

## THE CONVERSATION

# TRADUCTION AUTOMATIQUE : LA DANGEREUSE

## « SAGESSE DES FOULES »

13 octobre 2021

Guillaume Deneufbourg, Traducteur en exercice, titulaire d'une maîtrise de spécialisation en sciences du langage et traductologie, enseignant en traduction, Université de Lille.

Apôtre de l'intelligence collective, le journaliste américain James Surowiecki publie en 2004 un essai intitulé [« La sagesse des foules »](#), dans lequel il explique comment, en matière de prédiction, la capacité collective d'un groupe sera dans la plupart des cas supérieure aux aptitudes de chacun des individus qui le composent, y compris des plus fins connaisseurs.

Pour illustrer son propos, Surowiecki y relate l'histoire du statisticien britannique Francis Galton qui assiste à une foire agricole au début du XX<sup>e</sup> siècle. Un concours s'y tient, au cours duquel les participants sont invités à estimer le poids d'un bœuf. Rentrant chez lui, Galton se met à analyser les estimations des 787 paris.

Il note, à sa grande surprise, que la médiane des entrées est non seulement plus précise que les estimations individuelles de tous les bouchers et fermiers – censés avoir un œil affûté pour ce genre d'estimations – mais en outre que cette médiane correspond presque exactement au poids de l'animal, à une livre près.

Galton [publiera ses conclusions dans \*Nature\*](#), où il décrit le principe de l'intelligence collective : le plus grand nombre est bien souvent à l'origine des meilleures décisions.

### La force du collectif

Rapprochons l'histoire de Francis Galton des cours de traduction dispensés à l'université et des ateliers de perfectionnement pour professionnels en exercice : au cours d'échanges toujours féconds, chaque participant y trouve l'occasion d'exprimer ses bonnes idées et autres lumineuses trouvailles. Celles-ci sont débattues, décortiquées, critiquées par l'examen collectif. Les solutions retenues par l'ensemble du groupe sont alors compilées en une version finale, somme des meilleures inspirations de chacune des individualités. Fruit du travail d'équipe, cette traduction sera invariablement d'une qualité supérieure aux productions individuelles de chacun des participants, aussi talentueux soient-ils.

Par analogie, nous en venons à nous interroger sur la capacité de la traduction automatique – dont le modèle statistique reproduit peu ou prou le schéma de l'intelligence collective – à remplacer les traducteurs et traductrices de chair et d'os. Ainsi, à l'ère de l'intelligence artificielle et du big data, quid de mettre la force du collectif à profit pour traduire, comme si Internet était une immense salle de classe, un gigantesque projet collaboratif, une sorte de dream team composée de dizaines de millions de participants, où chaque texte déjà traduit constituerait une source d'inspiration ? Nous pouvons sans nul doute l'envisager.

### Pour le meilleur et pour le pire

Si l'idée est séduisante sur papier, je dois d'emblée décevoir les inconditionnels de l'automatisation. Si l'on peut facilement comparer les cours de traduction à l'université à un

rassemblement d'experts dont la mission serait d'apporter au plus précis l'ultrasolution à un problème donné, l'analogie avec la traduction automatique semble excessive, voire erronée.

Tout d'abord, parce que Word Wide Web compte certes de grands spécialistes, mais qui se trouvent représenter une infime minorité dans une foule d'internautes généralistes, voire néophytes, qui expriment eux aussi leur avis sur la question. L'intelligence artificielle essaiera, tant bien que mal, d'accorder la priorité aux sources identifiées comme fiables (grandes institutions, entreprises renommées), mais elle sollicitera à la vérité l'avis de toute... la planète, c'est-à-dire de tous ceux et celles qui auront déjà produit et publié du texte sur Internet. Et à cet égard, vous aurez certainement remarqué que ce que l'on trouve sur le net ne brille pas toujours par sa justesse.

