



## IDEES & DEBATS

**INFORMATIQUE** // Les logiciels de traduction ont fait d'énormes progrès depuis 10 ans... et ce n'est qu'un début. Mais seront-ils un jour capables de remplacer les interprètes ?

# L'ordinateur toujours plus doué en langues

Frank Niedercorn  
fniedercorn@lesechos.fr

L'ordinateur va-t-il faire disparaître le métier d'interprète ? On peut se poser la question au vu des progrès accomplis depuis dix ans par les logiciels de traduction automatique. Plus efficaces, plus rapides, ils deviennent même mobiles – lors du lancement du smartphone Galaxy S4, Samsung avait fait une démonstration de traduction de l'anglais vers le chinois. De quoi se demander s'il est encore utile d'apprendre une langue étrangère, et si le débat sur l'enseignement en anglais dans l'enseignement supérieur ne sera pas bientôt dépassé.

Professeur à l'université de Karlsruhe, Alex Waibel a fait l'an dernier au Parlement européen une présentation intitulée « Traduction simultanée automatique : une utopie ? ». Le chercheur allemand y expliquait comment, dans son université, certaines conférences étaient désormais traduites automatiquement et accessibles en temps réel depuis un ordinateur ou une tablette. « Nous avons un problème de fuite des cerveaux et nous voulons que les étudiants étrangers viennent dans nos universités malgré la barrière linguistique. Certes, notre système n'est pas parfait, mais il fournit une base de compréhension », expliquait Alex Waibel.

Une importante communauté scientifique s'intéresse au domaine, notamment en Europe, où coexistent 24 langues avec des coûts liés à l'interprétariat et à la traduction estimés à 1,3 milliard d'euros. Plusieurs projets ont été lancés : Casmacat ou Matecat dans la traduction professionnelle, EU-Bridge pour l'audiovisuel, ou Accept, destiné aux forums sur Internet. Une partie des projets du programme Quaero étaient aussi consacrés à la traduction.

### Statistiques ou linguistique ?

Le petit monde de la recherche en traduc-

tion automatique a vécu une profonde évolution il y a une dizaine d'années avec la prise du pouvoir par les statistiques. L'approche par règles à coups de dictionnaires, de grammaires, de lexiques et de transfert syntaxique, a pris alors un coup de vieux. A tel point que même les pionniers du secteur, comme l'éditeur Systran, ont intégré l'utilisation des statistiques dans leurs logiciels. Le principe : piocher dans des textes déjà traduits pour traduire les nouveaux. L'idée avait déjà été mise en œuvre chez IBM dans les années 1990, mais avec des résultats décevants. A l'époque, les ordinateurs ne savaient travailler que mot à mot et à partir d'un stock limité à 1 million de mots. Dans les années 2000, les chercheurs se sont remis au travail en utilisant des phrases entières, et à partir de volumes bien plus importants.

Si l'approche statistique a permis un bond en avant, elle touche désormais ses limites et commence même à tourner en rond.

Google, qui a embauché l'un des scientifiques inventeurs de cette méthode, a depuis 2007 industrialisé la méthode pour mettre au point son système de traduction. Celui-ci utilise des milliers d'ordinateurs pour aspirer le contenu de l'Internet et constituer des bases de données gigantesques. Le résultat, c'est notamment l'application Google Translate. Utile et efficace, même si sa fiabilité reste encore limitée. Acceptables pour acquérir de l'information, les logiciels de traduction sont impossibles à utiliser tels quels dans un cadre professionnel.

Si l'approche statistique a permis un bond en avant, elle touche désormais ses limites,





et commence même à tourner en rond. Ainsi certains chercheurs travaillent sur des systèmes destinés à prédire la qualité d'une traduction automatique, voire à en corriger automatiquement les erreurs. La plupart des scientifiques estiment qu'ils vont devoir remettre l'ouvrage sur le métier et revenir aux règles linguistiques. Même Philip Koehn, l'un des gourous de l'approche statistique, en convient : « *L'espoir est d'arriver à bâtir des machines de traduction statistique intégrant davantage la linguistique.* »

Il le faut, car les systèmes de traduction vont se multiplier. « *Ne serait-ce que parce que les besoins augmentent de 30 % par an et qu'on ne peut former autant de gens* », note Khaled Choukri, le secrétaire général de l'European Language Resources Association (Elra). « *Les militaires sont très intéressés, notamment dans le domaine des langues rares. Le problème est qu'il existe peu de données et que les statistiques n'apportent pas grand-chose* », explique Jean Senellart, directeur scientifique de Systran. En revanche, les applications pour smartphones, comme celle de Samsung, vont se multiplier pour les langues les plus courantes, avec lesquelles beaucoup de données sont disponibles.

