

PHRASEOLOGIE, TERMINOLOGIE ET EVOLUTION TECHNOLOGIQUE DE LA TRADUCTION :

TRADUCTION MACHINE VS. TRADUCTION HUMAINE ANGLAIS>FRANÇAIS ET
ANGLAIS>ARABE

Encadrant de thèse :

Yvon KEROMNES (Université de Lorraine & ATILF-CNRS)
yvon.keromnes@univ-lorraine.fr

Ecole Doctorale :

Ecole Doctorale Sociétés, Langages, Temps, Connaissances

Laboratoire d'accueil :

Analyse et Traitement Informatique de la Langue Française (ATILF), UMR 7118

Sujet

État des lieux

Longtemps un sujet de plaisanterie tant ses productions laissaient à désirer, la traduction machine (TM) n'a cependant jamais cessé de s'améliorer, avec le temps et le renouvellement des approches. Récemment, le remplacement de la traduction statistique par la traduction neuronale (TMN), adoptée par *Google Translate* ou *Babelfish* par exemple, mais popularisée par l'arrivée de la Startup allemande *DeepL* en 2017, a en effet permis une amélioration considérable des traductions (Wu et al. 2016). Différents programmes de TMN sont aujourd'hui gratuitement accessibles en ligne, si bien que certains pensent pouvoir prédire la fin prochaine de la « biotraduction », ou traduction humaine (TH), et son remplacement par des machines.

En réalité, nous n'en sommes sans doute pas là, pour plusieurs raisons. Tout d'abord, même les systèmes les plus performants sont loin d'être parfaits, et ce n'est d'ailleurs pas le but. Comme le remarque Koehn (2020 : 19), « The goal of current machine translation research is not to achieve perfect translation but to drive down error rates of machine translation systems. » Ensuite, cette technologie d'apprentissage automatique repose sur la mise à disposition d'importantes quantités de données sous la forme de corpus parallèles, ce qui implique généralement un investissement financier considérable. Plus la masse de données est importante, plus la qualité des traductions est élevée. Ainsi *DeepL*, qui devance tous ses concurrents, doit cette avance aux corpus parallèles de *Linguee*, qui n'atteignent une taille suffisante que pour un nombre de langues relativement réduit. Enfin, il faut noter que les spécialistes de l'apprentissage machine eux-mêmes n'ont pas accès au fonctionnement algorithmique d'apprentissage et ne font que constater empiriquement les effets induits par les modifications de différents facteurs (Burlot & Yvon 2018). Il ne s'agit donc pas encore d'une connaissance stable et objective.

Cependant ces bouleversements, qui s'inscrivent dans une évolution technologique plus générale (Kelly 2010, 2016) sont assurément en train de changer le rapport du public avec la traduction automatique, ce qui a des implications dans l'enseignement des langues étrangères (Borsatti 2021), pour lesquelles par exemple la pratique de la traduction pédagogique doit être repensée, mais leur impact se fait davantage ressentir dans le métier de la traduction, et par là-même dans la formation à ces métiers : si cette formation intégrait déjà largement l'usage de la traduction assistée par ordinateur (TAO) depuis une vingtaine d'années, elle doit aujourd'hui également familiariser les étudiant.e.s avec la TMN, et l'activité humaine qui découle de l'usage croissant de la TM, la post-édition (Nitzke 2019, Rossi 2019).

Il s'agit donc de mieux comprendre l'impact de cette évolution technologique dans la formation des traducteurs et traductrices : l'activité de traduction étant envisagée comme résolution de problèmes et

prises de décision liées à l'accès à des ressources terminologiques et lexicographiques, elles-mêmes en profonde mutation du fait de cette évolution, quelle est la spécificité des problèmes posés et des ressources nécessaires dans la post-édition comparée à la traduction humaine ?

Recherche

Le premier objectif de la thèse est de caractériser la TMN de l'anglais vers deux langues-cibles très différentes quant aux ressources dont elles disposent en termes de corpus parallèles (français et arabe), et ce dans différents domaines de spécialités parmi les plus demandés dans la traduction professionnelle, ceux auxquels les étudiants en traduction sont déjà confrontés : droit, médecine, technologie et tourisme. Comment la TMN gère-t-elle les deux impératifs contradictoires de la fluidité (idiomaticité de la langue, phraséologie) et de l'exactitude (structure des domaines scientifiques et techniques, traitement des unités terminologiques).

