Traduire

Revue française de la traduction

230 | 2014

À la croisée du texte et de l'image

Big data et cloud computing : une révolution dans l'univers de la traduction ?

Anne-Charlotte Perrigaud

Édition électronique

URL : <http://journals.openedition.org/traduire/637>

DOI : 10.4000/traduire.637

ISSN : 2272-9992

Éditeur

Société française des traducteurs

Édition imprimée

Date de publication : 15 juin 2014

Pagination : 101-107

ISSN : 0395-773X

Référence électronique

Anne-Charlotte Perrigaud, « *Big data et cloud computing : une révolution dans l'univers de la traduction ?* », *Traduire* [En ligne], 230 | 2014, mis en ligne le 15 juin 2016, consulté le 02 mai 2019.

URL : <http://journals.openedition.org/traduire/637> ; DOI : 10.4000/traduire.637

Big data et cloud computing : **une révolution dans l'univers de la traduction ?**

Anne-Charlotte Perrigaud

Depuis la pierre de Rosette, et même avant, on traduit. Avec l'apparition de la machine à écrire, puis de l'informatique, on s'est mis à traduire de plus en plus, et de plus en plus vite. La technologie n'a investi le monde de la traduction que tardivement, mais elle s'est bien rattrapée ensuite. L'arrivée de la TAO et d'autres outils ont au départ plutôt facilité le travail du traducteur qui a, à cette époque, connu une sorte d'âge d'or. Ce processus s'est encore accéléré avec l'avènement d'internet et des nouvelles technologies, bouleversant la manière dont nous travaillons et concevons l'opération traduisante. Les systèmes à reconnaissance vocale ont par exemple permis d'augmenter la vitesse de saisie du texte sans pour autant le dactylographier (diminuant du même coup le risque de troubles musculosquelettiques), de faire plus de recherches, plus vite et avec moins d'efforts (ce qui était mieux pour les poignets et le dos qu'au temps des encyclopédies et autres ouvrages papier). Il est devenu possible d'accéder à un plus grand nombre de dictionnaires sur CD sans problème de stockage dans le bureau, de faire des recherches à l'intérieur d'un texte très rapidement afin de déceler la fréquence ou non d'une expression, de repérer les répétitions et de faire toutes sortes de traitements sur les textes. Les nouvelles technologies nous ont également permis de communiquer beaucoup plus rapidement avec les confrères comme avec les donneurs d'ouvrage. Toutes ces innovations technologiques ont été plutôt bien accueillies et le consensus était que le « progrès » permettait une amélioration de l'ergonomie et de la qualité de vie au travail pour le praticien.

Avec l'arrivée de la TAO les réactions furent plus mitigées. Passée la réticence des premières heures, les systèmes de TAO ont été progressivement adoptés par les professionnels. Jusqu'à là, la traduction automatique n'était qu'une utopie d'informaticien. L'homme et la machine cohabitaient dans un équilibre qui satisfaisait de nombreux traducteurs et l'automatisation représentait alors une aide à la pratique plus qu'une menace. Les spécialistes du traitement automatique des langues faisaient bien des expérimentations et des tests, mais rien de concluant n'émergeait. Jusqu'à l'apparition puis la généralisation de l'exploitation par les entreprises de volumes de données massifs, tellement massifs qu'ils sont devenus difficiles à administrer *via* des outils traditionnels de gestion de bases de données ou de gestion de l'information : c'est ce qu'on

appelle le *big data*. Si elle a présenté de nouveaux défis en matière de stockage, de recherche et de partage d'informations par exemple, cette nouvelle donne en matière de données a favorisé dans le même temps un développement sans précédent des méthodes de traduction automatique dites « statistiques » ; sous cette influence, des millions d'utilisateurs se sont mis à cliquer sur le bouton « traduire » de leur moteur de recherche ; et soudain les lignes ont bougé. Bien sûr, les traductions fournies n'étaient pas de qualité publiable. Les erreurs étaient souvent risibles et le résultat peu fiable. Mais quelque chose avait changé : le grand public aimait le fait de contrôler l'opération et d'obtenir un résultat en temps réel, gratuitement qui plus est, avec les répercussions que l'on sait sur la vitesse de production demandée aux traducteurs face à cette « concurrence » rapide et sur le volume de mots à traduire sans cesse croissant. Les entreprises ont, elles aussi, rapidement adopté la traduction automatique. Très vite, il leur a semblé qu'un retour en arrière serait improbable car elles avaient adopté la traduction automatique comme un complément essentiel aux efforts de traduction humaine : de leur point de vue, elle permet désormais de résoudre, même partiellement et avec les failles que l'on sait, les problématiques du volume élevé des traductions et de production de contenu multilingue en temps réel, à toute heure du jour ou de la nuit.

