



HAL
open science

Que valent les traducteurs automatiques de la science-fiction ?

Frédéric Landragin

► **To cite this version:**

Frédéric Landragin. Que valent les traducteurs automatiques de la science-fiction ?. CNRS Le journal, 2020. hal-02934841

HAL Id: hal-02934841

<https://hal.science/hal-02934841>

Submitted on 9 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Que valent les traducteurs automatiques de la science-fiction ?

Frédéric Landragin

Lattice, CNRS, ENS, PSL University, Université Sorbonne Nouvelle

Dans la science-fiction, le traducteur automatique universel est une sorte de dispositif magique permettant aux civilisations humaines comme extraterrestres de communiquer sans effort et sans apprentissage préalable. Certains dispositifs sont clairement loufoques, mais d'autres ont-ils inspiré les chercheurs ?

Robots omniscients, pilules et bactéries à ingérer, ou même poisson « miraculeux », la science-fiction regorge d'étonnants systèmes de traduction automatique. Le célèbre *babel fish* imaginé par Douglas Adams dans *Le Guide du voyageur galactique* (1979) est ainsi une espèce de poisson que l'on glisse dans son oreille et qui traduit instantanément dans n'importe quelle langue tout ce que l'on vous dit. Tout comme le « comprimé informatico-traducteur » imaginé par Stanislas Lem dans *Les Voyages électriques d'Ijon Tichy* (1976) efficace dès son absorption : « *Instantanément, les bruits bourdonnants qui m'environnaient se transformèrent à mes oreilles en paroles parfaitement intelligibles.* » Sans oublier les bactéries traductrices de la série TV australienne *Farscape* (1999) que l'on s'injecte pour comprendre les langues étrangères. Hélas, les progrès des nanotechnologies, biotechnologies, informatique et sciences cognitives (NBIC) ne nous permettent pas encore de matérialiser ces propositions loufoques...

Du côté des robots, et donc de l'informatique, Robby de *Planète interdite* (1956) impressionne : « *si vous ne parlez pas anglais, je suis à votre disposition avec 187 autres langues.* » Et dans *Le Retour du Jedi* (1983) C-3PO se vante de connaître 6 millions de formes de communication, et donc d'être capable de traduire de l'une à l'autre ! On est ici dans le cas de traductions d'une langue connue vers une autre langue connue. Les films n'expliquent pas comment les robots procèdent mais la performance semble plausible.

Dans *Épigone. Histoire du siècle futur* (1659), l'abbé Michel de Pure décrit une sorte d'autel surmonté d'un dôme de cristal, doté d'un embout dans lequel il est possible de parler. Un « bon homme » utilise l'embout et le personnage principal, Épigone, peut alors comprendre ce qu'il dit. Cette communication avec l'*alien* – l'extraterrestre aussi bien que l'*étranger* – semble constituer la première

application de la traduction automatique en SF, bien avant avant son essor dans la dure réalité de la guerre froide. *Épigone* est aussi un exemple du cas de figure le plus complexe : celui de la traduction d'une langue source totalement inconnue vers une langue cible connue, sans aucune pierre de Rosette pour aider : question plausibilité, c'est moins convaincant...

C'est aussi le cas dans la nouvelle pionnière « Premier Contact » (1945) de Murray Leinster, à ne pas confondre avec le film du même titre réalisé par Denis Villeneuve et sorti en 2016. Elle montre des extraterrestres communiquant avec les humains via des ondes électromagnétiques, et développant un code artificiel qui joue le rôle d'intermédiaire, appelé de ce fait inter-langue ou *interlingua* et qui n'est pas une langue proprement dite (1). Deux traductions s'avèrent nécessaires à chaque échange : par exemple, de la langue des aliens vers le code artificiel d'abord, et du code artificiel vers la langue des humains ensuite. Ce code utilise des symboles pour désigner les objets du monde, ainsi que des diagrammes et images pour désigner les relations entre symboles. Le traducteur automatique, dont c'est peut-être la première apparition dans la SF moderne, est désigné en tant que machine à codage-décodage. Il ne s'agit pas d'un **langage machine**, mais bel et bien d'un langage intermédiaire conçu pour faciliter la communication, exactement comme l'idiome mécanique **« mechanese »** détaillé dans « *L'Enfant des étoiles* », langage à mi-chemin des capacités expressives des humains et des machines afin de permettre une communication efficace entre les deux.

