



HAL
open science

Traduction humaine et traitement automatique des langues - Introduction

Nicolas Froeliger, Claire Larssonneur, Giuseppe Sofo

► **To cite this version:**

Nicolas Froeliger, Claire Larssonneur, Giuseppe Sofo. Traduction humaine et traitement automatique des langues - Introduction. Traduction humaine et traitement automatique des langues, 2023, 978-88-6969-762-3. 10.14277/978-88-6969-762-3/001 . hal-04418864

HAL Id: hal-04418864

<https://hal.science/hal-04418864>

Submitted on 31 Jan 2024

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.

Introduction

Nicolas Froeliger

Université Paris Cité, France

Claire Larssonneur

Université Paris 8 Vincennes – Saint Denis, France

Giuseppe Sofò

Università Ca' Foscari Venezia, Italia

La décennie écoulée a vu se produire des changements technologiques majeurs dont les domaines de la traduction automatique (TA) et du traitement automatique des langues naturelles (TAL) ont été et demeurent un des théâtres privilégiés.

Les progrès remarquables de la recherche sur l'automatisation de la traduction de texte et de parole (traduction automatique, synthèse et reconnaissance vocales, articulation entre image et texte, métriques d'évaluation, etc.), articulés à ce que l'on regroupe sous la vaste appellation d'intelligence artificielle (IA), élargissent le champ d'application des outils de traduction. Ces évolutions et la démocratisation de la traduction neuronale nous invitent à repenser les modes opératoires et l'organisation même des métiers qui leur sont liés.

Cette dynamique est stimulée par la croissance rapide de l'industrie des langues (Hickey 2023) et par l'intérêt que lui porte le secteur technologique (GAFAM, BATX, iFLYTEK, etc.), dont les entreprises investissent massivement dans les technologies de traduction en vue de les intégrer à leurs diverses interfaces, plateformes et applications (Larssonneur 2020).

Depuis quelques années, donc, les effets de ces évolutions ne sont plus confinés à la recherche : la condition du traducteur professionnel est en train de changer, de même, peut-être, que la perception générale de ce qu'est la traduction (Lavault-Olléon, Zimina 2019). Deux grandes transformations sont en cours.

La première tient au rôle central que les données ont acquis dans le processus de traduction. La collecte, le nettoyage, l'annotation et la structuration de données issues de corpus gigantesques sont essentiels à l'entraînement efficace des algorithmes de traduction. Ces données peuvent au demeurant être considérées comme une marchandise ou comme un élément d'un droit fondamental et un bien commun (Moorkens, Lewis 2019). Autant d'aspects qui appellent évaluation, mise en perspective, voire réglementation en fonction des spécificités de la traduction et des intérêts de la société tout entière.

La deuxième grande transformation consiste dans le passage d'une approche centrée sur les contenus à une autre, axée sur l'utilisation des documents, qui sont désormais produits, partagés, évalués et recyclés en ligne d'une façon qui remet en cause la distinction entre les textes « pour information », « purement fonctionnels » ou « destinés à la publication ». En outre, la gestion agile des contenus, les mesures d'engagement des utilisateurs et les textes générés de manière automatique par de grands modèles de langue tels que le modèle GPT-3 (Brown et al. 2020) et sa très riche descendance supposent une articulation entre production linguistique humaine et traduction automatique, laquelle mérite là aussi d'être pensée et analysée (Ehrensberger-Dow, Massey 2019).

Le rythme et l'ampleur des évolutions technologiques, économiques et sociétales sont tels qu'ils soulèvent une série de questions. L'engouement récent autour de la TA neuronale contraste avec les nombreux domaines où les technologies de traitement des langues, bien qu'avancées, ne sont pas entièrement au point : on peut mentionner à cet égard, à des degrés divers, l'écart entre les langues bien dotées en ressources numériques et celles pour lesquelles ces ressources sont encore lacunaires, les contraintes inhérentes aux interfaces de TA ou encore l'opacité qui entoure l'élaboration des corpus d'entraînement.

