

## Sommaire

<b>Introduction.....</b>	<b>11</b>
--------------------------	-----------

<b>Théorie de la TA.....</b>	<b>17</b>
------------------------------	-----------

Jean-Louis Vaxelaire – Les progrès de la traduction automatique par le prisme du turc et du luxembourgeois.....	19
---	----

Aurélien Talbot – La « pensée-interprète », les appareils et la diversité linguistique.....	31
---	----

Jacques Coulardeau – Humaniser la machine à traduire ?.....	45
---	----

<b>TA, enseignement et formation.....</b>	<b>61</b>
---	-----------

Maria Zimina, Christopher Gledhill - L'impact de la traduction automatique sur les pratiques langagières et professionnelles des apprentis-traducteurs : entre apports en efficacité et menaces pour la diversité des discours.....	63
---	----

Janina Di Pierro Cárdenas, Renata De Rugeriis Juárez - Inteligencia artificial y SoftPower de la traducción asistida y automática: perspectivas en el proceso de enseñanza-aprendizaje de idiomas.....	83
--	----

Elena Kokanova, Aleksandra Epimakhova, Maxim Berendyaev, Nikolay Kulikov, Maria Evgrafova, Nata-	
--	--

Iya Pak – Traduction automatique, un nouveau défi pour la formation des traducteurs : le cas de l’Université fédérale Arctique (Russie).....	101
Eric Navé – Qu’en pensent-ils ? Réflexion sur les représentations d’apprenants et d’enseignants d’Arabie saoudite vis-à-vis du recours aux outils de traduction automatique en contexte FLE débutant.....	115
<b>TA et pratique de la diversité linguistique.....</b>	<b>131</b>
Nicolas Bacaër - Traduire automatiquement des articles dans les sciences dites dures.....	133
Christian Tremblay - La traduction automatique dans le contexte des institutions européennes - Essai de traduction automatique comparée.....	143
María Isabel Rivas-Ginel – La traducción automática en la localización de videojuegos.....	167
Claire Larssonneur – Alexa, Siri : la diversité linguistique au prisme des agents conversationnels.....	179
Séraphin Personne Feykéré – Dynamique du sängo : Fonction, procédures et méthodologie relatives à l’emploi étendu (production automatique et lecture automatique des textes écrits).....	199
<b>Conclusions : Jean-Gabriel Ganascia.....</b>	<b>211</b>
<b>Documents.....</b>	<b>225</b>

Jean-Gabriel Ganascia - Les langues comme composants des processus de créativité.....	227
---	-----

## Introduction

Depuis quelques années les logiciels de traduction automatique ont fait des progrès spectaculaires à la faveur de la constitution sur Internet d'immenses corpus de textes traduits par des traducteurs humains, tels que la base de données de l'Union européenne ou de l'Office européen des brevets, et de l'émergence informatique de la technologie des réseaux neuronaux ou réseaux de neurones artificiels, une des branches de l'intelligence artificielle.

Cet ouvrage, intitulé *Traduction automatique et usages sociaux des langues. Quelles conséquences pour la diversité linguistique ?*, recueille les communications présentées lors du colloque du même nom, organisé par l'Observatoire européen du plurilinguisme (OEP) et l'Université de Paris (Séminaire « Politiques linguistiques en Europe »), le 25 novembre 2020. Le colloque devait se dérouler en présentiel à l'Université de Paris mais, à cause des normes sanitaires de confinement, s'est déroulé finalement en visioconférence.

L'objet de ce colloque n'était pas prioritairement d'examiner les progrès techniques réalisés ni même de comparer l'efficacité respective des systèmes aujourd'hui présents sur le marché, mais il était évidemment nécessaire d'évoquer le contexte technique pour bien comprendre les nombreuses implications économiques, sociales, culturelles et politiques. Tous les usages et les pratiques professionnelles faisant appel à la traduction de manière régulière ou occasionnelle sont aujourd'hui concernés. Mais, au-delà du recours à la traduction, sont aussi concernés les emplois des langues dans la communication ordinaire où la Traduction Automatique peut permettre l'intercompréhension entre locuteurs qui n'ont pas de langue commune pour communiquer entre eux.

Il nous est donc apparu important de faire le point sur les progrès et les limites de la traduction automatique, mais surtout de susciter des réflexions sur ses effets multiples. Il est, en particulier, crucial de

problématiser les utilisations de la Traduction Automatique par rapport aux apprentissages des langues en vue des emplois effectifs de celles-ci.

Le développement de la traduction automatique concerne donc les particuliers, les entreprises et administrations, les professionnels de la traduction, les enseignants et les étudiants, les chercheurs, les médias, et un long etcétéra.

