



exit////

exit////

Quelles solutions pour la traduction automatique ?

Vos billets de blog mériteraient une plus large audience... Vous avez une notice d'emploi à internationaliser séance tenante... Ou une présentation d'entreprise à créer de manière urgente en Globish... Bref, comment traduire ou faire traduire un texte de manière ponctuelle grâce aux outils informatiques, services web ou logiciels spécialisés ? C'est ce que nous avons cherché à savoir ici.

Au milieu des années 90, Microsoft s'apprêtait à sortir Windows 95. Pour l'éditeur, chaque version de Windows fait l'objet d'une relocalisation plus que d'une traduction, car absolument tout doit être revu, des menus contextuels aux .dll du logiciel, le tout compilé et fonctionnel. Le fait est qu'un dimanche, veille de l'envoi des copies de Windows 95 pour duplication, un employé de l'agence de traduction réveillait son patron en pleine nuit... Problème : personne n'avait pensé à vérifier que l'OS fonctionnait après la compilation de tous les fichiers traduits en français. Le fait étant que Windows 95 ne fonctionnait pas, la faute à une dll de

trop sur le CD master ! Une petite erreur qui aurait pu coûter cher puisque Microsoft s'apprêtait à tout graver... On imagine facilement l'ampleur du problème, avec plusieurs millions de copies non-fonctionnelles. C'est dire l'importance que revêt aujourd'hui la traduction et pas seulement pour les grands éditeurs.

Les différentes solutions

Pour les particuliers comme pour les entreprises, il existe plusieurs solutions de traduction automatisée.

- Les plus simples et plus connues sont les outils disponibles sur le Web : l'incontournable Google Traduction, mais aussi Reverso ou Systranet. Ces derniers sont bien les plus simples car finalement



Guillaume de Brébisson, fondateur de l'agence AnyWord

les plus accessibles, mais ne constituent pas en soi de véritables solutions de traduction.

- Les logiciels de traduction automatique (Systran, Reverso, Promt, etc.) sont des outils qui analysent un texte et lui applique le dictionnaire, auquel est ajoutée une « intelligence ». On les appelle la plupart du temps des « moteurs » de traduction.
- Les logiciels à mémoire de traduction (SDL Trados, Word Fast, Déjà Vu, sont les trois principaux). Leur principe est différent puisqu'ils enregistrent chacun des phrases/mots que l'utilisateur tape pour s'améliorer continuellement, en termes de cohérence et de rapidité également.

Ceci vaut pour les solutions logicielles, que certains traducteurs professionnels utilisent, qui viennent épauler des solutions « humaines », à savoir les agences de traduction ou les freelances, spécialisés ou non.

Chacune des solutions citées ci-dessus fonctionne

Les agences de traduction

La traduction est une des rares activités qui ne connaît pas la crise. Avec un marché mondial estimé à 33 milliards de dollars en 2012, dont 16,5 milliards en Europe, il affiche une croissance annuelle d'une année sur l'autre de 12,17 % (source : Common Sense Advisory). 26104 agences de traduction sont aujourd'hui répertoriées dans le monde, dont environ 300 en France. La plupart d'entre elles sont d'ailleurs spécialisées dans un ou plusieurs domaines particuliers : finances-juridique (Translations, HL Trad, CPW), informatique (Datawords, Techword, Praetorius), marketing, médical (La Traduction Médicale)... D'autres agences sont quant à elles généralistes, capables de répondre à tout type de besoins (AnyWord) en faisant par exemple appel à des traducteurs extérieurs et spécialisés.



TRADUCTION AUTOMATISÉE exit////

exit////

donc différemment. Pour Google Traduction par exemple, l'outil « utilise la base de données créée au fur et à mesure par les utilisateurs, et cherche les termes correspondants », nous explique Guillaume de Brébisson, fondateur de l'agence AnyWord (www.anyword.fr). « Google propose de la traduction statistique et ne se préoccupe pas de la structure de la phrase, mais la compare aux autres entrées de sa base de données. » Un outil logiciel de bureau comme Systran va quant à lui utiliser un moteur de traduction qui bénéficie d'un entraînement. Ce qui lui permet de mixer traduction statistique et éléments linguistiques.

Quels choix ?

Le problème de ces outils est qu'aucun d'entre eux ne fournit une prestation « prête à l'emploi », notamment pour les textes longs ou comprenant des termes spécialisés. Pour Guillaume de Brébisson, la seule véritable solution lorsqu'on souhaite utiliser un logiciel est de l'accompagner par la vérification et la correction d'une personne bilingue dans la langue ciblée. « La traduction automatique vous fournit un brouillon, pas une copie parfaite », estime-t-il. C'est, semble-t-il, la solution la plus viable pour une PME qui a des besoins ponctuels en traduction.

