leurs connaissances préalables en matière de technologies de la traduction. Il nous permettra d'analyser une possible corrélation entre la perception de la qualité de la traduction automatique de la part des traducteurs et les résultats des mesures automatiques. Les données recueillies nous aideront à évaluer l'intérêt de l'intégration des deux technologies.

Après l'exposition de l'hypothèse du travail et des questions de recherche, passons maintenant à la description du milieu expérimental.

5.3 Le milieu expérimental

Les phases de traduction et post-édition vont se dérouler dans le système de mémoire de traduction *SDL Trados Studio 2017* (Freelance version). Pour la première partie, le participant va exploiter le module de traduction automatique personnel *Adaptive Engine* ; pour la deuxième, les participants vont utiliser uniquement la traduction automatique sous la forme du *Baseline Engine*.

Pour la première partie, nous allons soumettre à notre unique participante un texte extrait d'un guide utilisateur d'une imprimante d'une marque connue à traduire (ou post-éditer) de l'anglais vers l'italien en utilisant un module personnel de traduction automatique adaptative *Adaptive Engine*. Nous allons

enregistrer le comportement du participant à l'aide du logiciel d'enregistrement d'écran *BB Flashback Express*. Pour cette partie de l'expérience, nous allons nous intéresser à l'efficacité du logiciel. Nous allons donc analyser l'enregistrement pour vérifier le degré d'intervention de la fonctionnalité et la mémorisation des préférences du traducteur. Nous conduirons ensuite deux tests de retraduction pour vérifier la capacité de mémorisation et d'adaptation effectives aux préférences traductives de l'utilisateur de la part de l'*Adaptive Engine*, ainsi que l'apprentissage et l'intégration des corrections.

En ce qui concerne la deuxième partie, nous allons soumettre à nos participants le même texte à post-éditer (ou traduire) de l'anglais vers l'italien. Dans ce cas aussi, nous allons enregistrer le comportement de chaque participant à l'aide de *BB Flashback Express*. Pour cette partie de l'étude, cet enregistrement sera indispensable pour mémoriser les actions des participants, et nous aidera à analyser leur comportement, notamment l'approche de chacun à la proposition de traduction automatique pour chaque segment.

Dans la deuxième partie de l'étude, dédiée à l'évaluation de la traduction à l'aide de mesures automatiques TER (ainsi que sa variante HTER¹⁰) et BLEU¹¹, nous nous servons de deux logiciels gratuitement disponibles en ligne.

Après avoir exposé le milieu expérimental, nous allons présenter les données.

5.4 Les données

Pour chaque phase de traduction et de post-édition, nous avons utilisé des extraits de guides utilisateur d'imprimantes¹² de marques connues, sous la forme de fichier .doc. Autant pour les phases de traduction avec l'*Adaptive Engine* que pour les phases avec le *Baseline Engine*, nous avons utilisé la section relative à la résolution de problèmes (*Solve a problem*) qui comporte 244 segments au total. Le choix se justifie à partir d'une série de caractéristiques intrinsèques à ce type de textes. Comme nous l'avons déjà mentionné, afin de garantir une majeure objectivité à l'étude, nous avons sélectionné une typologie textuelle contenant un taux assez élevé de répétitions. En effet, la répétitivité des textes constitue un

¹⁰ <http://www.cs.umd.edu/~snover/tercom/> (dernière consultation : 24/07/18)

¹¹ <https://www.letsmt.eu/Bleu.aspx> (dernière consultation : 24/07/18)

¹² <https://support.hp.com/us-en/product/hp-deskjet-1510-all-in-one-printer-series/5157551/model/5157552/manuals> (dernière consultation : 26/07/18)

prérequis important aux fins de notre travail : d'un côté, elle assurerait des résultats plus facilement reproductibles, et de l'autre, elle faciliterait le test de l'efficacité de l'*Adaptive Engine*.

D'après une phase d'analyse effectuée à l'aide de l'outil d'analyse fourni par *SDL Trados Studio*, nous avons remarqué que cette section présente le taux de répétitions internes le plus élevé (environ 24%) par rapport aux autres sections du même document.

Similairement, dans le premier test de retraduction, nous avons extrait au hasard 50 segments issus du même document, afin de tester la mémorisation ; ensuite, dans la deuxième, nous avons extrait 40 segments de la section relative à la résolution de problèmes de documents similaires, afin de tester l'adaptation et l'apprentissage effectifs.

Les textes utilisés sont vraisemblablement rédigés en langage contrôlé (LC). Comme nous l'avons déjà mentionné à la section 4.1, la rédaction en LC présente un nombre de bénéfices pour la traduction automatique : la restriction du lexique et la simplification de la syntaxe, entre autres, facilitent le traitement du texte de la part de l'outil de traduction automatique et améliorent la qualité de la traduction sortante, en réduisant par conséquent l'effort de post-édition.

De plus, en vertu de la finalité à laquelle ils sont destinés – c'est-à-dire, fournir de l'aide à l'utilisateur dans la résolution de problèmes avec l'imprimante – ces textes sont structurés de façon très simple. Ils doivent pouvoir être compris par la majorité des utilisateurs, abstraction faite du niveau d'éducation de chacun. Aussi, ils sont conçus pour être lus pendant l'utilisation (et non pas avant) ; pour cela, chaque phrase ne contient qu'un seul concept et ne décrit qu'une seule action précise.

Après la description des données expérimentales, passons maintenant à la présentation du profil des participants.

5.5 Profil des participants

Huit étudiants en Maîtrise en Traduction, d'âge compris entre 23 et 27 ans, désormais jeunes traducteurs semi-professionnels, de langue maternelle italienne et ayant l'anglais parmi leurs langues passives dans leur combinaison linguistique ont participé à l'expérience.

D'après un petit sondage préliminaire, chacun des participants avait des connaissances de base de l'environnement *SDL Trados Studio* et de la traduction automatique, ainsi que des expériences préalables

avec la tâche de post-édition - même si, dans la plupart des cas, ces expériences étaient limitées à des exercices pratiques à des fins non professionnelles.

Le choix de ce profil pour nos participants se justifie pour un certain nombre de raisons. Premièrement, nous étions intéressée à évaluer le comportement d'un jeune traducteur, avec un niveau d'expérience professionnelle limité dans le domaine de la traduction, donc vraisemblablement plus ouvert aux nouvelles technologies et aux possibilités offertes par la traduction automatique. Un certain nombre d'études témoignent de la résistance opposée surtout par les traducteurs professionnels à la traduction automatique et à la tâche de post-édition. Comme nous l'avons vu, cette dernière est particulièrement perçue comme une activité fatigante et peu rentable. Aussi, les traducteurs qui ont le plus d'expérience manifestent parfois une attitude plus réticente à la traduction automatique. Ils ont tendance à rejeter la proposition du logiciel pour réécrire entièrement chaque segment. Les arguments qu'ils opposent le plus souvent sont relatifs à la qualité de la traduction automatique, qui est perçue comme « stylistiquement pauvre ». Maintenant, grâce aux derniers progrès dans le domaine des outils d'aide au traducteur, nous étions aussi intéressés à déterminer s'il y avait eu un éventuel changement dans la perception de l'environnement de travail et des outils de la part du traducteur, en observant l'approche ainsi que l'attitude de chaque participant envers l'activité et l'environnement de travail prévus dans l'expérience.

Deuxièmement, nous voulions simuler un contexte de travail vraisemblable : le cas d'un traducteur indépendant qui fait son entrée dans la vie professionnelle et qui traduit des textes techniques à l'aide d'un système de mémoire de traduction. Ce jeune professionnel décide de se tourner vers d'autres possibilités offertes par le domaine des technologies de la traduction – notamment les logiciels de traduction automatique – pour accélérer la production et optimiser le travail.

Après avoir présenté le profil des participants, passons maintenant à la description des logiciels utilisés pendant l'expérience.

5.6 Les logiciels

L'étude a eu lieu dans le cadre d'un projet de traduction effectué sur le système de mémoire de traduction *SDL Trados 2017*¹³ (Freelance version). Nous allons examiner la traduction automatique fournie par *SDL*

¹³ <https://www.sdltrados.com/products/trados-studio/> (dernière consultation : 06/03/2018)

*Language Cloud*¹⁴, soit sous la forme du *Baseline Engine*, soit sous la forme de l'*Adaptive Engine*¹⁵. Voyons maintenant dans les détails chaque logiciel.

SDL Trados Studio est le système de mémoire de traduction leader dans le marché. La fonctionnalité centrale de ce type de logiciel s'apparente à une base de données de textes pré-traduits et segmentés, stockés sous la forme de bi-textes. Cependant, dans le cadre de notre travail, nous ne nous intéressons pas à la fonctionnalité de mémoire de traduction mais *SDL Trados Studio* nous servira seulement comme environnement de travail pour l'intégration du module de traduction automatique fourni par *SDL Language Cloud*. Cette plateforme dans le cloud fournit plusieurs solutions de traduction automatique, entre autres. Nous allons intégrer le module de traduction automatique sous forme de *Baseline Engine*, ainsi que la fonctionnalité *Adaptive Engine*. Voyons dans les détails de quoi il s'agit.

L'*Adaptive Engine* est une nouvelle fonctionnalité fondée sur l'apprentissage incrémental, capable de mémoriser et de « s'adapter » au style, au ton, ainsi qu'à la terminologie du traducteur¹⁶. L'intégration dans l'environnement *SDL Trados Studio* se fait à travers l'activation d'un module de traduction automatique. Le traducteur a le choix entre utiliser celui proposé par défaut, le *Baseline Engine*, ou créer et activer un module personnel dit *Adaptive Engine*, qui est capable d'apprendre des corrections du traducteur dès la première utilisation. Dans cette optique, le processus de traduction se configure plutôt comme une activité de post-édition. Après la création et l'activation de l'*Adaptive Engine*, chaque segment est soumis au logiciel de traduction automatique, qui rend une traduction cible. À ce stade, le traducteur peut accepter la traduction telle quelle, ou la post-éditer en apportant des modifications. D'après (SDL, 2017b, non énuméré), chaque fois que nous modifions la cible de la traduction automatique et qu'ensuite nous confirmons la nouvelle traduction, l'*Adaptive Engine* apprend et mémorise les changements. Lorsque le même segment apparaît à nouveau à l'intérieur du même projet ou ailleurs et la fonctionnalité est activée, il ne sera plus soumis au système de traduction automatique : l'*Adaptive Engine* reconnaîtra la suite des caractères et rendra la traduction précédemment confirmée.

Les *Baseline Engines* représentent les systèmes de traduction automatique génériques développés par *SDL* pour toute paire de langues, et contiennent des centaines de millions de mots issus de corpus de données

¹⁴ <https://www.sdl.com/it/languagecloud/> (dernière consultation : 06/03/2018)

¹⁵ <https://www.sdltrados.com/products/trados-studio/adaptivemt/> (dernière consultation : 06/03/2018)

¹⁶ <https://gateway.sdl.com/apex/communityknowledge?articleName=000002881> (dernière consultation : 23/04/18)

bilingues. Ils constituent le point de départ pour la création de toute nouvelle direction de traduction. *SDL* utilise des bases de données de traduction existantes pour la création de nouvelles paires de langues. Ces données sont issues de sources fiables et publiquement accessibles, telles que la documentation informatique, les manuels techniques, ainsi que le matériel gouvernemental, et couvrent plusieurs domaines comme par exemple, entre autres, l'informatique, l'industrie automobile, l'actualité, le sport, l'électronique (SDL, 2017a, p. 58). *SDL* s'engage également à surveiller et à mettre à jour régulièrement ses *Baseline Engines*, ainsi qu'à garantir la protection et la confidentialité des données de traduction personnelles.

Le système de traduction automatique adopté par *SDL* est statistique, statique par définition ; par conséquent, lorsque les utilisateurs exploiteront les *Baseline Engine*, la base de données « centrale » ne sera pas mise à jour. Seulement *SDL* peut modifier directement les données d'entraînement pour influencer la traduction sortante, ce qui empêche toute interférence entre les utilisateurs. En effet,

[c]urrently, statistical machine translation engines learn during the engine training process by analysing the statistical relationships between large amounts of source and target data. Once an engine is created, it is static. Any changes made during post-editing will populate the updated translation memory but not the MT engine. MT engines are updated through a regular retraining process. This means that without AdaptiveMT, the machine repeats the same errors – the translator provides feedback during the post-editing process but the machine cannot benefit from the feedback (SDL, 2017a, p. 60).

Adaptive Engine est également fondé sur le système de traduction statistique actuellement exploité par *SDL* ; cependant, à la différence du *Baseline Engine*, l'*Adaptive Engine* adopte une approche dynamique ; en effet, d'après (SDL, 2017a, p. 61),

[w]ith AdaptiveMT, by contrast, the machine learns seamlessly and continuously from the user feedback, in real time, during the post-editing process. Any post-edits are fed back to the machine translation engine, hosted by SDL, and this engine is completely personal to the translator. (...) AdaptiveMT has evolved from a traditional static MT system to a highly scalable and responsive dynamic system capable of self-learning and self-improvement.

Après avoir présenté les logiciels utilisés dans l'expérience, passons maintenant à la description de la méthodologie de travail.

5.7 La méthodologie

D'abord, nous avons préparé le milieu expérimental. Comme nous l'avons déjà mentionné, l'environnement de travail pour les expériences de traduction et post-édition est le système de mémoire de traduction *SDL Trados Studio 2017*. Pour chaque participant, nous avons créé un nouveau projet de traduction, sélectionné les langues source et cible (anglais et italien, respectivement) et soumis le premier fichier à traduire en format .doc. Au moment de sélectionner les ressources à utiliser (depuis le même menu où l'on choisit les mémoires de traduction), nous avons ajouté le module de traduction automatique. Pour la première phase de la première partie (section 6.1.1), nous avons aussi créé et activé un module *Adaptive Engine* personnel pour notre participant. Pour la deuxième phase, nous avons intégré simplement le module de traduction automatique *Baseline Engine*.

Pour la première phase avec l'*Adaptive Engine*, ainsi que pour les traductions avec le *Baseline Engine*, le fichier traduit était le même pour tous les participants et comportait 244 segments au total ; en outre, d'après le rapport d'analyse de *SDL Trados*, il présentait un taux de répétitions internes de 24%.

Il est important de spécifier que, pour chaque projet de post-édition, nous n'avons pas tenu compte du respect du formatage ; en effet, il ne sera pas indispensable aux fins de l'évaluation de la qualité de la traduction automatique, réalisée à l'aide des mesures automatiques.

