

Georges Misri  
Maître de conférences émérite  
Université d'Alep  
Traducteur indépendant



**Résumé :** *Les problèmes posés par la traduction automatique sont nombreux et complexes. L'article de Georges Misri nous en propose une analyse claire, complète et rigoureuse.*

**Abstract :** *The automatic translation causes lots of problems. This contribution aims at giving a clear, complete and rigorous opinion about it.*

**Mots-clés :** *Traduction automatique / traduction humaine, homonymie, polysémie, complexité, interlangue et traitement du sens.*

## Introduction

S'il est vrai que la *traduction humaine* est aussi vieille que les langues naturelles elles-mêmes, il n'en est pas moins vrai que la *traduction automatique* est aussi vieille que les ordinateurs eux-mêmes. En effet, la naissance de la première *machine à traduire* date de 1933, année où Smirnov Trojanskij a déposé son invention d'une machine capable de «traduire» simultanément en plusieurs langues (Delavenay E., 1959, p. 35 et ss.).

Les recherches américaines dans ce domaine ont commencé dans les années quarante. Il s'agissait, à l'époque, de créer des dictionnaires automatiques dans les limites de la mémoire des premiers ordinateurs. Le travail des chercheurs se limitait donc aux équivalences lexicales, mais même ce niveau, apparemment simple, s'est révélé beaucoup plus complexe qu'on ne pensait car, les unités lexicales d'une langue ne sont pas toutes monosémiques et ne sont pas non plus toujours simples. Déjà, les problèmes d'*homonymie* et de *polysémie*<sup>1</sup> se faisaient nettement sentir, ainsi que les problèmes du figement linguistique qui caractérise les unités complexes plus ou moins lexicalisées, telles que *garde des sceaux*, *donner du fil à retordre* ou *à vol d'oiseau*. Les chercheurs ont essayé de s'attaquer à l'ambiguïté créée par ces phénomènes linguistiques et leurs efforts se poursuivent même aujourd'hui. S'ils ont réalisé beaucoup de progrès on ne peut pas encore affirmer qu'ils ont réglé tous les problèmes.

La présente étude s'inscrit dans le cadre des recherches portant sur les problèmes de traduction causés par l'homonymie et la polysémie, et sur les réactions du traducteur humain et de l'ordinateur face à ces problèmes. Nous y présenterons d'abord une analyse des notions d'*homonymie* et de *polysémie*, ainsi que les critères qui permettent de les distinguer. Ensuite, nous expliquerons comment le traducteur humain résout les problèmes causés par ces phénomènes linguistiques. Nous procéderons enfin à une analyse détaillée des erreurs de traduction commises par l'ordinateur, suivie d'une évaluation chiffrée de l'importance des faits observés et de leurs conséquences sur le produit de la traduction.

### Objectifs et méthode

Les objectifs de cette recherche peuvent se résumer dans les points suivants :

1. vérifier si l'homonymie et la polysémie posent autant de problèmes à l'homme et à l'ordinateur ;
2. évaluer l'importance de chacune de ces deux notions en traduction humaine et en traduction automatique ;
3. proposer des solutions susceptibles de réduire les erreurs dues aux ambiguïtés d'origine homonymique ou polysémique, et d'améliorer, par conséquent, la performance des systèmes de traduction automatique.

La méthodologie adoptée dans la présente recherche peut être qualifiée d'empirico-déductive, car nous ne partons pas d'hypothèse à confirmer ou à infirmer au cours de la recherche, mais de l'observation des faits traductologiques présents dans un corpus réel constitué de trois parties: l'original à traduire, la traduction automatique et la traduction humaine. Etant donné qu'il n'existe pas encore, à notre connaissance, de systèmes de traduction automatique français-arabe, nous avons travaillé sur des traductions anglais-français réalisées par le système SYSTRAN, utilisé par la Commission des Communautés Européennes depuis la fin des années soixante-dix. Le texte original que nous avons choisi pour notre corpus est extrait d'un document de l'UNESCO (référence: 155 EX/4), que l'on peut considérer comme un texte non technique car il peut être lu et compris sans problèmes par des non-spécialistes. La traduction automatique est faite par le système SYSTRAN (plus précisément par la version appelée «SYSTRAN Classic» et datée de 1998), sans aucune intervention de l'homme, et la traduction humaine est réalisée par les traducteurs et réviseurs de l'UNESCO. Pour des raisons pratiques, nous avons découpé le texte de départ en 60 segments, dont chacun correspond à une phrase entière, sauf dans les cas d'énumération où l'on ne peut pas vraiment parler de phrases.

