

Traduction automatique : ça marche ou non ?

Christian BOITET

Laboratoire LIG, GETALP – Université Joseph Fourier,
385 rue de la bibliothèque, BP 53,
38041 Grenoble, Cedex 9, France
Christian.Boitet@imag.fr

Interstices, 2008, v3

Dialogue imaginaire... synthèse de dialogues réels

Traduction automatique : alors, ça marche ou pas ?

Pas de doute, « ça marche » ! Des millions d'internautes demandent chaque jour la traduction de millions de pages à des « serveurs de TA » gratuits, comme ceux de Systran, Reverso, Google, ou au Japon Fujitsu, Toshiba, Nec, Oki...

Oui, bien sûr... mais ce n'est pas ce que je veux dire ! En fait, ça traduit n'importe comment, c'est souvent du charabia, ou encore c'est grammatical, mais les mots sont mal traduits...

Eh bien, oui... mais que peut-on espérer ? Si on doit traduire « they saw many arms », comme savoir si c'est « ils scient de nombreux bras » ou « ils ont vu de nombreuses armes » ou « elles ont vu de nombreux bras », « elles [scient / ont vu] de nombreuses armoiries¹ », etc. ? Bien des traducteurs humains, professionnels, font ainsi des contresens graves. En général, les textes qu'ils produisent sont grammaticaux, mais, au fond, cela contribue à « cacher » les contresens et faux sens ou omissions.

D'autre part, vous parlez ici de serveurs Web gratuits, dont le but est d'aider à comprendre « le mieux possible » des textes très divers dans des langues inconnues. Il ne s'agit pas de fournir des traductions parfaites, ni même de remplacer des traducteurs, en produisant des « premiers jets », utilisables par des réviseurs plus experts pour produire des traductions finales de qualité professionnelle.

Mais il existe aussi des systèmes spécialisés, inconnus du grand public, qui traduisent extrêmement bien, et en fait mieux, et pas seulement beaucoup plus vite, que des traducteurs humains.

Vraiment ? Alors il s'agit de progrès récents, je n'en ai pas entendu parler !

En fait, ça date déjà de 30 ans. Le système TAUM-météo, qui devint METEO en 1985, fut lancé de façon opérationnelle le 24 mai 1977 à Environnement Canada. Il a toujours été dédié au « sous-langage » des bulletins météo. Au début, il ne traduisait que de l'anglais en français. Une version français-anglais fut ensuite développée. Vers 1990, METEO traduisait 30 millions de mots par an (l'équivalent de 120 000 pages « standard »), 20 d'anglais en français et 10 de français

¹ par abréviation de « coat of arms ».

en anglais. Un bulletin envoyé de Winnipeg en anglais y retournait en version française 4 minutes plus tard, le résultat de la TA ayant été révisé en 1 minute ou moins. Le niveau de qualité de la TA peut ici être évalué :

- il faut environ 7 fois moins de temps pour réviser un bulletin traduit automatiquement que s'il est traduit par un traducteur « junior ».
- pour réviser 100 mots traduits et obtenir la qualité professionnelle, il faut en moyenne 3 opérations du traitement de texte, le remplacement d'un mot comptant pour 2 (suppression puis insertion). C'est pourquoi le « mainteneur » de ce système, John Chandio, parle de « 97% de qualité ».

Avant la TA, la traduction des bulletins météo était d'ailleurs le « purgatoire » des traducteurs juniors du Bureau de Traductions du Ministère du Travail Canadien, et c'est l'un d'eux qui avait été supplier le groupe TAUM (TA à l'Université de Montréal) de les délivrer de cette tâche ingrate. Ces traducteurs la « fuyaient » d'ailleurs après les 3 mois obligatoires...

Cet exemple illustre aussi le fait que la traduction professionnelle est non seulement difficile, mais souvent pénible psychologiquement. Contrairement à ce qu'on croit, il est en fait très difficile de bien traduire des bulletins météo ! Pour un traducteur formé, il n'y a bien sûr aucun problème de compréhension, mais il faut produire des traductions « dans le moule » du sous-langage des bulletins météo dans l'autre langue. Et ça prend bien 3 mois à temps plein pour devenir expert de cela. Difficile, donc. Mais aussi démoralisant, car un bulletin n'a qu'une durée de vie de 4 heures. Avant l'automatisation, le temps de validité d'un bulletin traduit n'était en fait que de 2 à 3h en moyenne...

Bref, ce système a économisé 17 traducteurs à plein temps depuis au moins 25 ans.

C'est votre seul exemple d'un système de TA de haute qualité ?