Nous précisons aussi que l'intervention de l'*Adaptive Engine* n'est jamais clairement signalée par *SDL Trados* : nous pouvons simplement la déduire en analysant le comportement du logiciel à travers ses suggestions de traduction, ainsi qu'en consultant les rapports d'analyse. Nous précisons aussi que l'*Adaptive Engine* peut intervenir soit au niveau du segment en entier, soit au niveau de sous-segments (mots, termes, collocations).

Dans le cas des traductions avec l'*Adaptive Engine*, ainsi que dans le cas des traductions avec le *Baseline Engine*, les suggestions proposées par le système étaient de deux types : soit la traduction automatique, indiquée par l'acronyme « AT », soit les répétitions à 100% pour les segments identiques et précédemment confirmés (*Perfect Match*), indiqués par le symbole « 100% ».

Lors de la préparation de chaque projet, autant dans le cas de traduction avec l'*Adaptive Engine* qu'avec le *Baseline Engine*, nous avons activé la mise à jour des données de traduction (uniquement pour le projet

courant) en cochant la case « Update ». Cette option garantit la mémorisation des corrections aussi pour le *Baseline Engine*, mais uniquement à l'intérieur du projet courant – ce qui devrait empêcher toute interférence entre les utilisateurs. Cependant, comme nous le verrons dans l'analyse, le système a donné des résultats différents pour les deux types de projets.

Avant de débiter la traduction, nous avons créé un nouveau projet et effectué une pré-traduction du fichier avec le module de traduction automatique *Baseline Engine*, pour ensuite exporter et sauvegarder cette version « brute ». Nous avons analysé manuellement cette traduction et l'avons ensuite comparé avec la référence pour calculer le TER (afin de déterminer la distance d'édition) ainsi que le BLEU (pour évaluer la qualité de la traduction automatique sortante). Cette phase nous a permis de répondre à la première sous-question de recherche.

Avant de procéder avec la traduction, nous avons lancé notre logiciel d'enregistrement d'écran *BB Flashback Express*. Dans le cadre de notre travail, nous avons demandé à chaque participant de démarrer l'enregistrement juste avant de débiter l'expérience. Une fois l'enregistrement activé, le participant pouvait commencer à traduire. Pour chaque segment, le logiciel de traduction automatique a donné une suggestion de traduction. Cette suggestion est affichée dans le panneau en haut et est signalée avec l'acronyme AT (en anglais *Automatic Translation*) sur fond bleu. Après avoir lu le segment source et la relative suggestion de traduction, le participant avait le choix entre accepter la proposition telle quelle – qui peut être insérée automatiquement avec un raccourci et ensuite confirmée – ou la post-éditer, ou encore la rejeter en retraduisant le segment dans son entièreté.

Lorsqu'ils choisissent la suggestion de la traduction automatique, nous allons demander à nos participants d'exécuter une post-édition « complète », c'est-à-dire, destinée à produire un texte de haute qualité, égal ou comparable à une traduction humaine, donc apte à être publiée ; à cet effet, dans les instructions, nous allons demander à nos participants de suivre les directives *TAUS* relatives à la post-édition complète, que nous allons exposer dans les instructions en annexe (Annexe II). D'après ces directives, l'objectif de la post-édition complète est de produire un texte compréhensible, capable de communiquer exactement le même signifié contenu dans la source, stylistiquement acceptable et correct au niveau syntaxique et grammatical. De plus, la ponctuation doit aussi être correctement reproduite.

Lorsque le traducteur décide de post-éditer la phrase, le symbole AT change sa couleur de fond de bleu à gris, ce qui indique qu'une modification a eu lieu. Pour la première phase de la première partie – celle qui exploite la fonctionnalité *Adaptive Engine* - une fois que le segment a été post-édité et confirmé, le logiciel

devrait apprendre des corrections du traducteur. Idéalement, lorsqu'un segment similaire revient, la proposition de traduction automatique du logiciel devrait correspondre à la traduction précédemment post-éditée. Une fois que le participant a traduit (ou post-édité) tous les segments, il doit finaliser le projet en exportant la version cible du texte. Au même temps, il s'arrête et sauvegarde l'enregistrement.

Ensuite, sur la base des informations mémorisées lors de la première phase de traduction avec l'*Adaptive Engine*, nous avons effectué deux tests de retraduction, toujours dans l'environnement *SDL Trados Studio*. Pour le premier, nous avons extrait de manière aléatoire 50 segments du même document, qui ont été soumis au logiciel pour pré-traduction. Les données recueillies dans cette étape nous ont servi pour déterminer le degré d'intégration effectif des préférences stylistiques du traducteur de la part du logiciel. Pour le deuxième, nous avons soumis au logiciel 40 segments extraits d'autres guides utilisateur ayant une terminologie et une syntaxe similaires à celles des premiers. Nous avons vérifié le degré d'adaptation du logiciel aux préférences de l'utilisateur.

Ensuite, nous avons analysé les données recueillies, notamment les enregistrements vidéo, les traductions post-éditées, ainsi que le projet de retraduction avec l'*Adaptive Engine*. Pour la première partie de l'expérience, d'abord, nous avons évalué la qualité de la traduction automatique « brute » en calculant le TER et le BLEU afin de répondre à la première sous-question. Comme nous l'avons mentionné à la section 3.2, les mesures utilisées dans l'expérience (BLEU, TER et sa version HTER) sont en effet les plus communément utilisées au niveau scientifique. Ensuite, nous nous sommes focalisée sur l'analyse de la fonctionnalité *Adaptive Engine* et le degré d'intégration des corrections. Cette phase nous a permis de répondre à la deuxième question de recherche. Ensuite, nous avons comparé le temps requis pour traduire le fichier à l'aide de la fonctionnalité *Adaptive Engine* avec le temps moyen mis par les participants qui ont utilisé le *Baseline Engine*. Cette phase nous a permis de répondre à la deuxième sous-question, relative à l'impact sur la productivité temporelle de l'*Adaptive Engine* par rapport au *Baseline Engine*.

Pour la deuxième partie, nous avons évalué l'effort de post-édition en analysant le taux de post-édition par nos participants. Nous avons aussi mesuré la qualité de la traduction automatique post-éditée en utilisant des mesures automatiques. Plus précisément, pour chaque participant, nous avons comparé la traduction brute avec les versions finales post-éditées (nos références) en calculant le HTER et le BLEU. Nous avons ainsi évalué le niveau d'acceptation et d'exploitation de la suggestion de traduction automatique, ainsi que la qualité de la traduction en relation à l'effort de post-édition. Nous rappelons que le fournisseur de la traduction automatique (*SDL Language Cloud*) est dans le cloud ; par conséquent,

le corpus d'entraînement du logiciel est le même pour tous les participants. Cette étape vise à répondre à la troisième sous-question de recherche. Finalement, par l'intégration des réponses obtenues de nos participants, qui constituent la troisième partie de l'étude (section 6.1.3), nous avons pu répondre à la première question, relative à l'intérêt de l'intégration d'un module de traduction automatique au flux de travail du traducteur. Vous trouverez le protocole détaillé en annexe (Annexe I).

Nous terminons le chapitre dédié à l'expérience. Maintenant, nous passons au chapitre dédié à la présentation, l'analyse et l'interprétation des résultats. Nous allons terminer la section avec un récapitulatif des réponses au questionnaire de la part des participants.

6. Les résultats

Nous présentons maintenant les résultats obtenus lors de l'expérience, ainsi que l'analyse et l'interprétation des données dans l'optique de notre étude. Sur la base de nos réflexions formulées dans la prochaine section, nous tirerons nos conclusions, qui seront exposées au chapitre 7.

6.1 Présentation et analyse des résultats

Notre travail comporte trois parties principales : les phases de traduction et de post-édition (6.1.1), les phases d'évaluation de la qualité de la traduction automatique à l'aide de mesures automatiques (6.1.2) et le questionnaire (6.1.3). Nous présentons maintenant la première partie de l'expérience.

6.1.1 Première partie : traduction et post-édition

Nous avons conduit notre étude en quatre phases : traduction avec l'*Adaptive Engine*, test de retraduction avec l'*Adaptive Engine* pour la vérification des mémorisations, test de retraduction avec l'*Adaptive Engine* pour la vérification de l'adaptation, tests avec le *Baseline Engine* pour la vérification du degré de post-édition.

D'abord, nous allons présenter les résultats de l'analyse manuelle de la traduction brute générée par le module *Baseline Engine*. Ensuite, nous allons examiner la première phase de traduction effectuée avec l'*Adaptive Engine* ; après, nous allons comparer ces données avec celles des projets effectués par nos 7 participants avec le *Baseline Engine*, qui seront décrits dans la quatrième phase. Nous pourrons ainsi évaluer la capacité d'adaptation et d'apprentissage du premier logiciel et déterminer les éventuels avantages par rapport au deuxième.

En outre, nous allons présenter les résultats de deux tests de performance de l'*Adaptive Engine* : pour le premier, décrit dans la phase 2, nous allons examiner la capacité d'apprentissage de la fonctionnalité en retraduisant 50 segments extraits de manière randomisée du premier projet. Dans le deuxième test, présenté dans la phase 3, nous avons soumis au logiciel 40 segments extraits d'autres guides utilisateur pour l'imprimante de marques connues pour examiner le degré d'adaptation de l'*Adaptive Engine* aux préférences traductives de l'utilisateur. Ces 40 segments contiennent des tournures ainsi que de la terminologie similaire à celles traduites dans les premiers projets.

Avant de nous focaliser sur la performance de l'*Adaptive Engine*, nous allons présenter les résultats de l'analyse manuelle de la traduction brute. Pour chaque segment, nous avons examiné et classé les erreurs repérées. Nous avons déterminé 7 catégories d'erreurs : **traduction, calque, capitalisation, accord** (genre

et nombre), **mode verbal**, **temps verbal**, **divers**. Les erreurs de traduction (mots erronés, erreurs de sens, erreurs de terminologie, erreurs de désambiguïsation) appartiennent à la première catégorie. La deuxième catégorie contient les calques de l’anglais. La catégorie « capitalisation » comprend tous les cas où le système n’a pas reproduit la majuscule du segment source. Les erreurs de type « accord » comprennent les cas où le système n’a pas respecté le genre et le nombre dans la traduction. Les catégories de « mode » et « temps verbal » comprennent les cas où le système n’a pas respecté le temps et le mode du verbe dans le segment source. Finalement, la catégorie « divers » regroupe toutes les erreurs non énumérables dans les classifications précédentes, telles que les erreurs dans l’ordre des mots et l’orthographe. En plus de ces 7 catégories, nous avons défini deux groupes à part : les **insertions**, qui regroupent les mots et les signes de ponctuation absents dans la source et qui ont été ajoutés par le système, et les **omissions**, qui en revanche regroupent les mots et les signes de ponctuation présents dans la source et qui ont été supprimés par le système. Voici un tableau récapitulatif des erreurs repérées.

TRADUCTION	CAPIT.	CALQUE	INSERTION	DIVERS	OMISSION	ACCORD	MODE	TEMPS
67	27	24	15	13	12	8	8	5

Tableau 1 : Nombre d'erreurs par catégorie repérées dans la traduction brute générée par le Baseline Engine

Globalement, les erreurs les plus fréquentes commises par le système étaient celles de traduction (67), capitalisation (27) et calque (24), suivies par les catégories insertion (15), divers (13), omission (12), accord (8), mode verbal (8) et, finalement, temps verbal (5). En particulier, le système a omis 6 mots, 3 signes de ponctuation et 3 espaces.

Nous analysons maintenant quelques exemples des principales difficultés traductives rencontrées par le système. La première catégorie que nous analysons est l’erreur de traduction. Nous présentons dans le tableau ci-dessous quelques cas d’erreurs de « désambiguïsation » des homonymes.

# SEG		
164	Point to or tap the upper-right corner of the screen to open the Charms bar (...).	Il punto a o toccare l'angolo superiore destro della schermata per aprire la barra di fascino (...).
186	If both ink cartridge lights are still blinking, then one or both cartridges may be defective.	Se entrambe le spie delle cartucce di stampa lampeggiano ancora, poi una o entrambe le cartucce potrebbero essere difettose.
176	Clear print carriage	Cancella carrello di stampa
174	Check the print queue again to make sure it is clear, and then try to print again.	Controllare nuovamente la coda di stampa per assicurarsi che sia chiaro, quindi provare a stampare di nuovo.
70	Open the cleanout door located on the bottom of the printer.	Aprire lo sportello di pulizia si trova nella parte inferiore della stampante.

Tableau 2 : Exemples d'erreurs de traduction

Dans le premier cas, « point to » (V suivi par C.O.D.) a été traduit par « il punto » (N). Une autre difficulté traductive très fréquente consiste dans le choix de la traduction appropriée pour le contexte donné : par exemple, « then » a été traduit par « poi » (« après »), alors que, dans le contexte, la traduction la plus appropriée serait « allora » (« cela signifie que » ou « donc »). Un autre exemple est la traduction du verbe « clear » dans « Clear print carriage » (« Vider ») par « cancella » dans « Cancella carrello di stampa » (« Supprimer »), reprise par la traduction erronée de l'adjectif « clear » par « chiaro » (« clair ») au lieu de « vuoto » (« vide »).

Dans le dernier cas, le système ne parvient pas à choisir la bonne traduction (mode et temps verbal) pour le participe passé « located » (littéralement « situé »), qui a été traduit à l'indicatif présent (« est situé »).

Une autre catégorie d'erreurs très fréquente est représentée par les calques.

# SEG		
8	Check the estimated ink levels to determine if the ink cartridges are low on ink .	Controllare i livelli di inchiostro stimati per determinare se le cartucce di inchiostro sono un basso livello di inchiostro .
10	If the ink cartridges are low on ink consider replacing them.	Se le cartucce di inchiostro sono un basso livello di inchiostro si consiglia di sostituirli.
27	Print a diagnostics page if the ink cartridges are not low on ink .	Stampare una pagina di diagnostica se le cartucce di inchiostro non sono un basso livello di inchiostro .

Tableau 3 : Exemples d'erreurs de calque

L'exemple le plus flagrant est représenté par « the ink cartridges are low on ink » (littéralement « les cartouches sont à court d'encre » ou « les cartouches n'ont plus beaucoup d'encre »), qui a été traduit par « le cartucce sono un basso livello di inchiostro » (littéralement « les cartouches sont un niveau d'encre bas »).

Considérons maintenant la catégorie temps verbal.

# SEG		
103	Point to or tap the upper-right corner of the screen to open the Charms bar (...)	Il punto a o toccare l'angolo superiore destro della schermata per aprire la barra di fascino (...)

Tableau 4 : Exemple d'erreurs de temps verbal

Comme nous pouvons l'observer, dans le cas de « tap », le système ne parvient pas à identifier la forme impérative (littéralement « touchez ») et traduit le verbe à la forme infinitive.