Comment en est-on arrivé à cette situation ? C'est en fait par le truchement du rapport que le *big data* entretient avec les systèmes à mémoire de traduction que la traduction automatique a connu son essor. Le *big data*, me direz-vous ? Mais quel rapport avec la traduction ? Je m'explique : pour schématiser, et sans prendre en compte les systèmes hybrides, on peut dire qu'il existe deux types de systèmes de traduction automatique : les systèmes basés sur des règles et les systèmes statistiques. Dans le cas des systèmes basés sur des règles, les traductions obtenues s'améliorent au fur et à mesure qu'un plus grand nombre de personnes rédige de meilleures règles et créent de meilleurs dictionnaires. En revanche, dans les systèmes statistiques, ce qu'il convient de faire c'est d'accumuler des données, plus précisément des textes alignés traduits précédemment par des traducteurs humains hautement qualifiés, puis de les diviser en « morceaux » plus petits, le but étant de les réutiliser pour créer de nouvelles phrases en essayant d'estimer les probabilités de traductions différentes d'un texte source, et en prédisant la plus probable. Mais comment obtient-on toutes ces propositions de traduction et sur quoi se base-t-on pour calculer toutes ces probabilités me direz-vous ? En analysant des millions de mots de textes traduits, générés par des traducteurs professionnels durant les dernières décennies : c'est le *big language* dont nous parlent de plus en plus les industriels.

Maintenant on comprend aisément pourquoi le modèle fondé sur les données obtient l'avantage par rapport au modèle fondé sur des règles : il paraît en effet plus facile d'augmenter la quantité de données fournies aux machines que de créer de nouvelles règles linguistiques. D'un côté, un processus qui peut être presque entièrement automatisé, de l'autre une rédaction longue et fastidieuse de règles linguistiques par des spécialistes du traitement automatique du langage avec la masse salariale correspondante. Un autre problème que l'on rencontre

avec le système basé sur les règles, au fur et à mesure que le nombre de règles linguistiques augmente, c'est celui de l'interférence des règles les unes avec les autres, qui plus est de façon totalement imprévisible.

Issus de ces deux méthodes, des systèmes hybrides ont également vu le jour. Ces systèmes ont été intégrés aux outils de TAO utilisés par les professionnels et sont déjà une réalité aujourd'hui. Ces nouveaux outils intégrés sont censés aider le traducteur dans sa tâche, mais ne risquent-ils pas aussi de révolutionner de façon incroyable le travail quotidien du traducteur, voire de le remplacer, interdisant tout recours à la créativité et réduisant le traducteur à un simple opérateur effectuant des contrôles sur un texte créé par une machine ? Les pratiques de post-édition, entre autres, sont-elles une menace pour notre profession ? La formation des traducteurs devra-t-elle être adaptée au regard des innovations technologiques ? Une dimension éthique devrait-elle être intégrée à la formation afin que les praticiens soient mieux armés pour faire leurs choix au regard des technologies ?

Toutes ces questions sont légitimes et font actuellement l'objet de débats animés dans la profession, tant sur les réseaux sociaux que lors des conférences et autres journées professionnelles. En revanche, d'autres questions, moins souvent énoncées et pourtant cruciales, se posent et doivent être prises en compte, en particulier dans le contexte de l'avènement des solutions d'aide à la traduction en ligne hébergées dans le *cloud*.

Le problème le plus visible est celui de la contamination des données par des données parallèles non fiables disponibles sur internet, qui aboutit à une déperdition de qualité dans la traduction produite statistiquement puisque les données n'étaient pas qualifiées. C'est ce qui se passe lorsque Google utilise des sites web traduits par un outil de traduction automatique de médiocre qualité pour remplir ses bases de données de textes parallèles et réutilise ensuite ces données pour produire une traduction statistique, que les internautes peuvent solliciter via Google Translate par exemple, puis publier sur leurs sites internet. En conséquence, le nombre de mauvaises traductions augmente encore ; celles-ci peuvent à nouveau être indexées, et la qualité générale des textes produits automatiquement décline inexorablement, mettant en évidence les limites de la solution statistique de traduction automatique lorsqu'elle est utilisée sans qualification préalable des sources.

