

# La traduction automatique : traduction machine?

Natalie Kübler  
CLILLAC EA 3967  
Université Paris-Diderot (Paris 7)  
[nkubler@eila.jussieu.fr](mailto:nkubler@eila.jussieu.fr)

**Résumé :** Cette contribution vise à situer le rôle de la traduction automatique (TA) parmi les outils d'aide à la traduction et évoque les besoins qu'elle a suscités dans le monde de la communication multilingue. Elle s'attache tout d'abord à replacer la TA dans son contexte historique en montrant ses apports dans le domaine du traitement automatique du langage. Se pose ensuite la question des besoins et des outils adaptés aux besoins. En effet, la TA peut répondre à des besoins spécifiques et bien cernés. Ceux-ci sont généralement répartis en quatre types : la diffusion, l'assimilation (ou veille), les échanges et l'accès de type base de données. La différenciation de ces besoins conduit à utiliser la TA de différentes manières. Des exemples concrets en démontrent les potentialités, mais aussi les limites. Ces exemples permettent aussi d'illustrer les raisons pour lesquelles la TA est plus efficace dans les domaines scientifiques et techniques que dans les domaines juridiques ou économiques. La TA peut par ailleurs répondre à de nouveaux besoins sans prendre la place du traducteur. Au contraire, cet outil est là pour s'intégrer dans un continuum entre traduction purement humaine et traduction machine, la machine restant au service de l'être humain.

**Mots clefs :** traduction automatique, traduction humaine, traduction spécialisée, besoins linguistiques, communication multilingue

*Abstract : This paper aims at analysing machine translation (MT) in the context of CAT-tools for translators, and tries to tackle the different needs it has generated in the world of multilingual information interchange. The history of MT has shown how advances in natural language processing are due to MT developments. Today, varied translation needs can be met using appropriate tools. Those needs have been split into four different types: Dissemination, assimilation, interchange, data base access. Taking those different needs into account leads to using MT in different ways. Concrete examples show the potentialities, but also the limits of MT, and help understand why it is more efficient in scientific and technical domains, than in economics or law. MT translation is there to answer new needs without replacing human translators. In fact, it is a tool that can take place in a continuum between human and machine translation, and stay a useful tool for translators.*

*Keywords: machine translation, human translation, specialised translation, linguistic needs, multilingual information interchange*

**Natalie Kübler** est linguiste, professeur à l'Université Paris-Diderot (Paris 7). Elle a travaillé dans le domaine de la correction automatique d'erreurs en anglais pour francophones, puis s'est tournée vers les langues de spécialité, la linguistique de corpus et la traduction spécialisée. Elle enseigne la recherche en linguistique de corpus et ses applications à la traduction et à la terminologie. Par ailleurs, elle donne des cours de traduction automatique aux traducteurs depuis 8 ans.

## **Introduction**

Notre but n'est pas de faire ici l'apologie de la traduction automatique, mais plutôt de la situer dans le domaine de la traduction parmi les outils d'aide à la traduction, et d'évoquer les besoins qu'elle a créés dans le monde du partage multilingue de l'information. La première partie de cet article replace la traduction automatique dans le contexte historique qui a permis son développement actuel. Nous la situerons ensuite en tant qu'outil pour aborder finalement ses applications actuelles. Cela nous amènera à poser des questions en vue d'aboutir à une conclusion provisoire, car le domaine des outils d'aide à la traduction évolue constamment.

## **1. Historique**

La traduction automatique (TA) est le domaine de recherche à la base du traitement automatique du langage. L'idée moderne en remonte aux années 1930 et on divise généralement son développement en quatre grandes périodes : les années 50-60, le rapport ALPAC (1966), et ses conséquences, la reprise des années 70-80 et la période actuelle qui commence au début des années 90.

### **1.1 Les années 50-60**

Durant la deuxième guerre mondiale les premiers ordinateurs permettent aux belligérants de déchiffrer leurs codes respectifs. On tente ensuite d'appliquer ces techniques de déchiffrement à la traduction automatique. L'expert mondial de la traduction automatique, John Hutchins (2001), évoque la première conférence sur la TA qui fut organisée en 1952 par Bar-Hillel. Dès cette époque, les chercheurs sont conscients des difficultés que présente le domaine : ils se sont en effet déjà accordés sur la conclusion qu'une traduction de bonne qualité et totalement automatique relève de la science-fiction, ou du moins, restait du domaine du virtuel. Ils se rendent compte que l'intervention humaine, déjà qualifiée de pré-édition ou de post-édition restera indispensable.

De plus, les premiers systèmes de TA partent avec de lourds handicaps : les connaissances en syntaxe et en analyse syntaxique sont encore très insuffisantes, les ordinateurs ont des capacités de stockage très limitées et sont peu puissants. On dispose avant tout de systèmes de mise en correspondance de dictionnaires, ce qui génère des résultats de traduction voués à l'échec ou alors des « systèmes-jouet », incapables de traduire plus de quelques phrases.

Le premier système de TA est présenté au public en 1954 ; il permet de traduire 49 phrases russes, soigneusement sélectionnées, vers l'anglais se servant d'un dictionnaire de 250 mots et de six règles de grammaire. Ce sont probablement des motivations stratégiques qui poussent alors les gouvernements américain et soviétique à financer ce type de recherches.

