Les deux voies de la traduction automatique ?

François Yvon

Les deux voies de la traduction automatique ?

François Yvon, LIMSI, CNRS, Université Paris-Saclay

La traduction automatique progresse, ses usages et ses utilisateurs se multiplient

Chacun peut désormais s'en convaincre en quelques clics : les logiciels de traduction automatique (TA) sont devenus suffisamment performants pour rendre de réels services, pour le plus grand bénéfice des traducteurs professionnels comme pour celui des utilisateurs profanes.

Les utilisateurs et les usages des outils de la TA sont multiples[1] : en premier lieu pour consulter des documents rédigés dans une langue inconnue, mais également pour diffuser des informations dans d'autres langues, pour communiquer avec des locuteurs parlant une autre langue, pour accélérer la production de traductions humaines, pour assister la rédaction en langue étrangère, pour apprendre de nouvelles langues, etc. Il existe même des usages qui n'impliquent aucun lecteur ni scripteur humain, la traduction permettant par exemple d'indexer ou de classer de manière homogène des documents rédigés en plusieurs langues, à des fins de recherche d'information ou de détection de contenus frauduleux ou haineux. Les documents soumis aux systèmes de TA ne sont pas moins variés, allant de mots ou de termes isolés à des écrits techniques ou littéraires, en passant par toutes les formes de contenus disponibles sur la toile : sites web, billets sur des forums, tweets, etc[2]. Il n'est donc pas surprenant que les services de traduction en ligne soient utilisés de manière toujours plus massive, servant chaque jour des milliards de demandes de traduction, concernant plusieurs milliers de couples de langues.

Cette amélioration des outils de traduction automatique est le résultat d'innovations technologiques relativement récentes[3] dans le domaine du traitement des langues et résultent de la conjonction de plusieurs facteurs. En premier lieu, le développement de modèles

---

[1] On s'appuie ici sur les principaux contextes de traduction (assimilation, dissémination, communication) identifiés par Hovy, King et Popescu-Bellis (2003) ; l'organisation des services offerts par *Google translate*, qui n'existent pas tous pour tous les couples de langues, est une autre manière d'appréhender la variété de ces usages (voir en ligne : https://translate.google.com/intl/en/about/languages/, consulté le 21 juillet 2019).

[2] Y compris des textes qui sont incrustés dans des images, comme des enseignes ou des panneaux de signalisation.

[3] On peut dater, avec Léon (2015), ce deuxième moment de l'empiricisme en traitement automatique des langues, qui touche d'emblée la traduction automatique, du début des années 1990.

computationnels, hier fondés sur des formalisations probabilistes, s'appuyant aujourd'hui sur des architectures de calcul « neuronales[4] », capables d'apprendre à mettre en correspondance une phrase écrite ou prononcée dans une langue « source » avec sa traduction dans une langue « cible ». Cet apprentissage, qui constitue l'essence des technologies de « l'Intelligence Artificielle »[5], nécessite de disposer d'exemples de telles correspondances, rassemblés dans un *corpus parallèle*. En second lieu, la possibilité, pour une large communauté de chercheurs et de développeurs de systèmes informatiques, d'accéder à de tels corpus, diffusés par exemple par des institutions multilingues telles que le Parlement européen ou l'ONU. En troisième lieu, l'augmentation continue des capacités de calcul et de stockage, qui permet le déploiement à très grande échelle de ces architectures de calcul et la production quasi instantanée de traductions. Enfin, il ne faut pas sous-estimer le rôle qu'ont joué les mesures automatiques de la qualité de traduction, qui servent à guider les optimisations internes aux systèmes de TA, mais également d'orienter l'effort de développement vers des architectures de plus en plus efficaces. Et ce, en dépit de la très faible ressemblance entre ces mesures automatiques, souvent très grossières[6] et les analyses multi-factorielles qui doivent être réalisées pour obtenir une évaluation approfondie de la TA (Hartley et Popescu-Belis, 2004).

Le besoin de données

Un premier enseignement est que la TA contemporaine se nourrit de grandes masses de données, ce qui permet d'opérer une première distinction entre les différents scenarios d'usage de la TA : ceux pour lesquels ces données existent et ceux pour lesquels elles n'existent tout simplement pas.

