



**HAL**  
open science

## **Littératie de la traduction automatique (TA) neuronale et traduction spécialisée : s'appropriier les outils de la TA au travers de projets de recherche interdisciplinaires**

Nicolas Ballier, Maria Zimina-Poirot

### ► To cite this version:

Nicolas Ballier, Maria Zimina-Poirot. Littératie de la traduction automatique (TA) neuronale et traduction spécialisée : s'appropriier les outils de la TA au travers de projets de recherche interdisciplinaires. Human Translation and Natural Language Processing Towards a New Consensus?, Fondazione Università Ca' Foscari, 2023, <10.30687/978-88-6969-762-3/011>. <hal-05028183>

**HAL Id: hal-05028183**

**<https://hal.science/hal-05028183v1>**

Submitted on 10 Apr 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# Littératie de la traduction automatique (TA) neuronale et traduction spécialisée

## S'approprier les outils de la TA au travers de projets de recherche interdisciplinaires

Nicolas Ballier, Maria Zimina-Poirot

Université Paris Cité, France

**Abstract** We present a set of approaches to facilitate the integration of Neural Machine Translation (NMT) in teaching and research favouring an interdisciplinary perspective. The study suggests that a nuanced comparison of NMT generated translations requires translation experts to move from passive consumption to active participation in the construction of controlled models. This transition requires an understanding of the fundamentals of NMT, which calls for changes in pedagogical models to promote digital literacy while maintaining the standards of professional translation education in line with translation competence frameworks.

**Keywords** NMT literacy. NMT specialisation. NMT toolkits. Translation workflow. Translation competence framework.

**Sommaire** 1 Introduction. – 2 Littératie de la TA à Université Paris Cité : principaux projets. – 3 Description des livrables : outils d'expérimentation destinés aux linguistes. – 3.1 Système d'aide à la décision pour la gestion de corpus d'entraînement. – 3.2 Environnements expérimentaux simplifiés. – 3.3 Écosystème de carnets Jupyter pour l'analyse du fonctionnement de la TA neuronale. – 3.4 Interface de calcul des scores de traduction automatique. – 3.5 Premiers résultats et perspectives de recherche. – 4 Gestion de projets en traduction automatique et traduction spécialisée : une approche axée sur les flux de travail (*workflow*). – 5 Conclusion.

C'est un truisme de dire que, si l'informatique devait modifier sérieusement, après-demain, la pratique professionnelle, elle modifierait aussi par contre-coup les objectifs et les modalités de formation des futurs traducteurs. Il faudrait alors, bon gré mal gré, changer d'objectifs et de contenus, jusqu'à ce que les outils informatiques deviennent des systèmes traducteurs, auquel cas la question de l'impact de l'informatisation de l'exécution de la traduction sur la formation ne se poserait plus.

Gouadec 1994, 74

## 1 Introduction

Les développements récents de la traduction automatique (TA) en font un outil grand public facile d'accès et simple à utiliser.<sup>1</sup> Les services de TA directement accessibles en ligne sont de plus en plus nombreux et diversifiés. Le processus d'utilisation de ces services est particulièrement facile et ne nécessite que très peu de compétences informatiques ou traductionnelles. Dans ce contexte, les utilisateurs qui déposent leurs textes sur les serveurs en ligne sont généralement peu informés des risques liés à l'utilisation de ce type de services et ignorent souvent comment les utiliser à bon escient (quel type de contenu privilégier, pour quel usage, comment le relire et le corriger, etc.). Par ailleurs, les progrès de la traduction automatique neuronale ont été spectaculaires ces dix dernières années et les textes traduits automatiquement paraissent de plus en plus fluides aux non-spécialistes. Les professionnels conseillent de ce fait d'utiliser des avertissements systématiques pour différencier les contenus multilingues générés automatiquement des traductions humaines.

