

SPRING, SUMMER, FALL, WINTER... AND SPRING. UNE HISTOIRE DE LA TRADUCTION AUTOMATIQUE

Pascaline MERTEN

ULB, Faculté de Lettres, Traduction et Communication,
École de traduction et interprétation – ISTI-Cooremans – Belgique,
pascaline.merten@ulb.be

Résumé : L'histoire de la traduction automatique (TA) est faite de hauts et de bas, d'espoirs déçus et de réalisations auxquelles on ne croyait plus. L'engouement de la première démonstration en 1954 a cédé la place à une déception qui marquera longtemps et profondément la recherche en traduction automatique. Différentes approches ont fait l'objet de recherches et ont débouché sur des réalisations de plus ou moins grande importance : d'abord les systèmes à base de règles (directs, de transfert ou à interlangue), ensuite les systèmes basés sur les données : les systèmes statistiques et à présent neuronaux. Certains concepts et certaines techniques n'ont été mis en œuvre que bien longtemps après avoir été proposés. Ainsi pourrait-on voir dans les propositions séminales de Warren Weaver en 1949 une amorce des systèmes à interlangue et statistiques. Quant aux réseaux neuronaux, qui ont fait sortir l'intelligence artificielle de son long hiver, ils remontent aux années 1980.

C'est pourquoi nous avons emprunté son titre au très beau film coréen de Ki-Duk Kim, pour rendre cette conception d'une histoire en spirale : la TA évolue, en passant par des périodes de succès et d'autres plus obscures, en reprenant des techniques qu'on avait abandonnées, parce qu'elles étaient immatures ou que la technologie n'était pas assez robuste pour les mettre en pratique.

Ainsi, au-delà des dates et des techniques, l'histoire de la TA est l'occasion de s'interroger sur le lien des technologies avec la vie intellectuelle d'une époque : en quoi la TA est-elle le fruit de son époque, et, en miroir, en quoi notre vision de la technologie, et de la TA en particulier en change-t-elle la réception ?

Mots-clés : Traduction automatique – Intelligence artificielle – Traduction assistée par ordinateur – Technologies de la traduction - TA à base de règles – TA statistique – TA neuronale

Abstract : The history of machine translation (MT) is one of ups and downs, of dashed hopes and disbelieved achievements. The excitement of the first demonstration in 1954 gave way to a disappointment that will leave a long and deep mark on machine translation research. Different approaches have been researched and have led to achievements of varying degrees of importance: first rule-based systems (direct, transfer or interlanguage), then data-based systems: statistical and now neural systems. Some concepts and techniques were not implemented until long after they were proposed. For example, Warren Weaver's seminal proposals in 1949 could be seen as the beginning of interlanguage and statistical systems. As for neural networks, which brought artificial intelligence out of its long winter, they date back to the 1980s.

This is why we have borrowed its title from the beautiful Korean film by Ki-Duk Kim, to convey this conception of a spiral history: MT evolves, going through periods of success and more obscure ones, taking up techniques that had been abandoned because they were immature or because the technology was not robust enough to put them into practice.

Thus, beyond dates and techniques, the history of MT is an opportunity to question the link between technologies and the intellectual life of an era: in what way is MT the fruit of its era, and, in mirror image, in what way does our vision of technology, and of MT in particular, change its reception?

Keywords : Machine Translation (MT) – Artificial Intelligence – CAT – Translation Technologies – Rule-Based MT (RBMT) – Statistical MT (SMT) – Neural MT (NMT).

1. Avant l'ordinateur (avant 1945)

1.1. Les langues universelles a priori et a posteriori

Nous n'avons pas attendu l'ordinateur pour essayer de surmonter nos différences langagières. Dans le monde occidental, le latin perd, à partir de la Renaissance, son statut de *lingua franca* et les langues vulgaires deviennent les langues de la communication intellectuelle et scientifique. Divers penseurs vont imaginer des systèmes pour franchir les barrières linguistiques de manière plus ou moins automatique. Descartes suggère de représenter les mots de différentes langues par un nombre, qui serait identique pour toutes les langues. Il propose même l'exemple, piquant si l'on songe à la polysémie du mot, du verbe *aymer* qui aurait la même représentation numérique que ses correspondants grec ou latin¹. Leibniz qui, comme Pascal ne sépare pas la pensée la plus spéculative des réalisations concrètes, imagine dans sa *Dissertatio de arte combinatoria* de 1666 une méthode de réduction des raisonnements à une combinaison d'éléments de base, qui permettrait un calcul sur les idées. Il construit en 1673 une machine à calculer qui améliore la *pascaline*. En 1668, John Wilkins publie *An Essay towards a Real Character, and a Philosophical Language* où il répertorie et classe tous les concepts et objets du monde en les ramenant à un arbre d'éléments premiers. Dans cette taxonomie, les objets prennent un nom qui correspond à leur place dans l'arborescence. Ainsi, la tulipe sera représentée par le mot « *gæde* » et le malt par « *glæde* ». Comme le souligne Umberto Eco, la caractéristique universelle viole la double articulation du langage. Ce faisant, elle « fait en sorte que la moindre variation de son ou de caractère impose un changement de sens » (Eco 1997 : 285), ce qui rend cette langue fort peu robuste². Wilkins, à la différence d'autres penseurs qui ont développé une langue universelle, a essayé de développer un système qui soit international. Quoi qu'il en soit, ce système est l'image de la science occidentale de son temps. Il n'imagine pas que d'autres peuples puissent avoir une conception différente ni que la connaissance évolue, ce qui devrait nécessairement se refléter dans la langue. Par ailleurs, même si la grammaire est également abordée, elle est le parent pauvre de cette approche, qui met l'accent sur le lexique.

À côté de ces langues universelles *a priori*, on a vu fleurir des langues universelles *a posteriori*, comme le volapük ou l'espéranto, projets généreux qui naissent à la fin du XIX^e siècle. On est passé d'un projet philosophique à des réalisations concrètes, créées dans un élan de fraternité des peuples.

1.2. Les précurseurs au XX^e siècle

Dans les années 1930, deux inventeurs développèrent des systèmes qui auraient peut-être abouti à des réalisations opérationnelles si la Deuxième Guerre mondiale n'était pas venue entraver leurs travaux.

Georges Artsrouni, un Français d'origine arménienne, commence à travailler à ses machines en 1932. Il dépose en 1933 un brevet pour « un appareil rendant mécanique et automatique l'emploi d'horaires de chemins de fer, d'annuaires téléphoniques, de dictionnaires, etc. » (Daumas 1965 : 284). Il le présente à l'exposition universelle de 1937. La guerre interrompt ses travaux ; ensuite, le manque de moyens financiers ne lui permettra pas de développer son invention. Toujours est-il qu'elle n'aurait permis que de développer des dictionnaires (Daumas 1965).

En Union soviétique, Petr Smirnov Troyanskii dépose en 1933 un brevet pour un « dispositif constituant un lexique automatique imprimant » (Daumas 1965 : 284). Le système comporte trois phases, dont seule la deuxième est automatisée : d'abord, un humain qui connaît la langue source lemmatise et catégorise les mots de l'entrée ; ensuite le dispositif les traduit ; enfin, un humain qui connaît la langue ciblée produit les formes fléchies. Selon John Hutchins, Troyanskii aurait pu être un des pionniers de la traduction automatique si les autorités avaient pris ses travaux en considération (Hutchins 1986 : 23-24).

2. Le concept (Memorandum de Warren Weaver) (1949-1954)

Même s'il y a des discussions sur la paternité et la date du premier ordinateur, on peut globalement considérer que c'est la réalisation de l'ENIAC qui, en 1945, signe les débuts de l'informatique³. En 1949, Warren Weaver publie un *Memorandum* resté célèbre sur l'usage des ordinateurs en traduction. Ses propositions sont générales, et à long terme, plutôt concrètes ou techniques. 1949 est considéré comme la date de naissance de la traduction automatique (TA).

Le *Memorandum* est un texte bref, qui lance des pistes de solution pour la TA. Il mentionne une expérience d'utilisation de l'ordinateur dans une traduction mot à mot, tentative que les journaux ne manqueront pas de tourner en dérision. Pour Weaver, il est clair – il y reviendra plusieurs fois – que la traduction automatique doit dépasser le niveau du mot et n'a pas pour objectif de traiter des textes littéraires. En effet, la plupart des mots sont polysémiques. Il suggère d'utiliser le contexte pour désambiguïser, la fenêtre de contexte nécessaire (le nombre de mots avant ou après, ou empan) pouvant être déterminée statistiquement.