La première situation présuppose l'activité préalable de traducteurs professionnels, opérant pour le compte d'une entité (une institution ou un industriel) solvable et devant produire des textes techniques ou littéraires à des fins de publication et donc soumis à de fortes exigences de qualité (par exemple pour des raisons réglementaires ou contractuelles). Ce contexte, exemplifié par l'activité des traducteurs de la direction générale de la traduction de l'Union Européenne, est idéal pour l'automatisation : les traductions à produire dans le futur ressemblent aux traductions passées, elles sont disponibles en grande quantité et l'apprentissage

[4] Les guillemets s'imposent ici : si initialement les spécialistes des réseaux de neurones pouvaient revendiquer de s'inspirer du fonctionnement du cerveau humain, ces références aux neuro-sciences ont presque entièrement disparu des travaux contemporains.

[5] Et qui n'est pas propre à la traduction automatique : les mêmes algorithmes sont également utilisés pour analyser des images, transcrire la parole, guider des voitures autonomes, etc.

[6] La plus usitée est BLEU (*Bilingual Evaluation Understudy)* introduite par Papineni et ses co-auteurs (2001).

réalisé par la TA produira souvent de bonnes traductions. Avant d'être publiées, elles devront toutefois être révisées (*post-éditées*) pour satisfaire les niveaux de qualité requis – suscitant des recherches visant à améliorer les environnements de traduction assistée par ordinateur (TAO). Ces recherches s'intéresseront, par exemple, à rendre l'outil à même de s'adapter à son utilisateur, d'expliquer ses décisions de traduction, d'intégrer des ressources terminologiques, etc.

Cette situation est toutefois **exceptionnelle et ne vaut que pour un nombre très restreint de types de documents et de couples de langues** : dans la majorité des contextes, les exemples de traductions n'existent pas[7] (ou dans des quantités dérisoires), soit qu'il n'existe aucune demande pour des traductions humaines, soit que cette demande ne soit pas solvable (l'immense majorité des requêtes sur les plateformes en ligne) et s'accommode de traductions approximatives ou médiocres, soit, enfin, que la notion même de traduction ne soit pas très bien définie, comme c'est le cas, par exemple, lorsque les textes à traduire sont des commentaires postés sans relecture sur des réseaux sociaux ou des petites annonces sur les sites de vente en ligne. On peut ainsi se demander comment apprécier ce résultat, calculé par DeepL[8], d'une critique (en français) de restaurant trouvée sur le site TripAdvisor[9] – même si la traduction vers l'anglais atteint ici indiscutablement son but, qui est de déconseiller le restaurant :

Service super désagréable, imblairable et très très lent. L'endroit se veut cool mais est très ringard. C'est très cher pour des cocktails dégueulasses et des apéros immangeables. A éviter ! / Super unpleasant, unbearable and very, very, very slow service. The place is supposed to be cool but it's very old-fashioned. It's very expensive for disgusting cocktails and inedible aperitifs. To avoid!

Apprendre à traduire sans exemple de traduction

Pour des raisons qui tiennent principalement aux besoins opérationnels des principaux acteurs du domaine, au premier rang desquels les opérateurs des grands services de l'internet, c'est

---

[7] La situation est bien sûr très évolutive et la demande de multilinguisme sur le web fait émerger de nouveaux contextes : ainsi la traduction collaborative de sous-titres de vidéos (tutoriels, cours et conférences) par des traducteurs amateurs produit de nouveaux corpus parallèles, grâce auxquels la traduction automatique de ces contenus, pour lesquels il n'existait jusqu'alors aucune bio-traduction, s'améliore rapidement. Voir par exemple l'organisation mise en place par les conférences TedTalks (https://www.ted.com/participate/translate, visité en ligne le 20 juillet 2019).

[8] https://www.deepl.com/translator

[9] http://tripadvisor.fr

cette seconde situation d'une traduction sans exemple qui mobilise aujourd'hui l'essentiel de l'effort de recherche, qu'il s'agisse d'offrir des services de traduction pour des langues ou paires de langues non encore outillées ; ou bien qu'il s'agisse de pouvoir traduire efficacement de nouveaux types de contenus : conversations dans des forums, négociations en ligne, légendes d'images et de vidéos, paroles de chansons, retranscriptions automatique de contenus audio, dialogues, etc.