En traduction professionnelle, plusieurs types de compétences sont exigés pour gérer les projets de traduction professionnelle. Les cinq domaines du référentiel de compétences EMT (2022) définissent cinq blocs de compétences majeurs : langues et culture, traduction, technologies, compétences personnelles et interpersonnelles, prestation de services. Le domaine de la technologie englobe les compétences relatives aux connaissances et aux aptitudes nécessaires pour que les traducteurs puissent efficacement utiliser les outils d'aide à la traduction et gérer les processus de traduction intégrant la traduction automatique (TA). Cette mise en pratique est impossible sans compétences solides en révision, qui sont incontournables pour

---

<sup>1</sup> Ce travail a bénéficié d'un financement dans le cadre de l'AAP émergence 2020 (ANR-18-IDEX-0001, Financement IdEx Université Paris Cité) et de l'AAP Plate-formes de recherche - Petits et moyens équipements 2021 coordonné par le Comité « Plate-formes & Équipements Scientifiques » d'Université Paris.

détecter et corriger les erreurs de traduction automatique. Parallèlement, les compétences linguistiques exigent une bonne compréhension des technologies utilisées. Les deux pôles de compétences (traduction-révision et nouvelles technologies de la traduction) sont alors incontournables pour déjouer les pièges des systèmes TA et pour savoir intégrer les résultats de la traduction automatique dans la gestion de projets professionnels. Il est donc incontournable aujourd'hui d'adapter la formation des traducteurs aux exigences du marché qui a amorcé un tournant technologique :

Nous sommes aujourd'hui face à un processus de traduction fortement outillé qui demande l'acquisition d'un savoir-faire technologique intervenant à toutes les étapes. Il est donc indispensable que les formations préparant à la traduction spécialisée transmettent aux étudiants non seulement les compétences en traduction mais également la pratique de la traduction sur un poste de travail équipé des derniers outils disponibles. (Martinez 2019, 310)

En ce qui concerne les technologies utilisées en traduction automatique, les systèmes récents utilisés sur le marché recourent généralement à des méthodes neuronales qui utilisent des techniques d'apprentissage automatique appliquées à des données issues des corpus de traductions existantes qui servent d'exemples. Peu enseignées dans les formations linguistiques, ces méthodes sont pourtant incontournables pour développer les modèles de traduction spécialisés à base de corpus de textes. Une bonne compréhension de leur fonctionnement est primordiale pour comprendre les lacunes des systèmes actuels et pour les perfectionner. Cette compréhension permet aussi de mettre en pratique des approches outillées de la traduction spécialisée, notamment pour capitaliser sur les ressources linguistiques disponibles, telles que les mémoires de traduction, les bases de données terminologiques, les corpus spécialisés du domaine. Elle permet également d'appréhender les erreurs de la traduction automatique, de les repérer, les analyser et les corriger de façon plus ciblée afin de produire des traductions de qualité optimale qui tiennent compte des spécificités discursives de chaque projet.

La complexité des mécanismes en jeu dans la traduction neuronale a mis en avant la nécessité d'une littératie de la traduction automatique (*machine translation literacy* ; Bowker, Buitrago 2019) dans les formations en sciences humaines. Pour avancer sur ce terrain, la mise en place des projets de recherches pluridisciplinaires (Jiménez-Crespo 2020) en collaboration avec des statisticiens, des informaticiens et des spécialistes de traitement automatique des langues permet de mieux appréhender les problématiques de la traduction neuronale et de mieux les intégrer dans les formations en traduction et dans les travaux de recherche en linguistique appliquée. En

d'autres termes, nous pensons voir émerger le besoin de sensibiliser des traducteurs, des terminologues, des traductologues et des linguistes de corpus à une nouvelle sorte de littératie numérique concernant l'ensemble des moyens mis à disposition pour faire fonctionner les systèmes de traduction automatique neuronale et pour s'approprier des problématiques liées à l'intégration de ces systèmes dans les projets d'enseignement et de recherche en traduction spécialisée.

## 2 Littératie de la TA à Université Paris Cité : principaux projets

### ENSUPECO

Le projet ENSUPECO porte sur la mise en œuvre d'équipements pédagogiques spécifiques au titre de la mise en œuvre des maquettes pédagogiques pour le contrat quinquennal 2019-23. Le financement obtenu<sup>2</sup> a permis la mise en place initiale de la plate-forme SYSTRAN Pure Neural Server sur les serveurs de l'Université proposant des accès internes pour les étudiants et les enseignants-chercheurs. Ce projet pédagogique vise à présenter la dernière génération d'outils TA-TAO aux étudiants du parcours ILTS (Industrie de la langue et traduction spécialisée). En M2, ces étudiants travaillent en alternance pour des agences de traduction, des départements de traduction institutionnels, etc. Ils s'orientent majoritairement vers la traduction spécialisée. Ce projet, porté par les enseignants-chercheurs de l'équipe pédagogique du Master ILT, a permis de développer une approche traductologique appliquée et pragmatique de l'enseignement de la traduction tenant compte des transformations technologiques, organisationnelles et sociales que connaissent les métiers de la traduction.