Il évoque aussi l'utilisation de la cryptographie. Il en parle d'abord au début du *Memorandum* en racontant qu'un message avait pu être déchiffré sans que le cryptographe ait su quelle était sa langue de rédaction initiale (le turc, en

l'occurrence). Les méthodes utilisées se basaient, dit-il, sur la fréquence des lettres, les combinaisons de lettres, les séquences de lettres, « which are to some significant degree independent of the language used » (Weaver 1949 : 2). Et on n'a pas tardé, en effet, à développer des algorithmes identifiant la langue sur la base de séquences de lettres. Il y revient à la fin de son texte, en mentionnant la théorie mathématique de la communication de Shannon. Celui-ci avait publié son article intitulé *A Mathematical Theory of Communication* en 1948. L'année suivante, il le republiera sous forme de livre sous le titre *The Mathematical Theory of Communication*. Weaver y publie une version vulgarisée de cette théorie. Si des moyens statistiques ont bien été utilisés durant des décennies, c'est le « noisy channel model » de Shannon qui permettra à la traduction automatique statistique de faire redécoller la traduction automatique dans les années 1990.

Il évoque aussi le caractère « logique » des langues. D'abord en mentionnant le théorème de McCulloch et Pitts :

This theorem states that a robot (or a computer) constructed with regenerative loops of a certain formal character is capable of deducing any legitimate conclusion from a finite set of premises. Now there are surely allogical elements in language (intuitive sense of style, emotional content, etc.) so that again one must be pessimistic about the problem of literary translation. But insofar as written language is an expression of logical character, this theorem assures one that the problem is at least formally solvable. (Weaver 1949 : 9-10)

Voit-il un lien, un rapport d'isomorphie entre les langues naturelles et le caractère « formel » de la méthode qui permettrait de les traiter ? Chomsky, sur les traces de Zellig Harris, allait en 1957, dans ses *Syntactic Structures* utiliser un formalisme commun pour décrire tous les langages, en mettant en lumière les différences entre langages naturels et langages formels. Ce faisant, il a jeté les bases de la linguistique formelle. Il allait aussi, pour un temps du moins, à travers la grammaire transformationnelle, populariser l'idée des structures profondes et des structures de surface. Y avait-il une prémonition du formalisme en linguistique chez Weaver ? Par ailleurs, à relire aujourd'hui son évocation du théorème de McCulloch et Pitts, on ne peut manquer de penser aux réseaux de neurones récurrents.

Une des idées les plus prometteuses, selon lui, est de retrouver au-delà de la diversité des langues, des « invariants ». Il imagine sans doute des invariants liés aux capacités physiques et cognitives ainsi qu'à l'expérience de vie, en plus des invariants statistiques (comme les séquences de lettres propres à une langue). Il mentionne le sémanticien comparatiste Erwin Reifler, qui deviendra d'ailleurs un des premiers chercheurs en traduction automatique (Hutchins 1986 : 31), et revient sur cette idée (« the most promising approach ») en conclusion :

Perhaps the way is to descend, from each language, down to the common base of human communication - the real but as yet undiscovered universal language - and then re-emerge by whatever particular route is convenient. (Weaver 1949 : 11)

Weaver aurait-il donc imaginé à la fois la linguistique formelle, l'approche à interlangue, la TA statistique et la TA neuronale ? Melby (2020 : 419-435) voit dans ses propositions aux moins le germe des développements futurs. On peut supposer qu'il y avait des idées dans l'air du temps, que Weaver a perçues. Mais

il faudra des décennies pour que certaines de ces idées soient comprises. Ainsi, voir la traduction comme une application de la cryptographie a pu paraître simpliste, pour qui connaît la complexité de la traduction. Mais c'est assimiler la cryptographie à un codage terme à terme, alors qu'elle est infiniment plus complexe. L'allusion au théorème de McCulloch et Pitts n'a pris son sens que récemment : il ne s'agit pas tant de dire que le langage a une base logique, que de dire qu'on peut développer des algorithmes qui le traitent et que ces algorithmes ont, eux, une base logique et formelle.

3. Les premières réalisations (1954-1965)

Un premier colloque consacré à la TA se tient en 1952. On y trouve entre autres Leon Dostert, Yehoshua Bar-Hillel et Erwin Reifler. Leon Dostert avait été interprète durant la Deuxième Guerre mondiale et avait contribué à la mise sur pied du système d'interprétation simultanée au procès de Nuremberg. Il travaillait alors à l'Institut des langues et de linguistique de l'université de Georgetown. Il semblait sceptique au début, fort de son expérience d'interprète, mais se laissa séduire au fil du colloque et plaida pour le développement d'un prototype qui permettrait de valider la faisabilité de la TA.

Yehoshua Bar-Hillel et Erwin Reifler suggérèrent d'inclure des phases de pré-édition et de post-édition. On peut dire que Reifler, sans le nommer, a conçu le concept de langage contrôlé. Bar-Hillel, quant à lui, lance le concept d'une « syntaxe opérationnelle » (Hutchins 1986 : 35).

Dostert met sur pied un groupe de recherche dans son université, en collaboration avec l'entreprise informatique IBM, et, le 7 janvier 1954, a lieu la première démonstration d'un système de TA qui ne soit pas de la traduction mot à mot et qui produise un résultat intelligible. Il s'agit d'un système russe-anglais, qui travaille sur un vocabulaire russe de 250 mots, avec une grammaire de 6 règles. Les 49 phrases en russe sont soigneusement choisies, et la démonstration répond aux attentes.

C'est le début d'une période faste pour la TA, spécialement aux États-Unis : les fonds affluent, les projets naissent, les approches se multiplient.

Dans les années 1950, tout est à faire : les dictionnaires électroniques n'existent pas, il n'y a pas d'analyseur morphologique, et on commence à se dire qu'une analyse syntaxique permettrait de résoudre certains problèmes de polysémie. On pourrait mettre toutes les formes fléchies dans un dictionnaire, mais les moyens techniques sont alors limités. L'algorithmique et les langages de programmation en sont à leurs balbutiements, et on ne sépare pas clairement les données et les méthodes. Aussi, le changement de position absolue de l'adjectif par rapport au nom entre l'anglais et le français était-il traité par un code dans le lexique (Hutchins 1986 : 45). On reste clairement ici dans une architecture directe, mot à mot.

Les projets de recherche vont alors s'orienter vers deux voies : l'une empirique, dans la lignée de Dostert, l'autre qui veut nourrir la pratique d'une réflexion plus fondamentale sur la langue. C'est ainsi qu'apparaît une approche formelle de la linguistique qui permettra l'élaboration d'autres modes de

représentation de la syntaxe. Les programmes d'analyse (*parsing*) devront mettre en lumière la structure syntaxique, en termes de constituants, de relations fonctionnelles ou de dépendance. Certains vont utiliser des automates à états finis, d'autres des grammaires de dépendance, d'autres enfin des grammaires de constituance. En ce qui concerne leur capacité générative, ces grammaires sont indépendantes du contexte (*context-free grammars*). Chomsky a étudié la puissance générative des différents formalismes et a montré les limites de ces outils. Il a proposé des règles de transformation pour relier des phrases de structures différentes, mais de même sens (comme la forme passive et la forme active). Sans entrer trop avant dans le débat sur le sens et le formalisme, notons que, dans les années 1980, Gazdar, Klein, Pullum et Sag « ont voulu montrer qu'il y a un moyen de décrire des phénomènes linguistiques avec un seul niveau de structure et que le but du linguiste ne doit pas être la recherche d'une quelconque pertinence psychologique du modèle, mais l'adéquation aux faits de langue ainsi que l'élégance et la concision du modèle. » (Merten 2005 : 336-337) Différents formalismes (GPSG, LFG...) utiliseront des traits linguistiques pour augmenter la capacité générative des grammaires indépendantes du contexte mais sans leur faire atteindre la complexité des grammaires transformationnelles.

On voit ici que, au-delà de la question de la traduction à proprement parler, la traduction automatique devient une coupole pour les recherches en linguistique formelle, quantitative, appliquée et computationnelle.