Cette approche présente-t-elle un intérêt pour concevoir des traducteurs automatiques réels ? Les chercheurs se sont en effet un temps demandé si pour plus d'efficacité et de souplesse par rapport aux langues à traduire, des systèmes qui passent systématiquement par une inter-langue étaient pertinents. Des recherches se sont développées autour d'une inter-langue conservant les principes essentiels des langues (dictionnaire, grammaire) tout en laissant de côté – par abstraction – les spécificités de chaque langue. On parle d'*Interlingual Machine Translation* et il semble bien que science et science-fiction suivent la même voie. Du moins dans les grandes lignes, car les chercheurs ne s'inspirent pas du *mechanese*, ni même des codes numériques universels de Descartes et Leibniz, mais au contraire innovent avec des méthodes informatiques puis avec l'introduction massive de statistiques. Le traitement se fait au niveau des phrases, tient compte des contextes et repose sur d'énormes stocks d'exemples. Il s'avère beaucoup plus efficace que le « mot à mot » qui a depuis longtemps montré ses limites, notamment avec le « cerveau », créé en 1954 par IBM et l'université de Georgetown, premier programme informatique conçu pour traduire une soixantaine de phrases du russe vers l'anglais américain. Cette méthode ne requiert qu'un simple dictionnaire bilingue mais elle tombe dans le panneau à chaque fois qu'une ambiguïté apparaît (lien web vers <https://lejournal.cnrs.fr/articles/les-androides-de-la-saga-alien-sont-ils-pour-2093>), c'est-à-dire à peu près à chaque phrase (2)...

Dans la pratique, plus il existe de données disponibles sur les langues (on peut parler de *big data*), plus la traduction de l'une à l'autre s'avère de qualité. En revanche, traduire depuis ou vers une langue peu décrite s'avère délicat. En fin de compte, c'est alors une langue bien connue, l'anglais, qui joue souvent le rôle d'inter-langue : on parle dans ce cas de **langue-pivot**.

Avec l'essor de l'**apprentissage profond** (ou *deep learning*) à l'aide des **réseaux de neurones** de l'intelligence artificielle (IA), les travaux actuels explorent de nombreuses voies qui cherchent toutes à minimiser l'intervention humaine. Personne ne construit plus d'inter-langue, et le développement de méthodes d'apprentissage multilingue conduit à se passer de langue-pivot. En revanche, l'apprentissage profond pose de nouvelles questions. Rappelons qu'avec ce type d'approche, la machine apprend à analyser un texte selon une façon qui lui est propre (sur la base de vecteur et d'opérations mathématiques) et qui n'a pas été dictée dans une suite de règles rédigées par un programmeur humain. La machine se construit ses propres représentations, et il est tentant d'y voir la matérialisation d'une sorte d'inter-langue, par exemple dans les couches cachées de Google Translate (*Google Neural Machine Translation*). On pourrait même étudier cette « inter-langue » en tant qu'œuvre de la machine, intéressante par son originalité voire son universalité. En fait, il ne s'agit que d'une représentation interne, comme il y en a dans tous les réseaux neuronaux profonds. L'intérêt ici, c'est que Google Translate a été conçu pour être multilingue et pas seulement dépendant de couples de deux langues. On peut interpréter cette avancée comme une capacité à transférer des connaissances d'un couple de langues à un autre, mais pas comme l'émergence – pour ainsi dire surnaturelle – d'une inter-langue inventée par la machine. Bref, si la réalité rattrape la fiction, elle prend aussi ses propres chemins...

- (1) Pour pouvoir parler de « langue », il faut que le système de signes soit propre à une communauté d'individus qui l'utilisent pour communiquer entre eux (dimension socio-politique). Sinon ce n'est qu'une sorte de code.
- (2) Il existe un célèbre exemple, inventé par les chercheurs en traduction automatique mais proche d'exemples réels : à un système prévu pour traduire mot à mot, on donne la phrase « *The spirit is strong, but the flesh is weak* » (« *L'esprit est fort, mais la chair est faible* »). Après une première traduction en russe, la phrase retraduite par le système dans l'autre sens donnerait : « *The vodka is good, but the meat is rotten* », c'est-à-dire « *La vodka est bonne, mais la viande est avariée* »... Dans *Le Guérisseur de cathédrales*, Philip K. Dick s'est amusé avec des exemples similaires.

