4

Traduction automatique
la révolution de l'intelligence artificielle

Grâce aux progrès de l'intelligence artificielle, la traduction automatique connaît un essor sans précédent. Quelles en sont les performances réelles ? Quels usages actuels et à venir ?

Enquête de Barbara Vignaux

Adobe Stock

En bref

Après les succès de la reconnaissance vocale, bienvenue dans la nouvelle ère de la traduction automatique ! Imaginez un monde où il ne serait plus nécessaire de connaître la langue de l'autre pour discuter avec lui… On en est encore loin, mais de nouveaux horizons s'ouvrent grâce aux récents progrès de l'intelligence artificielle. Dernier épisode en date : il y a tout juste un an, Google annonçait triomphalement le lancement de son nouvel outil de traduction en ligne, qui réduit les erreurs de traduction – selon la firme américaine – de 55 à 85 %. Basé sur la technologie des « réseaux de neurones » en partie inspirée du fonctionnement cérébral, cet outil se révèle particulièrement efficace pour traduire des textes du chinois vers l'anglais – une tâche que les précédentes technologies peinaient à réaliser. C'est un progrès réel. Pour autant, le rêve – ou le fantasme – d'un boîtier universel capable de tout traduire dans toutes les langues n'est pas pour demain. Et tant mieux peut-être…

La révolution des réseaux de neurones

Les outils récents de traduction automatique tirent leur architecture bio-inspirée de l'intelligence artificielle.

Depuis 2016, les « réseaux de neurones » équipent les traducteurs automatiques de Google (Google Translate) et Microsoft (Skype Translator). Depuis l'été 2017, c'est aussi le cas de Facebook – 2,5 milliards de traductions par jour – et de l'Européen DeepL (ex Linguee). Il s'agit d'algorithmes d'apprentissage dont l'architecture s'inspire du fonctionnement cérébral, avec ses cellules nerveuses ou « neurones » (les unités de calculs) reliés par des « synapses » qui assurent la transmission des informations. Les neurones sont répartis en couches – jusqu'à vingt dans certains réseaux – que les données d'entrée empruntent successivement jusqu'à la sortie du réseau. Ce processus permet de construire une représentation abstraite des unités de la langue – mots, groupes de mots, phrases –, transformées en valeurs numériques. Durant la phase d'apprentissage, les réseaux sont ainsi « nourris » de données disponibles en grandes quantités – documents multilingues des Nations unies ou corpus en ligne comme Wikipédia… Ils comparent les données d'entrée (texte en anglais, par exemple) et de sortie (même texte en français), vérifient que les paramètres de traduction sont justes et, si besoin, les corrigent. « *On ne donne donc pas à la machine la réponse, mais les moyens de la trouver* », résume l'expert Jean Senellart. Cependant, les informaticiens ignorent comment la machine procède précisément. En attendant, les réseaux de neurones n'ont pas éclipsé les deux technologies antérieures : traduction *rule-based* indispensable pour traduire des langues aux corpus limités (farsi, ourdou, somali…) ; et traduction statistique toujours largement diffusée.

Les trois méthodes de la traduction automatique

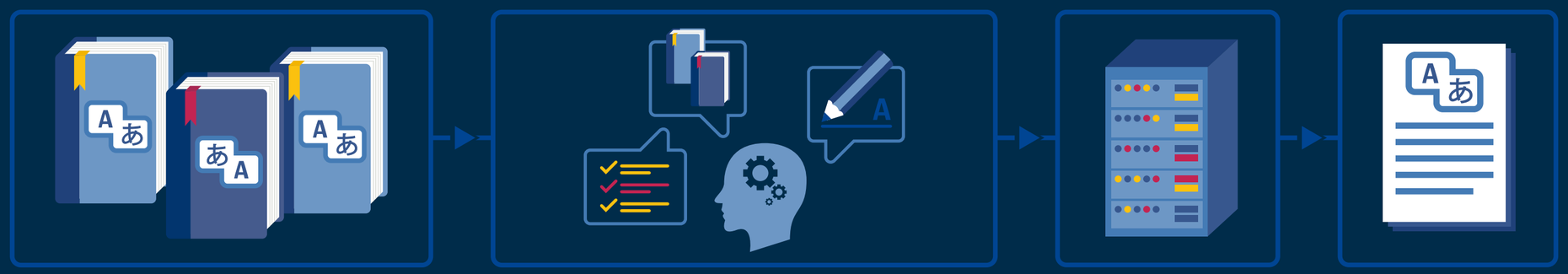
«RULE-BASED»

de 1950 à nos jours

00010110
01011011
10010011
01101001
11010101

L'être humain code les règles de traduction.