Selon le rapport que l'on entretient avec les langues, l'usage que l'on peut faire de la Traduction Automatique est très variable, et ses conséquences sur la connaissance et l'emploi des langues peuvent s'avérer très différentes. L'activité principale d'un traducteur professionnel, par exemple, peut finir par consister essentiellement à vérifier et à améliorer la traduction obtenue par la Traduction Automatique, plutôt que de traduire à proprement parler. Dans ce cas, c'est alors vers une redéfinition de ce métier que l'on s'orienterait à la fois en termes de productivité, mais aussi de méthodes ou même de compétences à acquérir et à mettre en œuvre, avec les conséquences que l'on doit imaginer sur la formation des traducteurs et des interprètes.

Cependant, la traduction automatique est susceptible d'engendrer de nouvelles inégalités entre les langues (ou de les reproduire) dans la mesure où sont et seront concernées en premier lieu les langues pour lesquelles on dispose déjà de bases de données importantes, nécessaires au traitement automatique, tandis que pour d'autres, pour diverses raisons, cela sera plus difficile à réaliser.

On peut se demander également quelles conséquences la Traduction Automatique pourrait avoir sur le désir d'apprendre des langues, sur les modalités d'acquisition de celles-ci, sur les compétences en langues à développer en compréhension et en production écrite et orale, et sur les compétences déjà acquises.

L'objet de ce colloque, comme cela a été souligné dans l'appel à communications, était donc de « caractériser les multiples aspects des emplois sociaux croissants de la Traduction Automatique, d'en

mesurer les enjeux pour les locuteurs, les institutions, les entreprises, en particulier en tant qu'elle pourrait favoriser ou non certaines formes de plurilinguisme individuel ou de multilinguisme sociétal. »

Les communications programmées au colloque ont été classées en trois groupes, qui correspondent, d'ailleurs, aux trois parties de l'ouvrage :

- 1) Les communications concernant les aspects théoriques de la traduction automatique
- 2) Les communications concernant les applications didactiques de la Traduction Automatique
- 3) Les communications concernant les répercussions de la traduction automatique sur la diversité linguistique et le plurilinguisme

La première partie de l'ouvrage, qui porte sur la théorie de la traduction automatique, souligne surtout les limites de la TA, en dépit des prodigieux progrès que nous devons à la technologie neuronale.

Jean-Louis Vaxelaire pointe les écarts de traduction, selon le couple de langues choisi (anglais-français, français-turc ou français-luxembourgeois, à titre d'exemple) et selon la diffusion des langues ou leurs caractéristiques intrinsèques.

Aurélien Talbot inscrit la TA dans le processus de recreation que suit le traducteur via la « déverbalisation », et voit dans la TA plus un partenaire du professionnel qu'une alternative au professionnel. Il met en lumière les zones sensibles de la TA que sont le contexte et le genre du texte.

Quant à Jacques Coulardeau, il identifie tout au long de la chaîne de la traduction, du concepteur de logiciel de TA, à l'utilisateur final en passant par l'entreprise chargée du développement et de la commercialisation, les différents points de biais sémantique et place ces biais sous l'œil de la responsabilité civile. Les risques ne cesseront de croître dans le contexte de la mondialisation avec le développement

des usages qui sont déjà omniprésents dans les modes d'emploi de tous les produits que nous utilisons.

La seconde partie de l'ouvrage traite de la traduction automatique dans l'enseignement et la formation.

Maria Zimina et Christopher Gledhill analysent au plan pédagogique, technologique, linguistique et de la gestion de projet des cours de niveau Master intitulés « Outils de traduction de site » et « Traduction de site web vers l'anglais » assurés en co-enseignement.

Janina Di Pierro Cárdenas et Renata De Rugeriis Juárez posent la question de l'utilisation de la TA dans l'enseignement-apprentissage des langues étrangères. À la lumière des développements actuels, il est peu vraisemblable que les modes d'apprentissage ne soient pas impactés et que la manière d'enseigner ne soit pas influencée, conduisant à une redéfinition du rôle de l'enseignant.

Elena Kokanova, Aleksandra Epimakhova, Maxim Berendyaev, Nikolay Kulikov, Maria Evgrafova et Natalya Pak font apparaître la tension existant entre une société russe où les outils de TA sont de plus en plus présents, notamment chez les jeunes lycéens ou étudiants et l'enseignement où les outils de TA n'ont pas encore fait leur entrée. Les auteurs développent l'impact de la traduction neuronale (TAN) sur le métier de traducteurs et par voie de conséquence sur la formation des traducteurs à tous les niveaux de cette formation.

Éric Navé, sur la base d'entretiens avec des étudiants et des enseignants de FLE en Arabie Saoudite, vient compléter opportunément les articles précédents, en cherchant à appréhender la perception qu'ont les étudiants et les enseignants de l'introduction de la TA dans les formations et sur les implications que peuvent avoir les avancées en TA sur l'apprentissage d'une langue étrangère.

La troisième partie de l'ouvrage est consacrée à la relation entre TA et la diversité linguistique.