Passons à la catégorie mode verbal.

# SEG		
124	Make sure the correct printer is set as the default printer.	Assicurarsi che la stampante corretta è impostata come stampante predefinita.

Tableau 5 : Exemple d'erreurs de mode verbal

Dans l'exemple, le prédicat « is set » (littéralement « est sélectionnée ») a été traduit à la forme indicative présente, alors que le verbe à la forme infinitive « assicurarsi » (« s'assurer ») requiert le conjonctif.

Considérons la catégorie divers. Le cas le plus fréquent est représenté par les erreurs dans l'ordre des mots.

# SEG		
84	Do not force paper too far forward in the input tray.	Non forzare troppo la carta in avanti nel vassoio di alimentazione.

Tableau 6 : Exemple d'erreurs de type "divers"

Dans l'exemple, la suite d'adverbes « troppo in avanti » a été séparée par l'introduction du substantif « la carta » (« le papier »).

En outre, le système montre souvent des difficultés à lier les mots éloignés : en particulier, nous avons repéré des erreurs dans les accords en genre et nombre. Ces difficultés s'expliquent à partir de la nature même du système de traduction automatique statistique, qui manque d'informations grammaticales et contextuelles.

# SEG		
10	If the ink cartridges are low on ink consider replacing them .	Se le cartucce di inchiostro sono un basso livello di inchiostro si consiglia di sostituirli.
237	If one ink cartridge is causing problems, you can remove it (...)	Se una cartuccia di inchiostro è la causa dei problemi è possibile rimuoverlo (...)

Tableau 7 : Exemples d'erreurs d'accord en genre et nombre

Le système montre parfois des difficultés dans le repérage, l'identification et la traduction des termes. L'exemple le plus clair est représenté par la traduction de « HP Print and Scan Doctor » par « HP di stampa e scansione di medico », où le système n'a pas identifié le terme (le nom de la fonctionnalité) et en donne une traduction littérale.

# SEG		
198	HP provides an HP Print and Scan Doctor (Windows only) (...)	HP fornisce una stampa HP Scan e medico (solo Windows) (...)
200	The HP Print and Scan Doctor may not be available in all languages.	La HP di stampa e scansione di medico potrebbe non essere disponibile in tutte le lingue.

Tableau 8 : Exemples d'erreurs de traduction (terminologie)

Du point de vue typographique, le système présente parfois des difficultés dans la reproduction de la majuscule de l'original, ainsi que dans la mémorisation et dans l'application du formatage dans la cible. Plus précisément, il a commis des erreurs dans la délimitation des mots concernés par le formatage et a

parfois tendance à appliquer le style à d'autres mots voisins, qui ne sont pas concernés dans la source. Cependant, ce dernier aspect ne sera pas pris en considération aux fins de l'évaluation de la qualité de la traduction.

# SEG		
2	NOTE:	Nota:

Tableau 9 : Exemple d'erreurs de formatage

Le système est parfois très incohérent dans la traduction ; le même terme a souvent été traduit en deux manières différentes à l'intérieur du même projet.

# SEG		
128	Restart the print spooler .	Riavviare lo spooler di stampa .
153	Double-click Services, and then select Print Spooler .	Fare doppio clic su Servizi, quindi selezionare Print Spooler .

Tableau 10 : Exemples d'erreurs de traduction (cohérence traductive)

Dans le cas de « Print Spooler », parfois le système traduit ce terme avec « spooler di stampa » (segments 128, 129, 135 141, 147), parfois avec « Print Spooler » (segments 153 et 154).

La grande quantité d'erreurs repérées est en corrélation avec les résultats très peu satisfaisants du TER (0,42), que nous analyserons dans la deuxième partie.

Phase 1 – Traduction avec l'Adaptive Engine

D'après l'analyse du fichier, sur 244 segments, 63 étaient indiqués comme répétitions à 100%.

Sur 244 segments traduits, le taux d'acceptation de la traduction automatique et des correspondances parfaites (*Perfect Match* - 100%) de la part du traducteur a été de 100% ; par conséquent, aucun segment n'a été retraduit de zéro. Effectivement, d'après le traducteur (nous-même), malgré la quantité d'erreurs, la qualité de la traduction automatique a été jugée globalement assez bonne pour le projet. Cela peut se justifier aussi par le fait que, comme nous l'avons déjà précisé, le texte était vraisemblablement rédigé en langage contrôlé, qui facilite le traitement de la part du système et qui a finalement un impact positif sur la qualité de la traduction sortante.

D'après une analyse plus détaillée du projet, 63 segments (des répétitions à 100%) ont été confirmés et acceptés sans modifications et 74 segments indiqués comme suggestion de la traduction automatique (AT)

ont été post-édités. Au total, 107 segments indiqués comme suggestions de la traduction automatique (AT) ont été confirmés et acceptés sans modifications.

Sur 244 segments traduits, 77 ont été correctement mémorisés (totalement ou au niveau de sous-segment) ; par conséquent, l'impact de l'*Adaptive Engine* a concerné environ le 32% du document. Plus spécifiquement, nous remarquons que le système mémorise plus facilement les segments entiers que les sous-segments ou les mots/termes/expressions considérés individuellement. Pour avoir une confirmation de l'impact de l'*Adaptive Engine* sur le projet, nous avons effectué une analyse du fichier final. Les résultats contenus dans le rapport d'analyse confirment nos calculs.

Sur 244 segments traduits, 8 n'ont pas été mémorisés (entièrement ou partiellement), ce qui représente environ le 3% du document. Dans ces cas, le système a proposé à nouveau la suggestion de traduction erronée, même après correction (à travers la post-édition) et confirmation du nouveau segment. Les résultats les plus pertinents sont résumés dans le tableau récapitulatif présenté ci-dessous.

TOT	100%	TA + PE	TA
244	63	74	107

Tableau 11 : Récapitulatif des nombres de segments par typologie

TOT	# segments mémorisés	# segments non mémorisés
244	77	8

Tableau 12 : Récapitulatif des nombres de segments mémorisés et non mémorisés

Phase 2 – Test de retraduction avec l'*Adaptive Engine* – vérification des mémorisations

D'abord, nous avons généré une série randomisée de 50 numéros à l'aide du logiciel en ligne *Random Integer Generator*¹⁷. Ensuite, nous avons repéré les segments correspondants dans le fichier source en constituant un nouveau fichier « artificiel ». Ensuite, nous avons créé un nouveau projet de traduction : nous avons sélectionné les langues source et cible (les mêmes que les autres projets), nous avons ajouté le fichier à traduire, la mémoire de traduction pour la segmentation, ainsi que le fournisseur de la traduction automatique *SDL Language Cloud* et le module *Adaptive Engine*. Finalement, dans les réglages

¹⁷ <https://www.random.org/> (dernière consultation : 28/05/2018)

des *Batch Tasks*, nous avons spécifié « Pre-Translate » et « Apply machine translation ». Nous avons donc lancé la pré-traduction du fichier sur la base des informations mémorisées par l'*Adaptive Engine* lors de la première phase de traduction. Nous avons ensuite généré les traductions cibles et nous avons exporté le fichier final. Ensuite, nous avons comparé les segments traduits avec ceux confirmés lors de la première étape.

Parmi les 50 segments sélectionnés, 15 étaient de type « TA avec post-édition », 12 de type « répétition à 100% confirmée » et 23 de type « TA sans post-édition ».

D'après les résultats, le système a correctement traduit 43 segments sur 50, ce qui représenterait apparemment un très bon résultat. Cependant, d'après une analyse détaillée des segments à travers une comparaison avec le projet précédent, nous avons remarqué un certain nombre d'aspects intéressants qui infirment sa performance. En effet, dans 5 cas de type TA + PE (139, 242, 8, 200 et 171) et dans 1 cas de type « répétition à 100% » déjà apparu et confirmé (168), le système n'a pas mémorisé les corrections du traducteur ; dans le dernier cas de type « TA sans PE » (83), le système a même inséré une erreur de majuscule.

Dans trois cas, le système a vraisemblablement retraduit le segment, en proposant la suggestion de la traduction automatique initiale : « Dal menu Start di Windows, fare clic su il pannello di controllo, Sistema e sicurezza, quindi su Strumenti di amministrazione. » (qui correspond aux segments 139 et 168) et « Se si trova scritto 'v1' diversi spazi a destra della data , allora la cartuccia è la versione aggiornata della versione più recente. » (segment 242).

# SEG		
3	From the Windows Start menu, click C ontrol P anel, System and Security, and then Administrative Tools.	Dal menu Start di Windows, fare clic su il pannello di controllo , Sistema e sicurezza, quindi su Strumenti di amministrazione.
26	If it says 'v1' several spaces to the right of the date, then the cartridge is the updated newer version .	Se si trova scritto 'v1' diversi spazi a destra della data , allora la cartuccia è la versione aggiornata della versione più recente .

Tableau 13 : Exemples d'erreurs de traduction

Dans le premier cas, le système a commis une erreur d'orthographe et n'est pas parvenu à mémoriser la structure correcte préposition + article « sul pannello di controllo » ; en outre, il n'a pas correctement reproduit la capitalisation de la source. En ce qui concerne le segment 242, le système n'est pas parvenu

à mémoriser l'ordre de mots imposé par les corrections du traducteur, et a maintenu l'espace supplémentaire.

Dans le segment 35 (correspondant au segment 8 de la première phase), le système a commis une omission (« livello »).

# SEG		
35	Check the estimated ink levels to determine if the ink cartridges are low on ink.	Controllare i livelli di inchiostro stimati per determinare se il di inchiostro nelle cartucce è basso.

Tableau 14 : Exemple d'erreurs d'omission

Dans le segment 36 (segment 200 de la première phase), le système a inséré un mot qui n'était pas présent dans la source (« Di »).

# SEG		
36	The HP Print and Scan Doctor may not be available in all languages.	Di HP Print and Scan Doctor potrebbe non essere disponibile in tutte le lingue.

Tableau 15 : Exemple d'erreur d'insertion

Dans le segment 39 (200 dans la première phase), le système n'a pas respecté la capitalisation précédemment confirmée (« Lettera »).

# SEG		
39	Load letter or A4 unused plain white paper into the input tray.	Caricare carta bianca normale formato lettera o A4 nel vassoio di alimentazione.

Tableau 16 : Exemple d'erreur de capitalisation

Finalement, dans le segment 49 (171 dans la première phase), le système a encore une fois commis une erreur de désambiguïsation et n'est pas parvenu à traduire correctement le verbe à la forme impérative « Double click », qui a été interprété à nouveau comme un substantif.

# SEG		
49	Double-click the icon for your printer to open the print queue.	Doppio clic sull'icona della stampante per aprire la coda di stampa.

Tableau 17 : Exemple d'erreur de désambiguïsation

Pour conclure, lorsqu'on considère que, sur 27 segments sur lesquels le traducteur est intervenu lors de la première phase de traduction (15 de type « TA avec PE » et 12 de type « répétition à 100% confirmée »), 21 ont été traduits correctement, nous pouvons affirmer que la performance du système a donné des

résultats plutôt appréciables. Voici un tableau récapitulatif des résultats plus pertinents pour notre analyse.

TOT	CORRECTES	TA + PE	rép. 100%	TA
50	43	15	12	23

Tableau 18 : Récapitulatif des nombres de segments par typologie

TOT	CORRECTES	TA + PE	rép. 100%
27	21	15	12

Tableau 19 : Récapitulatif des nombres de segments par typologie

Globalement, d'après ces résultats, la capacité d'apprentissage de l'*Adaptive Engine* a été assez satisfaisante.

Phase 3 – Test de retraduction avec l'*Adaptive Engine* – vérification de l'adaptation

D'abord, nous avons effectué une pré-traduction du fichier avec le module *Adaptive Engine*. En consultant le rapport d'analyse, sur 40 segments, 8 contenaient des mémorisations recueillies lors des phases de traduction précédentes. Globalement, l'impact de l'*Adaptive Engine* a été de 22,58%.

Sur 40 segments, 16 ont été traduits correctement et ne présentent aucun type d'erreur. Voici un tableau récapitulatif des erreurs par typologie.

TRAD.	ACCORD	CALQUE	INSERT.	DIVERS	MODE	TEMPS	OMIS.	CAPIT.
17	7	6	4	2	2	1	2	0

Tableau 20 : Nombre d'erreurs par catégorie repérées lors du deuxième test de retraduction

Au total, nous avons repéré 17 erreurs de traduction, 7 erreurs d'accord en genre et nombre, 6 calques, 4 insertions, 2 erreurs de type « divers », 2 erreurs de mode verbal, 1 erreur de temps verbal et 2 omissions. Nous présentons maintenant quelques exemples.

Encore une fois, la catégorie qui présente le nombre d'erreurs le plus élevé est « traduction ». En particulier, les principales difficultés traductives rencontrées par le système concernent le traitement des séries de noms, très fréquentes en anglais.

# SEG		
2	Check that the toner cartridges are not low and try again.	Controllare che il toner cartucce non è basso e riprovare.
18	The ink cartridge model number is not compatible with your machine.	La cartuccia di inchiostro il numero del modello non è compatibile con la vostra macchina.
25	The HP Print and Scan Doctor and the HP online troubleshooting wizards may not be available in all languages.	Di HP Print and Scan Doctor e HP online le procedure guidate per la risoluzione dei problemi potrebbero non essere disponibili in tutte le lingue.
38	Load Letter or A4 plain, white paper in the input tray.	Caricare carta formato Lettera o A4 carta bianca comune nel vassoio di alimentazione.

Tableau 21 : Exemples d'erreurs de traduction

« the toner cartridges » a été traduit par « il toner cartucce » (segment 2) ; « The ink cartridge model number » a été traduit par « La cartuccia di inchiostro il numero del modello » (segment 18) ; « The HP Print and Scan Doctor and the HP online troubleshooting wizards » (segment 25) a été traduit par « Di HP Print and Scan Doctor e HP online le procedure guidate per la risoluzione dei problemi » ; dans ce dernier cas en particulier, même s'il s'agissait d'une tournure déjà rencontrée lors des phases précédentes, le système n'arrive pas à lier le substantif « wizards » aux deux antécédents. En plus, une préposition absente dans la source a été insérée. Cependant, il a correctement mémorisé le terme technique « HP Print and Scan Doctor ». Finalement, au segment 38, « (Load) Letter or A4 plain, white paper (in the input tray.) » – encore une fois une tournure déjà rencontrée – a été erronément traduite par « (Caricare) carta formato Lettera o A4 carta bianca comune (nel vassoio di alimentazione) », avec la répétition de « papier ». Il faut toutefois préciser que le système a correctement reproduit la capitalisation.