### OCTAVES

Le projet OCTAVES (Outil de Collecte des Traductions des Apprenants en Vue de leur Exploration Scientifique) est un projet de création d'une application web permettant d'enrichir une base de données d'unités de traduction (phrases ou paragraphes) alignées. Ces données collectées sont alors mises à la disposition des enseignants-chercheurs avec comme objectif :

- d'améliorer les systèmes de traduction automatique dans les domaines de spécialité ;

---

<sup>2</sup> La demande de financement a été soutenue par le Conseil d'administration de l'UFR EILA, puis validée par la Commission de la Formation et de la Vie Universitaire (CFVU).

- de poursuivre les recherches sur les corpus d'apprenants de la traduction ;
- de poursuivre les recherches sur l'enseignement de la traduction
- de préparer des matériaux pédagogiques issus des corpus d'apprenants destinés à la formation en traduction des étudiants du Master Traduction et Interprétation (<https://u-paris.fr/eila/licences-et-masters>).

L'outil facilite ainsi la connexion entre les approches pédagogiques dans l'enseignement de la traduction et la recherche sur les défis de traduction posés par les textes spécialisés. Il permet de regrouper les archives des traductions réalisées par les apprenants de divers programmes de formation et diplômes. Grâce à un riche choix de métadonnées, il offre aux chercheurs la possibilité de trier les traductions selon les couples de langues, les domaines, les formations, etc. afin d'envisager des explorations scientifiques ciblées.<sup>3</sup>

## SPECTRANS

Le projet SPECTRANS a bénéficié d'un financement dans le cadre des appels à projets Émergence 2020 d'Université de Paris Cité. Il a notamment permis l'établissement de liens entre collègues informaticiens, mathématiciens et spécialistes de linguistique informatique pour appréhender la traduction spécialisée dans une perspective de traduction neuronale. La collaboration avec des informaticiens a facilité la mise à disposition d'outils simplifiés et l'exploration plus technique des mécanismes de la traduction neuronale. La collaboration s'est traduite par la création d'outils qui sont détaillés dans la dernière section, consacrée aux livrables.

## MutNMT

Université de Paris Cité a rejoint le consortium du projet Erasmus Plus sur la fin du projet MutNMT (<https://ntradumatica.uab.cat>) et a ainsi pu se servir des démonstrations des interfaces pour privilégier l'initiation aux méthodes d'apprentissage automatique sur corpus, notamment dans le cadre de projets d'initiation à la recherche en licence (voir également le chapitre coordonné par Caroline Rossi dans ce volume).

---

<sup>3</sup> Une documentation utilisateur est disponible sur la page d'accueil d'OCTAVES : <https://octaves.app.u-paris.fr>.

## PAPTAN

### Cadre général

Le projet PAPTAN (Plate-forme pour l'apprentissage profond pour la traduction automatique neuronale) a été financé en 2021 dans le cadre d'un appel à projets du Comité Plate-formes et Équipements Scientifiques d'Université Paris Cité. Voici la définition de la plate-forme de recherche donnée par le Comité :

Une plate-forme de recherche est le regroupement d'équipements scientifiques, numériques, documentaires et/ou d'expertises destinés à offrir à une large communauté d'utilisateurs des ressources technologiques de haut niveau autour d'une thématique commune. La plate-forme est ouverte de façon large au niveau régional et national, non seulement aux équipes du site, mais aussi aux expérimentateurs extérieurs, quel que soit leur rattachement (organismes publics, entreprises privées...) et dispose d'un modèle économique clair. (Comité Plate-formes et Équipements Scientifiques, Université Paris Cité 2023)<sup>4</sup>

Dans le cadre de la plate-forme PAPTAN, les équipements déployés servent à la fois à la recherche fondamentale pour entraîner des modèles multilingues spécialisés ainsi qu'aux applications des industries de la langue développées par un partenaire historique (SYSTRAN). La puissance de la plate-forme déployée permet à la fois de faire fonctionner le tout dernier écosystème de traduction neuronale, de mettre en production les ressources linguistiques (corpus spécialisés, bases de données terminologiques et phraséologiques) développées au sein de l'équipe de recherche CLILLAC-ARP,<sup>5</sup> et de conduire des expériences en apprentissage profond sur les processeurs graphiques (GPU) de la machine. PAPTAN est adossé à la plate-forme PNS-UP utilisée pour la formation et recherche en traduction spécialisée : <https://u-paris.fr/plateforme-paptan>.

