

ETAT DE L'ART

TECHNOLOGIES, OUTILS, PRATIQUES ET ENJEUX ACTUELS DE LA TRADUCTION SCIENTIFIQUE

Ce rapport est publié dans le cadre du projet Traductions et science ouverte sous la coordination d'OPERAS.

Ce projet est financé par le Fonds national pour la science ouverte, qui rassemble des contributions du ministère de l'Enseignement supérieur et de la Recherche, d'universités et d'organismes de recherche français.

Pour ce projet, le Fonds national pour la science ouverte reçoit également une contribution spéciale du ministère de la culture français.

Disclaimer

The ideas and views expressed in the exploratory reports only reflect those of the experts involved in the studies and may not be representative of the opinions or policies promoted by any specific organization, institution, or government entity. The present report is therefore only intended for informational purposes.

Avertissement

Les idées et les perspectives exprimées dans les rapports exploratoires reflètent uniquement celles des spécialistes ayant contribué aux études et ne sont pas nécessairement représentatives des opinions ou des politiques promues par une organisation, une institution ou une entité gouvernementale spécifique. Le présent rapport est donc uniquement diffusé à des fins d'information.

L'ESSENTIEL

1/ NOTRE MISSION : SCIENCE OUVERTE ET DIFFUSION DE LA RECHERCHE EN LANGUES NATIVES

- **Traduction assistée par ordinateur** pour accompagner les gestes des chercheurs, traducteurs et autres acteurs (enseignants, grand public,...) et qui pourra proposer l'aide d'un moteur de traduction automatique spécialisé par champ disciplinaire (entraîné spécifiquement sur des corpus disciplinaires).
- **Découvrabilité multilingue** des publications scientifiques dans d'autres langues grâce à la traduction automatique
 - indexation standard des métadonnées (mots-clés, titre) et des résumés en multilingue
 - survol des contenus grâce à la traduction, lire "en diagonale"
 - découverte d'autres textes par analogie avec le premier, par recommandation multilingue "vous pourriez aussi être intéressé par les articles suivants", sérendipité
- **Rédaction assistée** dans une langue cible

2/ L'ÉTAT DE L'ART

- **Dominance de l'anglais**, marginalisation (voire effacement) des autres langues et traditions de recherche
- Recherche en **traduction automatique** : développements importants

3/ NE PAS FAIRE / NOS POINTS D'ALERTE

- Se **limiter au bilingue**, type français-anglais
- **Du tout automatique**

4/ NOS RECOMMANDATIONS

- Un environnement numérique de traduction outillée qui s'inspire des pratiques effectives des traducteurs (leurs gestes)
- L'automatisme comme une des aides qui peuvent être sollicitées par l'humain (et qui doit rester une assistance) !
- Traçabilité de la traduction, qu'elle soit humaine ou automatique

LES AUTEURS

direction de l'étude

l'atelier universel

AGENCE DE DESIGN ET DE CONSEIL EN INNOVATION

- experte des démarches centrées utilisateurs,
- spécialisée dans les projets complexes, notamment dans les logiciels professionnels pour la recherche et la science ouverte
- croisant design désirable et intrants scientifiques
- habituée à travailler avec des chercheurs.

directeur scientifique

Philippe Lacour

DOCTEUR ET CHERCHEUR SPÉCIALISTE DE LA PHILOSOPHIE DU LANGAGE

Professeur titulaire au département de philosophie de l'Université de Brasilia (UnB, Brésil). Spécialiste de l'épistémologie des sciences de la culture (shs, humanités).

Responsable du projet Traduxio (<https://traduxio.org>), outil de traduction collaborative multilingue.

Co-responsable du projet de recherche "L'IA et ses critiques" (UnB).

Site personnel: <https://philippelacour.net>

comité d'experts

Claire Larssonneur

MAÎTRE DE CONFÉRENCE ET CHERCHEUSE
SPÉCIALISÉE EN ÉCONOMIE DE LA TRADUCTION,
TRADUCTION ET NUMÉRIQUE

Thierry Poibeau

DIRECTEUR DE RECHERCHE AU CNRS
ET DIRECTEUR ADJOINT DU LABORATOIRE LATTICE

SOMMAIRE

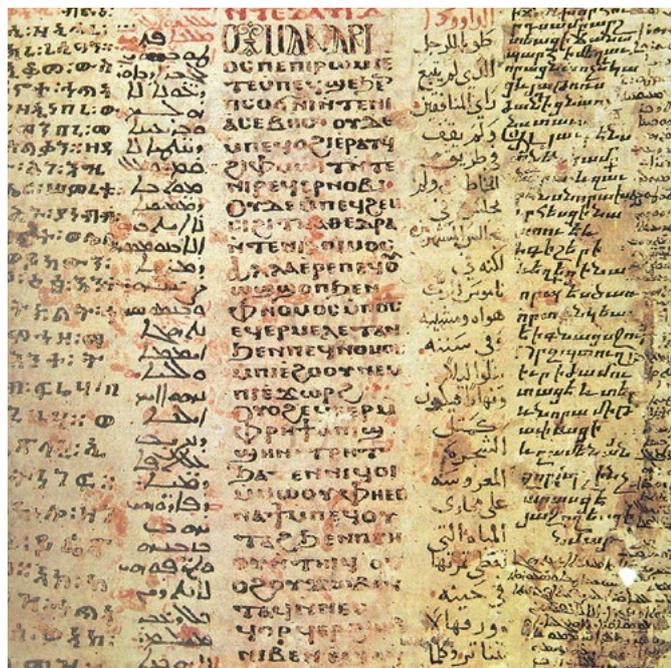
PRÉFACE	5
a_ Point de vue historique	
b_ Tendances actuelles autour de la traduction et de la recherche scientifique	
c_ La traduction automatique : une approche par probabilité qui pense l'acte de traduction comme un acte de "déchiffrement"	
d_ Couple de langues vs multilinguisme	
e_ IA VS travail de l'ombre ?	
INTRODUCTION	8
a_ Compréhension de la mission et parti pris	
b_ Les objectifs de ce premier rapport	
c_ Synthèse	
ÉTAT DE L'ART	12
01./ Définitions et objets	13
02./ La traduction scientifique, cartographie des pratiques et enjeux	17
a_ Les acteurs de la traduction scientifique	
b_ Les différents usages et pratiques de la traduction scientifique	
c_ Synthèse des entretiens: besoins remontés, enjeux et questions proposées par les personnes interrogées	
03./ Analyse d'outils	33
a_ L'outil type de TAO	
b_ Cartographie des outils - vue d'ensemble	
CONCLUSION DU RAPPORT : PREMIÈRES ORIENTATIONS DU FUTUR SERVICE DE TRADUCTION OUTILLÉE	54
a_ Incarnation du service de Traduction Outillée : un logiciel et ses extensions, les fonctionnalités envisagées	
b_ Les points critiques à prendre en compte / les questions à investiguer pour la suite	
BIBLIOGRAPHIE ET RÉFÉRENCES	57

PRÉFACE

Bref historique, notions et vision
autour de la traduction et la traduction automatique,
le point de vue de Philippe Lacour, directeur scientifique

a_Point de vue historique

Le célèbre récit de la Tour de Babel n'est qu'une des nombreuses tentatives de justification mythologique de la **diversité des langues**, en réponse à laquelle la traduction s'est vite imposée de façon privilégiée, dès lors que, malgré des illusions récurrentes, il n'existe pas de langue universelle. Historiquement, que ce soit dans la tradition occidentale (biblique) ou asiatique (textes bouddhiques), la traduction a souvent pris un tour **multilingue**, comme le montrent assez les manuscrits synoptiques et autres concordanciers.



▲
Figure: manuscrit du XIVe St-Macaire, Scété : syriaque copte arabe arménien (biblioteca apostolica vaticana).

Crédit: François-Xavier Fauvelle, *L'Afrique ancienne*, Paris, Belin, 2018, p. 17

Les différentes **techniques** appliquées au langage (alphabet, grammaire, imprimerie, numérisation) n'ont fait que confirmer cette tradition, comme l'attestent la persistance et même l'**intensification** des pratiques de traduction au cours de l'histoire.

b_Tendances actuelles autour de la traduction et de la recherche scientifique

Depuis la seconde guerre mondiale et l'invention de l'informatique, la traduction a été investie par l'ingénierie. On peut même dire que l'**intelligence artificielle** a son lieu de naissance dans les premiers programmes de traduction russe-anglais de la guerre froide. Le problème est que cette approche, s'inspirant des travaux de Turing pour décoder la machine nazie Enigma, ont tendance à **assimiler imprudemment traduction et cryptographie**. On peut bien sûr penser le langage naturel comme un code, mais toujours de façon partielle, car les langues comme l'Allemand ou l'Arabe ont bien d'autres propriétés qui risquent ainsi d'être négligées. Il faut ajouter qu'on a d'emblée demandé à la machine de se substituer au traducteur humain, dans une conception de la technologie qui privilégie la seule **automatisation** au détriment des autres (riches)

nuances de la relation homme-machine. De ce fait, le traitement automatique du langage naturel (TALN) représente aujourd'hui la quasi-totalité des technologies du langage. Comme le dit Thierry Poibeau, spécialiste du TALN, *«Il y a moins de travaux [de recherche] sur les aides pour les traducteurs au regard du nombre de travaux de recherche sur la traduction automatique.»*

Cette orientation a connu différentes phases : aux premiers programmes basés sur les **règles grammaticales** (le premier Systran des années 50) se sont substituées les tentatives utilisant les régularités **statistiques** (années 70), popularisées dans les années 90 par Google. Aujourd'hui, la plupart des logiciels de traduction automatique combinent les deux approches. L'intelligence artificielle a connu des progrès certains en recourant aux démarches de l'apprentissage profond (deep learning), qui utilisent toutefois les mêmes méthodes décrites ci-dessous. Ces pratiques placent les traducteurs humains en position de réviseurs (post-édition, finalisation), en risquant d'aspirer une partie de leur savoir-faire et, à terme, de remplacer certains d'entre eux (ou certaines de leurs tâches).

S'il existe certes des **règles** dans les langues naturelles, il existe aussi des phénomènes de **variation imprévisible**: idiotismes, évolution historique, registres de discours, néologismes, création de mots techniques, inventions (poétiques, humour, traits d'esprit, insultes), etc. C'est parce qu'ils ont toujours été sensibles à ces nuances que les traducteurs ont en général préféré les outils de **traduction assistée par ordinateur (TAO)**, parfois **collaboratifs**, qui fournissent une assistance à la réflexion, par la comparaison avec d'autres cas, l'examen d'occurrences en contexte, la consultation de définitions, mais aussi le travail en groupe, etc. Dans cette perspective, les **compétences humaines sont enrichies et non remplacées**.

c_La traduction automatique: une approche par probabilité qui pense l'acte de traduction comme un acte de "déchiffrement"

Qu'est-ce au juste que la traduction automatique neuronale ? C'est un calcul de probabilités visant à faire des conjectures sur des significations linguistiques, associé à un raisonnement inductif devant permettre une forme de prévision. Cette manière de penser le sens est parfaitement pertinente et rigoureuse, mais elle souffre aussi de **limitations intrinsèques indépensables**.

La notion de **calcul** échoue ainsi à rendre compte de la totalité des mathématiques (Gödel), et a fortiori du langage naturel. Les **probabilités** ne sont pas des fréquences et doivent donc être interprétées quand on les applique au monde empirique. La **conjecture** par calcul stochastique est donc possible, mais ne constitue pas la seule forme de présomption (on peut aussi procéder à des évaluations fines, par comparaison et pondération). Enfin, l'**induction** ne s'élève jamais qu'au général, et jamais à l'universel, de sorte que ses prévisions ne peuvent exclure les éventuelles exceptions. Aucune de ces limites, et a fortiori leur forme combinée, ne peut être dépassée par une augmentation de la puissance (ou de la vitesse) de calcul, du nombre de données ou de la souplesse des algorithmes. **Achille ne rattrapera pas la tortue** - ce qui ne veut pas dire qu'il ne sait pas courir.

d_Couple de langues vs multilinguisme

À cause des simplifications initiales, les ingénieurs ont décidé que traduire consistait à faire passer un message (codé) d'une langue A dans une langue B. De ce fait, la **quasi-totalité des interfaces** des outils de traduction automatique ou de traduction assistée par ordinateur sont **bilingues**: elles contiennent deux colonnes, pour la langue source et pour la langue cible. C'est la raison pour laquelle les gros acteurs multiplient aujourd'hui les couples de langues (ou les corpus bilingues), en utilisant parfois une langue pivot (en général, l'anglais).

Pourtant la **plupart des traducteurs connaissent plus de deux langues**, et se servent de leurs connaissances plurilingues pour affiner leur travail. Les parti pris de l'après-guerre (traduction comme cryptographie, technologie comme automatisation) ont empêché le développement de technologies de la traduction multilingues. Mais il **n'est pas du tout impossible de concevoir des outils de traduction numérique d'emblée multilingues**, à condition de s'émanciper des préjugés initiaux, et d'oser inventer des interfaces qui correspondent aux gestes traditionnels des traducteurs (consultation de dictionnaires (notamment de synonymes), lexiques, concordances, comparaison avec d'autres traductions, etc.). **Traduire consiste alors à trouver, pour la signification d'un texte singulier, des équivalents (sans identité) dans plusieurs langues**, en produisant autant d'autres versions de ce texte dans autant de langues qu'on le souhaite.

Bien entendu, **qui peut le plus peut le moins** : se donner dès le départ un outil complexe permet évidemment d'en faire des **usages simplifiés** pour la traduction anglais-français. Alors que l'inverse n'est pas vrai: il peut paraître plus économique et prudent de construire d'abord un outil de traduction avec deux langues, puis d'en rajouter ensuite, mais on se heurte alors à **des problèmes qui se révéleront vite très coûteux, voire insurmontables**. Imaginons en effet, comme c'est le cas à l'ONU, qu'un texte (de résolution) puisse être écrit dans une langue et doit immédiatement être traduit dans quatre autres langues officielles. La langue d'entrée n'étant jamais la même, il n'y a que deux solutions : soit se donner une langue pivot (interlangue, langue dite "universelle", en général tout simplement l'anglais), soit partir d'emblée d'une technologie multilingue. Sans quoi, **le risque est de devoir multiplier de façon exponentielle les couples de langues** - ce que peuvent se permettre les très gros acteurs (comme Facebook et Google, qui s'enorgueillissent d'une centaine de couples), quoique toujours dans certaines limites (on compte aujourd'hui quelques 6500 langues, dont environ 2500 sont décrites et seulement quelques centaines numérisées (transcription au standard Unicode). À l'époque où la mondialisation en tout anglais révèle ses limites, notamment par ses effets d'invisibilisation des écoles nationales, voire de constitution d'angles morts de la recherche, **la constitution d'une technologie d'emblée multilingue représente un investissement à la fois raisonnable et peu coûteux**.

Nous définissons le **multilingue** comme la possibilité d'écrire et traduire **directement** dans plusieurs langues. On permet ainsi **immédiatement** une multiplicité de regards (les différentes traductions) sur un même objet (le texte source). L'expression informatique de cette idée change la manière de construire une interface numérique, d'une représentation habituelle en 2 colonnes on peut passer à n colonnes.

On pourrait penser que, en multipliant les couples de langue, on obtient aussi un résultat multilingue. Mais ce n'est alors qu'un produit **indirect, coûteux** (gourmand en mémoire, en énergie, etc.), et très difficile à gérer (du fait de croissance exponentielle induite par l'adjonction de chaque nouvelle langue), à l'opposé de l'incrémentation "souple" que permet un système **d'emblée multilingue**. Concrètement, un traducteur doit donc pouvoir utiliser une interface présentant plus de deux fenêtres.

e_ IA VS travail de l'ombre ?

L'intelligence artificielle, dans le domaine de la traduction comme ailleurs, est une ingénierie des connaissances qui **consomme beaucoup d'énergie et de travail humain**, même si ce dernier est volontiers caché par la réputation d'automatisme des machines. Les réseaux de neurones doivent être entraînés, fût-ce de façon rapide et discrète (voire cachée), pour pouvoir fonctionner de façon correcte, et a fortiori s'ils visent une qualité optimale.

La plupart du temps, ce travail humain, situé en amont, en aval, et au milieu du processus algorithmique, est assumé, et constitue même un **nouveau corps de métier, en cours d'organisation** - c'est en ce sens que les étudiants en traduction sont désormais formés aux différentes formes de constitution et d'usage des intelligences artificielles.

Mais il arrive aussi que, pour des raisons commerciales, notamment, ce travail humain soit **invisibilisé** et surtout sous-traité et **sous-payé**. Comme le rappelle Thierry Poibeau, *«Il y a eu il y a quelques années la traduction automatique d'un livre entier, "Deep Learning"; annoncée comme étant réalisée en 12h; mais la traduction, au delà des expressions mathématiques était relativement simple, avec un dictionnaire pas si ambigu et, en pratique, il y a eu pas mal d'huile de coude pour sortir les formules et faire une passe derrière»*.

Mais le cas le plus célèbre est la plateforme de sous-traitance d'Amazon, le Mechanical Turk, où des milliers de petites mains préparent, nettoient ou rectifient les résultats des algorithmes du géant de la distribution, afin d'en améliorer les offres personnalisées. Les compétences humaines sont alors aspirées par la machine, au lieu d'être valorisées par elles.

INTRODUCTION

Compréhension de la mission et parti pris :

Quel est le contexte tel que nous le comprenons (les enjeux aujourd'hui en termes de diffusion de la recherche et de la traduction) ?

Quel positionnement pour le projet et quels ingrédients nous semblent a priori essentiels ?

Quel est le périmètre cible a priori du service ?

a_Compréhension de la mission et parti pris

Le but de la mission est de développer un **service d'intérêt général** pour que les différents acteurs concernés (traducteurs, éditeurs, auteurs, distributeurs) puissent diffuser la recherche scientifique dans différentes langues. En aidant à poser les bases d'un futur service de traduction scientifique, l'objectif est en effet de **soutenir la recherche dans la langue native des chercheurs**, et de permettre à ces derniers de ne pas être pénalisés sur la scène internationale, au motif qu'ils ne maîtrisent pas l'anglais ou toute autre langue "dominante". Comme le souligne Justine Fabre, Directrice du Patrimoine et des ressources scientifiques de l'Académie des sciences : *«un chercheur travaillera toujours mieux dans sa langue»*. Dans tous les cas, il est souhaitable que chaque chercheur soit libre de s'exprimer dans sa langue de prédilection (même si parfois, un chercheur peut être plus à l'aise dans une autre langue que sa langue native - selon sa formation, sa discipline, ses habitudes de rédaction...).

De façon symétrique, il semble également essentiel, dans cette logique de **diffusion des connaissances**, d'offrir aux chercheurs, aux acteurs de la recherche, aux organisations et à la société civile une plus grande **découvrabilité des productions scientifiques** réalisées à travers le monde dans des langues autres que l'anglais.

Le futur service devra donc permettre **la traduction et la diffusion** aussi bien que la **découvrabilité de productions scientifiques multilingues**.

Actuellement, l'équipe projet d'OPERAS a commencé à rassembler des corpora textuels disciplinaires qui doivent permettre d'entraîner un moteur de traduction neuronale, afin de proposer des suggestions terminologiques plus spécialisées et pertinentes. La mission vise, en complément, à penser les fonctionnalités d'un **service de traduction assistée par ordinateur TAO** ou **traduction outillée** qui proposerait l'assistance de ce moteur, avec **une entrée par discipline**. **Trois disciplines scientifiques** ont été choisies: une en sciences humaines et sociales (Human Mobility, Environment, and Space), une en sciences médicales (Neuroscience and Disorders of the Nervous System ou Immunity, Infection and Immunotherapy) et une en ingénierie et sciences de la Terre (Climatology and climate change) ; des corpora sont constitués dans les langues suivantes: français et anglais.

Positionnement du projet

Le futur service doit permettre aux productions de nos chercheurs d'avoir plus d'influence internationale, ainsi que de diffuser des connaissances issues de la recherche initialement publiée en anglais au bénéfice d'acteurs professionnels, d'entreprises, d'associations, de porteurs de politique publique ou de simples citoyens, en s'adressant à eux dans leur langue native grâce à la traduction. Cela passe:

- par un indexage multilingue des productions, via les **métadonnées** associées, et des **résumés**, pour lesquels les outils de traduction automatisée pourraient être utiles
- par des **fonctionnalités d'aide à la traduction** ou à la **rédaction multilingue**, pour des **chercheurs** ou **traducteurs** qui souhaitent traduire un texte, écrire directement dans une langue de destination ou relire une production: nous pensons au cas d'un chercheur non anglophone, qui rédige directement en anglais et qui souhaite revoir ou faire revoir son texte.

Le futur service pourrait donc proposer un **logiciel de traduction ou traduction-rédaction multilingue** ainsi que **ses extensions**, dans le navigateur ou proposé en plug-in aux éditeurs ou diffuseurs, qui permettraient la **traduction des métadonnées et des résumés**. Nous pourrions proposer une fonctionnalité dans **HAL** par exemple qui permettrait de consulter des résumés d'articles dans différentes langues.

Parti pris: un outil de TAO multilingue et un usage vigilant de la traduction automatique

Nous prenons le parti pris de travailler dès le départ dans un **contexte multilingue, plutôt que par couples de langues**, afin de poser les bases d'un service qui saura traiter la diversité des langues et langages sans devoir systématiquement passer par une langue pivot comme l'anglais, et qui pourra permettre de travailler de façon simultanée en plusieurs langues (par exemple, pour traduire du français au portugais, en s'aidant de mémoires de traduction français-espagnol). L'idée est de **prévoir le cas le plus complet** afin de poser les bases d'une interface numérique capable de traiter la diversité des situations de traduction et rédaction multilingue qui pourraient exister, y compris les plus complexes.

Les outils de traduction automatique peuvent parfois permettre de gagner du temps, surtout pour des textes assez descriptifs, ayant un contenu informatif - sans pour autant se bercer d'illusions, tant sont nécessaires les opérations de pré et post-édition. Structurellement, la machine n'est pas en capacité de se substituer au traducteur, de sorte que **l'humain doit toujours pouvoir garder la main** dans l'utilisation du futur service, notamment concernant l'assistance du moteur de traduction automatique. Il nous semble également essentiel, dans une logique de **traçabilité de la traduction automatique**, que des textes ou segments de textes qui auraient été traduits de façon automatique et non relus par un humain puissent bien être identifiés et affichés avec cette mention.

b_ Les objectifs de ce premier rapport

Approche méthodologique

ENTRETIENS FORMELS

Entretiens formels d'1h à 1h30 - dont une synthèse est disponible dans le présent rapport

- **3 entretiens avec 3 traducteurs** représentant chacun une discipline différente (SHS, climate change, médical)
- **1 entretien avec une équipe d'Erudit**, plateforme québécoise de diffusion dans le domaine des SHS
- **1 entretien avec une personne de l'Académie des Sciences**, organisme français d'expertise scientifique et de soutien à la recherche qui publie 7 revues de recherche de niveau international dans différentes disciplines de STM (Science Technologie & médecine).

SESSIONS DE TRAVAIL AVEC DES EXPERTS SCIENTIFIQUES

Direction scientifique :

- **Philippe Lacour**
Docteur et chercheur spécialiste de la philosophie du langage

Expertise scientifique :

- **Claire Larsonneur**
Maître de conférence et chercheuse, spécialisée en économie de la traduction, traduction et numérique
- **Thierry Poibeau**
Directeur de recherche au CNRS et directeur adjoint du laboratoire LATTICE

ENTRETIENS INFORMELS

Entretiens informels auprès de chercheurs ou enseignants-chercheurs aux disciplines variées (psychologie comportementale, mathématiques appliquées à la physique, politiques publiques, traductologie), notamment :

- **Franck Barbin**
Maître de conférences en traduction anglaise à l'université Rennes 2
- **Stéphane Pouyllau**
Ingénieur de recherche au CNRS, responsable de l'Huma-Num lab

PRÉSENTATION D'OUTILS

- Démonstration de Trados, commentée, par un traducteur.

Ce premier rapport est un état de l'art qui vise avant tout à comprendre les pratiques actuelles et les enjeux de la traduction scientifique. D'un point de vue méthodologique, nous avons pour cela mené des **entretiens formels et informels** avec des traducteurs, des chercheurs, des éditeurs ou diffuseurs, analysé **des outils de traduction assistée par ordinateur** afin d'identifier les fonctionnalités et structures types et notamment suivi une démonstration poussée de l'outil Trados, réalisé des sessions de travail avec le **directeur scientifique Philippe Lacour**, et deux chercheurs ayant une expertise spécifique et reconnue dans le domaine de la traduction automatique: Claire Larsonneur et Thierry Poibeau, afin de bénéficier d'un éclairage conceptuel rigoureux, fondé scientifiquement.

Priorité: comprendre l'acte et le parcours de traduction

Touchant les entretiens, le choix opéré a consisté à privilégier, dans un premier temps, les traducteurs, afin de bien comprendre leurs différents gestes et opérations professionnels. L'exploration des pratiques de la recherche a certes été abordée, mais sera complétée plus tard, notamment lors du rapport 2, grâce aux résultats de l'atelier avec les utilisateurs, auxquels chercheurs, traducteurs et éditeurs sont conviés.

c_ Synthèse

Principaux enseignements des entretiens, analyse d'outils et session de travail avec les experts scientifiques

Un parcours et des outils types pour la traduction de publication scientifique faite par des professionnels

Il convient de souligner que les professionnels de la traduction interrogés pour ce rapport utilisent des outils de TAO. À noter qu'un certain nombre de traducteurs spécialisés en sciences humaines et sociales contactés dans le cadre du projet Traductions et science ouverte n'utilisent pas de tels outils (certains traduisent sur Word ou même encore sur papier).

Ensuite, malgré une diversité des pratiques dont font état les différents traducteurs professionnels indépendants, nous avons pu identifier qu'ils observaient **un parcours type de traduction** et que les outils de TAO qu'ils utilisent présentaient des structures assez similaires (vue synoptique, découpage en segments, glossaires et mémoires de traduction, suggestions de correspondances par un moteur de traduction automatique, des outils pour la révision de la traduction).

