

La traduction planétaire avec Web to TM

La mondialisation est à l'origine d'une explosion des volumes de traduction. Même les PME de quelques salariés arrivent via Internet sur le marché mondial. D'où la nécessité de traduire toujours davantage. Compte tenu des résultats peu concluants de la traduction automatique, on devrait assister à une explosion parallèle de la population des traducteurs. Or il n'en est rien. Un petit nombre de traducteurs suffit pour répondre à des besoins croissants dans ce domaine. Comment cela est-il possible ?

Le recyclage avec la logique floue

L'invention de la mémoire de traduction (Translation Memory ou TM) a révolutionné la traduction traditionnelle, restée figée depuis Babylone, et qui semblait définitivement associée à l'utilisation de gros dictionnaires. Les mémoires de traduction tirent parti du caractère souvent répétitif des comportements humains. Les entreprises et les organisations peuvent en effet accomplir pendant des années des processus qui, linguistiquement parlant, se ressemblent comme deux gouttes d'eau. Ce « wording d'entreprise » – toujours le même bavardage sur un plan statistique – est capturé de façon optimale avec les TM. Ces petits miracles de la technique font appel à la logique floue pour identifier des modèles linguistiques variables. Si un chef d'entreprise prononce tous les ans un discours similaire devant l'arbre de Noël, sa société a la possibilité de réduire facilement ses frais de traduction d'une année sur l'autre.

Les premières mémoires de traduction commerciales sont originaires de Stuttgart en Allemagne. En 1984, les deux génies en informatique Jochen Hummel et Iko Knyphausen créent une petite agence de traduction et développent dès la fin des années quatre-vingts les premiers programmes d'aide à la traduction, également appelés outils de TAO (traduction assistée par ordinateur). Le principe est très simple. Des éléments de texte traduits sont stockés avec l'original dans une base de données, fournissant ainsi un corpus parallèle. Si Trados – c'est le nom de ce logiciel mondialement connu aujourd'hui – reconnaît un segment de texte ou une partie de segment, la traduction correspondante est alors extraite de la mémoire et récupérée (recyclée). Dans le cas idéal, le traducteur n'a plus qu'à valider le fragment de texte proposé, ce qui lui permet d'avancer plus vite dans son travail. Les outils de traduction assistée par ordinateur autorisent d'importantes économies avec les textes répétitifs. Par exemple, une fois que le constructeur Citroën a rédigé ses guides d'utilisation pour les modèles C2, C3, C4 et C5, le coût du manuel de la C8 est considérablement réduit.

Mais l'avantage subjuguant des mémoires de traduction, de nature invisible, réside dans l'unification du langage. Elle standardise la documentation, la

vente, le support et le site Internet. Tout devient simple et homogène. La première unification globale à grande échelle remonte à 1997, lorsque Microsoft entreprit de traduire son système d'exploitation aussi simplement que possible dans les langues les plus importantes au monde. Bill Gates fit alors l'acquisition d'une partie de Trados, introduisit le système et fit de Windows le logiciel le plus vendu de tous les temps. Dell et d'autres clients importants lui emboîtèrent le pas. En 2005, la société fut rachetée par SDL, un concurrent britannique, qui reprit la marque Trados pour ses propres produits logiciels.

De la solution isolée à la TM sur le cloud

La plupart des traducteurs travaillent en freelance chez eux à la maison. Dernièrement encore, les liaisons de données à haut débit étaient plutôt l'exception. Chaque traducteur créait par nécessité sa propre base de données, ce qui donnait naissance à une infinité de solutions isolées. Considérant sa mémoire comme sa propriété privée, le traducteur était peu enclin à collaborer. Pour contrer cette tendance, les grosses agences eurent recours à la collectivisation forcée et créèrent un nouveau métier: le gestionnaire de traduction. Tel un percepteur, celui-ci exigeait l'envoi des fichiers bilingues lorsque le travail était terminé. Depuis lors, traducteurs et réviseurs reçoivent leurs commandes sous la forme de packages de projet incluant les ressources de traduction, et livrent le fruit de leur travail dans des packages de retour, dont le contenu est ensuite transféré dans la mémoire principale du gestionnaire de traduction.

Les agences astucieuses ont pu constituer ainsi des mémoires principales colossales et développer leur part de marché. Mais leur zèle a été assombri par une faiblesse inhérente au système. Plus la mémoire de traduction principale est volumineuse, plus la qualité des packages diminue. La mémoire de concordance et les dictionnaires devenues gigantesques ne pouvaient plus être inclus dans le package, ce qui était un inconvénient notable pour la réception par courriel par rapport à l'exploitation d'une base de données en ligne.

