

Le projet européen MultiTraiNMT

Un premier retour sur les usages et les besoins au sein des formations universitaires françaises

Caroline Rossi

Université Grenoble Alpes, ILCEA4, France

Nicolas Ballier

Université Paris Cité, CLILLAC-ARP, France

Amélie Josselin-Leray

Université Toulouse Jean Jaures, CLLE, France

Sara Salmi, Hanna Martikainen

Université Sorbonne Nouvelle, CLESTHIA, France

Katell Hernandez-Morin

Université Rennes 2, LIDILE, France

Rudy Loock

Université Lille, STL, France

Abstract The MultiTraiNMT strategic partnership aims to develop, evaluate and disseminate open-access content to improve the teaching and learning of machine translation among language learners, language teachers, translator trainees, translation teachers and professional translators throughout Europe. This contribution introduces the project and its main deliverables, before analyzing several use cases in French universities.

Keywords Neural machine translation. Machine translation literacy. Translator training. MultiTraiNMT. Language teaching.

Sommaire 1 Introduction. – 2 Le sondage. – 3 Zoom sur le Master TIM de l'Université de Toulouse Jean Jaurès. – 4 Zoom sur la proposition de formation continue de l'ESIT. – 5 Conclusion.

Le partenariat stratégique MultiTraiNMT s'est donné pour but de développer, d'évaluer et de diffuser des contenus en libre accès afin d'améliorer l'enseignement et l'apprentissage de la traduction automatique chez les apprenant·e·s en langues, les professeurs de langues, les apprentis traducteurs et traductrices, les professeurs de traduction et les traducteurs et traductrices professionnel·le·s, dans toute l'Europe.

Après avoir présenté le projet et ses principaux livrables, nous analysons plusieurs cas d'usage. Pour chacune des formations partenaires du projet, l'analyse répond à trois questions que le projet MultiTraiNMT a traitées au moins en partie. La première concerne le rôle et la place d'un enseignement théorique sur la traduction automatique (histoire de la TA, explication de son fonctionnement) aux côtés de la formation aux outils et aux pratiques de post-édition : ces enseignements sont-ils assurés séparément, et le cas échéant, quelle est la part de l'un et de l'autre ? La deuxième question s'intéresse à la prise en main de solutions de traduction automatique susceptibles d'être entraînées et testées par les étudiant·e·s et/ou leurs enseignant·e·s. Enfin, nous nous interrogeons sur la formation des formateurs et des formatrices : l'équipe enseignante a-t-elle été formée à la traduction automatique, ou s'appuie-t-on sur un·e ou plusieurs expert·e·s ? Les présentations plus détaillées des pratiques de plusieurs des formations interrogées nous permettent de compléter l'analyse des réponses aux trois questions en apportant un regard plus qualitatif sur les usages et besoins en matière d'enseignement de la traduction automatique dans le contexte universitaire français.

1 Introduction

Le partenariat stratégique MultiTraiNMT (2019-22)¹ nous a permis de développer, d'évaluer et de diffuser des contenus en libre accès et des applications libres et *open-source* afin d'améliorer l'enseignement et l'apprentissage de la traduction automatique neuronale (*Neural Machine Translation, NMT*) auprès de publics variés, allant des apprenants de langues secondes aux apprentis traducteurs.

MultiTraiNMT² réunit des experts de quatre universités européennes - l'Universitat Autònoma de Barcelona, l'Université Grenoble-Alpes, l'université de Dublin (Dublin City University) et l'Universitat d'Alacant, et de deux entreprises - Prompsit Language Engineering et Xcelerator Machine Translations. Le partenariat a

1 Ce partenariat a bénéficié du soutien du programme Erasmus+ de l'Union européenne.

2 <https://multitrainmt.eu>.

également été soutenu par de nombreux partenaires associés dans l'enseignement et l'industrie de la traduction,³ soucieux de se former à une utilisation raisonnée de la traduction automatique neuronale.

L'objectif principal de MultiTrainMT était de développer un programme de cours innovant sur la traduction automatique en général, et sur la NMT en particulier. Au terme du projet, nous avons livré les éléments suivants :

- Un manuel de cours publié en accès libre par Language Science Press,⁴ qui traite à la fois des fondements techniques de l'apprentissage automatique (et en particulier de l'apprentissage profond) tel qu'il est utilisé par la NMT, et des implications éthiques, sociétales, didactiques et professionnelles de cette technologie.
- Une application pédagogique en ligne appelée MutNMT qui permet aux utilisateurs d'apprendre comment entraîner son propre système de traduction automatique neuronale et d'avoir un aperçu du fonctionnement interne des systèmes.⁵
- Des activités liées au manuel de cours et à MutNMT, qui permettent aux apprenants et aux traducteurs de co-construire des connaissances sur la NMT.⁶

Cette formation est conçue pour être suivie selon des modalités asynchrones et/ou synchrones. L'auto-apprentissage est possible, en s'appuyant sur le manuel de cours et sur les activités correspondantes. Tout enseignant intéressé pourra également utiliser le cours et les activités sous forme synchrone ou asynchrone avec ses élèves. Le manuel de cours et les activités d'apprentissage sont conçus en tenant compte de différents niveaux de progression, si bien que le cours peut être ajusté pour des profils d'étudiant·e·s différents. Si le projet est parvenu à son terme, nous souhaitons désormais que les usages des livrables produits dans ce cadre se diversifient et évoluent. Étant donné que chaque université associée au projet doit travailler avec ses propres contraintes, il nous a semblé pertinent d'engager un dialogue avec plusieurs partenaires universitaires français du projet. La présente contribution a donc été préparée avec plusieurs collègues qui se sont portés volontaires pour participer à cet échange, dont nous rendons compte ici.