Par ailleurs, pour filer la métaphore de nos constats agricoles, non seulement le monde entier vous donnera son opinion – pour le meilleur et aussi pour le pire donc – mais en outre, comme un ordinateur n'est pas en mesure de mettre du sens sur les solutions qu'il trouve, c'est un peu comme si vos pourvoyeurs d'informations se prononçaient sans même avoir identifié la bête soumise à leur examen. Ils auront certes une idée statistique du type d'animal, selon les caractéristiques détectées par la machine, mais rien de très précis, si bien qu'en plus d'estimations se rapportant aux races bovines, vous en aurez aussi potentiellement de l'ensemble de la faune terrestre, de la puce à la baleine bleue, avec toutes les incongruités qui peuvent en résulter.

À noter, enfin et surtout, que les traductions collectives humaines font toujours l'objet d'un arbitrage, soit par le professeur, soit par l'animateur de l'atelier, qui se charge de trancher, d'orienter. En d'autres termes, une entité supérieure se charge de trier les solutions provenant de la masse critique de traducteurs, comme un garde-fou veillant à la bonne marche de l'opération. En cas de recours à la traduction automatique sans intervention humaine a posteriori, cet arbitrage est inexistant.

### **Fluidité apparente et erreurs grossières**

Certes, il existe des sécurités. Les mots, tout d'abord, qui constituent tout de même une bonne indication du sens probable d'une phrase. Le contexte, ensuite, qui est désormais pris en compte par la technologie dite neuronale et réduit le champ des possibles à de grandes familles. Dans notre cas, la recherche sera circonscrite à tous les grands animaux de la ferme pour les moteurs les moins performants et à toutes les races bovines pour les technologies les plus abouties. Mais entre un veau angus maigrichon et un taureau charolais bien dodu, la marge d'erreur restera élevée !

Voilà ce qui explique qu'entre des phrases en apparence très fluides apparaîtront dans le texte : des erreurs grossières, des termes qui n'ont rien à voir avec le sujet du texte, des omissions d'éléments de sens, des [biais de genre](#), des phrases au sens complètement inversé – incapable de « comprendre » le sens de la phrase, le moteur de traduction opte pour la solution qu'il estime statistiquement la plus probable et parfois donc pour une tournure qui dit le contraire de l'original.

Ainsi, la phrase « UK car industry in brace position ahead of Brexit deadline », qui nous explique que l'industrie automobile britannique « craint le pire » (littéralement, elle se met en « position de sécurité » (brace position), se prépare à un scénario catastrophe, comme sont

invités à le faire les passagers d'un avion avant un crash), a été traduite dans [cette étude](#) par « L'industrie automobile britannique en position de force avant l'échéance du Brexit ».

Méfiance donc, car quelle que soit l'apparente fluidité de la traduction proposée, ce type d'erreur (terminologie défailante, omission, contresens) [reste omniprésent dans les productions automatiques](#).

### **Standardisation, nivellement et appauvrissement de la langue**

Un autre problème, moins connu du grand public, est celui de la standardisation. En effet, si les nouvelles traductions s'inspirent continuellement de ce qui existe déjà, cette mécanique peut, à terme, nuire à l'inventivité, à la créativité, à l'originalité – [comme le démontrent plusieurs études scientifiques](#). On le constate d'ailleurs très clairement dans un texte comportant des expressions idiomatiques colorées que la machine remplacera par des équivalents « explicatifs » – certes correct, mais plus terre-à-terre. La machine ne cherchera pas à « faire du beau », à chatouiller la poésie du verbe, mais se contentera de rendre le sens. « C'est assurément préférable à une traduction mot à mot sans queue ni tête ! » s'exclameront les nostalgiques des laborieux débuts de Google Translate. Nous ne pourrions pas leur donner tort, mais tout de même.

Dans le cas des textes d'auteur, qui, par définition, s'écartent de la norme pour acquérir un « relief littéraire » bien à eux, ce nivellement est très problématique, qu'il soit culturel, stylistique ou idéologique. À ce titre, [l'excellent texte de la traductrice Françoise Wuilmart](#), écrit une bonne décennie avant l'avènement de la technologie neuronale, en acquiert aujourd'hui des accents presque prophétiques : « Le phénomène de nivellement touche au cœur même du problème de toute traduction littéraire. Nivellement, ou encore 'normalisation', c'est-à-dire action de 'raboter' un texte ou de l'aplatir, c'est-à-dire en supprimer toutes les sortes de reliefs, y tronquer les pointes, y boucher les creux, y aplanir toutes les aspérités qui en font justement un texte littéraire. » Exactement ce que s'emploie à faire, malgré elle, la traduction automatique.

