



HAL
open science

Comparatif traduction automatique, traduction assistée par ordinateur et traduction humaine dans le domaine de la finance. Session 3 - TA et Biotraduction

Sandrine Peraldi, Muriel Bellamy, Marion Chesné, Camille Lehmann, Audrey Prost

► To cite this version:

Sandrine Peraldi, Muriel Bellamy, Marion Chesné, Camille Lehmann, Audrey Prost. Comparatif traduction automatique, traduction assistée par ordinateur et traduction humaine dans le domaine de la finance. Session 3 - TA et Biotraduction. Tralogy II. Trouver le sens : où sont nos manques et nos besoins respectifs ?, Jan 2013, Paris, France. 23p. hal-02497326

HAL Id: hal-02497326

<https://hal.science/hal-02497326>

Submitted on 3 Mar 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.

TRALOGY

Comparatif traduction automatique, traduction assistée par ordinateur et traduction humaine dans le domaine de la finance

Sandrine Peraldi

ISIT
s.peraldi@isit-paris.fr

Muriel Bellamy

Marion Chesné

Camille Lehmann

Audrey Prost

TRALOGY II - Session 3
Date d'intervention : 17/01/2013

lien video : http://webcast.in2p3.fr/videos-_comparatif_traduction_automatique



Introduction

La présente communication décrit la mise en œuvre d'un projet de recherche appliquée dans le domaine de la traduction automatique (TA). Il s'agit, plus spécifiquement, d'une analyse, menée notamment par des étudiants de Master 2 en traduction (Bellamy, Chesné, Lehmann, Prost, 2011) sous la direction d'un enseignant-chercheur, visant à déterminer l'efficacité d'une approche combinée entre traduction automatique et traduction assistée par ordinateur, et ce, dans le domaine de la finance. La traduction automatique est donc abordée ici sous un angle extrêmement pragmatique dans la mesure où ce projet répondait à un besoin concret énoncé par un véritable commanditaire.

En effet, sous l'impulsion d'une société de traduction spécialisée dans l'information financière réglementée, nous avons cherché à déterminer si la combinaison de plusieurs outils pouvait constituer une alternative fiable à la traduction humaine, tant en termes de qualité, de gain de temps et de coûts, tout en permettant de rationaliser au maximum le processus de traduction. Les sociétés de traduction spécialisée dans la finance sont contraintes de traduire des volumes extrêmement importants dans des délais très courts, tout en respectant une qualité optimale dans la mesure où les documents sont contrôlés et publiés par des organismes officiels. Or, malgré les récents progrès en traduction automatique (notamment par la mise en place de moteurs hybrides combinant une approche statistique et linguistique), les systèmes informatiques ne sont pas encore à même de produire des traductions de qualité optimale, nécessitant une révision (post-édition) plus ou moins importante de la part du traducteur, notamment face à des domaines de spécialité à forte densité terminologique et complexité syntaxique.

Cette mise en situation professionnelle dans le domaine de la traduction financière a donc donné lieu à une préconisation remise au commanditaire, tout en conférant à la recherche une forte dimension applicative. Aussi, après avoir évoqué la problématique de la traduction automatique et celle de la traduction assistée par ordinateur dans le marché actuel, nous passerons en revue les différentes combinaisons de méthodologies éprouvées au cours de ce projet ainsi que la typologie d'erreurs que nous en avons retirée afin d'évaluer la pertinence et la viabilité de la TA dans le domaine de spécialité précité.

1. Traduction assistée par ordinateur : approche restreinte versus approche étendue

Force est de constater que l'évolution constante des logiciels pour intégrer de nouvelles fonctions et gagner des parts de marché ne cesse de bouleverser le paysage de la traduction assistée par ordinateur. La frontière entre logiciel à moteur de traduction automatique et logiciel à mémoire de traduction s'estompe de plus en plus pour laisser place à des outils combinant ces différentes méthodes. La traduction assistée par ordinateur, communément abrégée TAO, revêt en réalité une acception large et une acception étroite, soit deux types de définitions qu'il convient de préciser ici.

Sauron (2004, 2007), dans le cadre de son enseignement à des étudiants de Master 2 en traduction, définit la traduction assistée par ordinateur comme « *l'ensemble des programmes et outils informatiques susceptibles d'aider le traducteur et le rédacteur à la fois dans ses recherches, dans la constitution de ressources et lors du processus de traduction* ».

Dans cette acception large, la TAO regroupe par conséquent l'ensemble des logiciels dont est amené à se servir le traducteur qui traduit à l'aide d'un ordinateur, quel que soit le degré de spécialisation desdits logiciels. À ce titre, les mémoires de traduction, les moteurs de traduction automatique, les bases terminologiques, les concordanciers et les extracteurs de terminologie font bien évidemment partie intégrante de cette définition (bien que ces deux derniers outils

relèvent plutôt du champ d'intervention du terminologue). Mais c'est également le cas des logiciels informatiques généraux, d'usage quotidien utilisés par n'importe quel rédacteur, à savoir : les outils de traitement de texte, les navigateurs web et, éventuellement, les logiciels de reconnaissance vocale. Cette définition a pour avantage de présenter la TAO comme un ensemble complexe d'outils que le traducteur pragmatique doit savoir maîtriser pour pouvoir mener à bien son travail. Dans ce cadre, la traduction automatique fait partie intégrante des outils d'aide à la traduction, elle en constitue un sous-ensemble, on ne peut donc opposer traduction assistée par ordinateur et traduction automatique à proprement parler.

Dans une acception restreinte et plus couramment employée, la traduction assistée par ordinateur désigne une « *technique permettant d'engendrer des textes en langue cible à partir d'une langue source par un système informatique avec l'assistance d'un traducteur humain* » (GDT, Office de la langue française, 1997).

Ainsi définie, la TAO recouvre l'utilisation des logiciels fonctionnant à l'aide d'une mémoire de traduction et met l'accent sur l'existence d'une intervention humaine au cours du processus de traduction. Le grand dictionnaire terminologique (GDT) précise que cette assistance se situe à trois niveaux : à la pré-édition (annotation ou transformation du texte source), durant la traduction (système interactif), en post-édition (révision de la traduction). En adoptant cette conception de la TAO centrée sur l'intervention humaine, il devient possible d'opposer traduction assistée par ordinateur et traduction automatique puisque cette dernière est définie comme une « *technique informatique permettant d'obtenir de façon automatique, sans l'intervention d'un traducteur humain, une traduction de textes d'une langue source vers une autre langue* » (GDT, Office de la langue française, 2006).

Pour être plus exact, l'intervention humaine est reléguée aux deux autres niveaux, en amont et en aval de la traduction, respectivement lors de la préparation des glossaires et lors de la post-édition. Si le texte est effectivement traduit de manière automatique, il nécessite cependant une révision d'autant plus poussée que la traduction n'est pas générée par le biais d'un traducteur humain. C'est une problématique essentielle de la traduction automatique qui resurgit avec force dans l'analyse à laquelle nous avons procédé.

Par commodité, nous avons choisi d'employer dans le cadre de cette communication, le terme de TAO pour désigner la fonctionnalité spécifique de mémoire de traduction et le terme de TA pour désigner la fonctionnalité spécifique de moteur de traduction automatique, sachant que les logiciels que nous avons analysés sont des programmes composites qui intègrent à la fois mémoire de traduction et moteur de traduction automatique. Néanmoins, à l'issue de notre analyse, nous essaierons de déterminer quelle acception semble la plus propice dans l'état actuel du marché de la traduction.

1.1 Contexte et enjeux du projet

L'entreprise à l'origine de ce projet de recherche appliquée édite, met en forme et traduit des documents financiers périodiques tels que rapports annuels, semestriels ou trimestriels, ou spécifiques à des opérations de marché, et ce, à destination de la majorité des entreprises françaises cotées en Bourse. L'ensemble du processus s'effectue sur une période très courte, étant donné que l'édition des documents de référence des entreprises intervient en même temps, en début d'année. Les traducteurs disposent donc d'un délai relativement court, généralement une dizaine de jours, contraignant le commanditaire à scinder les documents entre plusieurs traducteurs, se traduisant par un risque majeur en termes de qualité de la traduction, de cohérence et d'harmonisation terminologiques et de mise en page (essentielle dans ce contexte). Quant aux documents à traduire, ils sont extrêmement volumineux (plusieurs centaines de pages) et répétitifs, d'où l'intérêt pour les traducteurs de recourir à des outils d'aide à la traduction, tout en parvenant à terme à optimiser le processus d'alignement et de mise en mémoire extrêmement chronophage.