# SEG		
20	To prevent possible damage to the printhead, clear paper jams as soon as possible.	Per evitare possibili danni alla testina di stampa, eliminazione degli inceppamenti della carta al più presto possibile.

Tableau 22 : Exemple d'erreur de traduction

Au segment 20, en revanche, « clear paper jams » (V suivi par C.O.D.) a été erronément reconnu comme une série de noms et traduite par « eliminazione degli inceppamenti della carta », alors qu'il s'agissait d'un prédicat verbal suivi par complément d'objet direct ; cet exemple constitue une évidente erreur de

désambiguïsation. En outre, dans le même segment, la construction « as soon as possible » a été erronément traduite par « al più presto possibile ».

Un cas intéressant qui rend compte encore une fois de la nature statistique du système est représenté par l’erreur au segment 4.

# SEG		
4	Check printed pages, page counter, and display messages regularly.	Controllare le pagine stampate, contatore di pagina e visualizzare i messaggi regolarmente.

Tableau 23 : Exemple d'erreur de désambiguïsation

Dans cet exemple, le logiciel ne parvient pas à lier le verbe principal « check » aux deux compléments d’objet direct qui suivent. Il effectue donc une omission de l’article dans le cas de « page counter » et une faute de désambiguïsation dans le cas de « display messages » (erronément reconnu comme un prédicat, alors qu’il s’agit d’une suite de substantifs).

Les 7 erreurs d’accord en genre et nombre s’expliquent encore une fois à partir de la nature même du logiciel de traduction automatique statistique, incapable de déduire et exploiter les informations grammaticales. Le système se repose sur la probabilité de l’apparition du mot et retient le masculin plutôt que la forme correcte ici dans les exemples suivants :

# SEG		
15	If the paper is caught in the paper ejection area, gently pull it out.	Se la carta viene catturato nella carta area di espulsione, estrarla con delicatezza.
16	If the paper tears when you remove it or if the paper is caught beyond the feed slot, (...)	Se la carta si strappa durante la rimozione o se la carta è catturato al di là della fessura di alimentazione, (...)
34	Make sure the paper loaded in the input tray is aligned to the paper-width guides.	Accertarsi che la carta caricata nel vassoio di alimentazione è allineato per le guide di larghezza della carta.

Tableau 24 : Exemples d'erreurs d'accord en genre et nombre

Dans les segments 15 et 16, le système ne parvient pas à effectuer l’accord en genre entre « the paper » et « is caught », traduits par « se la carta è catturato » (au lieu de « catturata », « la carta » étant de genre féminin ». Il faut aussi noter, au segment 15, la traduction erronée de la suite de noms « paper ejection

area ». Dans le segment 34, à cause de la distance entre le substantif « the paper » et le prédicat « is aligned », qui sont séparés par 5 mots, le logiciel ne parvient pas à déterminer l'accord en genre.

De même, dans le cas des insertions, nous pouvons déduire que le système se limite à « reproduire » dans la cible une séquence de n-grams repérée dans les corpus d'entraînement ; voir, par exemple, les segments 28, 38, ainsi que 39.

# SEG		
28	If you have colored paper loaded in the input tray when you align the ink cartridges , the alignment fails.	Se si dispone di carta colorata carta caricata nel vassoio di alimentazione quando si allinea le cartucce di inchiostro , l'allineamento non riesce.
38	Load Letter or A4 plain, white paper in the input tray.	Caricare carta formato Lettera o A4 carta bianca comune nel vassoio di alimentazione.
39	If the diagnostic page shows streaks or missing portions of the text or colored boxes , clean the printhead automatically.	Se la pagina di diagnostica mostra striature o parti mancanti delle il testo o caselle colorate , pulire la testina di stampa automaticamente.

Tableau 25 : Exemples d'erreurs d'insertion

Les adaptations que nous avons repérées concernaient surtout les modes verbaux ; en effet, dans les segments 5, 10, 13, 33 et 35, le logiciel a correctement reproduit le conjonctif, alors que ce mode verbal posait beaucoup de problèmes traductifs lors de la première phase.

# SEG		
5	If your printer does not work correctly immediately after you install it, check that the printer has power and is properly connected to the computer.	Se la vostra stampante non funziona correttamente subito dopo l'installazione, verificare che la stampante sia accesa e che sia collegato correttamente al computer.
10	Check that the power cord is securely connected to both the AC inlet on the back panel and an electrical outlet.	Verificare che il cavo di alimentazione sia saldamente collegato alla presa CA sul pannello posteriore e una presa elettrica.
13	Also make sure your printer has paper in it, since many printers cannot operate without paper.	Accertarsi inoltre che la stampante disponga di carta in esso, poiché molte stampanti non possono funzionare senza carta.
33	Make sure the paper-width guides are set to the correct markings in the tray for the paper size you are loading.	Assicurarsi che le guide di larghezza della carta siano impostate per la marcatura corretta nel vassoio per le dimensioni della carta che si sta caricando.
35	If an error occurs after a cartridge is installed, or if a message indicates a cartridge problem, try removing the ink cartridges, verifying the protective piece of plastic tape has been removed from each ink cartridge, and then reinserting the ink cartridges.	Se si verifica un errore dopo aver installato una cartuccia, o se viene visualizzato un messaggio che indica un problema relativo alla cartuccia, provare a rimuovere le cartucce di inchiostro, verificando il nastro di protezione in plastica sia stato rimosso da ciascuna cartuccia di inchiostro, e quindi reinsertire le cartucce di inchiostro.

Tableau 26 : Exemples d'adaptations (conjonctif)

Comme nous l'avons déjà mentionné, aux segments 25 et 26, le logiciel a correctement mémorisé et traduit « HP Print and Scan Doctor », terme déjà apparu lors de la première phase de traduction et qui n'a pas été retraduit en italien. Aussi, au segment 40, il a correctement traduit le terme « HP support ».

Au segment 27, le système parvient à traduire correctement la série de noms « unused, plain white paper » ; au segment 30, dans le cas de « too far », il arrive à mémoriser pas seulement la traduction de

l’adverbe, mais aussi sa position dans la suite de mots – une structure déjà apparue et vraisemblablement mémorisée.

Similairement, au segment 36, le logiciel a correctement mémorisé la traduction de la série de substantifs « a set of gold-colored bumps ».

# SEG		
27	If the alignment process fails, make sure you have loaded unused, plain white paper into the input tray.	Se il processo di allineamento ha esito negativo, accertarsi di aver caricato la carta bianca comune non utilizzata in precedenza nel vassoio di alimentazione.
30	Do not force paper too far inside the input tray.	Non forzare eccessivamente la carta all'interno del vassoio di alimentazione.
36	The contacts look like a set of gold-colored bumps positioned to meet the contacts on the ink cartridge.	I contatti si presentano come un set di protuberanze di colore oro adiacenti ai contatti sulla cartuccia di inchiostro.
40	If print quality problems persist after cleaning and aligning, contact HP support .	Se i problemi di qualità di stampa persistono anche dopo la pulizia e l'allineamento, contattare il supporto HP .

Tableau 27 : Exemples d'adaptations (suites de noms)

Cependant, parfois le système n’est pas cohérent dans ses traductions. Considérons les deux premiers segments, qui partagent la même tournure « to be low », qui est déjà apparue dans plusieurs cas lors de la première phase ; si, dans le premier cas, le système parvient à traduire correctement la structure « is getting low », dans le deuxième, il reproduit le calque, déjà corrigé. De même, dans les segments 5, 13, 33 et 35, comme nous l’avons vu, il arrive à reproduire le conjonctif – requis en italien par le verbe « assicurarsi » (« s’assurer ») et « verificare » (« vérifier ») ; cependant, lorsque la même structure réapparaît aux segments 2 et 6, il traduit les verbes à l’indicatif.

# SEG		
2	Check that the toner cartridges are not low and try again.	Controllare che il toner cartucce non è basso e riprovare.
6	Make sure both the computer and the printer are off .	Assicurarsi che il computer e la stampante sono spenti .

Tableau 28 : Exemples de répétitions d'erreurs

Au segment 8, le verbe prépositionnel « to be on » pose à nouveau des difficultés. Sur la base du segment « Check that the power is on » de la première phase de traduction, le système omet la traduction de la

préposition « on » et ne parvient pas à lier le prédicat à la préposition, en traduisant par le calque « È la spia ».

Au segment 11, bien que cette construction soit déjà apparue dans la première phase, le système commet à nouveau un calque en traduisant « always wait » par « sempre spegnere » ; encore une fois, le logiciel n’a pas correctement mémorisé et reproduit l’ordre naturel des mots.

Finalement, aux segments 25 et 26, le logiciel insère à nouveau la préposition « Di », absente dans la source – une erreur déjà commise et corrigée lors de la première phase.

# SEG		
8	Is the power light on the computer’s front panel on ?	È la spia di alimentazione sul pannello anteriore del computer?
11	If you need to turn off the computer for any reason, always wait at least five seconds before switching it back on again .	Se è necessario spegnere il computer per qualsiasi motivo, sempre attendere almeno cinque secondi prima di passare di nuovo .
25	The HP Print and Scan Doctor and the HP online troubleshooting wizards may not be available in all languages.	Di HP Print and Scan Doctor e HP online le procedure guidate per la risoluzione dei problemi potrebbero non essere disponibili in tutte le lingue.
26	The HP Print and Scan Doctor is a utility that will try to diagnose and fix the issue automatically.	Di HP Print and Scan Doctor è un programma di utilità che tenta di diagnosticare e correggere il problema automaticamente.

Tableau 29 : Exemples d’erreurs d’insertion

D’après cette analyse, nous avons remarqué que le taux d’erreurs est encore très élevé, même à la troisième phase de traduction de documents similaires. Globalement, l’intervention de l’*Adaptive Engine* n’a pas eu un impact positif sur la qualité de la traduction automatique, qui n’a évidemment pas enregistré des améliorations progressives.

À partir de ces données, nous pouvons répondre à notre **deuxième question de recherche** : la capacité d’adaptation de l’*Adaptive Engine* ne serait pas suffisante pour justifier son intégration dans le flux de travail d’un traducteur indépendant.

Phase 4 – Traduction avec le *Baseline Engine*

Nous présentons ci-dessous un tableau récapitulatif des données de traduction avec l'*Adaptive Engine* (A) et le *Baseline Engine* (B1 à B7). Les phases de traduction considérées sont celles exécutées par nous-même avec l'*Adaptive Engine* et les 7 projets de traduction exécutés par nos participants avec le *Baseline Engine*. Le fichier était la section *Solve a problem* (244 segments).

Participant	Temps	TA	TA + PE	100%	100% + PE
A	1h 08m 10s	107	74	63	0
B1	1h 28m 00s	92	87	63	2
B2	0h 46m 53s	124	55	64	1
B3	1h 08m 25s	103	77	59	5
B4	0h 55m 55s	123	56	65	0
B5	0h 48m 04s	135	44	63	2
B6	0h 41m 16s	137	43	64	0
B7	1h 21m 17s	114	64	63	3

Tableau 30 : Récapitulatif des temps de post-édition par participant et des nombres de segments par typologie

Pour la traduction avec l'*Adaptive Engine*, le temps total de post-édition était de 1h 08m 10s ; lorsque nous comparons ce temps avec celui employé par les autres participants, nous remarquons que cette phase de traduction est parmi celles qui ont demandé le plus de temps. Par conséquent, l'*Adaptive Engine* ne garantit pas des gains temporels significatifs par rapport au *Baseline Engine*. Nous pouvons fournir des explications à ces données, comme par exemple le fait que la post-édition est plus lente et que le nombre de corrections effectuées est plus élevé. Aussi, il est important de spécifier que le module personnel *Adaptive Engine* était en phase d'entraînement, ce qui peut expliquer la difficulté de mémorisation initiale. Globalement, nous pouvons affirmer que, lors de la première utilisation, l'impact de la fonctionnalité *Adaptive Engine* sur la productivité temporelle par rapport au *Baseline Engine* n'a pas été déterminant. Nous pouvons donc répondre à la deuxième sous-question de recherche : l'*Adaptive Engine* n'assure pas une meilleure productivité temporelle par rapport au *Baseline Engine*.

Concernant la capacité d'apprentissage du système, nous avons remarqué que, grâce à la mise à jour (case « Update » cochée pour les deux types de projet de traduction et active uniquement pour le projet courant), pour les 7 projets, le *Baseline Engine* garantit également la mémorisation des segments précédemment confirmés et donne des résultats presque plus satisfaisants que l'*Adaptive Engine*. En effet, lorsque nous observons la colonne des « 100% », nous retrouvons presque toujours les mêmes résultats. À l'intérieur de chaque projet (et non pas d'un traducteur à l'autre), le *Baseline Engine* assure donc un degré d'intégration des corrections du traducteur parfois meilleur que l'*Adaptive Engine*. Par conséquent,

nous pouvons déduire que la capacité d'adaptation de l'*Adaptive Engine* n'est clairement pas systématique est immédiate : dans 8 cas, par exemple, le système n'arrive pas à mémoriser une correction, même après la troisième apparition à l'intérieur du même projet et même s'il s'agit de deux segments consécutifs (segments 8, 10, 27, 184 ; segments 89, 91, 93 et 201 ; segments 111 et 125 ; segments 141, 147 et 154 ; segments 160 et 161). D'après une analyse des erreurs, nous pouvons aussi remarquer que la fonctionnalité *Adaptive Engine* ne garantit non plus un impact progressif sur la qualité, étant donné que le système répète le même type d'erreurs ou en insère de nouvelles.

En définitive, du moins lors de la première utilisation et à l'intérieur du même projet, l'*Adaptive Engine* ne garantit non plus une performance meilleure par rapport au *Baseline Engine* au niveau de la capacité d'apprentissage des corrections.

Lorsque nous testons la capacité d'apprentissage de l'*Adaptive Engine* en retraduisant une partie des segments (50), nous obtenons des résultats plutôt satisfaisants.

Cependant, à la troisième phase de traduction, lorsque nous soumettons des segments similaires pour traduction, nous remarquons que la capacité d'adaptation du système est plutôt inadéquate et insuffisante : sur 40 segments, 16 seulement ont été traduits correctement. Aussi, d'après le rapport d'analyse, la fonctionnalité de mémorisation n'a concerné que 8 segments. En outre, la qualité de la traduction est parfois dégradée, alors que l'un des objectifs primaires de l'*Adaptive Engine* serait l'amélioration progressive de la traduction sortante.