Avec Nicolas Bacaër, nous entrons très vite dans le vif du sujet. Son article pourrait s'intituler « les tribulations d'un chercheur en sciences dures francophone et qui entend le rester avec la TA ». Il

nous fait la démonstration irréfutable qu'il est possible de lire en français des milliers d'articles initialement rédigés en anglais, qu'il est possible de traduire soi-même des articles en anglais vers le français mais surtout, que publier en français et traduire, en même temps ou plus tard, en une ou plusieurs langues, dont l'anglais, est un très bon moyen de conserver le français comme langue scientifique et d'assurer la diffusion de ses propres productions scientifiques.

Christian Tremblay se penche sur l'utilisation de la TA dans les institutions européennes. On ne sera pas surpris de constater que les institutions européennes font preuve dans ce domaine d'une prudence suspecte. Toutefois, la présidence allemande en cette fin d'année 2020 s'est montrée audacieuse en introduisant la TA sur son site pour les traductions vers 22 langues européennes, hors l'allemand, le français et l'anglais. C'est à une évaluation de cette expérience prometteuse que s'exerce Christian Tremblay, non sans évoquer quelques pistes de recherche pour améliorer le contrôle de la qualité de la traduction.

Dans un autre domaine qui intéresse une part importante de la population mondiale, celui des jeux vidéo, María Isabel Rivas-Ginel fait ressortir à la fois l'importance de la localisation qui par nécessité doit faire appel à la TA, mais aussi inversement le poids des contraintes techniques et le faible motivation des fabricants dans cette direction.

Claire Larssonneur s'attache à la question des « agents conversationnels », c'est-à-dire de ces programmes qui gèrent les interactions entre l'utilisateur et les plateformes auxquelles on accède sur smartphone, tablette ou ordinateur. Portés par des entreprises telles qu'Apple ou Amazon, ces agents conversationnels montrent une ouverture multilingue dictée par leur diffusion internationale. Mais il s'agit d'un multilinguisme restreint, contraint par le contexte nord-américain et à ne pas confondre avec une réelle diversité linguistique.

Il n'est pas inutile que la problématique de la TA soit évoquée dans un cadre très éloigné des contextes traités jusqu'à présent, celui des politiques linguistiques africaines et de la coexistence de langues na-



tionales avec les langues européennes devenues africaines, ici le français en République centrafricaine. C'est à cette incursion que nous invite Séraphin Personne Feykéré.

Enfin, pour terminer, nous avons fait appel à Jean-Gabriel Ganascia, expert internationalement reconnu en intelligence artificielle, à la fois pour introduire la problématique de notre colloque et pour observer son déroulement.

Nous le laissons dans les conclusions de cet ouvrage prendre tout le recul historique, scientifique et philosophique nécessaire et retracer les atouts, les limites et les grands enjeux sociétaux de la traduction automatique.

José Carlos Herreras et Christian Tremblay

## **Théorie de la TA**



# **Jean-Louis Vaxelaire – Les progrès de la traduction automatique par le prisme du turc et du luxembourgeois**

## **Résumé**

La qualité des traductions automatiques a fait un bond ces dernières années et il est possible que les moqueries des générations précédentes soient remplacées par un usage quotidien de la TA de la part des adolescents. Cependant, ces bons résultats semblent toujours être de ou vers l'anglais, il n'est pas certain qu'ils le soient entre des langues bien moins parlées. Nous observerons donc deux langues très différentes, une avec beaucoup de locuteurs mais un poids international faible, le turc, et une autre qui a peu de locuteurs et un poids international encore plus faible, le luxembourgeois. Le choix des langues, mais aussi la question du genre des textes, sont encore des obstacles pour une traduction automatique de qualité.

---

## **Abstract**

**The progress of machine translation through the prism of Turkish and Luxembourgish**

The quality of machine translations has jumped in recent years, and it is possible that the mockery of earlier generations is being replaced by everyday use of MT by teenagers. However, these good results always seem to be from or into English, it is not clear that they are between much less spoken languages. We will therefore observe two very different languages, one with many speakers but low international weight, Turkish, and another with few speakers and even lower international weight, Luxembourgish. The choice of languages, but also the question of the genre of the texts, are still obstacles for a quality machine translation.

---

## ***1. Une nouvelle aube***

L'histoire de la traduction automatique semble être faite de poussées d'optimisme puis de reculs face aux difficultés rencontrées. Le découpage chronologique en paradigmes (traduction par règle, traduction statistique et traduction neuronale) est schématique (Vaxelaire, à paraître) puisqu'il a toujours existé des chercheurs qui travaillaient dans l'une ou l'autre approche, mais il est évident que ces dernières

années ont vu tous les acteurs principaux du domaine se lancer dans la traduction neuronale.