### L'homonymie et la polysémie

Etant donné que notre sujet de recherche repose sur les notions d'*homonymie* et de *polysémie*, il convient d'établir dès le départ une distinction claire entre ces deux phénomènes linguistiques. Les linguistes s'accordent sur le fait que «l'*homonymie* est l'identité phonique (homophonie) ou l'identité graphique (homographie) de deux morphèmes<sup>2</sup> qui n'ont pas, par ailleurs, le même sens» (Dubois J. *et al.*, 1973), et que la *polysémie* est plutôt la «propriété qu'a un même signifiant de présenter plusieurs signifiés» (Mounin G., 1974). Mais qu'est-ce qui permet d'affirmer que dans tel cas il s'agit d'un terme à plusieurs significations

(polysémie) et non de deux ou plusieurs termes formellement identiques mais sémantiquement différents (homonymie) ? La distinction entre ces deux notions n'est pas toujours évidente: la preuve, c'est qu'un même vocable n'est pas toujours présenté de la même façon dans les dictionnaires de langue. Par exemple, le vocable *siège* est présenté sous une seule entrée dans le *Petit Larousse illustré* (édition de 1981), mais sous trois entrées dans le *Dictionnaire du français contemporain* (Dubois J. et al., 1971). Cela veut dire que les auteurs du premier dictionnaire considèrent *siège* comme un terme polysémique, alors que pour les auteurs du second dictionnaire, il s'agit plutôt de trois homonymes à ne pas confondre: **1. siège** («meuble ou autre objet disposé pour qu'on puisse s'y asseoir...»); **2. siège** («endroit où réside une autorité, où se réunit un Parlement, où fonctionne une société commerciale ou industrielle...»); **3. siège** («ensemble des opérations militaires exécutées pour s'emparer d'une place forte, d'une ville...»). Le *dictionnaire de la linguistique* propose un critère de distinction entre homonymie et polysémie auquel nous adhérons pleinement: «La polysémie diffère de l'homonymie en cela que les signifiés différents sont tout de même perçus comme présentant des traits sémantiques communs, alors que dans l'homonymie ils n'ont rien de commun: *le palais de Versailles, le palais de Justice* (polysémie)/*le palais (dans la bouche)* (homonymie).» (Mounin G., 1974).

### Le comportement du traducteur humain

L'observation du corpus nous a permis de faire une première constatation flagrante: c'est que les énormes problèmes que posent l'homonymie et la polysémie à l'ordinateur sont tous résolus par le traducteur humain professionnel; ce qui est réellement ambigu pour la machine ne l'est pas nécessairement pour l'homme. En effet, l'intelligence humaine étant largement plus développée que celle de l'ordinateur, le traducteur lève toute ambiguïté grâce à sa capacité de raisonnement qui permet de définir, cas par cas, l'information véhiculée par tel ou tel élément du texte de départ. L'examen des traductions humaines adoptées dans le texte d'arrivée permet de dégager les connaissances pertinentes qui permettent au traducteur de lever les ambiguïtés en question. Et comme sa démarche se fait assez souvent inconsciemment, le traducteur humain n'a pas le sentiment d'être en face d'ambiguïtés à lever; c'est en examinant la traduction automatique et les erreurs commises par l'ordinateur qu'il se rend compte que tel terme est polysémique et que tels autres sont des homonymes. Pour illustrer la démarche du traducteur, nous préférons partir de quelques exemples concrets du corpus; commençons par celui-ci:

N°	Original anglais	Traduction automatique	Traduction humaine
15	The development of early childhood education as an <u>integral part</u> <sup>3</sup> of basic education is important not only to cater for the developmental needs of young children, but also in preparing them for further learning.	Le développement de la première éducation d'enfance comme <u>partie intégrale</u> d'éducation de base est important non seulement pour couvrir les besoins développement aux des enfants en bas âge, mais également en les préparant pour davantage d'étude.	Le développement de l'éducation de la petite enfance en tant que <u>partie intégrante</u> de l'éducation de base est important non seulement pour répondre aux besoins de développement des jeunes enfants mais aussi pour les préparer aux étapes ultérieures de l'apprentissage.

Ce qui a permis au traducteur humain de traduire l'adjectif anglais *integral* par *intégrante* et non par *intégrale*, c'est la présence de *part* dans l'entourage immédiat de l'adjectif, ce qui donne en français la collocation *partie intégrante* (et non *partie intégrale*, comme c'est le cas dans la traduction automatique). On peut donc conclure que, dans ce cas, c'est le *micro-contexte* qui détermine le choix du bon équivalent malgré la prétendue ambiguïté. En est-il de même dans les exemples suivants ?