Sur la base de l'analyse des phases exposées, nous pouvons répondre à notre **deuxième sous-question de recherche** : en général, la fonctionnalité *Adaptive Engine* ne garantit aucune valeur ajoutée par rapport au *Baseline Engine*, ni en ce qui concerne la productivité temporelle, ni en termes de qualité de la traduction. En outre, comme nous l'avons découvert, elle ne donne non plus de résultats satisfaisants en termes de mémorisation et adaptation aux préférences stylistiques et terminologiques du traducteur.

6.1.2 Deuxième partie : évaluation automatique

Nous présentons un tableau récapitulatif des scores **TER**, **BLEU** et **HTER**, ainsi que le temps de post-édition par participant. Nous précisons que les scores ont été calculés sur le fichier en entier, sans donner un poids spécifique aux différents segments en relation à leur longueur. Pour l'analyse des données et le calcul des scores, nous nous sommes tenus à la segmentation effectuée par *SDL Trados*, sur la base de la liste d'abréviations ajoutée.

Pour le calcul du score **TER**, nous avons examiné la distance de post-édition entre la traduction « brute » du système et la traduction de référence, dans ce cas, la traduction publiée du guide utilisateur en italien. Comme nous l'avons déjà mentionné, d'après (Koehn, 2010, p. 238) et (Snover et al., 2006, p. 223), TER compare la traduction « brute » d'un système avec une traduction de référence en termes d'étapes d'édition. Ce score est basé sur la distance de Levenshtein, qui calcule le nombre minimal d'insertions, suppressions et substitutions requises pour transformer la traduction sortante dans une référence. Pour ce type de mesure d'erreur, plus le score est bas, plus la traduction est de bonne qualité.

Pour le calcul du score **BLEU**, nous avons comparé la traduction « brute » avec la traduction publiée, ainsi qu'avec les différentes post-éditions de chaque participant. Comme nous l'avons déjà mentionné, BLEU calcule le nombre de n-grams, de longueur différente, présents dans la traduction « brute » et dans la référence. Par conséquent, « the higher the score, the more the translation correlates to a human translation » (Snover et al., 2006, p. 224). L'objectif était, d'un côté, d'évaluer la qualité de la traduction sortante, et de l'autre, d'analyser le rapport entre effort de post-édition, temps de post-édition et perception de la qualité de la part des participants.

Pour le calcul de **HTER**, en revanche, nous avons examiné la traduction « brute » et les 8 versions post-éditées ; pour le projet A, la traduction a été effectuée avec l'*Adaptive Engine (A)*, tandis que les projets indiqués par la lettre B ont été effectués à l'aide du *Baseline Engine (B)* ; nous obtiendrons ainsi un indicateur de l'effort de post-édition de chaque participant. En outre, nous pourrions évaluer l'incidence de la fonctionnalité *Adaptive Engine* sur l'effort de post-édition, ainsi que comparer la performance des deux modules de traduction (A et B). Dans le tableau, nous avons associé chaque résultat au temps relatif de post-édition employé par chaque participant. Nous essayerons d'établir une corrélation entre **le temps employé** et **l'effort de post-édition**. L'intérêt de cette corrélation est justifié par un certain nombre d'aspects. D'après (Koehn, 2010, pp. 238-239),

“[a] post-editor would try to spend as little time as possible to correct the sentence.
(...) If the output is very muddled, human translators are justified in simply deleting the machine translation and writing a new translation from scratch, instead of fiddling with the broken output”.

PROJET	TER	BLEU	HTER	TEMPS
Tr. « brute » vs. référence	0,42	0,51	n/a	n/a
A		0,84	0,14	1h 08m 10s
B1		0,81	0,16	1h 28m 00s
B2		0,86	0,10	0h 46m 53s
B3		0,84	0,11	1h 08m 25s
B4		0,88	0,09	0h 55m 55s
B5		0,91	0,06	0h 48m 04s
B6		0,93	0,06	0h 41m 16s
B7		0,88	0,10	1h 21m 17s

Tableau 31 : Récapitulatif des scores et des temps de post-édition par projet

Considérons d’abord le résultat du score TER. Lorsque nous examinons notre première donnée (0,42), nous remarquons qu’elle n’est pas très satisfaisante. Ce résultat est confirmé par la grande quantité d’erreurs repérées dans la traduction brute lors de notre analyse manuelle.

Considérons maintenant le score BLEU (0,51). Lorsque nous comparons ce résultat avec celui du TER, nous remarquons qu’il affiche un résultat bien plus appréciable, ce qui rendrait compte d’une qualité initiale très bonne – un important indicateur de l’exploitabilité de la traduction sortante.

Pour résumer, cette première analyse montre des résultats mélangés. D’après le résultat du TER, la distance d’édition entre la traduction « brute » et la référence ne serait suffisamment basse pour justifier l’intérêt de l’intégration de la TA avec post-édition dans le flux de travail. Nous pourrions donc répondre à notre **première sous-question de recherche** : la qualité de la traduction automatique sortante ne serait pas satisfaisante pour justifier l’intégration de la TA dans le flux de travail d’un traducteur, relativement à la traduction de textes techniques.

Cependant, le résultat assez satisfaisant du score BLEU infirme cette première conclusion négative et nous encourage à poursuivre l’analyse. C’est pour cela que, afin d’affiner notre analyse des résultats, nous avons décidé de comparer les deux scores initiaux avec les HTER. En effet, il est important de ne pas considérer les mesures en termes absolus, mais plutôt d’insérer les données dans un contexte plus ample.

D’abord, nous pouvons remarquer que les scores les plus élevés de HTER sont associés à des temps de post-édition majeurs (A, B1, B3, B7). De même, les scores de post-édition les plus bas sont associés aux

temps de post-édition mineurs (B5 et B6). Ces résultats nous permettent de supposer qu'il existe une corrélation positive entre temps employé et effort de post-édition. Nous pourrions opposer à cette constatation le fait que les post-éditeurs ont eu un comportement différent dans l'exécution de la tâche. En effet, nombre de facteurs peuvent influencer la rapidité de post-édition : le niveau d'expérience du traducteur, l'attitude envers l'environnement de travail, le type de technologies utilisées, ainsi que la perception de la traduction automatique, entre autres. Les réponses au questionnaire nous aideront à répondre de manière exhaustive à ces questions.

Lorsque nous comparons le résultat du score TER avec les scores HTER, nous remarquons une baisse d'environ 0,32 en moyenne par participant. Nous pouvons en déduire que, malgré une qualité initiale relativement basse, l'effort de post-édition a été plutôt limité ; par conséquent, la perception de la qualité de la traduction automatique de la part de nos participants (y compris nous-même) a été assez positive. Nous avons trouvé une confirmation de plus de notre hypothèse en analysant les scores HTER et BLEU par participant. En effet, nous observons une corrélation positive entre les deux mesures : à l'amélioration de l'une (réduction progressive du score HTER et donc de l'effort de post-édition) correspond l'amélioration de l'autre (augmentation du score BLEU et donc meilleure perception de la qualité de la traduction). En particulier, lorsque nous observons les résultats des participants B1, B5 et B6, il est intéressant de noter qu'il existe une correspondance entre temps de post-édition, effort de post-édition et qualité de la traduction.

Globalement, malgré le résultat relativement peu satisfaisant du score TER (0,42), nous pouvons conclure que l'effort de post-édition, ainsi que la perception générale de la qualité de la traduction automatique sont tels que l'intérêt de l'intégration de la traduction automatique avec post-édition dans le flux de travail est justifié.

Considérons maintenant spécifiquement la performance de l'*Adaptive Engine*. Notre hypothèse était que, grâce à l'apprentissage incrémental des corrections de l'utilisateur, cette fonctionnalité garantirait une progressive amélioration de la qualité de la traduction sortante, avec une conséquente réduction du temps et de l'effort de post-édition.

Cependant, lorsque nous comparons les résultats obtenus par la post-édition avec l'*Adaptive Engine* par rapport au *Baseline Engine*, nous remarquons que l'impact de la fonctionnalité sur l'effort de post-édition n'a pas été déterminant. En effet, le score HTER pour ce projet est parmi les plus élevés, ce qui rend compte d'un effort de post-édition majeur. Inversement, le score BLEU est parmi les plus bas : la qualité de la

traduction automatique n'a pas été jugée meilleure par rapport aux autres. Encore une fois, nous rappelons que plusieurs facteurs peuvent influencer les résultats de la post-édition : certains traducteurs ont tendance à effectuer beaucoup de changements, parfois non requis, alors que d'autres préfèrent garder, le plus possible, la traduction sortante.

En définitive, les données recueillies nous permettent de répondre à notre **troisième sous-question de recherche** : nous pouvons affirmer que l'effort de post-édition est suffisamment bas pour justifier l'intérêt de l'intégration du flux TA + PE au travail du traducteur. Cependant, lorsque nous comparons la performance de l'*Adaptive Engine* et du *Baseline Engine* en termes d'effort de post-édition et perception de la qualité de la traduction automatique, nous remarquons que le premier ne garantit pas une valeur ajoutée par rapport au deuxième. Notre hypothèse est donc partiellement confirmée.

6.1.3 Troisième partie : récapitulatif des réponses des participants au questionnaire

Nous présentons maintenant le récapitulatif des réponses de nos 7 participants au questionnaire final, ainsi que nos considérations. L'analyse des réponses nous aidera à déterminer la perception des traducteurs envers les logiciels et les fonctionnalités exploités, et finalement à évaluer l'efficacité de l'intégration de la traduction automatique avec post-édition dans le flux de travail du traducteur.

Les premières questions concernaient les connaissances de nos traducteurs en matière de traduction automatique, post-édition et des technologies utilisées dans l'étude. Les suivantes visaient à recueillir l'opinion des participants au regard des typologies textuelles les plus adaptées au flux de travail en examen. Enfin, les questions ouvertes visaient à connaître la perception de nos participants relativement à la qualité de la TA et au flux de travail testé, ainsi que les aspects positifs et les éventuelles difficultés rencontrées lors des projets de traduction.

D'après l'analyse, la totalité des participants connaissait *SDL Trados Studio* et 85,7% d'entre eux avait déjà travaillé avec cet outil d'aide à la traduction. En revanche, la quasi-totalité d'entre eux (85,7%) ne connaissait pas le *SDL Language Cloud*. 85,7% avait déjà travaillé avec la traduction automatique, mais personne n'avait jamais utilisé le *Baseline Engine* auparavant. 85,7% avait des connaissances en post-édition et 57,1% avait déjà travaillé avec cette fonctionnalité.

À la question « En traduction technique, pensez-vous que le flux de travail traduction automatique + post-édition peut atteindre une qualité équivalente au flux classique de traduction humaine + révision ? », 85,7% a répondu « Plutôt oui » et 14,3% « En partie ».

À la question « En traduction technique, pensez-vous que le flux de travail traduction automatique + post-édition peut atteindre une productivité équivalente au flux classique de traduction humaine + révision ? », 57,1% a répondu « En partie », 28,6% « Oui » et 14,3% « Plutôt oui ».

D'après les réponses, les jugements étaient globalement positifs pour les deux questions, un aspect qui justifie l'acceptation à 100% de la proposition de la TA sans retraduction de la part de la totalité des participants. En général, nous pouvons supposer que les remarques parfois « peu enthousiastes » seraient justifiées par une connaissance relativement peu approfondie des capacités concrètes des outils d'aide à la traduction, un aspect qui trouve confirmation dans les erreurs conceptuelles parfois repérées dans les réponses. En effet, une connaissance limitée des outils peut donner lieu souvent à des attentes parfois excessives au regard de la performance des systèmes et, à l'inverse, à des jugements relativement « négatifs ».

Finalement, à la question « Pour quel type de textes parmi ceux énumérés ci-dessous (*techniques et scientifiques, médicaux, économiques et financières, juridiques et littéraires*) pensez-vous que la combinaison de *Baseline Engine* avec post-édition est utile ? », 100% des participants a sélectionné techniques et scientifiques, 57,1% médicaux et économiques et financières, 28,6% juridiques et 0% littéraires. D'après les réponses, nous pouvons déduire une claire préférence pour la première typologie textuelle, qui fait l'objet de notre travail, ainsi qu'un encouragement pour tester d'autres typologies textuelles dans le cadre des travaux futurs.

Passons maintenant aux réponses ouvertes. À la question « Quelles étaient vos impressions lors de la phase de traduction et post-édition avec *Baseline Engine* ? », la plupart des observations étaient très positives. En effet, la traduction sortante a été définie « plutôt bien faite et précise », « souvent correcte, avec très peu de problèmes terminologiques ou de syntaxe », voire « très bonne ». Une constatation intéressante était la suivante : grâce à la qualité déjà très satisfaisante, « le travail de post-édition n'est pas si lourd comme j'avais pensé avant de commencer à faire la post-édition » ; ce commentaire nous fournit une indication clé de la perception de l'utilisateur du flux de travail TA + PE. En outre, dans deux cas, les traducteurs ont aussi spécifié que ce flux de travail est très adéquat pour la traduction des textes techniques. Concernant le *Baseline Engine* spécifiquement, il a été défini « un bon programme de traduction automatique qui aide le traducteur à gagner du temps lorsqu'il effectue une traduction technique », ainsi que « assez performant vu que je n'ai pas post-édité trop de segments » ; dans ces deux

cas, nous remarquons une relation évidente entre qualité de la TA et productivité temporelle et, par conséquent, effort de post-édition.

En revanche, les aspects négatifs qui ont été soulevés concernaient la difficulté du système à traiter les suites de substantifs. Cette difficulté particulière est confirmée lors de l'analyse de la première phase de traduction avec l'*Adaptive Engine*. En outre, concernant le style de la traduction, un participant a remarqué que « La traduction était très basée sur l'Anglais et donc le texte en Italien était un peu artificiel. ». Il faut toutefois préciser que, par nature, les textes techniques accordent très peu de liberté à la créativité du traducteur ; un certain degré d'artificialité est donc acceptable.

En ce qui concerne la traduction effectuée par le logiciel, les principales difficultés rencontrées par nos participants concernaient la présence d'erreurs au niveau grammatical et syntaxique ; comme nous l'avons aussi remarqué lors de notre analyse, les traducteurs ont souligné que « Les accords entre nom et adjectif souvent étaient faux, donc il faut contrôler très attentivement » ; en outre, « La traduction des verbes et certaines unités que en Anglais étaient composés avec une structure qu'on ne retrouve pas en Italien étaient mal traduits. ».

Une observation intéressante concernait l'incohérence du système dans le traitement de la terminologie. En effet, comme nous l'avons aussi remarqué, « Parfois les traductions proposées pour le même terme étaient différentes. »

Même si cet aspect ne sera pas pris en compte dans l'évaluation, deux participants ont aussi remarqué l'incohérence dans la reproduction du formatage entre la source et la cible.

