

---

# Traitement de l'ellipse et des ambiguïtés catégorielles par deux programmes de TA neuronale en ligne de l'anglais vers le français, l'espagnol et l'italien

François Maniez<sup>\*†1</sup>, María Belén Villar Diaz<sup>‡2</sup>, and Sandra Garbarino<sup>§2</sup>

<sup>1</sup>Centre de Recherche en Linguistique Appliquée (CeRLA) – Université Lumière-Lyon 2 – France

<sup>2</sup>Centre de Recherche en Linguistique Appliquée (CeRLA) – Université Lumière-Lyon 2 – France

## Résumé

Les progrès de la traduction automatique (TA) et l'avènement de la traduction automatique neuronale (TAN) ont contribué à augmenter l'intérêt pour la post-édition dans les cursus d'enseignement de la traduction professionnelle. Les bons résultats obtenus pour certains couples de langues ont notamment fait émerger dans l'enseignement de la post-édition la pratique consistant à relire le texte cible sans systématiquement se référer au texte source (Ginovart 2020).

Les modèles de la TA traitent généralement les phrases de manière isolée et rejettent toute information contextuelle qui dépasserait les limites de la phrase (Tiedemann et Scherrer 2017), même si la TAN prend désormais en compte le contexte interphrastique (Voita *et al* 2019). Malgré les récents progrès de l'intelligence artificielle les traducteurs humains continuent de surpasser la TA pour les tâches suivantes :

- a) l'identification des référents dans l'anaphore (en particulier de type interphrastique) ;
- b) la résolution de l'ambiguïté sémantique générée par la polysémie ou l'homonymie ;
- c) la résolution des ambiguïtés syntaxiques qui ne peuvent être résolues par des méthodes statistiques (Maniez 2020).

Dans la prose scientifique, la première tâche pose peu de problèmes à la TA en raison notamment de l'absence de contraintes stylistiques qui empêcheraient la répétition (ce qui limite les cas d'anaphore). La prose journalistique est cependant plus riche en anaphores et en ellipses et peut donc provoquer davantage d'erreurs en TA.

A partir d'un corpus constitué de textes journalistiques traduits en 2016, 2019 et 2021 par les programmes de TA en ligne des sites Google Translate et DeepL, nous analysons l'évolution de la qualité des traductions proposées par ces deux programmes de l'anglais vers l'espagnol, le français et l'italien, notamment pour le traitement de l'ellipse intraphrastique et interphrastique et celui des ambiguïtés catégorielles. Nous examinons entre autres les cas

---

\*Intervenant

†Auteur correspondant: francois.maniez@univ-lyon2.fr

‡Auteur correspondant: maria-belen.villar-diaz@univ-lyon2.fr

§Auteur correspondant: sandra.garbarino@univ-lyon2.fr

où l'ellipse interphrastique implique la répétition partielle d'un terme complexe (*legal challenge* : recours en justice, impugnación legal), qui entraîne une traduction littérale erronée (*challenge* : défi, desafío) comme dans l'exemple suivant :

*A High Court legal challenge against Brexit will enter its second day on Monday [...]. The challenge will hear whether Theresa May has the power to take the UK out of the EU without MPs voting on it.*

Nos résultats font apparaître un saut qualitatif entre la TA statistique et la TAN, ainsi qu'une sensible amélioration des résultats de la TAN au cours des deux dernières années.

Maniez, F. (2020). "Identification of potentially ambiguous noun phrases in scientific English: a crucial aspect of translator and post-editor training", in Levey Daniel (dir.) *Strategies and Analyses of Language and Communication in Multilingual and International Contexts*, p. 168-176, Cambridge Scholars Publishing, Cambridge.

Voita, E., Sennrich, R., & Titov, I. (2019). "When a good translation is wrong in context: Context-aware machine translation improves on deixis, ellipsis, and lexical cohesion". *arXiv preprint arXiv:1905.05979*.

Tiedemann, J. and Scherrer, Y. (2017). "Neural Machine Translation with Extended Context". In *Proceedings of the Third Workshop on Discourse in Machine Translation, DISCOMT'17*, pages 82–92, Copenhagen, Denmark. Association for Computational Linguistics.

**Mots-Clés:** anaphore, ambiguïté, ellipse, post, édition, traduction automatique