---

<sup>4</sup> Site de l'appel à projets Plate-formes de recherche - Petits et moyens équipements : <https://u-paris.fr/appele-a-projets-plateformes-de-recherche-petits-et-moyens-equipements>

<sup>5</sup> CLILLAC-ARP (Centre de Linguistique Inter-langues, de Lexicologie, de Linguistique Anglaise et de Corpus-Atelier de Recherche sur la Parole) : <https://u-paris.fr/eila/clillac-arp>

## Architecture

La machine comporte trois processeurs A100 (GPU). Un des processeurs est piloté pour les entraînements par une composante logicielle qui permet un accès simplifié en interface graphique (ou plus complexe, en ligne de commande, dans le cas de SYSTRAN Model Studio Advanced).

Voici la configuration des composants logiciels du serveur :

- La plate-forme PNS-UP (SYSTRAN Pure Neural<sup>®</sup> Server) est un serveur de traduction qui offre la possibilité d'utiliser des modèles de traduction automatique spécialisée pour des projets de traduction, de post-édition et de révision de documents textuels, audio et de pages web. Elle permet une traduction instantanée, similaire aux interfaces web des systèmes de traduction, tout en intégrant des fonctionnalités de personnalisation des profils de traduction avec des ressources linguistiques telles que des dictionnaires, des mémoires de traduction, etc.
- SYSTRAN Model Studio Lite. Ce serveur d'entraînement de modèles spécialisés est configuré en mode graphique afin de répondre aux exigences des industries de la langue.<sup>6</sup>
- SYSTRAN Model Studio Advanced est un serveur d'entraînement de modèles spécialisés doté de fonctionnalités avancées qui offre un accès au GPU dédié de la plate-forme PAPTAN. Les deux autres processeurs ne sont accessibles qu'en ligne de commande, par le biais d'un protocole d'accès sécurisé (SSH) après l'installation d'un réseau virtuel privé (VPN) et la validation de la demande par les coordinateurs de la plate-forme.

L'accès aux GPU de la plate-forme PAPTAN facilite la réalisation d'expériences avec des technologies d'apprentissage profond. Ces technologies sont souvent complexes pour les novices, et cette complexité nous a incités à développer des outils spécifiques visant à rendre leur utilisation plus accessible aux linguistes qui cherchent à gagner en autonomie lors de la création de modèles spécialisés.

---

<sup>6</sup> SYSTRAN Model Studio Lite : <https://www.systran.net/en/modelstudio>.

### 3 Description des livrables : outils d'expérimentation destinés aux linguistes

Dans cette section, nous présentons une série de livrables de projets de recherche conçus pour créer des outils spécialisés destinés aux linguistes intéressés par le développement et la spécialisation de modèles de traduction. Étant donné que l'amélioration de la compétence en matière de traduction automatique implique la maîtrise de la programmation et de l'exécution de scripts en ligne de commande, notre travail vise à rendre l'accès aux technologies plus simple en développant des outils d'aide à la décision et d'expérimentation adaptés aux besoins des débutants.

#### 3.1 Système d'aide à la décision pour la gestion de corpus d'entraînement

Les stages réalisés dans le cadre du projet SPECTRANS ont permis de mettre au point une interface d'aide à la sélection de textes qui permet de collecter de façon ciblée les données d'entraînement pour la traduction spécialisée en anglais biomédical. Le système est conçu pour fonctionner avec les données PubMed<sup>7</sup> disponibles en ligne, mais pourrait aussi être adapté à l'exploration des données disponibles sur un serveur local. A partir d'un texte ou d'un ensemble de phases à traduire (par exemple, un échantillon utilisé dans la compétition WMT22 sur la traduction biomédicale),<sup>8</sup> l'outil permet d'identifier les documents disponibles dans PubMed qui seraient les plus proches sémantiquement des phrases à traduire. Les résumés de ces documents sont ensuite rapatriés pour compiler de nouveaux corpus d'entraînement ou de réentraînement sur des données spécifiques. Le système complet prêt à l'emploi est disponible en accès libre sur GitHub : [https://github.com/Fomba-Daouda/dashbord\\_pubmed](https://github.com/Fomba-Daouda/dashbord_pubmed) (Daouda 2023). Pour rendre ce type d'interfaces plus facilement utilisables par des linguistes, le déploiement a été facilité, notamment grâce à l'utilisation d'un environnement avec Docker qui simplifie le processus de gestion des processus d'application.<sup>9</sup>

---

<sup>7</sup> Archives PubMed Central® (PMC) : <https://www.ncbi.nlm.nih.gov/pmc/about/collections>.