Concernant la modalité d'exécution de la tâche de post-édition, un participant a remarqué que « Si on voulait faire un *post-editing* simplement sur le texte cible, parfois cette option n'était pas possible parce qu'il y avait des erreurs dont on pouvait s'apercevoir seulement en regardant le texte source aussi (ex. : les verbes modaux en anglais qui n'ont pas été traduits, comme *should* ou *may*, traduisibles en italien par « *dovere* » ou le mode conditionnel) » ; cette observation rend compte de l'intérêt d'introduire la possibilité d'effectuer une post-édition uniquement monolingue dans l'environnement *SDL Trados Studio*.

Finalement, un participant a remarqué que « Parfois les solutions que j'écrivais n'ont pas été sauvegardé par le software, donc j'ai changé des segments que j'avais déjà corrigés » ; en effet, à la différence de l'*Adaptive Engine*, le *Baseline Engine* ne devrait théoriquement pas garantir la mémorisation des corrections, un aspect qui pourrait déterminer un certain degré de frustration dans le traducteur.

Cependant, comme nous l'avons déjà mentionné, la fonctionnalité de mise à jour, uniquement activée à l'intérieur du même projet, a permis la mémorisation des segments dans une certaine mesure, avec des résultats presque meilleurs que ceux obtenus avec l'*Adaptive Engine*.

Quant aux aspects positifs, deux participants ont remarqué que les termes techniques récurrents ont généralement été bien traduits, et que « La plupart des mots du langage spécifique étaient déjà traduits, et vu qu'ils se répètent toujours, la TM les traduit directement. » ; il faut toutefois préciser que la mémoire de traduction présente nous a servi uniquement pour ajouter une règle de segmentation (à travers la liste des abréviations) et donc elle n'a pas été mise à jour ; les mémorisations sont uniquement imputables à la fonctionnalité de mise à jour (« Update »). En outre, concernant la capacité d'apprentissage, l'un des aspects positifs qui a été remarqué était « Le fait de ne devoir pas retraduire de segments similaires. ».

Concernant la qualité de la traduction sortante, un participant a remarqué que « La syntaxe était plutôt correcte, ainsi que le lexique. J'ai trouvé rarement des calques de la syntaxe anglaise. » En outre, deux observations rendent compte à nouveau de la relation entre qualité de la traduction sortante et effort de post-édition : « Les traductions proposées étaient très bonnes, ce qui rend la post-édition plus facile. » et « La traduction est de bon niveau et ne nécessite pas de beaucoup de post-édition au niveau de la langue. »

Sur la base de ces réflexions et en nous référant aussi au profil des participants à l'étude, nous allons maintenant tirer nos conclusions.

Comme nous l'avons déjà mentionné, le travail a été conduit avec des traducteurs non professionnels, dont la plupart avait toutefois travaillé avec la traduction automatique et la post-édition, mais qui ne connaissait pas le *Baseline Engine*. Nous précisons aussi qu'aucun des participants (sauf nous-même) n'était étudiant de la mention Technologies de la traduction de la FTI. Concernant leur perception relativement à la performance du flux de travail TA + PE du point de vue de la qualité et de productivité, par rapport au flux classique TH + révision et dans la traduction des textes techniques, les réponses étaient plutôt positives et assez encourageantes. Concernant la perception de l'environnement de travail, l'impression générale était aussi assez positive.

Concernant la typologie textuelle testée, les fréquentes observations positives sur l'adéquation du flux de travail pour la traduction des textes techniques soutiennent notre hypothèse.

Quant à la perception de la qualité de la traduction automatique sortante, les remarques ont été très encourageantes, ainsi que cohérentes avec les résultats des scores. En outre, nombre d'observations

soulevées par nos participants coïncidaient avec celles que nous avons faites lors de notre analyse, ce qui semble confirmer la fiabilité et l'objectivité de l'évaluation.

Sur la base de ces réflexions, nous pouvons répondre à notre **première question de recherche** : nous affirmons que l'intégration de la TA + PE dans le flux de travail d'un traducteur indépendant qui traduit des textes techniques est rentable.

En définitive, notre hypothèse, selon laquelle l'intégration d'un module de traduction automatique avec post-édition dans le flux de travail d'un traducteur indépendant, dans l'environnement de travail d'un système de mémoire de traduction, est rentable pour la traduction de textes techniques, est partiellement confirmée.

7. Conclusion

Dans ce chapitre, nous présentons les conclusions de notre étude phase par phase, ainsi que les principales limites et possibles améliorations du travail.

7.1 Conclusions de l'étude

Dans le présent mémoire, nous nous sommes proposée d'analyser l'efficacité de l'intégration d'un module de traduction automatique avec post-édition au flux de travail d'un traducteur indépendant, dans le cadre d'un système de mémoire de traduction (*SDL Trados Studio 2017 - Freelance version*) ; en particulier, nous avons examiné et comparé l'impact sur la qualité de la traduction et sur la productivité temporelle des modules de traduction automatique *Adaptive Engine* et *Baseline Engine* relativement à la traduction de textes techniques.

L'un des enjeux de l'étude était d'assurer l'objectivité maximale. Afin d'être en mesure de fournir des résultats reproductibles, dans l'évaluation de l'efficacité d'une fonctionnalité dans la traduction automatique, nous avons sélectionné des données expérimentales les plus simples et facilement mesurables possibles – en l'occurrence, un texte technique issu de guides utilisateur pour l'imprimante, ayant un taux de répétitions internes suffisamment élevé.

Après avoir présenté la traduction automatique et la post-édition, les relatifs aspects théoriques et les implications sur la pratique, ainsi que l'état de l'art en la matière, nous avons exposé notre expérience. L'étude comportait 3 étapes principales : traduction et post-édition, évaluation de la qualité de la TA, questionnaire.

D'abord, nous avons comparé la performance des deux fonctionnalités du module de TA à l'aide de 8 étudiants en Maîtrise de traduction (y compris nous-même). Ensuite, nous avons évalué la qualité de la traduction automatique. Dans l'étude, une place très importante a été aussi accordée à l'opinion de nos participants sur les technologies et le flux de travail testés, ainsi qu'à leur perception personnelle de la qualité de la traduction automatique par rapport à la typologie textuelle en examen ; à cet effet, nous avons prévu un questionnaire final.

Concrètement, nous avons analysé et comparé l'efficacité des deux modules de traduction en relation à la productivité temporelle et à la qualité de la traduction. Dans le cas de l'*Adaptive Engine*, nous avons évalué plus spécifiquement la capacité d'apprentissage et d'adaptation aux préférences stylistiques et terminologiques de l'utilisateur.

D'après les résultats, nous avons observé que la capacité d'apprentissage et d'adaptation de l'*Adaptive Engine* ne sont globalement pas assez satisfaisantes pour justifier son intégration. En outre, par rapport au *Baseline Engine*, l'*Adaptive Engine* ne garantit pas une performance meilleure ni au niveau de productivité temporelle, ni au niveau de la capacité d'apprentissage et d'adaptation.

En définitive, nous avons remarqué que l'*Adaptive Engine* ne représente pas une valeur ajoutée par rapport au *Baseline Engine* dans le cadre de notre expérience.

Pour vérifier l'intérêt effectif de l'intégration d'un module de traduction automatique, nous avons évalué la qualité de la traduction sortante à l'aide de mesures automatiques. L'objectif était, d'un côté, d'évaluer la distance de post-édition, afin de déterminer la qualité de la traduction brute ; de l'autre, de vérifier l'effort de post-édition en relation au temps employé, ainsi qu'au comportement et à la perception de nos participants. Nous avons comparé les résultats et nous les avons analysés à la lumière de nos réflexions.

Lorsque nous analysons la qualité de la traduction brute par rapport à une traduction de référence (la traduction en italien publiée du guide utilisateur) à l'aide de la mesure automatique TER, nous obtenons des résultats peu satisfaisants, en correspondance avec notre analyse manuelle. Cependant, le score BLEU affiche un résultat bien plus satisfaisant et nous encourage à poursuivre l'analyse. En effet, lorsque nous considérons les versions post-éditées par nos participants comme référence, nous remarquons que la perception de la qualité s'améliore. Cela est confirmé par la progressive amélioration des scores HTER et BLEU, qui dénotent une réduction de l'effort de post-édition et une meilleure perception de la qualité de la traduction. Nous pouvons déduire qu'il n'existe pas de corrélation entre qualité de la traduction automatique (évaluée à l'aide de mesures automatiques) et perception de la qualité de la traduction automatique de la part du traducteur humain. De plus, les mesures automatiques n'assignent pas un poids différent aux typologies d'erreurs sur la base de leur « gravité ». En définitive, les résultats des mesures automatiques, et en particulier du score TER, ne suffisent pas pour juger la qualité objective et l'exploitabilité de la traduction automatique : c'est la perception du traducteur humain qui prime.

En outre, d'après les réponses données par nos traducteurs au questionnaire final, la perception de la qualité de la traduction automatique fournie par le *Baseline Engine* est tout à fait positive. En particulier, la majorité des participants considère que le flux de travail TA + PE garantirait des résultats plus satisfaisants par rapport au flux classique TH + révision au niveau de la productivité et de la qualité, relativement à la traduction des textes techniques. Ces résultats confirment notre analyse.

Concernant l'environnement de travail, ainsi que le type de technologies et outils testés, l'impression de nos participants était positive.

En définitive, nous pouvons affirmer que **notre hypothèse est partiellement vérifiée** : l'intégration d'un module de traduction automatique (le *Baseline Engine*) avec post-édition dans le flux de travail d'un traducteur indépendant dans le cadre d'un système de mémoire de traduction et relativement à la traduction des textes techniques est rentable.

7.2 Limites, améliorations et perspectives futures

Premièrement, nous avons effectué l'expérience en testant qu'une seule paire de langue et une seule direction de traduction (En-US → It-IT). En ce qui concerne le module de traduction automatique fourni par le *SDL Language Cloud*, et plus spécifiquement la fonctionnalité *Adaptive Engine*, il serait intéressant de tester son comportement avec d'autres paires de langue et d'autres directions de traduction, afin d'examiner l'impact sur la qualité de la traduction sortante, ainsi que sur sa capacité d'apprentissage des corrections de l'utilisateur.

Deuxièmement, nous n'avons pris en compte qu'une seule typologie textuelle ; notre fichier, nous le rappelons, était extrait d'un guide utilisateur pour une imprimante ; en outre, il était vraisemblablement rédigé en langage contrôlé, ce qui a inévitablement eu un impact positif sur le traitement de la part du logiciel et finalement sur la qualité de la traduction sortante. Dans les travaux futurs, il serait donc intéressant de tester d'autres types de textes issus de domaines différents, par exemple économique et financier, juridique, ou encore général, pour citer quelques exemples. Spécifiquement en ce qui concerne la fonctionnalité *Adaptive Engine*, il serait intéressant de tester son comportement avec d'autres typologies textuelles, ainsi que déterminer l'impact du type de texte sur sa capacité d'intégration des corrections.

En outre, il serait intéressant d'analyser la qualité effective des versions post-éditées par nos participants et leur relation avec le temps de post-édition, afin de déterminer la qualité réelle du texte final.

En ce qui concerne le profil des participants, nous nous sommes limitée à des étudiants en Maîtrise de traduction entre 23 et 27 ans. Dans les prochains travaux, il serait intéressant de conduire l'étude sur d'autres catégories de traducteurs, notamment des traducteurs professionnels, afin d'analyser leur degré d'acceptation du flux de travail TA + PE, ainsi que leur perception de l'environnement de travail. Leur opinion représenterait un avis supplémentaire pour déterminer l'intérêt et l'efficacité de ce flux de travail.

Afin de rendre l'expérience plus « vraisemblable » et plus proche des conditions de travail réelles, il serait aussi intéressant de tester le flux de travail en intégrant aux projets une mémoire de traduction et éventuellement une base de données terminologique, afin d'étudier, d'un côté, le comportement du traducteur face à d'autres sources de suggestions de traduction, et de l'autre, plus généralement, l'intérêt et l'efficacité de l'intégration dans le flux de travail global.

En conclusion, il serait aussi intéressant d'introduire l'évaluation humaine dans l'analyse de la qualité de la traduction automatique. Nous pourrions ainsi mesurer l'accord inter-annotateurs, ainsi que, éventuellement, la possible corrélation des jugements humaines avec les résultats des mesures automatiques.

Bibliographie

- Allen, J. (2003). Post-editing. In *Computers and translation: a translator's guide* (pp. 297-317). Amsterdam: Harold Somers.
- Bouillon, P. (Éd.). (1993). *La traductique: études et recherches de traduction par ordinateur*. Montréal: Les Presses de l'Université de Montréal.
- Bouillon, P. (2016). *Notes de cours Traduction automatique*. Non publié, Université de Genève.
- Callison-Burch, C., Fordyce, C., Koehn, P., Monz, C., & Schroeder, J. (2007). (Meta-) Evaluation of Machine Translation. In *ACL 2007: proceedings of the Second Workshop on Statistical Machine Translation* (pp. 136-158). Prague, Czech Republic: Association for Computational Linguistics.
- Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is Neural Machine Translation the New State of the Art? *The Prague Bulletin of Mathematical Linguistics*, (108), 109-120.
- Daems, J., Vandepitte, S., Hartsuiker, R., & Macken, L. (2015). The Impact of Machine Translation Error Types on Post-Editing Effort Indicators. In *Proceedings of 4th Workshop on Post-Editing Technology and Practice (WPTP4)* (pp. 31-45). Miami, USA.
- Fiederer, R., & O'Brien, S. (2009). Quality and Machine Translation: A realistic objective? *Journal of Specialised Translation*, (11).
- Guerberof Arenas, A. (2009). Productivity and quality in MT post-editing. In *MT Summit XII - Workshop: Beyond Translation Memories: New Tools for Translators MT*. Ottawa, Ontario, Canada.
- Hassan, H., Aue, A., Chen, C., Chowdhary, V., Clark, J., Federmann, C., ... Zhou, M. (2018). Achieving Human Parity on Automatic Chinese to English News Translation. In *Microsoft AI & Research*. Consulté à l'adresse <https://arxiv.org/pdf/1803.05567.pdf>
- Hofmann, N., & Lepad, M. (2017). MT in real-world practice: Challenges and solutions at Swiss Federal Railways. In *Proceedings of The 20th Annual Conference of the European Association for Machine Translation* (pp. 36-40).
- Hutchins, W. John. (1986). *Machine translation: past, present, future*. Chichester [West Sussex] : New York: Ellis Horwood ; Halsted Press.
- Hutchins, W. John. (1996). ALPAC: the (in)famous report. *MT News International*, (14), 9-12.
- Hutchins, W. J., & Somers, H. L. (1997). *An introduction to machine translation* (2. printing). London: Academic Press.
- Jacqmin, L. (1993). Classification générale des systèmes de traduction automatique. In *La traductique: études et recherches de traduction par ordinateur* (pp. 43-63). Montréal: Les Presses de l'Université de Montréal.
- Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition* (2. ed., Pearson internat. ed). Upper Saddle River, NJ: Prentice Hall, Pearson Education Internat.