Non, bien sûr. Un exemple beaucoup plus récent est le système ALTFlash, construit par NTT pour la bourse de Tokyo (Nikkei). Il traduit les « brèves » (flash reports) du Nikkei de japonais en anglais, de manière également quasi-parfaite. Ses résultats sont fournis aux lecteurs sans aucune révision humaine, car la qualité « pro » n'est pas aussi cruciale que pour des bulletins météo. C'est un système « à 2 étages ». La brève à traduire est d'abord soumise à un système « à schémas bilingues », par exemple, en transposant au français-anglais :

- \$\$stock a fortement monté à \$\$value juste avant la clôture.
- \$\$stock rose sharply to \$\$value just before closing.

Si une phrase instancie un schéma (japonais), l'instanciation correspondante (éventuellement via le dictionnaire) en anglais donne une traduction parfaite. C'est une approche de type « livre de phrases avec variables ». Si aucun schéma ne correspond, la phrase est traduite par une version spécialisée du très gros système généraliste ALT/JE (400 000 entrées de dictionnaire) construit par NTT entre 1980 et 2000 en tant que système de recherche.

D'autres exemples sont des systèmes de TA destinés aux documentations techniques, et utilisés en interne. Le gain en temps humain est très important : sachant qu'en traduction traditionnelle il faut en moyenne 1 heure pour traduire 1 page de 250 mots, et 20 minutes pour la réviser, on arrive à 10—15 minutes de « post-édition » pour un système de TA spécialisé.

Attendez... pourquoi dites-vous « post-édition » et plus « révision » ?

On parle de révision quand la qualité du premier jet (dit aussi « prétraduction ») est suffisante pour qu'on puisse le corriger en ne regardant que très rarement le texte source, voire jamais si on ne connaît pas la langue source. Ainsi, un ingénieur français peut réviser une bonne prétraduction d'un manuel d'entretien, grâce à sa connaissance détaillée du domaine.

Par contre, on parle de post-édition quand il vaut mieux (et il faut très souvent) commencer par lire et comprendre chaque phrase à traduire, avant de regarder la prétraduction proposée, et de trouver comment la modifier *a minima* pour obtenir le sens désiré.

Exemple réel : comment réviser « voulez-vous un joint torique avec votre café ? » Même en sachant qu'on vient de l'anglais, on risque de ne pas trouver, et de s'énerver...

Mais si on post-édite, on lit d'abord « do you want a donut with your coffee » (écrire « donut » pour « doughnut » est possible aux USA), on rit un peu, on corrige, et on avance.

Mais alors, s'il faut connaître la langue source, ça ne peut servir qu'aux bilingues !

Eh bien... oui et non ! Il y a plusieurs sortes de « traduction automatique », ou plus généralement de « traduction automatisée » (TAO), avec des buts bien différents. par ordre de difficulté croissante pour les développeurs de systèmes :

- aider un vrai bilingue à produire des traductions de haute qualité en lui fournissant des « prétraductions » les plus utiles possibles, ainsi que des aides dictionnairiques. On parle de « TA du traducteur ».
- aider quelqu'un à accéder à une information dans une langue qu'il ne connaît pas. On parle de « TA du veilleur ».
- aider 2 personnes n'ayant pas de langue commune à communiquer (par oral, ou par écrit, en tchat). On parle de « TA de dialogue ».
- aider une personne monolingue ne connaissant pas ou que très peu une langue étrangère, mais très bien le domaine en cause, à produire des traductions de qualité dans sa langue. C'est une tâche envisagée assez récemment, et qui suppose de proposer une traduction mot à mot multiple, et plusieurs prétraductions complètes, factorisées, avec des dispositifs ergonomiques pour « faire apparaître » une bonne traduction, jugée plausible. On pourrait parler de « TA du transcripteur ».
- aider quelqu'un à produire des traductions de haute qualité dans une langue qu'il ne connaît pas, en le « consultant » (dans sa langue) en cas de doute, grâce à un dialogue de désambiguïsation le moins lourd possible : on parle de « TA de l'auteur »

Il n'y a pas d'exemples de TA de qualité pour des textes non techniques ?

Si. Mais la vraie question serait plutôt : peut-on construire des systèmes de TA de qualité ayant une « couverture » assez large, par exemple, capables de traduire des journaux ?