N°	Original anglais	Traduction automatique	Traduction humaine
1	Activities under this programme aim at assisting Member States in realizing the <u>Education for All (EFA)</u> goal set at the World Conference on Education for All (Jomtien, 1990).	activités dans le cadre de ce programme visent à aider des Etats membres en réalisant <u>l'éducation pour tout</u> l'objectif (EFA) fixé à la conférence du monde sur l'éducation pour tous (Jomtien, 1990).	Les activités de ce programme visent à aider les États membres à réaliser l'objectif de <u>l'éducation pour tous (EPT)</u> défini lors de la Conférence mondiale sur l'éducation pour tous (Jomtien, 1990).
2	The Jomtien Conference foresaw the need for a major assessment of progress achieved towards <u>EFA</u> goals during the 1990s.	La conférence de Jomtien a prévu le besoin d'évaluation importante de progrès réalisée vers des buts d' <u>EFA</u> pendant les années 90.	La Conférence de Jomtien avait prévu la nécessité de procéder à une grande évaluation des progrès accomplis dans la réalisation des objectifs de l' <u>EPT</u> au cours des années 1990.

Dans le segment 2, le sigle EFA a été gardé tel quel par l'ordinateur, parce que celui-ci ne l'a pas compris, bien que sa signification ait été explicitée dans la phrase précédente («Education for All»). Pour le traducteur humain, cela n'a posé aucun problème; car, même si celui-ci ne connaît pas le sigle en question, il peut facilement explorer le contexte pour comprendre de quoi il s'agit. Dans ce cas donc, ce n'est pas le micro-contexte mais le *contexte* qui garantit l'intelligibilité. Les limites du contexte peuvent être le paragraphe ou parfois le texte entier dans lequel figure l'unité en question.

Examinons maintenant le segment 35 qui comporte l'expression anglaise «clearing house» :

N°	Original anglais	Traduction automatique	Traduction humaine
35	In most cases, UNESCO's action focuses on awareness and fund-raising, technical assistance and <u>clearing-house</u> services.	Dans la plupart des cas, l'action d'UNESCO se concentre sur la conscience et la mobilisation de fonds, l'assistance technique et les services de <u>chambre de compensation</u> .	Dans la plupart des cas, l'action de l'UNESCO met l'accent sur la sensibilisation et la collecte de fonds, l'assistance technique et les services de <u>d'échange d'information</u> .

Dans un texte de type commercial, cette expression signifie «chambre de compensation», mais quand il s'agit d'un document qui parle des rôles joués par une organisation internationale comme l'UNESCO, le traducteur s'oriente plutôt vers l'idée d'«échange d'information» (ou «centre d'échange d'information»).

Dans un cas pareil, ce qui détermine le sens convenable, c'est la nature du texte ou le domaine auquel appartient le terme ou l'expression à traduire ; c'est ce que nous proposons d'appeler le *macro-contexte*.

Les ambiguïtés d'origine syntaxique auxquelles se heurte l'ordinateur ne posent pas non plus de problèmes au traducteur professionnel; qu'on en juge par l'exemple suivant :

N°	Original anglais	Traduction automatique	Traduction humaine
22	On the occasion of MINEDAF VII (Durban, South Africa)the Ministry of Basic Education of Maliand UNESCO signed an agreement to establish with Malia Regional <u>Early Childhood Resource Centre</u> for Francophone Africa.	A l'occasion de MINEDAF VII (Durban, l'Afrique du Sud) le ministère de l'éducation de base du Mali et de l'UNESCO a signé un accord d'établir avec le Mali <u>un centre tôt régional de ressource d'enfance</u> pour l'Afrique francophone.	A l'occasion de MINEDAF VII (Durban, Afrique du Sud), le Ministère de l'éducation de base du Mali et l'UNESCO ont signé un accord concernant la création au Mali d' <u>un Centre régional de ressources pour la petite enfance</u> en Afrique francophone.

Par sa connaissance intuitive de la syntaxe de l'anglais, le traducteur comprend que *early*, dans ce contexte, est un adjectif qui détermine *childhood*, et non pas un adverbe qui détermine *centre*, comme le laisse croire la grosse bêtise de la traduction automatique. Cette solution ne lui vient même pas à l'esprit; car, grâce à sa maîtrise, consciente ou inconsciente, de la syntaxe du français, il sait très bien que *tôt* appartient à une catégorie d'adverbes qui ne détermine jamais un nom.

Dans d'autres cas, ce sont les considérations *sémantiques* qui permettent au traducteur humain d'éliminer toute ambiguïté et de ne garder que l'information pertinente véhiculée par le texte de départ. L'exemple suivant illustre ce propos :

N°	Original anglais	Traduction automatique	Traduction humaine
30	Providing educational opportunities for the 'unreached' is an essential step towards ensuring education for all, and <u>combating poverty</u> and social exclusion.	Présenter des moyens éducatifs du 'unreached' est une étape essentielle vers assurer l'éducation pour tous, et <u>pauvreté combattante</u> et exclusion sociale.	Offrir des possibilités d'éducation aux «laissés-pour-compte» est une étape essentielle de la réalisation de l'objectif de l'éducation pour tous et de <u>la lutte contre la pauvreté</u> et l'exclusion sociale.