Guillaume de Brébisson est également l'auteur et le créateur de « L'observatoire de la traduction » (www.anyword.fr/blog/) et constate que « depuis la démocratisation des logiciels de traduction sur le Web, le degré qualitatif baisse constamment. » Une remarque qui a son importance puisque « plus on est technique, plus les conséquences sont graves et plus les volumes sont élevés ».

Plusieurs solutions techniques

Les logiciels de traduction se découpent en deux familles distinctes : les « moteurs » et les « mémoires de traduction ». La catégorie des « moteurs » se scinde elle-même en trois groupes :

- les moteurs traditionnels sont des logiciels comme Systran version Desktop, Power Translator mais aussi les outils en ligne comme Systranet, Reverso ou Babel Fish – pour ces trois derniers, sans personnalisation du dictionnaire. Ces logiciels ont une approche syntaxique d'une phrase. C'est-à-dire qu'ils la découpent de manière ordinaire en cherchant sujet, verbe, complément, proposition subordonnée, COD, etc. Le logiciel construit la phrase comme un « arbre » et pioche les mots correspondants dans le dictionnaire selon sa fonction et sa signification. Une fois ceci effectué, il crée une phrase cible en respectant la grammaire de la langue traduite. C'est généralement cette étape qui fait défaut à ce genre d'outils, qui ne proposent pas de personnalisation du dictionnaire. Ainsi, il butera souvent sur des mots spécialisés dans un domaine – informatique, médical, etc. L'autre gros problème de ces outils est qu'ils supposent que la phrase source respecte toutes les règles grammaticales de la langue. Ce qui est en fait très rare dans un texte. Parfois, une simple virgule peut changer entièrement le sens de la phrase par exemple... L'analyse ne sera donc pas la même.

Une nouvelle mode : le crowdsourcing

Facebook a été l'un des premiers grands sites à utiliser le crowdsourcing (le mélange de savoir, en quelque sorte, ndr) pour traduire son site et son interface en plusieurs langues. Il a donc snobé les agences de traduction, ce qui lui a permis d'économiser de l'argent et de gagner en rapidité, mais aussi de mobiliser des personnes en interne et d'avoir quelques surprises... Pour chaque pays, Facebook a découpé son interface en petits groupes de mots puis a demandé à sa communauté de les traduire. Après une première sélection, les internautes ont voté pour les meilleures traductions. Toutefois, toutes les pages de Facebook n'ont pas pu être traduites avec ce procédé. D'autant plus que le réseau social a eu droit à quelques farces, où des internautes traduisaient un message complètement différent, souvent à vocation humoristique...

- les moteurs statistiques, qu'utilise Google Traduction par exemple. Pour son outil, Google avait déterré une vieille technique de traduction étudiée par IBM dans les années 60 et 70. Google s'appuie en fait sur une gigantesque base de données alimentée quotidiennement par les utilisateurs. Quand vous entrez une phrase, il va chercher les nombreuses traductions correspondantes et, par occurrence, vous proposera comme résultat la plus fréquemment utilisée. Ainsi : pour une phrase source, Google analyse un grand nombre de phrases cibles avant de vous livrer le résultat final. Google Traduction fonctionne aussi sur des groupes de mots, et non pas sur une phrase comme les moteurs traditionnels. Ce système a l'avantage de très bien fonctionner pour la synthèse vocale, ce qui est le plus gros problème de la première famille de moteurs. Les moteurs hybrides : plus haut de gamme, ils combinent moteurs traditionnel et statistique. C'est le cas pour la version Serveur de Systran par exemple. Ils commencent par une analyse syntaxique de la phrase et proposent plusieurs solutions. Le moteur statistique prend ensuite le relais et trouve la meilleure phrase au

sein d'une base de données. Pour qu'un tel système fonctionne, il faut toutefois que la base de données soit assez importante... mais une post-traduction effectuée par un humain est également requise. Ce qui n'est donc pas à la portée de toutes les bourses...

La seconde famille de logiciels est constituée par les logiciels à mémoire de traduction. Ils sont utilisés par les professionnels et traducteurs qui traduisent chaque jour. Chaque phrase est entrée dans une base de données. Quand le logiciel reconnaît une phrase déjà traduite par le passé, il la traduit instantanément. Solutions de gain de temps pour les traducteurs, elles donnent lieu à trois situations :

- la phrase à traduire n'est pas dans la base de données, le traducteur fait seul son travail ;
- la phrase est dans la base de données, elle est instantanément traduite ;
- la phrase n'y est pas mais une autre lui ressemble. Si la phrase correspond à 75 %, le traducteur n'a plus que les 25 % restants à traduire, et à enregistrer une nouvelle phrase dans la base de données. ■

Emilien Ercolani