Deux remarques préalables s'imposent :

- Ces outils fonctionnent pour la plupart par "couple de langues" (interface à deux colonnes) ; nous envisageons de notre côté de proposer un espace de travail multilingue, avec par exemple une vue synoptique en N colonnes, permettant de voir et travailler plusieurs langues simultanément.
- Pour beaucoup de traducteurs professionnels, les suggestions de la machine "font écran" et il est ensuite difficile de s'en détacher. Nous souhaitons donc travailler à des principes ou astuces permettant de faciliter la prise de recul et la traçabilité de la traduction automatique (tout en ayant conscience que les usagers voudront in fine probablement "cacher" les parties ayant été proposées par la machine).

Il nous faut cependant prendre en compte **d'autres pratiques autour de la traduction et rédaction de textes scientifiques et leur diffusion**, notamment:

- **la rédaction directement dans une langue cible par un chercheur dont ce n'est pas la langue maternelle.** Le cas qui nous a été signalé à plusieurs reprises est celui d'un chercheur non anglophone qui décrit directement en anglais et qui va - ou non - se faire relire par un traducteur professionnel ou un traducteur "de circonstance" (le docteur irlandais de son laboratoire, par exemple).
- les usages liés à **la découvrabilité** des productions scientifiques, que nous mentionnions plus haut, et qui vont permettre de prendre connaissance du contenu d'une publication (au moins via la traduction de son résumé) ou même découvrir l'existence (notamment via l'indexation et la recherche par mots clefs).

Le périmètre du service visé reste encore relativement ouvert à ce jour. Il semble toutefois d'emblée inclure :

1. **la traduction des métadonnées et des résumés** avec l'aide de la traduction automatique pour augmenter l'impact et la découvrabilité des publications scientifiques;
2. **l'aide à la traduction (ou à la rédaction dans une langue cible) de productions scientifiques**, grâce notamment à l'assistance d'un moteur de traduction automatique qui a été entraîné spécifiquement, en fonction des différentes disciplines scientifiques;

La question des fonctionnalités d'aide à la traduction-rédaction reste posée. L'outil doit-il permettre de **rédiger directement dans une langue cible sans passer par l'étape préalable de rédaction dans sa langue** ? Propose-t-on des **aides à la rédaction académiques** qui seraient spécifiques à la recherche, selon les domaines ou les disciplines, comme des conventions sur la typographie (guillemets anglais, allemands, français), le chargement des guides de styles proposés par les différentes revues voire des templates de rédactions où il n'y aurait plus que des "blancs" à remplir ?

La question de **la nature des usagers prioritaires du service reste également ouverte**. Nous avons le sentiment que l'utilisateur cible est le chercheur qui doit produire, dans sa pratique quotidienne, des textes dans une langue qui n'est pas sa langue maternelle (pour un colloque, dans le cadre d'une pré-publication), usage pour lequel le niveau de langue n'est pas forcément discriminant, ou qui souhaite collaborer avec un traducteur, dans le cas où le niveau d'anglais requis nécessite l'intervention d'un professionnel. Nous pensons que les **traducteurs professionnels** ont déjà leurs outils dédiés, et ne seraient pas forcément intéressés par un usage direct du service, mais plutôt par la possibilité d'intégrer les moteurs de traductions entraînés par discipline à leurs outils actuels, ou encore d'accéder à des corpus de mémoires de traduction spécialisés par discipline ou domaine scientifique.

Enfin, encore une fois, nous croyons aux usages autour de la **découvrabilité** et notamment la traduction des métadonnées et des résumés qui pourrait se faire de façon automatique.

Ce rapport 1 sera augmenté par le rapport 2 (qui intégrera les ateliers) et ira plus loin dans les recommandations.

ÉTAT DE L'ART

Définitions et objets

La traduction scientifique, cartographie des pratiques et enjeux

Analyse d'outils

01

DÉFINITIONS ET OBJETS

Les principales définitions et objets que l'on retrouvera dans la future interface du service de Traduction Outillée (ou TAO, Traduction Assistée par Ordinateur).

Aligner

On parle d'aligner des segments, lorsqu'on part d'une unité de traduction (deux textes dont l'un est la traduction de l'autre) et qu'on la découpe en segments (la plupart du temps, des phrases), parallèles (la plupart du temps) ou non (dans de rares cas).

Base terminologique

La base terminologique est une organisation du glossaire en base de données exploitable. Elle permet d'effectuer des recherches d'occurrences d'un mot ou de proposer des correspondances terminologiques, en général multilingues.

Corpus (corpus parallèle et corpus comparable / corpus monolingue, bilingue, multilingue)

Un corpus est un ensemble de textes permettant d'identifier le style et la terminologie du texte cible. Contrairement à une conception volontiers répandue en informatique, un corpus n'est pas un sac de mots, mais un « regroupement structuré de textes intégraux, documentés, éventuellement enrichis par des étiquetages, et rassemblés : (i) de manière théorique réflexive en tenant compte des discours et des genres, et (ii) de manière pratique en vue d'une gamme d'applications » (F. Rastier). Il n'y a pas de corpus en soi, sans question qui lui préexiste et guidant sa constitution - les prétendus corpus "neutres" escamotent en fait le problème de leur constitution. Il existe des corpus parallèles qui rassemblent des textes (en général alignés par phrase) dans la langue source et leur traduction dans la langue cible (ex. les résolutions européennes qui doivent être traduites dans les langues des différentes nations membres), des corpus comparables (ex. Wikipedia où la page français ne correspond pas à la page anglaise). On peut également caractériser les corpus selon leur caractère monolingue*, bilingue voire multilingue*.

* un corpus multilingue peut s'avérer utile s'il existe moins de ressources dans la langue cible que dans une autre langue proche qui peut servir de langue intermédiaire. Ainsi on pourra traduire du français au portugais en passant par l'espagnol et se constituer un corpus potentiellement trilingue

Glossaire et base terminologique

Le glossaire définit les termes clés qui vont être utiles à la rédaction ou la traduction d'un texte en fonction de son objet et de son contexte de diffusion. Pour les disciplines scientifiques ou des sujets de recherche particuliers, le glossaire définit les termes conceptuels et techniques nécessaires à l'exactitude des propos et raisonnements en particulier le vocabulaire spécifique au domaine d'étude.

Un glossaire peut être monolingue ou multilingue.

Lexique

Le lexique, en traduction, désigne un ensemble de mots et leur correspondance dans une ou d'autres langues.

Mémoire de traduction (MT)

La mémoire de traduction est la mémoire de textes ou morceaux de textes (la plupart du temps, des phrases) et leurs traductions correspondantes. On désignera par unité de traduction ou (UT) un texte et son texte traduit. Il est possible de partir de zéro, et dans ce cas le mémoire de traduction se constitue au fur et à mesure de la traduction, ou alors d'utiliser des mémoires de traduction déjà existantes. La mémoire de traduction permet de proposer des correspondances qui ont déjà pu être faites au préalable. D'où leur contribution « *au processus de localisation, en améliorant considérablement la qualité, la rapidité, la cohérence et l'efficacité de chaque projet de traduction.* », comme le signale l'outil de TAO Trados studio, utilisé par les traducteurs professionnels.

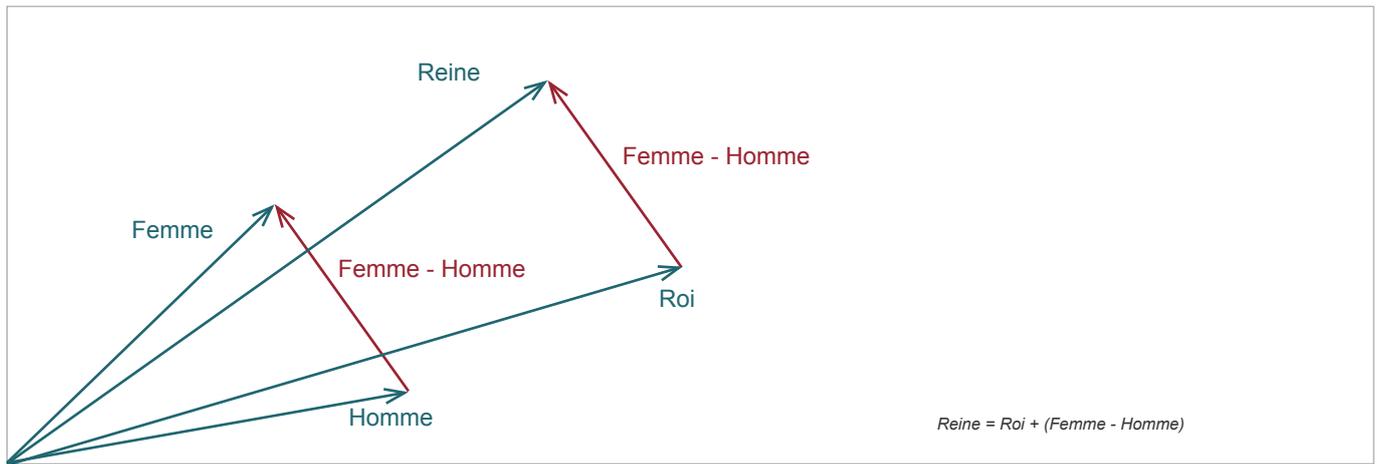
Moteur de traduction neuronale (MTN ou NMT)

Les moteurs de traduction neuronale utilisent l'intelligence artificielle et les réseaux de neurones pour proposer, au fil de leurs entraînements, des traductions de plus en plus adaptées. L'approche des réseaux de neurones diffère des approches statistiques et par règles syntaxiques en permettant d'identifier des motifs de correspondances à partir de corpus structurés, sans nécessiter au préalable l'écriture de ses règles statistiques de façon explicite. Leur fonctionnement reste fondé sur le calcul probabiliste, comme dans l'approche statistique ; ils n'ont donc pas de règles explicites et directement lisibles ("effet boîte noire").

Ces moteurs de traduction utilisent le concept de vecteur de mots ou ensemble de mots qui permet de décrire un mot à l'aide de variables (généralement 300 à 500). Cet encodage sous forme de vecteurs permet de les manipuler de manière extrêmement efficace, pour les comparer, pour calculer le sens d'une expression complexe, etc.

Pour aller plus loin au sujet de la notion de vecteurs <https://www.shanelynn.ie/get-busy-with-word-embeddings-in-troduction/>

→
Suite de la définition



Paramètres	Vecteur «Roi»	Vecteur «Reine»	Vecteur «Princesse»
Humain	$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$
Masculin			
Royal			
Adulte			

Exemple de vecteurs à 4 paramètres

▲
 Illustration du principe des vecteurs ; une Reine peut être décrite comme étant un “Roi Femme”, donc le vecteur “Roi” auquel on ajoute le vecteur “Femme - Homme”. Ce principe peut être généralisé à n dimensions.

Phraséologie / style

La phraséologie est le registre de langage, le style parlé ou écrit dépendant d'un contexte de diffusion (par époque, auteur, média, public cible,...). Concernant le domaine de la production scientifique, le texte pourra être plus ou moins narratif, avec un ton plus ou moins neutre et respectera certains formalismes, notamment orthographiques et typographiques.

Post-édition

La post-édition en traduction, désigne les opérations de re-traduction et révision qui ont lieu après l'acte de traduction humain ou automatisé. Il désigne généralement le travail effectué sur des textes ayant été traduits par des moteurs de traduction automatisée. Ce travail de nettoyage et de finition est de plus en plus enseigné dans les cours de traduction, notamment pour la traduction technique.

Prétraduction automatique

Le prétraduction automatique est un “premier jet effectué par une machine”, destiné à être relu et corrigé par un humain. (<https://journals.openedition.org/traduire/460>)

Révision de la traduction

Les opérations de révision visent à assurer la qualité de la traduction, tant du point de vue du sens que du point de vue des standards de style (le registre de langage, le vocabulaire type, les règles orthographiques et typographiques,...). Il s'agit par exemple de s'assurer de la cohérence à l'échelle du document (cohérence de point de vue, de style, de correspondance terminologique, d'usages typographiques,...).

Segment

Découpage permettant d'obtenir des morceaux de textes à traduire, et d'éviter de manipuler ainsi le texte dans son entier (lorsqu'on cherche un terme dans une concordance, par exemple). La segmentation la plus courante est le découpage du texte en phrases (en informatique : ce qui commence par une lettre majuscule et termine par un point) mais on trouve régulièrement des segmentations par paragraphe (notamment si les textes sont plus narratifs), voire par strophes, tercets ou quatrains (en poésie). On peut également imaginer des segmentations plus fines: par proposition, par mot (voire par préfixe ou suffixe,...). Dans le cas d'une traduction organisée en vue synoptique (alignement de différentes colonnes sur une même page), les segments du ou des texte(s) cible(s) et source(s) peuvent correspondre terme à terme (phrase à phrase, dans la plupart des outils de TAO) mais cela n'est pas obligatoire (dans le cas de TraduXio).

Texte source / texte cible (plutôt que langue source et langue cible)

Le texte source est le texte à traduire tandis que le texte cible est le texte en sortie de l'acte de traduction. On parlera de texte source et de texte cible plutôt que de langue source ou de langue cible, parce que l'idée que la traduction consiste à faire passer un message (codé) d'une langue A à une langue B est une simplification d'ingénierie, issue des premiers travaux sur la traduction automatique s'inspirant de la machine de Turing. Au contraire, concevoir la traduction comme consistant à trouver l'équivalent (sans identité) d'un texte singulier permet de la penser en direction d'un grand nombre de textes singuliers dans diverses langues. Par quoi on insiste sur le fait qu'il existe des singularités (et des variations) dans la langue (ce que l'approche par les règles a tendance à oublier), et sur la dimension immédiatement multilingue du travail de traduction. Remarque: il est possible de faire des traductions au sein d'une même langue. Par exemple, il est possible de partir d'un texte source rédigé en anglais par un non anglophone pour produire un texte cible avec la phraséologie et les règles convenant à une publication académique. Il est également possible d'être dans la situation de vouloir juste mettre à jour une ancienne traduction, voire de la refaire entièrement, puisque les langues (source et cible) évoluent dans le temps (on ne cesse de retraduire la Bible et Shakespeare).

Traduction automatique

La traduction automatique désigne la traduction brute d'un texte entièrement réalisée par un ou plusieurs programmes informatiques. On la distingue de la traduction assistée par ordinateur où la traduction est en partie manuelle, éventuellement de façon interactive avec la machine.

La traduction automatique existe depuis les années 50. Elle peut procéder par analyse syntaxique (à partir de règles, par analyse statistique ("ce terme français est généralement traduit par ce terme en espagnol avec telle récurrence), ou encore par un mixte des deux - les moteurs de traduction comme Systran ou Google translate reposent sur une hybridation de ces approches.

src. Wikipedia, page française de la "Traduction automatique", le 09-01-2023

Traduction outillée ou TAO

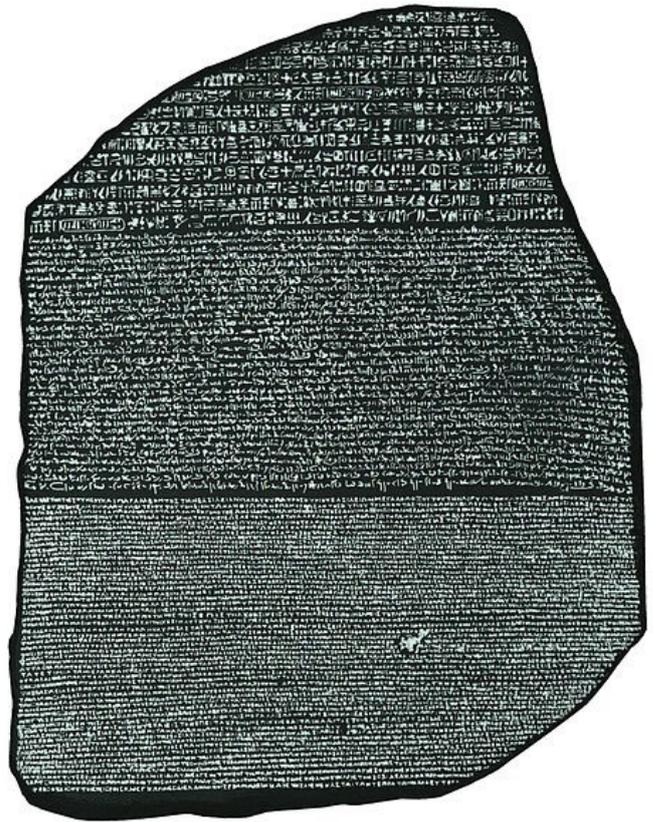
On parlera de traduction outillée ou de traduction assistée par ordinateur pour désigner une traduction qui a pu être facilitée par l'utilisation de logiciels. Un logiciel de TAO permet de faciliter et d'assister tout ou partie des opérations de traduction. En général on retrouve des fonctionnalités liées aux mémoires de traduction et glossaires, une vue synoptique permettant d'afficher de façon simultanée le texte source à traduire et le texte cible dans la langue de destination et des fonctions de relecture et contrôle qualité de la traduction.

La traduction outillée ou TAO peut faire appel à des outils de traduction automatisée.

Vue synoptique

La vue synoptique est une vue d'ensemble. En traduction, elle permet d'afficher sur une même page ou support les différentes versions d'un texte côte-à-côte. La présentation la plus courante est en colonnes verticales (même si un des traducteurs interrogés préfère un affichage dessus-dessous).

Verbatim: «*Même si je n'utilise pas la traduction automatique, je vais dans mon logiciel de TAO pour la vue en 2 colonnes*» (une traductrice). La vue synoptique est profondément liée dans l'histoire à l'acte de traduction et de diffusion multilingue.



▲
Pierre de rosette

TRADUCTION AUTOMATIQUE :

DISCUSSION AVEC

THIERRY POIBEAU

En quoi les réseaux de neurones changent-ils la donne par rapport aux précédents modèles probabilistes et fonctionnant par règles ?

Les systèmes à base de règles fonctionnaient avec des dictionnaires et des règles de transfert d'une langue à l'autre. Or, la traduction d'un mot peut être multiple et dépend du contexte, mais la notion de contexte n'est pas quelque chose que l'on sait bien modéliser par des règles (le contexte est une notion évanescence qui ne répond pas à une définition claire). C'est cette difficulté à modéliser la notion de contexte qui peut expliquer, entre autres, l'échec des systèmes à base de règles. Plus prosaïquement, traduire d'une langue à l'autre demande aussi des milliers de règles et les systèmes ainsi conçus ont toujours (et rapidement) été confrontés à d'immenses ensembles de règles, souvent ingérables, parfois contradictoires, etc.

Les systèmes statistiques sont popularisés dans les années 1990. Ils ont rapidement été plus performants que les systèmes à base de règles de par leur capacité à modéliser le contexte (c'est-à-dire, tout simplement, en tenant compte des n mots avant, et éventuellement après, un mot donné), ce qui permet de résoudre en partie la question de la traduction contextuelle. Ces systèmes étaient ainsi relativement robustes pour traduire mot à mot, et, au-delà, par segment (groupes de 2, 3, 4 mots...) mais ils ne pouvaient fournir directement une modélisation de la phrase entière. Pour obtenir la traduction d'une phrase, il fallait donc élaborer une phrase à partir de fragments de traduction parfois disparates, ce qui pouvait aboutir à des phrases bancales et/ou à une perte de sens lors de la traduction.

La traduction neuronale est fondée sur une modélisation de la phrase entière, ce qui a permis de résoudre la plupart des difficultés précédentes. On notera que, comme la traduction se fait encore le plus souvent au niveau de la phrase, les pronoms et autres phénomènes impliquant plus d'une phrase peuvent rester problématiques (par exemple, choix du pronom "le" vs "la" si l'antécédent est dans la phrase qui précède).

Quelles sont les grandes préventions que l'on peut avoir sur la traduction automatique ?

Les traductions offertes sont souvent de bonne qualité, mais elles dépendent quand même étroitement des langues considérées (la traduction est meilleure si la langue source ou la langue cible est l'anglais) et des corpus utilisés pour l'entraînement (cad pour la mise au point du système). La traduction automatique n'est pas toujours fiable pour des textes techniques, ni désirable pour des textes littéraires (le processus peut rendre le sens mais de façon très "plate", et ôter ainsi tout le charme d'une œuvre littéraire).

Du point de vue de l'usage, les systèmes de traitement des langues produisent souvent un texte "parfait" sur le plan syntaxique et sémantique. Il peut toutefois s'agir d'une mauvaise traduction (d'une traduction avec un contresens, ou avec un mot omis, etc). Or, il a été montré que des traducteurs, même chevronnés, avaient naturellement tendance à faire confiance à la machine et à ne pas détecter des erreurs de traduction si la phrase produite avait une apparence correcte dans la langue cible (c'est un type de biais d'automatisation).

Comment se passe la spécialisation des modèles ? Avez-vous des recommandations particulières à faire aux équipes d'Operas (qui sont en train de collecter des corpus disciplinaires dans la perspective d'entraîner un moteur à partir de ces corpus) ?

La spécialisation d'un système se fait toujours soit en entraînant le système sur un jeu de données représentatif, soit en adaptant un modèle générique sur un ensemble plus réduit de données représentatives (ce que l'on appelle en anglais le fine tuning).

Les recommandations que l'on peut faire sont classiques : s'assurer au maximum que les données d'entraînement (de mise au point du système) correspondent aux données à traduire. Mais c'est évidemment plus facile à dire qu'à faire. A mon avis, il ne faut pas réduire cela à un problème technologique. Il est aussi primordial de documenter le système : dire avec quelles données il a été entraîné, à quel type de traitement (quel domaine, quel type de texte, etc.) il s'adresse, etc.

La mise en œuvre de la traduction automatique n'est pas un problème technique (ou en tout cas pas seulement) : le plus important est - à mon avis - de bien mettre en avant l'intérêt et les limites de la technologie, les conditions et recommandation pour son usage, etc.

Qu'est ce qu'un modèle de langage (GPT3) ? Quels sont les impacts actuels ou à venir sur la traduction ?

Un modèle de langage est une ressource généralement élaborée à partir de milliards d'exemples (de phrases, de textes - au moins quand les corpus existent. Plus il y a de données, plus le système sera performant : c'est pourquoi les systèmes pour l'anglais sont meilleurs que ceux pour l'italien, qui sont eux-mêmes meilleurs que ceux pour l'estonien - par exemple ! La complexité de la langue joue aussi un rôle ; à ce titre, l'anglais est une langue très spécifique, ce qui lui donne un deuxième avantage : l'anglais a une morphologie très pauvre - les mots varient très peu et les informations sont encodées par de petits mots indépendants : articles, prépositions, etc.).

Cette ressource va emmagasiner des milliards de "connaissances" sous forme de valeurs (entre 0 et 1), que l'on appelle des paramètres. Ces connaissances permettent d'associer des profils sur les mots et leurs contextes d'apparition. Il s'agit donc d'objets abstraits (on peut dire pour donner une intuition que chaque valeur encodée correspondrait à une primitive sémantique) à partir desquels on peut effectuer des tâches très concrètes : génération de texte, analyse syntaxique, traduction automatique, etc.

Les modèles BERT sont plutôt tournés vers l'analyse (analyse syntaxique, reconnaissance d'entités nommées, analyse de sentiments). Les modèles GPT permettent de générer du texte (produire un texte à partir d'une amorce appelée prompt). Les modèles de type NMT sont eux tournés spécifiquement vers la traduction. ChatGPT, le dernier modèle d'OpenAI est avant tout tourné vers le dialogue (répondre à des questions), mais inclut aussi un modèle de traduction. Il existe enfin des modèles multimodaux (pour générer des images à partir de textes). Le point commun de tous ces modèles est de partager une même approche, ce que l'on appelle l'apprentissage profond (le "deep learning").

02

LA TRADUCTION SCIENTIFIQUE, CARTOGRAPHIE DES PRATIQUES ET ENJEUX

*Issu des entretiens, des ateliers avec l'équipe projet et experts,
de la veille documentaire*

a_ Les acteurs de la traduction scientifique

A l'aide des interviews réalisées, nous avons établi une première cartographie des acteurs en lien avec la traduction scientifique. A partir de cette cartographie des acteurs de la traduction, nous pouvons déjà tirer quelques enseignements utiles pour la conception à venir de notre service :

- La multiplicité et diversité des destinataires au sein de l'univers de la recherche (avec plusieurs disciplines, pratiques) mais aussi dans la société civile (pour laquelle la recherche doit pouvoir in fine être utile) avec chacun un rapport différent à la traduction (plus ou moins exigeant, mais toujours avec une volonté d'exactitude)
- Des pratiques de la traduction relativement morcelées : en agence, en freelance, parfois intégrée aux revues, ou bien encore des pratiques amateurs de traduction lorsque la production scientifique concernée et son contexte de diffusion le permet
- Au delà d'être intégré à une agence ou non, d'être professionnel ou amateur, il est important de noter une dimension disciplinaire de la traduction ; en effet, selon le type de traduction souhaitée (notamment selon le contexte de diffusion ou la discipline traduite) des traducteurs experts disciplinaires vont être sollicités (qu'ils soient traducteurs amateurs ou professionnels)
- Des postures différentes jouées alternativement par les traducteurs qu'ils soient professionnels ou amateurs : la posture de traducteur mais aussi celle de réviseur qui corrige
- Des acteurs logiciels historiques de la traduction qui perdurent (Trados) mais aussi des plateformes récentes (MateCat) d'une ergonomie apparente plus poussée et moderne.
- Une pratique de traduction qui recourt beaucoup aux indépendants (même si le contrat "initial" de traduction se fait avec une agence, des indépendants interviennent souvent de façon discrète).
- Un recours à l'humain encore systématique (cf. l'existence des acteurs de la post édition, qui démontre à elle seule certaines limites de la traduction automatique)
- Des acteurs de nouvelles technologies de la traduction encore relativement récents (DeepL est lancé le 28 août 2017 par l'équipe de Linguee), adoptées très rapidement par une partie significative des acteurs de la recherche.
- Des acteurs spécialisés par branche commencent ainsi à être formés, des étudiants spécialistes de l'utilisation des intelligences artificielles destinées à la traduction, des outils de TAO ou de navigation sur corpus. Mentionnons également l'existence de la jeune Association française de traductrices et traducteurs de sciences humaines et sociales, l'ATESS, <https://www.atesfrance.org/a-propos>. Dans le monde anglo-saxon, certains acteurs se sont spécialisés dans la traduction universitaire, comme par exemple les prestataires anglophones de Cairn.info ou lavidesidees.fr

→
Cartographie des acteurs

CARTOGRAPHIE

DES ACTEURS - VUE D'ENSEMBLE (1/2)

Les acteurs de la traduction

Les traducteurs

TRADUCTEURS / REVISEURS

LES AGENCES DE TRADUCTION

Les agences «traditionnelles»

Ces agences peuvent travailler avec leur propres traducteurs mais ont aussi souvent des "pools de freelances" avec lesquels ils sont habitués à travailler, soit pour disposer de personnes plus spécialisées sur un sujet, soit pour gérer la charge de travail.