Comme nous l'avons mentionné plus haut, la recherche en linguistique théorique n'a pas encore atteint toutes les connaissances en syntaxe que l'on a aujourd'hui. Ce n'est qu'avec les découvertes de Noam Chomsky dans *Syntactic Structure* (1957) et *Aspects of the Theory of Syntax* (1962) que l'on commence à avoir des modèles syntaxiques de la langue. C'est aussi dans ces années-là que Lucien Tesnière développe une approche alternative dans ses *Éléments de syntaxe structurale* (1959). Une description syntaxique de la langue donne une structure aux phrases qui apparaissent sinon comme une simple suite de mots sans ordonnancement. La syntaxe est par conséquent indispensable au traitement automatique du langage autre que statistique. On peut donc dire que la recherche en traduction automatique des années 50 et au début des années 60 vient stimuler et appuyer toute la recherche dans le domaine du traitement automatique du langage. En effet, les systèmes de traduction automatique de différents types nécessitent tous, entre autres, un module performant et robuste d'analyse syntaxique.

## 1.2. Le rapport ALPAC et ses conséquences

Mandatée en 1966 par le gouvernement des Etats-Unis, la commission ALPAC (*Automatic Language Processing Advisory Committee*) publie un rapport qui va avoir une influence déterminante et très négative sur le soutien à la recherche en TA. Ce rapport conclut que la TA est plus lente, moins efficace et deux fois plus chère que la traduction faite par des humains. Il recommande d'arrêter de financer la recherche dans ce domaine et de se tourner plutôt vers des aides à la traduction, comme par exemple les dictionnaires automatiques. Le rapport recommande cependant de continuer à soutenir la recherche fondamentale en traitement automatique du langage. On peut dire que la recherche en TA a été alors virtuellement stoppée pour une décennie aux Etats-Unis. Le rapport ALPAC peut être consulté sous <http://www.nap.edu/books/ARC000005/html/>.

## 1967-1976

Au cours de cette période, la TA est en sommeil aux Etats-Unis, mais d'autres pays continuent à développer la recherche que stimulent des besoins concrets. Si aux Etats-Unis la seule activité porte sur la traduction du russe en anglais de textes scientifiques et techniques, au Canada, en Europe et au Japon en revanche, le développement de la TA se poursuit. Citons le projet canadien TAUM-Météo (décrit dans Isabelle 1986 par exemple) ; il débute en 1970 et produit deux résultats importants : il mène au développement ultérieur du langage de programmation PROLOG et s'utilise par ailleurs avec succès dans le domaine restreint de la traduction des prévisions météorologiques. En France, on développe à Grenoble un système permettant de traduire du russe des textes mathématiques et physiques (Boitet & Nedobejkine 1980). Dans la communauté européenne, les besoins en traduction deviennent de plus en plus importants et le système Systran est mis en place à la Commission Européenne en 1976. Il est encore utilisé aujourd'hui.

## 1.3. Les années 80

A la fin des années 70 et dans les années 80, on commence à travailler sur de gros systèmes commercialisables, et l'activité de recherche dans le domaine de la TA connaît un nouvel essor. A partir de la première version de Systran produite en 1970 par Peter Toma et qui consiste à traduire du russe vers l'anglais, on assiste au développement et à l'installation de nouvelles versions auprès de la Commission Européenne, et de l'OTAN, chez Xerox notamment. Les systèmes Logos et METAL se développent et sont les principaux concurrents commerciaux de Systran. Du côté de la recherche, on peut citer le premier système Ariane, développé au GETA à Grenoble chez Christian Boitet (Boitet & Nedobejkine 1980), le projet européen EUROTRA (Maegaard 1989 basé sur un système à transfert multilingue, le Distributed Language Translation system à Utrecht (Maxwell 1990), construit sur un système à pivot en passant par l'espéranto, et d'autres encore qui utilisent l'intelligence artificielle.

## 1.4. Le développement des aides à la traduction

Dès les années 80, on commence à travailler sur des environnements d'aide à la traduction (Volanschi, ce volume) intégrant des dictionnaires, des concordanciers et des mémoires de traduction. Certaines entreprises qui construisent des systèmes de TA les abandonnent (ALPSystems aux Etats-Unis) et/ou développent des mémoires de traduction (SITE à Paris). Dans certains systèmes, on commence à travailler sur le couplage de ces outils avec la traduction automatique. Les mémoires de traduction connaissent alors un développement intéressant et prennent une place prépondérante parmi les outils d'aide à la traduction.

## 1.5. Des années 90 à aujourd'hui

A partir des années 90, ce sont les mémoires de traduction qui commencent à dominer le marché. Aujourd'hui ces outils font partie de la panoplie indispensable du traducteur professionnel. Avec la montée en puissance des ordinateurs personnels, les systèmes commerciaux de TA commencent à produire des systèmes individuels que l'on peut installer sur PC. Le développement d'Internet permet de mettre en ligne des systèmes de TA. Systran y a recours le premier en 1998 avec Altavista/Babelfish<sup>1</sup>.