D'ordinaire, lorsqu'on envisage de recourir à un outil de TA, deux critères textuels précis nous permettent de présager plus ou moins de la qualité de la traduction produite et donc de la pertinence du recours à ce type d'outils : la nature du texte à traduire et sa finalité. La prise en compte de ces indicateurs relève clairement d'une approche ciblée de la traduction automatique, c'est-à-dire déterminée en fonction du type de texte à traduire, par opposition à l'approche universelle qui a longtemps prévalu en la matière, à savoir la croyance selon laquelle la traduction humaine disparaîtrait au profit de traductions faites de manière instantanée par un logiciel.

La première question qu'il importe donc de poser est la suivante : le document à traduire est-il rédactionnel ? Spécialisé ? Rédigé en langue libre ou au contraire en langue contrôlée ? On estime habituellement que les textes répétitifs, sans ambiguïté, utilisant un vocabulaire restreint et une syntaxe simple, à l'exemple des langues contrôlées, constituent un terrain plus favorable à la traduction automatique. De la nature et de la complexité du texte dépend grandement la qualité de la TA.

Par ailleurs, il est essentiel de déterminer si le document cible doit être publié, requérant de fait une qualité irréprochable, ou s'il doit être seulement compréhensible, auquel cas on peut généralement se contenter d'une version imparfaite, voire parfois syntaxiquement incorrecte. La traduction automatique permet en effet de prendre rapidement connaissance de gros volumes de documents, pour peu que le lecteur sache combler les lacunes du discours produit automatiquement. En guise d'illustration, cet extrait du site internet de Microsoft, dont une partie de l'aide en ligne a été traduite à l'aide d'un logiciel de TA développé par Microsoft Research :

« La qualité linguistique de la traduction automatique ne peut bien sûr pas atteindre le même niveau de qualité que celle fournie par des traducteurs professionnels. [...] Néanmoins, la qualité de la traduction obtenue est très souvent suffisante pour vous permettre de répondre à votre question ou de résoudre votre problème technique. » (Aide et Support Microsoft [en ligne], 2011).

Dans le cas qui nous occupe, la recherche qualitative constitue un point essentiel dans notre réflexion et approche méthodologique. En effet, l'accessibilité, la précision et la disponibilité des documents produits par l'entreprise commanditaire sont, d'après cette dernière, indispensables à une transparence financière optimale. En l'occurrence, l'ensemble des documents à caractère financier délivrés par les sociétés cotées en Bourse doivent être systématiquement déposés auprès de l'Autorité des Marchés Financiers (AMF), qui supervise et contrôle à la fois la qualité et l'adéquation des informations ainsi émises. Par conséquent, l'éventuel recours à la traduction automatique doit non seulement ne pas générer de risque financier, dû à l'investissement dans des logiciels somme toute coûteux lorsqu'ils sont déployés à l'échelle d'une entreprise et/ou aux coûts liés à la post-édition, ni de déperdition de la qualité pour le commanditaire.

Or, les nouvelles générations de logiciels de TA intègrent désormais un certain nombre de règles et d'algorithmes visant à imiter au mieux des processus cognitifs à l'œuvre lors de la traduction et améliorer la qualité des traductions produites. Avec les progrès incontestables faits dans la conception des moteurs de traduction automatique, la TA a aujourd'hui la prétention de s'attaquer à des textes plus rédactionnels, appartenant à des domaines de spécialité de plus en plus techniques.

L'enjeu de ce projet consistait donc à déterminer s'il était possible de réaliser, par le biais d'une combinaison TA/TAO, une réduction du nombre d'intervenants dans la chaîne de traduction, ce qui entraînerait de fait un gain de temps et d'argent considérable pour l'entreprise, sans toutefois mettre en péril la qualité de la traduction.

1.2 Outils et méthodologies

Pour répondre à cette question, une méthodologie de recherche d'ordre comparatif a été mise en place afin d'identifier dans un premier temps la méthode logicielle la plus appropriée, suivie dans un deuxième temps de l'élaboration d'une typologie d'erreurs afin d'évaluer la qualité de

la méthode sélectionnée, pour finalement comparer cette dernière au processus de traduction actuel de l'entreprise, à savoir traduction humaine et outils d'aide à la traduction.

Qu'il s'agisse de la TA ou de la TAO, les deux outils sélectionnés correspondent aux deux logiciels les plus utilisés sur le marché français, logiciels que nous ne nommerons pas pour des raisons de confidentialité.

1.2.1 Traduction assistée par ordinateur

Le logiciel de TAO éprouvé repose très classiquement sur un système d'alignement et de mémoire de traduction, autrement dit une base de données linguistique qui stocke sous forme de paires (appelées « unités de traduction » ou UT) des segments de texte en langue source et les segments en langue cible qui leur correspondent, en vue d'une réutilisation ultérieure. L'unité de traduction la plus fréquente est la phrase, mais on rencontre également d'autres cas de figure : il est possible de lier plusieurs phrases sources à une phrase cible (ou vice-versa), ou de ne lier que des parties de phrase (ex. des éléments de liste). L'unité de traduction est avant tout une unité sémantique.

Les mémoires de traduction gérées par l'outil concerné contiennent des unités de traduction indépendantes les unes des autres. Il s'agit d'une approche micro-textuelle, qui décontextualise l'unité de traduction du texte d'origine, ce qui peut représenter un inconvénient lors d'une traduction ultérieure. L'option de recherche contextuelle pallie en partie cette faiblesse propre aux mémoires de traduction segmentées. La reconnaissance des unités de traduction présentes dans la mémoire de traduction s'appuie sur un système de paramétrage du seuil de similitude des segments de traduction, faisant la distinction entre concordances parfaites (le segment à traduire possède 100 % de similitude avec un segment source en mémoire), concordances partielles (degré de similitude inférieur à 100 % en raison d'un ajout, d'une suppression ou d'une interversion de mots qui entraînent des pénalités) et les nouveaux segments (degré de similitude inférieur à un degré minimum paramétrable, en général fixé à 75 %).

L'utilisation des mémoires de traduction s'avère particulièrement judicieuse lorsque les documents traduits présentent un taux de récurrence élevé, ce qui est souvent le cas pour les rapports annuels publiés par les grandes entreprises, dont les versions successives possèdent un fort pourcentage de segments similaires ou identiques d'une année sur l'autre. Il est à noter que la principale version du logiciel de TAO éprouvé (2007) ne possède pas de module intégré de TA, contrairement à la version 2009 évoquée en fin de communication.

Enfin, le logiciel a été paramétré de façon classique, notamment avec une valeur minimale d'analogie à 70 % afin que le système soit en mesure de récupérer un maximum de segments ou de fragments de segments déjà traduits et non seulement les segments « parfaits » (100 % de similitude), reconnaissables les uns des autres par le biais d'un jeu de couleur.

L'utilisation des transposables a également été privilégié étant donné l'importance des données chiffrées au sein du document.