En 2015, la société SDL, qui avait racheté le logiciel Trados, commercialise le premier serveur

GroupShare viable. Les segments de texte sont désormais stockés sur un serveur central, permettant ainsi à un nombre illimité de traducteurs d'accéder aux ressources de traduction. Grâce à la technologie de serveur, plus aucune limite n'entrave le recyclage des segments de traduction. Un petit pas seulement nous sépare encore de la traduction planétaire, une mémoire de traduction sur le cloud, qui engloberait tous les segments traduits sur le Web.

La révolution numérique, un véritable chasseur de coûts

TTN Translation Network, le premier réseau de traduction entièrement automatique, a été conçu à Genève en 1987, avant l'ère du Web. Son créateur Martin Bächtold avait eu l'occasion d'expérimenter les premiers réseaux interuniversitaires à l'université de Stanford, dans la Silicon Valley. Les cours sur le modèle de l'avantage comparatif avaient été pour lui une révélation : de toute évidence, la traduction et la communication seraient désormais indissociables. Les traductions doivent être réalisées sur le site le plus attractif sur un plan économique et qualitatif c'est-à-dire dans une région où la langue cible est pratiquée activement.

Avant de rentrer à Genève, Martin Bächtold a glissé l'un des premiers modems dans son bagage. Cette « boîte métallique » aux sifflements bruyants, qui est alors interdite en Suisse, lui permet d'installer le premier serveur de traduction au monde sur un PC Schnyder avec 10 mégas de disque dur. Mais cette innovation est beaucoup trop précoce pour le marché. Personne ne sait alors comment fonctionne un modem. La société doit donc emprunter de l'argent pour acheter des appareils bon marché à Taïwan et les expédier gratuitement aux clients et aux traducteurs. L'institut pour l'étude de la neige et des avalanches SLF de Davos, un service spécialisé dans l'alerte en cas de risques d'avalanches, est l'un des premiers clients. Les bulletins d'avalanches demandent à être traduits très vite et les textes doivent être transmis par voie numérique, et non par télécopie. Les traducteurs sont prévenus de l'arrivée d'un bulletin d'avalanche par une sonnerie sur le télécopieur, un système ancestral remplacé depuis bien longtemps par le SMS ou l'interface du smartphone.

En 1989, le Worldwide Web fait son apparition au CERN de Genève et révolutionne dans son sillage les technologies de communication avec un nouveau standard. TTN se lance sur Internet avec le 16 comme numéro client auprès des services de la Poste. Les gains générés avec le premier système sont utilisés pour financer le développement d'un réseau de type ARPA dédié à la traduction, en Inde,

où une très grosse équipe d'informaticiens travaille sur la programmation du code. A partir d'un réseau répliqué, il s'agit de créer un système sur le cloud pour le routage automatique de 165 langues. L'opération se solde par un échec. Le code est trop long et les problèmes bien plus complexes qu'on ne l'avait pensé.

La seconde tentative est plus fructueuse, même si son aboutissement demande beaucoup plus de temps que prévu. Progressivement, des pans de processus toujours plus importants sont automatisés, et les coûts de production peuvent être réduits de 30%. Il apparaît que les agences utilisant l'intelligence artificielle gèrent plus efficacement les gros portefeuilles de clients que celles qui les confient à des collaborateurs. Les programmes calculent la charge de travail des traducteurs en tenant compte des horaires et en intégrant les congés. Grâce à une gestion optimisée du temps, les traducteurs bénéficient d'un flux de travail plus constant, d'où moins de stress et une productivité accrue.



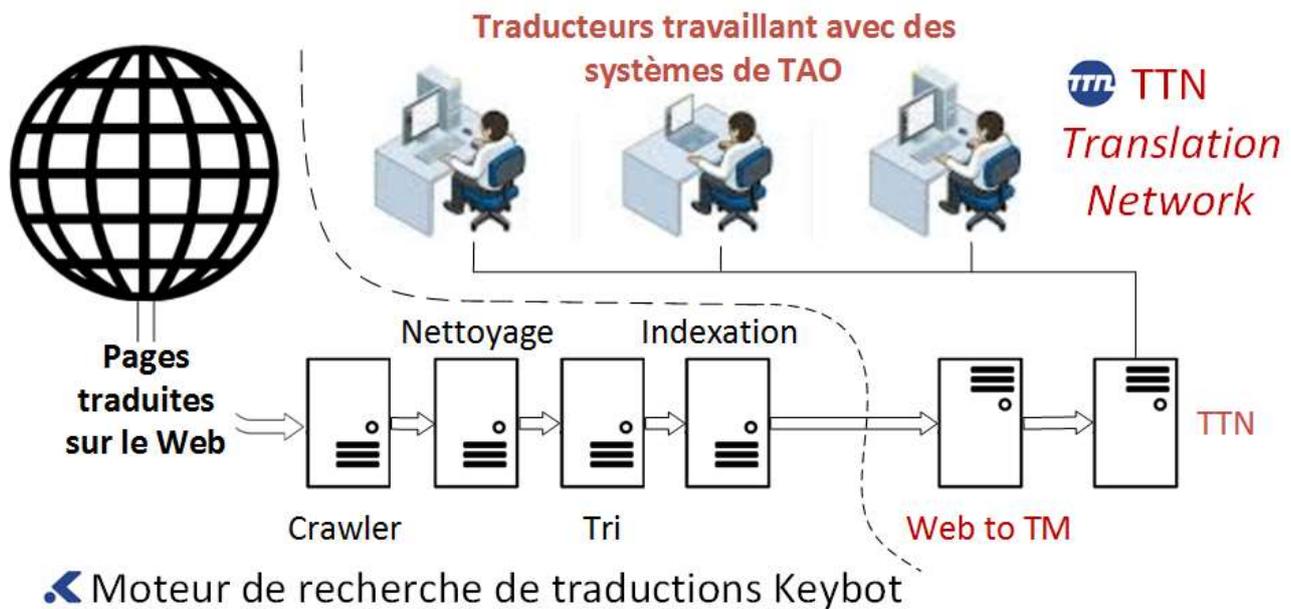
La nécessité d'une puissance de calcul élevée