Le tableau ci-dessous propose une brève description des formations concernées, et associe les noms des enseignant·e·s chercheurs·euses

3 La liste des principaux partenaires se trouve sur le site du projet : <https://multitrainmt.eu/index.php/fr/accueil/nous-rejoindre>.

4 <https://langsci-press.org/catalog/book/342>.

5 <https://ntradumatica.uab.cat/>.

6 <https://github.com/jaspock/mt4everyone>.

interrogés en tant que responsables de ces formations et partenaires associés du projet MultiTraiNMT.

Table 1 Les formations partenaires du projet MultiTraiNMT

Université	Intitulé de la formation	Responsable	Spécificités de la formation ou brève description
Toulouse Jean Jaurès	Master Traduction, interprétation, médiation (TIM)	A. Josselin Leray	Cours d'introduction à la traduction automatique en M1, puis de traduction automatique et post-édition en M2, ce dernier étant en partie assuré par une traductrice professionnelle.
Rennes 2	Master mention Traduction et interprétation	K. Hernandez Morin	Cours de traduction automatique et post-édition assuré par moitié, en M1 et en M2, par une enseignante chercheuse et par un professionnel.
ESIT	Master Traduction éditoriale, économique et technique	H. Martikainen	Cours de traduction automatique et post-édition en M2
ESIT	Formation continue	S. Salmi	Dans le cadre de sa thèse, Sara Salmi s'intéresse à la formation continue à la post-édition des traducteurs.
Lille	Master TSM	R. Loock	Formation aux outils de traduction, dont TA, en M1 et M2 (cours dispensé par un professionnel en M2).
Paris Cité	Projet étudiant L3 Initiation au <i>deep learning</i>	N. Ballier	Cours optionnel de L3 associant étudiant·e·s en informatique et étudiant·e·s en anglais ou en linguistique.

Le cas de l'Université Paris Cité est particulier : il s'agit de l'utilisation de la plate-forme MutNMT dans un cours optionnel de L3 associant des étudiant·e·s en informatique et des étudiant·e·s en anglais ou en linguistique. Nous en présentons les contenus et bénéfices en fin de parcours.

2 Le sondage

Pour dresser un panorama des pratiques, nous avons soumis aux partenaires du projet trois questions concernant la place et le fonctionnement des enseignements de traduction automatique dans leurs formations. Ainsi, pour chacune des cinq formations ci-dessus (Paris Cité étant un cas différent, traité séparément), l'analyse qui suit considère les réponses à trois questions qui nous ont été suggérées par le sondage EMT 2017 (Rothwell, Svoboda 2019) et que le projet MultiTrainMT a traitées au moins en partie. La première question concerne le rôle et la place d'un enseignement théorique sur la traduction automatique (histoire de la TA, explication de son fonctionnement) aux côtés de la formation aux outils et aux pratiques de post-édition: ces enseignements sont-ils assurés séparément, et le cas échéant, quelle est la part de l'un et de l'autre ? La deuxième question s'intéresse à la prise en main de solutions de traduction automatiques susceptibles d'être entraînées et testées par les étudiant·e·s et/ou leurs enseignant·e·s. Enfin, nous nous interrogeons sur la formation des formateurs: l'équipe enseignante a-t-elle été formée à la traduction automatique, ou s'appuie-t-on sur un·e ou plusieurs expert·e·s professionnel·le·s de la traduction ?

1.1 Rôle et place d'un enseignement théorique de la traduction automatique

La première question était formulée ainsi : vos formations intègrent-elles un enseignement théorique sur la traduction automatique (histoire de la traduction automatique, explication de son fonctionnement) aux côtés de la formation aux outils pratiques de post-édition ? Les réponses sont présentées ci-dessous.

Table 2 Réponses à la question « Vos formations intègrent-elles un enseignement théorique sur la traduction automatique (histoire de la traduction automatique, explication de son fonctionnement) aux côtés de la formation aux outils pratiques de post-édition ? »

Oui	Autre	Non
3	2	0

Les enseignements théoriques déjà présents dans la majorité des formations consultées (3/5) comprennent par exemple une introduction au fonctionnement de la TA et de l'IA (Rennes 2, M1) et un approfondissement précisant les grands principes de fonctionnement de la TA (Rennes 2, M2). Dans le Master TSM de l'Université de Lille, il s'agit d'un cours visant à présenter l'histoire de la TA depuis ses

balbutiements, et ce, dans l'optique d'avoir une vision globale de son développement et de resituer l'utilisation actuelle dans l'histoire. De même, dans le Master TIM de l'Université de Toulouse Jean Jaurès, l'introduction proposée en M1 se fonde sur l'histoire de la traduction automatique et de ses usages, nous y revenons de façon plus détaillée dans la partie suivante.

Les deux formations n'ayant pas répondu « oui » ont choisi l'option « autre » et ont fait état de propositions en cours, ou d'amorces de mise en place :

- Proposition d'intégration d'un enseignement théorique dans le cadre de la formation continue (ESIT, Doctorat).
- Enseignement théorique qui concerne principalement la séance d'introduction (ESIT, Master trad. édit.).