1.2.2 Traduction automatique

Le logiciel de traduction automatique sélectionné par le commanditaire repose, dans sa septième version, sur un moteur de TA hybride qui combine règles linguistiques et règles statistiques, associé par ailleurs à un module interne de mémoire de traduction qui permet de réduire le recours à la traduction automatique sur des textes qui possèdent un fort taux de récurrence d'une version à l'autre. Plus spécifiquement, ce moteur de traduction est d'abord fondé sur une analyse syntaxique du texte (également appelée méthode linguistique ou fonctionnement par règles). Autrement dit, le système commence par analyser la structure de la phrase à traduire :

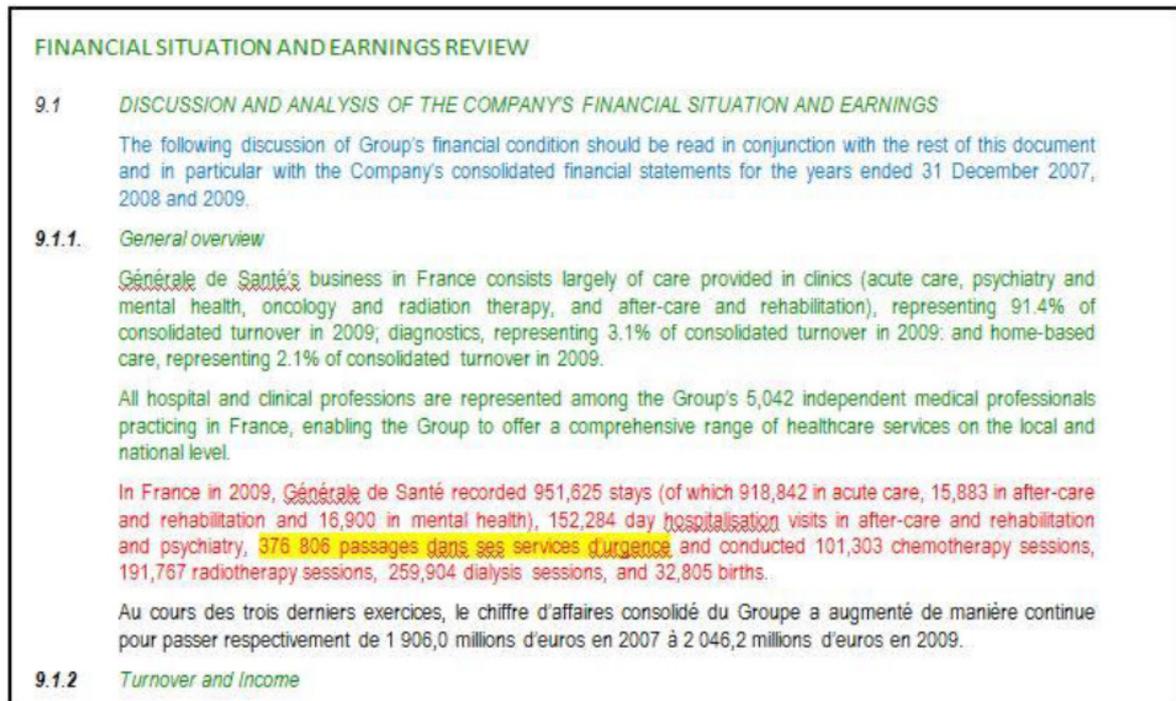


Figure 1 : Mise en exergue des segments en fonction du taux de similitude

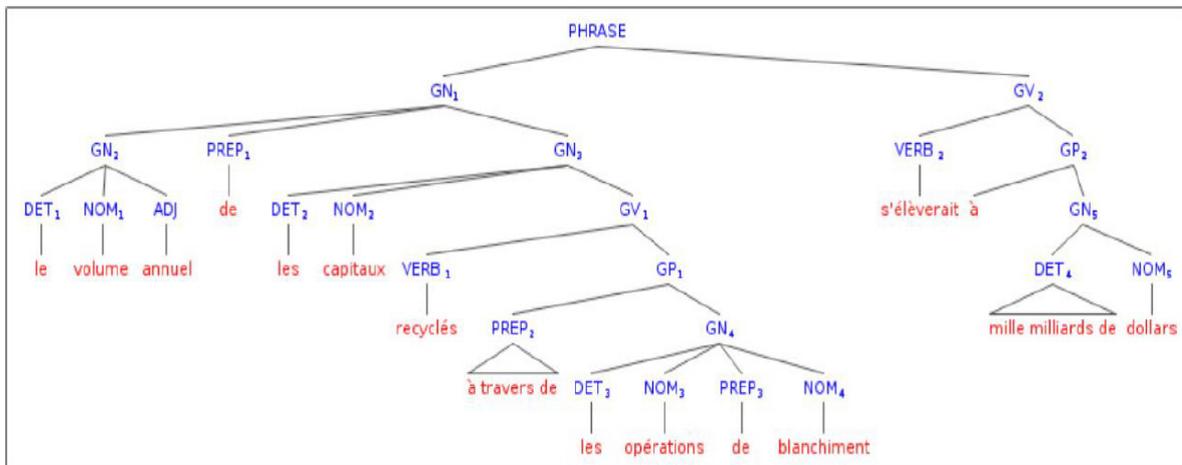


Figure 2 : Analyse linguistique - Exemple de décomposition de phrase opérée par le logiciel de traduction automatique.

il distingue les éléments qui la constituent, puis y applique un ensemble de règles de grammaire et de syntaxe afin de reconstruire la phrase source.

Cette représentation intermédiaire de la phrase, obtenue à l'aide de données syntaxiques, morphologiques et sémantiques, et de nombreuses règles linguistiques, permet au logiciel de « comprendre » le texte qu'il doit traduire, puis de transférer la structure grammaticale du texte source dans le texte cible et enfin de puiser dans des dictionnaires volumineux pour remplacer chaque mot par un équivalent en langue étrangère. Ainsi, pour la phrase *il fait chaud*, l'analyse linguistique, si elle est correcte, détectera un pronom (PRO), un groupe verbal (GV) composé d'un verbe (V) et un adverbe (ADV) et identifiera la règle suivante :

V(faire ; sujet → «il» ; MODIFIEUR = chaud, froid, gris, mauvais, beau).

En anglais, cette construction se traduira alors par « the weather is MODIFIER » (le dernier terme désignant le modifieur, traduit en anglais).

Néanmoins, en raison des nombreux particularismes locaux, des évolutions de la langue, de la complexité grandissante des langues et du degré de proximité structurelle plus ou moins fort entre les langues de départ et d'arrivée (selon la combinaison linguistique concernée), il est difficile pour un moteur de recenser toutes les règles nécessaires pour gérer n'importe quel type de phrase. Qui plus est, le moteur gère très mal l'ambiguïté et la polysémie, puisqu'il analyse chaque unité séparément avant même de mettre les mots en relation les uns avec les autres. La référence au contexte et la polysémie constituent en effet les plus grands obstacles auxquels fait face un logiciel de TA, qui produit une traduction littérale à l'extrême dès lors qu'il s'appuie exclusivement sur une analyse structurale du texte.

La traduction comme encodage et décodage de signes régis par des règles présente par conséquent de nombreuses limites. Dès lors, pour améliorer la qualité des traductions produites, une nouvelle approche est désormais associée au moteur linguistique : celle de l'analyse statistique. L'idée est d'indexer de grands volumes de textes ainsi que leur traduction dans des bases de données, puis d'appliquer à ces bases de données un modèle statistique pour déterminer quelle est la traduction la plus probable selon le contexte. Initialement développée par IBM, cette technique est popularisée par Google pour le développement de son outil de TA accessible gratuitement en ligne, Google Translate. Le quasi-monopole de Google sur le web lui permet de disposer de la plus grande base de données jamais constituée, car il indexe tous les contenus disponibles en libre accès sur internet.

Bien que cette méthode pose un certain nombre de problèmes en termes de droit d'auteur et de confidentialité, mais également en termes de qualité (le moteur absorbe également toutes les phrases qui lui sont soumises par des utilisateurs par le biais de Google Translate, sans égard quant à leur provenance), elle offre l'avantage non négligeable de prendre comme base de travail des traductions humaines, ce qui permet d'obtenir un rendu fluide à la lecture, proche de ce que pourrait écrire un traducteur humain. De même qu'un traducteur faisant une recherche sur internet ne s'arrête pas au premier résultat trouvé, le logiciel utilise les probabilités pour déterminer quelle est la traduction la plus pertinente en fonction du nombre d'occurrences trouvées (partant du principe que la traduction la plus fréquemment utilisée est la bonne traduction). La méthode statistique a donc pour elle une grande souplesse, car elle s'adapte aux innovations de la langue en recalculant régulièrement la fréquence d'apparition d'une traduction donnée.

Toutefois, les traductions produites par ce biais ne sont de qualité que si le corpus est volumineux et pertinent. Selon l'éditeur de logiciels de traduction automatique SYSTRAN (2011), l'entraînement d'un modèle bilingue nécessite au minimum deux millions de mots pour un domaine spécifique et bien plus pour le domaine général. Dans les faits, il est rare que les entreprises disposent d'un tel volume de documents traduits. Ces chiffres expliquent pourquoi le moteur de Google est probablement le plus performant à l'heure actuelle, particulièrement dans les domaines où les documents sont en libre accès (à titre d'exemple, leur base de données intègre l'ensemble des documents de référence traduits par les services de l'ONU). Par ailleurs, l'approche statistique de la TA bute sur les phrases courtes et décontextualisées, car le logiciel n'a pas suffisamment d'indices textuels à sa disposition pour trancher sur la traduction appropriée.