<sup>8</sup> Workshop on Machine Translation : <https://machinetranslate.org/wmt>.

<sup>9</sup> Pour plus d'informations sur les avantages de l'installation avec Docker, voir l'article de blog de Jean-Baptiste Yunès, IRIT : <http://yunes.informatique.univ-paris-diderot.fr/installation-dun-environnement-de-travail-avec-docker>.

### 3.2 Environnements expérimentaux simplifiés

Pour la partie la plus difficile de l'utilisation des systèmes de traduction en ligne de commande dont les installations ont également été simplifiées (voir Rossi dans ce volume), nous avons mis en place des environnements expérimentaux dédiés. Nous avons stabilisé une version pour les principaux systèmes de traduction utilisés dans le cadre expérimental à l'aide de Conda (un système de gestion de paquets et d'environnement Open Source qui fonctionne sous Windows, MacOS et Linux). Ce fonctionnement permet à chaque utilisateur qui se connecte à distance d'activer un environnement configuré pour les différents systèmes : JoeyNMT (Kreutzer, Bastings, Riezler 2019), OpenNMT (version 3.0.1, Klein et al. 2020), Fairseq (version 0.12.0, Ott et al. 2019) et Nematus (version. 0.5 : Sennrich et al. 2017) et de les utiliser sans avoir à gérer les incompatibilités éventuelles lors de l'installation.

### 3.3 Écosystème de carnets Jupyter pour l'analyse du fonctionnement de la TA neuronale

Les carnets Jupyter<sup>10</sup> proposent un environnement de travail qui permet de combiner l'utilisation du code informatique et du texte rédigé. L'approche consiste à utiliser du code dans un carnet (ce format convient particulièrement bien pour la diffusion des connaissances et la prise en main d'outils de programmation). Elle a été retenue, par exemple, par l'Université de Cologne pour le projet de littératie numérique de la traduction (Krüger 2021). Il s'agit de solutions clés en main où le code est rédigé, prêt à être exécuté. Il est possible de s'en servir pour piloter des GPU et pour entraîner des modèles de traduction.

Dans ce contexte, nos développements ont permis de simplifier l'accès à certaines technologies pour mieux comprendre le fonctionnement de la traduction neuronale, notamment la sous-tokenisation (Ballier 2022) et l'échantillonnage des corpus d'entraînement (Ballier et al. 2021). Nous avons aussi conçu des scripts adaptés à la prise en compte de problématiques spécifiques, telles que l'analyse des composés dans le cadre des recherches sur les erreurs de traduction automatique des groupes nominaux complexes dans les articles de recherche en langue de spécialité (Bénard, en préparation).

---

<sup>10</sup> Pour plus d'informations sur le projet Jupyter : <https://jupyter.org/try-jupyter/retro/notebooks/?path=notebooks/Intro.ipynb>.

### 3.4 Interface de calcul des scores de traduction automatique

Nous avons rassemblé plusieurs méthodes de calcul de scores automatiques : score BLEU, TER, Rouge, WER (Balvet 2020) qui nécessitent des traductions de référence. L'interface graphique permet de calculer plusieurs scores automatiquement et, si l'on dispose de plusieurs traductions produites au cours de l'entraînement (Ballier et al. 2020), il est possible de générer un tableau qui associe aux différentes phrases traduites les différents scores des traductions générées et de simplifier les comparaisons entre traductions. Le système TradEval complet prêt à l'installation (avec Docker, qui simplifie l'installation des dépendances) est disponible librement sur Github (Ballier et al. 2021).<sup>11</sup> Une interface web est en cours de développement.

### 3.5 Premiers résultats et perspectives de recherche

La production d'outils a permis de répartir en tâches les nombreux chantiers simultanés de la recherche en traduction automatique neuronale. On voit se mettre en place le cycle itératif de la production des données avec la participation des linguistes : ne pas seulement se situer en aval (c'est-à-dire analyser les erreurs dans les traductions une fois produites) mais intervenir sur tout le cycle de production de la donnée. Dès le début du processus, les outils sont employés pour influencer les décisions et identifier les textes pertinents en vue de réaliser des entraînements à des fins spécifiques. Parallèlement, nous construisons des interfaces (environnements simplifiés, carnets Jupyter, fichiers de configuration.yaml qui paramètrent les entraînements, etc.) pour faciliter la mise en place d'expériences et l'interprétation des données orientées vers des problématiques linguistiques.