Un exemple assez récent est celui du système développé par Comprendium en Espagne, dans l'environnement METAL initialement construit (1981—84) pour faire de la TA technique par l'université du Texas pour le compte de Siemens. Le système présenté à la conférence EACL-2005 traduit tous les jours plusieurs journaux d'espagnol (castillan) en catalan et en galicien. Le fait que ces langues soient lexicalement et syntaxiquement très proches a permis d'arriver à une qualité telle qu'il ne faut que 5 minutes de post-édition par page standard ! De façon traditionnelle, il faudrait au moins 1h de travail (sans révision), et, en utilisant une « mémoire de traduction », il faudrait au moins 30 minutes...

On dit souvent que les traducteurs ont peur de la TA, qui pourrait les remplacer

C'est vrai que beaucoup sont très réticents. Mais ils n'ont pas de quoi avoir « peur de la TA ».

D'abord, ils ne sont pas en concurrence pour la « TA du veilleur » : aucun traducteur humain ne serait en mesure de produire en une seconde la traduction d'une page Web, et même pas une traduction mot à mot... Et il ne le seront pas plus pour la « TA du transcripteur ».

Ensuite, les traducteurs professionnels indépendants et modernes utilisent en fait les systèmes de TA du commerce « configurables » (par choix des priorités de dictionnaires, insertion d'un « dictionnaire utilisateur », et nombreux autres paramètres) pour produire des prétraductions et les post-éditer. Le site <http://www.geocities.com/MTpostediting> de Jeff Allen donne des références et des études de cas prouvant qu'on arrive ainsi à traduire 6000 mots par jour au lieu de 2500 en moyenne.

L'exemple de TAUM-météo montre d'ailleurs que, parfois, ce sont des traducteurs qui poussent à la construction d'un système de TA. Bien sûr, il faut ensuite les laisser utiliser ou non le système selon qu'ils décident qu'ils vont ou non gagner du temps avec lui. C'est ce qui se passe à la PAHO (Pan American Health Organization, Washington) et à l'OMS (Genève) avec le système PAHOMTS (qui traduit entre anglais, espagnole et portugais), centré sur la politique et la santé, mais assez généraliste du point de vue de la syntaxe et des genres de textes.

Et qu'en est-il de l'aspect technique ? Est-il vrai que la TA a fait beaucoup de progrès depuis 20 ans grâce à des approches « statistiques » ?

Il faut y regarder de plus près, et préciser ce qu'on entend par « approche » de la TA, et par « progrès ».

Ce qui est derrière cette affirmation, c'est qu'on peut maintenant « générer » un système de TA à partir d'un très grand corpus parallèle bilingue, formés de segments source² et de leurs traductions (« segments cible »). Les systèmes de Google sont de ce genre, et ce sont pour l'instant les seuls utilisables en « TA du veilleur ». La firme LanguageWeaver, fondée en 2001, fait sa publicité en disant qu'elle peut construire un système de TA d'un jour à l'autre, mais elle n'en avait construit que 4 en 4 ans, et, près de 8 plus tard, elle n'a toujours pas de portail Web permettant d'essayer un de leurs couples.

Tout le monde peut constater que GoogleTranslate fournit des traductions apparemment « fluides », mais, au moins à 30%, incompréhensibles, ou fausses, et surtout où des parties d'information ont disparu. Au contraire, les évaluations d'utilisabilité des 4 ou 5 dernières promotions d'ingénieurs RICM3 de Polytech montrent que les résultats de Google et de Reverso (appelés un peu rapidement systèmes « à règles ») sont nettement plus « adéquats », en moyenne. D'après le chercheur Dong Zheng Dong, auteur du système « Longue Marche » et du réseau lexico-sémantique HowNet (≈100 000 entrées en chinois, anglais et concepts), ils sont plus « fiables » : si on voit une phrase cible qui a l'air parfaite, c'est le plus souvent une excellente traduction avec Systran ou Reverso, mais c'est assez souvent une phrase qui dit autre chose avec Google, en particulier une phrase qui omet de l'information, ou une négation, ou qui en introduit qui vient de nulle part...

Comment ça, un système de TA peut vraiment introduire une information qui n'est pas dans le segment source ?

Oui, mais cela peut être un bien comme un mal. Ainsi, si un portier d'hôtel voit arriver un bon client, disons M. Fischer, il lui dira en français « Bonjour Monsieur », et pas « Bonjour Monsieur Fischer », et encore moins : « Bonjour Monsieur Pêcheur ». Mais, en allemand, il devrait dire « Guten Tag Herr Fischer », car « Guten Tag, mein Herr » serait plutôt impoli de sa part. Un système de TA, comme un traducteur humain, doit donc parfois utiliser de l'information absente du texte (ou de l'énoncé) mais présente dans le contexte (ou plus généralement dans la situation).