Enfin, l'approche exclusivement statistique ne permet pas de traduire un texte totalement nouveau, qui ne possède pas de traduction antérieure, expliquant l'arrivée sur le marché de moteurs hybrides, couplant approche linguistique et statistique, à l'image du logiciel éprouvé dans le cadre de ce projet.

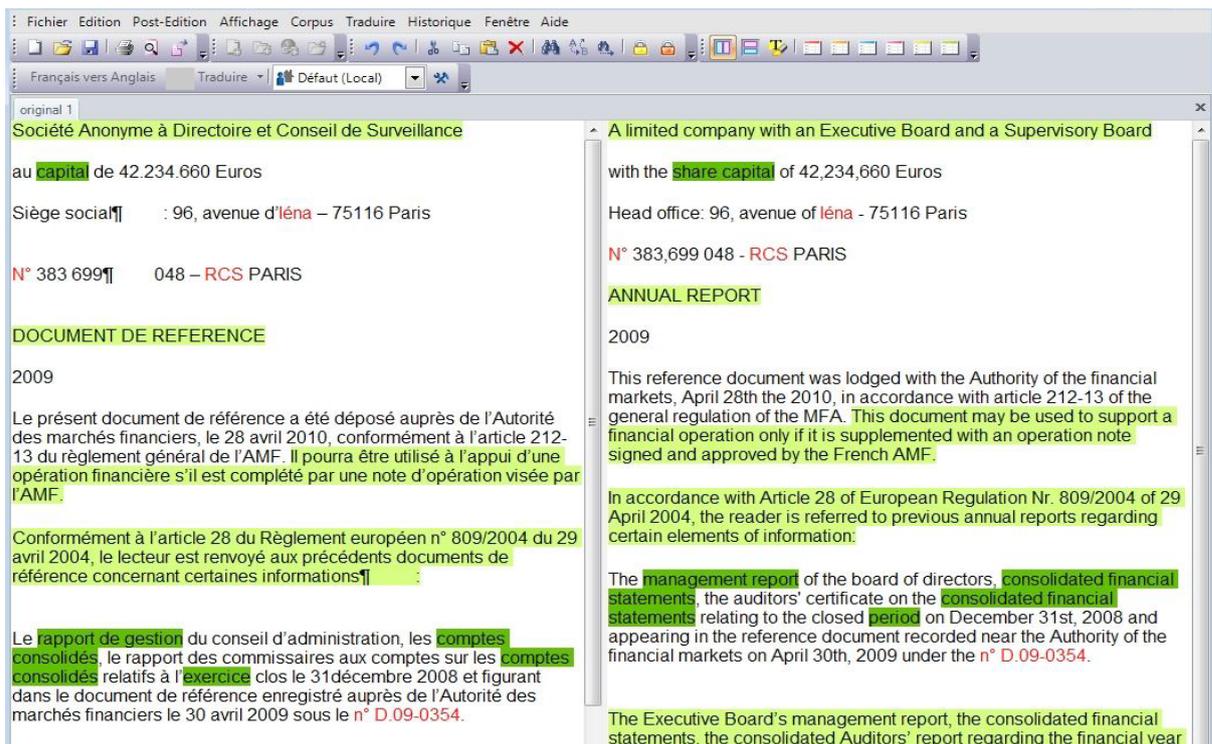


Figure 3 : Interface du logiciel de traduction automatique

Dans un premier temps, l'application des règles permet pour chaque mot ou groupe de mots de déterminer différentes traductions possibles. Les règles permettent de généraliser, c'est-à-dire de traduire une phrase jamais rencontrée en construisant une forme jamais produite. Dans la plupart des cas, plusieurs solutions sont trouvées. L'approche statistique vient alors en appui pour valider la meilleure proposition.

Dans la version que nous avons utilisée et qui est destinée aux entreprises, le système passe par une phase d'apprentissage, qui consiste à indexer les documents déjà traduits. L'entreprise doit en définir la liste, au sein de laquelle le logiciel bâtira son index, par auto-apprentissage.

Le panachage entre approche par règles et approche par statistique permet d'obtenir des traductions automatiques de bien meilleure qualité, qui nécessitent en principe une relecture allégée par rapport à la révision nécessaire lorsque l'on applique une méthode unique. Nous nous devons de spécifier toutefois qu'en raison des délais impartis, nous n'avons pas été en mesure d'entraîner le système, ce qui peut constituer une certaine limite à la représentativité et à la généralisation de nos résultats.

Par ailleurs, les options de personnalisation du logiciel permettent également d'enrichir les dictionnaires par défaut en ajoutant des glossaires utilisateurs comportant la terminologie spécifique à un domaine ou à une entreprise, afin d'optimiser la traduction sur les segments inconnus. Nous avons à notre disposition un glossaire de termes de finance, spécifique au commanditaire et élaboré dans le cadre d'un autre projet de recherche appliquée mené par nos soins.

Ainsi, le logiciel qui était auparavant exclusivement dédié à la traduction automatique devient un logiciel composite qui intègre une fonctionnalité de TAO, en permettant de créer ou d'utiliser des mémoires de traduction déjà existantes. L'utilisation conjointe de la traduction humaine (segments issus de la fonction TAO) et de traduction automatique optimisée (moteur de TA

hybride) est censée produire un texte de qualité optimale par rapport à un texte composé à l'aide d'une seule méthode.

1.3 Méthodologies

Nous avons donc cherché à déterminer si les documents traités par le commanditaire se prêtaient à l'utilisation combinée ou non de la traduction automatique, notamment en analysant le pourcentage de segments récurrents dans le texte source d'une année sur l'autre (critère d'adaptation à la TAO) ainsi que le style rédactionnel et la qualité générale de la traduction produite automatiquement (critère d'adaptation à la TA). Cette analyse a été complétée d'une simulation en termes de temps et de coûts pour chacune des méthodes testées. Plus spécifiquement, notre approche méthodologique a été scindée en trois phases distinctes :

1. Nous avons d'abord testé la combinaison TA/TAO sur un document de référence, soit en combinant des outils de TA et TAO distincts, soit en utilisant un outil composite, c'est-à-dire possédant ces deux fonctionnalités, en l'occurrence l'outil de TA présenté auparavant.

L'objectif de cette étape étant de comparer les deux méthodes en jeu et, à terme, de déterminer laquelle de ces deux méthodes paraissait la plus avantageuse et la plus efficace en termes de temps et de facilité de prise en main.

2. Une fois définie la méthode la plus pertinente, nous avons élaboré une typologie d'erreurs en catégorisant les erreurs rencontrées, mais également leur niveau de gravité, afin de nous permettre d'évaluer le temps de révision nécessaire pour produire une traduction de qualité. Cette étape est cruciale dans la mesure où la qualité de la traduction constitue le critère le plus essentiel pour l'entreprise commanditaire.

3. Enfin, la dernière étape a consisté à comparer la « solution préférée » avec la méthode actuelle (l'utilisation de la TAO seule) afin de déterminer, sous forme de préconisation à remettre au commanditaire, si cette nouvelle méthode analysée apportait véritablement une plus-value pour l'entreprise en termes de temps, tout en maintenant une qualité nécessaire à la traduction des documents financiers qui lui sont confiés..

1.4 Document de référence

Le rapport financier fourni par l'entreprise correspond au document de référence 2009 de l'entreprise Générale de Santé. Le style de ce document correspond en tous points à ce qu'on attend d'un document financier : phrases longues et complexes, terminologie spécialisée (financière bien entendu, mais aussi médicale et juridique) entre autres. Le document est également constitué de nombreux tableaux chiffrés ainsi que d'un certain nombre d'images comportant du texte, nécessitant un recours à la publication assistée par ordinateur.