Le deuxième objectif de la thèse est d'étudier l'activité de post-édition d'étudiant.e.s en traduction comparée à leur activité de traduction en termes de productivité et de besoins en ressources terminologiques et lexicographiques, ainsi que leur attitude face à ces deux activités.

Méthode

Dans un premier temps seront créés des corpus comparables (*WebBootCat*, Baroni *et al.* 2006) et des corpus parallèles dans les domaines étudiés afin d'établir, à l'aide de *SketchEngine* et de *WordSmith*, dans chacun de ceux-ci les spécificités de la TMN pour ce qui concerne la phraséologie au sens large (collocations, expressions polylexicales, constructions¹) et la terminologie. Des mesures statistiques seront effectuées pour évaluer ces différents paramètres (Baroni & Evert 2009, Brezina 2018).

Dans un deuxième temps, des groupes d'étudiant.e.s de première et deuxième année de Master traduction à l'Université de Balamand (Liban) et à l'Université de Lorraine se verront proposer en parallèle la traduction et la post-édition de textes de ces différents domaines, de l'anglais vers l'arabe (Liban) et de l'anglais vers le français (UL). Ce travail se fera sur le logiciel *TRADOS*, qui est le programme de TAO le plus employé dans le milieu de la traduction professionnelle. Différentes méthodes seront mises en œuvre pour ces observations, en particulier l'enregistrement de frappes (Key-logging) et des questionnaires (Alves 2003).

Profil attendu du candidat ou de la candidate

Le/la candidat.e devra posséder une affinité particulière pour la linguistique cognitive et la linguistique de corpus, ainsi qu'une expérience en langues de spécialité et en traduction (y compris l'utilisation des outils TAO).

L'équipe de recherche devant accueillir le/la doctorant(e) est l'équipe « Discours ». Cette équipe regroupe les compétences requises pour cette thèse.

Langues : maîtrise du français, de l'anglais et de l'arabe.

Références

- Almansor, E. H. (2018). *Translating Arabic as low resource language using distribution representation and neural machine translation models* (Thèse de doctorat).
- Alves, F. (2003). *Triangulating Translation*, Amsterdam : Benjamins.
- Baroni, M. et al. (2006). « WebBootCat : a web tool for instant corpora », *Proceedings Euralex 2006*.
- Baroni, M. & Evert, S. (2009). « Statistical methods for Corpus Exploitation », *Corpus Linguistics, an international handbook 2*, Berlin : de Gruyter, 777-803.
- Borsatti, D. & Blanco Riess, A. (2021). « Using machine translator as a pedagogical resource in English for specific purposes courses in the academic context », *Rev. Estud. Ling.*, Belo Horizonte, 29/2, 829-858.

¹ Au sens de Goldberg (1995, 2006), et en particulier ce que Tutin & Kraif (2016) nomment les « routines sémantico-rhétoriques », qui caractérisent différents domaines de spécialité.