Face à l'ambiguïté de *combating*, qui peut signifier «combattant(e)» (adjectif verbal) ou «combattre» (verbe), l'ordinateur a fait le mauvais choix, parce qu'il n'a pas pu faire la bonne interprétation sémantique (ou syntactico-sémantique). Quant au traducteur humain qui s'attaque à un cas pareil, il se trouve devant deux interprétations possibles : «la pauvreté combattante» ou «combattre la pauvreté». Sans hésiter, il en élimine la première, car il sait

que l'adjectif verbal *combattant(e)* doit déterminer un nom ayant le trait sémantique [+ humain], ce qui n'est pas le cas de *pauvreté*. Il retient alors la seconde solution, car la *pauvreté* étant un fléau social, il est plausible de la *combattre*, et la traduction humaine adoptée dans cet exemple («la lutte contre la pauvreté») va dans ce sens.

Il nous reste un exemple à examiner avant de clore cette partie qui porte sur le comportement du traducteur humain face aux cas d'homonymie et de polysémie :

N°	Original anglais	Traduction automatique	Traduction humaine
6	This is being undertaken through the International Consultative Forum on Education for All (EFA Forum) and a technical advisory group has been set up to advise and assist the EFA Forum Secretariat, based at UNESCO Headquarters, to coordinate this exercise.	Ceci est entrepris par le forum consultatif international sur l'éducation pour tous (forum d'EFA) et un groupe consultatif technique a été installé pour conseiller et aider le secrétariat de forum d'EFA, basé <u>aux sièges sociaux de l'UNESCO</u> , pour coordonner cet exercice.	Ils ont chargé de cette tâche le Forum consultatif international sur l'éducation pour tous (Forum EPT) dont le secrétariat, basé <u>au Siège de l'UNESCO</u> , est assisté d'un groupe consultatif technique ad hoc pour la coordination de cet exercice.

Pour traduire le groupe *at UNESCO Headquarters*, le traducteur humain a opté pour *au Siège de l'UNESCO*, et non pas pour *aux sièges sociaux de l'UNESCO*. Deux éléments d'information lui ont permis d'éviter cette erreur : il sait d'abord que l'UNESCO n'est pas une société commerciale ou industrielle, mais une organisation internationale ; il sait aussi que l'UNESCO a un seul siège, et non plusieurs. Dans le cas de cet exemple, les éléments pertinents qui ont permis de faire le bon choix ne sont pas d'ordre *linguistique*, mais de nature *extra-linguistique*.

On peut déduire de ce qui précède que le traducteur humain élimine, consciemment ou inconsciemment, toutes les ambiguïtés dues à l'homonymie et à la polysémie grâce à des critères *linguistiques et discursifs* (micro-contexte, contexte, macro-contexte, considérations syntaxiques, sémantiques, etc.), mais aussi grâce à ses *connaissances extra-linguistiques* (connaissance du monde, culture générale, connaissances thématiques, etc.). On peut même dire que l'ambiguïté n'existe pour le traducteur que quand il y a une défaillance dans le bagage linguistique ou cognitif de celui-ci.

### Le comportement de l'ordinateur

Avant d'analyser les erreurs commises par l'ordinateur à cause de l'homonymie et de la polysémie, il convient de donner une idée des systèmes de traduction automatique, afin de pouvoir replacer SYSTRAN dans l'ensemble des systèmes connus (Loffler-Laurian A.-M., 1996; Bouillon P. et Clas A., 1993) et de mieux évaluer les avantages et les inconvénients.

Les recherches portant sur la traduction automatique se sont développées notamment grâce à la concurrence scientifique entre les États Unis et l'ex-Union Soviétique, plus particulièrement dans le domaine de l'espace ; car chaque camp voulait savoir, le plus tôt possible, l'état d'avancement des recherches de l'autre. C'est ainsi que les Américains ont alloué, au début des années soixante, des sommes énormes aux recherches de la traduction automatique. Les enthousiastes qui croyaient à la possibilité de réaliser en quelques années des traductions entièrement automatiques de haute qualité étaient nombreux, mais l'expérience a prouvé que la FAHQT (Fully Automatic High Quality Translation) n'était qu'un rêve, et que les phénomènes linguistiques, en général, et traductologiques, en particulier, étaient beaucoup plus complexes qu'on ne pensait. En conséquence, la plupart des subventions dont bénéficiaient les recherches sur la traduction automatique ont été supprimées, notamment après la parution du rapport ALPAC<sup>4</sup> (*Report of the Automated Language Processing Advisory Committee*), qui a clairement montré que les recherches en question ne pourraient aboutir, dans un avenir prévisible, à la réalisation de traductions entièrement automatiques de haute qualité. A partir des années soixante-dix, les recherches se sont poursuivies avec moins d'enthousiasme et plus de réalisme, et des progrès importants ont été réalisés, sans pour autant prétendre atteindre la haute qualité dont il était question auparavant. (Nagao M., 1983).

La classification des systèmes de traduction automatique se fait selon plusieurs critères, notamment : la *génération* à laquelle appartient le système, la *technique* utilisée, le mode d'*intervention humaine* et l'*appartenance théorique*.