D'un point de vue technique, ce contexte est particulièrement excitant et propice à de nombreuses innovations qui touchent à la fois aux architectures des systèmes de traduction et à leur utilisation des ressources disponibles. On mentionnera pour l'exemple les travaux sur les systèmes de traduction multilingues intégrant au sein d'un unique logiciel la capacité de mettre en correspondance de multiples paires de langues. Ainsi, en réalisant un apprentissage utilisant simultanément des données parallèles pour les couples français vers anglais et espagnol vers l'allemand, on peut espérer traduire du français vers l'allemand sans avoir observé le moindre exemple de cette paire[10] de langues. Une seconde illustration, tout aussi frappante, concerne les efforts pour produire des traductions en n'utilisant aucune donnée parallèle. Partant d'un dictionnaire bilingue imparfait, lui-même extrait automatiquement de corpus monolingues, le système apprend progressivement à perfectionner ses traductions. Initialement réalisées mot-à-mot, les traductions vont s'améliorer itérativement dans un jeu « d'aller-retour » par lequel une phrase source est traduite en langue cible, puis rétro-traduite en langue source : la ressemblance entre la phrase originale et le résultat de la double traduction informe le système sur la conformité de son fonctionnement et l'oriente dans ses apprentissages[11].

Une troisième innovation récente concerne la prise en charge de l'imperfection des documents sources, et de la nécessité de traiter de manière robuste des entrées comprenant des coquilles, des emprunts, ou des abréviations inconnues – il s'agit alors de renoncer aux unités lexicales et de décomposer chaque phrase en unités plus petites, et moins nombreuses, dont la combinatoire permettra de prendre en charge des énoncés incorrects ou bruités et traduire un vocabulaire illimité. L'exemple suivant illustre l'effet de ce dispositif, testé de nouveau avec l'outil DeepL[12] :

[10] C'est la promesse des systèmes récemment proposés par Johnson et ses co-auteurs (2017).

[11] Voir par exemple Lample et co-auteurs (2018).

[12] On pourra apprécier la capacité du système de TA à reproduire des régularités morphologiques sur des termes qui sont inventés pour l'occasion (plastimère, pastimètre, véloquettevive) et penser qu'il ne s'en sort finalement pas si mal, faisant preuve indiscutablement d'une certaine forme de créativité.

Le plastimère-manucoïde ne vaut pas le pastimètre perpétoïde : oui mais cependant il le double mieux. Parfois il en dissimule les défauts et cependant j'en préfère la rivale : la serpentrière-véloquettevive qui coûte moins cher sans vaudre plus. / The manucoid plastimer is not worth the perpetual pastimeter: yes, but it doubles it better. Sometimes he hides its defects and yet I prefer the rival: the snakeback bicycle that costs less without more vaudre.

(Le texte français est extrait de « L'opérette imaginaire », de V. Novarina)

En s'affranchissant de la notion de mot, en mélangeant les corpus issus de domaines, voire de langues, différents, en s'appuyant sur de nouvelles architectures d'apprentissage automatique, les systèmes de TA neuronaux réalisent des tours de force techniques et sont à même de produire des correspondances entre énoncés sources et cibles qui ne s'inspirent directement d'aucun exemple de traduction humaine. Les messages ainsi produits sont incontestablement utiles, et reproduisent souvent (à gros grains) des équivalences sémantiques. Ils sont pourtant difficiles à évaluer en tant que traductions, faute de disposer d'un référentiel humain qui permettrait d'apprécier leur justesse.

Les traductions automatiques sont-elles des traductions ?

Pour discuter cette question, on s'appuiera sur les études maintenant classiques du « traductionnais »[13] ou *langue de la traduction* : de nombreux travaux ont montré qu'elle différait des énoncés naturels en langue cible à la fois par la présence d'interférences avec la langue source, par une moindre diversité lexicale, par une tendance à la normalisation, enfin par l'ajout d'explicitations nécessaires à assurer à une bonne réception du message source (Baker, 1993). Plusieurs de ces différences s'observent directement par des analyses de corpus, qui confirment que les textes cibles sont en moyenne plus longs que textes sources, que le rapport entre types et occurrences, qui mesure la diversité lexicale, y est plus réduit, etc. En se fondant sur de telles mesures de surface, il est possible de détecter avec une bonne confiance si un texte est une traduction ou bien un texte original (Baroni et Bernardini, 2003 ; Volansky et al, 2014).