Malgré l'intérêt croissant porté aux outils d'aide à la traduction, la recherche en TA ne cesse pourtant pas et certains de ses résultats sont prometteurs. Les systèmes les plus efficaces se basaient jusqu'alors sur des systèmes à transfert, à savoir, un module d'analyse de la langue source, un module de règles de transfert complexes entre la langue source et la langue cible et un module de génération de la langue cible. Ces systèmes fonctionnent avec des règles syntaxiques et de très gros dictionnaires. La quantité croissante de documents disponibles sur support électronique permet d'enrichir toujours davantage les dictionnaires et surtout de créer des dictionnaires spécialisés qui donnent des résultats de traduction appréciables dans les domaines techniques et scientifiques. La recherche est ainsi amenée à tenter de détecter le domaine d'un texte source afin d'indiquer automatiquement au système quels dictionnaires il doit activer (Lang E. & Jin Yang 1999)

On constate aussi que les systèmes mettent de plus en plus souvent l'accent sur le lexique et le lexique-grammaire ; on ajoute donc des informations syntaxiques toujours plus nombreuses aux éléments du lexique. C'est l'une des conséquences de l'évolution de la recherche en linguistique théorique et formelle, mais aussi de l'apport de la recherche en linguistique de corpus qui décrit la langue à partir des données concrètes du corpus (Scott & Tribble 2006 par exemple).

Par ailleurs, les outils de traitement statistique de corpus développés en TAL ont permis l'émergence d'une nouvelle direction de recherche : la traduction automatique « basée sur les exemples ». Cette approche permet de dégager les traductions les plus fréquentes en effectuant des analyses statistiques sur des corpus traduits et alignés. Ces systèmes ne sont pas commercialisés pour l'instant, mais ils présentent des résultats prometteurs.

Même s'ils sont encore bien loin d'imiter le traducteur humain, les systèmes de traduction automatique de textes écrits sont aujourd'hui reconnus et utilisés sur le marché de la traduction. La traduction de l'oral représente un défi passionnant pour le domaine du traitement automatique du langage ; on parle ici de traduction automatique du langage parlé ou encore d'interprétariat automatique. Le système ATR (Sumita 1999) par exemple, développé au Japon permet d'effectuer des inscriptions téléphoniques à des colloques internationaux ou des réservations d'hôtel par téléphone. Citons encore le projet allemand Verbmobil (1993-2000) dont l'objectif était de créer un traducteur oral transportable pour aider les locuteurs germanophones et japonais à mener des négociations commerciales en anglais. Enfin le projet TC-STAR<sup>2</sup> dont le but est d'effectuer la reconnaissance vocale des discours au Parlement Européen, traduit le discours et le restitue oralement dans la langue cible. Dans ce type d'outils, on combine donc la reconnaissance de la parole, la traduction automatique et la synthèse de la parole.

Un autre défi consiste à combiner deux approches en traduction automatique. Si les systèmes à transfert ont en effet montré leur efficacité, ils ont aussi révélé leurs limites. Nous l'illustrerons plus bas. Par ailleurs, les systèmes statistiques donnent des résultats prometteurs mais qui révèlent eux aussi un certain nombre de limites. La combinaison des systèmes basés sur exemples avec les systèmes à transfert représente peut-être une possibilité d'améliorer de

---

<sup>1</sup> Notons par ailleurs, que la première interface graphique d'Internet, le World Wide Web n'a été mise en place par Tim Burners-Lee qu'en 1993.

<sup>2</sup> On trouve une description complète du projet à l'adresse suivante : <http://www.tc-star.org>.

manière significative les résultats de la traduction automatique. La recherche en traduction automatique s'y attelle également.

Pour conclure, on peut donc relever que le domaine de la TA a représenté et représente un élément moteur en recherche sur le traitement automatique du langage et qu'il conduit au développement de nombreux résultats connexes. Tous les outils d'aide à la traduction que l'on utilise aujourd'hui sont issus des recherches et du développement en TA. Le développement des analyseurs syntaxiques ou de logiciels de génération automatique est lui aussi étroitement lié aux besoins de la recherche en traduction automatique.

## **2. La traduction automatique comme outil automatique ?**

Pour aborder la question de la nature de la traduction automatique, on doit poser la question de son utilité et de son utilisation. En effet, si la traduction automatique est « machinale » est-elle vraiment utile ?

Les systèmes de TA s'intègrent dans un ensemble plus vaste d'outils d'aide à la traduction. Par conséquent, on doit considérer que la TA est un outil d'aide à la traduction presque au même titre que les dictionnaires électroniques monolingues ou multilingues, les concordanciers, les mémoires de traduction, les vérificateurs orthographiques et grammaticaux, la fonction « recherche » sous Word, etc. (Volanschi, ce volume). On peut considérer l'utilisation de ces divers outils comme étant sur un continuum allant de l'être humain à la machine. Selon que l'on sera plus ou moins proche de l'intervention humaine ou de celle de la machine, la charge de travail sera répartie plus ou moins sur l'humain ou la machine.