Quel rapport avec le *cloud*, me direz-vous ? Ce phénomène se produit en fait également dans le monde des professionnels de la traduction, par le biais de l'utilisation de mémoires de traduction partagées dans le *cloud*, avec tous les risques avérés de « contamination » à grande échelle par des traductions médiocres (humaines ou non) qu'elle comporte. Mais il est un fait que certains traducteurs ignorent : en utilisant les systèmes basés dans le *cloud* pour traduire, le traducteur cède, le plus souvent sans même en avoir conscience, les droits sur les mémoires de traduction. Il fournit des données qui sont ensuite collectées, disséminées, cachées, réutilisées, sans que le traducteur sache quels traitements ces données subissent exactement, avec

tous les problèmes juridiques que cela pose en termes de statut des données de traduction : droit d'auteur, respect de la propriété intellectuelle, confidentialité des données du client, risques de fuites d'informations sensibles, pour n'en citer que quelques-uns.

Par choix ou par naïveté, voire manque d'information et de formation sur ces sujets, peu de traducteurs s'interrogent en effet sur l'ambivalence de ces outils « gratuits » se présentant comme étant au service des traducteurs, tout en ne leur permettant pas de gérer eux-mêmes leurs propres traces d'usage (effacement, prospection commerciale, désinscription entre autres). La gratuité n'est qu'apparente car en fait les utilisateurs « paient » avec leurs données (ou pire, avec celles de leurs clients) car l'utilisation de ces outils est généralement fondée sur un consentement implicite, rarement perçu par les traducteurs internautes, à céder leurs données de traduction en contrepartie des services rendus ; les utilisateurs « acceptent » ce faisant la perte de propriété sur les écrits, les documents, mémoires de traduction, etc. qu'ils déposent.

Ces outils basés dans le *cloud* présentent tous une même caractéristique : leur fonctionnement normal intègre des fonctionnalités qui ne sont pas au bénéfice des traducteurs qui les utilisent ni au bénéfice des clients de ceux-ci ; il proscriit même parfois, au contraire, des services indispensables à leur bonne utilisation par les utilisateurs (accès aux traces de fonctionnement par exemple).

Un autre point souvent oublié à propos de l'utilisation du *cloud* : quelles sont les conséquences des négligences ou des défauts de sécurité menant à la divulgation accidentelle de données confidentielles (données des clients ou données du cabinet de traduction), ou à d'autres dysfonctionnements préjudiciables pour les usagers ? Dans la plupart des cas, les personnes lésées ne parviennent pas à faire reconnaître leurs droits (que, de surcroît, elles connaissent souvent mal) face aux prestataires de solutions informatiques. Dans tous les cas, la responsabilité du traducteur est engagée, d'où l'intérêt de connaître et respecter la législation en vigueur, et de se doter d'une assurance de responsabilité civile professionnelle (RCP) convenable car, en toute logique, dans un tel cas, les clients finaux en cas de litige devraient vraisemblablement se retourner contre le traducteur. Face à ces nouveaux enjeux liés au statut des données, il est probable que les règles juridiques devront être adaptées à l'avenir afin de protéger les droits des traducteurs, mais aussi ceux de leurs clients.

Nous l'avons vu, de nombreuses failles existent, et les dérives sont toujours possibles. En dépit de cela, le travail dans le *cloud*, s'il est maîtrisé, peut être formidable en ce qu'il offre de nouvelles perspectives en matière de travail collaboratif ; du *crowdsourcing*⁽¹⁾ à la gestion termi-

(1) Le *crowdsourcing* est une méthode de sous-traitance qui consiste à faire appel à des contributeurs bénévoles en très grand nombre via internet afin de leur faire réaliser en un temps record des tâches qui étaient traditionnellement effectuées par un salarié ou par un entrepreneur sous-traitant. Chaque contributeur réalise une petite partie de la tâche qui est de ce fait collaborative.

nologique en temps réel, de la communication rapide à l'assurance qualité, les possibilités sont très grandes et mériteraient d'être explorées en détail. En tout état de cause, ce sont comme toujours la déontologie et l'éthique professionnelle qui offrent la meilleure solution pour notre profession.