4. La fin des illusions (1965)

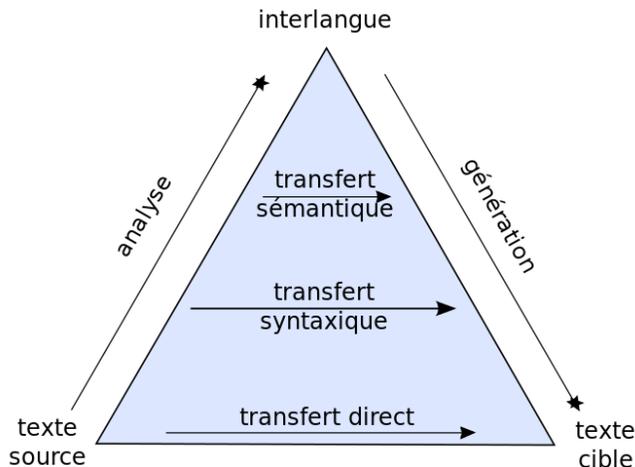
Cependant, le matériel, les formalismes et les algorithmes ne sont pas mûrs. En 1965, le groupe d'experts dépêché par le gouvernement américain rend un rapport qui, non seulement critique sévèrement les résultats obtenus, mais qui remet même en cause la faisabilité de la TA, en raison de la « barrière sémantique » infranchissable de la langue. Les experts du groupe ALPAC⁴ n'ont étudié que quelques groupes de recherche, sans prendre en compte qu'il s'agit, précisément, de projets. Ils en comparent les résultats au résultat d'une traduction humaine, sans tenir compte de ce que toute traduction humaine doit aussi être révisée... La TA est plus lente, plus chère et elle nécessite une postédition. Le rapport ALPAC, malgré ses limites et les nombreuses critiques qu'il suscita, jeta le discrédit sur la TA, freina la recherche pour des années, et pas seulement aux États-Unis — et ternit pour longtemps l'image de la traduction automatique. Sur une note plus positive, néanmoins, il suggère de développer d'autres outils, au premier chef des dictionnaires automatiques.

5. Les systèmes à base de règles (1954-)

Les premiers systèmes étaient, comme on l'a vu, des systèmes à architecture directe, ou mot à mot. Quelques procédures, dirigées par le lexique, permettaient de résoudre certains problèmes syntaxiques ou de traduction, comme la place de l'adjectif ou la traduction de certains mots (par exemple, *to know* qui se traduit notamment par *savoir* ou *connaître*).

La nécessité d'avoir des règles apparaît rapidement. La TA va alors être découpée en trois étapes, d'importance relative selon les systèmes : l'analyse, le transfert et la génération. À chaque étape correspondent un certain nombre de règles.

C'est ce que Vauquois a modélisé dans son triangle, repris ci-dessous.



L'analyse consiste à prendre une phrase en entrée et à produire une analyse plus ou moins poussée : la représentation peut être morphologique, syntaxique, sémantique, voire a-linguistique, dans une interlangue ou une langue universelle.

Le transfert consiste à se saisir de l'analyse réalisée en phase 1 et à produire une représentation dans la langue cible. On peut ainsi prendre un arbre d'analyse syntaxique de l'anglais et produire l'arbre correspondant en français. Les règles de transfert ne seront pas les mêmes entre deux paires de langues différentes. Pour reprendre l'exemple ci-dessus, la traduction de *know* est un problème pour la traduction vers le français, mais ne le sera peut-être pas vers d'autres langues.

La génération (ou synthèse) part du résultat du transfert (ou de l'analyse s'il s'agit d'un système à interlangue) pour produire la phrase en langue cible. Les règles de génération doivent produire les formes correctement fléchies (et choisir entre les formes *ce* et *cet*, *sa* et *son* en français), l'ordre correct des mots, la place de l'adjectif, la forme et la place de la négation, les pronoms...

Le triangle traduit l'importance relative des phases : si on se contente d'une analyse superficielle de la phrase, le travail du transfert sera important, celui de la génération sera proportionnel à la profondeur de l'analyse. Si on pousse l'analyse jusqu'à une représentation abstraite ou universelle, a-linguistique, il n'y a plus de transfert, mais la génération sera totale⁵.

5.1. Systèmes de transfert (1960-)

De nombreux systèmes de cette époque sont des systèmes de transfert, certains sont partis d'une architecture directe pour élaborer petit à petit des

règles plus complexes voire intégrer des éléments sémantiques : Geta-Ariane, Susy, Metal, Eurotra, Mu...⁶

Le système Meteo développé par le groupe TAUM (Traduction automatique à l'Université de Montréal) et opérationnel de 1981 à 2001⁷ est un système élaboré pour un sous-langage, celui des dépêches météorologiques. On parle de sous-langage, car les limites portent sur tous les niveaux : la syntaxe y est simple (on n'y trouve que des phrases nominales), la sémantique est étroitement encadrée, l'ambiguïté est réduite (*dépression* n'a qu'un sens ici), et la pragmatique tient peu de place puisqu'une dépêche météorologique se veut purement informative. On observe donc non seulement une terminologie spécifique, mais aussi des restrictions à différents niveaux linguistiques, ce qui différencie le sous-langage de la langue de spécialité ou du langage contrôlé.

Meteo utilise des règles d'analyse et de transfert qui combinent syntaxe et sémantique.

La règle :

MET1 (({ADJ/GN}, [CMOD]), [T], [L])

analysera la phrase *mainly sunny with moderate winds today*.

La règle :

MET2 (GN({*higs/lows*}), [T], [L], GN({Temp to Temp/Temp}), [T], [L])

analysera la phrase *higs today 15 to 18 in northern regions*.

Les règles de transfert suivantes permettent de traduire correctement l'adjectif *heavy*, selon son contexte :

heavy=ADJ((TOMBANT), abondant)

heavy=ADJ((STATIONNAIRE), dense)

heavy=ADJ((SOUFFLANT), fort)

Meteo produira ainsi *pluie abondante*, *brouillard dense* et *vents forts*.

Il est à noter que Meteo échouera à être adapté sur un autre type de textes, en l'occurrence des manuels d'aéronautique. Le problème viendra de syntagmes nominaux complexes, comme *left engine fuel pump suction line*⁸. On voit bien, une fois de plus, que c'est la sémantique et la connaissance du monde qui constituent la barrière infranchissable. Faute de pouvoir identifier un sous-langage, c'est l'utilisation d'un langage contrôlé par les rédacteurs techniques qui facilitera la traduction automatique de tels textes techniques, langage contrôlé qui sera mis en place dans l'industrie aérospatiale à partir de 1980 avec le *Simplified Technical English*.

5.2. La barrière sémantique et la représentation du sens

Déjà, en 1959, Bar-Hillel, avec l'exemple

Little John was looking for his toy box... Finally he found it: it was in the pen,

remettait en cause la faisabilité d'une « High Quality Fully Automatic Machine Translation » (traduction complètement automatique et de grande qualité) (Hutchins 1986 : 154).

On a vu que l'emploi de traits sémantiques permet à TAUM-Meteo de choisir la bonne traduction. D'autres systèmes, comme Metal, ont eu recours aux traits sémantiques, par exemple, pour sélectionner la bonne traduction française du verbe néerlandais *werken* : werken + sujet animé = travailler ;

werken + sujet inanimé = fonctionner. L'exemple de Bar-Hillel est intéressant en ce que sa résolution incombe à notre connaissance du monde, qui, pour évidente qu'elle soit à l'être humain, demande à la machine toute une série d'inférences : un objet ne peut être contenu que dans un objet plus grand ; or une boîte à jouets est plus grande qu'un stylo, donc elle ne peut pas être contenue dans un stylo ; en revanche, elle est plus petite qu'un parc d'enfant et peut donc s'y trouver.

Le problème était la représentation du sens, voire, en termes plus contemporains, de la compréhension. Il s'agissait donc au premier chef d'ajouter une composante sémantique à des systèmes essentiellement syntaxiques. On s'est tourné vers des approches linguistiques qui accordaient une place plus grande à la sémantique, comme la grammaire des cas ou la grammaire de dépendance de Tesnières, les dépendances conceptuelles de Schank ou la grammaire sémantique de Montague.

5.3. Systèmes à interlangue (1960-)

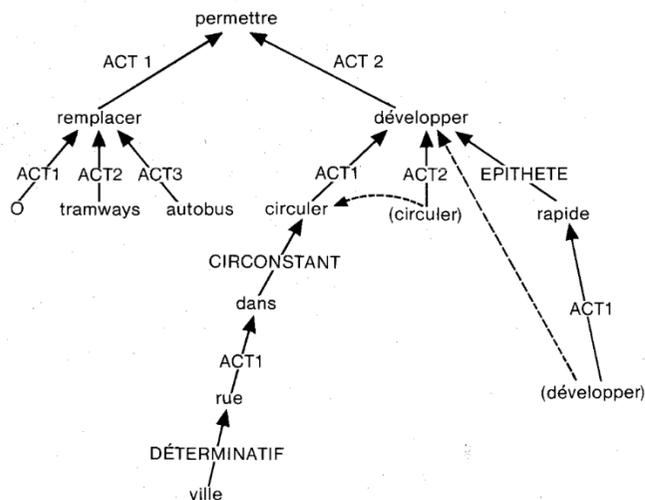
5.3.1 Langages pivots

Les systèmes à interlangue ont été un domaine de recherche et d'expérimentation à la lisière de la TA, de la linguistique et de l'intelligence artificielle (IA). En réalité, il y a un continuum entre les systèmes qui ajoutent une composante sémantique, ceux qui vont jusqu'à une représentation interlinguale, et les systèmes dits d'intelligence artificielle.