Se pose alors la question des besoins et des objectifs de la traduction que l'on a à réaliser. Le choix des outils serait donc fonction des besoins ? La faille dans cette logique est que les outils génèrent de nouveaux besoins. Commençons par les besoins et l'évolution du domaine de la traduction au niveau mondial qui justifient l'utilisation de l'ordinateur de manière générale. On trouve chez les experts en TA, mais aussi dans les argumentaires commerciaux des systèmes commercialisés, les explications suivantes :

- Le premier point concerne la quantité de texte à traduire. Aujourd'hui, la quantité de documents qui doivent être traduits est tellement importante et croît si rapidement qu'il n'est pas possible aux traducteurs de tous les traduire. L'ordinateur peut contribuer à augmenter la quantité de documents traduits.
- Nombreux sont les documents qui sont très techniques et assez laborieux à traduire. L'ordinateur peut aider les traducteurs dans cette tâche peut-être ingrate<sup>3</sup>.
- Les grosses entreprises ont pris conscience, depuis un certain nombre d'années maintenant, que la transmission de l'information multilingue demande une terminologie normalisée et cohérente dans toutes les langues utilisées. Les outils d'aide à la traduction proposent systématiquement les termes normalisés par l'entreprise. Cette cohérence est plus longue à réaliser avec des traducteurs humains qui travaillent en équipe et sans outils d'aide à la traduction.
- Les entreprises exigent toujours davantage de texte traduit, dans des délais de plus en plus brefs. L'ordinateur peut aider à intensifier le rythme et à augmenter le volume de texte traduit.

---

<sup>3</sup> Cet argument est cependant sujet à controverse et présente une certaine faiblesse ; en effet, il est avancé principalement par les entreprises qui développent des systèmes de TA et qui souvent connaissent mal le métier de traducteur professionnel.

- Enfin, dans certaines situations, il n'est pas toujours nécessaire d'avoir une traduction de haute qualité. La traduction automatique permet de traduire très rapidement des pages web rédigées dans une langue que l'on ne connaît pas, avec un résultat approximatif mais qui permet de comprendre globalement le contenu de la page.

Outre ces constatations sur les besoins au niveau industriel, Hutchins (2004) relève aussi des besoins à un niveau plus individuel. Il répartit donc ces besoins en quatre types qui devraient permettre de choisir le bon outil de traduction :

- La diffusion
- L'assimilation (ou la veille)
- Les échanges
- L'accès de type base de données

Cette classification reflète aussi l'influence d'un outil informatique sur la génération de nouveaux besoins. Nous verrons que le choix de la traduction automatique a des incidences sur la normalisation des documents source et donc probablement, un impact sur la rédaction technique.

## 2.1. La diffusion

La diffusion représente le niveau le plus élevé en traduction. Elle consiste à produire des traductions - généralement en quantités importantes et de haute qualité - qui seront publiées. Le niveau de qualité doit approcher la perfection : dans le domaine médical, par exemple, les instructions d'utilisation d'un instrument doivent être claires pour éviter les accidents et les litiges ; il est par ailleurs dans l'intérêt d'une entreprise et de son image de marque de publier des documents de qualité. Il est possible de choisir la TA pour la diffusion ; ce choix requiert cependant un certain nombre d'opérations autour de la traduction elle-même : la pré-édition du texte source avant la TA, la personnalisation du système de TA avant le processus et la post-édition du résultat de traduction.

### 2.1.1. Pré-édition du texte source avant la TA

Les systèmes de TA commercialisés sont des systèmes à règles. Bien que le nombre de règles soit très important et que celles-ci soient constamment corrigées en fonction des tests linguistiques effectués, certaines structures - comme les subordonnées - ou certains phénomènes - comme ceux de la dépendance à longue distance ou l'ellipse - posent problème. Par ailleurs, si le texte source contient des erreurs, le système ne pourra pas rétablir le texte correct. C'est donc à un être humain que revient la tâche de corriger le texte source et, plus radicalement, de le simplifier. La simplification est en effet en relation étroite avec le langage dit contrôlé. Le langage contrôlé est une forme de langue naturelle extrêmement prescriptive et simplificatrice au niveau de la syntaxe et du lexique. Il est largement utilisé dans la rédaction de documentation technique - par exemple, dans l'industrie aéronautique ou automobile - de manière à faciliter, entre autres, le processus de TA.

Les questions d'effacement sont difficiles à traiter par un système de TA. L'exemple suivant *cable and ADSL connexion* sera traduit par *le câble et la connexion de l'ADSL*, alors qu'il s'agit de *la connexion par câble et par ADSL*. Les recommandations d'un langage contrôlé demandent à ce que soit rétabli le terme effacé : *la connexion par câble et la connexion par ADSL*. C'est aussi ce qui est recommandé pour la pré-édition d'un texte à faire traduire par un système de TA. De même, l'ellipse du *that* en anglais doit être rétablie.

Ce type de problèmes est déjà évoqué par Austermühl (2001 :163) qui cite les recommandations faites par un système de TA commercialisé :

*Avoid idiomatic expressions*

*Avoid omitting pronouns before a verb*

*(...) Keep to standard, formal English in which grammatical connections are clearly expressed*

Ces recommandations sont encore valables aujourd'hui.

### 2.1.2. Personnalisation du système par la création d'un dictionnaire spécialisé

Un autre problème rencontré par la TA est en effet celui du lexique spécialisé (terminologie) ou tout simplement du figement des termes dans un domaine donné. L'exemple ci-dessous montre comment un système peut traduire un texte dans un domaine spécialisé, sans l'aide d'un dictionnaire machine spécialisé.