---

<sup>11</sup> Voir <https://github.com/pritie12/TradEval/blob/main/README.md>.

#### 4 **Gestion de projets en traduction automatique et traduction spécialisée : une approche axée sur les flux de travail (*workflow*)**

Au-delà du perfectionnement d'outils, l'utilisation réussie des moteurs de la traduction automatique neuronale en traduction spécialisée passe par le développement des méthodes de *workflow* éditorial qui reflètent les spécificités de chaque projet professionnel de traduction. Les traducteurs deviennent de plus en plus des experts en stratégies de traduction outillée, capables de concevoir des projets complexes sur mesure en fonction des attentes des clients et des contraintes locales (types de contenu, délais, formats, ressources et outils disponibles, politique éditoriale, mise à jour, archivage, confidentialité, etc.).<sup>12</sup>

Dans ce contexte, une très bonne compréhension des apports et des limites de la TA et des méthodes d'apprentissage automatique est cruciale pour la gestion intégrée des projets de traduction. Sur ce plan, les projets d'Université Paris Cité présentés dans le cadre de l'atelier SPECTRANS-PAPTAN du colloque Tralogy III ont permis d'avancer sur plusieurs chantiers afin de mieux cerner le spectre de l'utilisation de l'intelligence artificielle en traduction :

- Le rôle des méthodes d'apprentissage automatique dans le profilage et la sélection de textes pour l'entraînement de moteurs spécialisés (Ballier et al. 2022).
- Une meilleure compréhension des biais du fonctionnement des réseaux de neurones : données, architectures, etc. (Wisniewski et al. 2022).
- L'impact du pré-traitement des données et des annotations éventuelles sur les résultats de la traduction automatique spécialisés (Ballier et al. 2020).
- Les mécanismes de customisation et de spécialisation des modèles (Ballier et al. 2021 ; 2022).
- L'intégration des bases terminologiques disponibles dans plusieurs domaines spécialisés, par exemple dans le domaine médical (Ballier et al. 2021).
- La post-édition et la révision des sorties de la TA (Gledhill, Zimina 2019; Zimina, Gledhill 2021a; Kübler, Mestivier, Pecman 2022; Kübler et al., sous presse).
- L'analyse des erreurs de traduction et le perfectionnement de métriques (Ballier et al. 2021).
- L'intégration des guides de style dans les *workflows* avec des outils TA-TAO (Gledhill, Zimina 2019 ; Zimina, Gledhill 2021b)

---

<sup>12</sup> Voir, par exemple, Gledhill, Zimina 2019; Zimina, Gledhill 2021b.

- L'impact sociétal et la création d'outils pour la gestion de projet en traduction (voir le chapitre de Froeliger et al. dans ce volume).
- La nécessité de mutualisation des données, des modèles de langues et des ressources computationnelles (Namdarzadeh et al. 2022).
- Le retour critique sur les cursus et les prérequis des formations en traduction spécialisée (Zimina, Gledhill 2021a ; Looock, Lechaugnette 2021 ; Kübler et al., sous presse).

## 5 Conclusion

Nous avons essayé de montrer comment une approche pluridisciplinaire de la traduction permet le transfert de technologies des sciences dures vers les disciplines LSH. Ces transferts s'accompagnent de la mise à disposition des jeux de données produits, des scripts d'analyse et des accès centralisés à des plate-formes expérimentales. Cette pratique est importante pour la reproductibilité des expériences (Ballier et al. 2020).