Encore un petit exemple : si le portier demande en japonais à Mme Tanaka si elle a bien dormi, il lui dira « Tanaka san wa yoku nemashita ka » [Tanaka M./Mme thème bien avoir-dormi-poli est-

² « source » et « cible » sont invariables en usage adjectival, comme « marron », « pêche » ou « standard ».

ce que], et il faudra traduire par « Madame a-t-elle bien dormi ? » et pas par « Monsieur a-t-il bien dormi ? » ! Dans le sens inverse, on aurait non seulement le même problème que vers l'allemand (insérer le nom de famille correct), mais il faudrait utiliser le bon niveau de politesse (contexte « situationnel »), pour ne pas traduire de façon familière et inadéquate par « Tanaka san wa yoku neta ka ».

Mais je sais que les systèmes de TA traduisent phrase par phrase. Alors comment peuvent-ils utiliser de l'information absente de la phrase source ?

C'est vrai que la quasi-totalité des systèmes de TA opérationnels segmentent le texte source (ou la transcription obtenue par reconnaissance vocale) en « segments » correspondant idéalement à une phrase ou à un titre, puis les traduisent séparément. Parfois, ces segments sont trop petits, ce sont en fait des « infrasegments », comme les items d'une liste à puces dans une phrase qui en contient plusieurs.

Certains systèmes de TA font cependant un traitement global d'un document à traduire, par exemple pour essayer de déterminer le sens de certains mots ambigus en fonction de tous leurs contextes d'apparition. Par exemple, si notre premier exemple (« they saw many arms ») parle aussi beaucoup de tanks, et d'observateurs, diverses techniques de « désambiguïsation lexicale » permettront de conclure qu'il y a beaucoup plus de chances qu'il s'agisse de voir des armes que de scier des bras. Mais si le contexte est de la médecine médicolégale, la décision sera inverse.

Au niveau syntaxique, un système comme AS-Transac de Toshiba traite un document en deux passes, pour essayer de résoudre des ambiguïtés dites « d'attachement » : si « je vois des cahiers et des classeurs noirs », les cahiers sont-ils noirs ou non, et faut-il traduire par « I see notebooks and black folders » ou par « I see black notebooks and folders », introduisant l'ambiguïté similaire en anglais, ou par « I see black notebooks and black folders » ? Si on observe que, dans le texte, l'adjectif « noir » est souvent relié à « cahier », on préférera la seconde possibilité.

Je vois... mais pouvez-vous revenir aux « approches » et au « progrès » ?

Vous voulez donc savoir « comment ça marche » ? Allons-y. D'abord, il faut voir que, pour traduire par un programme une « unité de traduction », de quelque taille qu'elle soit (phrase, paragraphe, section, chapitre, document), on le fait en général en plusieurs étapes successives, chacune transformant une « représentation intermédiaire » en une autre. Par exemple, en « approche transfert multiniveau », on peut avoir les étapes et représentations suivantes :

<i>Étape</i>	<i>Représentation résultante</i>
Lecture et normalisation des chaînes (ex : ´ → é)	Segment normalisé
Segmentation	Liste (ou graphe) de mots orthographiques
Analyse morphologique	Structure « morphosyntaxique » contenant les « lemmes » et les catégories grammaticales, par ex. « avions » donnerait [avoir, verbe, indicatif, imparfait, p1, pluriel] et aussi [avion, nom, pluriel]

Analyse grammaticale	Arbre source « multiniveau concret ». Il peut contenir des fonctions syntaxiques, des relations logiques et sémantiques, des traits sémantiques, les unités lexicales peuvent être des couples <lemme, acception>, voir des familles dérivationnelles (réparer, réparation, réparateur, réparable, irréparable, irréparablement, avec la possibilité d'accepter aussi *réparabilité et *réparablement). Sa correspondance avec la phrase est simple, en particulier l'ordre est respecté.
Abstraction	Arbre ou graphe « abstrait » source. Il ne contient plus de « nœuds » correspondant aux auxiliaires, aux articles, aux balises de relief (italique, gras...), qui sont transformés en attributs), ni certaines ponctuations (parenthèses), représentées de façon structurale. L'ordre latéral ne correspond plus à celui de la phrase : par exemple, la particule séparable d'une verbe allemand le rejoint.
Transfert (lexical et structural)	Arbre ou graphe « abstrait » cible. Si l'arbre source est « multiniveau », il peut être considéré comme « plus général » que lui, et dit « générateur » si les niveaux de surface (fonction syntaxique, classe syntagmatique et syntaxique) sont proposés et non imposés au générateur.
Choix de paraphrase cible	Cette phase n'est présente que si on part d'une arbre « générateur ». Elle produit alors un arbre abstrait cible, en effectuant les choix possibles selon les critères (souvent stylistiques) du sous-langage cible visé. Par exemple, on produira l'arbre correspondant à « vérifier le fonctionnement correct du dispositif », ou bien l'arbre correspondant à « vérifiez que le dispositif fonctionne correctement ».
Génération syntaxique	Arbre cible « concret ». Il est de même nature qu'un arbre source concret, et peut ou non contenir les informations des niveaux « interlingues » (relations logiques et sémantiques, attributs sémantiques), mais elles ne sont pas utiles pour la suite.
Génération morphologique	Suite de mots, séparateurs et ponctuations formant le segment cible. Il est aussi possible de produire des traductions contenant des propositions multiples, ce qui revient à écrire un graphe avec des alternants, par exemple « [ils elles] [ont vu scient] de [nombreux bras nombreuses armes]. »