L'objet de cet article n'est pas, vous l'aurez compris, de porter un jugement de valeur sur les innovations technologiques mais plutôt de décrire les évolutions dont j'ai été le témoin au cours de ma pratique ces dix dernières années. Force est de constater pourtant que, pour le traducteur pragmatique, la nécessité se fait de plus en plus sentir de développer une nouvelle vision éthique de la traduction afin d'anticiper et d'accompagner les phénomènes décrits ci-dessus et, pourquoi pas, de les influencer de manière positive. Je suis convaincue que cela passera par une meilleure information de tous les acteurs du marché, y compris et surtout les clients finaux. Un premier pas pourrait être fait en travaillant à l'intégration de ces considérations dans les programmes de formation universitaire des futurs professionnels de la traduction, mais aussi en sensibilisant à ces questions les étudiants en écoles de commerce, de gestion et autres car ils sont les donneurs d'ouvrage de demain. De nombreuses possibilités s'offrent également dans le domaine de la formation continue tout au long de la vie professionnelle pour les praticiens déjà installés, afin que tous puissent prendre la mesure des enjeux et des défis à relever... mais ce secteur est encore jeune et de nombreuses voies restent à explorer.

Nous n'en sommes pour l'instant qu'au stade du défrichage, et c'est à chacune et à chacun qu'il appartient de prendre part à la révolution technologique que nous sommes en train de vivre et de l'accompagner d'un changement des mentalités. En effet, utilisé avec discernement et dans le respect de l'éthique⁽²⁾, le *cloud* pourrait changer la donne en permettant, pourquoi pas, l'émergence d'un nouveau modèle économique basé sur l'automatisation du flux de travail et

(2) Le minimum pourrait être :

1. S'assurer que l'on connaît et que l'on respecte la législation sur le droit d'auteur, le *copyright* et les droits voisins dans le contexte de production du texte source ainsi que dans le contexte de production mais aussi d'utilisation du texte cible, car ceux-ci peuvent être différents d'un pays à un autre.
2. Lire et comprendre les conditions d'utilisation des services que l'on se propose d'utiliser. Vérifier si celles-ci respectent les droits du donneur d'ouvrage, les droits du traducteur, la déontologie de la profession. Vérifier si on ne contrevient à aucune loi en utilisant ces services. Vérifier quel est le système de droit qui s'applique aux services qu'on souhaite utiliser.
3. Vérifier s'il sera facile ou non de récupérer ses données et celles de ses clients en cas de changement de prestataire. Vérifier aussi qu'on ne cède pas ses droits (ou ceux de ses clients, que par définition on ne possède pas) automatiquement en acceptant d'utiliser un service. S'informer sur ce que le prestataire fait des données que vous lui confiez (traitements, stockage, revente sous forme de mémoires de traductions, intégration à des mémoires de traduction partagées en libre accès sur la toile, etc.).
4. Communiquer et informer chaque client sur les risques et les enjeux des services dans le *cloud* et lui demander l'autorisation si on a l'intention d'utiliser des données qui lui appartiennent dans un *cloud* public (problèmes de confidentialité entre autres). Dans le cas d'un *cloud* privé les risques semblent moindres.

sur la gestion automatisée des projets de traduction en temps réel. Le concept des systèmes de gestion de projets de traduction basés sur le *cloud* pourrait à terme rendre obsolète le modèle actuel qui oppose traducteurs indépendants et agences de traductions ; il pourrait même faire contrepoids à la concentration croissante des agences de traduction, avec toutes les conséquences qu'elle implique souvent (disparition progressive des agences de traduction « à taille humaine » et des agences de traduction spécialisées, délocalisation des services de traduction par des poids-lourds imposant leurs conditions de façon unilatérale à une masse de petits prestataires indépendants, clauses abusives dans les contrats d'agences, tarifs bas, délais raccourcis, etc.). Les clients pourraient se voir proposer une véritable alternative : un nouveau modèle, basé sur la qualité, la confiance, le respect mutuel et, bien sûr, l'expertise de professionnels des langues exigeants ; tout cela facilité par des systèmes qui rendent redondants les gestionnaires de projets et automatisent toutes les tâches administratives et comptables, de la demande de devis au paiement en ligne par les clients, sans oublier le paiement automatique par virement bancaire de tous les acteurs intervenant sur un projet, permettant ainsi aux traducteurs et aux clients de travailler main dans la main et de se concentrer sur l'essentiel : les tâches à valeur ajoutée qui font toute la valeur de notre travail.