On le voit, la présence d'un enseignement théorique sur la traduction automatique est garantie (existante ou prévue) dans l'ensemble des formations interrogées, avec des formats variables. Même s'ils sont peu nombreux, les formateurs et formatrices interrogé·e·s reconnaissent tous·tes la nécessité d'expliquer la construction progressive de la traduction automatique, ses différentes étapes et les concepts clés de cette construction. Ces réponses viennent confirmer la pertinence des propositions faites dans les premiers chapitres du manuel issu du projet MultiTraiNMT, qui se fondaient sur des besoins déjà bien décrits dans la littérature : Olohan (2017) décrit bien les risques du déterminisme technologique et les apports d'une approche constructiviste et critique, tandis que Rossi (2017a) et Moorkens (2017) analysent la démoralisation des traducteurs se sentant aliénés par cette technologie opaque. Enfin, et pour appuyer encore ce constat, on peut remarquer que les infographies créées par Lynn Bowker pour développer les approches critiques de la traduction automatique (*Machine Translation Literacy*) et encourager les bonnes pratiques placent elles aussi ces considérations théoriques au cœur du dispositif.⁷

Une question complémentaire portait sur la place de cet enseignement théorique au sein de modules plus larges (souvent intitulés « traduction automatique et post-édition »). Elle était formulée de la façon suivante : Ces enseignements (théorie vs pratique) sont-ils assurés séparément, et le cas échéant, quelle est la part de l'un et de l'autre ? Les réponses obtenues sont présentées ci-dessous.

⁷ Voir par ex. <https://drive.google.com/file/d/1VHcng04SiBgSsGfQBYcizUTt-LLX6J3qV/view> et https://drive.google.com/file/d/1Jae6cVuSK__qJs2qBeTg-GJO-tDRzjled/view ainsi que le site dédié : <https://sites.google.com/view/machinetranslationliteracy/home?authuser=0>.

Table 3 Réponses à la question « Ces enseignements (théorie vs pratique) sont-ils assurés séparément, et le cas échéant, quelle est la part de l'un et de l'autre ? »

Même enseignant-e, même intitulé	Enseignant-e-s différent-e-s, intitulés différents	Autre (préciser)
2	2	1

Deux formations présentent la configuration la plus simple : même enseignant et même intitulé de cours. Dans le cas de l'ESIT (Master traduction éditoriale, économique et technique), le cours de traduction automatique et de post-édition est essentiellement axé sur des exercices concrets mais il intègre une séance d'introduction théorique (1h30) dans laquelle sont évoqués l'histoire et les principes de fonctionnement de la TA. Le fonctionnement de la TA (neuronale) est également abordé à travers les exercices pratiques (par exemple, lors d'exercices d'analyse et de catégorisation des erreurs de la TA). Dans le cas de l'Université Rennes 2, le même intitulé désigne un enseignement plus théorique en M1 et plus orienté vers la pratique en M2.

La configuration avec enseignements et enseignants dissociés (enseignants et intitulés de cours distincts) concerne également deux formations. À l'Université Toulouse Jean Jaurès, dans le Master TIM, le cours de théorie est enseigné par une enseignante-chercheuse, et le cours de post-édition est assuré par une traductrice. L'apport théorique se situe en amont des séances de post-édition, et il permet aux étudiant-e-s de mieux comprendre les erreurs commises par la machine à corriger lors de la phase de post-édition. Dans le cas de la formation continue à l'ESIT, s'agissant d'une formation qui s'adresse à des professionnels, il est plus difficile de prévoir des séances étalées dans le temps. L'enseignement prévu se déroule sur deux journées, et mobilise un enseignant (universitaire) pour la partie théorique et un formateur (professionnel de la traduction) pour la partie pratique.

La réponse « autre » correspond à une situation particulière : à Lille, le cours de traduction automatique est dispensé par trois enseignants. Deux enseignants se partagent un même cours, et le troisième intervient également sur la traduction automatique (TA), mais sous un intitulé différent. Le cours principal concerné étant celui partagé par deux enseignant-e-s, qui fonctionne de la façon suivante : un enseignant présente le cours théorique au premier cours. Puis le second prend le relais avec une approche d'analyse de projets (approche du questionnaire de projets) et de réflexion quant à la mise en place de la TA en situation professionnelle (les étudiant-e-s lisent des e-mails clients et doivent y répondre en apportant tous les éléments d'analyse requis en proposant une solution de mise en place, un processus, etc.). Le premier enseignant reprend ensuite le flambeau avec

la pratique de la post-édition (approche du post-éditeur). Les étudiant-e-s sont invité-e-s à déterminer les critères de post-édition selon les niveaux de qualité attendus, en fonction du projet confié, puis ils post-éditent, et participent à une mise en commun du travail accompli. L'idée est alors plutôt d'orienter les étudiant-e-s sur les modifications à apporter ou non, plutôt qu'à travailler la qualité de la langue (qui fait l'objet de cours de traduction distincts), afin qu'ils cernent bien les différences de niveau de post-édition et puissent appliquer ces connaissances sur le marché ultérieurement.