La suite des représentations intermédiaires par lesquelles on passe, et leur détail, constitue l'*architecture linguistique* d'un système de TA.

On distingue ainsi les systèmes « directs », « semi-directs », « à transfert » (syntaxique de surface, profond, multiniveau, sémantique), et « à pivot » (hybride ou interlingue). Un « pivot interlingue a son propre « espace lexical », et implique donc la présence de deux transferts lexicaux. Il peut être soit linguistique et basé sur une langue naturelle, comme UNL³, dit « anglo-sémantique », soit sémantico-pragmatique et basé sur un domaine restreint et un petit nombre de tâches (ex : IF⁴ pour les dialogues de réservation en tourisme dans le projet européen Nespole !).

³ Universal Networking Language, voir le site <http://www.undl.org>

⁴ Interface Format

La ou les techniques et ressources utilisées pour programmer les étapes réalisant le passage d'une représentation à une autre constitue l'*architecture computationnelle* d'un système de TA.

Une étape peut être réalisée avec une approche *experte* (par programmation directe, par automates, par règles de réécriture, par règles « statiques » de bonne formation, ou par programmation par contraintes), ou bien une approche *empirique* (statistique, ou par l'exemple).

Enfin, on peut définir l'*architecture opérationnelle* d'un système de TA comme l'ensemble des conditions de sa construction, de son utilisation, de sa maintenance et de son évolution. Autant les deux architectures précédentes sont indépendantes l'une de l'autre, autant celle-ci peut influencer les deux autres. Les techniques et méthodes à employer peuvent donc largement dépendre de la « situation traductionnelle ». Par exemple, on peut avoir à traduire d'une seule langue dans beaucoup d'autres (cas de la traduction de documents techniques), ou bien depuis beaucoup de langues vers une seule (cas des militaires américains).

Disons quelques mots des approches les plus récentes, dites empiriques.

L'approche statistique consiste à supposer l'existence d'un espace probabilisable, par exemple celui des « événements » comme les n-grammes de mots, pour $n \leq 4$ par exemple, et à calculer la sortie « la plus probable » correspondant à l'entrée à transformer. Les premiers systèmes de TA statistique fonctionnaient avec une architecture directe (au niveau des mots typographiques) ou semi-directe (en utilisant les mots et leurs catégories comme V, N, Adj, Adv...). Actuellement, on passe par des représentations hiérarchiques en « fragments » (« chunks ») souvent appelés « groupes » (« phrases ») sans en avoir vraiment le statut linguistique.

Il est cependant tout à fait possible, comme l'a fait Microsoft Research entre 1997 et 2002, de construire un transfert complet à un niveau de syntaxe profonde en alignant 150 000 paires d'arbres (produits par analyse « experte » de segments source et cible provenant de documents techniques déjà traduits). Un calcul statistique mène à la construction d'un « MindNet » dans lequel sont représentés les couples d'« arbrisseaux » (« treelets ») source et cible en relation mutuelle significative. Un arbre d'analyse étant donné, on lui associe une « couverture optimale » par des arbrisseaux source, et on « tricote » un arbre cible en parcourant l'arbre source de haut en bas et en remplaçant les arbrisseaux sources par des arbrisseaux cible correspondants. Il reste à faire une génération classique pour produire une traduction, en général assez bonne ou très bonne.

L'approche « par l'exemple » utilise directement les bi-segments d'un grand corpus bilingue. Là aussi, l'architecture linguistique peut être variée. À un extrême, la « TA par analogie » travaille directement sur les caractères des segments, sans aucun prétraitement. À l'autre, on utilise des bi-segments « préparés », comme par exemple dans le système EBMT de l'USM à Penang (Malaisie), où on utilise des alignements de « correspondances chaîne-arbre structurées ».