Plus généralement, il existe des incertitudes sur le rôle, le statut et l'avenir économique des traducteurs, ainsi que sur l'éco-durabilité du modèle de traduction (Cronin 2017) et la responsabilité sociétale des principaux fournisseurs de TA. Corollaire de ces mouvements tectoniques, la formation des traducteurs se trouve, elle aussi, à la croisée des chemins : il lui faut trouver les moyens de concilier ces nouvelles évolutions avec des compétences et des contenus plus traditionnels (Froeliger 2019 ; Massey, Kiraly 2019).

Conscients du rôle essentiel de la traduction et des technologies linguistiques dans les sociétés actuelles, nous proposons, à travers un volume comme celui-ci, de rechercher un nouveau consensus entre les utilisations humaines du langage et les apports de la machine, non seulement pour permettre les échanges et contribuer à l'élaboration et à la diffusion des connaissances, mais aussi pour exercer notre sens social des responsabilités (Martens, Hobbs 2015). Une culture

numérique plus développée et plus largement partagée dans le domaine des technologies linguistiques ainsi qu'une meilleure compréhension des implications d'ordre social, économique et éthique qui s'y rattachent nous apparaissent en effet comme impératifs (Koskinen, Pokorn 2021 ; Moorkens, Rocchi 2021).

Tel a notamment été l'objet du colloque international Tralogy (mot-valise formé en anglais à partir de *Translation* et *Technology*) III, organisé à Paris au printemps 2022, dont une partie des travaux sont venus nourrir le volume que nos lecteurs ont sous les yeux. Il s'agissait d'inviter les participants à identifier les angles morts du paysage actuel de la traduction et des technologies linguistiques multilingues sur les plans de la recherche, de la théorie, de la pratique et de la formation, et à proposer des perspectives critiques et des pistes novatrices pour y pallier.

Dans le sillage des deux premières éditions de Tralogy, en 2011 et 2013, également à Paris, cette manifestation visait à rassembler et faire dialoguer des spécialistes des différents domaines et disciplines participant des métiers de la traduction d'aujourd'hui et de demain : universitaires (traductologie et TAL), traducteurs professionnels (en particulier à travers la participation active de la SFT, Société française des traducteurs), utilisateurs de la traduction (notamment à travers la DGT, la Direction générale de la traduction de la Commission européenne) et enseignants de la traduction (à travers l'AFFUMT, l'Association française des formations universitaires aux métiers de la traduction, ainsi que l'Université Paris VIII et l'Université Ca' Foscari de Venise). Sont venues s'y ajouter, cette fois, des représentants des grandes entreprises technologiques, acteurs, redisons-le, désormais incontournables de ce secteur.

Les deux premières éditions de cette série avaient été des événements, en ceci qu'ils étaient parmi les tout premiers à rechercher un dialogue entre ces différentes communautés, jusqu'alors plus habituées à réfléchir chacune à l'intérieur de son propre créneau. La deuxième avait d'ailleurs expressément intégré ce dialogue en invitant ses contributeurs à présenter non pas tant ce qu'ils savaient faire, mais plutôt ce qu'il leur était impossible de faire seuls.¹ Depuis lors, les colloques et autres manifestations scientifiques reprenant un tel format se sont multipliés et il faut s'en réjouir, car la tâche est immense et le seul moyen de se montrer à la hauteur de la diversité des besoins est de frayer ensemble des pistes de réponse.

Nos lecteurs trouveront en tout état de cause dans ce volume bilingue une riche matière à réflexion sur l'avancement des efforts vers

1 Voir par exemple https://hal.science/hal-02497138v1/file/Tralogy_S0_A3_Froeliger.pdf (les autres présentations des deux premières éditions sont disponibles sur la plate-forme française de recherche HAL).

ce nouveau consensus auquel il nous paraît indispensable d'œuvrer entre biotraduction et traitement automatique des langues.

L'ouvrage est divisé en trois parties : dans la première partie, intitulée « Mise en contexte », un aperçu des rencontres entre les outils numériques et la traduction est proposé, analysant à la fois les succès et les échecs de cette rencontre, mais surtout les perspectives pour l'avenir.