- Kit Chunyu, & Tak-Ming, W. B. (2015). Evaluation in machine translation and computer-aided translation. In *Routledge Encyclopedia of Translation Technology* (pp. 213-236). Abingdon, Oxon: Routledge.
- Koehn, P. (2010). *Statistical machine translation*. Cambridge ; New York: Cambridge University Press.
- Koehn, P. (2017). Draft of Chapter 13: Neural Machine Translation. In *Statistical Machine Translation*.
- Koehn, P., & Germann, U. (2014). The Impact of Machine Translation Quality on Human Post-editing. In *Proceedings of the EACL 2014 Workshop on Humans and Computer-assisted Translation* (pp. 38-46). Gothenburg, Sweden: Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-0307>
- Koehn, P., & Monz, C. (2006). Manual and Automatic Evaluation of Machine Translation between European Languages. In *Proceedings of the Workshop on Statistical Machine Translation (HLT-NAACL)* (pp. 102-121). New York, USA.
- Koponen, M. (2012). Comparing human perceptions of post-editing effort with post-editing operations. In *Proceedings of the 7th Workshop on Statistical Machine Translation (WMT 2012)* (pp. 181-190). Montréal, Canada.
- Koponen, M. (2015). How to teach machine translation post-editing? Experiences from a post-editing course. In *Proceedings of 4th Workshop on Post-Editing Technology and Practice (WPTP4)* (pp. 2-15). Miami, USA.
- Koponen, M., Aziz, W., Ramos, L., & Specia, L. (2012). Post-editing time as a measure of cognitive effort. In *Proceedings of the Workshop on Post-Editing Technology and Practice (WPTP 2012)* (pp. 20-29). San Diego, California, USA: Sharon O'Brien, Michel Simard and Lucia Specia.
- Krauwier, S. (1991). Evaluation of MT systems : A programmatic view. *Proceedings of the Evaluators' Forum*, April 21st-24th, 1991, Les Rasses, Vaud, Switzerland, ed. Kirsten Falkedal. (Geneva: ISSCO); pp. 245-253.
- Krings, H. P., & Koby, G. S. (2001). *Repairing texts: empirical investigations of machine translation post-editing processes*. Kent, Ohio: Kent State University Press.
- Levin, P., Dhanuka, N., & Khalilov, M. (2017). Machine Translation at Booking.com: Journey and Lessons Learned. In *Proceedings of The 20th Annual Conference of the European Association for Machine Translation* (pp. 80-85). Prague, Czech Republic.
- Loffler-Laurian, A.-M. (1994). La traduction automatique : son utilisation par le « grand public ». *Langages*, 28(116), 87-94. <https://doi.org/10.3406/lgge.1994.1697>
- Nyberg, E., Mitamura, T., & Huijsen, W.-O. (2003). Controlled language for authoring and translation. In *Computers and translation: a translator's guide* (pp. 245-281). Amsterdam: Benjamins.
- O'Brien, S. (2002). Teaching Post-editing: A Proposal for Course Content. *Sixth EAMT Workshop "Teaching machine translation"* (pp.99-106). Manchester, England.
- O'Brien, S. (2004). Machine Translatability and Post-Editing Effort: How do they relate? In *Proceedings of the twenty-sixth International Conference on Translating and the Computer (Translating and the Computer 26)*. London: Aslib 2004.

- O'Brien, S. (2011). Towards predicting post-editing productivity. *Machine translation*, 25(3), 197.
- O'Brien, S., & de Almeida, G. (2010). Analysing Post-Editing Performance: Correlations with Years of Translation Experience. In *Proceedings of the 14th Annual Conference of the European Association for Machine Translation (EAMT 2010)*. Saint-Raphaël, France.
- O'Curran, E. (2014). Translation Quality in Post-Edited versus Human-Translated Segments: A Case Study. In *Proceedings of the eleventh conference of the Association for Machine Translation in the Americas (AMTA 2014)* (pp. 113-118). Vancouver, BC.
- Poibeau, T. (2017). *Machine translation*. Cambridge, Massachusetts: The MIT Press.
- Quah, C. K. (2006). *Translation and technology*. Houndmills [England] ; New York: Palgrave Macmillan.
- SDL. (2017a). SDL Certification: Post-Editing Certification.
- SDL. (2017b). The dawn of a new horizon - Adaptive MT White Paper.
- Shterionov, D., Nagle, P., Casanellas, L., Superbo, R., & O'Dowd, T. (2017). Empirical evaluation of NMT and PBSMT quality for large-scale translation production. In *Proceedings of the 20th Annual Conference of the European Association for Machine Translation* (p. 74-79). Prague, Czech Republic.
- Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Markhoul, J. (2006). A Study of Translation Edit Rate with Targeted Human Annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas, "Visions for the Future of Machine Translation"* (pp. 223-231). Cambridge, Massachusetts, USA.
- Somers, H. (Éd.). (2003). *Computers and translation: a translator's guide*. Amsterdam: Benjamins.
- Somers, H. L. (1993). La traduction automatique basée sur l'exemple ou sur le corpus. In *La traductique: études et recherches de traduction par ordinateur* (pp. 149-166). Montréal: Les Presses de l'Université de Montréal.
- Systran. (2017). Systran Pure Neural Machine Translation - White Paper. Consulté à l'adresse <http://www.systransoft.com/systran/resources/white-papers/>
- Tatsumi, M. (2009). Correlation between automatic evaluation metric scores, post-editing speed and some other factors. In *Proceedings of MT Summit XII* (pp. 332-339). Ottawa, Ontario, Canada.
- Trujillo, A. (1999). *Translation engines: techniques for machine translation*. London ; New York: Springer.
- Turian, J., Shen, L., & Melamed, I. D. (2003). Evaluation of machine translation and its evaluation. In *Proceedings of the MT Summit IX* (pp. 386-393). New Orleans, USA.
- Vasconcellos, M. (1987a). A comparison of MT postediting and traditional revision. In *Proceedings of the 28th Annual Conference of the American Translators Association* (pp. 409-415). Albuquerque, New Mexico: Medford, N.J. : Learned Information.
- Vasconcellos, M. (1987b). Post-editing on screen: machine translation from Spanish into English. In *A profession on the move: proceedings of translating and the computer 8* (pp. 133-146).

Vilar, D., Schneider, M., Burchardt, A., & Wedde, T. (2012). Towards the Integration of MT into a LSP Translation Workflow. In *EAMT 2012: Proceedings of the 16th EAMT Conference*. Trento, Italy.

Way, A. (2010). A critique of Statistical Machine Translation. In *Evaluation of Translation Technology* (pp. 17- 39). Antwerp; Brussels: Artesis University College, Dept. of Translators & Interpreters ; University Press Antwerp.

Wu, Y., Schuster, M., Chen, Z., Le, Q., Norouzi, M. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. CoRR, vol. abs/1609.08144 (2016)

Zaretskaya, A., Corpas Pastor, G., & Seghiri, M. (2015). Integration of Machine Translation in CAT Tools: State of the Art, Evaluation and User Attitudes. *SKASE Journal of Translation and Interpretation [online]*, 8(1).

Zhechev, V. (2012). Machine Translation Infrastructure and Post-editing Performance at Autodesk. In *Proceedings of the Workshop on Post-Editing Technology and Practice (WPTP 2012)* (pp. 94- 103). San Diego, California, USA: Sharon O'Brien, Michel Simard and Lucia Specia.

Sitographie

Adaptive Engine, SDL Trados : <https://www.sdltrados.com/products/trados-studio/adaptivemt/> (dernière consultation : 06/03/2018)

ASD Simplified Technical English : <http://asd-ste100.org/about.html> (dernière consultation : 09/03/2018)

Autodesk : <https://www.autodesk.ch/fr> (dernière consultation : 16/04/18)

BB Flashback : http://www.bbflashback.it/BBFlashBack_FreePlayer.aspx (dernière consultation : 16/03/2018)

HP Deskjet 1510 All-in-One Printer – User Guides : <https://support.hp.com/us-en/product/hp-deskjet-1510-all-in-one-printer-series/5157551/model/5157552/manuals> (dernière consultation : 26/07/18)

MetricsMaTr10 – The NIST Metrics for MACHine TRANslation 2010 Challenge : <https://www.nist.gov/sites/default/files/documents/itl/iad/mig/NISTMetricsMaTr10EvalPlan.pdf> (dernière consultation : 27/06/18)

Random.org – True Random Number Service : <https://www.random.org> (dernière consultation : 28/05/2018)

SDL Gateway : <https://gateway.sdl.com/apex/communityknowledge?articleName=000002881> (dernière consultation : 23/04/18)

SDL Language Cloud : <https://www.sdl.com/it/languagecloud/> (dernière consultation : 06/03/2018)

SDL Trados : <https://www.sdltrados.com/products/trados-studio/> (dernière consultation : 06/03/2018)

MT Post-Editing Guidelines – TAUS : <https://www.taus.net/academy/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines>

The Official Microsoft Blog : <https://blogs.microsoft.com/ai/machine-translation-news-test-set-human-parity/> (dernière consultation : 02/05/18)

Tilde Custom Machine Translation : <https://www.letsmt.eu/Bleu.aspx> (dernière consultation : 24/07/18)

Translation Edit Rate : <http://www.cs.umd.edu/~snover/tercom/> (dernière consultation : 24/07/18)

Index des figures

Figure 1 : Classification des types de traduction, Hutchins & Somers, 1997, p. 148..... 8

8. Annexes

I Protocole de travail

- 1) Créer une mémoire de traduction (vide, à ne pas mettre à jour). Ajouter la liste des abréviations.
- 2) Créer un nouveau projet de traduction. Ajouter le fichier à traduire. Spécifier la direction de traduction (English US → Italiano IT). Ajouter la mémoire de traduction vide (spécifier « à ne pas mettre à jour »). Cocher la case « Update ». Ajouter le module de traduction automatique *Baseline Engine*.
- 3) Effectuer la pré-traduction du fichier.
- 4) Exporter et sauvegarder la version pré-traduite (traduction brute).
- 5) Analyser la traduction brute sur la base des catégories d'erreurs définies.

Traduction avec l'Adaptive Engine

- 1) Créer un nouveau projet de traduction. Ajouter le fichier à traduire. Spécifier la direction de traduction (English US → Italiano IT). Ajouter la mémoire de traduction vide (spécifier « à ne pas mettre à jour »). Cocher la case « Update ». Ajouter le module de traduction automatique *Adaptive Engine*.
- 2) Avant de commencer à post-éditer, démarrer l'enregistrement d'écran.
- 3) Post-éditer le fichier.
- 4) Une fois la post-édition terminée, fermer l'enregistrement et sauvegarder le fichier vidéo.
- 5) Exporter et sauvegarder également la version finale post-éditée.

Test de mémorisation – Adaptive Engine

- 1) Générer une séquence randomisée de 50 numéros de 1 à 244.
- 2) Sélectionner les segments correspondants dans la source et créer le fichier « artificiel ».
- 3) Créer un nouveau projet de traduction. Ajouter le fichier à traduire. Spécifier la direction de traduction (English US → Italiano IT). Ajouter la mémoire de traduction vide (spécifier « à ne pas mettre à jour »). Cocher la case « Update ». Ajouter le module de traduction automatique *Adaptive Engine*.
- 4) Effectuer la pré-traduction du fichier.

5) Exporter et sauvegarder la version pré-traduite.

Test d'adaptation – Adaptive Engine

1) Sélectionner 40 segments avec terminologie et tournures similaires à celles déjà traduites lors de deux premières phases de traduction et créer le deuxième fichier « artificiel ».

2) Créer un nouveau projet de traduction. Ajouter le fichier à traduire. Spécifier la direction de traduction (English US → Italiano IT). Ajouter la mémoire de traduction vide (spécifier « à ne pas mettre à jour »). Cocher la case « Update ». Ajouter le module de traduction automatique Adaptive Engine.

3) Effectuer la pré-traduction du fichier.

4) Exporter et sauvegarder la version pré-traduite.

Traduction avec le Baseline Engine

1) Créer une mémoire de traduction (vide, à ne pas mettre à jour). Ajouter la liste des abréviations.

2) Créer un nouveau projet de traduction.

3) Ajouter le fichier à traduire. Spécifier la direction de traduction (English US → Italiano IT). Ajouter la mémoire de traduction vide (spécifier « à ne pas mettre à jour »). Cocher la case « Update ». Ajouter le module de traduction automatique *Baseline Engine*.

4) Avant de commencer à post-éditer, démarrer l'enregistrement d'écran.

5) Post-éditer le fichier.

6) Une fois la post-édition terminée, fermer l'enregistrement d'écran et sauvegarder le fichier vidéo.

7) Exporter et sauvegarder la version finale post-éditée.

8) Répondre au questionnaire.

Evaluation à l'aide des mesures automatiques

1) Comparer la version brute avec la référence publiée (IT). Calculer TER et BLEU.

2) Comparer la version brute avec les 8 versions post-éditées. Calculer HTER et BLEU. Analyser la relation entre les scores et le temps de post-édition.

Questionnaire

1) Analyser les réponses au questionnaire à la lumière des résultats obtenus lors des premières phases et en relation à l'hypothèse et aux questions de recherche.

II Instructions

Dans le cadre du présent travail, vous allez effectuer une post-édition bilingue d'un fichier issu d'un guide utilisateur pour l'imprimante de l'anglais vers l'italien. L'environnement de travail est le système de mémoire de traduction *SDL Trados Studio 2017*, et le fournisseur de la traduction automatique est *SDL Language Cloud*. Dans le travail, vous aurez la possibilité d'exploiter le module de traduction automatique *Baseline Engine*.

Nous vous demandons de post-éditer (ou retraduire) chaque segment proposé. Nous vous prions de reconstrôler et confirmer chaque segment avant de passer au successif. Nous vous demandons également de ne vous attarder trop longtemps sur l'adéquation du formatage, ainsi que sur la terminologie. Vous pouvez contrôler le progrès de la traduction et le pourcentage de segments restants dans l'onglet en bas à droite.

Lorsque vous décidez d'accepter la proposition de la traduction automatique et de post-éditer un segment, nous vous prions de vous tenir aux indications suivantes ; il s'agit des directives TAUS relatives à la post-édition complète. Nous vous prions de lire attentivement les directives avant de procéder avec la post-édition.