Langage machine

ou code machine : suite de bits (informations binaires composées exclusivement de 0 et de 1) interprétée par le processeur d'un ordinateur exécutant un

programme informatique. C'est le langage natif d'un processeur, c'est-à-dire le seul qu'il puisse traiter. Chaque processeur possède son propre langage machine.

Langue-pivot

Langue « intermédiaire » dans laquelle est traduit un texte avant de le traduire dans la langue voulue. Elle permet de s'affranchir d'un travail de « réglage » de tous les couples de langues possibles, par exemple du polonais vers le swahili, mais de se concentrer plutôt sur les liens entre chaque langue et la langue-pivot, ce qui requiert moins d'efforts.

« mechanese »

ou « idiome mécanique » : langue artificielle de fiction imposée à un humain dans le livre *L'Enfant des étoiles* (Frederik Pohl et Jack Williamson, 1965) pour pouvoir communiquer avec une machine toute puissante.

Apprentissage profond

Technique d'apprentissage machine, en intelligence artificielle (IA), utilisant des réseaux de neurones (fonctions mathématiques). Ceux-ci sont capables d'extraire/analyser/classer des caractéristiques abstraites des données qui leur sont présentées, sans production explicite de règle. On ignore pourquoi le système parvient au résultat, c'est une « boîte noire ». Alors que la démarche symbolique, autre grande approche utilisée en IA, utilise des règles logiques rédigées par l'humain programmeur, et donc parfaitement connues (« boîte transparente »).

Réseaux de neurones

Ces neurones sont des fonctions mathématiques qui, à partir des valeurs numériques reçues en entrée (connexions entrantes), calculent une valeur numérique de sortie (connexion sortante). Ils sont des modèles simplifiés des neurones biologiques du cerveau humain. Connectés en réseau, ils permettent de simuler numériquement le fonctionnement d'unités de calcul reliées entre elles par des connexions modifiables.

Bibliographie

Études scientifiques

Landragin Frédéric, *Comment parle un robot ? Les machines à langage dans la science-fiction*, Le Béliat', 2020.

Poibeau Thierry, *Babel 2.0. Où va la traduction automatique ?*, Odile Jacob, 2019.

Romans et nouvelles de SF

Adams Douglas, *Guide du Routard galactique (The Hitch Hiker's Guide to the Galaxy, 1979)*, traduit de l'anglais (Royaume-Uni) par Bonnefoy Jean, Paris : Denoël, 1982, coll. « Présence du futur ».

Dick Philip K., *Manque de pot ! / Le Guérisseur de cathédrales (Galactic pot-healer, 1969)*, traduit de l'anglais (États-Unis) par Lourbet François André, Paris : Champ Libre, 1977.

Leinster Murray, « *Premier Contact* » (« *First Contact* », Astounding Science Fiction, 1945), traduit de l'anglais (États-Unis) par Billon Pierre, in *Fiction spécial n°8 : L'âge d'or de la science-fiction*, Paris : OPTA, 1965.

Lem Stanislas, *Les Voyages électriques d'Ijon Tichy (Dzienniki gwiazdowe, 1976)*, traduit du polonais par Sila Dominique, Paris : Denoël, 1980, coll. « Présence du futur ».

Pohl Frederik et Williamson Jack, *L'Enfant des étoiles (Starchild, in If, 1965)*, traduit de l'anglais (États-Unis) par Billon Pierre, in *Galaxie (2^{ème} série) n° 23-24-25*, Paris : OPTA, 1966.

Films et séries de SF

Lucas G., *La Guerre des étoiles (Star Wars: Episode IV – A New Hope)*, États-Unis : Lucasfilm, 1977.

Marquand Richard, *Le Retour du Jedi (Star Wars: Episode VI – Return of the Jedi)*, États-Unis : Lucasfilm, 1983.

O'Bannon Rockne S., *Farscape (Farscape)*, Australie : Hallmark Entertainment, 1999.

Wilcox Fred M., *Planète interdite (Forbidden Planet)*, États-Unis : Metro-Goldwyn-Mayer, 1956.