Un historien de la TA pourrait sourire devant l'enthousiasme des équipes de recherche puisque ce n'est pas la première fois que l'on annonce un système qui traduit parfaitement un énoncé. La différence ici étant que les systèmes ne se contentent pas de traduire correctement un énoncé mais peuvent donner de bons, voire de très bons, résultats avec des textes entiers. Toutefois, lorsqu'une équipe de Google (Wu *et al.*, 2016 : 20) et une autre de Microsoft (Hassan *et al.*, 2018 : 21) affirment avoir de meilleurs résultats que les traducteurs humains, il faut noter que l'anglais est toujours une des deux langues employées dans les travaux (anglais-français et anglais-allemand dans le premier cas, chinois-anglais dans le second).

Une vision naïve pourrait laisser penser que si les résultats sont bons entre le français et l'anglais et entre l'anglais et le chinois, ils le seront forcément entre le français et le chinois. Rien n'est pourtant moins sûr.

## ***2. Les langues du monde ne se résument pas à l'anglais***

Pour ne pas inclure l'anglais dans l'équation, nous allons observer les traductions en français de deux langues très différentes sur le plan morphologique et du point de vue du nombre de locuteurs.

La première, le turc, est la langue maternelle de 90 % des Turcs. Elle fait partie du *World's Major Languages* de Comrie (republié par Routledge en 2009) en raison de son nombre important de locuteurs. En dehors de Turquie, on la retrouve dans des minorités de pays des Balkans, mais elle n'est langue officielle que d'un seul autre pays, Chypre, où le turc n'est réellement utilisé que dans la partie nord de l'île, dans une république qui n'est pas reconnue par les Nations Unies.

Bien qu'il y ait plusieurs dizaines de millions de locuteurs, les corpus entre le turc et le français sont limités, il est donc intéressant d'observer

ver quels sont les résultats donnés par la traduction automatique de Facebook (désormais FB) mais aussi de Google Translate (désormais GT), Translator.eu (un site qui utilise les systèmes de Microsoft, désormais TR.EU) et Reverso qui offrent le turc dans leur liste de langues.

Débutons avec une expression figée, « Denizden babam çıkırsa yerim » (74 000 résultats sur Google le 24/12/20), que l'on pourrait traduire par « je mangerais même mon père s'il sortait de la mer » pour indiquer son goût pour les fruits de mer. GT propose « Je mangerais si mon père quitte la mer. » le 19/05/20 et un encore moins clair « Je mangerais mon père de la mer. » le 24/12/20. Microsoft est plus proche de l'ancienne version de GT avec « Je mangerais si mon père sortait de la mer. ». Dans les deux cas, la traduction est trop littérale pour être compréhensible<sup>1</sup>.

Le turc est intéressant en raison de ses spécificités. C'est par exemple une langue qui n'a pas de genre grammatical, ce qui entraîne inévitablement des ambiguïtés. Prenons le cas d'une annonce pour la disparition d'une enfant :

!! LÜTFEN PAYLAŞALIM...ACİL !!

Bir Arkadaşımızın Kız Kardeşi KAYIP.

Gülce P., 11 yaşında.

10. Viyana'da kalıyor kendisi.

Bu sabah okula çıkıyorum diye çıkmış. Sonra bir daha haber alınamamış kendisinden (FB, 07/11/20)

La traduction en anglais de FB peine à régler ce problème de genre : le terme *kız* indique qu'il s'agit d'une jeune fille, de 11 ans comme l'indique la suite. Le premier pronom est faux : « he stays in Vienna<sup>2</sup>. » Le reste de la traduction alterne entre le féminin et le masculin : « She went out this morning because she went to school. Then

---

<sup>1</sup> Celle de Reverso est simplement indéchiffrable : « Si mon père sort de la mer, je suis un endroit. »

there was no news from him again ». Chez GT ainsi que chez T.EU, on se contente du masculin bien qu'il s'agisse d'une sœur : « Il reste à Vienne. Il est parti ce matin parce qu'il allait à l'école. Puis il n'y avait plus de nouvelles de lui » et « Il reste à Vienne. Il est parti parce que j'allais à l'école ce matin. Puis on ne lui a plus jamais entendu parler. » La présence d'un pronom de première personne chez Microsoft est surprenante. Enfin, Reverso alterne également les genres et ajoute une erreur : « Il reste à Vienne. Elle n'est pas allée à l'école ce matin. Puis il n'était pas au courant à nouveau ». Aucune de ces traductions n'est malheureusement d'une grande aide pour retrouver cette jeune fille.