DIRECTIVES TAUS RELATIVES À LA POST-ÉDITION « COMPLÈTE »

- Aim for grammatically, syntactically and semantically correct translation.
- Ensure that key terminology is correctly translated and that untranslated terms belong to the client's list of "Do Not Translate" terms.
- Ensure that no information has been accidentally added or omitted.
- Edit any offensive, inappropriate or culturally unacceptable content.
- Use as much of the raw MT output as possible.

- Basic rules regarding spelling, punctuation and hyphenation apply.
- Ensure that formatting is correct.

Cependant, nous vous demandons de ne pas tenir compte du respect du formatage ; en effet, cet aspect ne sera pas indispensable aux fins de l'évaluation de la qualité de la traduction automatique.

Avant de commencer l'expérience, démarrez l'enregistrement d'écran. Une fois que vous aurez terminé, fermez l'enregistrement et sauvegardez le fichier vidéo.

Pour chaque segment, le logiciel de traduction automatique donnera une suggestion de traduction. Cette suggestion est observable dans le panneau en haut, et sera signalée avec l'acronyme AT (en anglais, « Automatic Translation ») sur fond bleu. Après avoir lu le segment source et la relative suggestion de traduction, vous aurez le choix entre accepter la proposition telle quelle – qui pourra être insérée automatiquement avec un raccourci et ensuite confirmée – ou la post-éditer, ou encore la rejeter en retraduisant le segment dans son entièreté. Lorsque vous déciderez de post-éditer la phrase, le fond du symbole « AT » changera sa couleur de bleu à gris, ce qui indique qu'une modification a eu lieu.

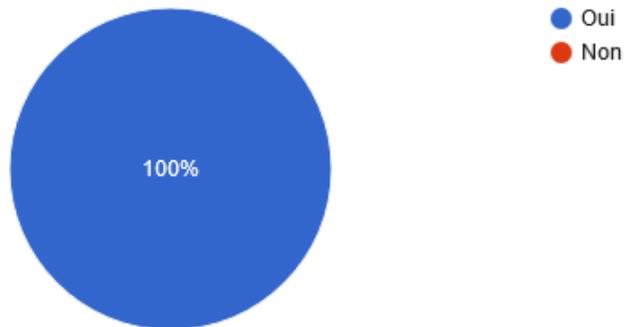
Finalement, nous vous demandons de répondre au questionnaire qui vous sera envoyé par courriel électronique. Les questions portent sur le niveau de vos connaissances préalables en matière d'outils d'aide à la traduction, traduction automatique et post-édition, ainsi que sur votre éventuelle expérience préalable avec dits outils et tâches.

Nous vous remercions de votre gentille participation.

III Questionnaire

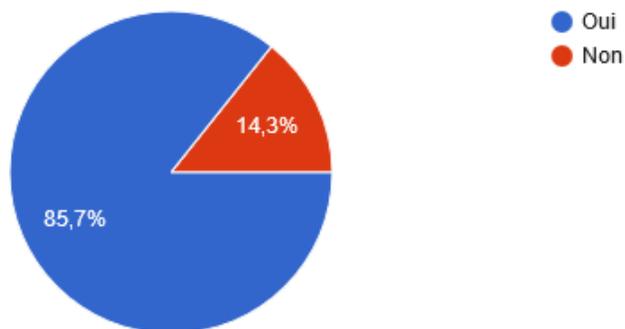
Est-ce que vous connaissez SDL Trados Studio (système de mémoire de traduction)?

7 risposte



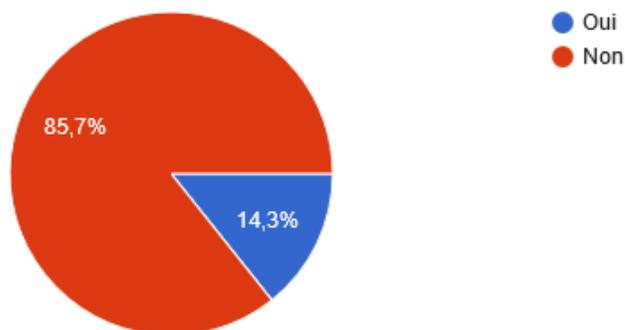
Est-ce que vous avez déjà travaillé avec SDL Trados Studio?

7 risposte



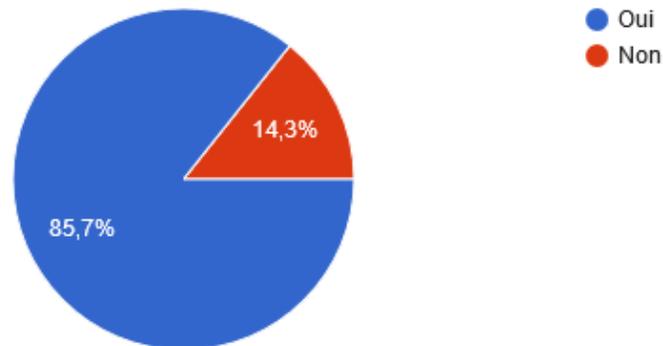
Est-ce que vous connaissez le SDL Language Cloud?

7 risposte



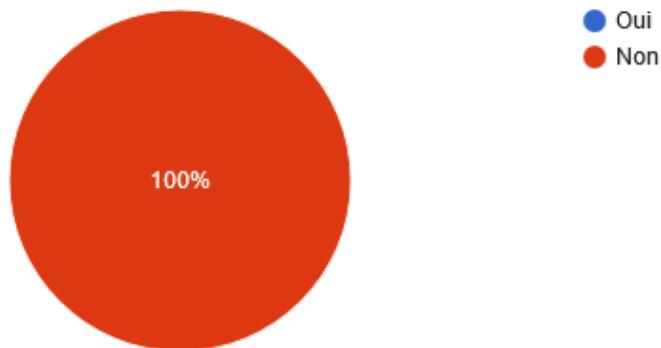
Est-ce que vous avez déjà travaillé avec la traduction automatique?

7 risposte



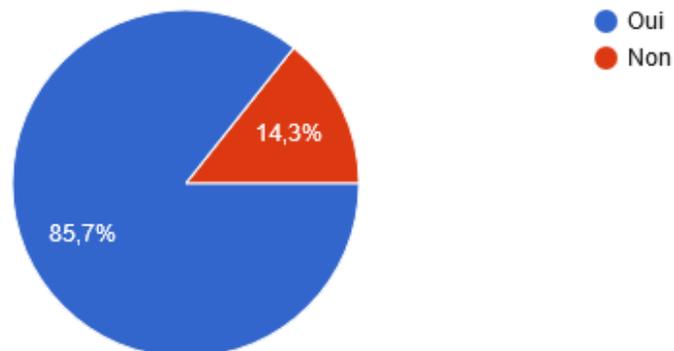
Est-ce que vous avez déjà travaillé avec la fonctionnalité Baseline Engine?

7 risposte



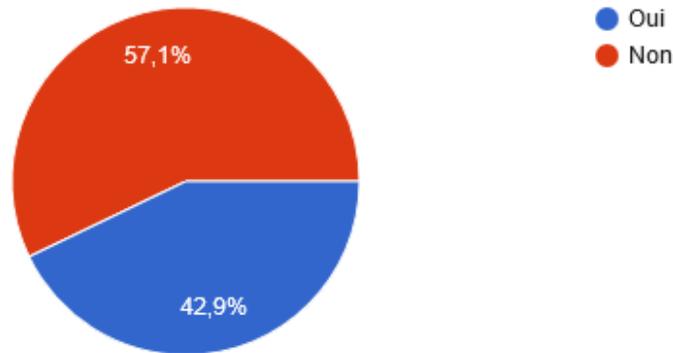
Avez-vous des connaissances en post-édition?

7 risposte



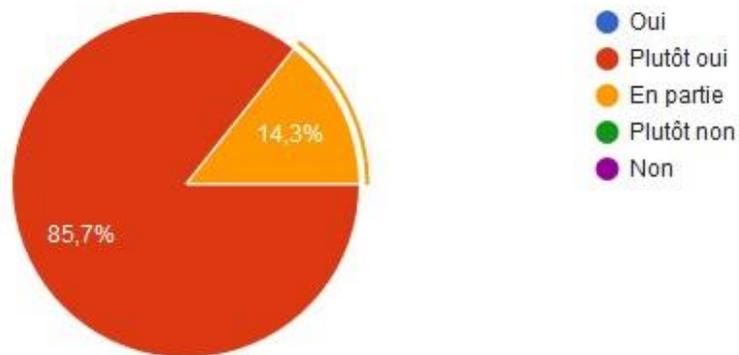
Est-ce que vous avez déjà travaillé avec la post-édition?

7 risposte



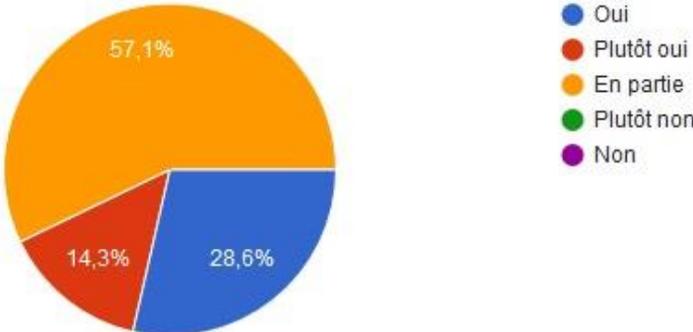
En traduction technique, pensez-vous que le flux de travail traduction automatique + post-édition peut atteindre une qualité équivalente au flux classique de traduction humaine + révision?

7 risposte



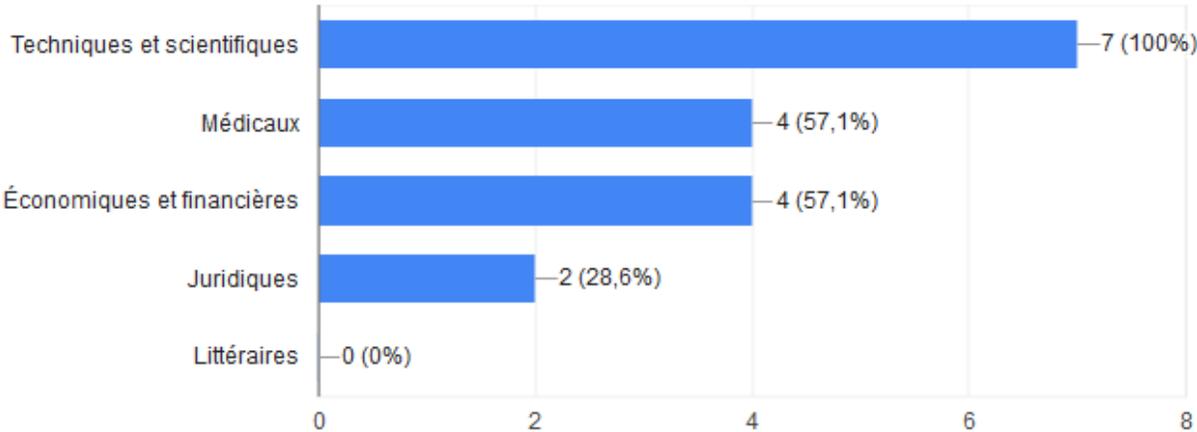
En traduction technique, pensez-vous que le flux de travail traduction automatique + post-édition peut atteindre une productivité équivalente au flux classique de traduction humaine + révision?

7 risposte



Pour quel type de textes parmi ceux énumérés ci-dessous pensez-vous que la combinaison de Baseline Engine avec post-édition est utile? (Plus d'une réponse possible)

7 risposte



Quelles étaient vos impressions lors de la phase de traduction et post-édition avec Baseline Engine?

7 risposte

Traduction automatique plutôt bien faite et précise. Parfois il y avait des problèmes de traduction avec des longues "noun strings" dans le texte source

Ils marchent très bien avec les textes techniques

La traduction automatique m'a fourni des traductions déjà très bonnes, donc le travail de post-édition n'est pas si lourd comme j'avais pensé avant de commencer à faire la post-édition.

La traduction proposée était souvent correcte, avec très peu de problèmes terminologiques ou de syntaxe.

La traduction était très basée sur l'Anglais et donc le texte en Italien était un peu artificiel.

il s'agit d'un bon programme de traduction automatique qui aide le traducteur à gagner du temps lorsque une traduction technique.

Je l'ai trouvé assez performant vu que je n'ai pas post-édité trop de segments

Quelles difficultés avez-vous rencontrées?

7 risposte

Si on voulait faire un post-editing simplement sur le texte cible, parfois cette option n'était pas possible parce qu'il y avait des erreurs dont on pouvait s'apercevoir seulement en regardant le texte source aussi (e.g: modal verbs en anglais qui n'ont pas été traduits, comme should or may)

Les accords entre nom et adjectif seurent étaient faux, donc il faut contrôler très attentivement

Quelques fois, les erreurs concernant la forme n'étaient pas soulignées, ce que je pensais aurait été remarqué : donc, j'ai du faire plus d'attention.

Parfois les traductions proposées pour le même terme étaient différentes. Il y a des incohérences dans le formatage.

La traduction des verbes et certaines unités que en Anglais étaient composés avec une structure qu'on ne retrouve pas en Italien et donc était mal traduits.

Du point de vu de la grammaire et de la syntaxe des phrases

Parfois les solutions que j'écrivais n'ont pas été sauvegardé par le software, donc j'ai changé des segments que j'avait déjà corrigés

Quels étaient, au contraire, les aspects positifs?

7 risposte

Les termes techniques récurrents ont généralement été bien traduits

La plupart des mots du langage spécifique étaient déjà traduits, et vu qu'ils se repètent toujours, la TM les traduit directement.

Les traductions proposées étaient très bonnes, ce qui rend la post-édition plus facile.

La traduction est de bon niveau et ne nécessite pas de beaucoup de post-édition au niveau de la langue.

Le fait de ne devoir pas retraduire de segments similaires.

les termes techniques étaient toujours bien traduits

La syntaxe était plutôt correcte, ainsi que le lexique. J'ai trouvé rarement des calques de la syntaxe anglaise.

Remerciements

D'abord, un très gros merci à ma directrice de mémoire Madame Marianne Starlander qui m'a guidée tout au long de cette expérience et qui, depuis le Baccalauréat, m'a transmis la passion pour les technologies de la traduction. Avec ses précieux conseils et ses attentives relectures, elle a permis la réalisation du présent travail.

Merci à Madame Pierrette Bouillon, qui a accepté de devenir ma jurée.

Merci à mes amis lointains mais proches à mon cœur.

Merci aux participants à l'expérience pour leur courtoisie et collaboration dans l'étude.

Finalement, merci à ma famille, pour le soutien et l'encouragement constants. Je vous en saurais gré à vie.