1.2 Prise en main de solutions de traduction automatique susceptibles d'être entraînées et testées

La deuxième question concernait l'intérêt porté à la prise en main de solutions de traduction automatique personnalisables, et était formulée ainsi : vos formations intègrent-elles la prise en main de solutions de traduction automatique susceptibles d'être entraînées et testées par les étudiant-e-s et/ou leurs enseignant-e-s ? Les réponses apportées montrent que seule une des cinq formations intègre déjà cette composante (il s'agit du Master de l'Université de Rennes 2), mais qu'il s'agit d'un aspect en cours de développement dans toutes les autres, à une exception près (formation continue à l'ESIT).

Table 4 Réponses à la question: « Vos formations intègrent-elles la prise en main de solutions de traduction automatique susceptibles d'être entraînées et testées par les étudiant-e-s et/ou leurs enseignant-e-s ? »

Oui	Non	Pas encore mais nous avons prévu de le mettre en place
1	1	3

Dans le Master TIM de l'Université Toulouse Jean Jaurès, les discussions en cours portent sur les corpus disponibles, en lien avec les spécialisations des étudiant et/ou des enseignants. Dans le Master TSM de l'Université de Lille, l'utilisation de MutNMT pour entraîner différents moteurs et analyser les différences des résultats avec un système entraîné/non entraîné est à l'étude, en sachant que les étudiant-e-s sont déjà sensibilisés à ce type de problématique, mais sans exemples concrets. Ainsi, au cours des exercices de post-édition, l'enseignant relève les points qui auraient pu être évités avec un système personnalisé. Enfin, dans le Master de l'ESIT, le cours intègre un projet final en groupes dans le cadre duquel plusieurs options seront proposées, dont l'entraînement d'un moteur (MutNMT) mais aussi la TA adaptative (systèmes ré-entraînés en temps réel, et

capables de s'adapter aux choix faits d'une phrase à l'autre, par ex. Lilt, ModernMT) ou encore l'alimentation terminologique (DeepL).

1.3 Formation des formateurs

La dernière question cherchait à cerner comment était construite l'expertise des équipes enseignantes. Elle était formulée ainsi : quelle a été la formation de vos formateurs: l'équipe enseignante a-t-elle été formée à la traduction automatique, ou s'appuie-t-on sur un-e ou plusieurs expert-e-s ?

Table 5 Réponses à la question: « Quelle a été la formation de vos formateurs : l'équipe enseignante a-t-elle été formée à la traduction automatique, ou s'appuie-t-on sur un-e ou plusieurs expert-e-s ? »

L'enseignement est assuré par des expert-e-s extérieur-e-s (traductrices et traducteurs professionnel-le-s	L'enseignement est assuré par un-e ou plusieurs collègues qui se sont formé-e-s en autonomie	Autre (précisez)
2	2	1

Les réponses font ressortir deux stratégies majoritaires : le recours à des expert-e-s extérieur-e-s à l'université, d'une part, et d'autre part, l'auto-formation des universitaires qui prennent en charge cet enseignement. Les deux réponses obtenues pour cette dernière proposition rejoignent notre propre expérience pour la création d'un nouveau cours sur la traduction automatique, statistique puis neuronale, dans le Master TSM de l'Université Grenoble Alpes (Rossi 2017b) et montrent que l'auto-formation des enseignants chercheurs reste une pratique courante. Dans le contexte des formations en question, deux remarques s'imposent : tout d'abord, leur forte articulation avec les travaux de recherche des enseignants-chercheurs qui s'y impliquent (et nous y reviendrons dans le point suivant, à propos de la formation toulousaine), mais aussi l'insertion de ces formations dans le très dynamique réseau des Masters européens de traduction (EMT). Ce réseau a notamment mis en place un groupe de réflexion sur les technologies de traduction, qui permet aux collègues d'échanger sur les dernières tendances et de construire ensemble une réflexion critique ainsi que des propositions de formation (voir par ex. le sondage réalisé par Rothwell et Svoboda et la publication de 2019 qui en est issue).

3 Zoom sur le Master TIM de l'Université de Toulouse Jean Jaurès

En complément des réponses au sondage, et parce que le petit panorama présenté ici s'apparente davantage à une série d'études de cas, nous proposons trois vignettes qui permettent de mieux saisir la diversité des propositions susceptibles d'être faites en prenant comme source les livrables du projet MultiTraiNMT.

La première vignette concerne le Master TIM de l'Université de Toulouse Jean Jaurès. Sa construction étant originale, nous en proposons dans ce qui suit une présentation plus détaillée.

La partie théorique du cours de M1 propose une progression en cinq étapes. L'introduction discute de la place de la TA dans le panorama général des 'technologies de la traduction', et va du logiciel de traitement de texte à la TA neuronale. La première partie du cours est une présentation des contextes dans lesquels la TA est massivement utilisée : plusieurs cas d'usage sont abordés et discutés, le plus important étant un historique de l'utilisation de la TA à la Commission européenne. Pour faire émerger les premières questions sur le fonctionnement de la TA, une comparaison de l'analyse de la langue par la machine et de l'analyse de la langue par l'humain est ensuite proposée, s'appuyant sur des exemples comme le traitement des ambiguïtés syntaxiques. L'histoire de la TA est ensuite présentée dans ses grandes lignes. Enfin, la dernière partie aborde les différents systèmes de traduction automatique, depuis les systèmes à base de règles jusqu'à la traduction neuronale. Cette partie permet non seulement d'explorer le fonctionnement des différents systèmes en les mettant en perspective, mais elle permet aussi de faire un rappel sur la différence entre TA et traduction assistée par ordinateur (TAO) avec un logiciel s'appuyant sur des mémoires de traduction.