Plusieurs millions de lignes de code



Rapidité, personnalisation possible des paramètres de traduction

Colossal travail de codage

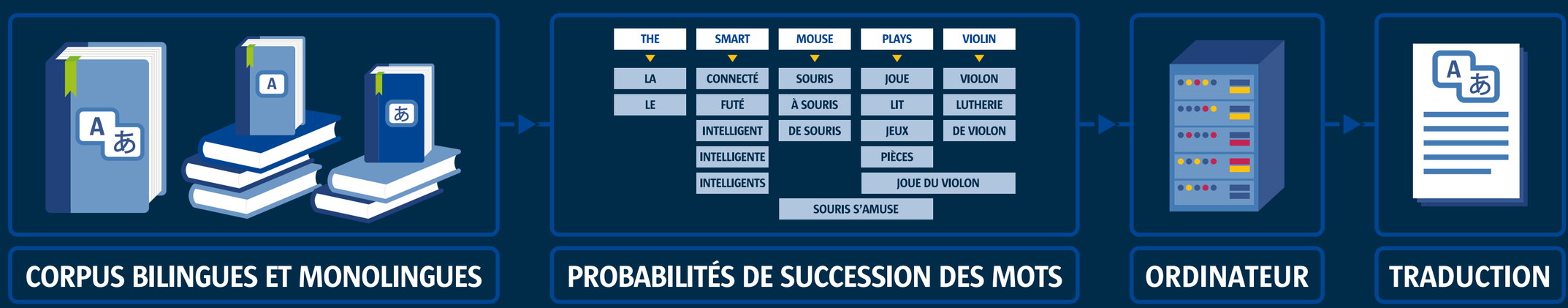
STATISTIQUE

de 1990 à nos jours

00010110
01011011
10010011
01101001
11010101

L'ordinateur tire les règles d'exemples de traductions manuelles.

1 million de lignes de code



Fluidité et volume des textes traduits

Faibles connaissances linguistiques

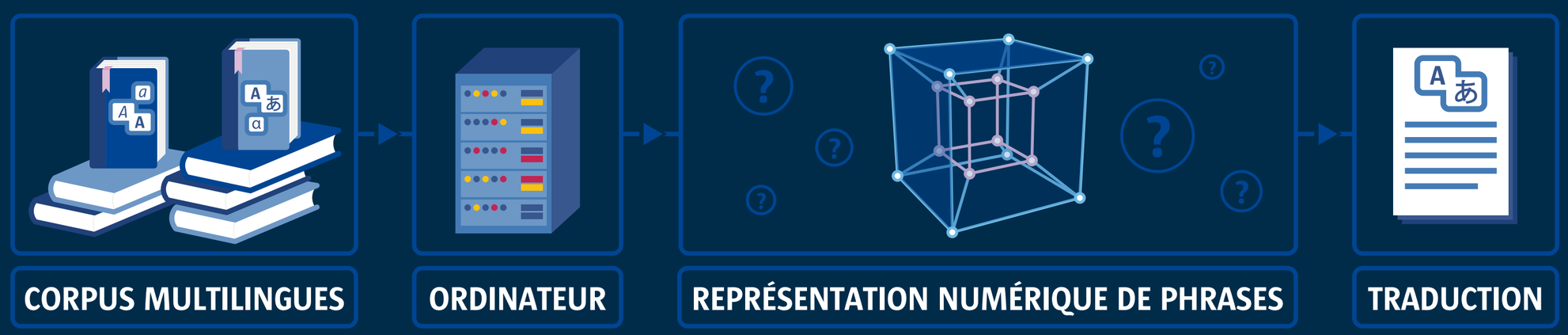
RÉSEAU DE NEURONES

de 2000 à nos jours

00010110
01011011
10010011
01101001
11010101

L'algorithme apprend à encoder et décoder des données.