Lorsque la communication des modèles ou des données d'entraînement pose des problèmes de place de stockage, une mise à disposition d'une partie des traductions produites et des scripts d'analyse en libre accès permet de travailler sur la diffusion des bonnes pratiques de littératie de traduction neuronale, par exemple dans le cadre de l'analyse des biais de genre fondés sur des jeux de données spécifiques (Wisniewski et al. 2022).<sup>13</sup>

A ce jour, les jeux de données hybrides, qui associent les traductions et leurs scores associés conçus pour simplifier la détection des exemples problématiques à analyser, ne peuvent pas être exploités sans participation active de traducteurs et linguistes qui restent indépassables pour mener des analyses des productions finies. Il nous semble néanmoins que l'analyse contrastive fine des traductions produites par les systèmes de la TA neuronale impose de passer de la consommation de la traduction automatique à la production plus contrôlée et ciblée de modèles de traduction avec la participation active des spécialistes de traduction. Cette transition exige de s'approprier les fondamentaux de la TA, et donc d'infléchir les maquettes pédagogiques pour désamorcer certaines craintes et pour inculquer les bonnes pratiques de littératie numérique, tout en préservant la qualité de l'enseignement de la traduction spécialisée en accord avec le cadre de compétences professionnelles (EMT 2022).

---

<sup>13</sup> Voir <https://naka.la.fr/10.34847/nkl.1209u49z>.

## Bibliographie

- Ballier, N. (2019). « R, Pour un écosystème du traitement des données? L'exemple de la linguistique ». Caron, P. ; Defiolle, R. ; Lay, M-H (éds), *Données, Métadonnées des Corpus et Catalogage des Objets en Sciences Humaines et Sociales*. Rennes : Presses universitaires de Rennes, 99-120.
- Ballier, N. et al. (2020). « The Learnability of the Annotated Input in NMT (Replicating Vanmassenhove & Way 2018 with OpenNMT) ». *Proceedings of the 12th Language Resources and Evaluation Conference*. Marseille : European Language Resources Association, 5631-40. <https://aclanthology.org/2020.lrec-1.691>.
- Ballier, N. et al. (2021). « The SPECTRANS System Description for the WMT21 Terminology Task ». *Proceedings of the Sixth Conference on Machine Translation*. Online : Association for Computational Linguistics, 813-20. <https://aclanthology.org/2021.wmt-1.80>.
- Ballier, N. (2022). « Faut-il former à ce que voit le réseau de neurones pour l'entraînement de la traduction ? ». *Actes de colloque Université libre de Bruxelles = Enseigner la traduction et l'interprétation à l'heure neuronale* (Université libre de Bruxelles, 28-29 septembre 2022).
- Ballier, N. et al. (2022). « The SPECTRANS System Description for the WMT22 Biomedical Task ». *Proceedings of the Seventh Conference on Machine Translation (WMT)*. Abu Dhabi : Association for Computational Linguistics, 895-900. <https://aclanthology.org/2022.wmt-1.82>.
- Balvet, A. (2020). « Métriques d'évaluation en Traduction Automatique : le sens et le style se laissent-ils mettre en équation ? ». Milliaressi, T. (éd.), *La traduction épistémique : entre poésie et prose*. Presses universitaires du Septentrion, 315-56.
- Bénard, M. (en préparation). *Difficulté de traduction automatique des groupes nominaux complexes dans les articles de recherche en langue de spécialité médicale et dans le domaine du TAL : aide à la formation des post-éditeurs professionnels et occasionnels* [thèse de doctorat en préparation sous la direction de Natalie Kübler]. Paris : Université Paris Cité.
- Bowker, L. ; Buitrago C.J. (2019). *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*. Bingley : Emerald Publishing Limited. <https://doi.org/10.1108/9781787567214>.
- Daouda, F. (2023). *Pubmed Dashboard*. [https://github.com/Fomba-Daouda/dashbord\\_pubmed](https://github.com/Fomba-Daouda/dashbord_pubmed).
- European Master's in Translation (EMT) (2009 ; 2017 ; 2023). *European Master's in Translation – Référentiel de compétences*. [https://commission.europa.eu/resources-partners/european-masters-translation-emt/european-masters-translation-emt-explained\\_fr](https://commission.europa.eu/resources-partners/european-masters-translation-emt/european-masters-translation-emt-explained_fr).
- Gledhill, C. ; Zimina, M. (2019). « The Impact of Machine Translation on a Masters Course in Web Translation : From Disrupted Practice to a Qualitative Translation/Revision Workflow ». *Proceedings = Translating and the Computer*, 41 (London, 21-22 November 2019). Geneva : Éditions Tradulex, 60-73. <http://www.tradulex.com/varia/TC41-London2019.pdf>.
- Gouadec, D. (1994). « Traduction et informatique : les implications pour la formation ». *Langages*, 59-74.
- Jiménez-Crespo, M.A. (2020). « The 'Technological Turn' in Translation Studies : Are We There yet? A Transversal Cross-Disciplinary Approach ». *Translation*