Pour prolonger cet enseignement théorique en M2, un état des lieux sur la recherche sur la traduction automatique est proposé. En groupe de trois ou quatre, les étudiantes et étudiant-e-s lisent deux articles de recherche ou chapitres d'ouvrage qui leur sont fournis par l'enseignante, et en font une présentation vulgarisée devant leurs camarades. Ces présentations sont suivies d'une discussion et de compléments éventuels apportés par l'enseignante.

À titre d'exemple, voici les sujets qui ont été abordés cette année:

1. La perception de la traduction automatique et des pratiques de post-édition par les traducteurs professionnels
2. Traduction automatique et traduction littéraire
3. Traduction automatique et sous-titrage
4. La traduction automatique en situation de crise
5. Traduction automatique et stéréotypes (par ex. les biais de genre, *gender bias*)
6. Approches critiques de la TA (*Machine Translation Literacy*)

7. Traduction automatique et qualité
8. Traduction automatique et pré-édition

Dans l'ensemble, ces sujets dessinent un tour d'horizon très large des pratiques et questions posées actuellement par la TA. Nous remarquons aussi que les deux derniers sujets correspondent à des chapitres de l'ouvrage issu du projet MultiTraiNMT, intitulé : *Machine Translation for Everyone* (Rossi, Carré 2022 ; Sánchez Gijón, Kenny 2022).

4 Zoom sur la proposition de formation continue de l'ESIT

Sara Salmi, doctorante à l'ESIT travaille sur les « défis et enjeux de l'élaboration d'un cadre pour la formation continue en post-édition des traducteurs déjà en exercice formation continue à la post-édition des traducteurs. » L'objectif de sa thèse est en effet d'élaborer un cadre de formation permettant à des traducteurs et traductrices, formatrices en post-édition, de créer des formations qui répondent aux besoins spécifiques des apprenants en formation continue. La méthodologie adoptée se compose de deux étapes: recueillir d'abord les besoins de formation à la post-édition des traducteurs et des donneurs d'ordre (type de contenu, mode de formation, formation générale ou spécifique à un domaine de spécialité) à travers deux questionnaires. La deuxième étape de cette recherche consiste à constituer un groupe de réflexion composé de différent·e·s acteurs et actrices du secteur de la traduction (post-éditeurs, formateurs, enseignants-chercheurs et donneurs d'ordre). L'objectif de ce groupe de réflexion est de définir, à partir d'un corpus de travail, les compétences devant être enseignées ainsi que le type de contenu que doit proposer la formation continue. Ce corpus de travail sera notamment composé du manuel de cours issu du projet MultitraiNMT et publié en accès libre par Language Science Press. Les modules traitant du choix des solutions de traduction automatique les plus appropriées, de la détection des erreurs en post-édition et de l'éthique de la traduction automatique seront plus particulièrement exploités. Pour mettre en avant les caractéristiques de ces chapitres et leur importance, un résumé est proposé ci-dessous.⁸

⁸ La version française du manuel étant en préparation, il s'agit ici de notre traduction des résumés de ces chapitres.

4.1 **Chapitre 3. Comment choisir une solution de traduction automatique (Rossi, Carré 2022)**

La traduction automatique (TA) évolue et se diversifie rapidement. Elle ne constitue pas une solution unique et uniforme, mais désigne plutôt une variété de solutions. Afin de choisir la bonne solution pour un projet donné, les utilisateurs doivent donc comparer et évaluer les différentes possibilités. Cette comparaison n'est jamais facile, surtout lorsque les résultats de la TA sont bons, et ils se sont considérablement améliorés ces dernières années, avec les systèmes de TA neuronale, ce qui rend les erreurs plus difficiles à repérer. Dès lors, une question se pose pour les utilisateurs de ces systèmes : comment pouvons-nous définir et évaluer au mieux la qualité d'une solution de TA neuronale, afin de faire les bons choix ? La première étape consiste à définir les besoins de la manière la plus précise possible. Après avoir défini une vision pragmatique de la qualité, nous présentons dans ce chapitre les notions clés de l'évaluation humaine et automatique de la qualité de la TA et expliquons comment elles peuvent être utilisées par les traducteurs.

4.2 **Chapitre 5. Comment aborder les erreurs de TA ? La post-édition (O'Brien 2022)**

Les sorties de traduction automatique peuvent être incorrectes et contenir des erreurs qui doivent être corrigées, en particulier si le texte est destiné à la publication. La tâche consistant à identifier et à corriger ces erreurs est appelée post-édition (PE). Dans ce chapitre, je présente une vue d'ensemble du processus de post-édition, en m'appuyant sur des sources universitaires et industrielles. J'explique que l'on distingue généralement la PE légère et la PE complète, et je décris les lignes directrices qui sont généralement utilisées pour chacun de ces deux types de pratiques, en me concentrant sur les questions qui se posent dans l'application de cette classification. Ce chapitre passe également en revue les différents types d'interface utilisés dans la PE, en allant des logiciels de traitement de texte et tableurs aux outils professionnels de traduction assistée par ordinateur (TAO). Les modes d'interaction possibles avec une sortie de TA sont également présentés : depuis l'interaction traditionnelle avec un texte statique, jusqu'aux interfaces permettant une présentation adaptative ou interactive. Enfin, les concepts et outils utilisés par les chercheurs qui travaillent sur la PE sont décrits, et un accent particulier est mis sur la mesure de l'effort temporel, technique et cognitif lors de tâches de post-édition.