5 000 lignes de code



Simplicité du codage, fluidité, corpus d'entrée moins volumineux

Impossibilité d'agir sur les paramètres de traduction maîtrisés par la machine seule

CERVEAU HUMAIN
100 milliards de neurones / 10 000 connexions (synapses) par neurone

RÉSEAU DE NEURONES
1 million de neurones / 100 à 1 000 connexions par neurone

Infographie : Julien Tredan-Turini

Des neurones artificiels, vraiment ?

Le terme « réseau de neurones » fournit une analogie descriptive – les neurones étant reliés par des synapses qui propagent des signaux – et en partie fonctionnelle : « *À l'image de notre cerveau, les réseaux neuronaux permettent d'effectuer beaucoup de tâches en parallèle, et non plus des tâches de type arithmétique, avec des opérations se succédant les unes aux autres* », explique Damien Querlioz, chercheur CNRS au C2N de l'université Paris-Sud. En outre, dans les réseaux de neurones artificiels et humains, les fonctions de calcul et de mémoire se trouvent sur les mêmes unités, contrairement aux ordinateurs ordinaires qui séparent les tâches.

Inserm/E. Eugène

Une intelligence militaire et artificielle

Apparue après la Seconde Guerre mondiale, la traduction automatique s'est développée au rythme des progrès informatiques.

À l'origine de la traduction automatique se trouve… la Guerre froide. Dans les années 1950-60, l'outil informatique apparaît en effet indispensable pour traduire les masses de publications scientifiques russes vers l'anglais. L'armée de l'air américaine fait alors appel à l'entreprise Systran : « *À l'époque, tout se fait sur de grosses machines et des cartes perforées !* », raconte Jean Senellart, directeur technique et innovation de cette entreprise successivement nord-américaine, française puis coréenne. Mais c'est bien plus tard, avec Internet, que la traduction automatique se développe à grande échelle. Dès 2003, l'application Babelfish est distribuée par le moteur de recherche américain AltaVista, pour enrichir la recherche documentaire en ligne. Dans la foulée, les géants du Web, comme Google ou Yahoo, mettent au point leurs propres technologies de traduction automatique (de type statistique). Moins d'une décennie plus tard, la diffusion de « l'apprentissage profond » (*deep learning*) – un concept qui remonte aux années 1980 – devient possible grâce à la puissance de calcul des nouvelles puces et à l'accès à des quantités colossales de données disponibles en ligne (dans ce qu'on appelle le *cloud*). Utilisée dans tous les domaines de l'intelligence artificielle – reconnaissance vocale, reconnaissance de formes, recherche en ligne, agents de dialogue virtuels ou *chatbots*… – cette technologie d'apprentissage informatique fondée sur les « réseaux de neurones » numériques bouleverse la traduction automatique.