Un problème similaire à celui qui est évoqué dans la note 2 se retrouve dans cette autre annonce :

Dr. Burhan Nalbantoğlu Devlet Hastahanesinde yatmakta olan annem Yüksel H. için 0rh+ kana ihtiyaç vardır. Verebileceklerin devlet hastahanesi kan bankasına bu isimle vermeleri rica olunur. (FB, 05/11/20)

Une femme demande que l'on donne du sang O+ pour sa mère Yüksel H. Le nom de l'hôpital est *Dr. Burhan Nalbantoğlu Devlet Hastahanesi*, mais FB et GT séparent *Dr.* du reste du nom et le conservent en début d'énoncé, T.EU fait une coupure<sup>3</sup>, seul Reverso maintient toutes les parties ensemble.

Le plus gros problème provient de FB qui efface complètement le nom de la mère, si bien que la conclusion devient problématique : « Those who can give it to the state hospital blood bank under this name. » Le message s'adresse évidemment à des gens qui parlent

2 Les noms propres posent également des problèmes comme le 10<sup>e</sup> arrondissement de Vienne dont le 10. sera séparé du reste du terme par tous les systèmes

3 « Le Dr Burhan Nalbantoglu 0rh+ sang est nécessaire pour ma mère Yüksel H., qui est couchée à l'hôpital d'État. Ceux qui peuvent donner à la banque de sang de l'hôpital d'État sont invités à le donner par ce nom. »

turc, mais celui qui, comme moi, reçoit ce message en anglais, ne peut rien y comprendre sans passer par l'original.

Les noms propres sont aussi la preuve que l'anglais est utilisé comme langue pivot entre le turc et le français. Ainsi, cette annonce pour un concert de Sifir à Istanbul en janvier 2020 :

En son, Türk Deneysel Müzik Antolojisi albümüyle Borusan Müzik Evi dinleyicileriyle buluşan SIFIR.

GT (16/01/20) : SIFIR a rencontré les auditeurs de Borusan Music House avec le dernier album de Turkish Experimental Music Anthology.

Reverso (24/12/20) : Le dernier est SIFIR, où Turkish Experimental Music Anthology rencontre les auditeurs de Borusan Music House.

Translator.eu (10/11/20) : Plus récemment, l'album Turc Experimental Music Anthology Borusan Music House rencontre le public ZERO.

La traduction de Google est compréhensible, si ce n'est que *en son* (« la dernière fois ») a disparu, mais on voit que le nom du disque et celui de la salle de concert passent du turc à l'anglais. Reverso est moins clair et T.EU se perd en ajoutant un peu de français (*Turc* dans le nom de l'album) et en allant jusqu'à traduire le nom du groupe. Si le recours à une langue pivot se comprend pour des raisons de corpus, il risque toujours d'entraîner des erreurs en ajoutant une couche linguistique dans le transfert.

Dans des écrits relâchés, on peut voir un certain nombre d'anglicismes, par exemple *track*, dans le sens de morceau de musique : Bir tane daha karantina yada koronavirus isimli bir track görürsem kusaçım<sup>4</sup> (FB, 27/03/20)

---

4 Facebook proposait alors un étrange : « I'll throw up if I see another track called koronavirus or another track ». La répétition de *another track* n'avait pas de raison et *karantina* avait disparu.



Le passage par l'anglais-pivot chez Reverso ne permet pas de conserver l'anglicisme : « Si je vois une autre voie appelée quarantaine ou coronavirus embrasser ». *kusacağım* pris isolément est bien traduit par « je vais vomir », mais devient étrangement *embrasser* dans ce contexte. T.EU traduit également *track* en français, mais le résultat est plus satisfaisant car *piste* peut être employé dans un contexte musical, contrairement à *voie* : « Si je vois une autre quarantaine ou une piste appelée coronavirus, je vais vomir ».

La présence d'un point final peut modifier la traduction, le problème est que les écrits relâchés que l'on peut trouver sur les réseaux sociaux ne contiennent pas toujours les marques de ponctuation attendues. Ainsi d'autres extraits qui contiennent des changements de casse, des émojis et sans doute des fautes de frappe sont incompréhensibles dans leurs versions traduites.

Prenons maintenant un cas différent, celui du luxembourgeois. Bien qu'il s'agisse d'une langue nationale au Luxembourg et qu'on la retrouve dans quelques villages en Belgique, France et Allemagne, le nombre de locuteurs est estimé entre 400 et 600 000 selon les sources.

L'arrivée de cette langue parmi la liste de Google Traduction avait été saluée au Grand-Duché car cela lui permettait de se retrouver sur le même plan que les grands voisins allemand et français. Les Luxembourgeois allaient vite déchanter car les résultats n'étaient pas à la hauteur. Les deux exemples que je vais prendre ont néanmoins vécu des évolutions différentes.