## La traduction de la parole spontanée reste inaccessible

Il est en revanche trop tôt pour rêver à des machines aux capacités équivalentes du robot C-3PO de « Star Wars », capable de comprendre les 6 millions de langages de la galaxie. « *La traduction de la parole spontanée, tout le monde en rêve, mais elle reste aujourd'hui inaccessible. D'une part, en raison des difficultés liées à la reconnaissance vocale, mais surtout par manque de prédictibilité* », résume François Yvon, chercheur au Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (Limsi).

Dépanner un touriste avec des phrases basiques tournant inévitablement autour des transports ou de l'hébergement ne pose guère de problème. En revanche, entrer dans le domaine d'une véritable conversation reste compliqué. Il y a quelques années, le Groupe d'étude en traduction automatique des langues et de la parole (Getalp) du Laboratoire d'informatique de Grenoble avait ainsi travaillé à un système destiné à renseigner les touristes de la région de

Trente, en Italie, offrant un patrimoine important avec beaucoup de châteaux et un riche contenu. « *Le système a vite explosé à cause du nombre trop élevé de domaines à couvrir : les arts, les sciences, l'architecture en passant par la littérature ou encore l'histoire* », explique Hervé Blanchon, chercheur au Getalp. Pour lui, l'avenir est aux systèmes spécialisés. A l'image du logiciel canadien qui traduit les bulletins météo de l'anglais vers le français et inversement, avec une fiabilité supérieure à 90 %. Tout comme son collègue japonais, Altflash, qui convertit les brèves boursières de la Bourse de Tokyo en anglais. France Télécom prépare ainsi le lancement d'un logiciel capable de traduire des SMS dans au moins 14 langues.

Le Getalp travaille lui sur des systèmes impliquant l'utilisateur afin de lui permettre de lever les ambiguïtés, et ainsi d'améliorer la qualité de la traduction. Un tel logiciel pourrait être intégré à un traitement de texte et permettrait de rédiger une lettre dans une langue étrangère avec une qualité acceptable. On peut aussi imaginer des systèmes de traduction simultanée qui afficheront le texte dans les lunettes. En revanche, il faudra toujours apprendre les langues étrangères... ■

## Des chiffres et des langues

### 7.100

Le nombre de langues vivantes dans le monde selon l'encyclopédie Ethnologue.com (SIL International).

### 2.000

Le nombre de langues qui sont parlées par moins de 1.000 personnes.

### 70

Le nombre de langues proposées par le site Google Translate.

### 23

Le nombre de langues officielles de l'Union européenne.





## 1.500

Le nombre de traducteurs employés par la Commission européenne.

## 1,9 million

Le nombre de pages traduites à la Commission européenne en 2010 (contre 1,1 million de pages en 1997)

## Quand Google Translate s'emmêle

Qui ne s'est jamais amusé à prendre le logiciel de traduction de Google en défaut ? Le système repose pourtant sur une technologie impressionnante mettant en œuvre des corpus bilingues constitués de millions de morceaux de phrases mis face à face. D'un côté, la version dans la langue source. De l'autre, celle dans la langue cible. A partir de ces tables de traduction, le logiciel établit un modèle de langage basé sur le nombre d'occurrences de mots. Toute la puissance du système vient de la taille de ce « corpus parallèle bilingue ». « *Etant donné le nombre de documents traduits en anglais, c'est vers cette langue qu'il est fort. Il est beaucoup moins performant vers l'allemand car on trouve moins de documents sur le Web et que la grammaire est beaucoup plus complexe* », explique Jean Sennelart, directeur scientifique de Systran. Lorsqu'un utilisateur lance une requête, Google Translate balaie ses bases de données à la recherche de la traduction la plus probable. Sans toutefois comprendre ce qu'il traduit – il s'emmêle les circuits entre le « canapé » sur lequel on s'assoit et celui qu'on mange. Le logiciel a aussi la mémoire très courte, ne tenant compte ni du contexte ni même de la phrase précédente : dans une recette de cuisine, il traduira alternativement le mot avocat par « avocado » ou « lawyer ».



**Casques pour la traduction au Parlement européen. Si les logiciels restent acceptables pour acquérir de l'information, l'humain reste indispensable à la traduction dans un cadre professionnel.**

*Photo Frederic Maigrot/RÉA*