Yann LeCun, père du réseau de neurones

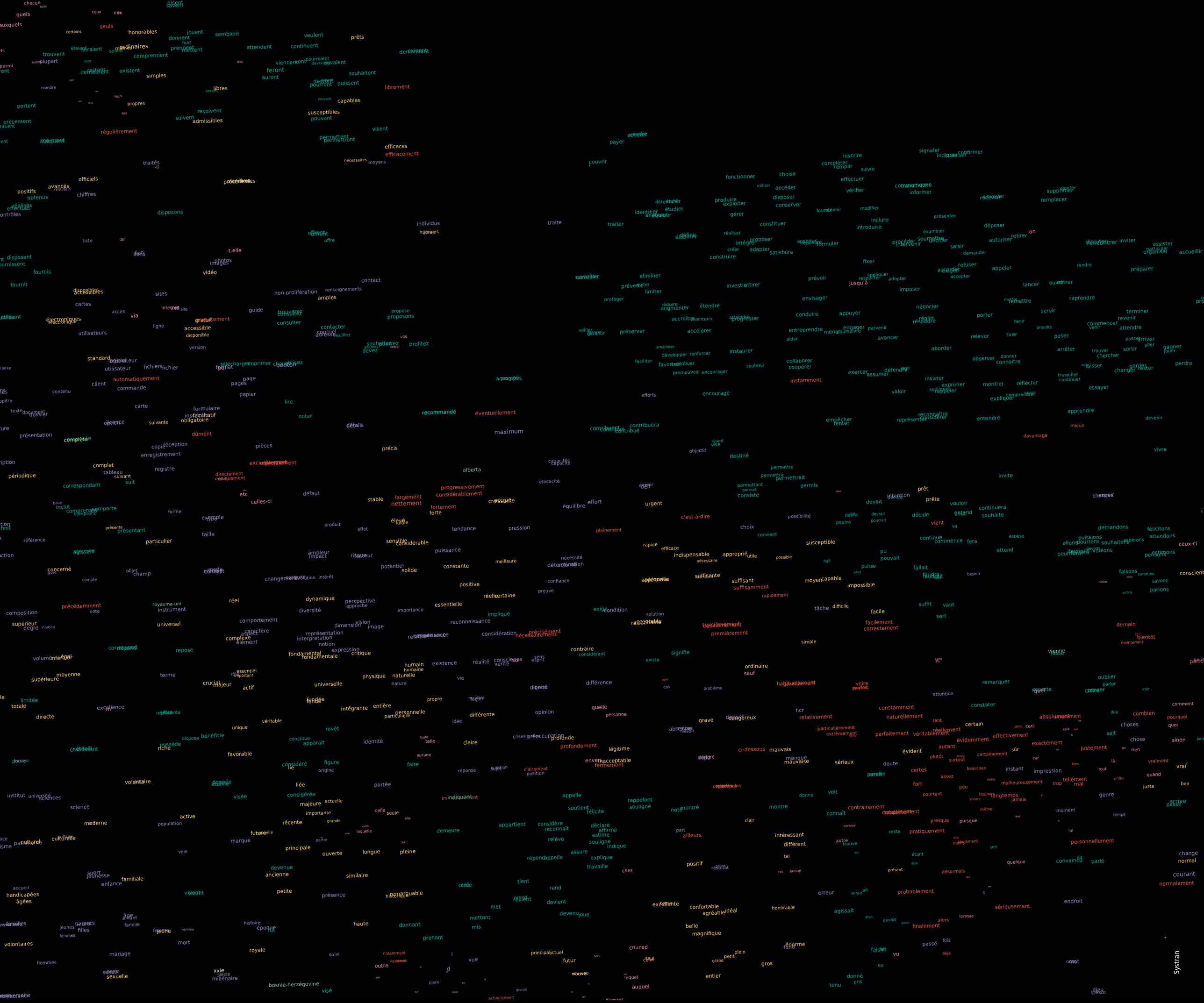
Cet ingénieur français est l'un des pères de l'intelligence artificielle (IA) : dès les années 1980, il a consacré sa thèse aux réseaux de neurones. Ayant achevé ses études universitaires en France puis travaillé dans le secteur privé aux États-Unis (Bell Laboratories, AT&T), ce génial touche-à-tout est aujourd'hui professeur à l'université de New York et directeur du centre de recherche de Facebook. Se méfiant de « *l'inflation de promesses* » née des prouesses de l'IA, il déclarait, durant sa leçon inaugurale au Collège de France, en 2016 : « *L'IA sera un amplificateur de notre intelligence, et non un substitut pour celle-ci* ».

James LEYNSE/REA

Géants du Web : une menace ?

Concentration des capacités de recherche, manque de confidentialité des données : les grands acteurs d'Internet suscitent des inquiétudes.