La taille des corpus nécessaires diminue avec la « préparation » associée. Pour la TA statistique destinée aux pages Web, comme GoogleTranslate, il faut au moins 50 M mots (dans chaque langue), soit environ 200 000 pages — qui ont nécessité autant d'heures de travail.

Quelques fausses idées

Terminons en « tordant le cou à quelques idées largement répandues, même (ou surtout ?) chez les chercheurs en TA, mais tout bonnement fausses.

La TA à pivot est nécessairement de la TA à règles

Faux, comme le montre par exemple le système MASTOR-I présenté par IBM en 2003. C'est un système de TA de parole entre anglais, mandarin et japonais, sur PDA, utilisant une représentation abstraite de type « liste de propriétés » (liste de couples <attribut, valeur>). Aussi

bien l'enconversion que la déconversion⁵ avaient été « apprises » à partir d'exemples, et étaient calculées par des automates munis de transitions et d'états « pondérés » par des poids assimilés à des probabilités.

En TA à transfert visant N langues, il faut construire N(N-1) étapes de transfert

Faux, car on peut évidemment choisir une langue comme « pivot textuel », et alors il n'y a que 2(N-1) transferts à construire, c'est linéaire et pas quadratique. On peut aussi « composer » des systèmes au niveau de leurs structures intermédiaires en source ou en cible. Par exemple, pour aller du japonais en français, on pourrait utiliser un système japonais-anglais, mais ne pas générer, et au contraire soumettre l'arbre anglais cible à un transfert anglais-français, puis à une génération du français. C'était d'ailleurs l'approche du projet DLT⁶, utilisant de « l'espéranto parenthésé » comme pivot.

Les systèmes de TA à règles ont un coût élevé

Faux. Quand on dit ça, on parle toujours du coût de construction, car ces systèmes sont, sur des machines comparables, beaucoup plus rapides que les systèmes empiriques. Quant à la préparation, il faut inclure le coût de la fabrication des corpus dans celui de la TA statistique, de même qu'on inclut le coût de la construction des dictionnaires et des programmes ou des règles dans ceux de systèmes « experts ».

Les systèmes de TA statistique ont un coût faible

Faux, à cause du coût de construction des très grands corpus parallèles nécessaires, et aussi à cause de l'énormité des calculs. On dit que Google a mis 3000 machines en ligne pour la TA !

La TA statistique est l'approche prépondérante en pratique

Faux. À part GoogleTranslate, et peut-être LanguageWeaver pour certains systèmes spécialisés, tous les systèmes opérationnels répandus ne contiennent aucune phase réellement statistiques.

BLEU mesure la qualité des traductions

Faux. La corrélation avec les jugements humains, excellente pour de très mauvaises qualités, diminue énormément quand la qualité perçue augmente. On a souvent observé que des systèmes commerciaux « à règles » (et même aussi des traducteurs humains très compétents) étaient très mal classés par BLEU, et très bien par des juges humains... Il y a aussi plusieurs arguments scientifiques qui prouvent que BLEU ne peut pas être un indicateur de « qualité ». Par contre, si on met au point un système de TA en cherchant à obtenir des traductions très proches de « traductions de référence » déjà effectuées, et garanties excellentes.

L'adéquation doit être mesurée par un nombre ≥ 0

On peut se poser la question, car un gros contresens est souvent de l'anti-information, et une traduction devrait donc pouvoir être dite « antiadéquate »... Le modèle d'évaluation serait alors le calcul du coefficient kappa d'accord entre la structure représentant le sens de la phrase, et des juges, qui diraient si oui ou non, ou à quel degré, les éléments de sens présents sont les mêmes que ceux de la phrase source (ou de la phrase cible de référence). Dans ce cas, on obtiendrait un score entre -1 et +1.

⁵ On utilise ces termes de préférence à « analyse » et « génération » car il s'agit en fait de traductions comportant un transfert lexical relatif à l'espace lexical (autonome) du « pivot interlingue » choisi.

⁶ Distributed Language Translation, BSO Research, Dordrecht, 1982-1989.