L'ouvrage débute ainsi par cette nécessaire mise en contexte, à partir de l'histoire de la rencontre entre traduction et intelligence artificielle retracée par Giuseppe Sofo. La mise en scène de la machine comme conversationnelle et traduisante, typique de l'anthropomorphisme des représentations associées à l'intelligence artificielle, mérite d'être questionnée : cette intelligence en est-elle une et s'apparente-t-elle à la cognition humaine ? Les articles de Michael Carl et de Roberto Laghi montrent la complexité des phénomènes cognitifs en jeu. Michael Carl, dans sa contribution, resitue et discute la part du corps (*embodiment*) : il propose un cadre théorique pour mieux comprendre les interactions cognitives avec l'environnement traductif. Roberto Laghi se penche sur l'articulation de l'informatique et de la cognition, à partir de la notion d'assemblage cognitif.

La deuxième section de l'ouvrage, « Traduction neuronale et formation des traducteurs », est consacrée aux enjeux de la traduction au sein de la formation des traducteurs.

Sara Cotelli Kureth, Alice Delorme Benites, Mara Haller, Hasti Noghrechi et Elizabeth Steele s'intéressent à la manière dont les étudiants s'emparent des plateformes de traduction gratuites et en mésusent, confondant recherche lexicale et pré-traduction automatique. Pour que les évaluations des travaux d'étudiants soient justes, mais aussi pour mieux les former, il est crucial de pouvoir distinguer entre textes traduits et textes post-édités à partir de pré-traductions automatiques : Hanna Martikainen examine la pertinence des métriques automatiques en la matière. A plus grande échelle, le projet européen MultiTraiNMT, présenté par Caroline Rossi, Amélie Josselin-Leray, Nicolas Ballier et Hanna Martikainen, a permis de développer, d'évaluer et de diffuser des contenus en libre accès et des applications libres et open-source, notamment pour former les formateurs. Enfin, à l'heure de la traduction neuronale, la conception même d'un programme de master doit évoluer : Elena Kokanova, Maya Lyutyanskaya et Aleksandra Epimakhova partagent leur expérience et leurs réflexions sur ce point.

La troisième section, intitulée « Recomposition des champs », examine l'intégration de la traduction automatique neuronale dans le domaine de la traduction spécialisée et institutionnelle.

Nicolas Froeliger, Christopher Gledhill et Maria Zimina-Poirot examinent ainsi l'impact sociétal de la traduction neuronale. Quelles sont les implications du passage au modèle neuronal de traduction

automatique au sein de la Direction générale de la traduction de l'Union européenne ? Caroline Rossi et Aurélien Talbot ont conduit une série d'entretiens semi-dirigés qui éclairent l'évolution des pratiques de traduction au sein des institutions internationales. Nicolas Ballier et Maria Zimina-Poirot montrent l'intérêt d'une approche pluridisciplinaire de la traduction, qui permet le transfert de technologies des sciences dures vers les disciplines LSH. La mise en place de la plateforme de recherche PAPTAN, en partenariat avec Systran, donne aux linguistes accès à des outils d'expérimentation (sur la gestion des corpus, le calcul des scores, etc.) très utiles.

Ce volume n'a bien évidemment pas l'ambition de dire un mot définitif sur la relation entre traduction automatique et traduction humaine, entre 'traducteurs automatiques' et 'traducteurs humains', ou 'biotraducteurs'. L'objectif de ce recueil d'essais est plutôt d'enregistrer un dialogue déjà présent dans la communauté scientifique et professionnelle – comme l'a très bien montré le colloque Tralogy III, qui a permis à ces deux domaines trop souvent distincts de dialoguer de manière constante et fructueuse –, et surtout de stimuler et d'encourager la poursuite de ce dialogue, à l'intérieur et à l'extérieur des universités.

On sait depuis longtemps que l'être humain n'est pas entièrement distinct des technologies qu'il crée et utilise, mais qu'il est en coévolution avec elles (Stiegler 2018). Il nous semble donc que le moment est venu de mieux comprendre l'apport des technologies numériques à la traduction et au traitement automatique des langues, et surtout de relever le défi d'orienter – ou au moins d'accompagner – leur évolution, afin qu'elles ne deviennent pas un mécanisme de contrôle et de normalisation, mais plutôt une arme pour enrichir nos connaissances et nos possibilités de développement culturel, social et humain.