4.3 Chapitre 6. Ethique et traduction automatique (Moorkens 2022)

La traduction automatique (TA) neuronale peut désormais faciliter la communication, ce qui n'était pas vraiment le cas des systèmes antérieurs de TA. Cependant, son utilisation a également des conséquences. Comme pour le développement de toute technologie, la TA n'est pas neutre d'un point de vue éthique, elle reflète plutôt les valeurs de ceux qui sont à l'origine de son développement. Dans ce chapitre, nous examinons les questions éthiques liées à la TA, en commençant par la collecte et la réutilisation des données et en examinant les correspondances et les écarts entre la TA et les valeurs et codes du traducteur. Si les machines et les systèmes reflètent des systèmes de valeurs, peuvent-ils être explicitement 'bons' et éliminer les préjugés de leurs résultats ? Quelle est la contribution de la TA aux discussions sur la durabilité et la diversité de nos pratiques ? Plutôt que de promouvoir une approche qui consiste à suivre une série d'instructions pour mettre en œuvre une technologie de manière irréfléchie ou peu critique, ce chapitre souligne l'importance d'un processus de prise de décision conscient lors de la conception d'un système s'appuyant sur la traduction automatique et le traitement de données.

L'autre aspect remarquable du développement des propositions de formation continue dans le cadre du travail de thèse de Sarah Salmi est leur aspect évolutif. Si la prise en main de solutions de TA n'était pas envisagée initialement, il est prévu que le groupe de réflexion discute de l'application pédagogique de MutNMT, afin de déterminer si le volet concernant l'apprentissage et l'entraînement des systèmes de traduction automatique neuronale doit être intégré ou non à la formation continue à la post-édition, de quelle manière et dans quel contexte (formation destinée aux traducteurs indépendants, aux traducteurs salariés dans le privé, aux traducteurs dans les institutions ou organisations publiques, etc.).

4.4 Étude de cas à l'Université Paris Cité

À l'Université Paris Cité, la plate-forme MutNMT a été utilisée en mode débutant et en mode expert. Cette double expérience permet d'aborder la formation des étudiant·e·s, dès la licence, et celle des enseignants-chercheurs, eux-mêmes souvent formateurs.

En mode expert, dans le cadre du projet de recherche SPECTRANS, la plate-forme a servi à tester les modèles entraînés sur la version complète de l'outil intégré dans MutNMT, à savoir le toolkit JoeyNMT (Kreutzer, Bastings, Riezler 2019). Dans le cadre de ces usages à Paris Cité, nous avons constaté qu'il n'était pas possible d'intégrer à l'interface

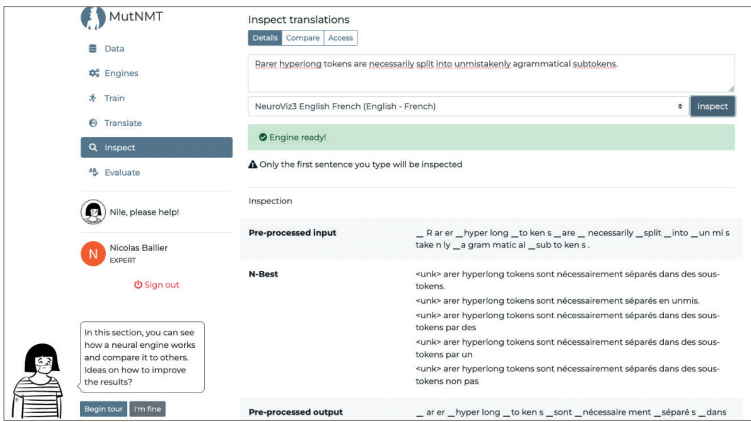


Figure 1 La fonction « inspect » de MutNMT. La plate-forme est disponible uniquement en anglais

les modèles déjà réalisés. Techniquement, les modèles se trouvent stockés sous la forme de fichiers qui ont une extension .chkpt (checkpoint). L'équipe de Paris Cité a donc procédé au réentraînement à partir d'une portion de données (conformément aux limitations de la plate-forme) et a laissé accessibles les modèles construits pour les traductions entre le français et l'anglais élaborés dans le cadre des projets Neuroviz (centré sur la problématique des biais de genre) et SPECTRANS (modèle plutôt centré sur la traduction de l'anglais biomédical). L'équipe a apprécié la possibilité de renseigner les métadonnées du modèle, même de manière succincte, remarquant que l'on pouvait ainsi savoir avec quelles données principales le modèle avait été entraîné.

Avec le profil débutant, la plate-forme a été utilisée avec des étudiant·e·s de licence LLCER motivés, dans le cadre du projet d'initiation à la recherche en Licence de Paris Cité, et l'interface graphique facilement utilisable s'est avérée très utile pour initier les étudiant·e·s aux différentes étapes de la traduction automatique neuronale. La possibilité de montrer l'étape de sous-tokenisation dans sa version la plus simple, illustrée par la capture d'écran ci-dessous, a été particulièrement appréciée.