La TA à pivot interlingue (IL), ça ne marche pas

Faux. Le premier système du CETA à Grenoble utilisait un « pivot hybride » et a produit près de 6000 pages de traductions de très bonne qualité. Au Japon, les systèmes ATLAS-II de Fujitsu et PIVOT de Nec sont des systèmes ancêtres d'UNL

La TA à pivot interlingue (IL), ça ne peut pas passer à l'échelle

Faux. ATLAS-II (Fujitsu) est le meilleur pour le couple J↔E depuis 20 an. Il couvre une grande variété de constructions syntaxiques, et surtout il a un énorme dictionnaire, avec au total environ ≈ 7 millions d'entrées dictionnairiques (ATLAS-II v.13, 2006), dont 5 440 000 entrées liées à des domaines techniques.

Ce qui va marcher

À mon avis, le futur est lié à l'architecture opérationnelle. Le voie à suivre, dès qu'on désire de la traduction de qualité, et qu'on ne peut à l'évidence pas payer des professionnels pour la faire, dans l'ensemble des couples de langues visés, est la *TAO contributive externe*. On entend par là qu'on remplacera le problème insoluble de la diffusion de traductions de qualité, très rapidement, et pour de nombreuses langues, par le problème soluble de *la TAO pour l'accès multilingue*.

Pour résoudre ce problème, nous avons proposé et prototypé le concept d'iMAG (interactive Multilingual Access Gateway). Il s'agit d'une « passerelle de traduction » fonctionnant presque comme celle de Google ou de Systran, mais

- dédiée à un « site élu » (en fait, au sous-langage associé à un site ou aux url utilisés par une communauté donnée) ;
- munie d'une mémoire de traduction et d'un lexique terminologique et phraséologique tous deux multilingues et spécialisés au sous-langage en question ;
- offrant, depuis l'environnement de lecture (on lit une page Web à travers un navigateur tout simple, et on n'a rien à installer), l'accès « sans couture » à un environnement de post-édition munis d'aides traductionnelles et dictionnairiques « proactives » ;
- associant différents « niveaux de qualité » et des « scores par défaut » aux producteurs de traductions (* pour le mot à mot, ** pour la TA, *** pour des bilingues, **** pour des traducteurs professionnels, et ***** pour des bilingues « agréés » par le site élu et pouvant certifier des traductions). Les contributeurs ont un niveau fonction de leur profil, et un score par défaut. Mais ils peuvent modifier le score qu'ils se donnent à eux-mêmes pour la traduction d'un segment (c'est ce qui est fait en pratique quand on post-édite un corpus à l'aide de l'outil SECTra_w que nous avons construit).

Dans un premier temps, on peut utiliser des systèmes existants pour proposer des « prétraductions », ensuite révisée *dans le contexte de lecture sur le Web* par les lecteurs, transformés en contributeurs bénévoles. On peut aussi, quand rien n'existe pour un couple de langues, demander à des humains de traduire.

Dans un second temps, on pourra utiliser les données collectées à l'occasion des post-éditions pour construire des systèmes de TA, en choisissant leur architecture linguistique et leur architecture computationnelle en fonction de la situation et des ressources humaines, lexicales et « corporales » disponibles.