Enfin, comme l'indique le tableau d'analyse ci-dessous (Figure 3), le document présente un fort taux de récurrence par rapport au document de référence de la Générale de Santé de l'année précédente (2008), avec seulement 20 % de segments non traduits. Il se prête donc parfaitement à une analyse combinée TA/TAO :

Start Analyse: 28/02/2011 17:20:12				
Analyse du texte de 250 pages				
Analyse Total	(1 file):			
Match Types	Segments	Words	Percent	Placeables
Context TM	0	0	0	0
Repetitions	338	2 122	2%	1
100%	3 432	52 929	56%	0
95% - 99%	328	10 352	11%	2
85% - 94%	276	6 543	7%	0
75% - 84%	190	3 458	4%	1
50% - 74%	76	1 468	2%	0
No Match	1 337	18 394	18%	0
Total	5 977	95 266	100%	4
Pages	250			

Figure 4 : Analyse du taux de récurrence du document de référence

L'ensemble des analyses ont été menées sur un échantillon de 50 % du document de référence afin de privilégier l'analyse qualitative sur l'étape proprement dite de traduction.

2. Analyse et résultats

2.1 Combinaison TA/TAO

La première méthodologie éprouvée a consisté à pré-traduire le document de référence avec le logiciel de TAO en effectuant le travail de correction classique d'un relecteur (transposables, chiffres, dates, etc.), puis à recourir au logiciel de TA pour traduire et éventuellement corriger les segments restants. Trois solutions étaient alors envisageables pour mener à bien la traduction automatique des nouveaux segments :

1. Exporter les segments non-traduits et les traduire avec le logiciel de TA.
2. Charger le fichier pré-traduit par l'outil de TAO dans le système de et ne traduire que les passages en français.
3. Utiliser les deux logiciels en parallèle.

2.1.1 Export des segments non-traduits

Malgré un certain nombre de problèmes de compatibilité de fichiers rencontrés entre les deux logiciels, il a été finalement possible d'exporter dans un nouveau fichier les segments non-traduits de la TAO et les charger dans le logiciel de TA (en utilisant le format TMX 1.1).

Cette solution s'est avérée inapplicable en l'état dans la mesure où il n'existe aucune fonctionnalité qui permette de réincorporer automatiquement ces segments dans le document pré-traduit, nécessitant dès lors un copié-collé extrêmement fastidieux, chronophage et à terme préjudiciable pour la qualité globale de la traduction. En effet, nous savons que l'activité traduisante n'est en aucun cas un simple transcodage de mots ou de segments d'une langue à l'autre, mais repose au contraire sur une indispensable contextualisation. La réintégration d'éléments tronqués traduits par le biais de la TA ne peut en aucun cas constituer une solution satisfaisante.

2.1.2 Traduction du document bilingue

Tel que spécifié précédemment, le logiciel de TA que nous avons exploité permet de ne pas traduire les passages apparaissant déjà en langue cible, en l'occurrence en anglais, après le passage du logiciel de TAO, grâce aux fonctionnalités « Détection d'une phrase en langue étrangère » et « ne pas traduire ». Toutefois, le logiciel de TA retouche systématiquement les mots présentant une similarité morphosyntaxique entre le français et l'anglais. Ainsi, le terme français *rapport* traduit correctement par l'anglais *report* par le biais de la TAO est systématiquement retraduit par le logiciel de TA par l'équivalent *carrying forward*.

Ni la modification de la police des passages concernés (sur les conseils du support technique), ni l'enrichissement du logiciel avec la mémoire de TAO fourni par l'entreprise ne permettent de régler le problème, générant de fait un temps de relecture trop importante.

2.1.3 Traduction au fil de l'eau

La dernière solution éprouvée a consisté à utiliser les deux logiciels en parallèle : l'outil de TAO constituant l'interface principale du traducteur et la TA, comme logiciel complémentaire, pour les besoins ponctuels de traduction automatique. Néanmoins, les incessants allers-retours entre les deux outils associés à la lenteur de l'ordinateur générée par cette double utilisation font de cette alternative une solution d'appoint.

Au regard de ces trois méthodologies, il apparaît clairement que l'utilisation combinée de deux logiciels distincts de TA et TAO ne constitue pas une solution viable, ni en termes d'efficacité, ni d'un point de vue qualitatif.

2.2 Recours à un outil composite

Face à l'échec des approches précédentes, la dernière méthodologie employée a consisté à recourir entièrement à l'outil de TA que nous avons à notre disposition, en activant le module de TAO disponible. Les ressources fournies par le commanditaire (mémoire de traduction et glossaire client) ont alimenté le système, outre l'utilisation du dictionnaire financier pré-intégré dans le logiciel.

Dans la mesure où près de 80 % du texte à retraduire donne lieu à des correspondances partielles ou totales, il nous semblait inconcevable de ne pas tirer parti de ces segments pré-

traduits. Comme présenté dans le Tableau 4, 56 % du texte ressort intégralement de la mémoire de traduction. C'est donc plus de la moitié du texte qui présente un niveau de qualité maximum dès l'étape de traduction (la mémoire de traduction est réalisée après la validation de la traduction de l'an passé par le client) tandis que les 22 % du texte présentant un taux de similitude allant de 70 % à 99 % ne requièrent que des retouches mineures, qui peuvent être effectuées par un non-spécialiste (Fig. 5). Cela signifie que seulement 1/5ème du texte nécessite une traduction par un spécialiste du domaine ou, en l'occurrence, peut être soumis à un processus de traduction automatique.

Segment français 2008 document de référence	Segment français 2009 à traduire	Segment anglais 2009 traduit avec	Pourcentage de reconnaissance	Remarques
Plan d'attributions gratuites d'actions	Plan d'attributions gratuites d'actions	Free share allocation plan	71	Problème de reconnaissance des balises de titre
Accompagner la mise en œuvre des démarches qualités et risques	Accompagnement de la mise en œuvre des démarches qualités et risques	Support the implementation of quality and risk initiatives	75	syntaxe légèrement différente
Ces actions seraient définitivement acquises au terme d'une période d'acquisition de deux ans sous réserve du respect d'une condition de présence à la date d'acquisition et de l'atteinte de	Ces actions seraient définitivement acquises au terme d'une période d'acquisition de deux ans sous réserve du respect d'une condition de présence à la date d'acquisition et de l'atteinte de critères de performance	These shares would be acquired definitively at the end of a two-year vesting period, subject to compliance with a condition of presence on the acquisition date and the	80	Il suffit, dans la version de 2009, de supprimer la fin du segment rendu par la mémoire. En effet, mise à part la fin, les deux phrases sont identiques.
A la date du présent document, Générale de Santé emploie plus de 21 420 salariés (E.T.P) étant précisé que ce chiffre ne prend pas en compte les médecins qui exercent leur profession à titre libéral et en	A la date du présent document, Générale de Santé emploie près de 21.000 salariés (E.T.P) étant précisé que ce chiffre ne prend pas en compte les médecins qui exercent leur profession à titre libéral au sein des établissements du Groupe (Cf. paragraphe 6.1.6)	On the date of this document, Générale de Santé employs more than 21,000 employees (full-time equivalent), it being specified that this figure does not include those medical doctors who have	81	adverbe différent, lié à un contexte différent.
Dans ce sens Générale de Santé a déjà mis en œuvre en faveur des femmes enceintes une réduction de la durée du travail dès le 2ème	Dans ce sens Générale de Santé a déjà mis en œuvre en faveur des femmes enceintes une réduction de la durée du travail dès le deuxième mois de grossesse	In that context, Générale de Santé has already implemented, for pregnant women, a working time reduction system as of the	83	différences orthographiques
Au 31 mars 2009, les actionnaires de la Société détenant plus de 3% du capital de la Société composé de 56.229.854 actions sont les	Au 31 mars 2010, les actionnaires de la Société détenant plus de 3% du capital de la Société composé de 56.314.182 actions sont les suivants :	On 31 March 2010, the Company shareholders holding over 3% of the Company's capital stock, composed of 56,314,182	98	les dates et les valeurs changent ; les transposables permettent de les modifier facilement.

Figure 5 : Exemple de segments entre 70 et 99 % de similitude

À l'instar du logiciel de TAO, il est possible d'ajuster dans l'outil de TA les paramètres de la mémoire de traduction pour qu'elle soit plus ou moins souple, c'est-à-dire qu'elle soit en mesure de réexploiter avec plus ou moins de latitude les segments enregistrés dans la mémoire. Toutefois, contrairement au système de traduction assistée qui peut être paramétré avec une grande précision grâce à l'utilisation des pourcentages, le système de traduction automatique présente une lacune importante puisque la mémoire peut seulement être paramétrée de trois manières, à savoir *stricte*, *normale* et *flexible*. Par défaut, la mémoire est placée dans une configuration *normale*. On notera l'impossibilité d'utiliser la mémoire souple en raison d'un message d'erreur récurrent ainsi que l'absence d'informations à ce sujet, car le réglage de la mémoire n'est expliqué ni dans le logiciel, ni dans l'aide, ni même sur les supports disponibles en ligne.