Débutons avec un extrait d'un discours du premier ministre, Xavier Bettel :

« Lëtzebuerg ass e Land, dat et an der Vergaangenheet  
ëmmer nees fäerdeg bruecht huet, sech ze erneieren,  
no vir ze kucken a gestäerkt aus Krisen eraus ze  
goen. »

Rendu lors de l'intégration de la langue dans la liste de Google par :

« L'Irlande est un pays qui est dans le passé toujours amélioré off, ont été persistante, Forward Look et amélioré des crises d'aller. » (GT, 2016)

Cette traduction est déstabilisante pour les personnes qui ne savent pas comment fonctionne la TA statistique. Pris isolément, GT propose bien *Luxembourg* pour *Lëtzebuerg*, mais dans ce contexte le pays devient l'Irlande. Ce choix ne prend du sens que si l'on comprend l'utilisation de la statistique dans ce type d'outil. La présence de termes anglais relève, elle, de l'emploi d'une langue-pivot.

Le passage à la traduction neuronale devait régler ces problèmes ; c'est effectivement le cas si on lit cette traduction datant de mars 2019 :

« Le Luxembourg est un pays qui, par le passé, a toujours réussi à se renouveler, à rechercher et à se renforcer en cas de crise. » (GT, 2019)

La version de novembre 2020 est différente mais également de très bonne qualité :

« Le Luxembourg est un pays qui a réussi à maintes reprises dans le passé à se renouveler, à regarder vers l'avenir et à sortir plus fort des crises. » (GT, 2020)

Les deux premières lignes de l'hymne national seront notre second exemple : « Wou d'Uelzécht duerch d'Wisen zéit, duerch d'Fielsen d'Sauer brécht ». On pourrait les traduire par « Là où l'Alzette traverse les prés, où à travers les rochers, la Sûre creuse son chemin. » La version de Google d'alors cite la première rivière mais oublie la seconde (avec un problème de conjugaison) :

« Où l'Alzette arrose champs et prés, les rochers se baigne. » (GT, 2016)

La traduction neuronale peut-elle être une solution dans un contexte où les noms propres posent des problèmes ? Ce que propose GT en mars 2019 pose plus de questions qu'elle n'offre de solution :

« Si l'ignorant parvient à voir les chuchotements, la tristesse s'éteint. » (GT, 03/19)

La version française n'a plus aucun lien avec le texte de départ...Une traduction plus récente revient à l'original mais a toujours du mal à insérer les noms de rivières :

« Là où la rivière coule à travers les prairies, les rochers se brisent à travers les rochers. » (GT, 11/20)

La « traduction » de mars 2019 interroge nécessairement. Une réponse peut être vue dans la pratique du Google Bombing. Ce phénomène a été noté sur GT, par exemple lorsque *Russie* devenait *Mordor* dans le passage de l'ukrainien au russe ou quand *Justin Bieber* en espagnol était traduit par *idiot* en anglais (exemples observés entre fin 2015 et début 2016). S'il est possible de détourner les résultats de GT entre l'espagnol et l'anglais, il s'agit probablement d'un jeu d'enfant pour une langue telle que le luxembourgeois. Si l'on demande aux utilisateurs de noter une traduction, voire de proposer leur propre solution, le risque d'un détournement est manifeste.

### ***3. Ce qui fonctionne***

Si les résultats pour ces langues ne sont toujours pas à la hauteur des promesses, il serait utile de se demander dans quels cas la TA fonctionne.

Un consensus net fait ressortir la question de la langue, l'anglais étant pour diverses raisons celle qui permet d'obtenir les meilleurs résultats. Le fait que la plupart des travaux de recherche emploient l'anglais joue évidemment un rôle, mais les exemples de traductions réussies présentés publiquement semblent toujours contenir cette langue dans l'équation.

Poibeau fait remarquer que les plongements de mots de la TA neuronale fonctionnent bien avec une « langue à morphologie pauvre avec beaucoup de données disponibles », mais moins pour « les langues à morphologie riche » (2018 : 145). Les résultats peuvent même être mauvais pour les langues agglutinantes (c'est le cas du turc) ou les langues à morphologie riche (slaves, voire l'allemand<sup>5</sup>) (Poibeau, 2018 : 165).

Le portrait-robot qui est fait d'une langue à morphologie pauvre avec énormément de données disponibles correspond évidemment à l'anglais, et puisque l'anglais est la langue la plus recherchée internationalement, il est logique d'adopter ce modèle, quand bien même il poserait des problèmes pour d'autres langues.

D'un point de vue historique, Météo est le premier système à pleinement répondre aux attentes. Développé par l'Université de Montréal dans les années 1970, le but est de traduire des bulletins météorologiques entre le français et l'anglais. Le résultat est très satisfaisant car les traductions sont tout à fait compréhensibles et le travail était rébarbatif pour les traducteurs humains. Toutefois, ce qui fait la particularité des textes à traiter est une récurrence des mêmes schémas, avec une syntaxe et un vocabulaire très restreints. Plus que le binôme des langues, c'est la simplicité du matériel à traiter qui explique le succès de ce système. Puisque la machine doit traduire des textes, le type de texte et, en conséquence le genre de ce texte, sont naturellement des points essentiels. Jean Senellart de Systran le résume d'un point de vue pratique :

« il est beaucoup plus facile d'entraîner un moteur neuronal spécialisé dans un domaine d'activité spécifique, comme le tourisme ou la santé, qu'un moteur générique. Avec 500 phrases extraites du corpus de traduction fourni par un client et deux jours d'entraînement, on peut aujourd'hui formater un NMT professionnel. Le tout avec des

---

5 Le luxembourgeois est très proche de l'allemand mais est plus analytique, donc pas nécessairement concerné par les propos de Poibeau.

résultats beaucoup plus performants que ceux de Google Translate. » (Senellart, in Lupieri, 2019).

