

DOSSIER SPÉCIAL : TRADUCTION AUTOMATIQUE

Traduction automatique en littérature :

l'ordinateur va-t-il nous voler notre travail ?

James Hadley

Après la Seconde Guerre mondiale, des expériences débutèrent dans le domaine de la traduction automatique. En 1947, Warren Weaver, mathématicien américain, énonçait, dans un mémorandum fondateur, l'idée selon laquelle un calculateur numérique pourrait servir à traduire le langage humain. Pendant la guerre, à Bletchley Park, haut lieu britannique du déchiffrement, les Alliés avaient cassé les codes de messages nazis à l'aide d'une batterie de machines à calculer (dont des « bombes » et des Colossus). En faisant un parallèle entre décodage et traduction, il était facile d'imaginer des machines aptes à restituer dans une langue un texte rédigé dans une autre.

Pendant les années 1950 et au début des années 1960, les recherches visant à la création de programmes de traduction automatique, principalement dans

la paire de langues anglais-russe, accédèrent de part et d'autre du rideau de fer au rang de priorité nationale en matière de sécurité. Cette période vit notamment, en 1954, l'expérience Georgetown-IBM, qui consista à faire traduire par un système par règles une soixantaine de phrases du russe vers l'anglais. Après pareille réussite pour l'époque, on proclama avec assurance qu'il suffirait de trois à cinq ans pour résoudre le problème de la traduction automatique.

Multiples exceptions aux règles linguistiques

Les systèmes par règles, y compris les dictionnaires bilingues, de même que les règles logiques du maniement d'informations textuelles, se fondaient sur les méthodes traditionnelles d'enseignement des langues. Or, comme le sait quiconque a appris une langue

étrangère, les règles linguistiques s'assortissent d'une multitude d'exceptions. Résultat : ces systèmes s'avèrent bientôt lourds, lents et truffés d'erreurs. En 1966, le Comité consultatif sur le traitement automatique de la langue ou ALPAC (*Automatic Language Processing Advisory Committee*) en vint à la conclusion que, malgré de substantiels investissements, les outils de traduction automatique ne pourraient dans un avenir proche atteindre le niveau de traducteurs humains. Mieux valait axer les recherches sur le développement d'outils d'aide à la traduction, lesquels seraient plus tard connus sous le nom de logiciels de traduction assistée par ordinateur (TAO), comme, par exemple, Trados.

C'est pourquoi, pendant une dizaine d'années, les recherches menées aux États-Unis en matière de traduction automatique avancèrent au ralenti. Elles se poursuivaient toutefois dans d'autres pays, sur un nombre restreint de langues comme l'anglais ou le français. Le Canada créa ainsi le programme MÉTÉO, qui servit de 1977 à 2001 à traduire les prévisions météorologiques entre les deux langues officielles du pays. À la même époque, les systèmes par règles cédaient la place aux systèmes de traduction automatique statistique (TAS), reposant non pas sur des règles codées manuellement mais sur l'alignement des phrases composant deux textes mis en parallèle. À partir des vastes corpus construits par ce procédé, l'ordinateur produit une traduction. Ces systèmes ont d'abord fonctionné mot à mot, puis phrase par phrase. Ils sont relativement efficaces pour des paires de langues assez semblables et pour lesquelles il existe une abondance de

sources d'où extraire en grand nombre les phrases à aligner pour construire le corpus. À l'inverse, les résultats obtenus sont moins satisfaisants pour des langues dans lesquelles l'ordre des mots diffère considérablement ou pour lesquelles on dispose de peu de données à mettre en parallèle.

Statistiques ou neurones ?

En 2014, ces systèmes ont été à leur tour détrônés par les systèmes de traduction neuronale (ou NMT, *Neural Machine Translation*). Ces nouveaux programmes reposent eux aussi sur de vastes corpus de phrases alignées, dans les deux langues concernées. La différence est que les systèmes neuronaux sont conçus sur le même modèle que celui par lequel communiquent les neurones du cerveau humain : le produit final résulte de la réunion d'un grand nombre de petits processus. De ce fait, et alors que les systèmes statistiques utilisent les corpus comme ingrédients de leurs traductions, les moteurs neuronaux s'en servent pour apprendre à traduire par eux-mêmes. Ces nouveaux systèmes travaillent plus vite et produisent des traductions de bien meilleure qualité, impossibles à distinguer de textes écrits par des traducteurs humains – à condition qu'on leur fournisse suffisamment de données de base pour leur apprentissage.

L'heure du *Game over* a-t-elle donc sonné pour les traducteurs humains ? Eh bien non.