TEXTE SOURCE	SANS DICTIONNAIRE SPECIALISE
This page contains a simple <b>cookbook</b> for <b>setting up Red Hat 6.X</b> as an <b>internet gateway</b> for a <b>home network</b> or small office network.	Cette page contient un <b>cookbook</b> simple pour le <b>chapeau rouge 6X d'établissement</b> en tant que <b>Gateway d'Internet</b> pour un <b>réseau à la maison</b> ou le petit réseau de bureau.

Lorsque le même extrait est traduit par le même système dans lequel le dictionnaire spécialisé est activé, on obtient un résultat bien meilleur :

TEXTE SOURCE	AVEC DICTIONNAIRE SPECIALISE.
This page contains a simple <b>cookbook</b> for <b>setting up Red Hat 6.X</b> as an <b>internet gateway</b> for a <b>home network</b> or small office network.	Cette page contient <i>des recettes</i> simples pour <i>l'installation Red Hat 6.X</i> en tant que <i>passerelle Internet</i> pour un <i>réseau domestique</i> ou un petit <i>réseau de bureau</i> .

Il en va de même pour l'exemple suivant :

TEXTE SOURCE	SANS DICO SPECIALISE.	AVEC DICO SPECIALISE
A series of <b>wideband</b> digital data transmission formats [...]	une série de formats <b>à large bande</b> de transmission de données numériques	une série de formats de transmission de données numériques <b>à large bande</b>

Les dictionnaires spécialisés permettent aussi d'intégrer un certain nombre d'informations syntaxiques attachées aux éléments du lexique, c'est-à-dire des informations de type lexique-grammaire comme par exemple la structure prépositionnelle d'un verbe, ou le type d'arguments pouvant entrer en position d'objet direct. Dans certains domaines, les résultats de traduction sont bien meilleurs.

La création d'un dictionnaire spécialisé ne suffit cependant pas à lever toutes les ambiguïtés au niveau du lexique. Le dictionnaire spécialisé fige l'emploi de certains termes dans une certaine acception. La conséquence en est que les questions de polysémie ne sont pas traitées. La décision d'insérer un seul emploi pour un mot dans un dictionnaire spécialisé repose sur des critères statistiques. On procède en général à une analyse statistique d'un corpus du domaine pour vérifier quelle est l'acception la plus usitée ; c'est celle qui sera intégrée dans le dictionnaire spécialisé, sachant qu'il faudra l'intervention d'un être humain pour rétablir les cas où une autre acception prime dans le texte. Ainsi, si dans le domaine de la terminologie

par exemple, le mot anglais *term* sera le plus fréquemment traduit par l'emploi de *terme* en français, il se peut que dans certains cas, il corresponde à un semestre universitaire<sup>4</sup>. Les dictionnaires spécialisés ne permettent pas non plus de traiter correctement la question des acronymes anglais qui n'ont d'équivalence en français que la forme développée. Cette question semble pourtant relever d'un problème de lexique. Mais si dans le texte anglais, on trouve la forme développée suivie de l'acronyme entre parenthèses, le système traduira la forme développée correctement et répétera celle-ci dans la parenthèse ! Un traducteur humain saura soit supprimer la parenthèse, soit conserver l'acronyme anglais à titre indicatif. C'est parce que ce type de question va au-delà du lexique et relève plutôt de la pragmatique que les dictionnaires spécialisés connaissent des limites.

### 2.1.3. Post-édition

Il reste cependant des problèmes que le système n'arrive pas à traiter malgré la pré-édition et la constitution de dictionnaires personnalisés. D'où la nécessité de la post-édition.

Outre des questions comme les anaphores ou tout ce qui dépasse le cadre de la phrase, on trouve aussi des erreurs générées par les règles du système qui sont surgénéralisantes, un peu à la manière dont un enfant en cours d'acquisition de sa langue maternelle surgénéralise les règles (« j'ai pris » pour « j'ai vu » par exemple). Les problèmes d'ambiguïté générés par une analyse syntaxique erronée sont aussi nombreux. Ces problèmes sont illustrés par les exemples ci-dessous qui sont tirés d'une traduction de l'anglais vers le français.

#### Surgénéralisation des règles de transfert :

Le système des déterminants présente un grand nombre de différences d'emplois entre l'anglais et le français. Un système de TA peut donc intégrer un certain nombre de règles de transfert lui permettant de savoir comment traduire un certain type de déterminant de la langue source à la langue cible. L'exemple ci-dessous illustre les difficultés de traduction du déterminant défini entre l'anglais et le français :

Texte source	Traduction par le système	Traduction corrigée
(...)decoded by specific individuals	(...) décodée par <b>les</b> individus spécifiques.	(...)décodée par <b>des</b> individus spécifiques.

Le système possède une règle de transfert qui spécifie que le déterminant zéro en anglais doit être traduit par un déterminant défini en français, comme c'est le cas dans des affirmations généralisantes (*Life is beautiful* => *La vie est belle*). Or, dans ce cas précis, il ne s'agit pas d'une généralisation mais justement d'une situation dans laquelle on ne connaît pas encore les individus dont il est fait mention. En français, il faudra donc utiliser un déterminant indéfini. Mais la règle de transfert citée ci-dessus a été appliquée de manière surgénéralisante.