Les agences intégrées chez les éditeurs

Les revues scientifiques qui proposent elles-mêmes leurs services de traduction spécialisée par discipline ou sujet. Ex. La revue de référence Springer Nature propose un service de traduction spécialisée par discipline, ce qui pose notamment la question de la captation des mémoires de traduction par une revue déjà dominante dans le paysage scientifique.

Les agences de «post-édition»

Ces agences associent des outils de traduction automatique, en général pour augmenter la productivité, puis passent les traductions automatiques à des humains (internes à l'agence ou indépendants) pour relecture et correction. La post-édition permet donc d'accélérer le rythme de traduction mais démontre bien la nécessité de recourir à l'humain pour corriger les erreurs.

exemple:

La performance de la traduction automatique en 12h de l'ouvrage de 800 pages "Deep Learning" par l'expert Quantmetry (avec la startup allemande DeepL), a finalement requis l'intervention de plusieurs relecteurs experts (chercheurs issus de l'ENSAI, de l'INRIA et du CNRS) et deux mois et demi de travail au total.

LES TRADUCTEURS INDÉPENDANTS

Missionnés directement par un client ou par une agence de traduction

LES TRADUCTEURS AMATEURS

Les traducteurs amateurs sont généralement les doctorants ou les relations personnelles des chercheurs qui maîtrisent la langue de destination

LES POSTURES

TRADUCTEURS / REVISEURS

Chaque traducteur peut adopter les postures de traduction et de révision, c'est-à-dire que les traducteurs peuvent fonctionner en binôme: un premier qui fait véritablement la traduction et un autre qui se charge de la relecture et des opérations de révisions. On trouve également le cas où la révision n'est pas faite par un traducteur mais plutôt par un expert maîtrisant la langue de destination et surtout le domaine scientifique ou technique. Il s'agit alors de corriger non pas le style mais surtout le bon emploi de la terminologie scientifique ou technique.

LES FOURNISSEURS D'OUTILS

De nouveaux outils, non exclusifs aux traducteurs

Les fournisseurs d'outils «métiers»

- Trados
- Mate Cat

Les fournisseurs de TA

- DeepL
- Google Traduction
- ModernMT

Les fournisseurs de modèles de langage

- GPT-3
- BERT

Les fournisseurs d'outils de description de contenus

- Open Calais (Reuters)

LES CHERCHEURS EN TECHNOLOGIE DU LANGAGE (DONT TAL)

Les auteurs - lecteurs

LES CHERCHEURS

Parmi les chercheurs, il faut distinguer les deux postures d'auteur et de réviseur. En effet, les chercheurs peuvent produire du contenu ou participer à leur évaluation au sein de structures éditoriales.

Les personnes

- Chercheurs
- Doctorants
- Stagiaires

Les institutions

- Laboratoires

LES AUTRES DESTINATAIRES

Les destinataires «professionnels»

- Les étudiants
- Les enseignants
- Les experts (médecins, avocats,...) ou les professionnels (bureaux d'études, architectes et paysagistes,...)

La «société civile»

- Le grand public
- Les vulgarisateurs scientifiques ou médiateurs
- Les journalistes et media

CARTOGRAPHIE DES ACTEURS - VUE D'ENSEMBLE (2/2)

Les acteurs de la diffusion

LES ÉDITEURS SCIENTIFIQUES

Il sont chargés du processus éditorial, qui inclut en général une révision par les pairs. Il peut s'agir de revues qui publient des articles mais aussi de maisons d'édition qui vont éditer des ouvrages. L'édition peut-être numérique, papier ou les deux. Les éditeurs ont différents modèles économiques, certains proposent aux auteurs et lecteurs de payer pour publier ou accéder à leurs contenus, d'autres proposent la gratuité aux lecteurs mais pas aux auteurs, tandis que d'autres proposent la gratuité totale.

Quelques exemples

- Académie des sciences
- Elsevier
- The Lancet
- Springer Nature
- Wiley
- Antropologica (revues dispo sur Erudit)
- PUF (Presse Universitaires de France)
- Cambridge University Press

LES PLATEFORMES DE DÉPÔT ET DIFFUSION

Leur métier est la diffusion de contenu scientifique auprès des publics, en utilisant les possibilités offertes par le numérique. Cette mission se décline en plusieurs axes, notamment :

- de veiller à la qualité des contenus hébergés avec des opérations de vérification avant d'accueillir une nouvelle revue, ou avant le dépôt d'un contenu;
- de faciliter la «découvrabilité» des contenus par des bonnes pratiques en termes d'indexation via des métadonnées et propositions systématiques de résumés (notices) ; il s'agit d'un travail de structuration;
- de rendre accessible le contenu en lui-même, via des modèles économiques reposant sur la gratuité et l'accès libre;
- d'archiver les contenus ou les notices - description des contenus - de façon pérenne.

Ces plateformes ne sont pas «éditeurs» et ne sont donc pas responsables des opérations propres à l'édition (Érudit: "Nous intervenons quand le processus éditorial entre la revue et le ou les chercheurs est terminé). Elles ne se chargent pas de la révision par les pairs par exemple. Ces plateformes peuvent être disciplinaires (ArXiv, Erudit, HAL SHS), institutionnelles (portails HAL) ou nationales (HAL).

Archives ouvertes

- Hal
Archive ouverte française
- arXiv
Archive ouverte de pré-publications

Les plateformes de diffusion de revues et livres scientifiques

- OpenEdition
Portail de revues en SHS
- Erudit
Portail francophone canadien de revues SHS
- Cairn
Portail de revues françaises en SHS
- Persée
Portail diffusant en accès ouvert des collections patrimoniales de publications scientifiques françaises en sciences humaines et sociales et sciences de la Terre et de l'environnement.
- Centre Mersenne
Plateforme de diffusion de revues scientifiques.
- ...

Les fonctions supports de la recherche

INFRASTRUCTURE DE SOUTIEN À LA RECHERCHE

- HumaNum
- Operas
- MSH (Maison des Sciences de l'Homme)
- ...

RESSOURCES DIVERSES D'AIDE AUX CHERCHEURS

- Optimice
guide en libre accès pour traduire en anglais les métadonnées des articles scientifiques.
- ...

LES BIBLIOTHÈQUES UNIVERSITAIRES

Les pouvoirs publics

- Ministère de l'Enseignement Supérieur et de la Recherche
- Institutions européennes
- ...

b_ Les différents usages et pratiques de la traduction scientifique (les étapes du parcours de traduction, les données manipulées, les outils utilisés,...) et le(s) parcours type(s) de traduction

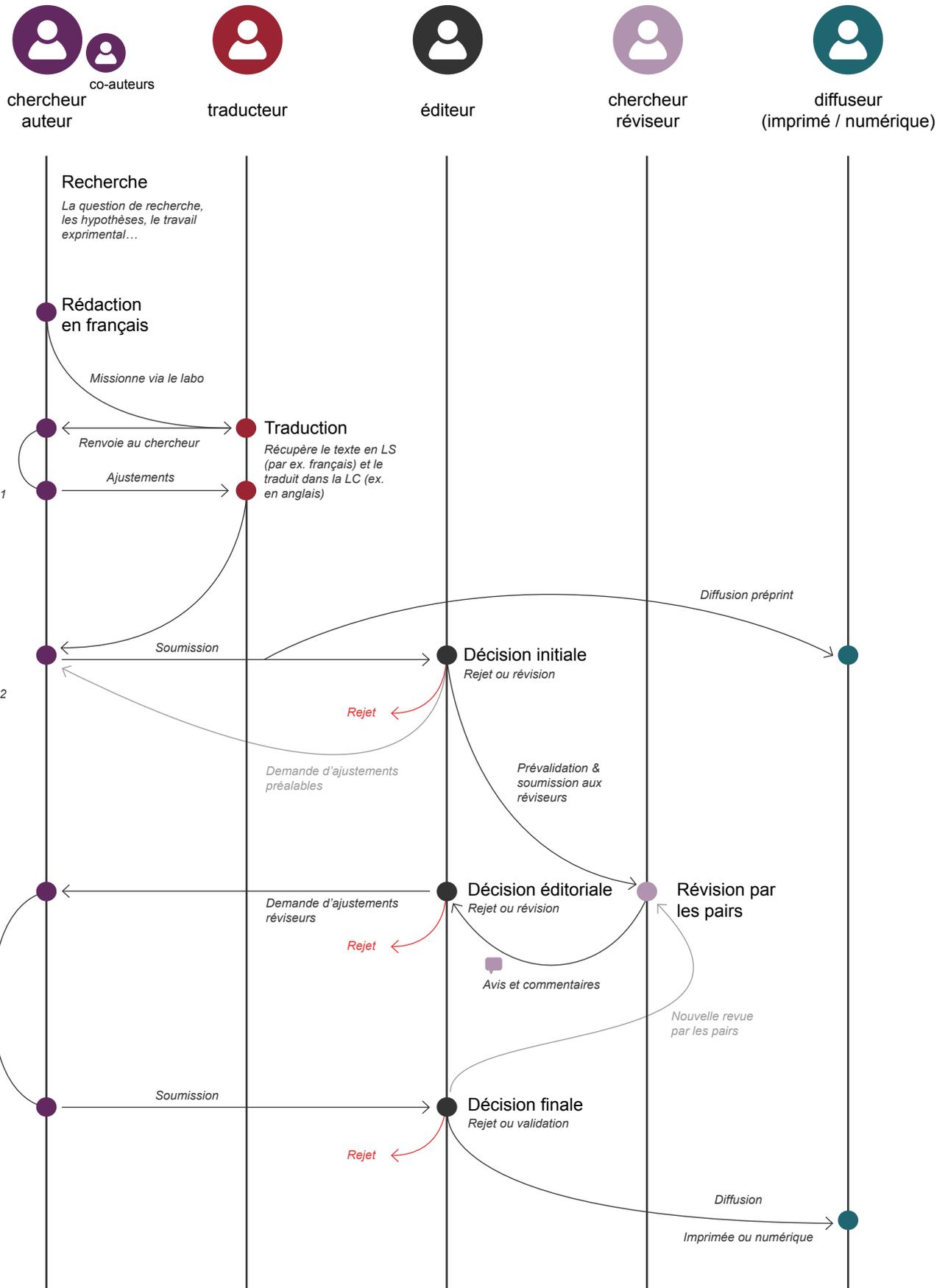
Si le parcours type de traduction professionnelle nous apparaît, à la suite de cette première étude, relativement standard concernant l'acte de traduction à proprement parler, les pratiques de la traduction et rédaction scientifique présentent un spectre beaucoup plus large qui plaide pour une future **plasticité** du service et du logiciel de traduction.

Au-delà du parcours type de traduction, il faudra notamment distinguer les différents **niveaux de langue** qui sont tolérés en fonction des contextes de diffusion, avec le recours à des traducteurs "amateurs" et de circonstance, ou des pratiques de rédaction-relecture uniquement dans la langue cible (plutôt que de traduction).

*les étapes et le parcours global pour un article scientifique et où intervient la traduction ;
le cas type d'une traduction par un traducteur en amont du dépôt à l'éditeur.*

La traduction d'un article scientifique au sein du parcours global de publication

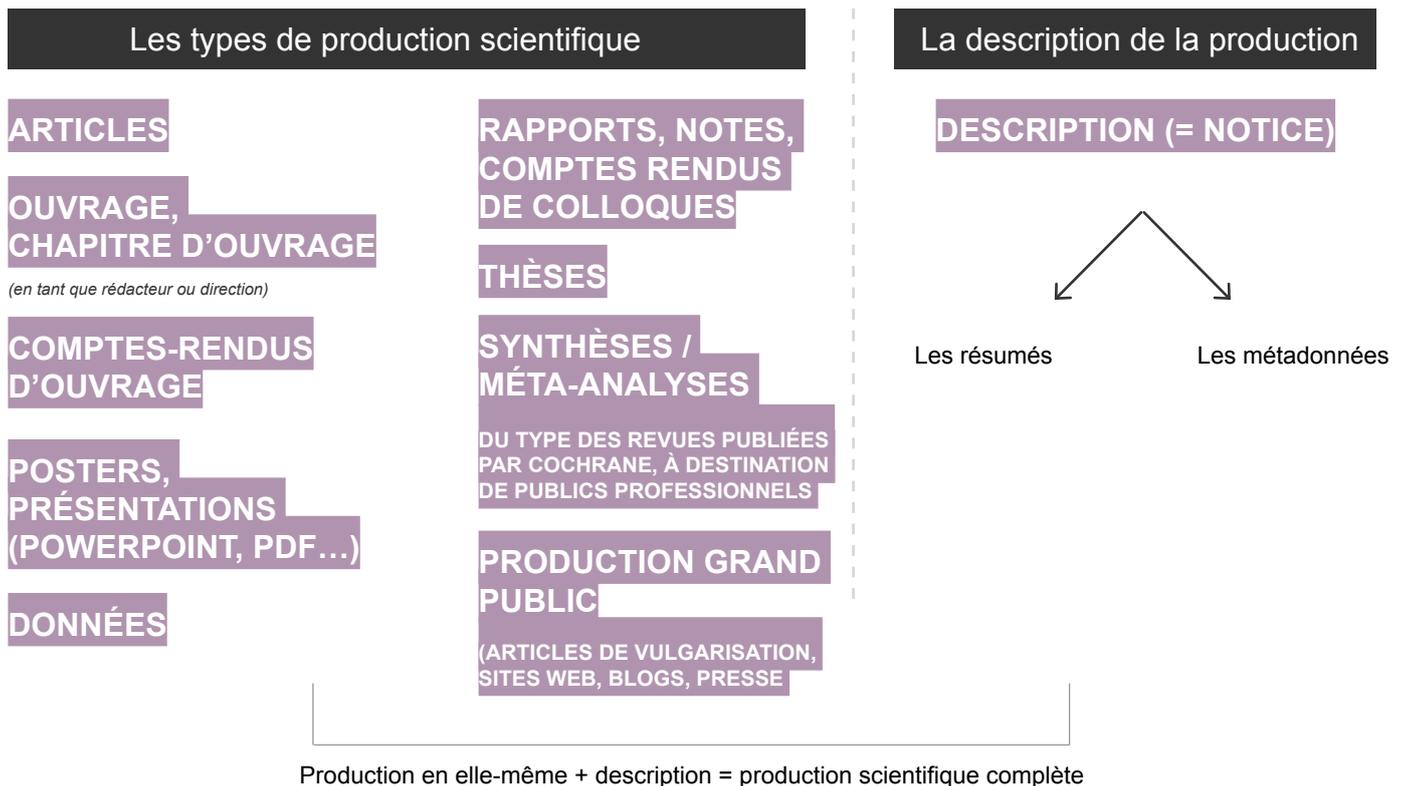
Avant de détailler le parcours de traduction, nous présentons ici le parcours de publication d'un article (avec revue par les pairs) afin de situer l'acte de traduction au sein de cet exemple emblématique de la recherche scientifique..



La diversité de la production scientifique

Nous avons pu relever lors de cette première phase d'étude une **diversité des productions scientifiques**, avec une grande variété des disciplines scientifiques et des sujets, des types de productions mais également des formats et des contextes de diffusion, des pratiques d'autorat, de styles de rédaction - plus ou moins formels et "académiques", standardisés ou libres car la variation stylistique est au contraire encouragée,...

La cartographie ci-dessous vise à définir le spectre de possibles que le futur service devra possiblement prendre en compte et les critères qui vont influencer sur le parcours de traduction :



Autres éléments relatifs à la publication scientifique

STADE DE MATURITÉ DE LA PRODUCTION

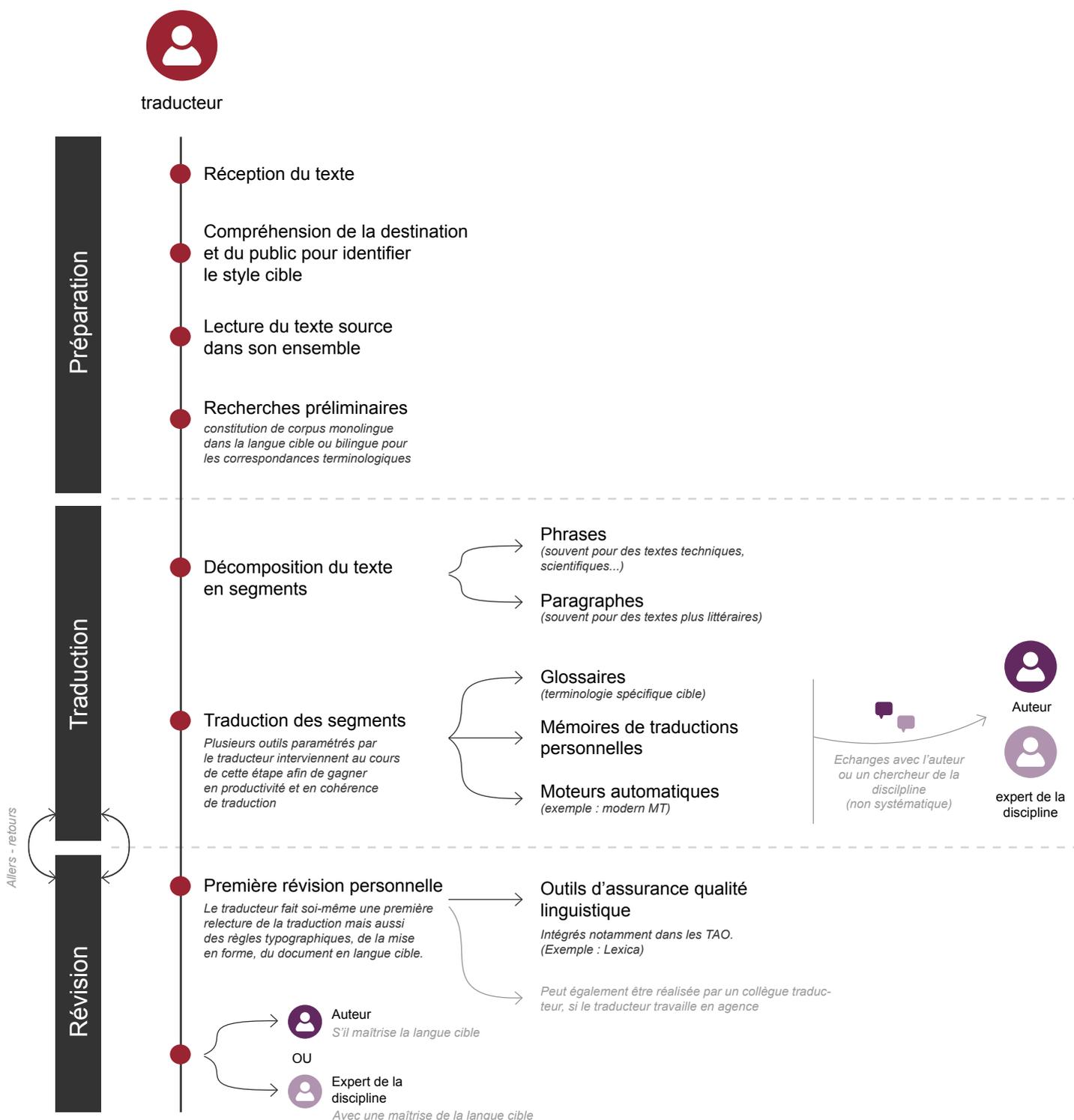
- Prépublication ou preprint
- Version soumise pour publication
- Version acceptée pour publication (après révision par les pairs)
- Version finale publiée (version of record)

LES FORMATS

- Word
- Pdf
- Html
- LateX
- Excel
- Powerpoint
- Epub
- ...

Le parcours type de traduction professionnelle

Nous avons pu identifier un **parcours type de traduction professionnelle** qui passe par la lecture du texte, une phase de recherche et de constitution de corpus de référence dans la langue cible ou de corpus multilingue permettant d'identifier des correspondances terminologiques et phraséologiques, le découpage en segments, l'utilisation et la constitution de glossaires et mémoires de traduction, la traduction à proprement parler et les opérations de révision qui sont faite par le traducteur et généralement un autre traducteur qui se met dans la posture de "réviseur".



Les différentes pratiques de la traduction-rédaction scientifique dans une langue cible ou étrangère

Le parcours type représenté ci-dessus ne représente qu'un type de parcours possible parmi les pratiques qui nous ont été remontées. En effet, si certains chercheurs travaillent dans leur langue puis font traduire, **le niveau de maîtrise de la langue cible du traducteur diffère** : on trouvera des traducteurs professionnels mais également des traducteurs amateurs "de circonstance" (un doctorant du laboratoire ou un proche natif de la langue cible), voire le chercheur lui-même qui se place en position de traducteur pour sa propre production et dont le degré de maîtrise de la langue cible peut varier.

Enfin, selon les disciplines, il va possiblement y avoir un nombre important de **chercheurs qui rédigent directement dans une langue qui n'est pas leur langue maternelle** (le cas d'un chercheur français qui écrit directement en anglais), et dans ce cas, **le texte source n'existe pas, il ne s'agit donc plus de traduire mais de relire un texte** pour corriger le style, lui conférer une tournure idiomatique et vérifier que la terminologie scientifique a du sens (ce qui pour ce dernier point, peut s'avérer difficile car on n'a pas accès à la version originale et que le chercheur n'est pas toujours disponible pour expliquer ce qu'il a voulu dire).

Les critères qui influencent le parcours de traduction - rédaction

Plusieurs critères vont influencer le parcours de traduction, notamment parce que certaines situations vont tolérer un niveau de langue cible moins exigeant ; ces situations vont donc être plus sujettes au recours des logiciels de TAO.

Parmi les critères que nous avons pu identifier lors des interviews, on peut trouver :

Des critères qui vont concerner le contexte de diffusion et impacter le parcours car changeant les attendus en termes de qualité :

- Le **type de production** : prépublication, article, présentation diapositives pour un colloque, compte rendu de colloque, données... On sera moins exigeant pour une prépublication ou pour une présentation diapositives que pour un article publié dans la revue Nature. Dans ce cas, la rédaction en mode "globish" pourra suffire.
- La **diffusion attendue** (cible, type de média et impact potentiel). Par exemple, un article de vulgarisation scientifique pour le collègue différera d'un article destiné à une revue académique qui exige un style plus formel.
- Le **budget** disponible, que ce budget soit pris par l'auteur, sa structure de rattachement ou encore la structure d'édition / publication. Ce budget est souvent directement proportionnel au contexte de diffusion.
- Les **disciplines et le type d'article** (descriptif ou plus narratif, plus conceptuel) peuvent flécher ou non l'usage de l'assistant automatique de traduction.

Des critères plus "techniques" qui vont changer légèrement la façon d'utiliser les outils :

- Les **types de fichiers échangés** (word, LaTeX, InDesign...) différent selon que la revue est imprimée ou purement numérique. La présence de balises HTML ou de formules peuvent indiquer l'usage de logiciels comme LaTeX. En pratique, les outils de traduction professionnels proposent des fonctionnalités permettant de tenir compte de ces balises en les représentant explicitement.
- la **longueur du texte** certains articles de mathématiques font 30 à 50 pages et les chercheurs vont avoir tendance à produire une seule version de texte et privilégier la version anglaise au détriment de la version en langue native.

Enfin, un dernier critère essentiel :

- **L'usage attendu de la traduction**

Les différents usages de la traduction

Il existe plusieurs usages ou finalités de la traduction, qui vont de la **recherche de références** au besoin d'**accéder véritablement au fond** de la publication scientifique en passant par le **survol rapide**. Ces différents usages impliquent des **niveaux de qualité de traduction différents** ; la compréhension fine d'un texte et des concepts qui y sont mentionnés nécessite effectivement une qualité de traduction plus importante, requérant par exemple l'intervention d'un traducteur expert de la discipline, que son survol pour simplement "savoir de quoi ça parle". Le besoin de rendre facilement découvrable le texte sur des moteurs de recherche va nécessiter une pratique de traduction centrée sur les métadonnées, en mobilisant les bons lexiques disciplinaires et structures de métadonnées.

Ainsi, ces différents usages vont avoir une influence sur l'acte de traduction ; on va s'autoriser à traduire soi-même, sans être natif, ou à recourir à des outils de TA pour permettre de parcourir rapidement le texte, mais pas pour faire exister une pensée dans toute sa finesse : pour cela on aura besoin d'un traducteur professionnel et même d'un expert disciplinaire (ces deux rôles pouvant être joué par une seule et même personne si on la trouve...). Dans le cas de la traduction de métadonnées, la traduction peut être plus "technique" et aidée par des experts de l'information scientifique et technique qui ne seront pas forcément des traducteurs professionnels.

Pour préciser les différents (3) usages de la traduction qui sont apparus lors des premiers entretiens :

- **Découvrabilité** : Faire découvrir l'existence. Il s'agit de traduire pour rendre une production découvrable, notamment via les moteurs de recherche ou catalogues. Dans ce cas l'essentiel du travail se fait sur la traduction des métadonnées qui sont à la base de nombreux systèmes de recherche (mais pas que, certains moteurs de recherche pouvant faire du "plein texte").

« Chez Erudit par exemple on réfléchissait à traduire toutes les métadonnées pour faire en sorte que dans les outils de recherche internationaux - principalement anglophones, quand on cherche en anglais on trouve des articles en français ! »

« Si ce qu'on veut, c'est une traduction pour faire en sorte que les articles d'érudit qui sont en français soient repérables, quand on cherche en anglais, ... une bonne traduction, ce sont des métadonnées qui sortent bien ! » - Erudit

- **Survol** : Permettre de parcourir. Le besoin de survol consiste en la compréhension générale d'une ou plusieurs productions scientifiques. C'est typiquement ce qu'un chercheur fait au début d'une recherche : il veut savoir quels champs de recherche ont été explorés ; il n'a pas besoin de comprendre dans le détail de la langue ce dont il est question, juste la méthodologie et quelques éléments...