Patrick Boulmier, spécialiste du Big Data de la société Infologo, paramètre avec Martin Bächtold, CEO de Keybot, les nouveaux ordinateurs ultra-puissants pour la machine linguistique planétaire. Plusieurs centaines de pages Web doivent être converties à la minute en mémoires de traductions.

Keybot : Web to TM

Dans de nombreux groupes internationaux, la numérisation est souvent chaotique dans le secteur de la traduction. Ces organisations mettent en œuvre des applications Web contenant des milliers de pages traduites mais ne disposent pas de mémoires de traduction pour stocker proprement ces textes dans un corpus parallèle. La négligence qui préside au choix des prestataires de traduction a souvent des conséquences désastreuses. Lorsqu'elles souhaitent réviser leurs sites Internet, les sociétés mal organisées paient le prix fort pour chaque page, puisque le travail déjà effectué ne peut être récupéré et recyclé. Le remplacement de ces connaissances malencontreusement perdues a un coût.

Web to TM : le Web est aspiré et transformé en gigantesque mémoire de traduction (TM)



◀ Moteur de recherche de traductions Keybot

Web to TM a été conçu pour venir en aide à ces entreprises. La société Keybot, une filiale de TTN, a développé un moteur de recherche de traductions qui porte son nom. Celui-ci passe le Web au crible, tout comme Google. Il ne mémorise que les pages traduites en plusieurs langues et les indexe ensuite dans un corpus parallèle. Un réseau de serveurs complexe effectue l'exploration de données et analyse les pages Internet de clients potentiels, à la recherche des éléments de texte traduits. Les informations ainsi aspirées, ou Big Data, doivent être ensuite nettoyées et triées, puis soumises à une analyse statistique. De même, les répétitions doivent être répertoriées, leur pertinence calculée, puis sauvegardées. Une fois ce lourd processus achevé, la machine peut envoyer les contenus, segment par segment, à une batterie de serveurs GroupShare. Lorsqu'un traducteur ouvre ensuite un projet avec un logiciel de TAO, tous les fragments de texte retrouvés sur l'application Web du client par le moteur de recherche, sont traduits automatiquement. Le traducteur dispose toujours de la version publiée la plus récente, et non d'une version obsolète qui a été remaniée en interne entretemps.

Pour pouvoir classifier les éléments linguistiques, il est nécessaire à Keybot d'avoir une première base d'apprentissage, c'est pourquoi il intègre toutes les pages Wikipedia comme les traductions de textes bibliques et des droits de l'homme en 165 langues. Chaque langue possède son propre gencode, qui peut être extrait sous la forme de n-grammes. Keybot entend exploiter ces propriétés statistiques pour identifier et mettre en parallèle les segments de texte. Le système se trouve encore en phase bêta. Jusqu'à présent, il n'a été possible de générer des

mémoires de traduction fiables que lorsque le site Web du client était organisé de façon à éviter tout faux-pas au crawler pendant la phase d'aspiration des données. La mémoire de traduction la plus importante à ce jour compte 23 langues. Elle a été produite pour une société américaine.

Keybot a pour but de transformer le Web plurilingue en une gigantesque mémoire de traduction : « Web to TM ». La transformation étant extrêmement exigeante en termes de puissance de calcul, elle nécessite par conséquent une batterie de serveurs suffisamment puissante. Pour réunir les fonds nécessaires, Keybot prépare en Allemagne son introduction à la bourse des PME et tentera de financer une partie du parc de machines avec le crowdfunding.