L'exemple suivant montre comment une règle de transfert sur la traduction du passif anglais par un verbe réfléchi en anglais s'applique à tort :

Texte source	Traduction par le système	Traduction corrigée
(...) which has been termed hypermedia	(...) qui s'est nommée hypermedia	(...) qu'on a appelée hypermedia

<sup>4</sup> Sans compter la polysémie de *terme* qui ne recouvre pas celle de *term* de manière bi-univoque

## Problèmes d'analyse syntaxique

Comme mentionné plus haut, les systèmes à transfert peuvent être décrits de manière simplifiée comme fonctionnant avec trois modules : un module d'analyse syntaxique de la langue source qui produit une représentation abstraite de la phrase, un module de transfert contenant des règles de traduction et un module de génération qui, à partir d'une représentation abstraite de la langue cible produite par le module de transfert, génère la phrase de la langue cible. Des erreurs peuvent donc se produire à tous les stades du processus. Nous illustrons ci-dessous des erreurs qui interviennent dès la phase d'analyse de la langue source et se propagent dans les deux autres phases pour donner un résultat de traduction complètement erroné.

Dans l'exemple ci-dessous, on peut penser à première vue que le mot *light* a été mal analysé par le système puisqu'il est traduit par un adjectif, alors qu'il s'agit du nom *light* qui se traduit par *lumière*. Cette erreur provient en fait d'une erreur d'analyse syntaxique de toute une partie de la phrase et ne porte pas seulement sur un mot. C'est tout le groupe nominal complexe *the phase of the light passing through* qui a été mal analysé. Le mot *passing* qui ici fonctionne comme la forme verbale du verbe *to pass* (en fait *to pass through*) a été analysé comme une nominalisation.

Texte source	Traduction par le système	Traduction corrigée
The liquid crystal changes the phase of the <b>light</b> passing through it	Le cristal liquide change la phase du <b>dépassement léger</b> par lui	Les cristaux liquides modifient la phase de la <b>lumière</b> qui les traverse.

De même, dans la phrase suivante, le mot *can* a été analysé comme un nom, rendant l'analyse correcte de la phrase totalement impossible.

Texte source	Traduction par le système	Traduction corrigée
If such a unit is placed between a pair of plane polariser plates then <b>light can pass</b> through it only if the correct voltage is applied.	Si une telle unité est placée entre une paire de <b>passage léger de bidon</b> de plaques plane polarisées puis par elle seulement si la tension correcte est appliquée.	Si on place cette unité entre deux plaques planes polarisées, alors la <b>lumière peut</b> les <b>traverser</b> seulement si on applique le voltage correct.

Ces deux exemples démontrent aussi que si la TA ne va pas au-delà de la phrase, elle ne va parfois même pas jusqu'à la phrase. En effet, certains systèmes de TA commercialisés effectuent une analyse de la phrase lorsque c'est possible, mais sinon, ils fonctionnent plutôt avec l'analyse « par chunks », c'est-à-dire, par morceaux de phrase (la plupart du temps en tentant de reconnaître des syntagmes plus ou moins complexes). Cela expliquerait tout à fait les erreurs d'analyse ci-dessus.

On voit ici que la production d'une traduction de haute qualité à l'aide de la TA ne peut se faire sans intervention humaine. Rappelons cependant que la révision est un processus qui s'applique aussi aux traductions effectuées par les traducteurs. On peut donc envisager d'utiliser la TA pour produire des traductions publiables, mais il est évident que ce choix est aussi dépendant du domaine dans lequel on traduit. Les dictionnaires spécialisés montrent une bien plus grande utilité dans certains domaines spécifiques par exemple, scientifiques et techniques, que dans d'autres – par exemple, dans les domaines économiques et juridiques.

Cela tient sans doute au fait que les systèmes de TA ont été surtout développés pour traduire de la documentation technique ; leurs modules d'analyse, de transfert et de génération ne contiennent donc pas les structures de la langue juridique, souvent obsolètes en langue générale ou scientifique et technique. Par ailleurs, ce n'est pas seulement le type de domaine qui doit déterminer l'emploi ou non d'un système de TA, c'est aussi le genre du texte : des documents techniques, comme des manuels d'utilisation ou des spécifications techniques rédigés en langage contrôlé, ainsi que des articles scientifiques très spécialisés et de haut niveau sont bien plus adaptés à la TA que des documents de vulgarisation scientifiques, des cours en ligne, des textes de marketing. Le travail de révision est en effet bien moins important dans le premier type de textes.

## 2.2. La veille ou l'assimilation d'information

Passons maintenant à la deuxième utilisation possible de la TA : l'assimilation, ou veille.

C'est l'un des besoins qui ont été générés par l'existence de logiciels de traduction automatique, et notamment par l'installation de ceux-ci sur le Web. Le développement du Web est bien sûr lui aussi à l'origine de ce besoin, en raison de la quantité exponentielle de documents qui sont mis en ligne dans une diversité de langues de plus en plus grande.