Le Centre d'études pour la traduction automatique (CETA), conduit par Vauquois à Grenoble, développe à la fin des années 1960 le concept de « langage pivot ». Ainsi, la phrase

Le remplacement des tramways par des bus a permis un développement rapide de la circulation dans les rues de la ville.

sera-t-elle représentée par l'arbre :



On voit que le système repose sur un schéma actanciel : un verbe, un nom ou un adjectif appellent un certain nombre d'actants⁹. Le verbe *remplacer* appelle trois actants : quelqu'un remplace quelque chose par quelque chose. Ici, le premier actant est absent, de plus *remplacer* est lui-même le premier actant du sommet de l'arbre et sera nominalisé : *le remplacement des tramways par des autobus*.

5.3.2 Langages universels a posteriori (1975-)

L'idée d'utiliser une langue universelle *a posteriori* a aussi été évoquée. C'est ainsi que le projet DLT (*Distributed Language Translation*) a tenté d'utiliser l'espéranto. Mais cette langue aussi a ses limites : elle n'échappe pas à l'ambiguïté, au manque de vocabulaire technique et est peu encline à la néologie.

5.3.3 Systèmes d'intelligence artificielle (1970-)

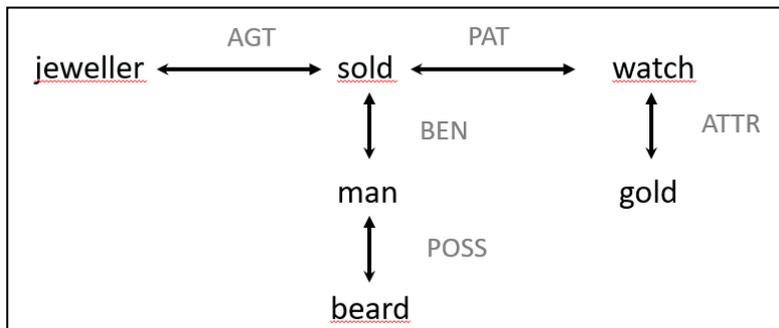
Le but de l'intelligence artificielle (IA) n'est pas, au premier chef, de traduire, mais de comprendre, de synthétiser et de hiérarchiser, à des fins de dialogue personne-machine, de compréhension, d'extraction de l'information, de synthèse automatique...

Comme en TA, les techniques sont aujourd'hui différentes, mais l'idée était de modéliser une compétence humaine pour la faire reproduire par la machine : un système expert en médecine, en droit, un système de dialogue personne-machine qui interprète la question de l'utilisateur, extrait l'information d'une base de données et produit une réponse. Les techniques d'analyse du langage en intelligence artificielle font abstraction de la dimension syntaxique pour repérer le sens, l'organisation du texte et les relations sémantiques.

Voici la représentation donnée à la phrase :

The golden watch was sold by the jeweller to a man with a beard

dans l'approche de Yorick Wilks (Hutchins 1986 : 275) :



Cette représentation en termes de dépendances conceptuelles ne semble pas très éloignée de celle de Vauquois. Néanmoins, les méthodes et les objectifs des approches « TAL » (traitement automatique du langage) et « IA » varient. En traduction, on doit *a priori* garder des informations de surface, comme l'organisation thème/rhème. Il n'empêche que c'est le renouveau de l'IA qui donnera son troisième souffle à la TA.

6. Les outils d'aide à la traduction et le poste de travail du traducteur (1990-)

Certains systèmes à base de règles sont parvenus au stade industriel, comme le célèbre Systran, système direct d'abord, puis de transfert, hybride en fait et souvent *ad hoc*, qui a passé les âges en intégrant chaque fois les technologies nouvelles. Metal aussi, d'abord développé à l'université du Texas, racheté par Siemens, puis par Lernhout et Hauspie, a eu un certain succès dans les années 1980. L'approche sous-langage de Meteo semblait prometteuse. Face à l'explosion combinatoire des problèmes de syntaxe et de sens, circonscrire un sous-domaine qui impose des limitations morphosyntaxiques, syntaxiques, sémantiques et discursives semblait la seule alternative à des systèmes généralistes aux résultats insatisfaisants, même en travaillant en pré-édition ou en imposant un langage contrôlé et en travaillant en postédition. Les traducteurs ne voyaient pas la TA comme une menace, mais comme un embarras, s'ils avaient à postéditer des textes qu'ils auraient plus rapidement traduits. Dans les années 1980, on travaillait encore sur des terminaux informatiques reliés à une unité centrale. Ni les systèmes d'exploitation ni les traitements de texte n'étaient multilingues, ce qui compliquait encore l'usage d'un système de TA.

Le rapport ALPAC, dont on a vu les conséquences sur les recherches en TA et sur son image, préconisait de développer d'autres types d'outils, utiles au traducteur, comme des dictionnaires électroniques. En 1970 naît TERMIUM qui centralise la terminologie anglais-français au Canada, puis, en 1975 EURODICAUTOM, le *dictionnaire automatique européen*, ancêtre de l'actuel IATE, voit le jour. Depuis, les dictionnaires de langue ou traductifs, sur CD-ROM d'abord, en ligne ou accessibles grâce à une application sur smartphone ensuite, se sont répandus.

Dans les années 1980, la micro-informatique se répand, avec, en particulier, le PC d'IBM (1981) et le Macintosh d'Apple (1984). Le travail sur ordinateur ou sur traitement de texte n'est plus réservé au travailleur en entreprise, mais devient progressivement accessible à tous. Dès lors, de plus en plus de textes sont en fait des textes numériques, des fichiers, donc autant de données qu'un algorithme peut traiter. Dans le domaine de la traduction, on va ainsi commencer à disposer de textes parallèles, qui sont la traduction l'un de l'autre. La demande est alors d'avoir un environnement de travail facilitant la tâche du traducteur, un bon traitement de texte, facilitant la mise en forme et la mise en page, des correcteurs d'orthographe, de grammaire et de style, l'accès à des dictionnaires, à des bases terminologiques locales ou en ligne et la possibilité de travailler sur des langues à jeux de caractères différents. On envisage le « poste de travail du traducteur » (*Translator's Workstation*), qui ne sera pleinement réalisé qu'avec la micro-informatique et avec l'avènement des mémoires de traduction. La traduction automatique cède la place à la traduction assistée par ordinateur (TAO) et aux outils d'aide à la traduction (OAT).

L'idée de réutiliser les traductions antérieures a été émise dès 1979 par Peter Arthern qui avait observé la répétitivité des textes dans le contexte européen et qui soulignait combien les traducteurs perdaient leur temps à

retraduire des textes déjà traduits (Hutchins 1998 : 293). Ce seront néanmoins et ironiquement les algorithmes d'alignement statistiques qui permettront de concrétiser les projets conçus par Arthern, Kay ou encore Melby.

Dans les années 1990 apparaissent les systèmes de mémoire de traduction de la société Trados (Translator's Workbench, 1992), d'IBM (Translation manager/2, 1992), de Star (Transit, 1992), d'Atril (déjà Vu, 1993) et d'Eurolang (Eurolang Optimizer, 1994), pour ne citer que les premiers.

Le développement de corpus et de corpus bilingues ou multilingues numériques a aussi permis le développement de tout le champ de la linguistique de corpus et ensuite de la traductologie de corpus. Des auteurs comme Mona Baker, Nathalie Kübler ou Rudy Loock utilisent des corpus monolingues ou multilingues dans une approche traductologique soit théorique soit empirique : analyser la langue traduite et ses variations par rapport à la langue non traduite, identifier les « universaux de la traduction » (Baker 1993) ; dans la formation des futurs traducteurs, identifier les problèmes et les erreurs de traduction, les problèmes de terminologie, aider à la décision et à la révision (Loock 2016).

7. Les systèmes basés sur les données (1990-)

7.1. La traduction basée sur l'exemple

L'idée de réutiliser des traductions existantes pour traduire prend d'abord la forme de traduction basée sur l'exemple, comme suggéré par Makoto Nagao en 1984. Lorsqu'il traduit, le traducteur réutilise des mots, des expressions, des segments qu'il a préalablement traduits et les assemble. Mais c'est le développement des corpus et leur utilisation par des algorithmes statistiques qui vont marquer une nouvelle étape.

