

L'intelligence artificielle et la traduction générée automatiquement

Lingua e traduzione – lingua francese – corso avanzato (RISE-SA), a.a. 2022-2023

Prof.ssa Alida Maria Silletti

SOURCES: François Yvon. « Les deux voies de la traduction automatique ? ». *Hermès, La Revue-Cognition, communication, politique*, CNRS-Editions, 2019, 85, pp.62-68 ;

Nathalie Kübler, *La traduction automatique : traduction machine?*,
<https://core.ac.uk/download/pdf/47087967.pdf>

AUX ORIGINES DE LA TRADUCTION AUTOMATIQUE (TA)

- La traduction automatique (TA) est le domaine de recherche à la base du traitement automatique du langage

1940-1960 :

- Les premiers ordinateurs permettent aux belligérants de déchiffrer leurs codes respectifs et on tente d'appliquer ces techniques de déchiffrement à la traduction automatique ;
- dans les premiers systèmes de TA, les connaissances en syntaxe et en analyse syntaxique sont encore très insuffisantes, les ordinateurs ont des capacités de stockage très limitées et sont peu puissants ;
- des systèmes de mise en correspondance de dictionnaires sont créés, ce qui génère des résultats de traduction voués à l'échec, incapables de traduire plus de quelques phrases
- 1954 : le premier système de TA, financé par les gouvernements russe et américain, est présenté au public en 1954 ; il permet de traduire 49 phrases russes, sélectionnées au préalable, vers l'anglais se servant d'un dictionnaire de 250 mots et de six règles de grammaire ;
- 1957-1962 : des modèles syntaxiques de la langue sont créés, par lesquels une description syntaxique de la langue donne une structure aux phrases - les systèmes de traduction automatique de différents types nécessitent tous, entre autres, un module performant et robuste d'analyse syntaxique

AUX ORIGINES DE LA TA

1966 : le rapport ALPAC

- le gouvernement des Etats-Unis demande à la commission ALPAC (Automatic Language Processing Advisory Committee) un rapport qui conclut que la TA est plus lente, moins efficace et deux fois plus chère que la traduction faite par des humains et recommande d'arrêter de financer la recherche dans ce domaine et de se tourner vers des aides à la traduction, comme les dictionnaires automatiques - la recherche en TA est temporairement stoppée aux Etats-Unis

1967-1976 :

- d'autres pays continuent à développer la recherche en TA que stimulent des besoins concrets ;
- aux Etats-Unis la seule activité porte sur la traduction du russe en anglais de textes scientifiques et techniques ;
- 1970 : au Canada, le projet TAUM-Météo mène au développement ultérieur du langage de programmation PROLOG et s'utilise par ailleurs avec succès dans le domaine restreint de la traduction des prévisions météorologiques ;
- en France, on développe à Grenoble un système permettant de traduire du russe des textes mathématiques et physiques ;
- 1976 : le système Systran, produit par Peter Toma, est mis en place à la Commission Européenne en 1976 et il continue à être actuellement utilisé

AUX ORIGINES DE LA TA

Années 1980 :

- à partir de la première version de Systran, qui consiste à traduire du russe vers l'anglais, de nouvelles versions auprès de la Commission Européenne et de l'OTAN sont créées et installées
- des systèmes concurrents de Systran sont développés : Logos et METAL
- les premiers systèmes multilingues sont créés : le projet européen EUROTRA est basé sur un système à transfert multilingue, le Distributed Language Translation system à Utrecht (est construit sur un système à pivot en passant par l'espéranto)
- des aides à la traduction sont développées intégrant des dictionnaires, des concordanciers et des mémoires de traduction

AUX ORIGINES DE LA TA

Années 1990 :

- des mémoires de traduction, indispensables pour la traduction professionnelle, commencent à dominer le marché
- la montée en puissance des ordinateurs personnels permet aux systèmes commerciaux de TA de produire des systèmes individuels que l'on peut installer sur PC
- le développement d'Internet permet de mettre en ligne des systèmes de TA : Systran est le premier à s'en servir en 1998 avec Altavista/Babelfish
- la recherche en traduction automatique progresse : des systèmes à transfert sont utilisés, où un module d'analyse de la langue source, un module de règles de transfert complexes entre la langue source et la langue cible et un module de génération de la langue cible interagissent à partir de règles syntaxiques et de très grands dictionnaires
- les documents disponibles sur support électronique permettent d'enrichir toujours davantage les dictionnaires et surtout de créer des dictionnaires spécialisés qui donnent des résultats de traduction appréciables dans les domaines techniques et scientifiques – l'idée est de tenter de détecter le domaine d'un texte source pour indiquer automatiquement au système quels dictionnaires il doit activer

