

# HIGH-TECH

## L'INTELLIGENCE ARTIFICIELLE MUSCLE LES SITES DE TRADUCTION EN LIGNE

Le Figaro · 12 mar. 2018 · DIDIER SANZ @sanzdidier

Lorsqu'on travaille sur ordinateur ou avec un smartphone, il faut bien avouer que les applications et les sites Web de traduction nous facilitent la vie quand on doit répondre à un correspondant allemand, comprendre une documentation en japonais, remplir un formulaire en espagnol ou rédiger un courrier à un commerçant chinois... Ces services ont évolué de façon spectaculaire.



On se souvient encore des premiers traducteurs automatiques, qui se contentaient de remplacer les mots d'une phrase par leur équivalent dans la langue souhaitée. Une traduction mot à mot qui donnait souvent des résultats comiques, comme «How tall are you ? » traduit en « Combien grand es-tu? ». Les techniques basées sur des approches statistiques ont amélioré les résultats. Elles comparent la phrase à traduire avec une multitude de phrases stockées dans un registre et fournissent la traduction qui correspond le mieux. La base de données est, en général, construite à partir de documents officiels de l'ONU ou du Parlement européen dont les textes sont publiés en plusieurs langues. Seulement, cette approche, qui demande beaucoup de ressources et de temps, atteint vite ses limites dans le cas de tournures inusuelles ou si l'une des langues ne figure pas dans la base de données, auquel cas le système doit passer par une langue intermédiaire, ce qui altère forcément le sens de la phrase.

Pour résoudre ces problèmes, les méthodes les plus récentes, elles, s'appuient sur l'intelligence artificielle. Le principe: on entraîne un réseau de neurones artificiels à coder des millions de phrases de manière à produire pour chacune une séquence numérique unique. Dès qu'un utilisateur soumet une phrase au système de traduction, celui-ci va alors récupérer la séquence numérique correspondante et fournir instantanément la traduction dans la langue souhaitée. D'autres astuces complètent le traitement, parmi lesquelles des processus de prédiction, de mémorisation et de déduction. «De cette façon, le réseau de neurones aborde le problème de la traduction selon une approche traditionnelle basée sur les phrases, expliquent des chercheurs de Google. Il peut comprendre des dépendances inhérentes aux langues, par exemple les accords de genre, les structures syntaxiques, etc. »

Le principe est tellement prometteur qu'il est désormais adopté par plusieurs systèmes de traduction informatisée. Outre Google Traduction, on le retrouve notamment dans le projet Open NMT, un

outil de traduction neuronale en Open Source, lancé fin 2016 par Systran et des chercheurs de Harvard. Principalement utilisé dans les secteurs de l'enseignement et de l'entreprise, il peut être entraîné pour s'adapter au vocabulaire juridique, médical, financier ou autre. De son côté, DeepL, créé par l'équipe européenne qui a lancé le service de traduction Linguee, est davantage conçu pour le grand public et se présente comme le concurrent direct de Google Traduction.

Pour évaluer l'efficacité des deux principaux acteurs en présence, nous leur avons soumis un extrait de texte littéraire et des phrases courantes en anglais et en espagnol, puis nous avons inversé le sens de traduction et demandé à l'un de traduire la réponse de l'autre. Bilan: on mesure assurément les progrès réalisés par ces outils. Non seulement la traduction s'effectue à grande vitesse, mais le résultat est parfois impressionnant. Certes, le style n'est pas toujours très élégant, mais on peut clairement saisir le sens original. Aussi bien Google Traduction que DeepL ont remarquablement restitué un passage du Parfum de la dame en noir, de Gaston Leroux, qu'ils ont d'abord traduit vers l'anglais puis retraduit en français. À quelques détails près («terrible» ou «affreux» à la place d'«effroyable», par exemple), on retrouve pratiquement toutes les expressions de l'auteur. En ce qui concerne les phrases courantes, les résultats sont plus mitigés. Confronté à «He swims across the river», Google traduit par «Il nage à travers la rivière», alors que DeepL propose plus convenablement «Il traverse la rivière à la nage». Ce dernier marque aussi des points en traduisant «This is not a good time for us to talk» par «Ce n'est pas le bon moment pour parler» quand Google s'embrouille avec «Ce n'est pas un bon moment pour nous de parler». Et DeepL l'emporte encore en traduisant «I am heading to tennis practice» par «Je vais à l'entraînement de tennis», alors que Google se satisfait de «Je vais à la pratique du tennis».

En espagnol, les deux traducteurs ont quelques difficultés avec des phrases courantes. Quand on leur soumet «Son las cuatro menos cinco» («Il est quatre heures moins cinq»), Google propose «Il est quatre à cinq», tandis que DeepL préfère «C'est cinq contre quatre»... Une phrase un peu épineuse de Jorge Luis Borges ne les départage pas plus: «Uno no es lo que es por lo que escribe, sino por lo que ha leído», qu'on peut traduire par «On n'est pas ce qu'on est par ce qu'on écrit, mais

Pour l'instant, DeepL se limite à une demi-douzaine de langues, alors que Google en propose une bonne centaine

par ce qu'on a lu», devient, avec Google, «L'un n'est pas ce qu'il est par ce qu'il écrit, mais par ce qu'il a lu» et, avec DeepL, «L'un n'est pas ce qu'il écrit, mais ce qu'il a lu». Enfin, DeepL montre encore sa supériorité dans le test littéraire, restituant correctement le passage de Gaston Leroux après traduction en espagnol, tandis que Google peine sur la structure des phrases et confond certains mots avec d'autres, ce qui complique la lecture.

Pour l'instant, DeepL se limite à une demi-douzaine de langues (français, anglais, allemand, espagnol, italien, néerlandais, polonais), alors que Google en propose une bonne centaine, de l'afrikaan au zoulou, en passant par le bengali, le créole haïtien, le hongrois et le tamoul. Google Traduction peut également énoncer le texte dans la langue souhaitée, ce que ne propose pas DeepL. Du fait de leur conception, les deux services vont inévitablement évoluer et devraient s'approcher de plus en plus des traductions humaines. La barrière de la langue ne sera bientôt plus un obstacle.