Pour parvenir véritablement à ce 'nouveau consensus', prudemment accompagné d'un point d'interrogation dans le titre de cet ouvrage, il faudra en effet que tous les acteurs du monde de la traduction – professionnels, formateurs, chercheurs, mais aussi les utilisateurs de la traduction – collaborent dans le sens d'une interaction toujours plus profonde et plus significative entre les machines et les humains.

Introduction

Nicolas Froeliger

Université Paris Cité, France

Claire Larssonneur

Université Paris 8 Vincennes – Saint Denis, France

Giuseppe Sofò

Università Ca' Foscari Venezia, Italia

The last decade has seen major technological changes related to artificial intelligence, particularly in the fields of machine translation (MT) and natural language processing (NLP).

The remarkable progress made in research into the automation of text and speech translation (machine translation, speech synthesis and recognition, image/text linking, evaluation metrics, etc.), combined with what is known as artificial intelligence (AI), is broadening the field of application of translation tools. These developments and the democratisation of neural translation are leading us to rethink our working methods and the organisation of the professions themselves.

This dynamic is fuelled by the rapid growth of the language industry (Hickey 2023) and the interest shown by the technology sector (GAFAM, BATX, iFLYTEK, etc.), whose companies are investing heavily in translation technologies in order to integrate them into their various interfaces, platforms and applications (Larssonneur 2020).

In recent years, therefore, the impact of these developments have no longer been limited to research: the status of the professional translator is changing, as perhaps the general perception of the nature of translation (Lavault-Olléon, Zimina 2019). Two major transformations are underway.

The first is the central role that data has acquired in the translation process. Collecting, cleaning, annotating and structuring data from gigantic corpora is essential to the effective training of translation

algorithms. Moreover, data can be considered as a commodity or as part of a fundamental right and a common good (Moorkens, Lewis 2019). All these aspects need to be assessed, put into perspective and even regulated according to the specificities of translation and the interests of society as a whole.

The second major change is the shift from a content-centric approach to one that focuses on the use of documents, which are now produced, shared, evaluated and recycled online in a way that challenges the distinction between ‘for information’, ‘fit for purpose’ or ‘ready to publish’ texts. In addition, agile content management, user engagement metrics and the automated generation of content by large language models such as the GPT-3 model (Brown et al., 2020) rely on a combination of human linguistic production and machine translation. The nexus of relations between human and machine translation deserves reflexion and analysis (Ehrensberger-Dow, Massey 2019).

The pace and scale of technological, economic and societal developments are such that they raise a series of questions. The recent hype over neural MT contrasts with the lack of maturity of language processing technologies in some areas: we might mention in this respect, to varying degrees, the gap between high-resource languages and low-resource languages, the affordances of MT interfaces and the lack of information as to how training corpora are obtained.

More generally, the role, status and economic future of translators seem uncertain, the eco-sustainability of the translation model (Cronin 2017) is a moot point and the social responsibility of the main MT providers remains murky at best. As a consequence of these tectonic shifts, translator training is also at a crossroads: it needs to find ways of reconciling these new developments with more traditional skills and content (Froeliger 2019; Massey, Kiraly 2019).

Convinced that translation and language technologies play an essential role in society, we propose here to seek a new consensus between the human uses of language and the contributions of the machine; our aim is not only to enable exchanges and contribute to the development and dissemination of knowledge, but also to exercise our social sense of responsibility (Martens, Hobbs 2015). Furthering digital literacy in the field of language technologies and promoting a better understanding of the social, economic and ethical stakes are indeed imperative (Koskinen, Pokorn 2021; Moorkens, Rocchi 2021).

Such was the remit of Tralogy III, an international conference held in Paris in the spring of 2022, part of the proceedings of which have been included in this volume. The aim was to invite participants to identify blind spots in the current landscape of translation and multilingual language technologies, in terms of research, theory, practice and training, to suggest critical perspectives and devise innovative ways of overcoming them.