« Une bonne traduction pour faire découvrir différents types de recherches, des sujets... c'est une traduction qui n'a pas forcément besoin d'avoir les concepts ou les niveaux de précisions. Dans ce cas là on cherche une vision plus globale ; on est au stade exploratoire du cycle de recherche » Erudit

«par exemple dans Firefox, j'ai un petit truc pour que toutes les pages s'affichent en français : ça aide pour l'entièreté du texte dans une phase de recherche exploratoire, sans toutes les subtilités des concepts» - Érudit

- **Pensée** : Faire exister une pensée
C'est le "stade ultime" de la traduction : on traduit dans le texte vraiment, pour faire comprendre une pensée et la rendre existante dans une autre langue.

«Il s'agit de faire exister dans un environnement intellectuel (par ex la France) une œuvre écrite d'abord en italien par exemple. On la lit pour la lire en tant qu'œuvre (pas juste pour chercher différents types de méthodologies ou faire une recension de la littérature...)» - Érudit

La mesure de la qualité de traduction

Nous avons vu que plusieurs critères influent sur la qualité attendue d'une traduction. Il reste alors à préciser comment on peut évaluer cette qualité de traduction.

Plusieurs types de mesures de qualité de traduction existent, notamment pour évaluer la recherche sur les algorithmes de TA. Par exemple, la mesure BLEU (bilingual evaluation understudy) est un algorithme d'évaluation de la qualité du texte qui a été traduit mécaniquement d'une langue naturelle à une autre. La qualité est considérée comme la correspondance entre la production d'une machine et celle d'un humain : « plus une traduction automatique est proche d'une traduction humaine professionnelle, mieux c'est » - c'est l'idée centrale de BLEU. Il s'agit d'une des premières métriques à revendiquer une corrélation élevée avec les jugements humains de qualité, et reste l'une des mesures automatisées les plus populaires et les moins coûteuses. ([source wikipedia : https://fr.wikipedia.org/wiki/BLEU_\(algorithme\)](https://fr.wikipedia.org/wiki/BLEU_(algorithme)) consulté le 15 janvier 2023)

Cependant, plusieurs personnes mentionnent que cette approche est de moins en moins utilisée et que les évaluations faites dans le cadre de la recherche sur la traduction automatique se font aujourd'hui plutôt par des traducteurs humains, même si beaucoup s'accordent à dire que cela rend l'exercice plus subjectif.

Sur la **qualité** de la traduction on peut estimer finalement qu'une bonne traduction doit respecter à la fois des critères formels, potentiellement **mesurables et automatisables** et ceux relatifs au contenu, qui doivent **rester humains**.

a) les critères formels

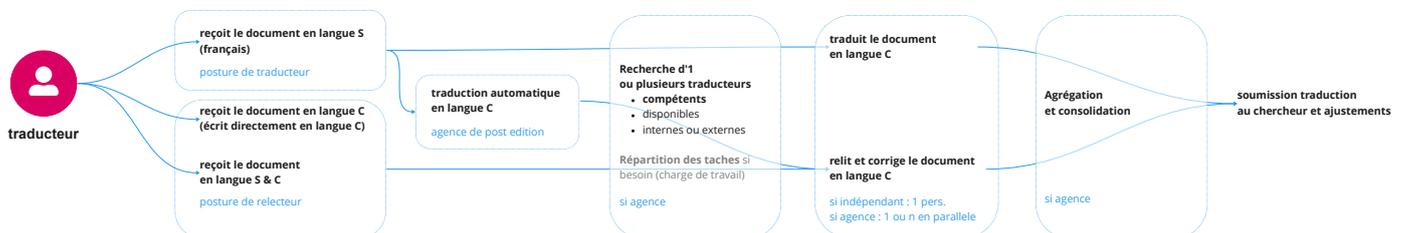
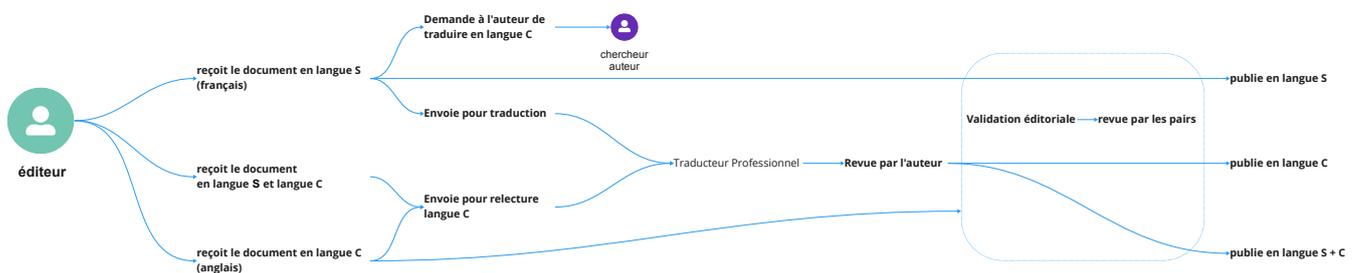
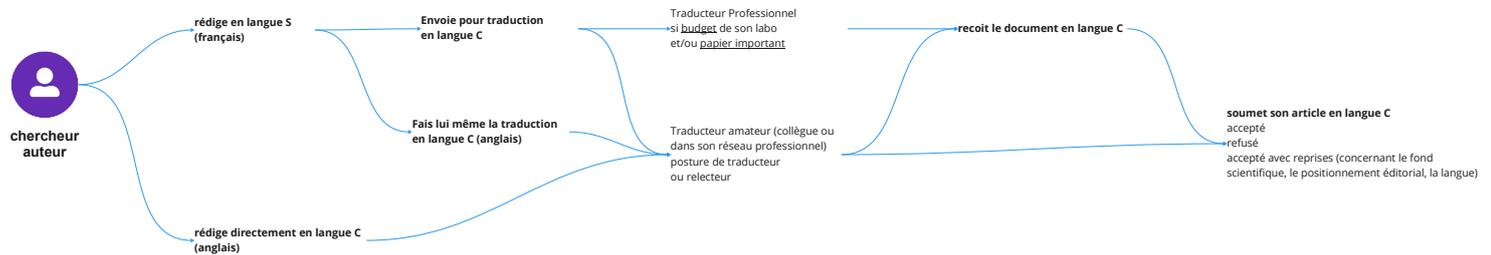
- tout le texte
- rien que le texte (pas d'ajouts par rapport à l'œuvre originale)
- depuis la langue originale (par opposition à une traduction de traduction)
- identification de l'auteur(e) de la traduction, humain ou machine (par exemple, le traducteur a-t-il traduit vers dans sa langue ? Est-il spécialisé dans ce domaine ? Quel moteur de traduction automatique a été utilisé ?)
- date

b) les critères de **contenu**

Si l'appréciation de la qualité des traductions a évolué, ajoutant aux évaluations purement "automatiques" un recours à l'humain, c'est le signe qu'il est difficile d'objectiver totalement la mesure de la qualité des traductions. De fait, le recours à l'humain est fondamental pour prendre en considération le contexte d'une œuvre, ainsi que sa signification d'ensemble - l'approche "holistique" (pour laquelle le tout est plus que la somme des parties, comme lorsqu'une révélation finale vient changer radicalement le sens d'une histoire) s'opposant à l'approche "atomistique" (qui compose - par calcul - le sens du tout comme somme de la signification des parties). C'est ce qui donne notamment sa valeur au fait qu'un même traducteur (ou groupe de traducteurs) traduise les œuvres complètes d'un auteur.

Récapitulatif des situations clefs et variations (1/2)

Langue S : langue Source (ex français)
Langue C : langue de Destination (ex anglais)



▲
les différentes situations rencontrées par les acteurs de la rédaction-traduction.

Récapitulatif des situations clefs et variations (2/2)



chercheur
auteur

Les possibilités (ce qui va varier)

rédige en langue S
(français)
rédige directement en langue C (anglais)
sollicite un traducteur professionnel
sollicite un traducteur amateur
utilise ou non un système de TA

les critères (pourquoi cela va varier)

Sa capacité à écrire et penser en langue C
Son réseau personnel en langue C
Ses ressources budgétaires
ou celles de son laboratoire
L'impact attendu de la diffusion (échelle
et réputation)
La discipline



éditeur

Les possibilités (ce qui va varier)

reçoit le document en langue S
reçoit le document en langue C
sollicite un traducteur professionnel
ne sollicite pas de traducteur

les critères (pourquoi cela va varier)

La capacité de l'auteur à écrire et penser
en langue C
Les ressources budgétaires
L'impact attendu de la diffusion (échelle
et réputation)
La discipline



traducteur

Les possibilités (ce qui va varier)

reçoit le document en langue S
(traduction)
reçoit le document en langue C (manuel
ou auto) : review
reçoit le document en langue S & C
(traduction ou review)
traduit/relit seul ou à plusieurs

les critères (pourquoi cela va varier)

La capacité de l'auteur à écrire et penser
en langue C
Les ressources budgétaires (et le délai ?)
L'impact attendu de la diffusion (échelle
et réputation)
La discipline



les différentes situations rencontrées par les acteurs
de la rédaction-traduction.

c_ Synthèse des entretiens: besoins remontés, enjeux et questions proposées par les personnes interrogées

Les entretiens menés avec les traducteurs et les éditeurs / diffuseurs ont permis d'identifier un certain nombre de pratiques, **besoins et enjeux spécifiques liés à la traduction scientifique.**

Impact des publications / anglais dominant / inégalité d'accès à la diffusion des productions de la recherche

Tout d'abord, il apparaît que pour nos chercheurs européens, l'anglais est aujourd'hui la langue dominante, ce qui pose des problèmes d'accès à la diffusion des productions scientifiques. Certains chercheurs se trouvent en effet en difficulté pour publier dans des revues du fait de leur niveau d'anglais et parce qu'ils n'ont pas accès aux services de traducteurs professionnels. Selon certains témoignages :

"En tant que reviewer, il m'arrive régulièrement de retoquer des articles en anglais de confrères parce que le niveau d'anglais est trop mauvais".

(Un chercheur en psychologie comportementale)

"Pour les sciences "dures", on juge sur l'impact, sur le fait d'être hautement cité et actuellement on est plus cité si on est publié en anglais.

Même si on est attaché au français et qu'on doit le défendre comme langue scientifique, pour ne pas porter préjudice à nos chercheurs, il faut diffuser en anglais." (Un diffuseur)

Un éditeur témoigne du suivi de nombre de vues ou de téléchargements qui mesurent effectivement une différence d'impact significatif des articles en anglais. Beaucoup de directeurs éditoriaux demandent en conséquence des articles soumis en langue anglaise. Dans ce cas, lors de la "lecture critique" faite par un chercheur spécialisé du domaine, il sera demandé d'évaluer, au-delà du fond, le niveau d'anglais. Si le niveau d'anglais est trop faible, l'article pourra être refusé et parfois "sans sommation, car nous n'avons pas les moyens de faire autrement" regrette un éditeur.

Une traduction à la charge du chercheur ou son laboratoire

De ce que nous comprenons, les revues ne prennent pas en charge la traduction des articles, ou si elles proposent de le faire, ce service est généralement monnayé. C'est in fine au chercheur qu'il revient de fournir son article en anglais. Selon les moyens de son laboratoire, ou ses propres moyens, il fera appel ou non à un traducteur pour traduire son article ou le corriger.

Des traducteurs qui pratiquent différents types de révision de la traduction: a minima la vérification de la terminologie scientifique et technique

Les différents traducteurs témoignent tous de l'importance selon eux de la révision des traductions qu'ils produisent, mais cette traduction peut être de deux sortes : une révision du style, de l'idiomaticité et de la phraséologie du texte cible ou une révision du fond et du bon emploi de la terminologie scientifique. Certains traducteurs procèdent aux deux types de révisions car ils travaillent d'un côté en relation étroite avec l'auteur du texte, et de l'autre avec un confrère traducteur qui relit le texte traduit, mais dans d'autres situations, seule la vérification de la terminologie scientifique et technique est faite. Comme le reconnaît une traductrice : *"Pour moi la révision est indispensable dans le domaine. Je traduis de façon beaucoup plus sereine sachant qu'il va y avoir une review."*

Des enjeux pour la vérification terminologique scientifique et technique: la disponibilité de l'auteur dans la langue source ou d'un relecteur dans la langue cible.

Un traducteur témoigne qu'il traduit des articles déjà publiés et que les chercheurs auteurs des articles ne se rendent pas toujours disponibles. Dans ce cas, le traducteur va aller chercher d'autres spécialistes du domaine qui pourront l'éclairer sur la signification des termes et des concepts dans la langue source. L'idée est de bien comprendre ce que l'auteur a voulu dire dans la langue source.

Dans d'autres situations, on demande à un spécialiste du domaine dont la langue maternelle est la langue cible si le texte cible est correct d'un point de vue scientifique et technique. On ne cherche pas dans ce cas à remonter à ce que l'auteur aurait pu vouloir dire mais bien à l'exactitude des propos dans la langue cible. Il faut toutefois noter que cette situation nous a été rapportée par une traductrice qui traduit des modes d'emplois techniques dans le domaine médical et non des articles scientifiques à proprement parler, articles scientifiques pour lesquels la compréhension de la pensée originelle de l'auteur est fondamentale.

Des opérations de contrôle spécifiques pour les articles scientifiques

La spécificité de la production scientifique et notamment d'un article scientifique nous semble tenir du fait que l'auteur est identifié et responsable de son texte.

"Si on revient à la traduction scientifique, elle suppose que l'auteur soit identifié pour qu'on puisse échanger avec lui, critiquer, tenter de refaire ses expérimentations,..." Claire Larssonneur

La priorité pour un article scientifique nous semble être de pouvoir traduire fidèlement la pensée de l'auteur et voici une première priorisation des révisions de la traduction que nous pouvons proposer :

- échange entre le traducteur spécialisé et le chercheur auteur pour vérifier ce que l'auteur a voulu dire dans sa langue source
- échange entre le traducteur et un autre spécialiste (dans la langue source) à défaut de l'auteur
- si le traducteur n'est pas disponible, échange avec un chercheur spécialiste du domaine dans la langue cible pour vérifier l'exactitude scientifique
- révision par un traducteur natif dans la langue cible

"Sur les sujets techniques, comme par exemple sur l'énergie nucléaire, au choix, je ferai relire par une personne familière du sujet, plutôt que par un traducteur." une traductrice

Les règles de rédaction pour les publications académiques

Les traducteurs et éditeurs ont souligné l'exigence particulière de l'édition académique. Certaines revues proposent des normes d'écriture (des guides de style, des plans types de rédaction,...).

Il y a, par exemple, en fonction de la langue et des revues, certaines normes orthographiques et typographiques à respecter pour les citations et l'usage des guillemets, les notes bibliographiques, et certaines revues encouragent ou à l'inverse normalisent le plan et la rédaction des articles. Comme le dit Claire Larssonneur : *"Certains articles de chercheurs en informatique sont hyper formatés, avec un plan type standardisé, et il y a déjà 15 à 20% des phrases que l'on peut réutiliser dans tous les articles."* Les outils de TAO pourraient alors être utiles pour faire les opérations de vérification typographique de façon semi-automatique. Claire Larssonneur pose aussi la question des outils d'aide à la rédaction du futur service: ne proposerait-on pas, dans les situations où les articles sont très standardisés,

des "templates de rédaction ou il n'y aurait finalement plus que les blancs à remplir"?

Textes scientifiques: quand la traduction automatique paraît indiquée

Les différents traducteurs interrogés ont, à plusieurs reprises, fait remonter que les outils de TAO et plus particulièrement la traduction automatique étaient intéressants pour des textes assez descriptifs, avec un style relativement neutre, assez longs et répétitifs, pour lesquels souligne une traductrice, il peut y avoir peu de valeur ajoutée ou de plaisir à traduire en tant que traducteur.

"Je n'utilise pas toujours la traduction automatique: sur des documents plus rédactionnels, je prends plus de plaisir à traduire à la main...sur du technique, il n'y a pas de plaisir particulier à manipuler la phrase." une traductrice

L'assistant de traduction est alors utilisé pour aller chercher rapidement des correspondances dans des glossaires ou des mémoires de traduction ou comme outil d'aide à la frappe (fonctionnalités de suggestion et autocomplétion).

"La traduction automatique me fait gagner énormément de temps et ce temps gagné, je l'utilise pour faire mes recherches terminologiques plus précisément. Il y a des phrases, je n'ai aucun intérêt à les taper. J'ai plus de temps pour faire ce qui m'importe pour la traduction." une traductrice

Pour des textes plus littéraires, la machine peut se retrouver davantage en difficulté. Une autre traductrice souligne que la segmentation par phrase proposée par son outil de TAO ne convient pas à certains textes et qu'elle préfère traduire paragraphe par paragraphe, de façon plus libre.

Le choix de la traduction automatique ? Par critères plutôt que par discipline scientifique

Pour l'instant, il ne nous a pas été permis de dire si la traduction automatique était davantage indiquée pour telle ou telle discipline. Il y a l'intuition que les sciences humaines et sociales, parce que le style peut y être plus libre, plus littéraire pourrait être un moins bon candidat que les sciences formelles par exemple. Cependant, même au sein d'une discipline, la variété des productions scientifiques (prenons l'exemple d'un article de vulgarisation mathématique pour le grand public), des pratiques d'atorat et des sujets, ne nous permet pas de favoriser un type de science ou de discipline plutôt qu'une autre. Une approche par critères nous semble donc plus pertinente: ce texte est-il plus ou moins littéraire, est-il au contraire très descriptif ? Y a-t-il beaucoup de concepts dont la bonne traduction est critique ? Y a-t-il des formules qui ne sont pas balisées et donc que le moteur va traiter comme du texte et non une formule ?...

Ce qui va par contre changer est le fait que les chercheurs de telle ou telle discipline vont davantage écrire directement en anglais ou non, ou à défaut d'écrire correctement en anglais, auront certaines facilités à lire en anglais.

"Dans les sciences dures, et tout ce qui est plus technique (physique, chimie, médical) le niveau d'anglais est meilleur en général. Ce sont des mondes qui sont habitués à lire et publier en anglais: la médecine en particulier. Ils lisent tout le temps en anglais à défaut de pouvoir toujours bien s'exprimer en anglais." un traducteur

Les points de vigilance concernant l'usage de la traduction automatique

Des traducteurs ont souligné qu'une fois la suggestion proposée par la machine, il est difficile de s'en détacher, la suggestion fait écran.

"Le problème c'est quand on nous amène à une phrase "comme ça", ça influe sur notre regard. La phrase a l'air correcte, c'est OK, c'est bon je vois tous les mots." un traducteur

"Dans les opérations de post-édition, parfois on se rend compte le lendemain, qu'on est resté trop près de la machine. Il y a le "halo de la machine", le texte pré-traduit fait écran." Claire Larsonneur

Ces difficultés de prise de recul critique rencontrées par les traducteurs professionnels doivent nous alerter sur l'utilisation de la traduction automatique par des non experts (traducteurs amateurs ou chercheurs): comment ceux-ci peuvent-ils être critiques face aux propositions de la machine, quels garde-fous doit-on mettre en place pour que ces suggestions ne soient pas prises pour argent comptant ?

Les stratégies de post-édition en réponse à la traduction automatique

Un traducteur nous a dit tester certains morceaux de textes dans le logiciel de traduction automatique DeepL pour voir si cela donnait des résultats corrects et décider ou non de continuer à travailler son texte avec l'assistance de la traduction automatique.

Deux traductrices nous ont confié qu'elles n'affichent pas la traduction automatique de façon systématique dans la colonne du texte cible afin de garder de la distance vis-à-vis des suggestions faites par le moteur de traduction neuronale.

"J'utilise la traduction automatique en mode consultation [- pas dans la colonne de gauche, mais dans l'espace supérieur de Trados]. Je ne la propage pas automatiquement." Une traductrice

Ecrire ou non dans sa langue, précision de la pensée et autres enjeux

Si l'on comprend qu'on réfléchit mieux dans sa langue, doit-on encourager la rédaction dans sa langue maternelle plutôt que de rédiger directement dans la langue cible ? Que faire face aux pratiques qui consistent à écrire directement en anglais alors qu'on est un non-anglophone par exemple ? Doit-on rester sur un outil de traduction car on réaffirme l'importance de penser dans sa langue ou se met-on également au service des chercheurs qui rédigent directement dans la langue cible ?

"Quand j'écris un article en anglais ou en français, je n'aborde pas les problèmes de la même manière. En anglais, on suit le détail des processus, des actions-réactions des objets tandis qu'en français on s'intéresse au résultat final, le you en anglais est beaucoup plus utilisé tandis que les textes français sont plus impersonnels,... Chaque langue correspond à un fonctionnement cognitif particulier" Claire Larsonneur

La rigueur scientifique nécessite enfin une précision qui ne peut pas être laissée à l'appréciation seule de la machine.

"Human in the loop. Dans les sciences il y a une question d'exactitude et de précision qui est importante à garder." une traductrice

Les pratiques collaboratives de la recherche et de la traduction.

Il est courant que les chercheurs travaillent à plusieurs à un article et soient dans une posture de co-auteurs, tandis que les traducteurs peuvent être amenés à travailler à 4 mains sur un article (en traduction et en révision), ce qui plaide pour un futur outil de TAO qui soit collaboratif et dans le "cloud".

"On est toujours 2 ou 3 auteurs sur un article et on utilise OpenLeaf, une sorte de LaTeX collaboratif en ligne" un chercheur spécialisé dans le domaine des mathématiques appliquées à la physique

FOCUS :

DISCUSSION AVEC

CLAIRE LARSONNEUR 1/3

Les tendances

Quelles sont les tendances dans le domaine de la traduction et notamment la traduction assistée par ordinateur ?

La tendance lourde est à l'utilisation de **moteurs de traduction automatique**, désormais accessibles en ligne et intégrés dans différentes applications (réseaux sociaux, traitement de texte).

De nombreux traducteurs utilisent encore les **logiciels de traduction assistée** (et non automatique) tout simplement parce qu'ils les ont achetés et que ces logiciels sont encore la base de la facturation. Toutefois, **l'industrie des langues est passée à la traduction automatique suivie de post-édition**.

Ceci pose la question de la formation des traducteurs à l'utilisation de ces outils. Il faut former à bien retravailler un texte pré-traduit, ce qui passe par former à faire de bonnes traductions et pas seulement apprendre à corriger les erreurs de la machine. Parfois, après avoir utilisé la machine, on se rend compte le lendemain qu'on est resté trop près de la traduction automatique. On n'a pas réussi à **se détacher du "halo" de la machine et du texte pré-traduit qui fait écran**.

Le grand fantasme de la traduction automatique est la transversalité du modèle. En réalité, il va falloir être capable de **choisir quelle intelligence artificielle pour traduire tel type de texte**, en fonction des disciplines. Il y aura un travail de "DJ" pour pouvoir choisir quel disque prendre. Il y aura une connaissance des machines car il y aura différentes machines et des effets de spécialisation qu'il faudra connaître.

Autre tendance, depuis 2 ou 3 ans, la traduction automatique **ne se contente plus de faire du texte à texte** : elle perfectionne désormais les fonctions texte/voix ou même voix/voix (cf Apple translate).

De plus, il existe désormais des logiciels qui permettent de faire de la **rédaction assistée** directement (cf ChatGPT, GPT 2 ou 3, Facebook etc).

J'ouvre mon application, je dicte et le téléphone me dit la traduction ou, dans notre cas, le chercheur dicte son texte à la machine (2 ou 3 phrases) et la transcription peut être faite dans le texte cible. Pour l'instant, sur des textes longs, la machine va être en difficulté et il faut s'arrêter "phrase à phrase". En soi, cela n'est pas prévu pour faire de la rédaction longue mais pour des énoncés courts (pour des réseaux sociaux). On voit cependant certains utilisateurs qui détournent l'usage de dictée vocale pour faire de la traduction.

Fonctionnement MTN et IA pour la traduction

Comment, sur quels matériaux sont entraînées les IA ? (des traductions parallèles ? des traductions comparables ?)

Les IA sont entraînées sur des **corpus**, c'est-à-dire des bases de données qui croisent des textes ou des énoncés vocaux. L'entraînement se fait principalement sur des corpus bilingues : à savoir des bases de données qui croisent des textes et leur traduction existante (ce qui pose un problème, car les énoncés d'entraînement sont des traductions et non pas des productions natives, or il y a une différence linguistique et stylistique

entre les deux). Dans le cas des langues pour lesquelles il existe peu de données bilingues disponibles, les chercheurs créent des données artificielles et recréent des corpus bilingues de façon artificielle à partir de corpus monolingues (back translation), ce qui évidemment n'est pas l'idéal.

Comment les ressources sont-elles moissonnées/collectées ?

Le grand angle mort de ce domaine ! Google et Facebook (et les autres) utilisent massivement toutes les données postées en ligne par les internautes et très certainement tout ce qui a été numérisé par Google Books. Les chercheurs en général utilisent beaucoup les **TED Talks**, le corpus **Europarl** de l'Union européenne ainsi que les corpus montés régulièrement par l'Association internationale de la recherche sur la traduction automatique pour comparer les performances (les **WMT 2014**, **WMT 2018** etc.). Un grand nombre de corpus sont disponibles en ligne, pas toujours de bonne qualité. On ne sait pas ce qu'utilise **DeepL** mais je parierai qu'ils utilisent aussi les **données en ligne APRES les avoir nettoyées de manière plus approfondie que leurs compétiteurs**.

Un des problèmes en terme de qualité des recommandations est quand la machine parle à la machine et que les moteurs apprennent sur des données créées par des IA.

Comment se fait l'apprentissage (par probabilité, par feedback humain,...) ?

Les deux (unsupervised et supervised training). Le principe est de récompenser toutes les productions qui s'approchent d'un énoncé existant.

Après, il faut distinguer les corrections faites par une personne ou un groupe de personnes en local qui sont de l'ordre de l'enregistrement de préférences (par exemple la frappe prédictive ; ma fille s'appelle Constance, dès que je tape "Con" sur mon téléphone, on me le propose !)