7.2. Les systèmes statistiques (1990-)

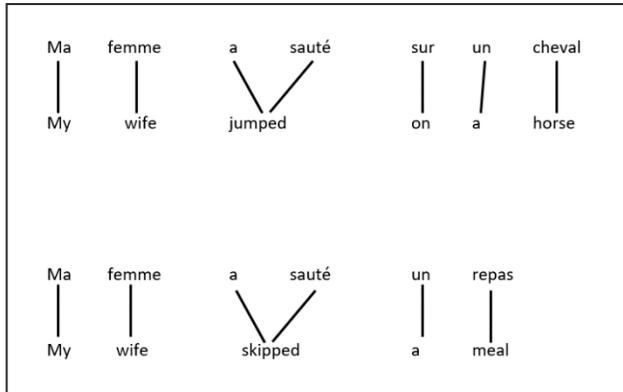
7.2.1 Fonctionnement

À la fin des années 1980, une équipe d'IBM qui avait travaillé sur la reconnaissance et la synthèse de la parole reprend ses algorithmes pour les appliquer à la traduction automatique (Brown *et al.* 1988). Ils se servent du corpus Hansard, qui rassemble les documents du parlement canadien, soit un corpus bilingue de textes anglais et français. Ces premiers systèmes sont fondés sur un alignement mot à mot, jusqu'à ce que Koehn, Och et Marcu (2003) proposent une approche fondée sur des *n-grammes*, soit des séquences de *n* mots..

Un système statistique se divise en deux processus : l'entraînement (*training*) et la traduction (*decoding*).

L'entraînement du système produit, d'une part, un modèle de traduction fondé sur un corpus bilingue et, d'autre part, un modèle de langue à partir d'un corpus monolingue.

Soit les deux phrases française et anglaise ci-dessous :



(Li 2013)

Le système va apparier des n-grammes, ici 1 et 2 mots, mais n peut être plus important. Il va en déduire une table de probabilités, par exemple :

Ma	My	0,8
Femme	Wife	0,8
a sauté	jumped	0,85
a sauté	skipped	0,15
Un repas	a meal	1,0
a sauté un repas	skipped a meal	1,0

On conçoit aisément que *a sauté* se traduit beaucoup plus fréquemment par *jumped* que par *skipped*.

À partir d'un corpus monolingue, le système va établir un modèle de langue, indiquant la probabilité qu'un n-gramme soit un n-gramme plausible de la langue cible. On aura par exemple la table de probabilités suivantes :

Le grand chat	0,4
Le gros chat	0,5
La gros chat	0,1

On trouvera surtout « le gros chat », ensuite « le grand chat », et de façon anecdotique l'incorrect « la gros chat ».

La phase de traduction est appelée « décodage », car le processus se fonde sur le modèle de la théorie mathématique de la communication de Shannon, dit « *noisy channel* », évoqué plus tôt. Le modèle de Shannon entend trouver comment transmettre un message (les données à transmettre) sur un canal (à l'époque les lignes téléphoniques) en maximisant le *rappel* (la quantité de données correctes transmises) et en minimisant le *bruit* (les données incorrectes) et le *silence* (l'absence de données). La TAS représente la traduction comme un processus dans lequel le texte source doit être décodé, par élimination du bruit (les divergences syntaxiques, sémantiques...), pour produire le message escompté (c'est-à-dire la traduction). (Hearne et Way 2011)

Selon une première approche, la traduction proposée sera, parmi toutes les traductions possibles, celle qui aura la probabilité la plus élevée d'être le correspondant traductionnel de la phrase source ($P(S|T)$) et la probabilité la plus élevée d'être effectivement une phrase plausible de la langue cible ($P(T)$), soit la formule suivante :

$$Translation = \operatorname{argmax}_T P(S | T).P(T)$$

$P(S|T)$ mesure l'adéquation ; $P(T)$ mesure la fluidité.

Un modèle plus sophistiqué, le modèle log-linéaire, reprend une somme de probabilités :

$$Translation = \operatorname{argmax}_T \sum_{m=1}^M \lambda_m \cdot h_m(T, S)$$

La meilleure traduction possible sera celle qui a le meilleur résultat pour une somme de caractéristiques. Ce peut être simplement $P(S|T)$ et $P(T)$, mais on peut ajouter d'autres caractéristiques (h_m), comme un corpus spécialisé, une source terminologique... et on peut également pondérer chacune des caractéristiques (h_m).

La traduction automatique statistique a donné un nouvel élan à la TA et à la recherche dans le domaine. Les résultats demandent toujours une part d'intervention humaine, mais ils étaient incontestablement meilleurs. Un autre avantage considérable de cette approche est qu'elle permet de développer assez rapidement un système de TA pour une paire de langues, à condition, bien sûr, de disposer de corpus de taille suffisante. Un des corpus utilisés est le corpus européen EUROPARL qui comptait en 2012 60 000 000 de mots par langue, pour 21 langues. Cela présente un avantage considérable sur les systèmes à base de règles, qui demandent un temps de développement considérable. Le projet Euromatrix, sous la conduite de Philip Koehn, a débouché sur le moteur de TAS ouvert et libre Moses, ce qui a stimulé la recherche dans ce domaine.

7.2.2 Les forces et les faiblesses de la TAS

Même si la TAS se révèle meilleure que la TA à base de règles, elle est loin d'être parfaite : des syntagmes peuvent être corrects, mais la phrase ne le sera

pas forcément dans son ensemble. La dimension textuelle n'est pas prise en compte, et cela se ressent dans le manque de fluidité du texte, dans la faiblesse des liaisons et des reprises pronominales. Elle gère mal les syntagmes discontinus, mais peut être efficace sur le plan terminologique.

Ainsi, si nous reprenons l'exemple skip/jump, une bonne traduction est possible en cas de syntagme continu :

La femme de ménage a sauté un repas → *The maid skipped a meal*

Mais ce sera plus compliqué si le syntagme est discontinu :

La femme de ménage a sauté, encore une fois, un repas →

The maid jumped again a meal (Li 2013)

On a vu qu'un avantage des méthodes fondées sur les corpus est de permettre un développement rapide d'un système de TA, à condition de disposer de corpus de taille raisonnable. L'absence de corpus bilingue suffisant est à l'inverse un obstacle. La recherche actuelle se penche d'ailleurs sur des méthodes permettant de faire de la TA avec des langues disposant de peu de corpus écrits.

7.2.3 Les systèmes d'évaluation automatique (2000-)

En parallèle à la TAS se sont développées les méthodes d'évaluation automatique.

Pour tester rapidement l'effet d'une modification apportée au système, pour tester l'emploi d'une variable ou son réglage, il fallait développer des systèmes permettant une évaluation rapide et automatique de la traduction produite. Une des premières métriques à cet effet est la métrique BLEU (*Bilingual Evaluation Understudy*), proposée par Papineni, Roukos, Ward et Zhu (2002). L'idée fondamentale de BLEU est que, plus un texte traduit automatiquement est proche d'un ensemble de textes de référence, traduits humainement, meilleur il est. BLEU va donc comparer les phrases d'un texte à un ensemble de phrases de référence, et ensuite calculer le score pour l'ensemble du texte. Il faut un texte d'une certaine longueur pour que l'évaluation soit pertinente, typiquement, un millier de phrases (Hearne et Way 2011 : 16).

Plus précisément, le score représente la proportion de séquences de n mots présents dans la traduction candidate et également présents dans la traduction de référence. Le résultat sera un nombre compris entre 0 et 1 ; plus la valeur est élevée, plus la traduction candidate est considérée comme bonne. BLEU a pu être critiqué à la fois pour ses limites techniques (il est moins efficace pour des phrases courtes) et pour des raisons de fond (l'amélioration du score ne signifie pas nécessairement que la traduction est meilleure, car il peut y avoir différentes manières de bien traduire ; il privilégie les correspondances de longueur, lexicales et d'ordre des mots et le rappel de manière générale), il n'en reste pas moins le plus utilisé, malgré le développement d'autres métriques : NIST, METEOR, ROUGE, WER...