Cette standardisation pose un autre problème, plus conséquent, encore amplement méconnu. Tout indique qu'une utilisation exponentielle de la traduction automatique initie un cercle vicieux qui génère, à terme, un appauvrissement de la langue : la machine produit, comme on l'a vu, des textes toujours plus standardisés, qui servent ensuite eux-mêmes de matière première pour alimenter d'autres moteurs, qui raboteront à leur tour les textes, et ainsi de suite. La perte de richesse lexicale des textes [ayant traversé le filtre de la traduction automatique a été démontrée](#), on peut donc extrapoler cette tendance à l'ensemble des textes traduits de cette façon.

Or, s'exposer à une standardisation toujours plus grande du langage, c'est réduire notre capacité d'expression, et donc notre pensée. Orwell avait déjà saisi tout l'enjeu de cette question. Pour contrôler la pensée, diminuer le langage peut s'avérer très efficace. Ainsi, dans 1984, Big Brother interdit l'usage de mots qui pourraient véhiculer des pensées interdites. « Nous détruisons chaque jour des mots, des vingtaines de mots, des centaines de mots, nous taillons le langage jusqu'à l'os » s'y réjouit un spécialiste de novlangue. [« À la fin, nous rendrons littéralement impossible la pensée, car il n'y aura plus de mots pour l'exprimer. »](#) Sans vouloir paraître défaitiste, je ne suis pas sûr que la peur de voir se réaliser ce scénario, aussi glaçant soit-il, puisse enrayer le processus. Il est en tout cas tout sauf irréaliste. Même si son

développement comporte quelques raccourcis et faiblesses, [ce n'est pas Christophe Clavé qui me contredira](#).

### **Problèmes de confidentialité**

Comme nous le rappelle également le traductologue Rudy Loock dans [son article](#) sur la plus-value de la traduction humaine face à la machine, l'utilisation d'outils de traduction automatique en ligne n'est pas non plus sans risques en termes de confidentialité. On se souviendra notamment du cas du géant pétrolier norvégien Statoil/Equinor, dont des données ultraconfidentielles ont « fuité » sur l'Internet à la suite du recours [au service de traduction en ligne translate.com](#). Le site spécialisé Slator a ainsi retrouvé, après quelques rapides recherches Google, une quantité étonnante d'informations en libre accès. Dans tous les cas, des noms complets, des courriels, des numéros de téléphone et d'autres données très sensibles furent révélés.

Sans vouloir nous montrer alarmistes, rappelons le flou entourant les conditions d'utilisation des [données que vous cédez à Google](#) ou les secrets de fabrication du moteur DeepL, [impénétrable boîte noire](#) dont on ne sait quasiment rien.

### **Le risque d'une mauvaise publicité**

En dépit de tous ces dangers, on constate aussi et surtout une grande incompréhension dans le chef de certains acheteurs de services de traduction, qui n'ont pas toujours conscience des risques auxquels ils s'exposent et dont une part grandissante voit dans la traduction automatique la solution miracle à tous leurs problèmes. À titre d'exemple, la Direction Générale de la Traduction (DGT) de la Commission européenne met son moteur de traduction automatique etranslation gratuitement à disposition afin de [« briser la barrière de la langue »](#).

Empressons-nous d'ajouter : ne voyez pas dans la critique qui suit la grogne épidermique d'un professionnel aux abois anxieux pour son avenir. L'évolution technologique est inéluctable et il est indiqué de cultiver l'ouverture (adossée à une posture critique réfléchie). L'inquiétude procède du fait que les informaticiens de la Commission ont mis cet outil en ligne à la disposition du plus grand nombre en oubliant un élément fondamental : éduquer le public.