Quels sont les points clefs de ces techniques ? données d'entraînement, explicabilité (traçabilité des données qui ont servi à entraîner les IA), limites...

Les logiciels ne peuvent rien faire de bon s'ils ne sont pas entraînés sur des données de qualité. Le nœud du problème est là et c'est pourquoi il faudrait **entraîner les algorithmes spécifiquement sur les textes de recherche**, idéalement en **séparant l'outil pour la médecine de l'outil pour les SHS**, etc. Outre ce point concernant la production, l'enjeu majeur dans ce secteur d'activité tient à l'**exploitation des données** (en clair, Google, Facebook et Microsoft ne s'intéressent pas à la traduction en soi, mais à la traçabilité des interactions des internautes que cela leur permet).

Une limite importante tient à ce que la traduction automatique produit un calque massif des structures de la langue de départ : or, on ne présente pas un fait, on ne décrit pas un événement et on ne raconte pas une histoire de la même manière dans des langues différentes. Du coup, les textes traduits sont en fait plus difficiles à comprendre pour les locuteurs natifs de la langue cible.

FOCUS :

DISCUSSION AVEC

CLAIRE LARSONNEUR 2/3

Quelle spécificité de la traduction scientifique (articles, résumés de colloques, résumés d'articles) ? Et y a-t-il des différences disciplinaires qui impactent la traduction (entre SHS et sciences formelles par exemple, ou entre de la médecine et des mathématiques) ?

Quand j'écris un article en anglais ou en français, je n'aborde pas les problèmes de la même manière. En anglais, on suit le détail des processus, des actions-réactions entre les objets tandis qu'en français on s'intéresse au résultat final. Autre exemple, le you en anglais est beaucoup plus utilisé tandis que les textes sont plus impersonnels en français. **Chaque langue correspond à un fonctionnement cognitif particulier et c'est cela qu'on écrase quand on utilise la traduction automatique.**

La machine fait très bien la traduction de textes stéréotypés, répétitifs mais elle **ne va pas bien traduire la manière dont on va approcher un problème.**

Enfin, **certains concepts n'ont pas d'équivalence** dans la langue cible et ne sont pas compréhensibles sans la note du traducteur.

Sur les différences disciplinaires

Oui, il y a d'**importantes différences disciplinaires** et un seul outil ne conviendrait sans doute pas. Par exemple, j'ai observé que la structure d'un **article en sciences informatiques est très normée** : ceux-ci doivent se caler sur un **modèle** qui peut facilement être reproduit en multilingue. La langue en soi est caractérisée par un vocabulaire spécialisé (jargon technique) et une nécessaire simplicité stylistique (exposer des étapes, proposer des causalités). En revanche, ces articles comprennent de **nombreux tableaux et graphiques** qui posent des **problèmes spécifiques** (par exemple la quantité de texte qu'on peut mettre dans une cellule ou encore la traduction des nombres). Inversement, en SHS et particulièrement en art, on nous incite désormais à utiliser des **formats créatifs** en intégrant des images, des vidéos, des narrations différentes. La fluidité du style est souvent importante et on nous incite aussi à créer de nouveaux concepts, lesquels ne seront pas répertoriés dans la littérature préexistante.

Tous les articles toutefois sont accompagnés de **résumés et de métadonnées, plus faciles à traduire automatiquement.**

Enfin la spécificité de la traduction en recherche est que les articles visent la production et la transmission de savoirs et de techniques susceptibles d'avoir un effet direct sur nos sociétés. **Une mauvaise traduction est donc un risque pour tous.** Cela se comprend aisément en médecine si on se trompe par exemple de nom de médicament. Moins évident, le risque en SHS est néanmoins présent : une mauvaise traduction d'un article de sociologie sur la ségrégation urbaine peut inspirer des politiques publiques mal fondées. Sans parler des négationnismes divers, propagande etc.

Le plus simple serait d'utiliser la formule du "opt out" à savoir de rédiger une notice d'utilisation qui prévient les déposants de la traduction automatique des résumés et des métadonnées. **S'il s'agit de traduire l'intégralité de l'article, il serait sans doute sage d'en avvertir l'auteur par courriel et de l'inviter à relire et à valider la traduction.**

En deux mots, traduire les résumés et les métadonnées c'est bien, mais s'il s'agit de traduire des articles attention.

Enfin, de façon plus large, l'algorithme, c'est un peu comme l'ouvre bouteille, on peut l'utiliser sur pleins de bocaux différents. Si on veut des bonnes traductions de qualité, le corpus d'entraînement est essentiel et cela vaut le coup d'avoir un même logiciel entraîné par corpus disciplinaires. Les points clés des techniques d'IA, ce sont les corpus.

IA & humain. Traduction assistée par ordinateur : quels rapports ordinateur / humain

Freins et motivations à utiliser des outils d'assistance que vous pouvez observer ; les angoisses ou au contraire les fantasmes ?

Du côté des traducteurs expérimentés, l'angoisse de perdre leur source de revenu est un sujet brûlant. Cette angoisse est à mon sens fondée. Un deuxième frein est l'habitude d'utiliser d'autres outils (TAO), donc un frein classique de résistance au changement. Un troisième frein serait le sentiment de déclassement : corriger une machine au lieu de créer.

Du côté des clients (éditeurs, institutions), il y a le fantasme de la traduction instantanée et gratuite.

Des exemples d'erreurs typiques de l'IA appliquées à la traduction ?

La traduction automatique peine sur les phrases longues, les mots rares, l'attribution du genre et de manière générale, les références au réel (la machine ne peut pas savoir que le soleil se lève à l'est par exemple). Un cas particulier est celui des hallucinations, où le moteur va proposer des solutions décalées, très étranges, lorsqu'il ne trouve pas de correspondances dans son corpus. Par exemple, lors d'une traduction d'un extrait de *The Night Watchman* de Louise Erdrich, roman situé dans une réserve indienne en Amérique, dotée de lacs, rivières et marais, l'expression "slough-dotted reservation" est devenue "réserve à pois oblique".

Quand l'humain est-il meilleur ? comment machine et humain collaborent ?

A mon sens, envisager une collaboration homme-machine est la bonne approche. Par exemple, moissonner les articles recensés dans HAL pour établir des outils terminologiques assortis de références : quelles traductions pour la notion de leuzienne de dispositif dans quels articles ?

Ou bien en sciences proposer des templates d'articles pré-rédigés dans lesquels les chercheurs vont insérer leurs données. On peut proposer un assistant pour identifier la bonne terminologie, gérer les références, aider à la rédaction.

Il faut des fonctions spécifiques pour les chercheurs, sinon ceux-ci risquent d'aller au plus simple: Google Translate...

FOCUS :

DISCUSSION AVEC

CLAIRE LARSONNEUR 3/3

Quelles sont les critiques faites à l'IA et quels sont les contre-arguments ?

Il y a des centaines de pages sur ce point... voir les travaux de Vanessa Nurock par exemple. L'intelligence artificielle : enjeux éthiques et politiques (Cités 2019/4, PUF). Côté grand public, il y a la **peur du remplacement de l'homme par la machine**. Côté experts, je citerai l'**opacité des procédés** et l'**oligopole des prestataires**. Mon avis est qu'il faut encadrer certains usages (notamment l'exploitation des données), préciser quelle est la responsabilité des prestataires et former les personnes à utiliser ces outils.

Références

1_ Une ou deux références / liens / articles pour comprendre la traduction neuronale, ses (bons) usages et limites

L'ouvrage le plus complet est celui de Thierry Poibeau chez Odile Jacob : Babel 2.0. https://www.odilejacob.fr/catalogue/sciences/informatique/babel-20_9782738148490.php

Pour les usages et les limites, voir les travaux de Joss Moorhens ou les miens (en anglais) : sur l'exploitation des données, sur la dimension éthique, sur la distorsion de marché.

Par exemple : **Translation Spaces - Volume 9, Issue 1, 2020** <https://www.jbe-platform.com/content/journals/2211372x/9/1>

2_ Les outils et les services à regarder (DeepL, MateCat,...?) pour faire une analyse critique.

Si vous entendez par là une évaluation des performances des moteurs de traduction, la plupart des chercheurs utilisent le score BLEU (proximité avec des traductions humaines existantes), lequel est toutefois limité et critiqué.

3_ Les moteurs de traduction neuronales ? (au-delà des outils, des briques techniques incontournables : ModernMT...).

En termes de moteurs et si on reste dans le cadre des langues européennes, les plus intéressants sont **DeepL, Google Translate, Microsoft Word et Lingvanex**. Tout simplement parce que ce sont les plus accessibles et qu'ils sont produits par des équipes de recherche de très haut niveau. Autre ressource clef à mon avis : **eTranslation**, produit par l'Union européenne et ouvert depuis peu gratuitement aux institutions et entreprises européennes. Il faut s'inscrire et être une institution, mais eTranslation est déjà multilingue et pourrait être entraîné sur d'autres sujets que celui des politiques publiques.

Attention, en fonction du contexte d'utilisation, les outils de références ne sont pas les mêmes. Pour le Coréen, on regardera par exemple <https://papago.naver.com/>

Quelles sont les incitations à publier la recherche en français ? Y a-t-il des obligations à publier la recherche en français ? Est-ce que des obligations de publier la recherche en français existent ailleurs ?

Pas à ma connaissance (attention, comme je travaille sur un domaine anglophone, la publication en anglais est naturelle). Cela dit, j'ai regardé le cas canadien où il existe une obligation légale pour que les textes officiels soient publiés dans les deux langues, et il apparaît que les chercheurs francophones préfèrent là aussi publier en anglais. Ce qui se pratique est la publication bilingue quasi-systématique des résumés et des mots-clefs.

<https://francopresse.ca/2021/06/08/la-recherche-en-francais-au-canada-publier-en-anglais-ou-perir/>

Valeur et modèle économique

Quels "objets" peuvent être monétisés ? Peut-on imaginer permettre aux utilisateurs de valoriser (réputation, points, troc,...) ou monétiser leurs corpus de traduction générés à partir de leurs traductions ? Peut-on valoriser ou monétiser les corpus d'entraînement spécialisés par disciplines ?

La vraie question n'est-elle pas QUI paye ? je ne vois pas l'Etat français ou les universités rémunérer les chercheurs pour leur visibilité accrue...

Open Data / open science / transparence des algorithmes et communs numériques

Que montre-t-on, que rend-on disponible ? Le code, les corpus et mémoires de traduction qui servent à entraîner le moteur ? Peut-on rendre utilisable par d'autres les mémoires de traduction sans permettre de reconstituer les textes / articles scientifiques complets ? Y a-t-il des solutions (sur le découpage des textes, leur échantillonnage) ou défensive (repérer un nombre de requêtes) ?

Excellentes questions sur lesquelles je pense que Lionel Mauriel aura des réponses plus précises que moi. Il existe déjà des algorithmes open source de traduction automatique OpenNMT par exemple <https://aclanthology.org/P17-4012.pdf>. La clef du problème est toutefois l'accès à des corpus spécialisés et les ressources humaines pour entraîner et ajuster l'outil.

Sur les mémoires de traduction, il faudrait parvenir à créer des corpus accessibles aux chercheurs et spécialisés par grands domaines scientifiques. Cela implique des équipes.

Mon avis est qu'il serait coûteux et sans doute peu efficace de développer un algorithme spécifique pour la traduction scientifique : pourquoi ne pas chercher plutôt un partenariat avec eTranslation (UE), qui est un outil public et sécurisé où ils ont déjà l'expertise multilingue. Et mobiliser cette ressource pour l'entraîner sur les corpus de recherche.

03

ANALYSE D'OUTILS

a_L'outil type de TAO

Le parcours type de traduction peut être facilité par l'usage d'outils de TAO qui travaillent pour la majorité d'entre eux par **couple de langues avec une vue synoptique en 2 colonnes** où le texte source à traduire est affiché dans la colonne de gauche et découpé en segments (par défaut par phrase ou proposition).

Au delà de cette vue synoptique et du découpage, les fonctionnalités clefs proposées sont :

- **la constitution et la consultation de glossaires et mémoires de traduction**, que ceux-ci soient déjà prêts (voire téléchargeables) ou soient au contraire en cours de construction (de façon incrémentale).
- **la suggestion par un moteur de traduction automatique**
- **un suivi de l'avancée de la traduction** (avec le pourcentage de mots traduits par exemple)
- **des outils pour la correction de la traduction**: correction orthographique et typographique, vérification des correspondances (un terme est-il toujours bien traduit par le même terme)
- **un espace collaboratif pour la révision de la traduction** dans l'idée qu'un tiers puisse reprendre la traduction pour révision.

Avant même de parler d'utilisation d'un moteur de traduction automatique qui fait des suggestions, les fonctionnalités utiles des outils de TAO visent à proposer un espace ergonomique pour la traduction grâce à la vue synoptique et la recherche dans des glossaires et mémoires de traduction organisés en bases de données.

"A la base ce qui est intéressant, ce sont les mémoires de traduction, vous pouvez réutiliser les mêmes segments et gagner du temps, on peut faire des glossaires, pour retrouver des mots et aller plus vite. Ils sont utilisés pour faire le contrôle qualité, pour faire de la vérification."
une traductrice

Si la structure du futur outil reprendra probablement les fonctionnalités ci-dessus, il nous semble important de noter d'ores et déjà que la vue synoptique devrait pouvoir **proposer plusieurs colonnes** afin d'ouvrir la possibilité de travailler dans un espace multilingue.

On pense qu'il est également important que la traduction automatique ne soit pas par défaut proposée et, si elle est activée, que la suggestion ne se fasse pas par défaut dans la ou les colonnes cibles pour ne pas faire "écran" et permettre aux utilisateurs de **se détacher** de la suggestion proposée par la machine.

Par ailleurs, il serait bon d'inclure, dans le futur service, des fonctionnalités que certains outils de TAO ne proposent pas, parce qu'ils sont fournis par d'autres prestataires : recherche de synonymes (synonymo.fr), que DeepL propose toutefois, définitions contextuelles (de type Linguee) qui sont la plupart du temps indexées sur des mémoires de traduction (il s'agit de consulter des concordances), dictionnaires (monolingues ou bilingues), encyclopédies, grammaires (règles de syntaxe, de conjugaison), manuel de "bon usage" (type Grévisse), etc.

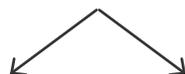
b_Cartographie des outils - vue d'ensemble

Note

Ces outils proposent d'outiller différents gestes qui participent du parcours de traduction ou rédaction et proposent différentes fonctionnalités qui pourraient faire partie du service de Traduction Outillée ou être utilisés en complément de notre futur service.

Les outils de traduction

Les outils «métiers» des traducteurs



LES OUTILS D'AIDE À LA TRADUCTION

Les modèles d'outils correspondant à une pratique «historique»

- MateCat
- Trados
- TraduXio
- DéjàVu
- MemoQ
- Systran
- Le kit de traduction de Google
- Microsoft translate
- Omega-T

Les outils de «post traduction»

- Lexica

LES OUTILS COMPLÉMENTAIRES

Les outils de création / gestion de corpus

- Antconc
- Sketchengine

Les corpus bilingues ou monolingues disponibles

- Linguee
- Commn crawl
- Oxford recollection
- U Align

banques de terminologie, dictionnaires et glossaires

- Termium Plus
- IATE (Interactive Terminology for Europe)

LES OUTILS DE TRADUCTION - RÉDACTION AUTOMATIQUES

De nouveaux outils, non exclusifs aux traducteurs

Les moteurs MTN

- DeepL
- Google Traduction
- Modern MT
- eTranslation (service de TA fourni par la Commission Européenne)
- LINGVANEX

Les modèles de langage

- GPT-3
- BERT
- API Muse (par LightOn, startup française)

Les logiciels de synthèse / description de contenu

- Open Calais (développé par Reuters)

Les outils connexes

LES OUTILS DE POST-EDITION

- Antidote, Language Tool, Grammarly (correction ortho-typo)
- XBench (vérifier la cohérence des formules mathématiques)

LES OUTILS DE PAO

- Suite Office (Word, Powerpoint...)
- LaTeX ("le Word des articles scientifiques")
- HTML
- Indesign, Framemaker (et autres logiciels de mise en page)

LES OUTILS DE PARTAGE / COLLABORATION

- WordPress, Open Journal System (pour l'édition et la Revue par les pairs)

LES AUTRES OUTILS

- Zotero
- Open Journal System
- Scopus
- Les guides proposés par les revues

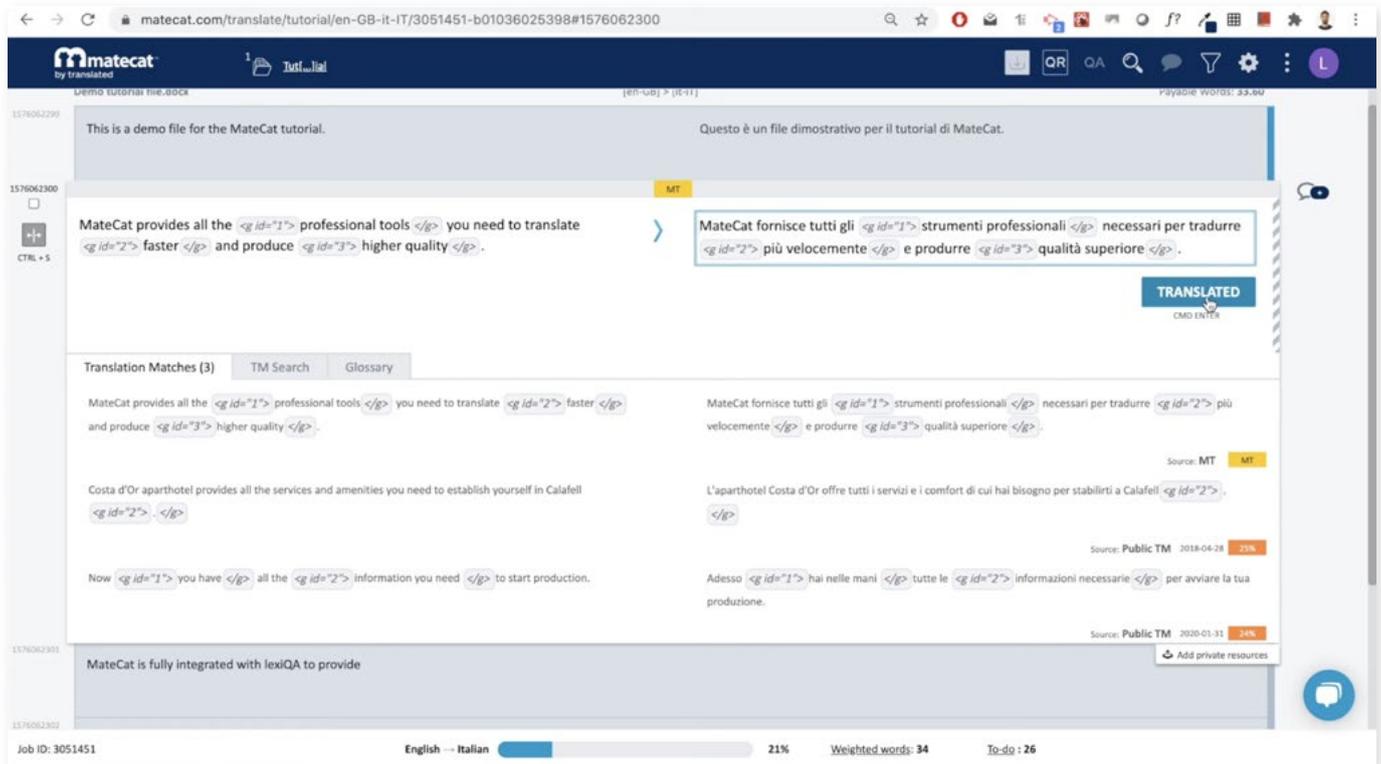
Des outils d'aide à la rédaction, pour le traducteur

Liste des gestes de rédaction-traduction :

- **Recherche d'une définition**
("que veut dire ce terme ?"), dans un dictionnaire (monolingue, type Larousse) ou une encyclopédie (type Universalis ou Wikipedia), dans la langue source
- **Consultation d'un dictionnaire bilingue**
(plusieurs sont disponibles sur lexilogos.com).
 - Variante: **dictionnaire bilingue contextuel** (Linguee, Reverso+contexte)
 - Variante : recherche dans un **dictionnaire monolingue** de la langue cible (pour vérifier si notre idée intuitive de traduction est bonne)
 - Variante: recherche d'une **définition en contexte**: avec des exemples d'extraits de textes, type Linguee
- **Recherche d'un synonyme**
("comment traduire cela autrement, car j'ai déjà utilisé un premier terme et j'aimerais varier un peu"), type <https://www.cnrtl.fr/synonymie/> ou <https://synonymo.fr>
- **Recherche d'occurrence**
("comment ai-je déjà traduit cela plus haut, au fait?")
 - Variante: recherche par **troncature** : du type génér*: peut viser généreux, généreuse, général, générique, générer, etc.
- **Recherche par champ lexical approchant**
type thesaurus (le Roget, en anglais): ("À quels autres mots est en général associé ce terme ?")
- **Comparaison multilingue**
comment ceci a-t-il été traduit dans une autre langue : consultation d'une traduction tierce, éventuellement sous la forme d'une concordance multilingue (type TraduXio)
- **Consultation d'un glossaire**
("comment l'équipe de traducteurs a-t-elle convenu de traduire ce terme de façon fixe ?")
- **Consultation d'un lexique technique**, pour une spécialité
(extraction pétrolière, chirurgie dentaire, prosodie grecque, etc.)
- **Consultation d'une grammaire**,
pour vérifier les règles de syntaxe, de conjugaison, etc.
- **Consultation d'un manuel de "bon usage"**
de type Grévisse, recensant les erreurs de rédaction les plus fréquentes ; ces manuels peuvent aussi concerner la traduction (pour éviter les calques, les faux amis, etc.)
- **Traquer les erreurs éventuelles**
doublons, fautes d'orthographe et de grammaire : ceci peut se faire par assistance automatique
- **Garder en mémoire/sauvegarder**
ce serait intéressant d'avoir aussi un historique des modifications, afin de pouvoir revenir à une version antérieure
- La **traduction et la rédaction automatique de textes** ou de fragments de textes.

LES OUTILS D'AIDE À LA TRADUCTION - POUR LES TRADUCTEURS

MateCat



MateCat est un **outil de TAO en ligne gratuit et open source**, développé par la FBK (Fondazione Bruno Kessler), la société de traduction Translated srl, l'Université du Maine et l'Université d'Édimbourg.

Cet outil a vu le jour en 2014, après 3 années de recherches, financé par le septième programme-cadre de l'Union européenne pour la recherche, le développement technologique et la démonstration.

La philosophie de cet outil est de rester gratuit et en licence open source pour **partager les avancées technologiques avec tous ceux qui peuvent y contribuer ou en bénéficier**.

De plus : *Matecat envisage un avenir où les machines et les humains travaillent ensemble. Les machines prendront en charge les tâches répétitives et monotones et les humains se concentreront sur les parties créatives et plus complexes du processus de traduction. (> issu du site internet de l'outil)*

En effet, cette plateforme de traduction est construite sur l'**apprentissage automatique et l'analyse de données volumineuses**. L'objectif est d'identifier les tâches qui peuvent être automatisées et effectuées par une machine, de la création de projet à la post traduction.

Que permet cet outil ?

Les grandes fonctionnalités :

- **La mise en place d'un projet** : son nom, et la paire de langues sont obligatoires. Des options sont possibles : ajout de ressources personnelles (glossaire, mémoire de

traduction...), les moteurs utilisés en traduction automatisé (il s'agit de google traduction par défaut), le type de segmentation. Il est également possible de choisir l'utilisation de Lexica, un outil d'assurance qualité linguistique.

- **L'analyse du projet** : une fois les paramètres établis et le/les documents chargés, l'outil fait une analyse (nombre de mots, partie effectuée par un TA...). A ce stade, le projet peut déjà être partagé avec d'autres traducteurs (jusqu'à 50)
- **La traduction du projet** : il est possible, comme dans les autres logiciels de TAO, de voir les correspondances avec les ressources qui ont été ajoutées et d'insérer des termes dans un glossaire. Il est également possible de télécharger un aperçu à n'importe quel moment pour s'assurer de la bonne mise en page du fichier cible, mais aussi de télécharger un fichier bilingue .xliff afin de continuer la traduction hors ligne. Si le projet a été partagé, il est possible de laisser des commentaires et suivre les actions de traduction.
- **Étapes de post-traduction** : si l'option d'assurance qualité linguistique a été cochée, il est possible de corriger les différents problèmes mis en évidence par un code couleur et accéder à un rapport d'assurance qualité. Il s'agit d'un plug-in Lexica intégré. [Non testé]

Autres fonctionnalités :

- **Révision de la traduction automatique** : les traducteurs peuvent noter la qualité de la traduction automatique en indiquant le type et la gravité des erreurs. Ces catégories à cocher calculeront un « score » obtenu par le traducteur quant à la qualité de sa traduction. [Non testé]
- **«Externalisation» de la traduction** : Si un traducteur manque de temps pour un projet, la plateforme propose un réseau de traducteurs à contacter pour externaliser la production. Il semble que l'outil de facturation et de commande de traduction soit directement intégré à la plateforme. [Non testé]
- **Traduction d'un texte dicté**, une fonctionnalité appréciée par certains traducteurs, permettant un usage spontané et instantané, notamment lors de la rédaction de «premiers jets».
- [Margaux, Pirog, « J'ai testé pour vous : MateCat », dans le Blog du Master «Traduction Spécialisée Multilingue» (TSM) de l'Université de Lille] **Guess tag position** : tous les traducteurs seront d'accord pour dire que les tags sont très pénibles lorsque l'on traduit dans un logiciel de TAO (ils peuvent être nombreux, fractionnent les segments et il est indispensable de les réintégrer au bon endroit dans la cible). Sur MateCat, la fonctionnalité guess tag position vous permet de masquer les tags dans la langue source afin de voir l'intégralité de votre segment et de le traduire. Une fois votre segment traduit, cliquez sur le bouton Guess tags et un algorithme replacera les tags masqués en se basant sur leur position dans la source. Si le placement n'est pas correct, vous pouvez le modifier. N'est-ce pas merveilleux ?

Quels sont les points forts de Matecat ?