On touche ici le principal problème théorique de la traduction automatique. Tous les modèles de TA neuronale, comme les systèmes précédents, parlent de mots (corpus de millions de mots, plongement de mots, etc.) alors qu'on ne traduit pas des mots mais des textes. La notion de genre semble ignorée : on traduit des textes à partir de corpus dont on ne connaît pas le(s) genre(s) et on se comporte comme si tous les textes se traduisaient sur le même principe. La TA fonctionne généralement « en aveugle ». Le traducteur humain a une vision globale du texte, ce que n'a pas le logiciel. Pourtant, comme l'écrit Rastier (1998), le global détermine le local ; notre problème est que la TA ne connaît que le local.

Les travaux sur les genres (Gérard, 2019, Rastier, 2002) indiquent que chacun a ses particularités, un lexique spécifique, des schèmes récurrents, etc. Les exemples de traduction réussie donnés par les chercheurs sont le plus souvent des textes journalistiques, peut-être parce qu'il s'agit le plus souvent du type de texte qui constitue le corpus d'entraînement des systèmes. Si je reprends mes deux exemples luxembourgeois, le texte politique, plutôt proche de ce que l'on peut trouver dans la presse, est désormais correctement traduit ; l'hymne national, plus poétique, pose toujours un problème.

Les exemples turcs proviennent principalement de Facebook et, comme le note Poibeau, « les textes sur ces réseaux (tweets, story d'Instagram, messages échangés sur FB...) sont écrits dans une langue très différente de celles que l'on trouve sur le Web en général. » (2018 : 142). On y retrouve énormément d'emplois non standards, d'utilisation de la 1<sup>e</sup> personne, des invectives, etc. (Poibeau, 2018 : 147), bref des éléments qui ne se retrouvent pas dans les corpus d'entraînement. La distance entre les deux types de matériau pourrait être une explication aux résultats peu probants.

## ***Conclusion***

Si les progrès de la TA sont incontestables pour l'anglais, le passage par d'autres langues indique que la situation n'est pas si rose que ne le prétendent les différentes compagnies qui proposent des services de TA. Nous avons pu voir qu'avec des écrits quotidiens en turc ou des textes bien plus normés en luxembourgeois, les résultats étaient parfois loin d'être recevables.

Le passage du statistique au neuronal ne correspond pas à un bouleversement visible dans ce petit corpus. De toute manière, comme pour les méthodes précédentes, la traduction neuronale est plus efficace avec certaines langues qu'avec d'autres, l'anglais étant la langue parfaite. Pour autant, l'utilisation d'une langue pivot, si elle a des avantages dans ce cadre, a aussi des inconvénients. Facebook a annoncé un nouveau système qui ne passe pas par l'anglais à partir de novembre 2020, les différences ne sautent pas aux yeux pour l'instant.

Enfin, la question du genre implique que des systèmes spécialisés sont plus efficaces qu'un traducteur automatique généraliste. La TA neuronale en prenant de plus en plus en compte le contexte va dans la bonne direction, mais doit encore intégrer que les types de texte sont un facteur important pour le traducteur humain et qu'on ne traduit pas un énoncé d'une manière identique entre, par exemple, un texte didactique et un pamphlet. Les terminologues ont depuis longtemps noté qu'un même terme ne se traduisait pas nécessairement d'une manière identique selon les domaines, le problème est analogue pour les énoncés selon les genres.

## **BIBLIOGRAPHIE**

Hassan, H. *et al.*, 2018, « Achieving Human Parity on Automatic Chinese to English News Translation », <https://arxiv.org/abs/1803.05567>