Si les moteurs neuronaux ont démontré leur extrême efficacité lorsqu'il s'agit de traduire certains types de textes, surtout composés de phrases courtes et d'expressions toutes faites, le champ de leurs exploits demeure limité. Cela

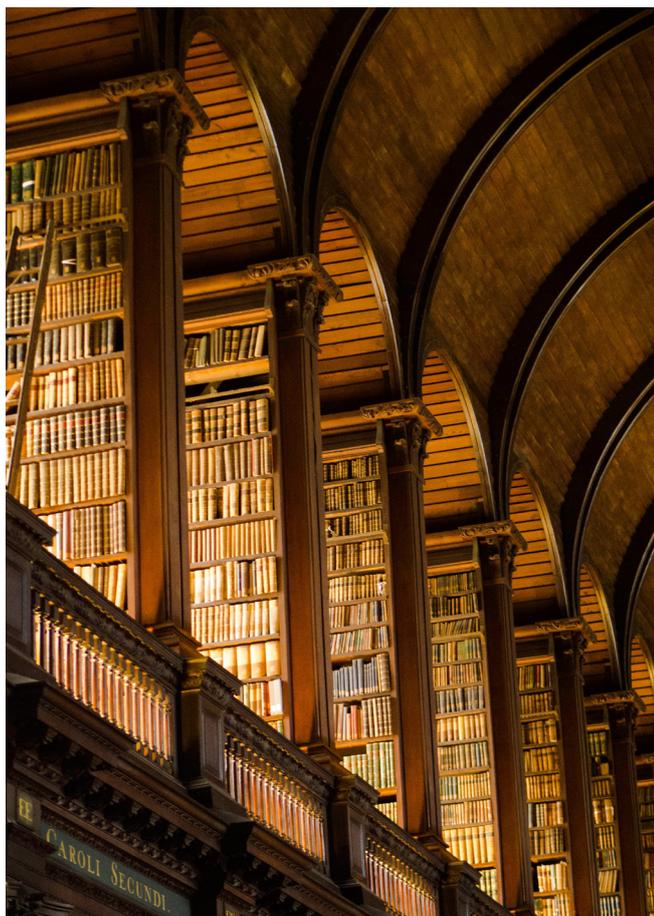
s'explique par le fondement technique de cette méthode selon laquelle, pour exercer un système, il faut un important corpus de phrases alignées. Le moteur offrira de meilleurs résultats s'il fait son apprentissage sur le type de phrases qu'il sera par la suite appelé à traduire. Par exemple, entraîné sur un ensemble de phrases mises en parallèle à partir de manuels d'automobile, il excellera à traduire des manuels d'automobile. Mais pour ce qui est des livres de recettes de cuisine, il sera moins doué. Pour pallier cet inconvénient, il ne suffit pas d'entraîner un système sur toutes sortes de textes, car la machine, qu'elle s'entraîne ou qu'elle traduise, ne sait pas faire la distinction entre un domaine et un autre. C'est pourquoi, si l'on exerce

un système sur des textes très divers, les résultats obtenus dans un des domaines traités ne seront pas aussi bons que si on ne l'avait exercé que sur un seul.

Des millions de phrases alignées

Pour la plupart des textes techniques, ce problème est sans gravité. En effet, les conventions d'écriture applicables aux livres de cuisine ne sont pas si éloignées de celles qui régissent les manuels d'automobile. Donc, si un système entraîné sur toutes sortes de textes techniques ne peut, statistiquement parlant, aboutir à d'aussi bons résultats que si on l'entraînait sur un seul sujet, la différence est souvent trop mince pour causer de réels inconvénients. On ne saurait en dire autant en matière de littérature, car à cet égard, les conventions d'écriture diffèrent amplement de celles existant en technique. De plus, elles varient énormément d'un auteur, d'une époque, d'un genre et d'une forme littéraire à l'autre.

Même si ce sont tous deux des types de poèmes, un sonnet est très différent d'un *limerick*. Et, même s'ils se classent l'un et l'autre au rayon *fantasy*, *Harry Potter* est très différent du *Seigneur des anneaux*. L'ennui, avec les systèmes de traduction automatique, est qu'en littérature en général, le style de l'auteur n'est pas forcément transposable et qu'il n'y a pas de précédent sur lequel fonder un système. S'il existe, par exemple, de nombreux exemples de contrats alignables dans les deux langues d'entraînement d'un système, où trouver le parallèle de Dante en swahili, ou de Tolstoï en vietnamien ? Le plus approchant serait une traduction humaine de Dante en swahili ou de Tolstoï en vietnamien. Or, pour



La « Salle longue » de l'ancienne bibliothèque, Trinity College, Dublin
Photo : Sarah Shaffe, Unsplash

fonctionner correctement, un corpus d'apprentissage doit comporter des millions de phrases alignées, soit l'équivalent de centaines de livres. C'est beaucoup plus que ce qu'un auteur écrira en toute une carrière ! En outre, s'il existe déjà des traductions humaines de ces œuvres, à quoi bon exercer un système informatique à les traduire de nouveau et de la même manière ?