- La gratuité du service et la licence opensource (évolution continue des fonctionnalités et technologies)
- L'accessibilité : accéder n'importe où, n'importe quand à son travail et ses projets (via un page web safari ou google chrome), sur ordinateur et sur tablette.
- La collaboration facilitée entre pairs : le partage des fichiers et des projets est facilité. La mise en relation entre traducteurs est également favorisée.
- L'outil facilite la gestion de projet, avec une interface claire dédiée à la gestion.
- La pédagogie : tutoriels youtube de démonstration de l'outil, articles avec screenshots et modes d'emploi ; tout est fait pour accompagner l'utilisateur.
- Une assistance technique gratuite est assurée par mail en semaine, de 9h à 19h.
- La possibilité de masquer les étiquettes («tags») lors de la traduction
- La possibilité d'inclure un ou plusieurs moteur(s) de traduction automatique (TA)
- Utilisation de correspondances, de glossaires

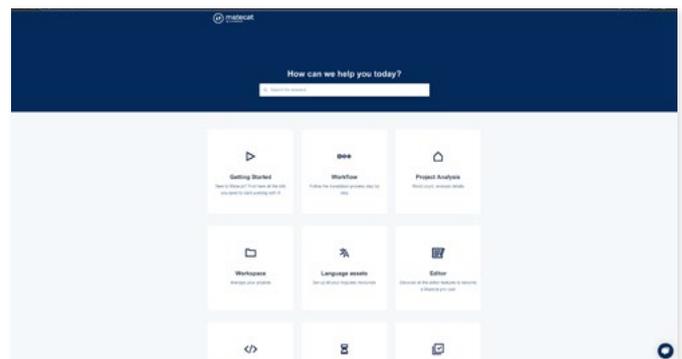
Quels sont les points faibles de Matecat ?

- L'outil ne possède pas d'extension ou de plug in pour être utilisé directement dans Microsoft Word ou d'autres logiciels de PAO.
- L'outil ne permet pas d'usage hors ligne.
- Fonctionne selon les principes un peu rigides du TALN: couple de langues, segmentation par phrases, insistance sur les règles au détriment des singularités

Quelques extraits de screenshots :



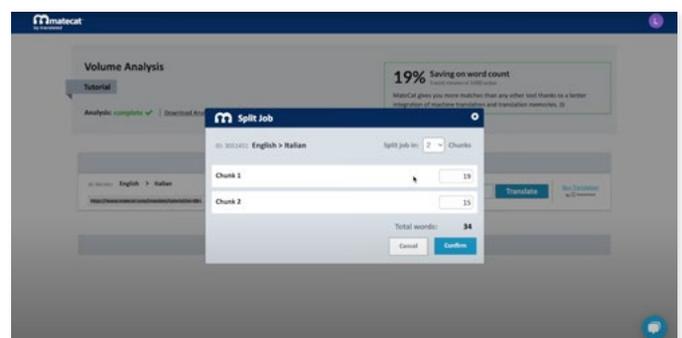
^ Vue des formats de document pris en charge par la plate-forme



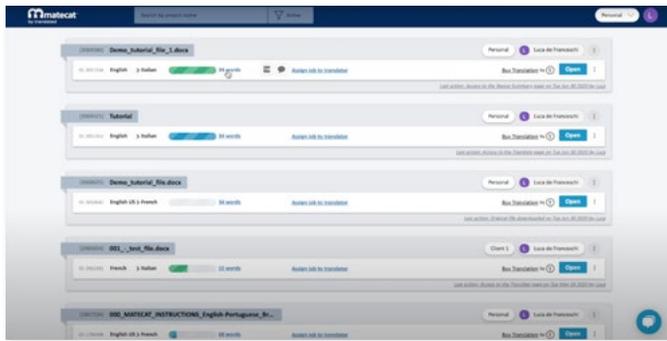
^ Vue d'une page «mode d'emploi» permettant de guider les usagers de la plate-forme, les modes d'emploi comprennent des screenshots détaillés. De plus, le chat visible en bas à droite de l'écran est présent sur toutes les pages et permet d'accéder à des ressources d'aide à tout moment du parcours.



^ Vue d'analyse de qualité linguistique, pour la post traduction.



^ Vue de la séparation d'un projet en plusieurs parties pour les attribuer à des collaborateurs.



^ Vue du dashboard de suivi et gestion de projets de traduction.

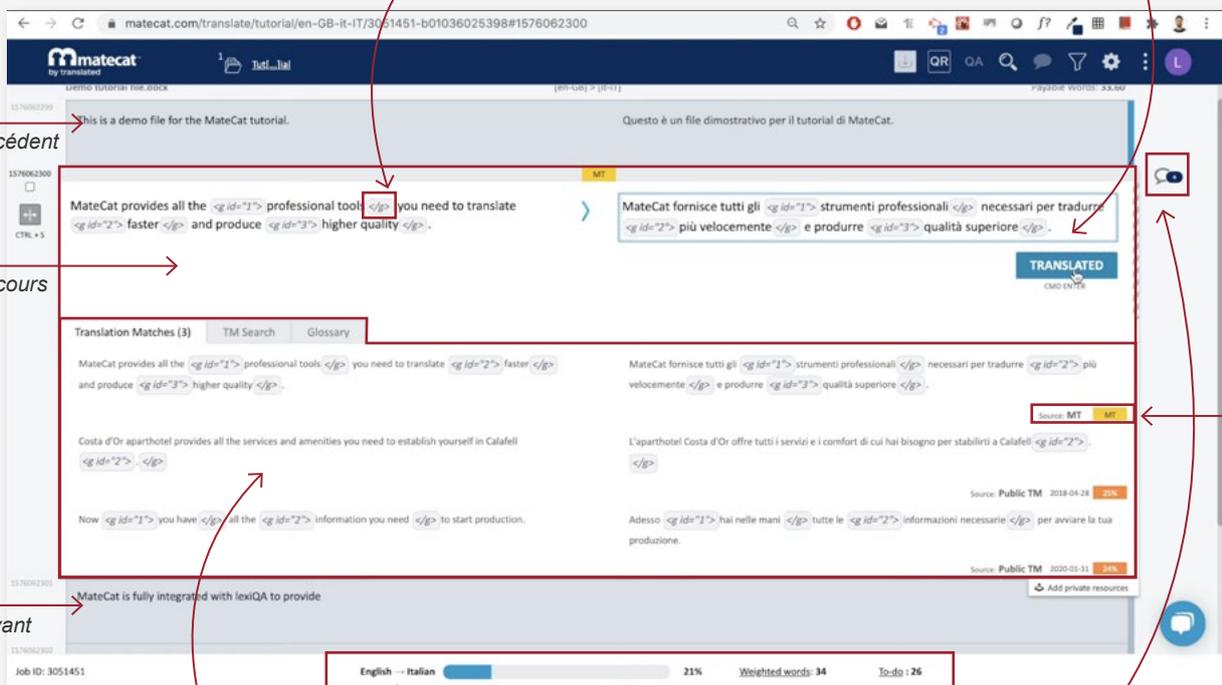
MateCat apparaît donc comme un outil en ligne gratuit ayant simplifié et recentré sur l'essentiel les fonctionnalités présentes dans les logiciels historiques de TAO ; intégrant les usages du cloud, notamment le nomadisme et la collaboration simplifiée et fluidifiée entre collaborateurs.

Focus : l'interface de traduction de MateCat

Zone d'édition de la traduction :
Texte en langue cible.

Le traduction peut être préremplie à partir des données personnelles de traduction importées (mémoire de traduction, glossaire) ou de moteur MTN

Balise ou Etiquette : Eléments relatifs à la mise en forme du texte



Segment précédent

Segment en cours de traduction

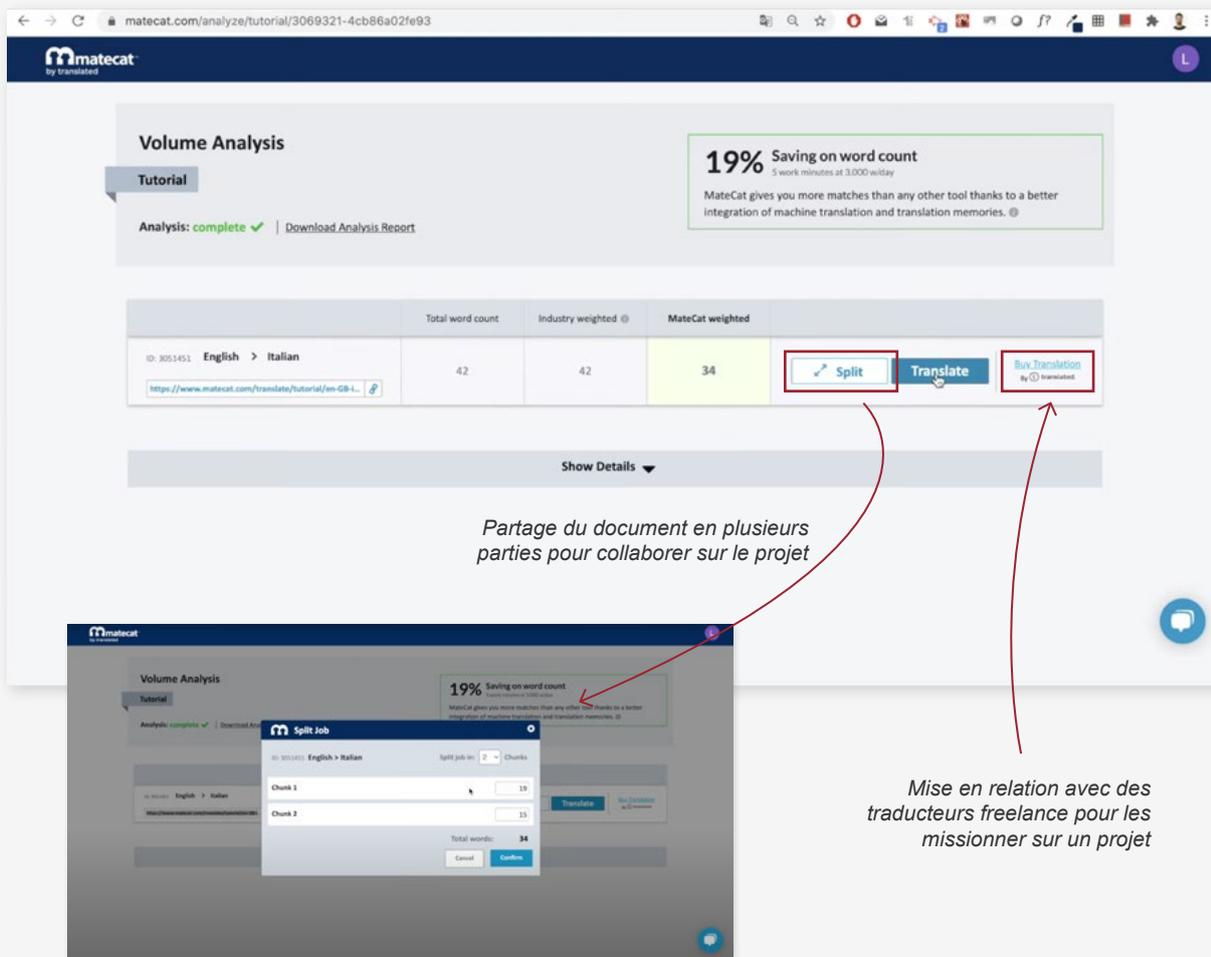
Segment suivant

Traçabilité de la source

Traduction provenant de la mémoire, du glossaire ou de moteur MTN
(toujours en vue bilingue, à gauche langue source / précédente traduction faite dans la langue cible)

Barre de progression / suivi de l'avancée de la traduction

Ajout de commentaires (notamment pour les documents partagés entre collaborateurs) :
Les commentaires sont relatifs à chaque segment, et sont présentés de la même manière que sur un PDF commenté.



Les éléments à retenir

Représentation :

- Un espace de travail structuré, avec une représentation segment par segment, rattachant directement les suggestions de traduction à l'espace d'édition.
- Prise en charge des balises afin de garantir une cohérence de mise en forme entre le fichier cible et source. (Une option permet de masquer/montrer les balises pour plus de lisibilité)
- La fonction d'aperçu est présente, mais demande de télécharger un fichier d'aperçu.

Suivi, analyse et traçabilité :

- Avant même d'accéder à l'interface de traduction, l'outil procède à une analyse indiquant le nombre de mots à traduire et le nombre de mots restant à traduire.
- Une barre de progression est présente en bas de l'interface de traduction permettant d'avoir une vision permanente de l'état d'avancement de la traduction.
- Des vues à l'échelle des projets et fichiers permettent de suivre les avancements, statuts et ainsi faciliter la gestion de projet.

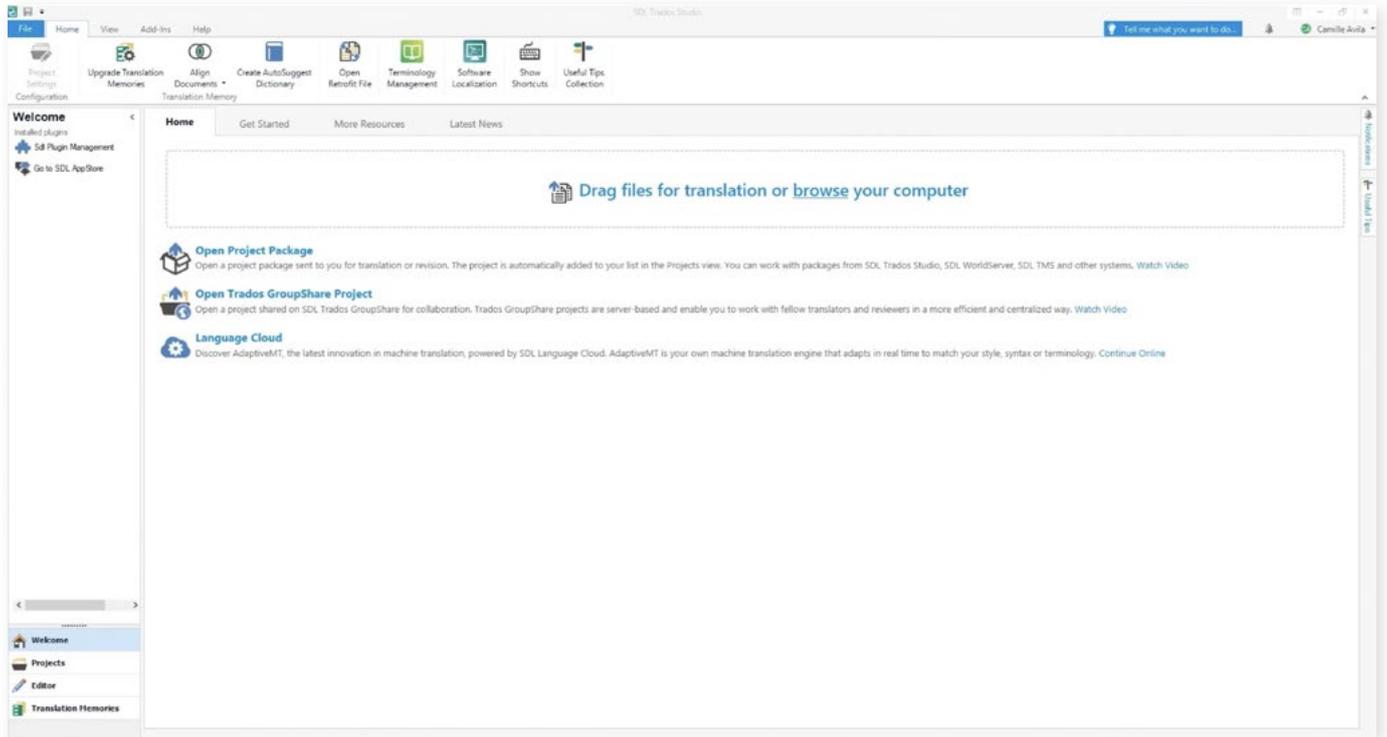
Aide à la traduction :

- Des suggestions de traduction issues de :
 - Mémoire de traduction
 - Bases terminologiques
 - Moteurs MTN (TA)
- Une vue des propositions de traduction issue de la mémoire de traduction, directement rattaché au segment en cours de traduction.

Collaboration :

- La première étape de la création du projet de traduction est l'analyse. A partir de cette étape, l'utilisateur peut diviser le fichier en plusieurs parties et inviter des collaborateurs à travailler sur une ou plusieurs parties du fichier. (Cf. ci-dessous)
- L'interface prévoit également à cette étape la possibilité de mettre en relation le traducteur avec d'autres traducteurs freelance, pour les missionner sur tout ou une partie du projet.

SDL Trados Studio



SDL Trados Studio est un logiciel de traduction assistée par ordinateur (TAO), **lancé pour la première fois en 1984. Depuis 2005 Trados fait partie du groupe SDL plc, un fournisseur de solutions cloud pour la gestion la gestion de contenu mondial et traduction.**

C'est un **leader du marché**, y compris chez les traducteurs indépendants, prestataires de services linguistiques, services internes de traduction et établissements universitaires.

Ce logiciel fait parti d'une suite, **SDL Trados, qui développe tout un écosystème de logiciels et services permettant d'accroître la productivité en traduction.**

Les autres produits de l'écosystème sont notamment :

- SDL MultiTerm : Un outil pour le stockage et la gestion de la terminologie multilingue.
- SDL Language Cloud : Une plate-forme cloud pour un accès sécurisé à la traduction automatique et manuelle, avec de nouveaux moteurs spécialisés, depuis SDL Trados Studio.

La philosophie de Trados, **l'outil métier de référence sur le marché**, est de proposer un **environnement de travail complet, fluide, permettant d'accompagner les traducteurs dans l'ensemble de leur processus / parcours de traduction.**

Que permet cet outil ?

Les grandes fonctionnalités de Trados :

- **Mémoires de traduction** : Une technologie de mémoire de traduction (MT) intelligente permet de réutiliser facilement du contenu déjà traduit et approuvé, afin d'augmenter la productivité et la cohérence dans la traduction.
- **Gestion terminologique** : Garantir un ton homogène dans toutes les langues en créant, en gérant et, si nécessaire, en partageant une terminologie approuvée.
- **Gestion de projets** : un module de gestion de projet est

intégré dans l'outil de traduction afin de pouvoir garantir une gestion efficace de projets, même complexes (multilingues / volumineux, partagés avec une diversité de collaborateurs...)

- **Traduction automatique** : L'outil permet d'accéder à la traduction automatique neuronale (NMT) de Language Weaver, l'outil de TA de la suite SDL Trados. Le traducteur peut également importer un moteur NMT qui lui est familier.

Quels sont les points forts de Trados Studio ?

- Une large **communauté de professionnel** : un outil de référence sur le marché, partagé par un grand nombre de traducteurs
- Permet certains usages **"plurilingues"** (par multiplication des couples de langues)
- Facilite les **usages collaboratifs** (exportation et partage de certaines mémoires de traduction ou glossaires)
- Une **assistance technique** efficace
- La **pédagogie** : des tutoriels échelonnés en fonction du niveau de prise en main du logiciel.
- La prise en charge d'un grand nombre de formats de fichiers y compris des fichiers issus de scan (pdf, images...) grâce à de la reconnaissance de caractères.

Quels sont les points faibles de Trados Studio ?

- L'outil est payant. (licence perpétuelle)
- Fonctionnement par couple de langues, principalement
- L'outil n'est pas disponible en version tablette / mobile, pour un usage plus nomade (avec une adaptation de l'interface et des fonctionnalités).
- L'outil n'est disponible que sur Windows

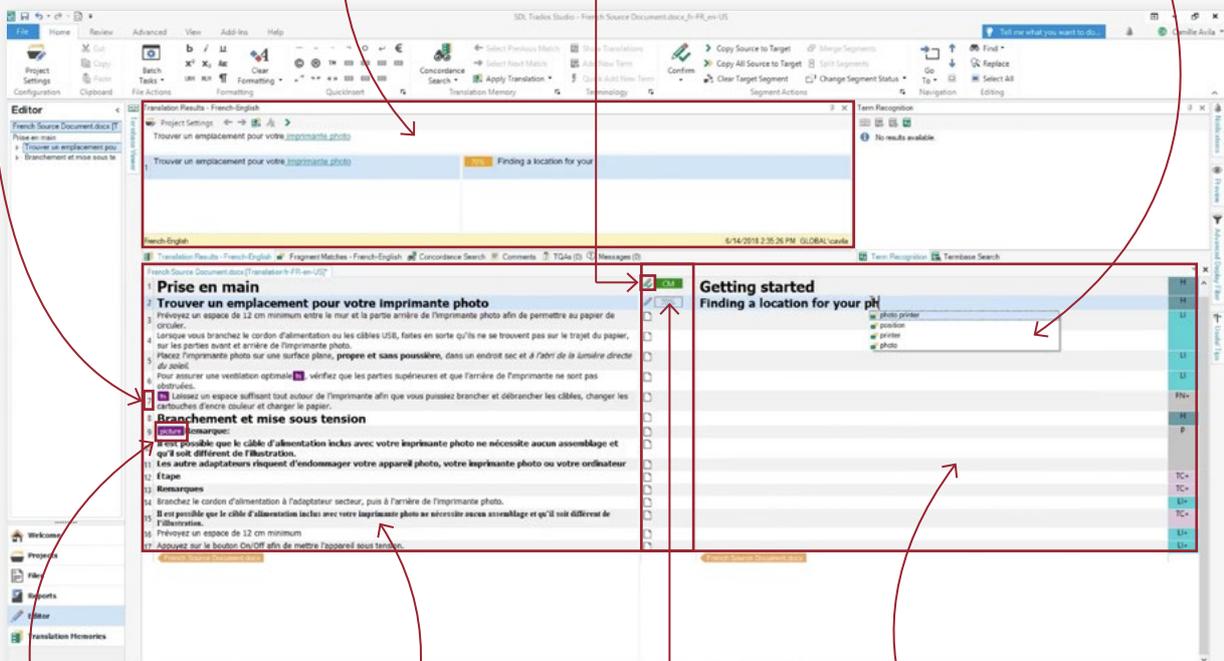
Focus : l'interface de traduction de Trados Studio

Traduction provenant de la mémoire, toujours en vue bilingue (à gauche langue source / à droite, précédente traduction faite dans la langue cible)

Numérotation des segments : cela peut être un point de repère important pour des documents volumineux, afin notamment de partager sa progression

Validation de la traduction : La coche verte indique que la traduction est validée (suivi du projet) mais également que cette traduction est alors importée dans la mémoire de traduction

Suggestions : ces éléments de suggestion proviennent des mémoires de traduction, bases terminologiques et même moteurs de traductions automatique importés



Balise ou Etiquette : Eléments relatifs à la mise en forme du texte (exemple ici : icône)

Espace principal de travail (1) : grille permettant de différencier chaque segment (qui sont numérotés).

Le texte source est découpé en segment (ici par phrase)

A gauche : langue source

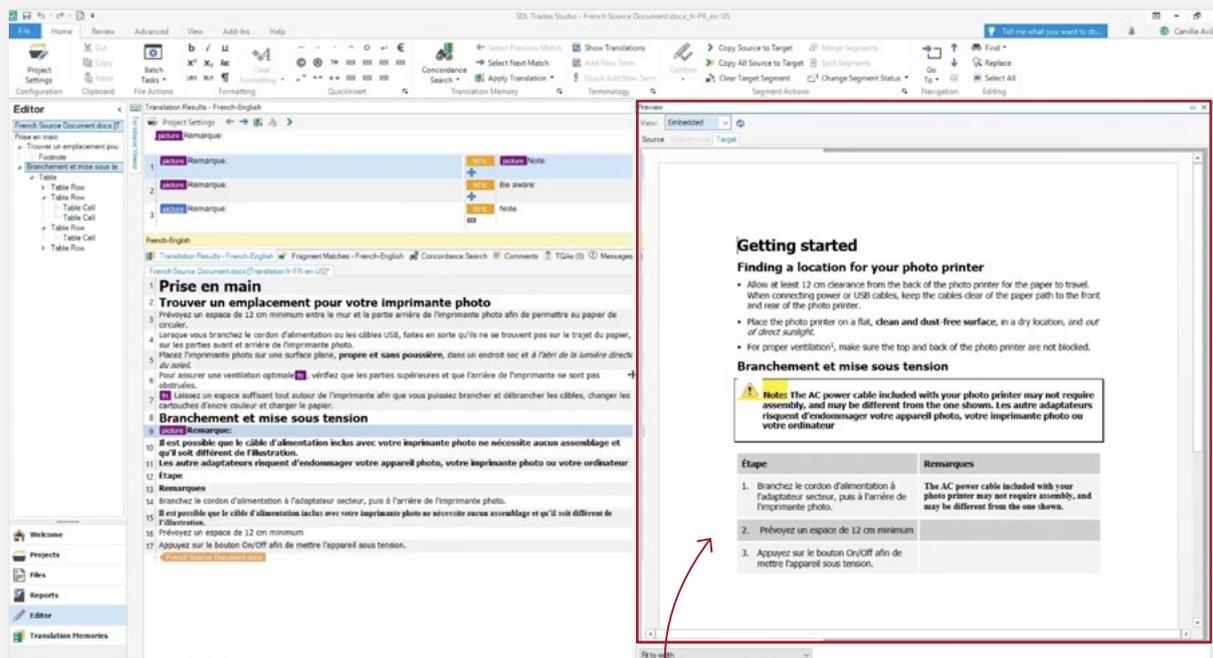
«Analyse» et traçabilité : Permet d'avoir une vision du pourcentage de traduction «pré traduite» à partir des données personnelles importées.

Indique le % de traduction pré remplie et son origine (Mémoire de traduction ou MTN)

Espace principal de travail (2) : Zone d'édition, de traduction -> Texte en langue cible.

Le traduction peut être pré-remplie à partir des données personnelles de traduction importées (mémoire de traduction, glossaire)

Focus : l'interface de traduction de Trados Studio



Fonction aperçu :

A tout moment, l'utilisateur peut avoir un aperçu du fichier mis en forme dans la langue cible. Cet aperçu s'actualise au fur et à mesure de la traduction des segments.

Les éléments à retenir

Représentation :

- Une vue synoptique des langues sources et cibles côte à côte
- Un découpage numéroté en segments
- Prise en charge des balises afin de garantir une cohérence de mise en forme entre le fichier cible et source.
- Une fonction d'aperçu, permettant de contrôler la mise en forme dans le texte cible.