Selon le premier critère, on distingue d'habitude deux générations: développés surtout dans les années soixante, les systèmes de *première génération* adoptent une méthode *directe* qui consiste dans deux opérations linéaires, une analyse et une synthèse, qui se font par recours aux dictionnaires et grammaires de la langue source et de la langue cible. Sont considérés comme appartenants à cette génération les systèmes SYSTRAN et TAUM-Météo.

Les systèmes de *deuxième génération* ont été élaborés à partir des années soixante-dix. Leur classification fait intervenir le critère de la *technique* utilisée. Ainsi distingue-t-on les architectures à *interlangue* et les architectures de *transfert*. La première comporte deux modules, analyse et génération, liés par une langue pivot, ou interlangue, constituée d'une représentation abstraite du monde en catégories sémantiques et/ou logiques indépendantes des langues de départ et d'arrivée. Les systèmes suivants adoptent cette technique: ULTRA, KBMT-89, TITUS, ROSETTA et DLT. Quant aux systèmes de transfert, ils comportent trois modules. Un troisième module, dit de *transfert*, s'ajoute à l'*analyse* en langue source et à la *génération* en langue cible. Il traite les équivalences entre les deux langues en question, et il change selon les paires de langues, alors que l'*interlangue* reste la même quelles que soient les langues de traduction. Mais «malheureusement, dans la pratique, l'élaboration de l'*interlangue* s'avère très ardue. De plus, l'extraction des éléments de sens du texte source et leur organisation en une représentation cohérente alourdissent considérablement l'analyse. C'est pourquoi, la plupart des systèmes dits «de deuxième génération» choisissent une architecture de transfert (SUSY, METAL,

EUROTRA, CRITTER...» (Jacqmin L., 1993).

En marge de cette classification en générations, on trouve une classification chronologique en décennies, proposée par Jacques ANIS :

- «la première décennie (≅1945-1955) : les premiers pas» ;
- «la deuxième décennie (≅1955-1965) : de l'enthousiasme à la déception» ;
- «la troisième décennie (≅1965-1975) : la période calme » ;
- «la quatrième décennie (≅1975-1985) : le réveil» ;
- «la cinquième décennie (≅1985-1995) : la maturité». (ANIS J., 1994).

Selon le mode d'intervention humaine, on distingue généralement trois types de systèmes de traduction automatique ou assistée par ordinateur :

- les systèmes où l'intervention humaine, appelée *préédiction*, a lieu avant l'opération traduisante. Cette intervention a pour objectif de réduire les ambiguïtés et d'adapter le texte de départ aux possibilités de l'ordinateur. Les systèmes KANT et ULTRA utilisent la préédiction ;
- les systèmes de *traduction interactive*, où l'intervention humaine s'effectue au cours du processus de traduction : «l'interactivité homme/machine permet dans un certain nombre de cas ambigus de recourir à l'intelligence de l'homme. Ainsi pour lever l'indétermination de signification d'un terme polysémique, un message question apparaît sur le terminal, mentionnant les courtes définitions numérotées des différents sens possibles et connus dans le lexique. L'utilisateur doit seulement taper le numéro du sens qu'il veut attribuer à ce terme polysémique.» (Ducrot J. M., 1989, p. 62). Les systèmes ALPS, KBMT-89 et ROSETTA utilisent l'*interaction* ;
- les systèmes à *postédiction*, où l'intervention humaine s'effectue après la traduction pour corriger les erreurs de la machine. La plupart des systèmes utilisent cette méthode: LOGOS, SYSTRAN, WEIDNER, DLT, EUROTRA, SUSY, etc.

Il convient de préciser que ces trois procédés sont cumulables, car certains systèmes utilisent à la fois la préédiction et l'interaction, comme TITUS IV, BABEL-2 et LIDIA, ou la préédiction et la postédiction, comme METAL.

Enfin, selon l'appartenance théorique, on peut distinguer trois groupes de systèmes :

- les systèmes à dominante informatique, comme SYSTRAN et LOGOS ;
- les systèmes à dominante linguistique, comme CRITTER, DLT, EUROTRA, METAL et ROSETTA ;
- les systèmes fondés sur l'intelligence artificielle, comme KBMT-89 et ULTRA.

Pour résumer ce qui a été exposé à propos de SYSTRAN, nous pouvons dire que c'est un système de *première génération*, qui utilise la méthode *directe*, et qui recourt à la *postédiction* comme intervention humaine destinée à la correction des erreurs de l'ordinateur. SYSTRAN n'est pas fondé sur une théorie linguistique; il a plutôt une base *informatique*: «Dans son brevet, Peter Toma<sup>5</sup> dit clairement que son approche a pour objectif de réduire le processus de traduction aux possibilités et aux limites de la logique informatique. L'approche de SYSTRAN a donc un aspect «terre à terre», pragmatique, efficace.» (Loffler-Laurian A.-M., 1996, p. 52).

