



HAL
open science

La traduction automatique neuronale, un nouveau tournant ?

Franck Barbin

► To cite this version:

Franck Barbin. La traduction automatique neuronale, un nouveau tournant ?. Palimpseste. Sciences, humanités, sociétés , 2020, 4, pp.51-53. <halshs-03603588>

HAL Id: halshs-03603588

<https://shs.hal.science/halshs-03603588v1>

Submitted on 18 Mar 2022

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.



HAL Authorization

Palimpseste

sciences • humanités • sociétés

numéro 4

printemps-été 2020



Archives de la recherche
à l'université Rennes 2

PALIMPSESTE :

grec παλίμψηστος (*palimpsèstos*)

« qu'on gratte pour écrire à nouveau »

DOSSIER

4 De la marginalisation linguistique et sociale

Philippe Blanchet, Stéphanie Clerc Conan, Gudrun Ledegen, Claire Lesacher, Matthieu Marchadour, Nadia Ouabdelmoumen & Pauline Rannou

10 Langues et cultures minoritaires : *haq a-nevez* ?

Erwan Hupel

14 La littérature arthurienne tardive en Europe : un objet de recherche marginal ?

Christine Ferlampin-Acher

18 L'art comme sociologie du salariat

Pascale Borrel

24 Design et grande précarité

Rahaf Demashki

27 *Underground*. Ethnographier la carrière de musicien du métro

Alina Bystritskaya-Maintenant

30 L'organisation du travail des web-créatifs nomades

Anne-France Kogan & Marcela Patrascu

33 Travailler à la marge ?

Gerhard Krauss & Anne-Laure Le Nadant

36 La deuxième chance : une marge qui tient la page ?

Maël Loquais

40 Migrations et inégalités sociales. Une relation complexe entre le Mexique et les États-Unis

Pascal Sebille & René Zenteno

46 Maladies rares et sciences humaines : un enjeu scientifique et social

Agnès Lacroix, Mélissa Burgevin, Virginie Dardier & Sylvie Odent

VARIA

51 La traduction automatique neuronale, un nouveau tournant ?

Franck Barbin

54 *Marginalia*

Ann-Marie Hansen

La traduction automatique neuronale, un nouveau tournant ?

PAR FRANCK BARBIN*

LE MYTHE DE LA MACHINE supplantant l'humain n'a jamais été autant d'actualité. Il faut avouer qu'un pas de géant a été franchi en 2015 avec le passage de la traduction automatique statistique (TAS) à la traduction automatique neuronale (TAN). Se démarquant clairement de la TAS, la TAN fonctionne sur le modèle du cerveau humain au travers de l'intelligence artificielle. Le saut qualitatif est indéniable : la TAN permet de générer des traductions de meilleure qualité (plus précises et naturelles) qui se rapprochent de la traduction humaine.

Cette amélioration provient du fait que les moteurs de traduction automatique sont désormais basés sur l'apprentissage alors qu'ils étaient basés auparavant sur des dictionnaires et des modèles statiques construits à partir de corpus découpés en unités de traduction. Ils sont désormais capables d'apprendre en temps réel : le logiciel s'améliore au fur et à mesure que l'on insère des segments à traduire. Cet apprentissage profond est rendu possible grâce aux réseaux de neurones [POIBEAU, 2019]. Il convient toutefois de l'alimenter en permanence avec un volume considérable de données de qualité bilingues afin de continuer à améliorer la fiabilité des résultats obtenus.

La TAN repose sur trois principes essentiels : le prolongement de mots, le modèle d'attention et les réseaux neuronaux récurrents (RNN). Le prolongement de mots permet d'obtenir une représentation continue (carte langagière) des mots qui apparaissent dans des contextes semblables. Le mécanisme attentionnel rend possibles une focalisation sur les éléments linguistiques importants et une collecte des informations sur eux, ce qui offre une meilleure prise en compte du contexte. Les RNN, en reproduisant le système d'apprentissage par essai-erreur, permettent de diffuser l'information et d'apporter ainsi de la connaissance contextuelle. En d'autres termes, la TAN imite le fonctionnement du cerveau humain

(création de voies neuronales) pour générer des traductions plus naturelles et idiomatiques.