Références

- Boitet C. (1986) Current Machine Translation systems developed with GETA's methodology and software tools. *Proc. ASLIB Conf.*, Nov. 1986, 12 p.
- Boitet C. (1986) The French National MT-Project: technical organization and translation results of CALLIOPE-AERO. *Computers and Translation*, 1, pp. 281—309.
- Boitet C. (1988) L'apport de Bernard Vauquois à la traduction automatique et au traitement automatique des langues naturelles. *Proc. Colloque sur l'Histoire de l'Informatique en France.*, 3-5 mai 1988, P. Châtelin, ed., vol. 2/2, pp. 63—82.
- Boitet C. (1988) PROs and CONs of the pivot and transfer approaches in multilingual Machine Translation. *Proc. Int. Conf. on "New directions in Machine Translation"*, 18–19 August 1988, BSO, ed., Foris Publications, pp. 93—108.
- Boitet C. (1988) Representation and Computation of Units of Translation for Machine Interpretation of Spoken Texts. *Computers and Artificial Intelligence*, 8/6, pp. 505—546.
- Boitet C. (1988) Software and lingware engineering in modern M(A)T systems. In *Handbook for Machine Translation*, Bátor, ed., Niemeyer.
- Boitet C. (1993) La TAO comme technologie scientifique : le cas de la TA fondée sur le dialogue. In *La traductique*, A. Clas et P. Bouillon, ed., Presses de l'Université de Montréal, Montréal, pp. 109—148.
- Boitet C. (1993) TA et TAO à Grenoble... 32 ans déjà ! *T.A.L. (revue semestrielle de l'ATALA)*, 33/1—2, Spécial Trentenaire, pp. 45—84.
- Boitet C. (1995) Factors for success (and failure) in Machine Translation — some lessons of the first 50 years of R&D. *Proc. MTS-V (Fifth Machine Translation Summit)*, 11—13 July 1995, CEE, 17 p.
- Boitet C. (2001) Machine Translation. In *Encyclopedia of Cognitive Science*, Nature Publishing Group, London, (in manuscript form) 24 p.
- Boitet C. (2002) A rationale for using UNL as an interlingua and more in various domains. *Proc. LREC-02 First International Workshop on UNL, other Interlinguas, and their Applications*, Las Palmas, ELRA/ELDA, pp. 23—26.
- Boitet C. (2003) Automated Translation / Traduction Automatique. *Revue française de linguistique appliquée*, VIII-2, pp. 99-121.
- Boitet C. (2005) New architectures for "democratic" tunable quality MT systems. *Proc. Pacific Association for Computational Linguistics (Proc. PACLING-2005)*, Meisei daigaku, Hino campus, PACLING Association, pp. 33—57. (invited communication)
- Boitet C., Bey Y., Tomokiyo M., Cao W. et Blanchon H. (2006) IWSLT-06: Experiments with Commercial MT Systems and Lessons from Subjective Evaluations. *Proc. IWSLT-06 (International Workshop on Spoken Language Translation)*, Kyoto, ATR, pp. 23—30.
- Boitet C. et Blanchon H. (1994) Promesses et problèmes de la "TAO pour tous" après LIDIA-1, une première maquette. *Langages*, 116, pp. 20—47.
- Boitet C., Boguslavskij I. et Cardenosa I. (2007) An Evaluation of UNL Usability for High Quality Multilingualization and Projections for a Future UNL++ Language. In *Computational Linguistics and Intelligent Text Processing (Proc. CICLING-2007)*, A. Gelbukh, ed., Springer (LNCS 4394), pp. 361-373. (ISBN-10: 3-540-70938-X Springer, ISSN: 0302-9743)

Boitet C., ed. (1988) BERNARD VAUQUOIS et la TAO, vingt-cinq ans de Traduction Automatique, ANALECTES. BERNARD VAUQUOIS and MT, twenty-five years of MT. Ass. Champollion & GETA, Grenoble,

Boitet C. et Gerber R. (1986) Expert Systems and other new techniques in MT. In *Neue Ansätze in maschineller Sprachübersetzung*, Niemeyer, Tübingen, pp. 103—119.

Boitet C. et Tsai W.-J. (2002) La coédition langue \leftrightarrow UNL pour partager la révision entre les langues d'un document multilingue : un concept unificateur. *Proc. TALN-02*, Nancy, ATALA, 12 p.

Eisele A. (2005) *Exploiting Multilingual Corpora for Machine Translation*. (JRC Enlargement and Integration Workshop on Exploiting parallel corpora in up to 20 languages), Arona, Saarland University & DFKI, (slides)

Hutchins W. J. (1986) *Machine Translation : Past, Present, Future*. Ellis Horwood, John Wiley & Sons, Chichester, England, 382 p.

Hutchins W. J. et Somers H. L. (1992) *An Introduction to Machine Translation*. H. B. Jovanovich, ed., Academic Press, 362 p.

JEIDA (1989) *A Japanese view of Machine Translation in light of the considerations and recommendations reported by ALPAC, USA*. Japanese Electronic Industry Development Association, Tokyo.

Kraif O. (2006) *Corpus multilingues — multilingual corpora*. 22/11/06.

http://w3.u-grenoble3.fr/kraif/index.php?option=com_content&task=view&id=20&Itemid=36

Lafourcade M. et Boitet C. (2002) UNL Lexical selection with conceptual vectors. *Proc. LREC-02 (International Conference on Language Resources and Evaluation)*, Las Palmas, vol. VI/VI, pp. 1958—1964.

Sénelart J., Boitet C. et Romary L. (2003) XML Machine Translation. *Proc. MTS-IX (Machine Translation Summit)*, New-Orleans, 9 p.

Thurmair G. (2006) Using corpus information to improve MT quality. *Proc. LR4Trans-III (3rd International Workshop on Language Resources for Translation Work, Research & Training)*, LREC 2006, Genoa, ELRA / ELDA, 4 p.

Uchida H. (2004) *The Universal Networking Language (UNL) Specifications Version 3 Edition 3*. UNL Center, UNDL Foundation, December 2004.

<http://www.undl.org/unlsys/unl/UNLSpecs33.pdf>

Vauquois B. et Boitet C. (1985) Automated translation at Grenoble University. *Computational Linguistics*, **11**/1, January-March 85, pp. 28—36.