Abordons maintenant l'analyse de notre corpus de traduction automatique brute effectuée par SYSTRAN. Il convient de rappeler ici que notre objectif n'est pas de faire l'éloge ou la critique de ce système, mais de dégager des caractéristiques du comportement de l'ordinateur face aux ambiguïtés causées par l'homonymie et la polysémie.

Pour mettre en ordre les résultats de notre analyse du corpus, nous les avons classifiés selon trois critères essentiels: *l'origine de l'erreur*, le *type d'erreur* et les *conséquences de l'erreur*. Les erreurs commises par l'ordinateur peuvent être d'origines très diverses, mais nous n'avons considéré, bien entendu, que celles qui entrent dans le cadre de notre étude, à savoir les erreurs causées par l'homonymie ou la polysémie.

### Les erreurs dues à l'homonymie

L'observation du corpus nous a permis de distinguer *homonymie d'origine syntaxique*, *homonymie d'origine lexicale* et *homonymie d'origine lexico-syntaxique*. Les erreurs dues à une homonymie d'origine syntaxique se divisent en deux types: la *confusion de classes syntaxiques* et l'*homonymie de structure*. Pour illustrer le premier type, examinons l'exemple suivant :

N°	Original anglais	Traduction automatique	Traduction humaine
21	Thanks to this seminar, the experience of the 'Clos d'enfants' <u>project</u> (in Mali), which offers training to mothers and grandmothers in childcare based on traditional practices, is being extended to other African countries.	Grâce à cette conférence, l'expérience des 'd'enfants de Clos <u>projette</u> (au Mali), qui offre la formation aux mères et aux grands-mères dans le childcare basé sur des pratiques traditionnelles, est étendue d'autres pays africains.	Grâce à ce séminaire, l'expérience du <u>projet</u> «Clos d'enfants» (au Mali) qui offre aux mères et grand-mères une formation en puériculture basée sur les pratiques traditionnelles est actuellement étendue à d'autres pays d'Afrique.

L'ordinateur confond ici entre deux classes syntaxiques: il prend *project* pour un verbe («projette») alors qu'il est utilisé comme un nom («projet»). Deux considérations permettent d'éliminer cette mauvaise interprétation: d'abord, si *project* était en rapport sujet-verbe avec *experience*, le verbe *is being extended* resterait sans sujet; ensuite, l'absence d'*s* final dans *project* prouve que celui-ci n'est pas un verbe qui se rapporte au sujet *experience*. Une analyse syntaxique adéquate permettrait d'éviter ce genre d'erreurs. Vu les progrès actuels de la traduction automatique, cette analyse ne serait pas impossible à réaliser, mais elle ne serait pas simple non plus.

Dans le second type, il s'agit d'une ambiguïté créée par une homonymie de structure syntaxique, qui permet d'avoir deux interprétations possibles. L'erreur se produit quand l'ordinateur attribue à un monème un rapport syntaxique erroné, comme le montrent les deux exemples suivants :

N°	Original anglais	Traduction automatique	Traduction humaine
15	early (childhood education)	première (éducation d'enfance)	(éducation de la) petite (enfance)
52	special (education needs)	(besoins d'éducation) spéciale	(besoins éducatifs) spéciaux

Il s'agit ici de la structure anglaise *Adjectif + Nom 1 + Nom 2*, où l'adjectif peut déterminer le premier nom ou le second. Dans le segment 15, l'ordinateur considère que *early* détermine le second nom (*education*) alors qu'il détermine le premier (*childhood*). Inversement, dans le segment 52, il attribue l'adjectif *special* au premier nom (*education*) alors qu'il est en rapport avec le second (*needs*). Dans des cas pareils, l'absence de repères formels complique énormément la tâche de l'ordinateur. En effet, il faudrait explorer le *contexte* dépassant les limites de la phrase pour découvrir, par exemple, que le groupe *early childhood* apparaît ailleurs, et à plusieurs reprises, indépendamment du lexème *education*, et que, par conséquent, il serait plausible de privilégier l'hypothèse que ces deux termes seraient également en rapport dans le contexte du segment 15. Le second exemple présente un cas encore plus délicat; car, même si on dépasse les limites du contexte pour explorer le macro-contexte, c'est-à-dire le domaine de connaissance auquel appartient le texte, à savoir *l'éducation*, on trouvera que les deux groupes *special education* («éducation spéciale») et *special needs* («besoins spéciaux») sont assez fréquemment utilisés dans les ouvrages portant sur l'éducation. Rien jusqu'ici ne nous permet de trancher, mais le contexte où apparaît cette expression ambiguë nous fournit tout de même un élément utile :

N°	Original anglais	Traduction automatique	Traduction humaine
52	<i>Regional meetings/workshops on: Media and Education</i> (Abidjan, Côte d'Ivoire, March 1998); <i>Changing Management of Special Education Needs</i> (Santiago, Chile, April 1998); [...]	<i>Meeting/workshops régional sur: des médias et l'éducation</i> (abidjan, Côte d'Ivoire, mars 1998); <i>Gestion changeante des besoins d'éducation spéciale</i> (Santiago, Chili, avril 1998); [...]	<i>Réunions/ateliers régionaux sur les thèmes suivants</i> - Les médias et l'éducation (Abidjan, Côte d'Ivoire, mars 1998) - <i>Changer la gestion des besoins éducatifs spéciaux</i> (Santiago du Chili, avril 1998); [...]