Première constatation issue de l'analyse : le logiciel de TA se révèle incapable de reconnaître les segments à similitude partielle, ce qui constitue une faiblesse majeure étant donné le pourcentage de segments concernés. Dans un deuxième temps, nous avons observé que, sur les trois extraits analysés (978 segments), le logiciel est seulement en mesure de suggérer au traducteur 303 segments à 100 % sur les 535 proposés par le logiciel de TAO :

Statistiques	Nb	%
Fichiers	1	
Paragrapes	697	
Phrases	845	
Phrases uniques	774	
Complexité moyenne d'une phrase	7	
Ambiguïté moyenne d'une phrase	13	
Mots	12421	
Mots uniques	3030	
Caractères	79596	
Mots inconnus uniques	74	
Mots inconnus	96	
Mots dans les mots inconnus	104	0,84%
Caractères dans les mots inconnus	648	0,81%
Concordances du dictionnaire utilisateur par entrée unique	22	
Concordances du dictionnaire utilisateur	116	
Couverture de dictionnaire utilisateur par signes	116	0,93%
Couverture de dictionnaire utilisateur par caractères	1176	1,48%
Concordances de la mémoire de traduction par entrée unique	271	
Couverture de mémoire de traduction par phrases	281	33,25%
Couverture de mémoire de traduction par mots	5269	42,42%
Couverture de mémoire de traduction par caractères	34732	43,64%

Figure 6 : Analyse statistique de l'extrait 1

Reportés en pourcentage, ces chiffres nous montrent que le système n'effectue un travail de TAO que sur 31 % des extraits en moyenne (33 % pour l'extrait 1, 36 % pour l'extrait 2, 26 % pour l'extrait 3), tandis que la TAO couvrirait plus de 78 % des extraits (56 % de segments à 100 % et 22 % de segments entre 70 et 99 % de similitude). Les segments non reconnus sont systématiquement retraduits par le logiciel de TA, ce qui a une conséquence non négligeable sur la qualité de la traduction produite. Il importe de préciser ici que l'opération a été menée à plusieurs reprises pour s'assurer qu'il ne s'agissait pas d'un dysfonctionnement passager.

2.3 Bilan post-traduction

Le bilan des différentes méthodologies est relativement contrasté. S'il est évident que l'utilisation d'outils distincts ne constitue en aucun cas une solution envisageable pour le commanditaire, le défaut de comptabilité de la mémoire de traduction au sein de l'outil de TA représente également une gageure.

Depuis une perspective chronologique, l'utilisation du logiciel de TA semble néanmoins gagnante dans la mesure où il faut 15 minutes pour traduire l'ensemble des extraits contre environ 30 heures, soit 4,5 jours homme pour les segments compris entre 70 % et 100 % de similitude (traduisibles par un non-spécialiste) auxquels s'ajoutent environ 72 heures pour les segments nouveaux ou à faible degré de similitude (0 - 69 % de similitude, traduisibles uniquement par un traducteur professionnel spécialiste du domaine), lorsqu'on fait uniquement appel à la TAO. Ce calcul a été effectué en prenant pour base la moyenne suivante : un traducteur professionnel traduit entre 1 500 et 2 000 mots par jour, soit 250 mots par heure. Soit au total une centaine d'heures pour le document entier (segments nouveaux et segments déjà présents dans la mémoire). Notre analyse de traduction n'ayant porté que sur 50 % du document, nous avons multiplié toutes les durées obtenues par deux pour obtenir une moyenne sur le document entier.

Il s'agit à présent de contrôler la qualité des segments traduits par le biais de la traduction automatique et d'évaluer le temps de post-édition nécessaire à leur correction.

3. Analyse qualitative

Type d'erreur	Fréquence (%)*
Erreurs de terminologie	
Nom propre mal traduit	5,89
Terminologie générale	20,23
Terminologie financière	33,72
Terminologie médicale	8,80
Erreurs de langue	
Style/Syntaxe	27,86
Grammaire	23,17
Erreurs de sens	
Contresens	19,35
Non-sens	27,27
Dysfonctionnements du logiciel	
Mauvaise distinction du segment (liée aux balises...)	0,59
Non-reconnaissance des chiffres	0,88
Mot non traduit	16,42
Passage non traduit	0,59

Figure 7 : Typologie d'erreurs

3.1 Typologie d'erreurs

Document de référence 2009 à traduire	Traduction automatique	Document de référence traduit et validé
<i>Nombre d'actions pouvant être souscrites</i>	<i>Number of actions which can be subscribed</i>	<i>Number of shares that may be subscribed</i>
Document de référence 2009 à traduire	Traduction automatique	Document de référence traduit et validé par le client
<i>Le Groupe a recours à la sous-traitance dans le domaine de la restauration et du nettoyage pour respectivement 70% et 39% des établissements</i>	<i>The Group resorts to subcontracting in the field of the restoration and cleaning for respectively 70% and 39% of the establishments.</i>	<i>The Group uses subcontractors in the area of catering and cleaning for respectively 70% and 39% of facilities.</i>

Figure 8 : Exemples d'erreurs à caractère terminologique

L'analyse qualitative a porté sur 341 segments traduits par le logiciel de traduction automatique. Chaque segment cible a été analysé, évalué et comparé à sa version corrigée et validée, ce qui a permis d'établir une typologie d'erreurs, classées par nature, puis ramenées à un pourcentage. La fréquence de chaque erreur, exprimée en pourcentage, est calculée en divisant le nombre de segments affectés par l'erreur par le nombre de segments traduits en TA :

Document de référence 2009 à traduire	Traduction avec SYSTRAN	Document de référence traduit et validé par le client
<p>Chaque année <i>Générale de Santé</i> met en œuvre une politique d'encouragement au développement de la <i>Validation des Acquis de l'Expérience</i> avec des <i>résultats encourageant</i> puisque <i>Générale de Santé</i>, à <i>elle seule</i>, représente près de 50% de la totalité des VAE <i>aides -soignants</i> du secteur privé correspondant.</p>	<p>Each year implements a policy of encouragement at the development of <i>Assets of the Experiment with earnings encouraging since</i>, with it <i>only</i>, represents nearly 50% of the totality of the VAE <i>assistances - looking after</i> corresponding private sector.</p>	<p>Every year, <i>Générale de Santé</i> implements a policy to encourage the development of the <i>Validation of Experience</i> with <i>encouraging results</i> since <i>Générale de Santé alone</i> represents nearly 50% of the <i>Health Aide</i> VAEs in the corresponding private sector.</p>

Figure 9 : Exemple d'erreurs de langue/syntaxe

Les erreurs de terminologie figurent parmi les erreurs les plus récurrentes, et ce, malgré l'intégration au sein du logiciel d'un glossaire client. L'absence de terminologie spécialisée conduit le logiciel à traduire un certain nombre de termes par leur acception courante, menant à un nombre important de contre-sens :

Malgré l'association des approches linguistique et statistique, de nombreuses erreurs de langues sont également commises par le logiciel, principalement en raison d'une traduction « mot à mot », génératrice d'erreurs syntaxiques et sémantiques, ainsi qu'illustré dans l'exemple suivant (Fig. 9). La longueur et la complexité des phrases propres au domaine financier tendent à perturber considérablement le fonctionnement de l'outil. On notera également l'omission du sujet de la phrase par le moteur :

3.2 Post-édition

Après ce travail d'analyse typologique des erreurs commises par le logiciel de TA, l'ensemble des segments analysés ont été classés en quatre catégories, afin de rendre compte du niveau de gravité de ces erreurs et ainsi être en mesure d'estimer le travail de relecture et de correction que la TA nécessite sur ce type de document :

- Les segments bons/passables : ils peuvent être conservés tels quels. Bien qu'écrits dans un style imparfait, ils sont compréhensibles.