Cette montée en puissance de la traduction automatique (TA) est flagrante dans le dernier rapport de l'EUATC (European Union of Associations of Translation Companies) publié en 2018 : il n'y a plus que 31 % des sociétés de traduction qui n'utilisent jamais la TA pour leurs prestations, alors que 22 % l'emploient quotidiennement. Cette évolution se fait également sentir du côté des clients : une part croissante exige l'utilisation de la TA et demande de réaliser des prestations de post-édition [VOIR L'ENCADRÉ P. 52]. Cette tendance devrait encore s'accroître dans les rapports à venir. Il est indéniable que la traduction automatique gagne du terrain dans le secteur de la traduction, pour répondre aux exigences de coûts et de délais des clients. Les économies engendrées par la post-édition l'emportent souvent sur la baisse de qualité du texte final.


Il semble évident, toutefois, que les prestations de post-édition ne pourront pas s'appliquer à tous les types de documents : les documents fortement rédactionnels (comme ceux du domaine marketing) et les supports nécessitant un travail de création et d'adaptation culturelle (par exemple les jeux vidéo) ne seront guère concernés, au contraire de certains documents techniques dont la structure est souvent identique (manuels d'utilisation, aides en ligne). L'agence de traduction devra également assumer son rôle de conseil auprès du client, en le mettant en garde contre les conséquences du choix de la post-édition plutôt que de la traduction sur la qualité finale du texte, ce dont il n'a pas forcément conscience.

Après l'apparition des outils de traduction assistée par ordinateur (TAO) dans les années 1990, le marché de la traduction doit faire face à un nouveau défi. Une fois n'est pas coutume, le développement de la TA, et plus particulièrement de la TAN, va modifier le modèle traditionnel dans l'industrie des langues et accélérer « le passage d'une économie de produits à une économie de services » : le traducteur va devenir « un véritable consultant en ingénierie linguistique » [PERRIGAUD, 2014]. Il va devoir se réinventer et trouver sa légitimité dans ce nouveau processus. Même si la traduction automatique exerce une pression croissante sur les traducteurs, il revient aux

*Maître de conférences en traduction, membre de l'unité de recherche Linguistique, ingénierie, didactique des langues (LIDILE, EA 3874). — On trouvera les références complètes des travaux cités p. 53.

traducteurs humains de faire valoir leur expertise linguistique et disciplinaire : ils doivent démontrer leur « valeur ajoutée » par rapport à de tels systèmes [Loock, 2018].

Cette adaptation aux nouvelles tendances du marché de la traduction passe également par une évolution des formations de traducteurs. Il est nécessaire que les futurs traducteurs soient conscients des avantages et des limites de la TAN. En outre, leur formation devra se focaliser

encore davantage sur les compétences connexes au métier de traducteur : communication interculturelle, rédaction ciblée utilisateur, transcréation, assurance qualité, etc. [MASSEY & EHRENSBERGER-Dow, 2017]. Il s'agit enfin de développer ces fameuses compétences interpersonnelles ou comportementales (*soft skills*) qui font partie du savoir-être professionnel tant attendu sur le marché du travail. 

« Il me semble (...) que la traduction ne pose pas seulement un travail intellectuel, théorique ou pratique, mais un problème éthique. Amener le lecteur à l'auteur, amener l'auteur au lecteur, au risque de servir et de trahir deux maîtres, c'est pratiquer ce que j'aime appeler l'*hospitalité langagière*. »

Paul Ricœur

LA TAN COUPLÉE À LA POST-ÉDITION HUMAINE

La traduction automatique neuronale (TAN) ou *Neural Machine Translation* (NMT) est une technologie basée sur les réseaux de neurones artificiels. Elle a fait des progrès importants ces dernières années grâce à l'intelligence artificielle, par rapport à d'autres formes de traduction automatique. Les meilleurs outils permettent ainsi d'obtenir sur certains segments de traduction une qualité comparable à une bonne traduction humaine. Cette avancée est rendue possible par la capacité de la machine à perfectionner constamment son apprentissage d'une langue, à la manière du cerveau humain.

La post-édition est l'étape durant laquelle des traducteurs professionnels relisent le texte produit par la traduction automatique et le corrigent pour supprimer notamment les erreurs sémantiques et linguistiques. Même s'il est encore difficile de prédire avec précision les erreurs produites par la traduction automatique neuronale (cette technologie étant récente), des éléments tels que l'exactitude sémantique, la cohérence terminologique, l'harmonisation du style ou l'adaptation culturelle, seront à vérifier systématiquement par les traducteurs-correcteurs.