Pour l'essentiel, le « e-traductionnais »[14] présente les mêmes caractéristiques qui le rapprochent d'une traduction littérale : moindre variété lexicale, sur-représentation des vocables très fréquents, tendance à reproduire l'ordre des mots des textes sources, auxquelles

[13] Traduction libre de l'anglais « translationese ».

[14] Le traductionnais de la machine.

s'ajoutent les effets des erreurs propres aux traductions automatiques, relatifs, par exemple, à la faible cohésion lexicale ou à la méconnaissance de termes ou d'idiomes. Entraînés, au moins en partie, sur des textes qui sont des traductions, il n'est pas surprenant que les systèmes de TA manifestent – en les amplifiant – les biais du traductionnais; il n'est guère surprenant non plus que le repérage automatique des productions de la TA soit encore relativement aisé (Haroni et al, 2014)[15].

Il reste pourtant une différence de taille toutefois entre traductionnais et e-traductionnais, l'absence d'explicitation[16] dans les textes traduits automatiquement, manifestée par leur plus grande similarité avec le texte source (Burlot et Yvon, 2018). Ignorant des différences culturelles et linguistiques entre émetteur(s) et récepteurs, indifférents aux asymétries entre les deux côtés des textes parallèles, les traducteurs automatiques, qui manipulent les textes au niveau des caractères, ne fournissent ni compréhension, ni explication. Ce constat doit être mis en relation avec les méthodes du travail des praticiens de la TA, qui ne distinguent qu'exceptionnellement dans leurs apprentissages et dans leurs évaluations les textes sources originaux et ceux qui sont des traductions[17]. Ce constat doit être également relié avec le postulat sous-jacent que la traduction automatique se formalise par une opération mathématique *inversible* : en passant de source à cible, puis de cible à source, *on devra retrouver le même texte*. Cette idée est directement à l'œuvre dans les méthodes d'apprentissage non supervisé : en l'absence de corpus parallèle, on entrainera le système de TA à transformer (encoder) chaque phrase source en une phrase cible sémantiquement équivalente, à partir de laquelle la traduction inverse devra reconstruire le plus exactement possible la phrase d'origine (Lample et co-auteurs, 2018).

Conclusion : le futur des deux TA

En s'appuyant sur les méthodes d'apprentissage automatique (profond) qui exploitent des grands corpus parallèles, les systèmes de traduction automatique ont fait des progrès remarquables et tangibles qui les rendent utiles à une grande variété d'utilisateurs. Dans la

[15] Ce constat, qui valait pour la génération précédente des systèmes de TA statistique, mériterait d'être réévalué pour la nouvelle génération de systèmes neuronaux.

[16] Pour nuancer cette observation, on rappellera que la plus grande majorité des exemples de traduction que les systèmes de TA cherchent à reproduire sont des textes relativement techniques, qui proposent peut-être moins d'occasions d'expliciter le texte source.

[17] Au risque de présenter des performances exagérément optimistes de leurs systèmes, comme le montre l'analyse réalisée par Toral et ses co-auteurs (2018) des évaluations du premier système neuronal de Google.

plupart des cas, ces traductions restent pourtant imparfaites et nécessitent toujours une intervention humaine : celle du post-éditeur qui en corrigera les erreurs quand la traduction doit être publiée ; celle du lecteur qui rectifiera et complètera – quand c'est possible - les informations erronées ou manquantes, lorsqu'elle ne mérite pas de l'être. Ces deux grands contextes d'usage diffèrent de bien des manières et ne pas les distinguer est source de multiples confusions. Pour servir ces usages, les différents types de systèmes de TA doivent continuer à progresser, dans des directions qui commencent, peut-être, à diverger.

Pour l'usage « grand public », les recherches actuelles s'attachent en premier lieu à contourner la pénurie de données parallèles dans le but d'améliorer la qualité moyenne des traductions automatiques pour le plus grand nombre possible de genres, de registres, de domaines et de couples de langues, avec pour visée un système unique capable de traduire toutes les directions de traduction (Johnson et al, 2017). Ces recherches auront un fort impact sur tout le domaine du traitement des langues, permettant, par traduction et par transfert, de construire des données annotées pour toutes grandes les applications du domaine, et ce dans un maximum de langues. Les nouvelles architectures de traduction progressent également dans la fluidité et la cohérence des textes générés, qui les rendent de plus en plus indiscernables d'authentiques textes en langue cible, ce qui n'est pas sans poser d'autres problèmes[18] : qui garantira à l'internaute que ce qu'il lit dans une langue en apparence soignée retranscrit fidèlement le message de la langue source – que, par définition, il ignore ?