- Gérard, C., 2019, « Linguistique des genres : objet et méthode. Statut culturel des genres et variétés génériques », *Linx* 78, <http://journals.openedition.org/linx/3030>.
- Lupieri, S., 2019, « La traduction automatique fait des pas de géants », *Les Échos*, <https://weekend.lesechos.fr/business-story/innovation/0600802801462-la-traduction-automatique-fait-des-pas-de-geant-2248830.php>.
- Poibeau, T., 2019, *Babel 2.0 : Où va la traduction automatique ?*, Paris, Odile Jacob
- Rastier, F., 1998, « Rhétorique et interprétation, ou le Miroir et les Larmes », in Ballabriga, M. (éd.) *Sémantique et rhétorique*, Toulouse, Éditions universitaires du sud, p. 33-57
- Rastier, F., 2002, « L'accès aux banques textuelles – des genres à la doxa », *Texte !* [http://www.revue-texto.net/Inedits/Rastier/Rastier\\_Acces.html](http://www.revue-texto.net/Inedits/Rastier/Rastier_Acces.html).
- Vaxelaire, J.L., à paraître, « Regard linguistique sur la traduction automatique », Actes du colloque Trax 2019
- Wu, Y. et al., 2016, « Google's neural machine translation system: Bridging the gap between human and machine translation », CoRRabs <http://arxiv.org/abs/1609.08144>

# **Aurélien Talbot – La « pensée-interprète », les appareils et la diversité linguistique**

## **Résumé**

Le présent article s'interroge sur les effets des progrès récents de la traduction automatique sur la pratique de la traduction professionnelle. Pour ce faire, nous revenons sur la manière dont la théorie interprétative a pu représenter le savoir-faire du traducteur professionnel en mettant au cœur de son modèle la « pensée-interprète » ou « déverbalisation ». Puis, nous interrogeons le bouleversement que la traduction automatique est susceptible de produire sur la représentation de ce savoir-faire ainsi que les risques associés concernant la diversité linguistique et les manières de traduire. Enfin, nous invitons à ouvrir une réflexion sur la bonne distance à trouver par rapport à la traduction automatique en vue d'éviter le double écueil de la technophobie et de la technophilie.

---

## **Abstract**

### **“Interpretive thinking”, devices and linguistic diversity**

The present paper examines the effects of recent advances in machine translation on the practice of professional translation. To do so, we look back at how the interpretive theory has represented the know-how of the professional translator by placing “interpretive-thought” or “deverbalisation” at the heart of its model. We then question the disruption machine translation is likely to produce on the representation of this know-how, as well as the associated risks concerning linguistic diversity and ways of translating. Finally, we propose a reflection on the right distance to be found with regard to machine translation in order to avoid the double pitfall of technophobia and technophilia.

---

## **Resumen**

### **“Pensamiento interpretativo”, dispositivos y diversidad lingüística**

En este artículo se examinan los efectos de los recientes avances en traducción automática sobre la práctica de la traducción profesional. Para ello, nos fijamos en cómo la teoría interpretativa ha representado el saber hacer del traductor profesional al colocar el “pensamiento-intérprete” o “desverbalización” en el centro de su modelo. Luego, nos preguntamos sobre el trastorno que probablemente produzca la traducción automática en la representación de este saber hacer, así como sobre los riesgos conexos en relación con la diversidad lingüística y las formas de traducir. Por último, proponemos una reflexión sobre la distancia adecuada que hay que encontrar

---



### ***1. Le savoir-faire du traducteur et « l'intelligence artificielle »***

À l'École Supérieure d'Interprètes et de Traducteurs (É.S.I.T), la traduction est présentée comme un exercice d'*intelligence*<sup>6</sup>, au sens où une bonne traduction ne reposerait pas uniquement sur des connaissances ou des *savoirs* linguistiques<sup>7</sup>, mais consisterait avant tout en un acte de compréhension et de réexpression entre lesquels se situerait une étape cruciale qualifiée de « déverbalisation ». Selon la théorie interprétative, au cœur du *savoir-faire* du traducteur, à l'oral comme à l'écrit, réside un phénomène que Danica Seleskovitch, la fondatrice de cette théorie, avait pu appeler, dès 1968, la « pensée-interprète »<sup>8</sup>. Par « pensée-interprète » ou « déverbalisation », il s'agit d'indiquer le moment du processus de traduction où le message à traduire en langue cible ne subsisterait qu'à « l'état de concept non formulé »<sup>9</sup> dans l'esprit de l'interprète-traducteur. Jean-René LADMIRAL, évoquant cette étape qu'il propose de qualifier à son tour de « dissimilation », a pu déclarer qu'elle lui « paraît représenter l'essence même de la traduction et elle se situe de plain-pied avec le vécu de celui qui traduit »<sup>10</sup>. Karla Déjean Le Féal, enfin, est allée jusqu'à affirmer, en 1993, que la « déverbalisation », constitue « le *savoir-faire* propre du traducteur » et que « toutes les connaissances et les capaci-

---

6 SELESKOVITCH, D. et LEDERER, M., 2014, p. 152.

7 SELESKOVITCH, D., LEDERER, M., 2014, p. 242.

8 SELESKOVITCH, D., 1968, p. 160.

9 *Ibid.*

10 LADMIRAL, J.-R. *Sourcier ou cibliste : les profondeurs de la traduction*. Paris : Les Belles Lettres, 2016, p. 95.