« *Ils ont les données, ils ont les meilleurs chercheurs, ils ont la puissance de calcul : difficile de rivaliser lorsqu'on est universitaire !* », lance Laurent Besacier, professeur au laboratoire d'informatique de Grenoble (CNRS/Inria/UGA). Les Gafam – Google, Apple, Facebook, Amazon et Microsoft, auxquels s'ajoute Baidu, en Chine – dominent la traduction automatique comme d'autres secteurs de l'intelligence artificielle et jouent un rôle moteur dans la recherche. Quitte à débaucher des universitaires confirmés ou en thèse, comme l'a appris à ses dépens le Centre de recherche en informatique, signal et automatique de Lille. « *On voudrait mettre en place un master de* machine learning*, mais avec quels enseignants ? Avec trois chercheurs en moins en un an, notre vivier est affaibli et amoindri* », s'inquiète Olivier Colot, son directeur. *A contrario*, la recherche conduite chez les grands du Web profite aussi à toute la communauté scientifique, grâce à la mise à disposition d'algorithmes de traduction automatique en open source, c'est-à-dire librement modifiables – une quinzaine aujourd'hui. La seconde interrogation que soulève l'omniprésence des Gafam est celle de la confidentialité des données. Plutôt satisfait de ses procédures de sécurité (messagerie sécurisée, identification par badge à l'entrée des bureaux), le responsable informatique d'un important groupe pharmaceutique a un jour déchanté en réalisant que tous ses brevets étaient traduits en ligne sur Google Translate, sans aucune protection ! Les opérateurs privés préfèrent donc s'adresser à des entreprises de traduction automatique qui assurent la confidentialité des données.



Des nuages de mots aux logiques diverses

Dans son travail de traduction, le réseau de neurones crée des espaces multidimensionnels dans lesquels il répartit le vocabulaire sous forme de données numériques. Ce classement prend la forme de « nuages de mots », c’est-à-dire d’associations entre termes obéissant à diverses logiques : grammaticale (noms, verbes, prénoms…), sémantique (géographie, institutions politiques…), fonctionnelle (mots de liaison…). Ici, une expérience conduite après l’entraînement d’un réseau de neurones sur la base de données constituée par Wikipédia en français et représentée en deux dimensions ; à vous de découvrir les clés du classement !

Quels usages et limites ?

La traduction automatique excelle dans certaines tâches, mais échoue dans d'autres. Tout dépend du besoin.

« *Ce qui fait la force de la traduction automatique, c'est moins sa qualité que sa valeur d'usage* », constate François Yvon, chercheur au LIMSI/CNRS. Le service offert aux internautes par Facebook ou TripAdvisor suffit à traduire des contenus simples – informations touristiques ou produits grand public, par exemple. Ses usagers en connaissent implicitement les limites et s'amusent d'erreurs parfois grossières. Voire de contresens ; *smoke eater scented candle*, par exemple, devient : « bougie parfumée à la fumée » (au lieu de bougie parfumée anti-tabac). La traduction automatique peut toutefois s'avérer très efficace pour des usages professionnels comme la mise à jour annuelle de manuels techniques ou de documents internes d'entreprise, ou appliqué à un domaine précis comme la météo. À l'instar de Systran, des opérateurs spécialisés proposent d'ailleurs d'adapter les algorithmes aux besoins sectoriels de leurs clients – un secteur au chiffre d'affaires mondial estimé à 150-200 millions d'euros annuels. Dans tous les cas, l'algorithme de traduction automatique décompose le texte en phrases et se révèle donc, selon François Yvon, « *incapable de traduire des phénomènes qui dépassent la phrase, par exemple les références pronominales (*it *en anglais,* versus *elle ou il en français), les registres de langue ou les choix lexicaux* ». Les textes destinés à la publication doivent donc toujours être relus et corrigés par des traducteurs professionnels. Enfin, la traduction automatique échoue – par nature – à restituer les termes inventés ou poétiques, les expressions orales ou les abréviations, les effets humoristiques ou les jeux de mots. Inutile d'y recourir pour traduire des œuvres littéraires !

L’humain vainqueur (pour le moment)

L’ordinateur a peut-être battu l’humain aux échecs et même au jeu de go (photo), mais il a échoué à remporter la palme de la meilleure traduction. Lors de la compétition organisée par l’Association internationale d’interprétation et de traduction, à l’université de Sejong, en Corée du Sud, en février 2017, les traducteurs professionnels ont en effet vaincu les trois programmes de traduction neuronale : Google Translate, Systran et Papago (application développée par le géant de l’Internet coréen Naver). Il s’agissait de traduire huit textes (littéraires et non-littéraires) en 50 minutes de l’anglais vers le coréen et du coréen vers l’anglais.