7.3. Les systèmes neuronaux (2016-)

7.3.1 Fonctionnement

Les années 2000-2010 voient donc le développement des systèmes statistiques, popularisés par Google Translate et Bing. Certes, les spécialistes de la traduction se méfient de la TA et en pointent les erreurs, souvent risibles ; il n'empêche qu'elle devient incontournable et que le traducteur devient de plus en plus souvent postéditeur. L'acceptation des mémoires de traduction aura été lente, celle de la TA le sera plus encore et devra attendre des systèmes d'une qualité suffisante pour que la postédition ne soit pas ressentie comme injurieuse par un traducteur chevronné. En revanche, nombreux sont les utilisateurs non professionnels de Google, Bing ou Systran qui, de toute façon, ne passeraient pas par les services d'un traducteur professionnel pour comprendre le sens d'une page, d'un message ou d'une publication sur Facebook.

Fin 2016, les utilisateurs de Google Translate remarquent un changement net de qualité. À la même période apparaît un nouveau moteur de traduction, DeepL, qui impressionne par ses résultats.

Ces systèmes, auxquels les autres ont rapidement emboîté le pas, se fondent sur des algorithmes d'apprentissage profond neuronaux (*deep neural learning*), méthode d'intelligence artificielle où des réseaux de neurones profonds apprennent, s'entraînent sur des données.

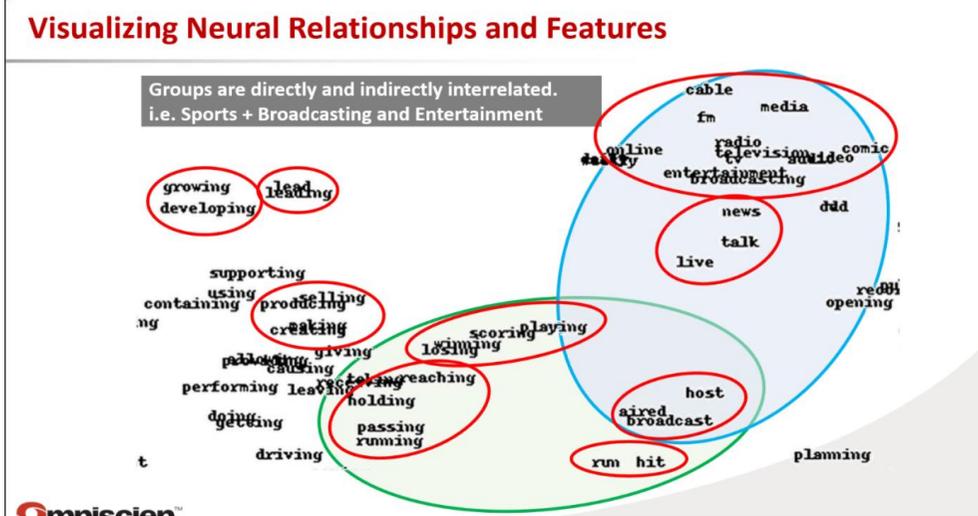
En fait, l'intelligence artificielle, dont on n'entendait plus beaucoup parler, a fait son grand retour en 2012, lorsqu'un algorithme appelé Supervision gagne un concours de reconnaissance d'images¹⁰. Des algorithmes nourris des milliards d'images disponibles sur Internet parviennent à distinguer une voiture d'un bus, un chat d'un chien, le rouge du bleu...

L'usage des réseaux de neurones est pourtant beaucoup plus ancien. En 1989, Yann LeCun et d'autres chercheurs chez AT&T Bell Labs étaient parvenus à appliquer un algorithme de rétropropagation à un réseau neuronal multicouche pour lui faire reconnaître des codes postaux. Mais, vu les limites du matériel de l'époque, il fallait trois jours pour accomplir la tâche. Ce qui a permis le renouveau de l'IA au début des années 2010, c'est l'amélioration des algorithmes, mais aussi du matériel informatique (on utilise désormais des processeurs graphiques cadencés à un milliard d'instructions par seconde) et enfin la disponibilité d'immenses corpus de données. Supervision s'est nourri des milliards d'images étiquetées disponibles sur des banques comme Imagenet. Car ces systèmes sont des systèmes d'apprentissage *supervisés* : le réseau de neurones, qui comprend plusieurs couches, calcule un résultat probable, renvoie le résultat qui est comparé au résultat réel (ce qu'on appelle la rétropropagation). Petit à petit, le système peut affiner ses critères, sa pondération et se rapprocher du résultat attendu.

En matière de traitement du langage naturel et de traduction, le texte va être modélisé par *plongement lexical* (*word embedding*). Un réseau de neurones artificiels à deux couches va représenter le contexte linguistique de mots ou de séquences de mots. Le modèle Word2vec, utilisé par Google, se nourrit de grand corpus de textes et produit un espace vectoriel, comprenant en général plusieurs centaines de dimensions, chaque mot unique dans le corpus se voyant attribuer

un vecteur correspondant dans l'espace. Les vecteurs de mots sont positionnés dans l'espace vectoriel de telle sorte que les mots qui partagent des contextes communs dans le corpus sont situés à proximité les uns des autres dans l'espace.

Prenons comme exemple cet espace vectoriel (Koehn et Wiggins 2018 : 46) :



On observe que des mots ou expressions de sens proche seront proches dans l'espace vectoriel : *growing* et *developing* ; *news*, *talk* et *live*.

Voici un autre exemple proposé par LeCun, Bengio et Hinton (2015 : 441) :

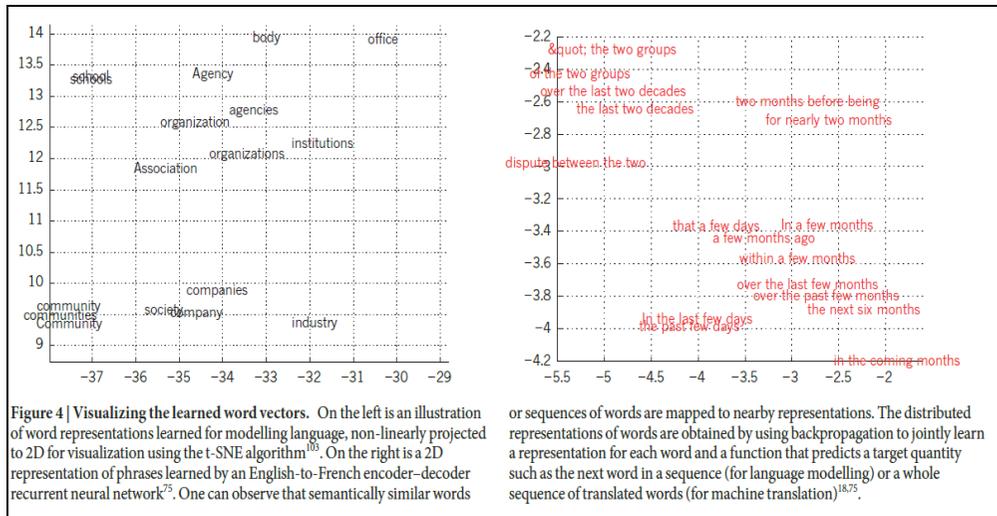


Figure 4 | Visualizing the learned word vectors. On the left is an illustration of word representations learned for modelling language, non-linearly projected to 2D for visualization using the t-SNE algorithm¹⁸. On the right is a 2D representation of phrases learned by an English-to-French encoder-decoder recurrent neural network¹⁹. One can observe that semantically similar words

or sequences of words are mapped to nearby representations. The distributed representations of words are obtained by using backpropagation to jointly learn a representation for each word and a function that predicts a target quantity such as the next word in a sequence (for language modelling) or a whole sequence of translated words (for machine translation)^{18,75}.

Ici aussi, on observe que des mots, termes ou syntagmes proches de sens sont proches dans l'espace vectoriel.

7.3.2 Forces et faiblesses

La TAN a d'emblée impressionné par sa fluidité. En revanche, elle est peu précise en terminologie. Et cela se comprend : des mots ou expressions proches de sens seront proches dans l'espace vectoriel. Dès lors, on peut avoir des contresens tels que

La TAN est plus sensible à la qualité des corpus. Et, comme elle ne se décompose pas en une série de sous-systèmes clairement identifiés et paramétrables comme la TAS, il est d'autant plus important pour un utilisateur industriel de disposer de corpus importants, de bonne qualité, dans son domaine de spécialité. La TAN demande des corpus plus grands et de meilleure qualité que la TAS. On compte des corpus d'au moins 1 à 5 millions de phrases bilingues, correctes et spécifiques au domaine considéré en TAS ; en TAN, l'ordre de grandeur est d'un milliard de mots (Koehn et Wiggins 2018 : 21). Ces corpus seront déterminants pour donner un contexte, déterminer des caractéristiques stylistiques et augmenter la précision terminologique et la fluidité.

La TAN utilise des réseaux de neurones profonds. Le système est un système global, obscur, et il est donc plus difficile de le paramétrer finement, sauf à reconstruire l'ensemble. On a aussi observé des biais de genre ou de race, qu'il est difficile de lever.