Document de référence 2009 à traduire	Traduction avec SYSTRAN	Document de référence traduit et validé par le client
<i>Le plan de formation 2010 constitue un des enjeux stratégiques du Groupe inscrit dans la poursuite des actions pluriannuelles mises en œuvre</i>	<i>The training program 2010 constitutes one of the strategic issues of the Group registered in the continuation of the multiannual actions put in work.</i>	<i>The 2010 training plan is a strategic stake for the Group and falls within the pursuit of the multi-annual actions implemented.</i>

Figure 10 : Exemple de segments bons/passables

- Les segments à corriger : ils sont pour la plupart écrits dans un style moyen, mais compréhensible ; en revanche, ils contiennent notamment des non-sens, des erreurs de terminologie et des omissions que le relecteur doit corriger.

Document de référence 2009 à traduire	Traduction avec SYSTRAN	Document de référence traduit et validé par le client
<i>les options de souscription et/ou d'achat d'actions de la Société seront consenties ;</i>	<i>the options of subscription and/or purchase of actions of will be authorized;</i>	<i>Company share subscription and/or purchase options will be granted;</i>
<i>Cette conduite doit amener le secteur à définir une rénovation de la convention collective visant la convergence salariale avec le secteur public, un déroulement de carrière motivant pour ses personnels, une politique sociale active vis-à-vis de l'emploi des personnes handicapées et des séniors, une prise en compte de la pénibilité au travail plus globale.</i>	<i>This control must lead the sector to define a restoration of the collective agreement aiming at wage convergence with the public sector, a course of career justifying for its staffs, an active social policy with respect to the employment of the handicapped people and the seniors, a taking into account of more total painfulness to work.</i>	<i>This policy should cause the sector to determine a renovation of the collective agreement regarding wage convergence with the public sector, a motivating career path for its personnel, an active social policy vis-à-vis the employment of disabled and senior individuals, and more consideration for the overall difficulties of the work at stake.</i>

Figure 11 : Exemple de segments à corriger

- Les segments à retraduire : ces segments sont incompréhensibles ; le relecteur devra se référer au texte source et retraduire ces segments.

Document de référence 2009 à traduire	Traduction avec SYSTRAN	Document de référence traduit et validé par le client
<i>un effet de glissement vieillesse technicité (« G.V.T. ») de +1% par an compte tenu de la structure conventionnelle des salaires ;</i>	<i>an effect of slip old age technicality ("G.V.T. ") of +1% a year taking into account the conventional structure of the wages;</i>	<i>The effect of automatic seniority increases by age (known by the French acronym GVT) of about +1% per annum in terms of the conventional salary structure;</i>
<i>Un plan d'épargne d'entreprise de groupe a été mis en place par la Société le 15 mai 2001 soit au moment de son introduction en bourse, et fin 2006, un avenant au PEG a été signé permettant d'y intégrer deux nouveaux FCPE (Lever GDS et GDS Italie) au travers desquels les salariés ont pu participer à l'augmentation du capital social qui leur était réservée ouverte le 4 décembre 2006 et réalisée le 9 février 2007</i>	<i>A plan of savings of company of group was set up by on May 15th, 2001 is at the time of its going public, and at the end of 2006, an endorsement with the PEG was signed making it possible to integrate into it two new FCPE (Lever GDS and GDS Italy) into through which the employees could take part in the increase in the share capital which to them was reserved open on December 4th, 2006 and was carried out on February 9th, 2007.</i>	<i>A group company savings scheme was set up by the Company on May 15, 2001, when the company was floated on the stock exchange and, at the end of 2006, an addendum to the group savings scheme was signed allowing the integration of two new FCPEs (mutual funds, GDS and GDS Italie) through which the employees were able to participate in the capital increase reserved for them, opened on December 4, 2006 and carried out on February 9, 2007.</i>

Figure 12 : Exemple de segments à retraduire

- Les segments à risque d'erreurs : ces segments ne sont pas systématiquement mal rédigés ; cependant, ils sont susceptible de renfermer des contresens auxquels le lecteur doit rester vigilant.

Document de référence 2009 à traduire	Traduction avec SYSTRAN	Document de référence traduit et validé par le client
<i>la présentation de projets collectifs évalués par un jury composé du comité de Direction de Générale de Santé et d'HEC</i>	<i>the presentation of collective projects evaluated by a jury made up of the steering committee of and HEC.</i>	<i>the presentation of collective projects assessed by a jury made up of the Board of Directors of Générale de Santé and HEC.</i>
<i>Le Groupe est présent dans la conduite de la politique sociale de la Branche de l'Hospitalisation Privée</i>	<i>The Group is present in the control of the social policy of Private Hospitalization.</i>	<i>The Group is a player in the social policy of the Private Hospitalization Sector.</i>

Figure 13 : Exemple de segments à risque d'erreurs

Ainsi, nous obtenons la répartition suivante :

	En part de la TA (%)	En part des trois extraits, représentatifs du document total (%)
Segments ressortis de la mémoire	-	31
Segments bons/passables	16,3	11
Segments à corriger	30,5	21
Segments à réécrire/retraduire	38,6	27
Segments à risques d'erreurs	14,5	10

Figure 14 : Répartition des segments en fonction de leur degré de gravité

Nous pouvons, à partir de cette analyse, faire un certain nombre de remarques sur le travail de relecture nécessaire pour un document traduit à l'aide d'un outil de TA générique :

- 31 % du texte présente un niveau de qualité élevé, puisque issu de la mémoire de traduction. En revanche, aucun segment traduit en TA (soit 69 % du texte) n'est de qualité équivalente à celle attendue par le client. Nous rappelons que le logiciel de TAO conférait à 78 % du texte une qualité élevée dès l'étape de traduction.

- Le relecteur est contraint de se référer systématiquement au texte source : puisque 10 % du texte présente un risque d'erreurs graves (contresens), le relecteur doit se montrer attentif en permanence.
- Le relecteur doit idéalement être un spécialiste du domaine afin d'être en mesure de repérer les erreurs de terminologie et les non-sens ou contresens commis par le logiciel et les corriger.
- Un travail de traduction est nécessaire sur 27 % du texte. Avec la TAO, le traducteur ne devait traduire « manuellement » que 20 % du texte.

Enfin, si on estime qu'un relecteur corrige 1 500 mots par heure pour une traduction de bonne qualité (réalisée manuellement par un traducteur spécialisé avec le recours à la TAO), il corrigerait, pour un document de référence traduit de manière automatique :

Type de segment	Temps de relecture estimé
Segments ressortis de la mémoire	1000 mots à l'heure
Segments bons/passables	750 mots à l'heure
Segments à corriger	750 mots à l'heure
Segments à risque	750 mots à l'heure
Segments à retraduire	250 mots à l'heure

Figure 15 : Répartition des segments en fonction de leur degré de gravité

Par conséquent, le travail du relecteur est plus complexe (en raison de la démultiplication et complexité des erreurs de traduction) et pénible (lecture difficile) tandis que sa vigilance doit être constante sur l'ensemble du texte.

3.3 Comparaison des méthodes

Nous constatons par conséquent que le recours à la traduction automatique présente de nombreux inconvénients : outre une prise en main fastidieuse du logiciel, près de la moitié des segments à 100 % en mémoire ne sont pas réexploitables, rendant l'activité de traduction et de relecture laborieuse et chronophage.

Pour clore notre étude, nous avons finalement procédé à une dernière analyse, en matière de coûts et d'organisation du travail, pour évaluer la rentabilité du logiciel.

Étape	Durée (heures)	Coût total (€)*
Traduction	102	2 599
Relecture	64	2 223
Total	166	4 822

Figure 16 : Estimation des coûts engendrés par la TAO

Le temps total de traduction avec le logiciel de TAO du document de référence a été estimé à 102 heures, soit près de 5 jours de travail par traducteur (au nombre de trois) selon les propres chiffres donnés par le commanditaire. Il en va de même pour le temps de relecture globale qui a été évaluée à 64 heures, ou 9 jours, au rythme de 1500 mots relus par heure ; les trois relecteurs dédiés ont donc dû travailler chacun environ 3 jours sur ce document. Nous obtenons les données suivantes :

S'agissant des coûts liés à la traduction automatique, on estime à 185 heures le temps nécessaire à une relecture attentive du document de référence traduit par la TA :

Nature des segments	Part du document	Nombre de mots	Mots relus à l'heure*	Temps de relecture (h)
TAO	31 %	29 628	1 000	29,6
TA	69 %	95 266	-	185
Bons	11 %	10 670	750	14
A corriger	21 %	20 006	750	27
A risque	10 %	9 527	750	13
A retraduire	27 %	25 341	250	101
Total	100 %	95 266	-	185

Figure 17 : Estimation du temps de post-édition requis

Ce qui nous donne en termes de coûts :

Étape	Durée (heures)	Coût total (€)
Traduction	0,2	0
Relecture/réécriture*	185	6 475
Relecture	64	2 223
Total	249	8 698

Figure 18 : Estimation des coûts engendrés par la TAO

* Nous avons pris pour base un tarif horaire de 35 € ; cependant, le travail étant plus complexe et nécessitant un fort degré de spécialisation de la part du relecteur, ce tarif pourrait être revu à la hausse.