AUX ORIGINES DE LA TA

- l'apport de la recherche en linguistique de corpus qui décrit la langue à partir des données concrètes du corpus permet de travailler sur le lexique et d'y ajouter des informations syntaxiques toujours plus nombreuses
- les outils de traitement statistique de corpus développés en traitement automatique du langage permettent l'émergence d'une nouvelle direction de recherche : la traduction automatique « basée sur les exemples », qui permet de dégager les traductions les plus fréquentes en effectuant des analyses statistiques sur des corpus traduits et alignés
- la traduction de l'oral représente un défi passionnant pour le domaine du traitement automatique du langage : la traduction automatique du langage parlé ou interprétariat automatique
- des projets combinant la reconnaissance de la parole, la traduction automatique et la synthèse de la parole sont créés : le projet allemand Verbmobil (1993-2000), pour réaliser un traducteur oral transportable pour aider les locuteurs germanophones et japonais à mener des négociations commerciales en anglais ; le projet TC-STAR2 dont le but est d'effectuer la reconnaissance vocale des discours au Parlement Européen, traduit le discours et le restitue oralement dans la langue cible

LES ATOUTS DE LA TA

- consulter des documents rédigés dans une langue inconnue
- diffuser des informations dans d'autres langues
- communiquer avec des locuteurs parlant une autre langue
- accélérer la production de traductions humaines
- assister la rédaction en langue étrangère
- apprendre de nouvelles langues
- indexer ou classer de manière homogène des documents rédigés en plusieurs langues, à des fins de recherche d'information ou de détection de contenus frauduleux ou haineux
- les services de traduction en ligne sont utilisés de manière toujours plus massive, servant chaque jour des milliards de demandes de traduction pour plusieurs milliers de couples de langues

LES APPROCHES ACTUELLES ET LES ATOUTS DE LA TA

- Des systèmes computationnels sont créés, basés sur des architectures de calcul « neuronales », capables d'apprendre à mettre en correspondance une phrase écrite ou prononcée dans une langue « source » avec sa traduction dans une langue « cible » - cela nécessite de disposer d'exemples de telles correspondances, rassemblés dans un corpus parallèle
- La possibilité, pour une large communauté de chercheurs et de développeurs de systèmes informatiques, d'accéder à de tels corpus, diffusés par exemple par des institutions multilingues telles que le Parlement européen ou l'ONU
- L'augmentation continue des capacités de calcul et de stockage permet le déploiement à très grande échelle de ces architectures de calcul et la production quasi instantanée de traductions
- La TA contemporaine se nourrit de grandes masses de données, qui devraient être disponibles pour les différents d'usage de la TA : soit ces données existent soit elles n'existent pas

LES APPROCHES ACTUELLES ET LES ATOUTS DE LA TA

- Si les données existent déjà, des traductions professionnelles sont opérées pour le compte d'une entité solvable et devant produire des textes techniques ou littéraires à des fins de publication et soumis à de fortes exigences de qualité – ex. la direction générale de la traduction de l'Union Européenne, dont les traductions à produire dans le futur ressemblent aux traductions passées, elles sont disponibles en grande quantité et l'apprentissage réalisé par la TA produira souvent de bonnes traductions
- Un travail de révision en post-édition sera cependant requis pour satisfaire les niveaux de qualité requis
- Si les données ne sont pas disponibles, cela peut dépendre de diverses raisons : soit il n'existe aucune demande pour des traductions humaines ; soit cette demande n'est pas solvable : soit les langues ne sont pas disponibles ; soit la notion même de traduction n'est pas bien définie, par exemple, si les textes à traduire sont des commentaires postés sans relecture sur des réseaux sociaux – le résultat : des traductions approximatives ou médiocres ... bien que la traduction atteigne son but de faire comprendre