Enfin, ce sont des systèmes exigeants en développement et consommateurs d'énergie. Il s'agit bien de réseaux de neurones artificiels, et ils sont au moins 10 fois plus consommateurs d'énergie que le cerveau du traducteur humain (LeCun 2017).

La TAN est aussi plus sensible à la qualité de l'entrée et au domaine : sa qualité diminue si une phrase contient des erreurs, un message à la syntaxe orale, une coquille... ou si un texte ne correspond pas au domaine qui a servi à entraîner le système. Elle est plus sensible également à la longueur des phrases, et produit parfois ce que certains ont appelé, non sans humour, du « neurobable » :

There was no clear correlation between the measured mass density and the measured mass density, and neither experiment A or B. (Hasler 2018: 14)

Malgré toutes ces précautions, il semble bien que l'on soit arrivé – pour combien de temps ? – à un modèle de TA dont les résultats sont utilisables en postédition. L'inconvénient de la TAN serait aussi sa fluidité : la traduction donne l'impression d'être correcte, mais elle demande d'autant plus d'attention. L'observation du comportement de nos traducteurs apprenants montre que le temps d'attention est plus important sur une TAN que sur une TAS, mais que le temps d'édition est plus bref. Ceci correspondrait à l'intuition qu'il faut à un traducteur non expérimenté davantage de réflexion pour identifier les problèmes, mais qu'il lui faut moins de temps pour postéditer à proprement parler.

8. Conclusions

8.1. Quelle TA pour quel usage ?

La TA et les OAT, en règle générale, ont la réputation de fonctionner dans certains domaines et pour certains types de textes, de préférence non littéraires. Si l'on se réfère aux fonctions du langage de Jakobson, ce sont les textes référentiels qui se prêtent le mieux à ces techniques. Les bulletins

météorologiques se prêtent à la TA, davantage qu'un article scientifique, car le bulletin est référentiel alors que l'article relèvera davantage de la fonction conative, si l'auteur entend démontrer une hypothèse. De même, il peut être extrêmement difficile de faire traduire un article de journal de type éditorial, qui fait allusion à une réalité immédiate, use de jeux de mots ou exprime la conviction de l'auteur.

Quoi qu'il en soit, on a vu des propositions d'emploi de la TA en littérature. Cela peut faire sourire, mais on pourrait inverser la proposition et s'interroger sur une littérature traduisible par machine. De plus en plus de colloques et de revues consacrent un numéro ou une session à cette question. Ce fut le cas des *Assises d'Arles* en 2018, du très important *Machine Translation Summit* de 2019 (EAMT 2019), ce sera le cas du prochain numéro de *La Main de Thôt*.

De toute façon, on sait que la TA, malgré ses limites, est utilisée, que ce soit comme une première étape de traduction ou brute, car dans certains cas, « c'est de la TA, sinon rien ». On a pu ainsi être très heureux de profiter de la TA dans des situations d'urgence. Mais c'est aussi la TA que chacun utilise chaque jour pour juger de l'utilité d'un courriel, pour comprendre un mode d'emploi, pour traduire un passage en chinois ou en coréen qui s'affiche dans une série, ou pour sous-titrer en direct un documentaire ou une série.

Le fait est que la TAN fait moins rire les traducteurs, car dans certains cas, et avec une certaine dose de postédition, elle devient un véritable outil d'aide à la traduction.

La question de la postédition devrait être plus largement développée. Elle fait aujourd'hui l'objet d'une norme ISO (18587), qui définit deux niveaux de postédition :

Légère : le sens doit être correct et la phrase compréhensible ;

Complète : le résultat doit être comparable à celui d'une traduction humaine.

De façon très pragmatique, le traducteur à qui l'on propose un tel travail devra évaluer en fonction du prix proposé et de la qualité de la TA, en faisant un test pour quantifier sa charge de travail, si la postédition est une option rentable. (Vandervorst 2020).

8.2. Des systèmes et des erreurs

Comme la TAS a supplanté les systèmes à base de règle, la TAN tend à supplanter la TAS. Cela dit, pour des langues qui disposent de peu de corpus écrits, ou pour des combinaisons rares, les systèmes plus anciens restent une possibilité. Les systèmes à base de règles ont également des vertus pédagogiques et scientifiques, car ils se fondent sur une vraie interrogation de ce que sont les mécanismes langagiers de la compréhension, de la traduction et de la production.

On a vu que les problèmes ne sont pas les mêmes, et le développeur comme le postéditeur devront y être attentifs.

TA à base de règles	TA statistique	TA neuronale
Erreurs flagrantes	Problèmes des	Fluidité → les erreurs

	syntagmes discontinus	peuvent passer inaperçues ; nécessite une bonne connaissance du domaine
La terminologie peut être adaptée	Précision	Manque de précision
Temps de développement humain important	Exigeant au niveau des performances	Très exigeant au niveau des performances
N'utilise pas de corpus	Demande de grands corpus, de bonne qualité (min. 1 million de phrases bilingues)	Demande de très grands corpus, de très bonne qualité (min. 1 milliard de phrases bilingues)
Sensible à la qualité de la source (ce qui n'a pas été décrit ne peut être traité)	Plus robuste, moins sensible à la qualité de l'input	Sensible à la qualité de la source, au domaine et à la longueur des phrases
Erreurs prédictibles	Erreurs relativement prédictibles et améliorations possibles	Erreurs imprédictibles et améliorations difficiles à apporter rapidement
Erreurs syntaxiques flagrantes et nombreuses	Erreurs syntaxiques flagrantes	Neurobale évident

8.3. L'image de la TA

L'image de la TA a parcouru des montagnes russes. À l'optimisme excessif du début a succédé la déception. Diverses tentatives fondées sur la recherche en linguistique computationnelle, en statistique et en intelligence artificielle ont offert tour à tour de l'espoir et des déceptions. Même si le grand public pouvait rire des erreurs de la TA et la balayer comme une alternative à la traduction humaine, les spécialistes mesuraient les progrès ou sa faisabilité dans certains domaines ou pour certaines applications précises. La TA devient une phase dans un processus de traduction, éventuellement précédée d'une pré-édition et d'une description terminologique et phraséologique, et nécessairement suivie d'une phase de postédition.

Un des grands malentendus a certainement été de voir en la TA une « FAHQMT », une traduction complètement automatique et de haute qualité, à savoir une traduction automatique équivalente à une traduction humaine suivie d'une révision. On a vu que les chercheurs lucides ou honnêtes n'ont pas prétendu à un tel niveau de qualité. Il n'empêche qu'aujourd'hui, si DeepL ou Google Translate ne remplacent pas le traducteur professionnel, leur qualité est suffisante dans bien des situations de communication.

8.4. Où se situe-t-on sur le triangle de Vauquois ?

Vauquois a développé son triangle comme une représentation des systèmes à base de règles, pour visualiser la proportion entre l'analyse, le transfert et la génération. Dans la pratique, aucun système à base de règles n'a jamais été tout à fait pur. Beaucoup ont fait une analyse syntaxique en intégrant des traits sémantiques.

Ce triangle a été conçu avant l'apparition des méthodes statistiques et neuronales, mais on peut assez facilement situer les systèmes statistiques sur la ligne du triangle qui réalise une analyse syntaxique, même si les règles en sont fort différentes.

En revanche, si on voit les plongements lexicaux, se pourrait-il que les systèmes neuronaux se rapprochent d'une interlangue ?

8.5. Warren Weaver et les méthodes symbolistes

Dans son *Memorandum*, Warren Weaver hésite entre deux conceptions d'une TA qui descendrait dans les profondeurs des langues. D'un côté, il pense à des universaux de la langue, qui seraient liés au fait que tous les humains partagent un environnement, une existence et un corps assez proches ; de l'autre, il fait allusion à des propriétés statistiques, à la théorie mathématique de Shannon, à l'expérience d'un cryptographe qui parvient à décoder un message sans savoir qu'il a été écrit en turc, et enfin au théorème de McCulloch et Pitts de 1943 (voir plus haut p. 4).

On peut donc s'interroger sur ce qu'il entend par « logique » à propos du langage. S'il s'agit de retrouver à travers la diversité des langues une langue commune ou des universaux, il est à craindre que l'histoire de la linguistique et du traitement automatique des langues ne lui ait guère donné raison. S'il s'agit de retrouver des régularités formelles qu'un algorithme puisse percer, il a peut-être eu l'intuition des systèmes statistiques et neuronaux.