Le plus grand service de traduction au monde

Avec 2 300 employés, le service de traduction de la Commission européenne est le premier au monde, car l'Union européenne reconnaît les 24 idiomes de ses États membres (à titre de comparaison, les Nations unies n'ont que cinq langues officielles). L'ensemble des directives et des règlements doit donc être traduit dans ces 24 langues. Baptisé MT@EC, le système de traduction automatique communautaire repose sur les méthodes statistiques, mais évolue progressivement vers les réseaux de neurones, notamment pour certaines paires de langues (français-allemand par exemple). Mais les traducteurs professionnels relisent et corrigent tous les textes avant publication au *Journal officiel*.

iStock

À terme, un assistant virtuel ?

L'avenir de la traduction automatique pourrait bien être de devenir une fonction, parmi d'autres, d'un super assistant virtuel.

Plusieurs pistes sont actuellement explorées pour améliorer la qualité de la traduction automatique : intégrer des éléments contextuels, comme le niveau de langue (soutenu ou familier) ou les références aux phrases antérieurement traduites ; accroître la mémoire, actuellement très limitée, des algorithmes… À plus long terme, la traduction automatique pourrait être incluse dans les tâches d'un assistant virtuel également capable de rédiger des messages, prendre des rendez-vous, identifier des individus sur une photo, lire des QR codes… tout cela, sur tablette, téléphone portable ou objet connecté. Un scénario qui fait écho à l'ambition de Google, énoncée par son directeur de la recherche à Zurich, Emmanuel Mogenet : « *Le but ultime, c'est de permettre une conversation en langage naturel avec une intelligence artificielle. C'est le futur du moteur de recherche Google* » (magazine suisse *Bilan*, 2 février 2017). Pour y parvenir, il faudra toutefois doter les machines de la « *base de données du sens commun qui est au cœur de notre compréhension mutuelle entre humains* ». Une tâche colossale mais qui, à terme, pourrait remettre en cause une des définitions les mieux admises de l'être humain : « *Bientôt, le langage cessera sans doute d'être le propre de l'Homme* », prédit l'ingénieur Jean Senellart (Systran).

Des réseaux énergivores

Pour être entraînés, les réseaux de neurones nécessitent d'importantes capacités de calcul et sont donc voraces en énergie : ils en consomment 10 000 fois plus qu'un cerveau humain. L'accès à de telles puissances « *constitue un goulot d'étranglement de la recherche*, remarque le professeur en informatique François Yvon, *mais il devrait être résolu d'ici cinq ans* ». Côté grand public, la démocratisation des réseaux de neurones et donc des applications d'intelligence artificielle sera facilitée par la diffusion des GPU, processeurs très performants et peu énergivores. Certains téléphones portables en sont déjà pourvus.

Adobe Stock

Le fantasme du traducteur universel

Deux oreillettes, une connexion bluetooth et une application de traduction automatique et simultanée sur téléphone portable : c'est le dispositif imaginé par la start-up américaine WaverlyLabs, qui promet de « *faire tomber les barrières entre les langues* ». Un dispositif à 249 dollars déjà commandé par 25 000 personnes dès avant son lancement, à l'automne 2017. Ses limites prévisibles ? « *Le nombre de langues (cinq disponibles, versus 6 000 langues actives dans le monde), l'autonomie d'un dispositif nécessairement très consommateur d'énergie et enfin, la qualité de la traduction* », résume le chercheur Laurent Besacier. On est donc encore loin d'un boîtier universel de traduction !

Au service des langues « sous-dotées » ?

Selon Laurent Besacier, professeur au laboratoire d'informatique de Grenoble, la traduction automatique est autant un risque qu'une opportunité pour les langues peu présentes sur la Toile.

Durée : 2 min 20

In French only

DR

Demain, plus besoin d’être polyglotte ?

Comment fonctionne la reconnaissance vocale ? Quelle place occupent les logiciels de traduction automatique dans notre quotidien ? Un extrait de Futuremag (Arte France et Effervescence Doc)

Durée : 13 min

In French only