SLOTT Translation

Dans le domaine de la traduction automatique, les innovations décisives sont venues du secteur de la météorologie. Les bulletins d'alertes donnent lieu à un dilemme, avec d'un côté l'obligation d'une diffusion rapide et, de l'autre, l'impérieuse nécessité d'exclure toute erreur de traduction. L'approche statistique de Google Translate n'est ici d'aucun secours, en raison de son manque d'exactitude d'une part, mais aussi parce qu'elle ne permettra jamais de restituer la précision horlogère des bulletins d'alerte.

Jörg Kachelmann, un météorologue astucieux, qui a étudié les mathématiques à Zurich, a été le premier à résoudre ce dilemme. A partir d'une simple feuille de calcul Excel, il a « bricolé » un système destiné à gérer la production d'énoncés linguistiques par l'intermédiaire de cellules. Dès les années quatre-

vingts, le directeur du SLF avait tenté de bâtir un système de traduction automatique. Le système statistique basé sur le principe de probabilité et les modèles de Markov, qui avait été imaginé quelque temps après par une université allemande, fut également un échec. Plusieurs années plus tard, lorsque Kurt Winkler, de l'Institut pour l'étude de la neige et des avalanches SLF à Davos, adresse une feuille de calcul Excel à première vue totalement farfelue à la métropole linguistique de Genève, il est tout d'abord la risée des linguistes qui le prennent pour un fou. Son projet est déposé tout au fond d'un tiroir, tel un mauvais roman policier. Il lui faut insister pour qu'un collaborateur de TTN, connaisseur en mémoires de traduction, soit mandaté pour infirmer la thèse de l'absurde. Une phrase incorrecte, et bye bye le système Winkler, qui serait mort-né.

Trois jours plus tard, c'est toujours silence radio. Aucune erreur n'a été identifiée. Même un logiciel spécialement conçu par TTN à cet effet ne peut démontrer l'existence de la moindre erreur. Étonnant ! Kurt Winkler, qui n'y connaissait rien à la linguistique, avait analysé les bulletins d'avalanches et leurs traductions sur les dix dernières années pour déceler les possibilités de mutation, et consigné ce travail dans une base de données Excel que personne ne comprenait.

Personne ? Peut-être pas. Quelque cent ans plus tôt en effet, Ferdinand de Saussure, originaire de Genève et père fondateur du structuralisme, avait déjà attiré l'attention dans ses cours sur la dimension syntagmatique du langage. Il avait été le premier à définir les possibilités de mutation pouvant se présenter dans une structure linguistique, sans toutefois établir un lien avec d'autres langues. Kurt Winkler avait découpé les énoncés selon les mêmes principes et dégagé des règles de transformation pouvant être appliquées pour transposer des éléments de texte dans une autre langue.

A partir d'une liste de phrases répertoriées par Kurt Winkler, des millions de phrases idiomatiques et grammaticalement parfaites peuvent être générées en quatre langues. Le système ne fonctionne, il est vrai, que pour les bulletins d'alerte avalanches en Suisse, et les phrases doivent être produites à partir d'une liste à l'écran. Cela n'est guère pratique et l'usage est très limité.

TTN expérimente un système analogue, appelé SLOTT Translation. Tout comme les bulletins météorologiques, les traductions ne doivent comporter aucune erreur car cela saperait la confiance des clients. A partir d'un catalogue de 20 phrases types dans un premier temps, il s'agit de standardiser la communication avec les clients, de façon à pouvoir

répondre correctement et sans erreur à des requêtes dans toutes les langues.



Des traductions automatiques pour plus de sécurité

Kurt Winkler, travaille à l'Institut pour l'étude de la neige et des avalanches SLF. Il est à l'origine d'une avancée étonnante dans le secteur de la traduction automatique. Les alertes avalanches sont traduites en quelques fractions de secondes.

Nul ne sait encore si le système SLOTT parviendra à s'imposer sur un plan commercial. Par contre, il ne fait aucun doute que les mémoires de traduction de demain présenteront une organisation hiérarchique des phrases, dans la mesure où leur potentiel s'en trouvera ainsi considérablement augmenté. Les systèmes de TAO de la prochaine génération seront capables de traduire avec précision non seulement des textes stockés dans une TM, mais aussi des millions de variantes.

Le traducteur est-il appelé à disparaître ?

Les mémoires de traduction ne sont que des machines de recyclage. Elles n'ont aucune aptitude intellectuelle ni créative. De par leur architecture, les processeurs actuels ne peuvent en aucun cas rivaliser avec les synapses du cerveau humain lorsqu'il s'agit de produire des énoncés porteurs de sens. Le traducteur reste indispensable. Seul l'ordinateur quantique pourrait y changer quelque chose.

Utilisation libre pour publication (2639 mots)

Martin Bächtold, Keybot Sarl, Genève, mai 2017