Following on from the first two editions of Tralogy, in 2011 and 2013, also held in Paris, the aim of this event was to bring together and promote dialogue between specialists from the various fields and disciplines involved in the translation professions of today and tomorrow: academics (from both Translation Studies and Computational Linguistics), professional translators (through the involvement of SFT, Société française des traducteurs), users of translation (the European Commission's Directorate-General for Translation) and translation trainers (AFFUMT, the French Association of University Courses in Translation, University of Paris VIII and Ca' Foscari University in Venice). This time, they were joined by representatives of the major technology companies, which are now key players in this sector.

The Tralogy I and II conferences were among the very first events to set the stage for a dialogue between these different communities, which until then mostly kept to their own pursuits. In fact, for Tralogy II, contributors were invited to present on what they could not achieve alone, rather than on what they could already achieve on their own.¹ Since then, many conferences and scientific events have adopted this format, a welcome evolution, because the task is immense and the only way to rise to the diversity of needs is to work together to find solutions.

The present bilingual collection will provide readers with a wealth of resources on the relations between human translation and automatic language processing.

The book is divided into three parts. The first part entitled “Setting the Scene”, gives an overview of the encounters between digital tools and translation, analysing both the successes and failures of these encounters, and charting future developments.

To start with, Giuseppe Sofò recounts the history of the encounter between translation and artificial intelligence. Then we move on to the characterisation of artificial intelligence as anthropomorphic, in which algorithms are described as conversational agents engaged in translation and text generation. This representation deserves to be questioned: is the so-called artificial intelligence actually a form of intelligence, is it similar to human cognition? Michael Carl and Roberto Laghi highlight the complexity of the cognitive phenomena involved. Michael Carl's contribution resituates and discusses *embodiment*, proposing a theoretical framework for a better understanding of cognitive interactions with the translational environment. Roberto Laghi investigates the relationship between computer science and cognition, through the notion of cognitive assembly.

1 See for example https://hal.science/hal-02497138v1/file/Tralogy_S0_A3_Froeliger.pdf (other presentations from the first two editions are available on the French HAL research platform).

The second section of the book, “Neural Translation and Translator Training”, is devoted to the challenges of translation within translator training.

Sara Cotelli Kureth, Alice Delorme Benites, Mara Haller, Hasti Noghrechi and Elizabeth Steele examine students’ usage and mis-usage of free translation platforms, notably when they treat automated translation as a reliable source of lexical information. Distinguishing between human translated texts and texts post-edited from automatic pre-translations is crucial in terms of student performance assessment and training: Hanna Martikainen addresses the relevance of automatic metrics in this area. On a larger scale, the European MultiTraiNMT project, presented by Caroline Rossi, Amélie Josselin-Leray, Nicolas Ballier and Hanna Martikainen, has made it possible to develop, evaluate and disseminate open-access content and free and open-source applications, particularly for teacher training. Finally, the very design of a master’s programme needs to evolve to make up for neural translation: Elena Kokanova, Maya Lyutyanskaya and Aleksandra Epimakhova share their experience and thoughts on this point.

The third section, entitled “Recomposing Fields”, examines neural machine translation integration in the field of specialised and institutional translation.

Nicolas Froeliger, Christopher Gledhill and Maria Zimina-Poirot address the societal impact of neural translation in general. From an institutional perspective, what are the implications of the recent shift to a neural translation tool within the European Union’s Directorate-General for Translation? Caroline Rossi and Aurélien Talbot conducted a series of semi-structured interviews that shed light on changes in translation practices within international institutions. Nicolas Ballier and Maria Zimina-Poirot demonstrate the value of a multidisciplinary approach to translation, which enables the transfer of technology from hard sciences to the humanities. The establishment of the PAPTAN research platform, in partnership with Systran, gives linguists access to very useful experimental tools (for corpus management, score calculation, etc.).

Of course, the reader will not find here a definitive statement on the relationship between machine translation and human translation, between ‘machine translators’ and ‘human translators’, or ‘bi-translators’. Rather, the aim of this collection of essays is to record a dialogue already taking place in the scientific and professional communities – as the Tralogy III conference so clearly demonstrated, enabling these two fields, which are too often separated, to enter into a continuous and fruitful dialogue – both inside and outside academia.