Aide à la traduction :

- Des suggestions de traduction issues de :
 - Mémoire de traduction
 - Bases terminologiques
 - Moteurs MTN (TA)
- Une vue des propositions de traduction issue de la mémoire de traduction, au-dessus de la vue synoptique.

Suivi, analyse et traçabilité :

- Numérotation des segments permettant une vue rapide de l'état d'avancement d'une traduction
- Pour chaque fichier texte traduit, une analyse est disponible et indique le % de traduction pré-remplie à partir de données de traduction importées et l'origine de ces traductions.
- Des vue à l'échelle des projets et fichiers permettent de suivre les avancements, statuts et ainsi facilité la gestion de projet.

Omega-T

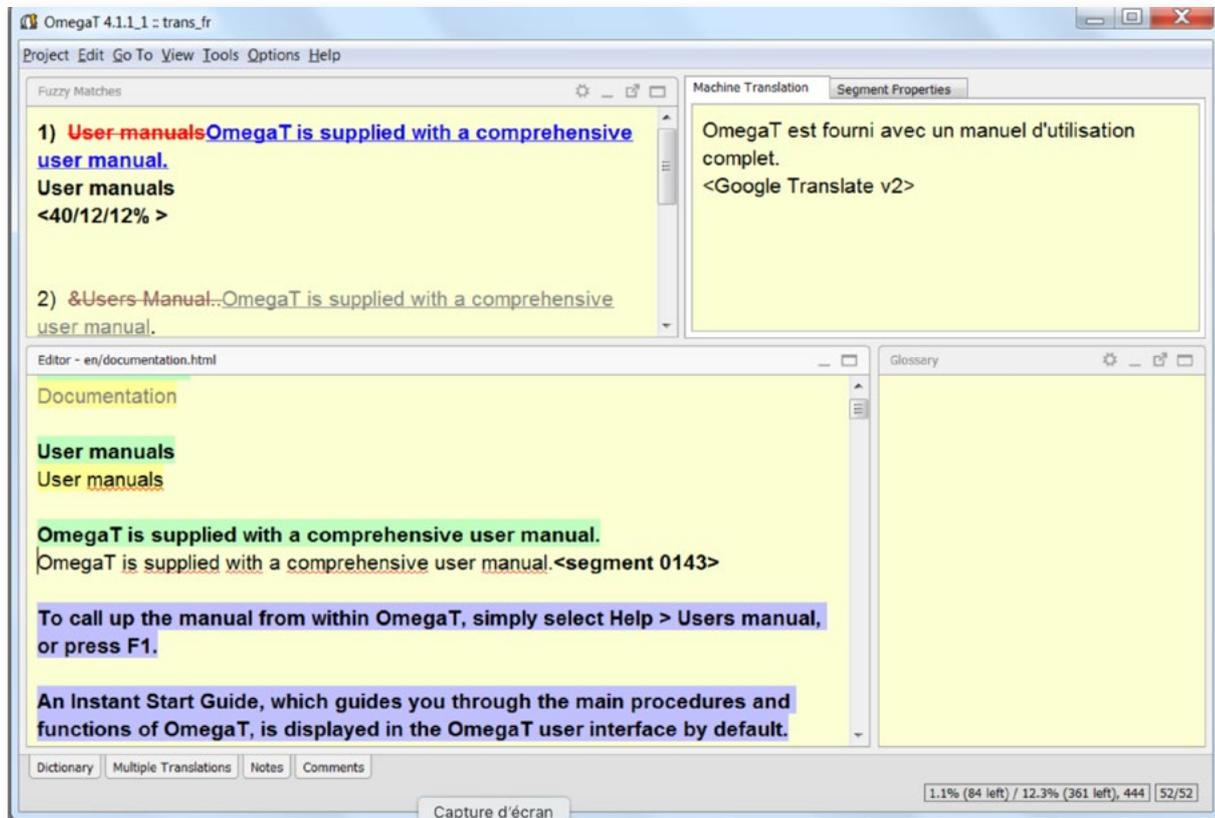
Outil de mémoire de traduction gratuit et open-source.

Avantages :

Gratuité et code modifiable pour réaliser des plug-ins.

Inconvénients :

Interface peu soignée et pas complètement intuitive.



TraduXio: “manuel”, multilingue, collaboratif

TraduXio est un environnement numérique pour la traduction collaborative et multilingue. D'utilisation gratuite et de code ouvert (licence GNU-GPL3), l'outil fait la promotion des licences ouvertes (de type Creative Commons) pour les contenus (les textes traduits), dans la perspective de la circulation des textes à l'époque de la science ouverte. Créé en 2006, grâce à des financements publics (ONU, Francophonie, Ministère Français de la Culture), il existe aujourd'hui dans une version 2.0, qui est utilisée par différents groupes de recherche dans le monde (Korea University, Inalco, ISIT, universités de Porto, Brasilia, Paris 8, Florence, NYC, etc).

Le choix en faveur du multilinguisme se traduit par une interface en forme d'accordéon, où les utilisateurs peuvent à volonté ouvrir et fermer autant de fenêtres/colonnes qu'ils le souhaitent, pour les diverses traductions considérées.

TraduXio fonctionne avec toutes les langues numérisées (sous format Unicode), y compris les langues sémitiques (s'écrivant de droite à gauche) ou asiatiques, contemporaines ou anciennes, y compris avec certaines qui ne sont même pas acceptées par le logiciel Word, comme le birman.

Sans aucune prétention à l'automatisation, et ne fonctionnant pas du tout selon les principes du TALN, l'environnement TraduXio est une technologie de l'assistance à l'interprétation qui

stimule l'intelligence collective grâce à la comparaison et aux suggestions pertinentes d'une concordance multilingue.

TraduXio promeut tout particulièrement le travail collaboratif, notamment grâce à la gestion des privilèges (de lecture, d'écriture), ou à la constitution de glossaires, particulièrement utiles pour les traductions de textes longs.

<p>Leur père, le duc de San-Sévéro, qui avait été longtemps favori du roi de Naples, tombé tout à coup dans la disgrâce, était mort de chagrin après avoir dissipé toute sa fortune. Césaró aurait souffert la misère avec courage s'il avait été seul à la supporter, mais il ne pouvait s'accoutumer à voir Thérésina, si belle, si fière, se servir elle-même, et s'imposer toutes sortes de privations. Il passait des nuits entières à se tourmenter l'es- prit pour trouver un moyen de gagner leur vie; c'est pourquoi ses joues étaient si pâles, quoiqu'il fût jeune et bien portant. L'idée de voir entrer sa sœur au couvent lui déchirait le cœur; car il savait que Thérésina faisait un grand sacrifice en prenant le voile, puisqu'elle n'avait point de vocation. Il n'avait plus qu'elle au monde, et pour elle, qu'il aimait tant, il aurait tout sacrifié.</p>	<p>Their father, the duke of San Severo, who had long been the king's favorite, had suddenly fallen into disfavor, died of grief after having lost his fortune. Césaró would have suffered misery with courage if he had been alone bearing it, but he couldn't get used to seeing Thérésina, so beautiful, so proud, herself in servitude? and enforcing all sorts of privations upon herself. He spent nights tormenting his spirit in order to find a way of earning a living. This is why his cheeks were so pale, although he was so young and in good health. The idea of seeing his sister in a convent broke his heart; for he knew that Thérésina was making a great sacrifice taking the veil, as she had no vocation. He only had her in the world, and for her, whom he loved so much, he would have sacrificed everything.</p>	<p>de Girardin, D. Contes d'une vieille fille à ses neveux Trad. Diana Thow 2019-02-15</p>
<p>Leur père, le duc de San-Sévéro, qui avait été longtemps favori du roi de Naples, tombé tout à coup dans la disgrâce, était mort de chagrin après avoir dissipé toute sa fortune. Césaró aurait souffert la misère avec courage s'il avait été seul à la supporter, mais il ne pouvait s'accoutumer à voir Thérésina, si belle, si fière, se servir elle-même, et s'imposer toutes sortes de privations. Il passait des nuits entières à se tourmenter l'es- prit pour trouver un moyen de gagner leur vie; c'est pourquoi ses joues étaient si pâles, quoiqu'il fût jeune et bien portant. L'idée de voir entrer sa sœur au couvent lui déchirait le cœur; car il savait que Thérésina faisait un grand sacrifice en prenant le voile, puisqu'elle n'avait point de vocation. Il n'avait plus qu'elle au monde, et pour elle, qu'il aimait tant, il aurait tout sacrifié.</p>	<p>died of grief after having lost his fortune</p>	<p>de Girardin, D. Contes d'une vieille fille à ses neveux Trad. Amanda GILBERT 2019-04-12</p>
<p>dans les brouillards de chagrin crus</p>	<p>in raw mists of sorrow</p>	<p>Gaston Miron LE DAMNED CANUCK Trad. Marc Plourde 1980</p>
<p>dans les brouillards de chagrin crus</p>	<p>비애의 거친 안개 안에서</p>	<p>Gaston Miron LE DAMNED CANUCK Trad. 한대균 2011</p>
<p>--- Elle est d'abord tourmentée sur l'abus de confiance de ses amis contre elle, mais après elle se rend compte de la pression inévitable et du chagrin d'amour qui s'amène à ses amis.</p>	<p>She agonizes, at first, over the breach of trust her friends committed against her, but then realizes the inescapable pressure and heartache informing brought to her friends.</p>	<p>Katherine Verdery My Life as a Spy: Investigations in a Secret Police File 2018 Trad. Haby Gadio 2019-03-22</p>
<p>--- Elle est d'abord tourmentée sur l'abus de confiance de ses amis contre elle, mais après elle se rend compte de la pression inévitable et du chagrin d'amour qui s'amène à ses amis.</p>	<p>Elle est d'abord tourmentée par la trahison de ses amis dans la confiance qu'elle leur portait, avant de réaliser la pression et la déchirure implacables que leurs délations provoquaient en eux.</p>	<p>Katherine Verdery My Life as a Spy: Investigations in a Secret Police File 2018 Trad. Haby Gadio 2019-03-22 Trad. Gabrielle Pauvert</p>

▲
exemple de résultat de recherche dans la concordance pour le terme de "chagrin"

DéjàVu par ATRIL

Un des grands acteurs de la TAO, après Trados, bien connu des traducteurs professionnels.

Avantages :

Fonctionnalités de mémoire de traduction, glossaires, collaboration (partage des mémoires de traduction).

Inconvénients :

Payant, code propriétaire, fonctionne par couples de langues; à télécharger et à installer sur son micro-ordinateur.

MemoQ

Outil de TAO compatible avec la suite Windows, fonctionne avec des mémoires de traduction et des glossaires.

Avantages :

Compatible avec différents fichiers, mais aussi interopérable avec divers logiciels de traduction automatique voire de TAO.

Inconvénients :

Payant, propriétaire ; fonctionne par couples de langues.

Microsoft Traduction

Microsoft Traduction est un outil de traduction multilingue fourni par Microsoft. L'API de Microsoft Traduction est intégrée dans de multiples produits, pour les développeurs et les entreprises.

Microsoft vient d'investir 10 milliards de dollars dans OpenAI, la start-up qui construit Chat-GPT.

Systran

(qui propose de rentrer par discipline)

The screenshot shows the SYSTRAN Translate website. At the top, there is a navigation bar with the SYSTRAN logo (tagline: beyond language) and links for SOLUTIONS, INDUSTRIES, PRODUITS & SERVICES, À PROPOS DE SYSTRAN, and CONTACT. There are also icons for SYSTRAN TRANSLATE, BLOG, SUPPORT, and LANGUAGES. The main heading is "TRADUCTION EN LIGNE". Below this, there is a form to "Choose source and target languages" with dropdown menus for "French" and "English". A text input field says "Type or paste your text here". A green "Translate" button is at the bottom of the form. To the right, there is a paragraph: "Pour toutes vos traductions en ligne, ayez le réflexe SYSTRAN ! Leader mondial de la traduction automatique, SYSTRAN vous permet d'effectuer toutes vos traductions gratuites en ligne grâce à SYSTRANet." Below this is a section titled "Le service de traduction en ligne SYSTRAN Translate" with a sub-heading "Véritable traducteur instantané, SYSTRAN Translate est un outil linguistique disponible gratuitement en ligne. Vous pouvez vous en servir comme d'un dictionnaire, mais aussi pour traduire une phrase ou même une page web complète. Ultra performant, SYSTRANet est un traducteur en ligne qui vous épaula au quotidien pour tous vos besoins en traduction simultanée de langue étrangère." Another section is titled "SYSTRAN et la traduction en ligne" with the text: "Avec plus de 40 ans d'expérience, SYSTRAN est devenu le spécialiste de la traduction instantanée. Au service de nombreuses multinationales et d'organisations publiques..."

Le kit du traducteur de Google

Cet outil a été proposé par Google entre 2009 et 2019, date à laquelle il a été retiré, car jugé peu utilisé, du fait notamment du succès des autres outils de traduction automatique (dont Google Translate). Il ajoutait à la fonction de traduction automatique, avec post-édition, la convocation de mémoire de traduction, dont on pouvait noter la pertinence des résultats. Ses limites étaient notamment liées à celle du TALN: fonctionnement par couple de langues, segmentation par phrase, traduction selon des règles ; en outre, tout ce qui est déposé sur le logiciel appartient à Google (aucune protection des données personnelles).

The screenshot shows the Google Translator Toolkit interface. At the top, it says "Google translator toolkit 0001-longtemps-je-me-suis-couche-de-bonne-heure" and "modified Aug 27, 2010 by philacour". There are buttons for "Share", "Show toolkit", "Edit", "View", "Save", and "Save & Close". The main area is split into two panes. The left pane is titled "Original text:" and contains a list of links, including "1281 - Les phrases de Vinteuil me firent penser à la petite phrase et je dis à Albertine" and "A LA RECHERCHE DU TEMPS PERDU - I - II - III - IV - V - VI - VII". Below this is the text "0001 Longtemps, je me suis couché de bonne heure" and a paragraph of French text. The right pane is titled "Translation: French » English" and shows "7% complete, 1739 words." The translated text is "Long, I went to bed early." Below this is a small window showing the original text "Long time, I went to bed early." and a "Use suggestion" button. There is also a "Glossary (0)" section with the text "No glossary matches available." At the bottom, there are sections for "Automatic Translation Search" and "Custom Translation Search". The "Automatic Translation Search" section shows "Translation Search Results" with a list of results, including "Longtemps, je me suis couché de bonne heure." and "For a long time, I went to bed early." The "Custom Translation Search" section shows "0001 Longtemps, je me suis couché de bonne heure" and "For a long time, I went to bed early."

LES CORPUS BILINGUES OU MONOLINGUES DISPONIBLES

Linguee



Linguee est un **dictionnaire bilingue gratuit en ligne, paru en 2010** ; proposant à l'internaute de comparer plusieurs textes en différentes langues afin de trouver la signification du mot ou de l'expression qu'il cherche. Linguee est le corpus à la base de l'outil de TA DeepL (Linguee appartient à la même entreprise).

En haut de l'écran, l'utilisateur accède au résultat de sa recherche : le mot ou l'expression recherchée traduit dans la langue cible. Souvent, le dictionnaire propose plusieurs résultats, du plus fréquent au plus rare.

En bas, l'utilisateur accède à plusieurs dizaines de textes bilingues issus du net contenant l'expression recherchée. Linguee n'est pas un traducteur automatique. Le contenu des textes bilingues est en totalité une production humaine, donc non robotique.

Fonctionnement :

Un programme détecte les sites bilingues, souvent des textes du parlement européen, des brevets et de l'Unesco. Le contenu traduit va dans la base de données de Linguee, puis un algorithme d'apprentissage automatique juge la qualité et fait un tri selon le niveau de qualité.

Les grandes fonctionnalités :

- La **sélection de langues** (21 langues cibles, si l'utilisateur utilise le français en langue d'origine).
- La fonction de **dictionnaire bilingue** : des résultats de

recherches permettant de **traduire un mot**, d'avoir sa **définition** et un **exemple d'utilisation** dans une phrase.

- La **prononciation enregistrée** des termes recherchés.
- L'accès à un **corpus de textes bilingues** permettant de consulter des expressions / mots traduits dans le contexte d'un domaine d'expertise.
- La **redirection vers la plateforme de traduction** automatique DeepL. L'interface propose plusieurs boutons pour «basculer» sur l'outil de TA de l'entreprise, permettant ainsi de «jongler» entre les deux usages.

Quels sont les points forts de Linguee ?

- L'accès à un **corpus de textes bilingues afin de pouvoir affiner le résultat de recherche du dictionnaire**.
- Le dictionnaire possède une **application pour tablettes et mobiles, permettant d'utiliser le dictionnaire hors ligne**.



< Vue : choix du couple de langues dans le dictionnaire Linguee

Le nombre et les langues cibles disponibles ne sont pas les mêmes en fonction de la langue source.

demander (qqch.) *verbe*

ask v (asked, asked)
 Le juge a demandé au témoin de parler. The judge asked the witness to speak.
 Le logiciel m'a demandé de saisir mon mot de passe. The software asked me to enter my password.

request v (requested, requested)
 Le client a demandé plus d'informations sur le produit. The customer requested more information on the product.

ask for v
 L'étudiant a demandé la permission de parler. The student asked for permission to speak.
 Mon ami avait soif et a demandé un verre d'eau. My friend was thirsty and asked for a glass of water.
 Nous avons demandé à l'expert une analyse complète. We asked the expert for a comprehensive analysis.

require v (required, required)
 Conduire de nuit demande de la vigilance. Driving at night requires vigilance.
 Apprendre une langue demande une pratique constante. Learning a language requires constant practice.

< Vue : présentation des résultats de recherche pour la partie dictionnaire

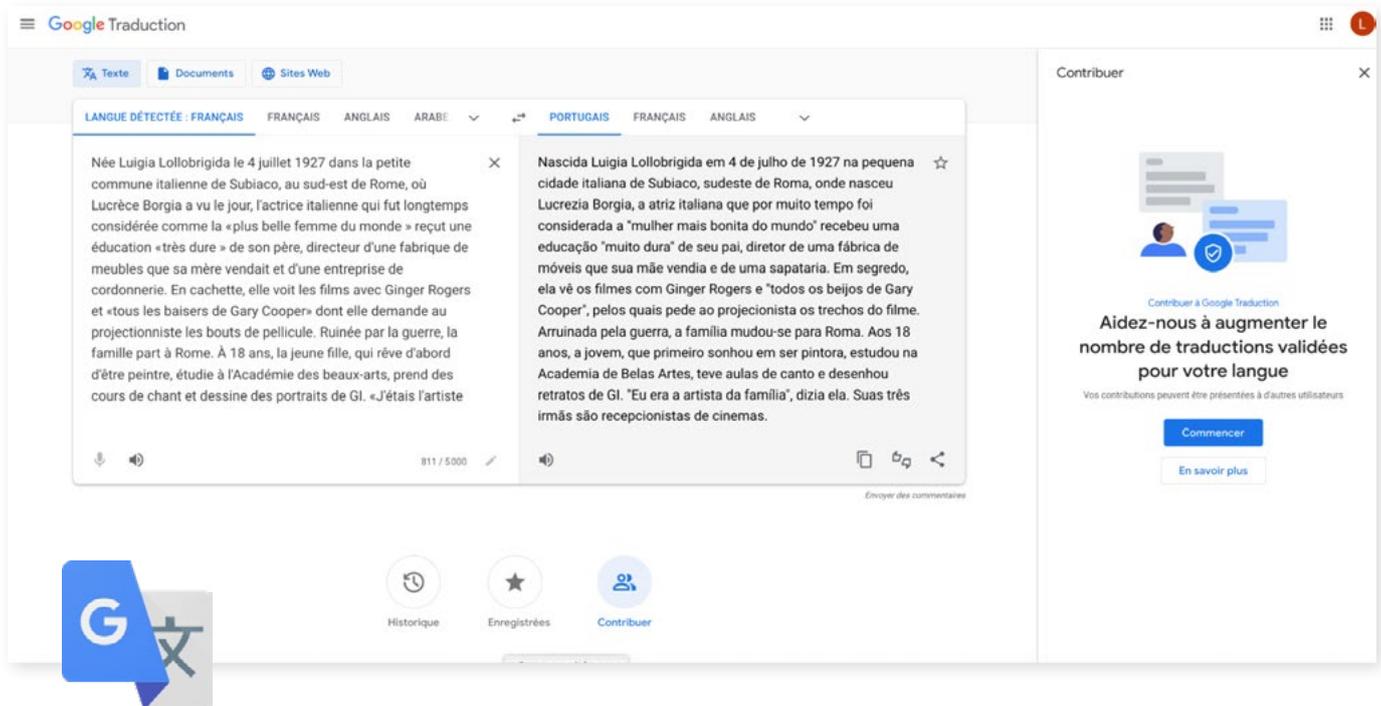
^ Sources externes (non révisées)

[...] ont été établis que le mot comitologie a été inventé pour décrire ce système si difficile à comprendre et à appréhender. ↳ europarl.europa.eu	[...] name comitology was invented to describe a system that was so difficult to fathom and so difficult to get to grips with. ↳ europarl.europa.eu
Comprendre les objectifs du programme de travail est essentiel pour la préparation d'une proposition de qualité. ↳ cordis.europa.eu	Understanding the objectives of the work programme is essential for preparing a good proposal. ↳ cordis.europa.eu
Un groupe ICC peut également comprendre des comptes MP qui sont compris dans un groupe CL. ↳ ecb.europa.eu	A CAI group may also include PM accounts which are included in an AL group. ↳ ecb.europa.eu
[...] substances doivent être enregistrées, il est important de bien comprendre ces termes ainsi que la façon dont sont traités les préparations et les articles. ↳ guidance.echa.europa.eu	[...] require registration it is important to have a clear understanding of these terms and how preparations and articles are dealt with. ↳ guidance.echa.europa.eu
De telles dispositions devraient pouvoir comprendre les règles nationales qui prévoient qu'en l'absence de réponse de l'autorité [...]. ↳ eur-lex.europa.eu	Such different arrangements could include national rules according to which, in the absence of a response of the competent [...]. ↳ eur-lex.europa.eu
Pour beaucoup d'entre nous, la guerre est un phénomène difficile, voire impossible, à comprendre . ↳ www2.parl.gc.ca	For many of us, war is a phenomenon that is difficult, if not impossible, to comprehend . ↳ www2.parl.gc.ca
Je ne suis pas bien sûr de comprendre de quelle transparence il veut parler. ↳ daccess-ods.un.org	I am not quite sure exactly what sort of transparency he has in mind. ↳ daccess-ods.un.org
J'espère que mes remarques aujourd'hui contribueront à mieux faire comprendre un sujet qui présente un intérêt de plus en plus grand pour votre travail. ↳ edps.europa.eu	I trust that my remarks today will contribute to a better understanding of a subject that is increasingly relevant for your work. ↳ edps.europa.eu

< Vue : présentation du corpus de textes bilingues utilisant le mot recherché dans le dictionnaire

LES OUTILS DE TRADUCTION AUTOMATIQUE

Google Traduction



Google Traduction est un service de **traduction automatique gratuit fourni par Google et lancé en 2006**.

C'est une des références de la traduction automatique depuis les années 2000. L'interface est simple, intuitive, rapide et relativement efficace. Google a repris une technique des années 1970 (la traduction par règles statistiques), en a amélioré les résultats grâce à la masse des données moissonnées (sur Google Books, les différents services Google, et même le Web lui-même). Concurrencé par DeepL dans les années 2010, Google a su s'adapter aux modèles neuronaux en les intégrant à son service déjà existant. La différence entre les deux (l'avantage de DeepL) tient sans doute au travail humain caché (de nettoyage et correction) chez DeepL. Mais Google propose beaucoup plus de langues de traductions que DeepL (ourdou, tamoul, etc., avec des efforts de recherche pour les langues africaines). Certaines fonctionnalités permettent d'affiner la traduction (consultation de dictionnaires inclus par double-clic, historique de vos traductions faites sur la plateforme, collaboration par validation des traductions). Très puissant, le service permet "d'aspirer" directement des documents, des pages Web, voire des sites Web entiers, et de partager rapidement ses traductions. Toutes les données insérées dans l'interface sont soumises aux CGU Google (qui peut les réutiliser).

La philosophie :

Etre **LE traducteur automatique à l'échelle mondiale** (avec ses 133 langues disponibles à la traduction). C'est aujourd'hui l'un des outils Google les plus utilisés au monde.

Que permet cet outil ?

Les grandes fonctionnalités :

- La **traduction automatique de texte** (jusqu'à 5000 caractères), mais aussi **de documents** (PDF, word..) et **de contenus de site web**.
- La traduction automatique **à partir de texte dicté** (audio)
- **L'écoute de la prononciation** de la traduction dans la langue cible.
- Un **dictionnaire** est affiché sous le résultat de la traduction automatique, avec une proposition de résultats alternatifs pour la langue cible.
- La possibilité d'**enregistrer une traduction** pour avoir accès à un historique.
- Des fonctions de **copie** et de **partage «en un clic»** de la traduction.
- Une fonction d'**évaluation de la traduction par les usagers**.
- Une **extension de navigateur** «Google Translate» permet de traduire instantanément une page web.
- Des **applications pour tous les devices** (ordinateurs, tablettes, smartphones)

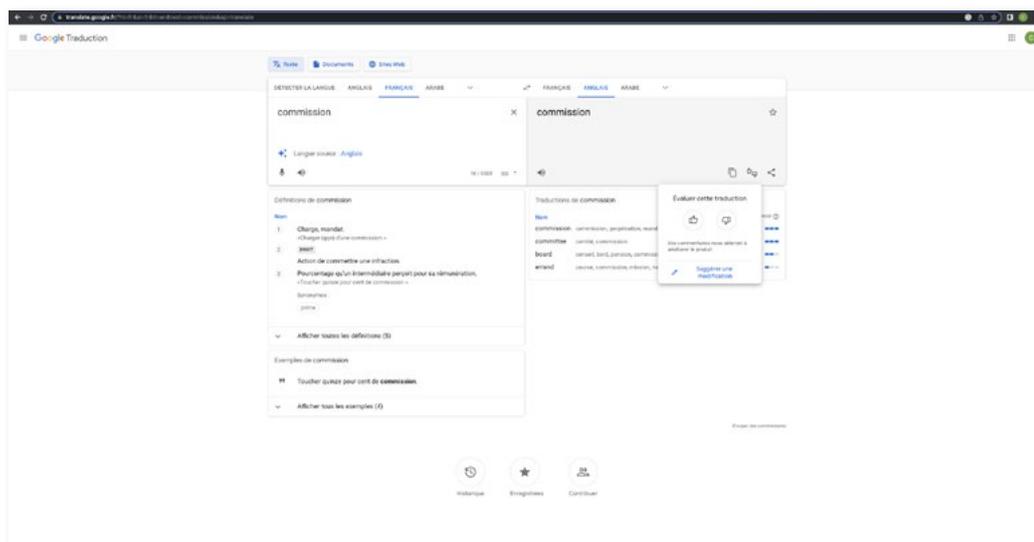
- La possibilité de «télécharger des langues» pour avoir accès au service de traduction via l'application en mode hors connexion.