La fonction de veille de la TA permet par exemple de traduire un site Web afin d'en comprendre le contenu global. Si la TA n'existait pas, il serait probablement beaucoup plus difficile de faire de la veille dans la mesure où tout devrait être traduit par des traducteurs humains. Il est fort probable que les entreprises ne pourraient pas investir suffisamment pour faire traduire les sites intéressants par des traducteurs humains. Il semble donc que cette fonction de veille ne prenne pas la place d'un traducteur humain et soit un avatar de la TA. Il en va de même pour les particuliers qui encore moins qu'une entreprise, ne pourraient pas faire appel à un traducteur pour les aider à comprendre le contenu d'un site rédigé dans une langue qu'ils ne connaissent pas. A titre d'exemple, voici ce que donne la traduction automatique en français d'un extrait du site de la ville de Saarbrück en Allemagne (<http://www.saarbruecken.de/>):

### ALLEMAND

#### AKTUELL

#### **Neuer Kulturdezernent gewählt Veranstaltungshalle an die Camphauser Straße**



Rathaus St. Johann

Erik Schrader (35) aus Berlin wird **neuer Dezernent für Bildung und Kultur**. Er wurde am Dienstag mit deutlicher Mehrheit vom Stadtrat gewählt.

Außerdem sprach sich der Rat mehrheitlich dafür aus, die vom Land geplante **neue Veranstaltungshalle** am Ort der heutigen Saarlandhalle zu errichten. Sollte dies nicht zu realisieren sein, käme die Messe als Standort zum Zuge.

Schließlich wurde der Saarbrücker Tierarzt Dr. Richard Francke zum neuen hauptamtlichen **Zoo-Direktor** gewählt. Er leitet den Zoo derzeit kommissarisch.

[mehr >>>](#)

## FRANÇAIS

### ACTUELLEMENT

#### **Nouveau kulturdezernent culturel choisi Hall d'organisation aux Camphauser route**



Hôtel de ville St. Johann

L'Erik Schrader (35) devient de Berlin **nouveau Dezernent de formation et de culture**. Il a été choisi le mardi à la majorité claire par le conseil municipal.

En outre, le conseil a préconisé à la majorité de construire le **nouveau hall d'organisation** prévu par le pays à la place au Sarre-hall actuel. Si cela ne devait pas être réalisé, la foire comme emplacement viendrait au train.

Enfin, le Saarbrücker a été choisi vétérinaire Dr. Richard Francke au nouveau **directeur de zoo** professionnel. Il conduit actuellement le zoo kommissarisch.

[plus >>>](#)

Bien que la traduction soit loin d'être d'une qualité acceptable, on peut cependant comprendre que M. Schrader a été élu à la direction de la formation et de la culture, qu'une nouvelle construction devrait se faire et qu'un nouveau vétérinaire a été nommé pour diriger le jardin zoologique.

### **2.3. Faciliter la communication interlangue**

La TA peut aussi faciliter la communication dans les entreprises multinationales dont les différentes implantations dans le monde travaillent dans différentes langues. Comme pour les sites Web, les courriers électroniques peuvent être traduits automatiquement pour donner une idée générale de leur contenu. Ensuite, si cela s'avère nécessaire, on fera appel à un traducteur pour rendre la traduction exacte et éviter tout malentendu. Il en va de même pour tous les documents de travail que peuvent s'échanger les différents sites d'une multinationale.

### **2.4. La TA comme base de données linguistiques**

Utiliser la TA comme un dictionnaire en ligne s'est révélé être l'un des emplois les plus récents de cet outil. Lorsqu'il s'agit de systèmes en ligne, ce ne sont pas forcément des professionnels de la traduction, mais plutôt des individus devant rédiger des textes dans une langue seconde, qui en font le plus fréquemment usage. L'outil de TA est alors utilisé comme une sorte d'aide à la rédaction en langue seconde. L'inconvénient de cette approche concerne les expressions ou collocations du vocabulaire général ; en effet, comme celles-ci dépendent

du contexte, il y a donc un risque de se voir proposer une traduction littérale. Les systèmes personnalisés seraient plutôt utilisés par le traducteur qui va chercher comment traduire une expression ou des termes dans un domaine spécialisé. Il fera appel pour cela aux dictionnaires spécialisés inclus dans le système.

### 3. Peut-on dépasser les limites de la TA ?

Nous avons vu quelles étaient les limites des systèmes de traduction automatique. Les fonctions de veille, d'échange d'information par courrier électronique ou de dictionnaire/base de données linguistiques sont faciles à mettre en œuvre avec un système de TA en ligne ou un système installé sur un ordinateur personnel. Il est en revanche évident que pour obtenir une traduction de haute qualité, l'outil de TA doit être intégré dans un environnement plus diversifié et faire partie d'un « flux de production » déterminé.

Les systèmes de TA actuels ne sont pas interactifs – ce qui n'a pas toujours été le cas<sup>5</sup>. Aujourd'hui cependant, ils fonctionnent comme une boîte fermée que l'on nourrit d'un texte en langue source et qui produit un texte en langue cible. En intégrant un système de TA dans un environnement d'outils plus vaste, on permettrait au traducteur d'interagir avec le processus de traduction. La partie automatique n'interviendrait qu'au cours d'un processus de traduction plus complexe. Dans ce cas, la TA serait intégrée avec une mémoire de traduction. Celle-ci produirait les segments qu'elle reconnaîtrait et les segments inconnus seraient traités pas la TA. Les ressources linguistiques - dictionnaires monolingues et bilingues - devraient être accessibles autant au traducteur qu'au système de TA. Cette approche, qui consiste à faire de la TA un outil parmi d'autres et participant avec les autres au processus de traduction, commence à être mise en œuvre. C'est d'ailleurs dans cette optique qu'il faut considérer la formation à l'utilisation de la TA comme faisant partie intégrante de la formation de traducteur spécialisé. A notre avis, c'est une approche qui loin de viser à remplacer le traducteur humain, vient bien plutôt lui offrir son assistance.