Far from being immune to the technologies they create and use, human beings co-evolve with technology (Stiegler 2018). Now is the

time to strive to understand the contribution of digital technologies to translation and automatic language processing, and above all to take up the challenge of guiding - or at least accompanying - their evolution, to keep in check further control and standardisation, and to spur on collective knowledge together with cultural, social and human empowerment.

In order to truly achieve this 'new consensus', all the stakeholders in the field of translation - professionals, trainers, researchers, and end-users - will have to work together to foster more and better interaction between machines and humans.

Bibliographie | Bibliography

- Bowker, L.; Buitrago-Ciro, J. (2019). *Machine Translation and Global Research. Towards Improved Machine Translation Literacy in the Scholarly Community*. Bingley: Emerald Publishing. <https://doi.org/10.1108/9781787567214>.
- Brown, T.B. et al. (2020). *Language Models Are Few-Shot Learners = Proceedings of the Conference on Advances in Neural Information Processing Systems (Vancouver, 6-12 December 2020)*. *NeurIPS*, 33. <https://papers.nips.cc/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf>.
- Cronin, M. (2017). *Eco-Translation. Translation and Ecology in the Age of the Anthropocene*. London; New York: Routledge. <https://doi.org/10.4324/9781315689357>.
- Ehrensberger-Dow, M.; Massey, G. (2019). « Le Traducteur et la machine. Mieux travailler ensemble? ». *Lavault-Olléon, Zimina 2019*, 47-62.
- Froeliger, N. (2019). « At a Certain Stage, One Has to Deliver. Why Professional Translation Masters' Matter ». *Cultus. The Journal of Intercultural Mediation and Communication*, 12, 35-55. http://www.cultusjournal.com/files/Archives/Cultus_2019_12_003_Froeliger.pdf.
- Hickey, S. (2023). *The 2023 Nimdzi 100. The Ranking of Top 100 Largest Language Service Providers*. <https://www.nimdzi.com/nimdzi-100-top-lsp/#market-size-growth>.
- Koskinen, K.; Pokorn, N.K. (eds) (2021). *The Routledge Handbook of Translation and Ethics*. London; New York: Routledge. <https://doi.org/10.4324/9781003127970>.
- Larssonneur, C. (2020). « Neural Machine Translation. From Commodity to Commons? » Desjardins, R.; Larssonneur, C.; Lacour, P. (eds), *When Translation Goes Digital. Case Studies and Critical Reflections*. Cham: Palgrave-Macmillan, 257-80. <https://doi.org/10.1007/978-3-030-51761-8>.
- Lavault-Olléon, É.; Zimina, M. (éds) (2019). « Traduction et technologie, regards croisés sur de nouvelles pratiques ». *Monogr. no., Des mots aux actes*, 8. <https://dx.doi.org/10.15122/isbn.978-2-406-09779-2>.
- Martens, H.; Hobbs, R. (2015). "How Media Literacy Supports Civic Engagement in a Digital Age". *Atlantic Journal of Communication*, 23(2), 120-37. <https://doi.org/10.1080/15456870.2014.961636>.
- Massey, G.; Kiraly, D. (2019). « The Future of Translator Education. A Dialogue ». *Cultus. The Journal of Intercultural Mediation and Communication*, 12, 15-34. https://www.cultusjournal.com/files/Archives/Cultus_2019_12_002_Kiraly_Massey.pdf.
- Moorkens, J.; Lewis, D. (2019). « Copyright and the Re-Use of Translation as Data ». O'Hagan, M. (ed.), *The Routledge Handbook of Translation and Technology*. London; New York: Routledge, 516-30. <https://doi.org/10.4324/9781315311258>.
- Moorkens, J.; Rocchi, M. (2021). « Ethics in the Translation Industry ». *Koskinen, Pokorn 2021*, 320-37.
- Stiegler, B. (2018). « *La Technique et le temps* » suivi de « *Le nouveau conflit des facultés et des fonctions dans l'anthropocène* ». Paris: Fayard.