Après avoir réalisé cette analyse sur trois extraits du document de référence de la Générale de Santé, force est de constater que la Traduction automatique ne répond pas à nos attentes : la qualité de la traduction laisse à désirer (incohérences terminologiques, contresens, omissions, etc.) et sa fonction de TAO est peu concluante. Alors qu'avec la TAO, nous parvenions à une qualité élevée (avant relecture) sur 78 % du texte et une qualité moindre (parce que traduit

manuellement) sur 22 % du texte, les chiffres s'inversent en traduction, avec 31 % du texte traduit en TAO et 69 % traduit en TA.

Enfin, en termes de temps et de coûts, il est manifeste que la TA engendre un travail plus fastidieux et donc plus coûteux que la traduction humaine associée à la traduction assistée par ordinateur.

4. Conclusion

Au terme de notre analyse, il semble évident que l'association de deux outils distincts ou l'utilisation de l'outil de TA, malgré la présence d'un module de TAO, ne constitue pas une solution adéquate pour notre commanditaire puisque la post-édition démultiplie le temps de relecture et donc les coûts, tout en plaçant le traducteur/relecteur dans une démarche traductive contraignante, voire peu satisfaisante intellectuellement parlant. Comme indiqué au début de cette communication, l'absence de possibilité d'entraîner l'outil de TA nous porte à relativiser quelque peu les conclusions de notre analyse. Il apparaît néanmoins indispensable d'étoffer considérablement les glossaires destinés à alimenter la machine dans la mesure où la terminologie constitue une source importante d'erreurs de traduction. Nous souhaitons également rappeler que le logiciel de TA testé appartient à la famille des outils génériques. Nous ne prétendons pas à la généralisation de nos résultats, notamment à l'aune des outils de TA spécifiques (c'est-à-dire développés par des laboratoires/entreprises de recherche et adaptables aux besoins des utilisateurs). De même, ces conclusions s'appliquent en l'état au domaine financier et ne sauraient être étendues à d'autres domaines.

Enfin, une autre perspective plus prometteuse semble s'ouvrir aux utilisateurs. En effet, en fin de projet, nous avons l'occasion d'éprouver brièvement la nouvelle version du logiciel de TAO testé, qui contient désormais un module de traduction automatique. Étant donné le niveau élevé de performance de la fonctionnalité de TAO de cet outil, il semble logique de s'appuyer sur ce logiciel comme environnement de travail privilégié afin d'exploiter pleinement les concordances totales et partielles.

Outre une interface simplifiée et intuitive, cette nouvelle version offre la possibilité de s'appuyer sur trois moteurs de traduction automatique distincts. Mais avant tout, elle permet de réaliser une traduction du document au fil de l'eau, c'est-à-dire segment par segment, qu'il s'agisse de TAO ou de TA. Le traducteur reste donc maître de son travail : il prend ou rejette les segments de traduction automatique proposés par le logiciel, au même titre que ceux récupérés par le biais de la TAO. La phase de relecture de la TA se confond dès lors avec la phase de traduction, permettant une traduction simultanée TAO/TA par un spécialiste du domaine, puis une relecture normale. En d'autres termes, le logiciel fournit au traducteur un environnement de travail beaucoup plus « naturel » dans la mesure où la traduction automatique « se fond » dans l'activité globale de traduction humaine. Et si l'on reprend notre problématique initiale, la traduction automatique se présente dès lors comme un outil d'aide à la traduction à part entière, participant de fait à une vision élargie de la traduction assistée par ordinateur.

En dernier lieu, une rapide estimation du temps de relecture et donc des coûts générés par cet outil présage un rapport temps/coût avantageux pour l'entreprise commanditaire :

Nature des segments	Part du document	Nombre de mots	Mots traduits par heure	Temps de traduction (h)	Mots relus à l'heure	Temps de relecture
TAO	80 %	72 582	2 400	30	1 500	48 h
Segments à 100 %	58 %	54 351	-	-	1 500	-
Segments entre 70 et 99 %	22 %	20 353	-	-	1 500	-
TA	20 %	22 684	-	48	1 500	15 h
Bons	3 %	2 858	750	4	1 500	
A corriger	6 %	5 716	750	8	1 500	
A risque	2 %	1 905	750	3	1 500	
A retraduire	9 %	8 574	250	34	1 500	
Total	100 %	95 266	-	78		64 h

Figure 18 : Estimation du temps de relecture (nouvelle version du logiciel de TAO)

Étape	Durée (heures)	Coût (€)
Traduction	78	2 476*
Relecture	64	2 223
Total	142	4 699

Figure 19 : Estimation des coûts (nouvelle version du logiciel de TAO)

Dans tous les cas, il a été démontré que, pour un document de ce type, au style complexe et à la terminologie précise, très récurrent d'une année sur l'autre et qui nécessite un niveau de qualité élevé, il faut privilégier le recours à la TAO. Quant à la traduction automatique, elle peut induire pour le traducteur un certain gain d'efficacité sur les phases de traduction « manuelle » à condition d'être perçue comme un outil d'aide et non un outil destiné à remplacer le traducteur humain.

Bibliographie

Autorité des Marchés Financiers (2011), « Présentation de l'AMF », In Les Missions de l'AMF », http://www.amf-france.org/affiche_page.asp?urldoc=lesmissionsamf.htm&lang=fr&Id_Tab=0 [en ligne] dernière consultation le 26.06.2011.

Bellamy, Muriel, Chesné, Marion, Lehmann, Camille et Audrey Prost (2011), « Projet de Recherche Appliquée en Traduction Automatique », Mémoire de recherche de Master 2, sous la direction de Sandrine Peraldi, soutenu en juin 2011.

Drothier, Yves (2008), « Au cœur de la traduction automatique de Systran », In Journal du Net, <http://www.journaldunet.com/solutions/intranet-extranet/reportage/au-coeur-de-la-traduction-automatique-de-systran/l-analyse-statistique-renforce-la-qualite-des-traductions.shtml> [en ligne], dernière consultation le 26.06.2011.

Goudet, Jean-Luc (2008), « Traduction automatique : les années où tout a changé », In Futura-Techno (FuturaSciences), http://www.futura-sciences.com/fr/doc/t/informatique-2/d/traduction-automatique-les-annees-ou-tout-a-change_831/c3/221/p3/ [en ligne], dernière consultation le 27.06.2011.

Office de la Langue française (2012), « Grand dictionnaire terminologique », <http://gdt.oqlf.gouv.qc.ca/> [en ligne], Gouvernement du Québec.

Microsoft (2011), « Informations sur la traduction automatique », In Aide et Support Microsoft, <http://support.microsoft.com/gp/mtdetails> [en ligne], dernière consultation le 27.06.2011.

Peraldi, Sandrine (2010), « Traduction assistée par ordinateur : entre théorie et pratique. » In Actes des journées d'études de Strasbourg 18 et 19 septembre 2008, Les Cahiers du GEPE, N° 2/2010. Outils de traduction - outils du traducteur ? (<http://www.cahiersdugepe.fr/index1771.php>)

Sauron, Véronique (2004), « Wordfast™ et Trados™ ou le combat de David contre Goliath », Hieronymus, Bulletin trimestriel de l'Association suisse des traducteurs, terminologues et interprètes (ASTTI), 1/2004.

Sauron, Véronique (2007), « Les outils d'aide à la traduction : ennemis ou alliés du traducteur ? » in I Saperi dele Tradurre : analogie, affinità, confronti, a cura di Clara Montella et Giancarlo Marchesini, ed. Franco Angeli, Milan 2007, pp199-226

SDL (2011), « Traduction automatique » in SDL Translation Zone, <http://www.translationzone.com/fr/translator-solutions/automated-translation> [en ligne], dernière consultation le 26.06.2011.