La question de savoir si la machine va remplacer l'humain ne se pose donc plus en ces termes maintenant. Nous partageons l'avis de N. Froeliger (ce volume) concernant la traduction pragmatique : elle va certes bien au-delà de ce que peut faire une machine. Nous pensons donc qu'il s'agit davantage ici de trouver un juste équilibre entre traduction purement humaine et traduction machine. Une évolution qu'on ne peut pas nier a envahi le marché mondial de la traduction. Les besoins de texte traduit ont augmenté de manière exponentielle et les entreprises ne mettent pas forcément toujours à disposition les moyens financiers de leurs besoins linguistiques. Ainsi, selon les besoins, on ferait appel à un type de traduction ou à un autre : les interventions humaines et celles de la machine – rappelons-le- se placent, dans cette optique, sur un continuum qui permet d'en varier les proportions. Nous sommes bien sûr consciente que le recours à la machine crée de nouveaux besoins et pousse le traducteur à s'approprier de nouvelles compétences. La guidance de l'entreprise reste et restera entre les mains de l'être humain comme les nouveaux outils à disposition<sup>6</sup> Pour un professionnel de la traduction dans les domaines scientifiques et techniques en priorité – s'ouvre une nouvelle perspective : celle du travail de défrichage, parfois ingrat, vers une activité critique de

---

<sup>5</sup> Dans les années 80, le système ALPS proposait des alternatives au traducteur en cours de traduction ou posait des questions visant à clarifier les structures ambiguës.

<sup>6</sup> Contrairement à ce qui se passe dans le film *Terminator* où les machines ont pris le pouvoir et dominent la Terre.

révision, de recomposition et de supervision. Le moissonneur a jadis quitté la faucille et dorénavant coupe les blés du haut de sa moissonneuse-batteuse, le traducteur d'aujourd'hui peut dorénavant partager sa tâche avec l'ordinateur, mais seulement quand et où – c'est-à-dire pour quel type de texte et que type de traduction – il le veut. Du moins est-ce ce que nous espérons...

## **Références bibliographiques**

Austermühl F. 2001. *Electronic Tools for Translators*. St Jérôme Publishing: Manchester.

Isabelle P., 1986. « Le Projet TAUM de Traduction Automatique », *La Revue Québécoise de Linguistique*, 5:2,

Boitet C., Nedobejkine N. 1980. "Russian-French at GETA: outline of the method and detailed example", *Proceedings of the 8th conference on Computational linguistics (COLING)* Tokyo, Japan, pp. 437 – 446.

Froeliger N. "Pourquoi avoir peur de l'informatisation en traduction", voir dans ce même volume.

Gross M. 1992. "Quelques réflexions sur le domaine de la traduction automatique". *Revue T.A.L.* Vol . 33, N° 1-2.

Hutchins, J. 2001. "Machine translation over fifty years" in Jacqueline Léon (ed) *Histoire, Epistémologie, Langage*. Vol. 23 (1), *Le traitement automatique des langues*, 7-31.

Hutchins J. 2004. Current commercial machine translation systems and computer-based translation tools: system types and their uses. *International Journal of Translation* vol.17, no.1-2, pp.5-38

Kübler N. 2002a. "Teaching Commercial MT to translators: Bridging the Gap between human and machine". In H. Somers (ed.) *Proceedings of the EAMT workshop on MT*, Manchester, UMIST. 155-162

Kübler N. 2002b. "Creating a Term Base to Customize an MT System: Reusability of Resources and Tools from the Translator's Point of View". In *Proceedings of the Language Resources for Translation Work and Research Workshop of the LREC Conference*, 28 May 2002, Las Palmas de Gran Canarias, Espagne. 44-48

Lang E. & Jin Yang 1999. "Automatic Domain. Recognition for **Machine Translation**". In *Proceedings of. the MT Summit VII*, Singapore.

Maegaard B. 1989. EUROTRA: "The machine translation project of the European Communities". In Campbell J. & Cuenca J. (eds) *Perspectives in artificial intelligence vol.2: machine translation, NLP, databases and computer-aided instruction*. Prentice-Hall, Upper Saddle River, pp39-47.

Maxwell, Dan: *Distributed Language Translation: A Multilingual Project*. Utrecht: BSO Research. Reproduced by The Indiana University Linguistics Club. Bloomington. 1990.

Sumita E. & al. 1999 “Solutions to Problems Inherent in Spoken-language Translation: The ATR-MATRIX Approach.” *Proceedings of the MT-Summit VII*, Singapore, September 13-15, vol. 1/1.

Scott, M. & Tribble, C., 2006, *Textual Patterns: keyword and corpus analysis in language education*, Benjamins, Amsterdam.

Volanschi, A., “Outils informatiques et ressources électroniques pour les traducteurs”, voir dans ce même volume.