- Brezina, V. (2018). *Statistics in Corpus Linguistics – A Practical Guide*, Cambridge : Cambridge University Press.
- Brunner, A. & Steyer, K. (2007). « Corpus-Driven Study of Multi-Word Expressions Based on Collocations from a Very Large Corpus », *Proceedings of the 4th Corpus Linguistics Conference*, Birmingham : Birmingham University Press.
- Burlot, F. & Yvon, F. (2018). « Évaluation morphologique pour la traduction automatique : adaptation au français », JEP-TALN-RECTAL.
- Faber, P. (éd.) (2012). *A Cognitive Linguistics View of Terminology and Specialized Language*, Berlin: de Gruyter.
- Gablasova, D., Brezina, V., & McEnery, T. (2017). « Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence », *Language learning*, 67(S1), 155-179.
- Hanks, P. (2009). « The impact of corpora on dictionaries », *Contemporary corpus linguistics*, 214-236.
- Goldberg, A. (1995). *Constructions – A Construction Grammar Approach to Argument Structure*, Chicago : University of Chicago Press.
- Goldberg, A. (2006). *Constructions at Work – The Nature of Generalization in Language*, Oxford : Oxford University Press.
- Kadhim, K. A., Habeeb, L. S., Sapar, A. A., Hussin, Z., & Abdullah, M. R. T. L. (2013). « An Evaluation of Online Machine Translation of Arabic into English News Headlines: Implications on Students' Learning Purposes », *Turkish Online Journal of Educational Technology-TOJET*, 12(2), 39-50.
- Kelly, K. (2010). *What Technology Wants*, New York : Viking Press.
- Kelly, K. (2016). *The Inevitable – Understanding the 12 Technological Forces that Will Shape our Future*, New York : Viking Press.
- Keromnes, Y. (2007). « Bases de données et traductions : aspects quantitatifs et qualitatifs », In : G. Williams (éd.), *Actes des 4^{èmes} Journées de Linguistique de Corpus*, Université Bretagne Sud, 231-238 (Actes publiés en ligne : <http://www.licorn-research.fr/jlc4/acteJLC2005.pdf>)
- Keromnes, Y. (2010). « Traduction professionnelle et enseignement de la traduction en France », in : La traduction : de la linguistique à la didactique, Université Charles de Gaulle-Lille 3 <<http://stl.recherche.univ-lille3.fr/textesenligne/textesenlignecadreprincipal.html>>.
- Keromnes, Y. (2013). « Expressivité et économie des expressions idiomatiques », *Annales littéraires de l'Université de Franche-Comté. Série Littérature et histoire des pays de langues européennes*, 3, 81-91.
- Keromnes, Y. (2016). « Where linguistics meets translation studies – a mootable point », in: M. Boisseau, C. Delesse, C. Chauvin & Y. Keromnes. (éds.), *Linguistique et traductologie : les enjeux d'une relation complexe*, Arras : Presses Universitaires d'Artois, 39-53.
- Koehn, P. (2020). *Neural Machine Translation*. Cambridge : Cambridge University Press.
- Lerat, P. (2016). *Langue et Technique*. Paris : Hermann.
- Lewandowska-Tomaszczyk, B., & Pezik, P. (2018). « Parallel and comparable language corpora, cluster equivalence and translator education », in : *Общество и языки в третьем тысячелетии. Коммуникация. Образование. Перевод= Society and Languages in the Third Millennium. Communication. Education. Translation*, 131-143.
- Liu, Z. (2020). « Research on the Extraction Method of Translation Equivalence in Chinese and English Comparative Corpus », *International Journal of Early Childhood Special Education (INT-JECSE)*, 29(5), 1651.
- McCarthy, B. (2004). « Does online machine translation spell the end of take-home translation assignments », *CALL-EJ Online*, 6(1), 6-1.
- Nitzke, J. (2019). *Problem Solving Activities in Post-Editing and Translation from Scratch – A Multi-Method Study*, Berlin : Language Science Press.
- Philip, G. (2008). « Arriving at equivalence. Making a case for comparable general reference corpora in Translation Studies », in : A. Beeby et al. (éds), *Corpus Use and Translating*, Amsterdam : Benjamins, 59-73..
- Rossi, C. (2019). « L'apprenti traducteur et la machine: des connaissances aux perceptions de la traduction automatique ». *Des mots aux actes* 8, 93-105.
- Seiler, M. & Markantonatou, S. (éds) (2018). *Multiword Expressions – Insights from a multi-lingual perspective*, Berlin : Language Science Press.
- Tutin, A. & Kraif, O. (2016). « Routines sémantico-rhétoriques dans l'écrit scientifique de sciences humaines : l'apport des arbres lexico-syntaxiques récurrents », *LIDIL* 53, 119-141.

- Váradi, T., & Kiss, G. (2001). « Equivalence and non-equivalence in parallel corpora », *International journal of corpus linguistics*, 6 (special issue), 167-177.
- Wu Y. et al. (2016). « Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation », *arXiv preprint arXiv: 1609.08144*.