8.6. L'avenir

Alan Melby dans un article récent sur la postérité des idées de Warren Weaver, affirme que sans représentation symbolique du sens, la TAN connaîtra à son tour ses limites (2020 : 435). Cependant, la visualisation des espaces vectoriels des plongements lexicaux fait bien penser à une représentation symbolique, et d'une langue à l'autre, ces espaces sont superposables, ce qui permet potentiellement de traduire entre des couples de langues pour lesquels on n'a pas de corpus parallèle (Hutson 2017). Est-ce que ces mots et ces expressions qui voisinent dans un espace avec d'autres mots et expressions de la même famille sémantique ne constitueraient pas une certaine représentation du sens ? Est-ce qu'un système à base de règles, avec une vraie représentation symbolique à différents niveaux linguistiques, représentation symbolique qui a été codée par un être humain, comprend le sens de la phrase qu'il traduit mieux qu'un système neuronal ?

Il n'en reste pas moins vrai qu'ils sont des boîtes noires, que le résultat est non déterministe et qu'il est difficile d'interagir en cas d'erreur, sauf à réentraîner le système. Une autre limite est qu'ils ne sont pas tout à fait non supervisés.

Certes, ils apprennent, mais ils n'apprennent que des données qu'on leur soumet. Les spécialistes de l'intelligence artificielle, comme LeCun, Bengio et Hinton (2015 : 442) soulignent qu'un véritable apprentissage non supervisé, c'est celui du bébé qui vient de naître et qui pourra un jour mettre l'objet carré dans l'ouverture carrée, construire un jouet en pièces détachées, lire, écrire, traduire, conduire une voiture, écrire un roman d'anticipation... Les systèmes de TAN, aussi impressionnants soient-ils, ne font jusqu'à présent qu'une seule tâche : traduire.

Notes

¹ Lettre au Père Marin Mersenne du 20 novembre 1629 (Mounin 1964, 16)

² Louis Couturat et Léopold Leau *Histoire de la langue universelle* 1903 ainsi qu'Umberto Eco (*La recherche de la langue parfaite dans la culture européenne*, 1997) développent ces diverses langues universelles.

³ Il y a effectivement des discussions sur la paternité et la date du premier « ordinateur ». Comme le soulignent Lazard et Mounier-Kuhn : « S'il [l'ENIAC] n'est ni le premier calculateur électronique (l'ABC et le Colossus lui sont antérieurs), ni le premier calculateur programmable (le Z3 et le Mark 1 l'étaient avant lui), son importance est proportionnelle à l'ampleur du projet et à son retentissement, à la publicité faite après la guerre autour de lui. Il permet également de démontrer la viabilité de l'électronique, et inspire à plusieurs chercheurs le désir d'aller plus loin. » (2016 : 86)

⁴ *Automatic Language Processing Advisory Committee*.

⁵ Ceci fait inévitablement penser à la théorie interprétative de la traduction développée par Danica Seleskovitch et Marianne Lederer dans les années 1970 et qui débouchera en 1984 sur l'ouvrage *Interpréter pour traduire*. Cependant, le modèle de Vauquois est antérieur et repose sur des procédures d'analyse linguistiques, alors que Seleskovitch et Lederer veulent s'écarter de l'approche linguistique. Sans doute leur théorie du sens serait-elle plus proche des modèles utilisés en intelligence artificielle, comme on le verra plus bas.

⁶ Voir (Hutchins 1986) pour une description détaillée de tous ces systèmes.

⁷ Un prototype avait été créé en 1975-1976.

⁸ Mais, selon Hutchins (1986 : 231), c'est en raison des coûts de développement que le système a été refusé. L'approche sous-langage exige en effet le développement de dictionnaires très importants.

⁹ Le terme vient de Lucien Tesnière, dans sa grammaire de dépendance, exprimée dans ses *Éléments de syntaxe structurale* (1959).

¹⁰ Geoffrey Hinton, Alex Krizhevsky et Ilya Sutskever, à l'université de Toronto.

Références

- Brown, Peter F. et al. (1988) : « A statistical approach to language translation », in *Coling Budapest: Proceedings of the 12th International Conference on Computational Linguistics*, Dénes Varga ed., Budapest, ACL, pp. 71-76.
- Couturat, Louis et Leau, Léopold (1903) : *Histoire de la langue universelle*, Paris, Hachette
- Daumas, Maurice (1965) : « Les machines à traduire de Georges Artsrouni », in *Revue d'histoire des sciences*, tome 18, n°3, pp. 283-302.
- EAMT (2019) : *Machine Translation Summit XVII - The Qualities of Literary Machine Translation*, Dublin, ADAPT, DCU, EAMT.
- Eco, Umberto (1997) : *La recherche de la langue parfaite dans la culture européenne*, traduit par Jean-Paul Manganaro, Paris, Éditions du Seuil.

- Gazdar, Gerald, Klein, E., Pullum, G. et Sag, I. (1985) : *Generalized Phrase Structure Grammar*. Cambridge, Massachusetts, Harvard University Press
- Harris, Brian (1988) : « Bi-Text: A New Concept in Translation Theory », in *Language Monthly*, n°54, pp. 8-10.
- Hasler, Eva (2018) : « Turning NMT research into commercial products. », in *Smart Computer-Aided Translation Environment Final Workshop*, http://www.ccl.kuleuven.be/scateworkshop/slides/11%20-%20SCATE_EHasler.pdf, consulté le 15 mars 2020.
- Hearne, Mary et Way, Andy (2011) : « Statistical Machine Translation: A Guide for Linguists and Translators » in *Language and Linguistics Compass*, n° 5, pp. 205-226.
- Hutchins, John (1986) : *Machine Translation. Past, Present, Future*, New York - Chichester - Brisbane – Toronto, Ellis Horwood Limited.
- Hutchins, John (1998) : « The origins of the translator's workstation », in *Machine Translation* 13, n°4, pp. 287-307.
- Hutson, Matthew (2017) : « Artificial intelligence goes bilingual—without a dictionary », in *Science*, 11, <https://www.sciencemag.org/news/2017/11/artificial-intelligence-goes-bilingual-without-dictionary>, consulté le 15 mars 2020.
- Koehn, Philipp et Wiggins, D. (2018) : *Secrets to Building a High-Quality Neural Machine Translation Engine*, <http://omniscien.com/more/resources/webinar/>, consulté le 15 mars 2020.
- Koehn, Philipp, Och, F. J. et Marcu, D. (2003) : « Statistical Phrase-Based Translation », in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - NAACL 2003*, pp. 48-54.
- Lazard, Emmanuel et Mounier-Kuhn, Pierre (2016) : *Histoire illustrée de l'informatique*, Paris, EDP Sciences.
- LeCun, Yann (2017) : « Yann LeCun et l'intelligence artificielle », conversation dans le cadre de *Le Monde Festival*, 3 octobre 2017, <https://www.ultimedia.com/deliver/generic/iframe/mdtk/01637594/src/rfk8km/zone/1/showtitle/1/>, consulté le 15 mars 2020.
- LeCun, Yann, Bengio, Y. et Hinton, G. (2015) : « Deep Learning », in *Nature* 521, n°7553, pp. 436-444.
- Li, Gong (2013) : *La traduction automatique statistique, comment ça marche ?* <https://interstices.info/la-traduction-automatique-statistique-comment-ca-marche/>.
- Melby, Allan K. (2020) : « Future of machine translation: musings on Weaver's memo », in *The Routledge Handbook of Translation and Technology*, Minako O'Hagan ed., Abingdon-Oxon; New York, Routledge, pp. 419-435.
- Merten, Pascaline (2005) : *La caractérisation multiple en français. Description, comparaison avec d'autres langues et formalisation XML*, 3 vols., thèse de doctorat, Bruxelles, ULB, 2005.
- Mounin, Georges (1964) : *La machine à traduire; histoire des problèmes linguistiques*, The Hague, Mouton.

- Papineni, Kishore, Roukos, Salim, Ward, Todd et Zhu, Wei-Jing (2002) : « BLEU: A Method for Automatic Evaluation of Machine Translation », in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL 2002*, Philadelphia, ACL, pp. 311-318.
- Tesnières, Lucien (1959) : *Éléments de syntaxe structurale*, Paris, Klincksieck.
- Vandervorst, Sarah (2020) : *Séminaire de postédition, ULB 6 mars 2020*, Bruxelles, ULB, 6 mars 2020.
- Vauquois, Bernard (1971) : « Modèles pour la traduction automatique », in *Mathématiques et sciences humaines* 34, pp. 61-70.
- Weaver, Warren (1949) : *Memorandum on translation from one language to another*, New York, The Rockefeller Foundation.