Pour ce qui concerne les traducteurs professionnels, on peut penser que les améliorations[19] à venir sont d'un autre ordre : amélioration de la prédictibilité et de la systématicité des choix de traduction, introduction de capacités d'explication, d'apprentissage à la volée pour aller vers des TA spécialisées à des contextes particuliers de traduction, voire intégralement personnalisées, meilleure ergonomie de la collaboration humain-machine. En parallèle à ces recherches, des efforts supplémentaires sur la mesure automatique de la qualité des traductions et le repérage automatique des erreurs doivent être réalisés.

---

[18] Les textes engendrés automatiquement par les modèles neuronaux « dernier cri » de la société OpenAI sont tellement réalistes que ses concepteurs ont décidé de ne pas les rendre disponibles à la communauté scientifique, de peur qu'ils soient détournés pour produire des infox (https://openai.com/blog/better-language-models/, visité le 20 juillet 2019).

[19] Ceci bien entendu sans préjuger des effets, qui sont loin d'être tous positifs, de la diffusion de ces nouveaux outils sur les conditions d'exercice du métier de traducteur. Pour certains couples de langues ou certains types de documents, il est prévisible que la qualité croissante des traductions automatiques va continuer d'exercer une pression à la baisse sur les tarifs des traducteurs, qui sont de plus en plus souvent sollicités pour réviser à bas prix des traductions automatiques plutôt que pour produire en toute autonomie des bio-traductions.

En conclusion, la polysémie du mot « traduction » dans « traduction automatique », conduit à distinguer au moins deux[20] conceptions de la TA : il semble que de la première, les traducteurs humains n'aient pas beaucoup à craindre pour leur activité, et que de l'autre, ils aient peut-être beaucoup à attendre.

RÉFÉRENCES BIBLIOGRAPHIQUES

Mona Baker. Corpus linguistics and translation studies: Implications and applications. In Gill Francis Mona Baker and Elena Tognini-Bonelli, eds, *Text and technology: in honour of John Sinclair*, pages 233–252. John Benjamins, Amsterdam, 1993.

Marco Baroni et Sylvia Bernardini (2006). *Literary and Linguistic Computing,* 21 :3, 2006.

Jacqueline Léon (2015). Histoire de l'automatisation des sciences du langage. ENS Editions, 2015. URL : https://books.openedition.org/enseditions/3737

Franck Burlot et François Yvon (2018). Using Monolingual Data in Neural Machine Translation: a Systematic Study. *In Proceedings of the third Conference on Machine Translation (WMT), pages 144-155, Brussels, Belgium, 2018.*

Roee Haroni, Moshe Koppel, Moshe et Yoav Goldberg (2014). Automatic Detection of Machine Translated Text and Translation Quality Estimation. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 289-295, Baltimore, Maryland.

Antony Hartley et Andrei Popescu-Belis (2004). Évaluation des systèmes de traduction automatique, S. Chaudiron (éd), Évaluation des systèmes de traitement de l'information, Paris, Hermès, 2004, p. 311-335.

Eduard Hovy, Margaret King et Andrei Popescu-Belis (2002). Principles of Context-Based Machine Translation Evaluation. *Machine Translation,* 17: 43 (2002).

Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, Jeffrey Dean (2017). Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. *Transactions of the ACL*, 5:339-351, 2017.

Kishore Papineni, Salim Roukos, Todd Ward, et Wei-Jing Zhu (2002). Bleu : a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 311–318, Philadelphia, PA.

Guillaume Lample, Marc'Aurelio Ranzato et Ludovic Denoyer (2018). Unsupervised Machine Translation Using Monolingual Corpora Only, in *Proceedings the international conference on representation learning* (2018).

[20] On pourrait également évoquer les usages de la TA à des fins d'interprétation, d'apprentissage des langues, d'indexation automatique, etc.

Antonio Toral, Sheila Castilho, Ke Hu et Andy Way (2018). Attaining the Unattainable? Reassessing Claims of Human Parity in Neural Machine Translation. *In Proceedings of the third Conference on Machine Translation*, Brussels, Belgium, 2018.

Vered Volansky, Noam Ordan et Shuly Wintner (2013). On the features of Translationese. *Digital Scholarship in the Humanities,* 30(1):98-118