Associé à une démarche active d'information auprès des clients et du grand public sur les enjeux de la traduction, sur les droits et les devoirs de chacun en termes de confidentialité, de respect de la propriété intellectuelle et de qualité entre autres, le *cloud* pourrait alors devenir un pont entre les clients (directs) et les traducteurs, rendant du même coup redondants les intermédiaires sans valeur ajoutée et autres « boîtes aux lettres », au profit d'acteurs orientés vers la qualité et travaillant dans le respect de l'éthique et de l'équité envers chacune des parties impliquées dans le marché de la traduction, qu'il s'agisse des clients, des traducteurs indépendants ou des réseaux de traducteurs. Cet avenir est possible, à nous de le construire.

contract@e-frenchtranslations.com

Bibliographie indicative

Baker M. (ed.) 1998. *Routledge Encyclopaedia of Translation Studies*, Routledge, London — New York.

Bowker, L. 2002. *Computer-aided translation technology: a practical introduction*, Ottawa, University of Ottawa Press.

Dunne, K.-J. (ed.) 2006. *Perspectives on localization*, Amsterdam (Pays-Bas), John Benjamins.

Dunne, K.-J. and Dunne, E. S. (eds.) 2011. *Translation and localization project management: the art of the possible*, Amsterdam (Pays-Bas), John Benjamins.

- Esselink, B. 2000. *A Practical Guide to Localization*, Amsterdam (Pays-Bas), John Benjamins.
- Froeliger N. 2005. *Placer le traducteur au cœur de la traductologie*, Méta, volume 50, n° 4, décembre, Montréal (Canada), Les Presses de l'Université de Montréal, <http://id.erudit.org/ide-ridut/019838ar>, consulté le 10 décembre 2013.
- Gouadec D. 2009. *Guide des métiers de la traduction-localisation et de la communication multilingue et multimédia*, Paris, La Maison du Dictionnaire.
- Ladmiraal J.-R. 1994. *Traduire : théorèmes pour la traduction*, Paris, Gallimard.
- O'Hagan, M. and Ashworth, D. 2002. *Translation-mediated communication in a digital world: facing the challenges of globalization and localization*, Bristol (Royaume-Uni), Multilingual Matters.
- Perrigaud, A-C. 2013. *About the Need for IT Security Awareness in the Language Industry*, Wiesbaden (Allemagne), Proceedings tekomp-Jahrestagung 2013, Tekom.
- Pym, A. 2004. *The moving text: localization, translation, and distribution*, Amsterdam (Pays-Bas), John Benjamins.
- Quah, C. K. 2006. *Translation and technology*, London, Palgrave MacMillan.
- Somers, H. (ed.) 2003. *Computers and translation: a translator's guide*, Amsterdam (Pays-Bas), Philadelphia (États-Unis), J. Benjamins Pub. Co.
- Toudic Daniel, de Brébisson Guillaume (2010). *Poste du travail du traducteur et responsabilité : une question de perspective*, <http://ilcea.revues.org/1043>, consulté le 10 décembre 2013.
- Toury Gideon (1980). *In Search of a Theory of Translation*, Tel Aviv (Israël), Porter Institute for Poetics and Semiotics.

Sites internet

- Tradumatica : revue en ligne n° 1 à n° 8,
<http://www.fti.uab.cat/tradumatica/revista/hemeroteca.htm>, consulté le 10 décembre 2013.
- TAUS : divers articles, <https://www.taus>, consulté le 10 décembre 2013.

Anne-Charlotte Perrigaud est traductrice diplômée de l'ULIP, anciennement Institut Britannique de Paris. Elle exerce en tant que traductrice libérale depuis 2002. Ses domaines de spécialité sont l'informatique et le droit des NTIC. Depuis 2013, elle est également chargée d'enseignement en traduction à l'Institut Catholique de Rennes.