Nous avons bien expliqué que seul un des algorithmes possibles pour la réduction du vocabulaire inhérente à la sous-tokenisation est ainsi utilisé. Il s'agit de l'algorithme SentencePiece (Kudo, Richardson 2018), là où d'autres systèmes sont possibles tel que Subword-NMT (Sennrich, Haddow, Birch 2015). Avec l'algorithme SentencePiece, c'est le tiret du bas qui indique le début du token, les espaces marquent la séparation entre les sous-tokens. Comme l'exemple l'illustre, les tokens les moins fréquents sont subdivisés en sous-chaînes de caractères qui ne correspondent pas nécessairement à des unités

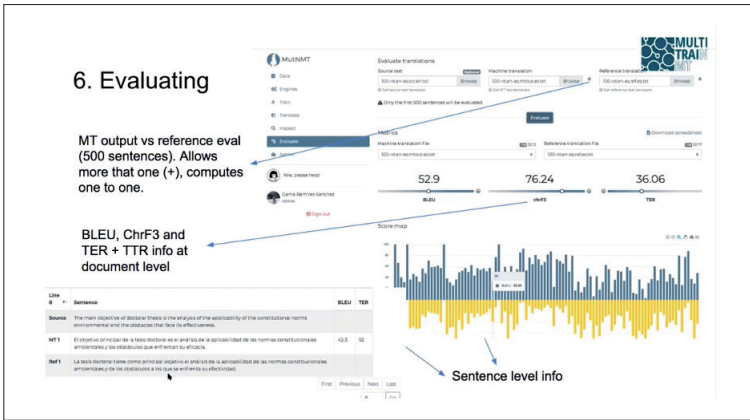


Figure 2 Les fonctionnalités de scores dans MutNMT

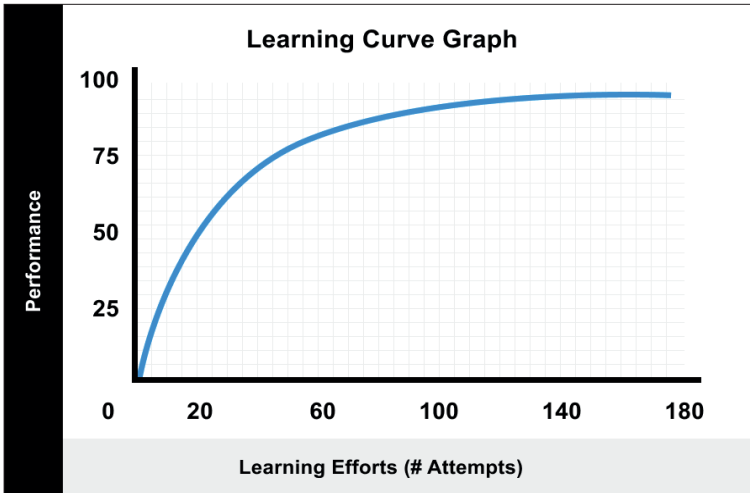


Figure 3 La courbe d'apprentissage des outils de la traduction automatique neuronale

morphologiques. *'Hyperlong'* est bien divisé en *hyper | long* mais *tokens* est divisé en *to ken s*. À l'inverse, nous n'avons pas particulièrement utilisé les activités de suivi de l'entraînement sur le tableau de bord (les fonctionnalités du dashboard étant plus complexes à appréhender).

Dans le cadre du projet de recherche SPECTRANS, consacré à la traduction spécialisée à base de traduction automatique neuronale, nous avons montré les différentes fonctionnalités disponibles, telle que l'inspection des données qui permet de visualiser les scores



Figure 4 Un exemple de carnet Jupyter

obtenus. Nous avons sensibilisé les doctorants du projet à l'existence de nombreux scores possibles pour l'analyse des traductions. L'interface graphique telle qu'elle a été conçue permet d'évaluer les traductions produites par le modèle en les comparant à une traduction de référence et en affichant trois métriques principales: le score BLEU, le score CHRF3 et le TER (Translation Error Rate). Si le système ne permet pas d'exporter sous forme de fichier.csv l'ensemble des scores associés à des traductions, nous avons apprécié que le graphe des scores mette en évidence la variabilité des scores selon les phrases traduites.

- Utiliser Pip pour installer Joeynmt <https://github.com/joeynmt/joeynmt>
- Installer un environnement conda, Python >= 3.7 PyTorch, torchtext et JoeynMT
- Programme en ligne de commande pour lancer des jobs
- Utiliser les carnets Jupyter (exécutables sous Google Collab)
- Utiliser la plate-forme mutNMT
- Utiliser les interfaces web (DeepL.com)

La véritable fracture numérique se situe pour nous au niveau de la pratique de la ligne de commande et MutNMT constitue la dernière étape (ou l'avant dernière étape) avant de se lancer dans la ligne de commande. Dans la courbe d'apprentissage des outils la traduction automatique neuronale que nous pouvons actuellement proposer, le premier niveau repose sur les interfaces classiques web que les étudiant·e·s pratiquent déjà largement, telles que les sites de DeepL ou de Google Translate. La plate-forme MutNMT permet de réaliser les actions par clic de souris (le « cliquodrome » disent les informaticiens férus de la ligne de commande) et l'étape suivante est le carnet Jupyter, qui mélange des commentaires des données de cours et du code. Avec le système des carnets, les étudiant·e·s sont dispensés de programmation et doivent juste exécuter le code en appuyant sur