UN PROJET D'OUTIL DE TRADUCTION AUTOMATIQUE DE MÉTADONNÉES ET SON INSERTION DANS LA CHAÎNE ÉDITORIALE (2019-2021)

La collaboration entre trois partenaires – LIDILE, MSHB et PUR* – ambitionne de mettre au point une méthode économe et efficace de traduction qui pourra bénéficier à l'ensemble des chercheurs et des professionnels au service des revues scientifiques. Il s'agit de concevoir une méthode qui combine la traduction automatique neuronale ou TAN (DeepL) [VOIR PAGE CI-CONTRE] et la post-édition humaine pour améliorer la qualité des métadonnées des articles (résumés, mots-clés, titraire, etc.) du français vers l'anglais dans le processus éditorial des revues. L'objectif est de développer une méthodologie de traduction qui peut être reproduite et transférée à d'autres revues et domaines disciplinaires.

Sur la base des métadonnées d'articles publiés en 2017 dans quatre revues des PUR appartenant respectivement aux domaines de l'histoire, de l'archéologie, de la didactique et de la géographie (*Annales de Bretagne et des Pays de l'Ouest*, *ArcheoSciences*, *Éducation & didactique* et *Norois*), il s'agit de comparer la traduction anglaise de ces métadonnées précédemment publiée (et jugée perfectible) avec la traduction des mêmes données générée par la TAN par le biais de la grille d'évaluation de la qualité TRASILT (Toudic *et al.*, 2014), afin de déterminer les éléments qualitatifs et les limites de chaque production (humain vs TAN). Les métadonnées d'autres articles de 2017 seront ensuite traduites par DeepL puis seront révisées et améliorées par des traducteurs professionnels (post-édition humaine de la TAN). Notre objectif est de concevoir la méthode de traduction la plus appropriée pour ce type de publication scientifique. Cette méthode sera ensuite testée sur les numéros des quatre revues à paraître en 2020.

* Respectivement, l'unité de recherche Linguistique, ingénierie, didactique des langues (LIDILE, EA 3874), l'unité de service et de recherche Maison des sciences de l'homme en Bretagne (MSHB, USR 3549, CNRS) et le service d'activités industrielles et commerciales Presses universitaires de Rennes (PUR, SAIC édition).



Source : www.deeppl.com. Capture d'écran réalisée le 16 juin 2020.

EXEMPLE DE TRADUCTION VIA DEEPL

Est reproduite ci-dessus une capture d'écran du site DeepL (www.deeppl.com), un des services de traduction automatique neuronale (TAN) en ligne. Dans la colonne de gauche, on a saisi le résumé d'un article de recherche publié par les Presses universitaires de Rennes intitulé « Des consignes aux activités des élèves. Une étude de cas sur les phases de la Lune en Cycle 3 » (Karine Bécu-Robinault et Olivier Robin, *Éducation & didactique* [en ligne], 2017 / 3 [vol. 11], p. 31). Dans la colonne de droite, on lit la traduction générée automatiquement par DeepL. Le résultat est dans l'ensemble fidèle au texte de départ, même si l'on peut déplorer une erreur de sens majeure : « *instructions, considered as incentives for teacher*

action ». Les consignes de l'enseignant doivent être considérées comme des incitations à l'action (pour les élèves) et non pour l'enseignant. L'intelligence artificielle se heurte également à des problèmes typiques d'adaptation culturelle : la « classe de CM2 » est rendue par « *in class of CM2* », ce qui devient incompréhensible pour un public anglophone. Il faudrait parler, si l'on se situe dans un contexte éducatif américain, de « *in a 5th grade class* ».

SYSTÈME DE TRADUCTION AUTOMATIQUE NEURONALE DEEPL

AVANTAGES

- Qualité de la traduction reconnue
- Entraînement de ses propres réseaux neuronaux sur des milliards de segments de traduction
- Protection des données communiquées (cryptage des données et absence de stockage)
- Version gratuite adaptée aux besoins des chercheurs
- Génération de documents entièrement éditables
- Intégration à des logiciels de traduction assistée par ordinateur (TAO)

INCONVÉNIENTS

- Entraînement limité à l'alimentation de l'outil
- Pas de possibilité d'adosser un glossaire ou une base terminologique
- Pas de possibilité d'accès au code source