D'après le contexte, on comprend que notre expression fait partie de l'intitulé du thème d'une réunion régionale qui a eu lieu à Santiago, en avril 1998. La solution serait alors de consulter les documents issus de cette réunion pour pouvoir décider s'il s'agit d'*éducation spéciale* ou de *besoins spéciaux*. On peut déduire de ce qui précède que, dans ce cas, la clé de la solution est de nature *extra-linguistique* (les informations sur le lieu, la date et le thème de la réunion régionale); mais cette clé, en elle-même, n'est pas suffisante: elle suscite un comportement (consulter les documents issus de la réunion), qui va conduire à la bonne interprétation. Le traducteur humain est capable de faire ce raisonnement et de suivre la démarche requise, mais l'ordinateur d'aujourd'hui est loin d'être doté de ce genre de performances.

L'homonymie, qui cause des ambiguïtés pour la machine, peut être aussi d'*origine lexicale*. Dans ce type d'erreurs, il s'agit d'une confusion entre deux lexèmes de forme identique et appartenant à la même classe syntaxique, mais ayant deux significations différentes, ne présentant pas de traits sémantiques communs. En voici un exemple :

N°	Original anglais	Traduction automatique	Traduction humaine
3	While a mid-decade <u>review</u> , held in Amman (June 1996), took stock of the progress made and identified the shortcomings that needed to be <u>addressed</u> , in December 1997 the United Nations General Assembly adopted resolution A/52/84, which called on Member States to provide information on the implementation of their EFA strategies.	Tandis qu'une <u>revue</u> de la mi-décennie, tenue à Amman (juin 1996), examinait progrès accomplis et identifiait les imperfections qui ont dû être <u>adressées</u> , en décembre 1997 la résolution adoptée A/52/84 de l'Assemblée générale des Nations Unies, qui a invité des Etats membres pour fournir des informations sur l'exécution de leurs stratégies d'EFA.	Après l' <u>examen</u> à mi-parcours de la Décennie (Amman, juin 1996), qui a permis de faire le point des progrès réalisés et des lacunes à <u>combler</u> , l'Assemblée générale des Nations Unies a adopté en décembre 1997 la résolution A/52/84 dans laquelle elle a recommandé à tous les Etats membres de fournir des informations sur le stade de mise en œuvre de leurs stratégies de l'EPT.

Cet exemple comporte deux cas d'homonymie lexicale : *review* («revue»/«examen») et *addressed* («adressées»/«combler»). Dans le premier cas, le micro-contexte représenté par *held in Amman* permet d'éliminer l'interprétation selon laquelle *review* serait «une revue», et pousse à favoriser la seconde interprétation («l'examen»), parce qu'il s'agit d'une réunion d'évaluation où l'on *examine* les progrès et les problèmes. De même, dans le deuxième cas, la considération du micro-contexte ne mène pas directement à la solution, mais permet d'éliminer la mauvaise interprétation: le «bon sens» n'admet pas l'expression *des imperfections à adresser*, ce qui pousse le traducteur humain à chercher d'autres termes dans le même champ lexical, d'où le lexème *lacunes*, qui conduit, par conséquent, à l'expression connue *des lacunes à combler*. Mais la machine n'étant pas dotée de ce «bon sens», elle a besoin d'une analyse lexico-sémantique des deux homonymes (*address* = «adresser» et *address* = «s'attaquer à, aborder, traiter (un problème)») et des compléments d'objet que chacun admet: *une lettre, une enveloppe, etc.*, dans le premier cas, *un problème, des lacunes, etc.*, dans le second. Dans l'état d'avancement des systèmes actuels de traduction automatique, ce genre d'analyse ne serait pas impossible à réaliser. Son application à un groupe limité de verbes semble tout à fait possible, mais sa réalisation à l'échelle d'une langue entière serait extrêmement complexe, et nécessiterait des efforts énormes.