un bouton. Le carnet Jupyter est un mode de distribution qui a été exploité pour le toolkit à proprement parler. L'étape suivante et ultime est la maîtrise sous forme de ligne de commande, nécessaires pour installer le programme sur un serveur distant et lancer les opérations (on parle de *jobs*). Les universitaires français peuvent avoir accès à la plate-forme HUMA-NUM et bénéficier ainsi d'un compte à l'INP3.⁹ Quand bien même le toolkit dispose d'un système de pré-installation et d'une commande simplifiée en '`pip install joeyNMT`' il est rare d'arriver à faire fonctionner du premier coup cette commande. Typiquement, si l'on dispose d'un accès à une carte graphique qui fera les calculs plus rapidement (ce qu'on appelle un GPU, Graphic Processing Unit), on a toutes les chances d'avoir des conflits avec le pilote de cette carte (CUDA) ou entre les bibliothèques de python. C'est toute la difficulté de l'incompatibilité entre certaines bibliothèques Python qui rend cette technologie particulièrement rebutante pour les traducteurs de bonne volonté et c'est là que réside la grande force de MutNMT en intégrant toutes ces difficultés et en évitant à l'utilisateur de d'avoir à installer le système 'derrière' l'interface graphique.

Enfin, pour le moment, l'interface de démonstration repose sur un serveur mais il est possible avec le code en accès libre d'adapter le système sur un serveur équipé d'un GPU; c'est la solution qui a été adoptée à l'Université Grenoble Alpes [fig. 4].

Même s'il n'est pas transposable partout, cet exemple montre que le projet ouvre aussi le monde de l'apprentissage automatique aux étudiant-e-s en langues et en traduction, et à leurs enseignants, améliorant par-là même leur capacité à fonctionner en tant que citoyens informés et compétents sur le plan technologique dans une Europe multilingue.

5 Conclusion

Conformément aux objectifs du partenariat stratégique MultiTrainMT, qui avait pour but de développer, d'évaluer et de diffuser des contenus en libre accès afin d'améliorer l'enseignement et l'apprentissage de la traduction automatique, nous avons interrogé plusieurs responsables de formation issus de contextes universitaires français variés. Par-delà la diversité des pratiques, les résultats de nos échanges font ressortir plusieurs points communs : l'articulation d'enseignements théoriques et pratiques pour permettre aux apprentis traducteurs de

⁹ Cette procédure est tout de même plus complexe, indépendamment de la disponibilité et de la réactivité des personnels d'HUMA-NUM, car il faut généralement installer un environnement conda, une version de Python supérieure à 3.7, des systèmes tels que pytorch, torchtext et un toolkit tel que JoeyNMT.

tirer le meilleur de la traduction automatique, mais aussi la mise en place progressive d'enseignements qui se fondent sur la prise en main de systèmes de traduction automatique neuronale, pour construire une compréhension toujours plus fine de leur fonctionnement et améliorer les interactions entre le traducteur et la 'machine'. Nous avons aussi présenté d'autres utilisations de la plate-forme MutNMT, dans le cadre de projets de recherche et d'une proposition originale de cours d'initiation à la recherche en troisième année de licence à l'Université Paris Cité, qui montre comment des étudiant·e·s en langue peuvent s'acheminer progressivement vers la prise en main de systèmes d'IA, des interfaces aux lignes de commandes.

Bibliographie

- Kenny, D. (ed.) (2022). *Machine Translation for Everyone : Empowering Users in the Age of Artificial Intelligence*. Berlin : Language Science Press. <https://zenodo.org/records/6653406>.
- Kreutzer, L. ; Bastings, J. ; Riezler, S. (2019). « Joey NMT : A Minimalist Nmt Toolkit for Novices ». *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) : System Demonstrations* (Hong Kong, Novembre 2019). Hong Kong : Association for Computational Linguistics, 109-14.
- Kudo, T. ; Richardson, J. (2018). « Sentencepiece : A Simple and Language Independent Subword Tokenizer and Detokenizer for Neural Text Processing ». *arXiv preprint arXiv : 1808.06226*.
- Moorkens, J. (2022). « Ethics and Machine Translation ». Kenny 2022, 121-40.
- Moorkens, J. (2017). « Under Pressure : Translation in Times of Austerity ». *Perspectives*, 25(3), 464-77.
- O'Brien, S. (2022). « How to Deal with Errors in Machine Translation : Postediting ». Kenny 2022, 105-20.
- Olohan, M. (2017). « Technology, Translation and Society : A Constructivist, Critical Theory Approach ». *Target*, 29(2), 264-83.
- Rossi, C. ; Carré, A. (2022). « How to Choose a Suitable NMT Solution? Evaluation of Mt Quality ». Kenny 2022, 51-80.
- Rossi, C. (2017a). « De quoi la traduction automatique est-elle le nom ». *En attendant Nadeau*. <https://www.en-attendant-nadeau.fr/2017/07/25/traduction-automatique/>.
- Rossi, C. (2017b). « Introducing Statistical Machine Translation in Translator Training : From Uses and Perceptions to Course Design, and Back Again ». *Revista Tradumàtica : tecnologies de la traducció*, 15, 48.
- Rothwell, A. ; Svoboda, T. (2019). « Tracking Translator Training in Tools and Technologies : Findings of the Emt Survey 2017 ». *Journal of Specialised Translation*, 32, 26.
- Sánchez-Gijón, P. ; Kenny, D. (2022). « Selecting and Preparing Texts for Machine Translation: Pre-editing and Writing for a Global Audience ». Kenny 2022, 81-104.
- Sennrich, R. ; Haddow, B. ; Birch, A. (2015). « Neural Machine Translation of Rare Words with Subword Units ». *arXiv preprint arXiv : 1508.07909*.