Références bibliographiques

- EUROPEAN UNION OF ASSOCIATIONS OF TRANSLATION COMPANIES, *Language Industry Survey. Expectations and Concerns of the European Language Industry*, 2018 ; en ligne : ec.europa.eu/info/sites/info/files/2017_language_industry_survey_report_en.pdf.
- LOOCK, Rudy, « Traduction automatique et usage linguistique : une analyse de traductions anglais-français réunies en corpus », *Meta. Journal des traducteurs*, vol. 63, n° 3 : *Traductologie de corpus : 20 ans après*, décembre 2018, p. 786-806 ; en ligne : erudit.org/fr/revues/meta/2018-v63-n3-meta04634/1060173ar.pdf.
- MASSEY, Gary & EHRENSBERGER-DOW, Maureen, « Machine learning : Implications for translator education », *Lebende Sprachen*, vol. 62, n° 2, septembre 2017, p. 300-312 ; en ligne : digitalcollection.zhaw.ch/bitstream/11475/3597/2/Machine_learning_implications.pdf.
- PERRIGAUD, Anne-Charlotte, « Mémoire de traduction, TAO, post-édition et traitement automatisé de la langue : le traducteur en quête de sens... », *Tralogy, Tralogy II, Session 3 : Machine and Human Translation : Finding the Fit ? / TA et Biotraduction*, 2013, mis à jour le 1^{er} octobre 2014 ; en ligne : lodel.irevues.inist.fr/tralogy/index.php?id=300.
- POIBEAU, Thierry, *Babel 2.0. Où va la traduction automatique ?*, Paris, Odile Jacob, 2019.
- TODIC, Daniel, Katell HERNANDEZ MORIN, Fabienne MOREAU, Franck BARBIN & Gaëlle PHUEZ, « Du contexte didactique aux pratiques professionnelles : proposition d'une grille multicritères pour l'évaluation de la qualité en traduction spécialisée », *ILCEA*, n° 19 : *Acceptabilité et transgression en langues et cultures de spécialité*, 2014 ; en ligne : journals.openedition.org/ilcea/2517.



PLACE DU RECTEUR HENRI LE MOAL
CS 24307 – 35043 RENNES CEDEX
+33 (0)2 99 14 10 00
WWW.UNIV-RENNES2.FR

Palimpseste

sciences • humanités • sociétés

DIRECTEUR DE LA PUBLICATION Olivier David, président de l'université Rennes 2
RÉDACTEUR EN CHEF Leszek Brogowski, vice-président SHS et Politique de site
SECRÉTARIAT DE RÉDACTION & CONCEPTION GRAPHIQUE Cyrille Habert

COMITÉ ÉDITORIAL

REPRÉSENTANTES DE LA COMMISSION DE LA RECHERCHE

Gwénola Druel • Gudrun Ledegen • Claudia Zudini

REPRÉSENTANTS DES UNITÉS DE FORMATION ET DE RECHERCHE

Christine Ferlampin-Acher (UFR Arts, lettres, communication) • Anne Goarzin (UFR Langues)
Stéphane Héas (UFR Sciences et techniques des activités physiques et sportives)
Gaël Henaff (UFR Sciences sociales) • Geneviève Lameul (UFR Sciences humaines)

REPRÉSENTANTES DES CHERCHEURS CNRS

Anne Atlan & Anne-Cécile Hoyez (UMR 6590 ESO)

REPRÉSENTANTS DE L'ÉQUIPE DE DIRECTION

Marc Bergère (vice-président Documentation et Transition numérique)
Benoît Bideau (vice-président Valorisation)
Philippe Blanchet (président du conseil académique)
Lesley Lelourec (vice-présidente Internationalisation)
Gaïd Le Maner-Idrissi (vice-présidente Recherche)
Jacques Oulhen (vice-président Formation et Vie universitaire)

REPRÉSENTANTS DES DOCTORANTS

François Bignon • Charleyne Caroff • Solenne Derigond

DIRECTION DES CHAMPS LIBRES

Corinne Poulain

SERVICES

Service communication

Anaïg Blum Le Coat & Reine Paris

Direction de la recherche et de la valorisation

Yvan Coadour

Ce quatrième numéro de *Palimpseste* a été imprimé sur les presses de la DILA
(26, rue Desaix, 75015 Paris) en septembre 2020
pour le compte de l'université Rennes 2.

La version numérique est disponible à l'adresse suivante :

www.univ-rennes2.fr/recherche/palimpseste

ISSN 2680-4549 (IMPRIMÉ) • ISSN 2592-6705 (EN LIGNE) • DÉPÔT LÉGAL : SEPTEMBRE 2020

drv-palimpseste@univ-rennes2.fr