Un corpus n'aide en rien pour le style

On serait tenté de croire qu'une perte en style, ce n'est pas la fin du monde. Le « sens » passe avant tout, n'est-ce pas ? En fait, ce n'est le cas ni en littérature ni en traduction automatique, deux domaines où forme et fonction sont étroitement liées. J'en ai récemment fait l'expérience en traduisant des poèmes extraits des *Mille et une nuits* à l'aide d'un système exercé avec le seul corpus existant de correspondances en arabe-anglais, qui compile pour l'essentiel des traductions du Coran et des données publiées par les Nations Unies. Une grande majorité des mots contenus dans les poèmes figuraient dans le corpus. Mais le style des textes était à ce point différent que le système a surtout produit des blancs.

En outre, les outils de traduction automatique fonctionnent actuellement phrase par phrase : ils en traduisent une, de façon isolée, puis l'oublient dès qu'ils passent à la suivante. Là encore, ce n'est en général pas bien gênant pour un texte technique. Mais en littérature, où une idée, une métaphore, une image ou une allusion peut se retrouver plusieurs phrases, paragraphes, voire chapitres plus loin, la machine a encore de vastes progrès à accomplir avant que ses aptitudes approchent seulement celles d'un traducteur littéraire humain.

L'informatique, assistante du traducteur littéraire

Pour ces raisons et pour bien d'autres, les programmeurs de traduction automatique ne savent se prononcer sur les futures capacités de leurs systèmes et sur le moment où elles seront exploitables. C'est pourquoi ils travaillent à l'heure actuelle sur des outils spécialement destinés aux traducteurs littéraires. Si, parmi ces derniers, certains ont déjà recours à des logiciels de TAO comme par exemple Memo-Q, nombre d'entre eux, contrairement à leurs collègues techniques, ne leur trouvent guère d'utilité. Pourtant, l'informatique permet réellement de traiter des questions spécifiques à la littérature.

« Le style de l'auteur n'est pas forcément transposable »

Par exemple, le projet de recherche **QuantiQual** étudie les traductions littéraires indirectes produites par des humains et par des machines. Une traduction indirecte est une traduction de traduction. Dans le cas où l'on ne peut effectuer une traduction directement de la langue A à la langue C, on passe par une traduction dite « relais », en langue B. Les discussions sur la légitimité même du procédé occultent depuis des lustres le fait que dans la pratique, celui-ci est monnaie courante. Quoi qu'il en soit, ce projet a le mérite de montrer que la traduction indirecte permet de diffuser des connaissances et de la littérature dans des langues traditionnellement



James Hadley est professeur-assistant « Ussher » de traduction littéraire au Trinity College de Dublin (Irlande), où il dirige un master en traduction littéraire primé. Il est également à la tête du projet de recherche **QuantiQual**, généreusement financé par le programme **COALESCE** de l'Irish Research Council.

James Hadley
Photo : collection privée

marginalisées. QuantiQual vise entre autres à déterminer en quoi les possibilités de la traduction automatique (exploitation d'un très large éventail de sources d'information, catégorisation des aspects techniques, identification de modèles) peuvent rendre service au traducteur humain. L'équipe cherche ainsi comment aider une traductrice ou un traducteur qui se trouverait, par exemple, face aux poèmes des *Mille et une nuits*. Les chercheurs développent actuellement un système qui ne crée pas lui-même une traduction de poésie mais fournit en un clin d'œil au traducteur humain des informations précieuses sur le texte source. Son travail y gagnera donc en rapidité et en efficacité. Par exemple, le logiciel peut indiquer le schéma de rimes de chacun des poèmes, repérer allitérations et assonances, calculer le nombre de mots et la longueur moyenne des vers, établir une liste de synonymes en langue cible pour chaque mot du texte. Dans ces conditions, le traducteur humain reste celui qui décide de la traduction la plus appropriée, mais le logiciel l'aide à se concentrer sur son travail créatif, au

lieu de passer du temps à se documenter dans de multiples sources. Comparée au traitement du poème, type de texte d'une complexité extrême, l'adaptation d'une solution comme celle-là à des romans, par exemple, afin d'aider le traducteur à conserver certains aspects du style, comme la longueur des phrases, l'utilisation de pronoms ou l'emploi de mots idiosyncrasiques, représente une avancée relativement modeste.

Il ne faut jamais dire : « Fontaine... » Si l'adage garde toute sa pertinence, rappelons toutefois que des traducteurs pessimistes prédisaient déjà l'avènement de leurs substituts mécaniques en 1954. En soixante-dix ans de recherches sur la traduction et sur la traduction automatique, plus nous en avons appris sur la première, plus nous avons saisi la mesure de sa complexité. Les concurrents sérieux des traducteurs littéraires humains sont encore bien loin. Mais nous assistons déjà à l'émergence d'outils propres à les aider dans leur travail.

*Traduit de l'anglais par
Marie-Christine Guyon*