Quels sont les points forts de Google traduction ?

- **133 langues** sont disponibles à la traduction depuis le 5 décembre 2022.
- Une **interface intuitive**, facile à prendre en main, accessible.
- La traduction instantanée de contenus issus de page web, même très longs.
- Un service pouvant être utilisé hors connexion, sur une multitude de supports numériques.

Quels sont les points faibles de Google traduction ?

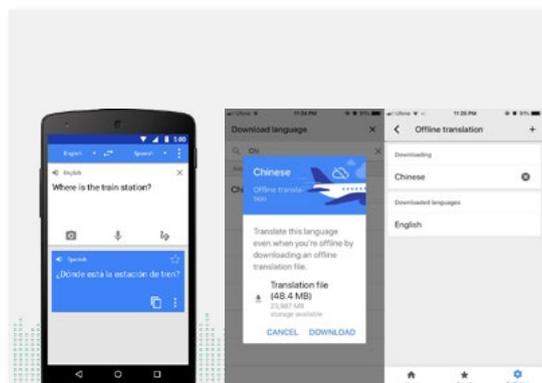
- Des **traductions parfois approximatives**, ou «mot à mot», des fautes grammaticales qui se font particulièrement ressentir lorsque les textes sont longs ou ont un certain niveau de technicité.
- Des **traductions non sécurisées**. Si un projet est confidentiel, on ne peut pas se servir de cet outil de traduction automatique, puisque les données sont récupérées par Google.



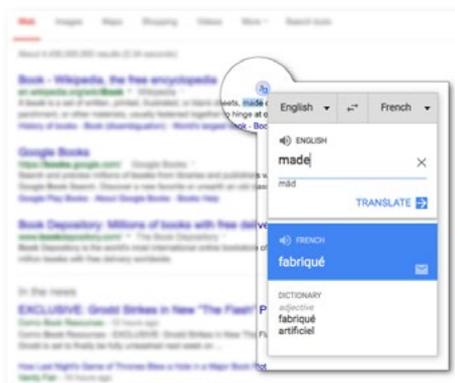
< Vue d'un résultat de traduction automatique de Google Traduction.

En partie haute, la traduction bilingue ; en partie basse le dictionnaire.

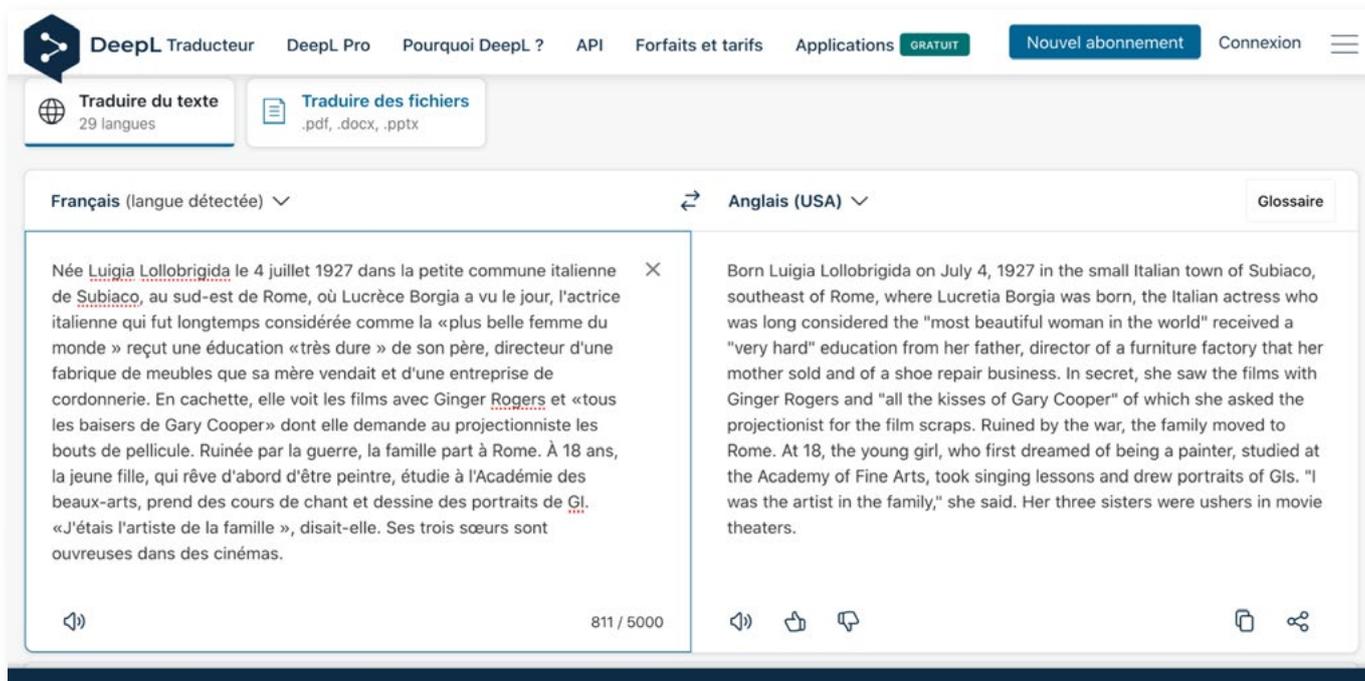
De plus, la popup montre la fonction d'évaluation de la traduction par les usagers.



^ Vue de l'application mobile de google traduction et de la fonctionnalité permettant d'utiliser l'outil hors connexion.



^ Vue d'un usage de l'extension Google Traduction permettant de surligner un mot sur un résultat web et de le traduire instantanément via une popup.



DeepL est un service de traduction automatique en ligne de la société DeepL GmbH, lancé en **2017** par l'**équipe de Linguee**. Le service permet de traduire **25 langues**.

Fonctionnement :

Ce service utilise des réseaux de neurones convolutifs constitués sur la base de données Linguee.

La philosophie :

Etre **LE traducteur automatique le plus précis et le plus subtil grâce à une IA de pointe**. Proposer une qualité de traduction inégalée pour un outil de TA.

Que permet cet outil ?

Les grandes fonctionnalités :

- La **traduction automatique de texte** (jusqu'à 5000 caractères en version gratuite), mais aussi **de documents** (PDF, word..).
- La possibilité de créer un **glossaire même en version gratuite** (10 entrées)
- L'**écoute de la prononciation** de la traduction dans la langue cible.
- L'accès en survolant la traduction en langue cible, au **dictionnaire intégré Linguee**.
- La possibilité de **choisir le «ton adéquat»** (registre formel ou informel), pour la version payante
- Des fonctions de **copie** et de **partage «en un clic»** de la traduction.
- Une fonction d'**évaluation de la traduction par les usagers**.
- Des **applications pour tous les devices** (ordinateurs, tablettes, smartphones)
- Une API pour intégrer l'outil de TA au sein de produits et plateformes

- La possibilité d'intégrer DeepL aux outils de TAO, proposant ainsi une solution tout en un pour les traducteurs professionnels, pour une expérience «sans couture»

Quels sont les points forts de DeepL ?

- La qualité de la traduction automatique, notamment pour les traductions de ou vers l'anglais, l'allemand, le français.
- Une **interface intuitive**, facile à prendre en main, accessible.
- Une version gratuite et payante, qui permet d'adapter les fonctionnalités à des usages professionnels (sécurisation des données, glossaire sans restriction, pas de restriction de nombre de caractère)

Quels sont les points faibles de DeepL ?

- Un nombre de langues limité.
- Il n'est aujourd'hui pas possible de traduire instantanément le contenu d'une page web.
- Pour la version gratuite, les données peuvent être récupérées par DeepL.
- Outil propriétaire
- Fonctionne (comme Linguee) sur le corpus des documents de l'Union européenne, avec de bons résultats de et vers l'anglais, mais une qualité moindre depuis et vers les autres langues
- Aucune information n'est donnée quant aux méthodes utilisées pour obtenir les résultats (aucune transparence)

Les outils de génération de textes

Ces outils, de type GPT3 (et bientôt GPT4) sont des technologies d'intelligence artificielle basées sur de très grands modèles de langage (plusieurs milliards de paramètres). Ils sont capables de générer du texte ("répondre" à des questions), y compris traduire. Issus de GPT3, testé et rectifié par des humains pendant trois ans, ChatGPT défraie aujourd'hui la chronique pour ses capacités d'analyse et de réponse (relativement) pertinentes, quel que soit le domaine envisagé, d'une question technique d'informatique ou de droit à l'écriture d'une dissertation. GPT 3 est développé par la société OpenAI, a été annoncé le 28 mai 2020 et ouvert aux utilisateurs via l'API d'OpenAI en juillet 2020.

DeepL vient de lancer (en version test, dite "bêta") son propre générateur de texte pour l'aide à l'écriture en langue étrangère :

https://www.deepl.com/write?utm_campaign=ROL_DeepLWrite_PS_TrPr_eml_sgs_PR_M01_v01&utm_medium=email&_hsmi=241863928&hsenc=p2ANqtz-8fWHG5sY4EYXytl2L2ZvOT-dlqeoP1WgDLVXKbWHaz5wfZSYcwNG37V62S0jkhT-CiBAeUYtVB439hi_XPnR7IBoD0qGtYjZmNcwmoB6hr-jGwIYTcfk&utm_content=241863926&utm_source=hs_email

Ces outils récents permettent d'imaginer des situations où l'outil pourrait générer un résumé à partir d'un article et de mots clés renseignés par le chercheur dans HAL, que le chercheur pourrait corriger et autoriser à la publication, et qui proposerait aussi la traduction multilingue de ce résumé.

La start-up française Light-On (<https://lighton.ai/about-us/>) travaille exactement dans cette perspective, privilégiant une approche multilingue et désireuse de s'associer à des projets de recherche.

La startup Hugging Face (au départ française, aujourd'hui très largement internationale) a réussi à rallier un millier de chercheurs et chercheuses pour proposer, à partir des usages du supercalculateur Jean Zay, un projet de mégascience, lequel a permis de développer un nouveau très grand modèle de langage naturel (170 milliards de paramètres): BLOOM (BigScience Large Open-science Open-access Multilingual Language Model). Cette solution « open science » a été développée grâce aux moyens publics français du ministère de l'enseignement supérieur et de la recherche et du CNRS. La qualité des résultats semble toutefois inférieure à celle de LightOn. Par ailleurs, les données insérées dans le modèle (textes traduits, corpora) sont par définition (open science) rendues disponibles à tous, y compris les concurrents potentiels du service de traduction scientifique à venir.

Open Calais, développé par Reuters il y a une dizaine d'années et désormais disponible sur Refinitiv, permet d'extraire les grandes lignes et les points saillants d'un article. Il extrait des informations clés d'un texte sous format « Resource Description Framework » (lieux, dates, personnes, institutions, mots-clés etc.).

LES AUTRES OUTILS

Zotero

Zotero fait partie des outils dont se sert un traducteur pour traduire les références (par exemple mention de la version américaine d'un livre de Foucault).

Pour préciser, Zotero est un outil de constitution de bibliographie par insertion manuelle, semi-automatique ou automatique de références. C'est le modèle du "butinage" qui fournit le miel de sa propre base bibliographique. Très pratique (pour automatiser la mise aux normes des différentes feuilles de style des divers éditeurs), facile d'utilisation, open source, c'est un outil de plus en plus répandu, et qui permet aussi de partager des bibliographies de façon collaborative.

Open Journal System

Système de gestion de contenu destiné tout particulièrement à la publications de revue scientifiques universitaires. Equivalent international de, par exemple, Lodel (créé et utilisé par OpenEdition). Gratuit et en code ouvert; promotion de l'accès libre et de la science ouverte. Utilisé notamment au Brésil (cf en bas de cette page: <https://periodicos.unb.br/index.php/fmc/index>)

Scopus

Base de données de citations et de résumés ; moteur de recherche intelligent et système de mise en relation. Groupe Elsevier: accès gratuit, logiciel propriétaire

CONCLUSION DU RAPPORT : PREMIÈRES ORIENTATIONS DU FUTUR SERVICE DE TRADUCTION OUTILLÉE

a_Incarnation du service de Traduction outillée : un logiciel et ses extensions, les fonctionnalités envisagées

Un environnement numérique de traduction outillée:

1. qui facilite le geste de traduction, de façon collaborative (dans le cloud) avec des fonctionnalités comme la vue synoptique, les mémoires de traduction et glossaire.
2. Multilingue d'emblée, même si n'utilise que deux langues au début
3. qui permet la rédaction directement dans la langue cible et la relecture par un traducteur

Les extensions possibles

- un moteur (de traduction automatique) spécialisé par discipline à intégrer dans les autres logiciels de TAO
- une version simplifiée de l'outil de traduction outillée, à l'image de la version grand public de DeepL ou Google Translate
- des plug-ins pour traduire automatiquement les résumés (titres et mots-clés) dans HAL ou d'autres plateformes et revues

b_Les points critiques à prendre en compte / les questions à investiguer pour la suite

Les questions juridiques

Ce premier travail d'exploration a permis de mettre le doigt sur des enjeux juridiques qui vont potentiellement affecter les fonctionnalités proposées par le futur service de traduction.

- *A-t-on le droit de charger un texte dont on n'est pas l'auteur dans un logiciel de traduction automatique?*
- *Dans la mesure où une traduction humaine est une œuvre de l'esprit, peut-on garder, ou à défaut exploiter, les mémoires de traduction qui sont générées par les utilisateurs de l'outil ? (Un projet de traduction aboutit à la constitution d'une mémoire de traduction constituée par le texte source et le texte cible).*
- *Faut-il passer un accord particulier avec l'auteur de la traduction (par exemple: "aidez la communauté de la recherche scientifique à proposer des traductions plus pertinentes en nous autorisant à partager et utiliser vos mémoires de traduction") ?*
→ EXPLORER LA PISTE DU DROIT DE CITATION COURTE ET LES "WAIVERS" DU TYPE CREATIVE COMMONS ZERO (<https://creativecommons.org/publicdomain/zero/1.0/deed.fr>)
- *Peut-on traduire automatiquement des résumés d'articles ? Dans quelles conditions ? (acceptation tacite ou validation explicite ?)*
- *Pourrait-on traduire à l'aide d'un moteur de traduction automatique tout ou partie d'articles scientifiques pour son usage personnel (lecture / consultation) ? Y a-t-il en l'occurrence une taille de texte qu'il est possible de traduire sans enjeux liés au droit d'auteur dans l'idée de pouvoir proposer au lecteur de découvrir certains passages grâce à la traduction automatique, par droit de citation courte, notamment ?*
- *Serait-il envisageable de proposer des traductions de résumés voire d'articles directement dans les plateformes de diffusion en open access (type HAL) ? Dans quel cadre ? Autorisation demandée à l'auteur, ou signature d'une licence CCby ou autre au moment du dépôt de l'article ?*
- *Toujours dans le cas d'une extension de type Plug-In dans HAL, peut-on distinguer le cas où les résumés traduits automatiquement sont disponibles uniquement à la lecture et en mode survol via une action du lecteur vs un affichage proposé spontanément par la plateforme?*
- *Y a-t-il une taille de texte qu'il est possible de traduire sans enjeux liés au droit d'auteur dans l'idée de pouvoir proposer au lecteur de découvrir certains passages grâce à la traduction automatique?*

Les outils et besoins spécifiques des chercheurs avec un périmètre du futur service qui reste à discuter

Forts des premières fonctionnalités envisagées dans ce rapport, il s'agira pour la suite de la mission de confronter le positionnement du futur service au besoin et retours faits par les usagers potentiels dans le cadre d'une série de 2 ateliers et d'un premier travail de co-design avec le directeur scientifique Philippe Lacour, les experts et l'équipe projet Operas :

- **Comment se positionne-t-on face aux différents outils déjà disponibles pour la traduction, la rédaction et la découverte de productions scientifiques ?**
Propose-t-on un DeepL spécialisé car entraîné sur des corpus disciplinaires spécifiques ? Un outil de TAO type MateCat ou Trados (dans sa version multilingue) qui intègre également des fonctionnalités de glossaire, mémoire de traduction et révision ? Traite-t-on le cas des articles rédigés directement dans une langue cible par des chercheurs dont ce n'est pas la langue maternelle pour offrir un service, non pas de traduction, mais de rédaction ? Si l'on accompagne la révision, propose-t-on de charger des "guides de styles" propres à des normes académiques ? Va-t-on jusqu'à proposer des templates d'article avec des phrases types et des "blancs" qu'il faudrait remplir ?
- **Quelles extensions de l'outil peut-on proposer ?**
Propose-t-on un moteur de traduction automatique et entraîné spécifiquement sur des corpus disciplinaires à intégrer par les traducteurs à leurs outils de TAO à l'image de Trados Studio qui intègre ModernMT. Propose-t-on des plug-in à intégrer à son navigateur, ou dans HAL et dans les autres plateformes de diffusions et revues en accès libre ?
- **Quels autres services peut-on rendre aux traducteurs spécialisés dans la traduction scientifique ou technique (chargement de mémoires de traduction en open access, de glossaires, de corpus monolingues ou multilingues spécialisés...)?**

BIBLIOGRAPHIE & RÉFÉRENCES

Bibliographie consultée

- Bouillery, Carine, Marie-Céline Georg, et Elaine Holt. « Édito ». *Traduire. Revue française de la traduction*, no 246 (15 juin 2022): 3-4.
- Cloiseau, Gilles. « La traduction automatique en 2021. Qui, quoi et comment ? Une enquête sociolinguistique ». *Traduire. Revue française de la traduction*, no 246 (15 juin 2022): 84-97. <https://doi.org/10.4000/traduire.2853>.
- Conjard, Lucie, "Ethnographie de l'IA", intervention au séminaire "Code Source", 12 mai 2022, <https://listes.services.cnrs.fr/www/arc/athe-na/2022-04/msg00013.html>
- Davat, Ambre, "la notion de biais en IA", 08/01: Séminaire IA Grenoble; chaire IA, <https://www.ethics-ai.fr/ambre-davat-biais-un-concept-tech-nosolutionniste/>;
Vidéo: <https://www.ethics-ai.fr>
- Fiorini, Susanna. « Traduction automatique et édition scientifique ». *Traduire. Revue française de la traduction*, no 246 (15 juin 2022): 36-45. <https://doi.org/10.4000/traduire.2805>.
- Foti, Markus. « eTranslation. Le système de traduction automatique de la Commission européenne en appui d'une Europe numérique ». Traduit par Noëlle Brunel. *Traduire. Revue française de la traduction*, no 246 (15 juin 2022): 28-35. <https://doi.org/10.4000/traduire.2793>.
- François, Floriane. « Post-édition, mode d'emploi. Retour sur une expérience personnelle ». *Traduire. Revue française de la traduction*, no 246 (15 juin 2022): 5-9. <https://doi.org/10.4000/traduire.2754>.
- Filière, Carole : « Sommaire : n° 9 - La traduction littéraire et SHS à la rencontre des nouvelles technologies de la traduction : enjeux, perspectives et défis (2021) », <https://revues.univ-tlse2.fr/lamaindethot/index.php?id=899>.
- Gonse, Angèle. « Traduction automatique et usages sociaux des langues. Quelles conséquences pour la diversité linguistique ? Jean-Claude Beacco et al. (dir.) ». *Traduire. Revue française de la traduction*, no 246 (15 juin 2022): 24-27.
- Grass, Thierry. « L'erreur n'est pas humaine ». *Traduire. Revue française de la traduction*, no 246 (15 juin 2022): 10-23. <https://doi.org/10.4000/traduire.2763>.
- Hurot, Laura. « Vers une slow translation ? Ralentir pour mieux traduire ». *Traduire. Revue française de la traduction*, no 246 (15 juin 2022): 109-17. <https://doi.org/10.4000/traduire.2869>.
- Kosmopoulos, Christine, Natacha Aveline, Colette Cauvin-Reymond, Bernard Elissalde, Maria Gravari-Barbas, Margaux Hardy, Nathalie Lemarchand, et al. « CybergeoNet – Traductions scientifiques ». *Cybergeo: European Journal of Geography*, 23 février 2022. <http://journals.openedition.org/cybergeo/38309#tocto2n4>.
- Lacour Philippe ; Béné Aurélien- TraduXio Project: Latest Upgrades and Feedback, jdmdh:6733 - *Journal of Data Mining & Digital Humanities*, 8 janvier 2021, Atelier Digit_Hum - <https://doi.org/10.46298/jdmdh.6733>
- Mion, Enrico Antonio. « Un dialogue de sourds ». *Traduire. Revue française de la traduction*, no 246 (15 juin 2022): 46-54. <https://doi.org/10.4000/traduire.2814>.

Poibeau, Thierry (2019). Babel 2.0. *Où va la traduction automatique*. Odile Jacob, Paris. https://www.odilejacob.fr/catalogue/sciences/informatique/babel-20_9782738148490.php

Russo, Nicola Pascal. « Faut-il craindre l'interprétation automatique ? De la fiction à la réalité, le point sur les « traducteurs électroniques » ». *Traduire. Revue française de la traduction*, no 246 (15 juin 2022): 98-108. <https://doi.org/10.4000/traduire.2863>.

Tagand, Hélène. « Traducteurs, quel est votre métier ? » *Traduire. Revue française de la traduction*, no 246 (15 juin 2022): 55-64. <https://doi.org/10.4000/traduire.2823>.

Vidrequin, Magali. « Acceptabilité de la traduction automatique. Le cas de la post-édition chez les traducteurs médicaux ». *Traduire. Revue française de la traduction*, no 246 (15 juin 2022): 77-83. <https://doi.org/10.4000/traduire.2844>.

Villani, Cédric, "Donner un sens à l'Intelligence artificielle", rapport du 28 mars 2018, https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

Autres ressources: blogs, colloques, séminaires, émissions (ordre chronologique)

- Blog d'Adrian Acolier (The morning paper). Article sur les plongements de mots (word embeddings) : <https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/>
- Blog de Mark Liberman. Article sur les limites de la traduction neuronale en ce qui concerne les expressions idiomatiques : <http://languagelog.ldc.upenn.edu/nll/?p=40602>.
- "Les humanités numériques en langue", Atelier DigitHum, ENS Paris, 17 **octobre 2019**, <https://digithum.huma-num.fr/atelier/2019/>
- Xavier De La Porte, avec T. Poibeau (invité). Comment la traduction automatique s'est-elle mise à (mieux) marcher ? Podcast "Le Monde a changé", France Inter, 11 **septembre 2020**. <https://www.radiofrance.fr/franceinter/podcasts/le-code-a-change/la-traduction-automatique-avec-thierry-poibeau-9802626>
- Translation Spaces - Volume 9, Issue 1, **2020** <https://www.jbe-platform.com/content/journals/2211372x/9/1>
- "IA et cancer : le diagnostic infaillible ?", émission radio du 21 **avril 2021**, <https://www.franceculture.fr/emissions/la-methode-scientifique/la-methode-scientifique-emission-du-mercredi-21-avril-2021>
- Séminaire "Objectivité et big data en médecine", **avril-juin 2021**, <http://poincare.univ-lorraine.fr/fr/seminaire-objectivite-et-big-data-en-medecine-objectivite-graphique-en-sciences>
- Colloque "Qu'est ce qui échappe à l'IA ?", Ecole Polytechnique, 20-21 **septembre 2021**, <https://www.polytechnique.edu/actualites/retour-sur-le-colloque-quest-ce-qui-echappe-lintelligence-artificielle>
- "Le numérique dans les sciences humaines: édition et visualisation", congrès de l'ACFAS (association canadienne pour la recherche francophone), 9-10 **mai 2022**, <https://www.acfas.ca/evenements/congres/programme/89/300/310/c?ancre=23314>
- "La formation en traduction à l'ère du numérique", congrès de l'ACFAS (association canadienne pour la recherche francophone), 12-13 **mai 2022**, <https://www.acfas.ca/evenements/congres/programme/89/300/306/c>

- Colloque: “L’IA au prisme des sciences humaines et sociales”, Paris, EHESS, 13-14 **oct 2022**, <https://www.ehess.fr/fr/journ%C3%A9es-d%C3%A9tude/ia-prisme-sciences-humaines-et-sociales>
- Colloque: “Objectivité et big data en médecine”, Univ. de Strasbourg, 14 et 15 **déc. 2022**, <https://www.misha.fr/agenda/evenement/colloque-objectivite-big-data-et-medecine-objectivite-graphique-en-sciences>
- “Après ChatGPT : où en est-on avec les modèles de langage ?”, demie journée d’étude (org. Thierry Poibeau), ENS Paris, 11 **janvier 2023**, 14h, <https://www.risc.cnrs.fr/echos/420146>
- Séminaire et carnet de recherche “CulturIA”, **2022-3** <https://cultureia.hypotheses.org/703>
- Rapport de l’enquête 2022 sur les pratiques professionnelles en traduction, Société Française des Traducteurs, **juillet 2022**: https://www.sft.fr/sites/default/files/2022-11/2022_SFT_resultats-enquete-statistiques-metiers-de-la-traduction.pdf
- Barbin, Franck ; Hernández Morin, Katell ; Phuez-Favris, Gaëlle (2022) «Rédaction et traduction des métadonnées. Guide à destination des auteurs et comités éditoriaux de revues» [guide issu du projet Optimice]: <https://doi.org/10.34847/nkl.1b145pz7>