- Spaces, 9, 2 (31 December 2020), 314-41. <https://doi.org/10.1075/ts.19012.jim>.
- Klein, G. et al. (2020). « The OpenNMT Neural Machine Translation Toolkit : 2020 Edition ». Denkowski, M.; Federmann, C. (eds), *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas* (October 2020). Vol. 1, *Research Track*, 102-9. Association for Machine Translation in the Americas. <https://aclanthology.org/2020.amta-research.9>.
- Kreutzer, L. ; Bastings, J. ; Riezler, S. (2019). « Joey Nmt : A Minimalist Nmt Toolkit for Novices ». Pado, S. ; Huang, R. (eds), *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* (November 2019). Hong Kong : Association for Computational Linguistics, 109-14.
- Krüger, R. (2021). « An Online Repository of Python Resources for Teaching Machine Translation to Translation Students ». *Current Trends in Translation Teaching and Learning*, 8, 4-30. <https://doi.org/10.51287/cttle20212>
- Kübler, N. ; Mestivier, A. ; Pecman, M. (2022). « Using Comparable Corpora for Translating and Post-editing Complex Noun Phrases in Specialised Texts : Insights from English-to-french in Specialised Translation ». Granger, S. ; Lefer, M-A. (eds), *Extending the Scope of Corpus-based Translation Studies*. London ; New York : Bloomsbury Academic, 237-66. Bloomsbury advances in translation series.
- Kübler, N. et al. (sous presse). « Post-editing Neural Machine Translation in Specialised Languages : The Role of Corpora in the Translation of Phraseological Structures ». Monti, J. ; Corpas Pastor, G. ; Mitkov, R. (eds), *Recent advances in Multiword Units in Machine Translation and Translation Technology*. Amsterdam ; Philadelphia : John Benjamins Publishing Company.
- Loock, R. ; Lechauguette, S. (2021). « Machine Translation Literacy and Undergraduate Students in Applied Languages : Report on an Exploratory Study ». *Tradumàtica*, 19, 204-25.
- Martínez, L. (2019). « L'impact de la technologie sur les traductions spécialisées : Le cas de la traduction médicale ». *Des mots aux actes*, 8 : *Traduction et technologie, regards croisés sur de nouvelles pratiques*. Classiques Garnier, 309-26. <https://hal.science/hal-02927915>.
- Namdarzadeh, B. et al. (2022). « Toward a Test Set of Dislocations in Persian for Neural Machine Translation ». *Proceedings of the Third International Workshop on NLP Solutions for Under Resourced Languages (NSURL 2022) Co-Located with ICNLSP 2022*. Trento : Association for Computational Linguistics, 14-21. <https://aclanthology.org/2022.nsur1-1.3>
- Ott, M. et al. (2019). « Fairseq : A Fast, Extensible Toolkit for Sequence Modeling ». *Proceedings of the 2019 Conference of the North*, 48-53. Minneapolis : Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-4009>
- Sennrich, R. et al. (2017). « Nematus : A Toolkit for Neural Machine Translation ». *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 65-8. Valencia : Association for Computational Linguistics. <https://aclanthology.org/E17-3017>.
- Wisniewski, G. et al. (2022). « Biais de genre dans un système de traduction automatique neuronale : une étude des mécanismes de transfert cross-langue ». *Revue TAL : Traitement Automatique des Langues*, 63(1), 37-61.
-

Zimina, M. ; Gledhill, C. (2021a). « L'impact de la traduction automatique sur les pratiques langagières et professionnelles des apprentis-traducteurs : entre apports en efficacité et menaces pour la diversité des discours ». Beacco J-C. ; Herreras J.C. ; Tremblay C. (éds), *Traduction automatique et usages sociaux des langues. Quelle conséquences pour la diversité linguistique ?* Observatoire européen du plurilinguisme, 63-81. <https://doi.org/10.3917/oep.beacc.2021.01.0063>.

Zimina, M. ; Gledhill, C. (2021b). « Comment les étudiants révisent-ils les sorties de la traduction automatique neuronale ? Quelques réflexions sur l'application des directives de la DGT et l'impact de l'anglais lingua franca sur les processus de traduction ». *L'intelligenza artificiale per la traduzione: verso una nuova progettazione didattica?* = *Translating Europe Workshop* (Torino, 3 dicembre 2021).