Il ressort ainsi d'une série d'études que la plupart des consommateurs ne s'appuient que sur des impressions très subjectives, rarement sur des faits. [Une autre étude](#) a analysé le traitement médiatique de la traduction automatique, mettant en lumière que la majorité des articles de presse présentaient la technologie sous un jour ([parfois très](#)) positif, le spectre de l'homme remplacé par la machine ayant de tout temps été un [sujet porteur](#). Couplée à la facilité d'utilisation et à la gratuité, cette subjectivité ambiante fait barrage à une approche raisonnée de ces outils par le grand public.

Autre preuve, [cette campagne](#) lancée par un organe public belge actif dans le développement économique, nous expliquant qu'« aujourd'hui, la traduction automatique constitue une solution rapide, bon marché et facile d'utilisation. » Et l'organisme d'annoncer dans un mailing envoyé à des milliers de petites et moyennes entreprises que « la Commission européenne compte [les] soutenir en leur proposant des solutions automatisées pour la traduction de sites Web. Ces solutions seront fondées sur eTranslation, le service de traduction automatique de la Commission, actuellement utilisé par les institutions européennes et des milliers d'administrations publiques et de PME dans l'ensemble de l'UE ».

La tournure commerciale du texte de présentation n'aura pas manqué de piquer les yeux des professionnels de la traduction. Pas un mot sur les dangers de la technologie, pas l'ombre d'une explication sur le pourquoi du comment. Juste un message alléchant.

Or, lorsqu'elle dépasse le cadre de l'usage privé en tant qu'outil d'intercompréhension, la traduction automatique peut rapidement se transformer en une redoutable arme de destruction massive, capable de ruiner une réputation en moins de temps qu'il n'en faut pour partager sur les réseaux sociaux une perle produite par Google Translate. Mon confrère anglophone Ben Karl en reprend quelques exemples sur [son site Web](#), dont le [cas de l'Office du tourisme du Mexique](#), sur le site duquel le nom de la ville d'Aculco avait été (automatiquement) traduit en anglais par « J'accuse ». Autre perle magnifique : le nom du président de la République populaire de Chine, [traduit du birman à l'anglais](#) par le très élégant « Monsieur Trou du cul ».

### **Un recours inévitable à l'expertise humaine**

Plus personne dans le secteur ne nie aujourd'hui l'évolution technologique dont la traduction fait l'objet, comme tant d'autres métiers. Il est évident que la traduction-machine est de plus en plus utilisée comme outil d'aide à la traduction. Il est clair également qu'un moteur statistique peut produire des résultats sans cesse plus exploitables.

Mais encore trop d'utilisateurs oublient que ce contenu traduit automatiquement reste (potentiellement) criblé d'erreurs, de tous types et de toutes natures, que des énormités peuvent se cacher partout, entre des phrases en apparence fluides et cohérentes. Aussi, seule l'expertise d'une ou d'un professionnel de la traduction permettra d'évaluer la qualité de cette matière première. Seul un professionnel de chair et d'os pourra faire le choix – ou non – d'y recourir, à l'instar d'une photographe qui sélectionnera l'appareil photo le plus approprié aux conditions ou qu'un expert-comptable qui décidera du moyen de saisie des écritures le plus adapté à son mode de fonctionnement.

Comme tous les métiers, la traduction n'échappe pas à l'automatisation. Je ne cesse de clamer qu'il faut embrasser cette évolution, ne serait-ce que parce qu'elle est inéluctable. On peut même se réjouir de cette mutation, qui peut parfois servir la qualité, en ce sens où elle permet aux praticiens de faire véritablement valoir leur savoir-faire, d'éviter les tâches répétitives et de se concentrer sur celles à plus forte valeur ajoutée.

Gardons quoi qu'il en soit à l'esprit qu'en matière d'automatisation, la prudence reste plus que jamais de mise. Il convient à ce titre d'éviter tout recours « aveugle » à la traduction-machine. Les vrais professionnels sauront choisir avec vous la meilleure méthode de travail, en fonction de vos priorités et du célèbre triptyque qualité – délai – budget. Conseillers linguistiques et culturels avertis, ils seront les seuls garants d'une communication multilingue irréprochable. Comme aurait sans doute pu l'avancer en 1906 le garçon boucher victorieux du concours de la foire agricole de Plymouth, l'expertise humaine restera votre seul et unique moyen de faire, à tous les coups, un effet... bœuf.