Nous avons rattaché à l'homonymie un autre phénomène constaté dans le corpus: c'est l'ambiguïté produite par ce qu'on appelle les *expressions figées*, qui sont des groupes de monèmes plus ou moins lexicalisés, et présentant assez souvent une différence nette entre leur signification globale (ou *signification figée*) et leur signification monème par monème (ou *signification non figée*). L'homonymie serait ici entre l'expression considérée globalement et le groupe de monèmes considéré élément par élément. Nous avons considéré ce

phénomène comme une homonymie d'origine lexico-syntaxique: d'abord parce qu'il s'agit de groupes plus ou moins lexicalisés, donc pouvant être considérés comme des unités lexicales complexes; ensuite parce qu'ils se caractérisent par un comportement syntaxique particulier (Martinet A., 1967). Voici quelques exemples qui illustrent ce type d'homonymie :

N°	Original anglais	Traduction automatique	Traduction humaine
13	following up	suivre vers le haut	suivi
18	non-state sectors	secteurs d'non-état	secteur privé
18	state funding	l'état plaçant	fonds publics
28	clearing-house	chambre de compensation	échange d'information

Comme on peut le constater, les trois premiers exemples sont traduits mot à mot par l'ordinateur ; c'est ce que nous appelons *traduction par défigement*, parce que l'expression figée est considérée comme un groupe non figé, ce qui mène à une déformation de l'information véhiculée par le message de départ. Dans le quatrième exemple, il ne s'agit pas de défigement, parce que l'expression de départ (*clearing-house*) est traduite par une expression figée (*chambre de compensation*); mais celle-ci, qui serait acceptable ailleurs, ne constitue pas un bon équivalent dans ce contexte. Un enrichissement du dictionnaire des expressions intégré au système de traduction automatique, et une amélioration des algorithmes de reconnaissance de ces expressions aideraient l'ordinateur à réduire le nombre d'erreurs de ce type, en évitant les traductions par défigement.

### Les erreurs dues à la polysémie

Les erreurs dont l'origine est la polysémie peuvent être divisées en deux groupes: celles qui sont liées à une *polysémie caractérisant une unité indépendante*, et celles qui sont dues à une *polysémie caractérisant un élément dépendant d'une expression figée*. L'exemple suivant illustre les erreurs du premier groupe :

N°	Original anglais	Traduction automatique	Traduction humaine
32	Various <u>sponsorship schemes</u> have been concluded involving the media, private enterprise, and public figures associated with UNESCO's work and ideals.	On a conclu de divers <u>arrangements de patronage</u> impliquant les médias, l'entreprise privée, et les figures publiques associées au travail et aux idéaux d'UNESCO'S.	Divers <u>projets de parrainage</u> ont vu le jour grâce à des médias, des entreprises du secteur privé et de personnalités associées à l'action et aux idéaux de l'UNESCO.

Deux cas de polysémie d'unités indépendantes peuvent être relevés dans cet exemple: *sponsorship* («patronage»/«parrainage») et *schemes* («arrangements»/«projets»). On peut noter ici que les mauvais choix de l'ordinateur n'aboutissent pas à une véritable déformation du message de départ ; ils conduisent plutôt à des termes sémantiquement proches des lexèmes convenables, mais qui ne constituent pas des équivalents propres dans les contextes en question. Choisir

parmi les paronymes celui qui correspond le mieux à un contexte donné ne serait pas une tâche facile, même pour le traducteur humain. Réexaminons le cas de *sponsorship* : le dictionnaire bilingue *Robert & Collins* (édition de 1987) décrit ce lexème de la façon suivante: «*[loan]* cautionnement *m* ; *[child, member]* parrainage *m* ; *(Rad, TV)* commande *f* publicitaire; *[appeal, announcement]* patronage *m* ; *(Sport)* sponsoring *m*». Ce dictionnaire propose donc cinq équivalents possibles. Il les distingue selon les lexèmes susceptibles d'apparaître dans le micro-contexte ou le contexte de l'unité en question (*loan, child, etc.*), présentés entre crochets carrés, mais aussi selon le macro-contexte, c'est-à-dire le domaine où le lexème peut être utilisé (*Rad, TV; Sport*), présenté entre parenthèses. Même si on réussit à doter l'ordinateur de ces critères de distinction, il serait impossible de prévoir tous les contextes possibles pour une unité ; la preuve, c'est que l'emploi du lexème *sponsorship* dans notre exemple ne correspond à aucun des contextes ou des macro-contextes prévus par le dictionnaire consulté.

Les exemples suivants illustrent le deuxième type de polysémie, celui qui caractérise un élément dépendant d'une expression figée :

N°	Original anglais	Traduction automatique	Traduction humaine
13	non-governmental organizations	organismes non gouvernementaux	organisations non gouvernementales
14	primary education	l'éducation primaire	l'enseignement primaire
55	equalization of opportunities	égalité [sic] des occasions	égalité des chances

Dans ces cas, les solutions adoptées par l'ordinateur n'empêchent pas la compréhension du message, mais un monème de l'expression (ou parfois plusieurs) est traduit sans tenir compte de la globalité de l'expression, ce qui donne des formes complexes proches des expressions équivalentes, que le traducteur humain trouve, d'ailleurs, sans beaucoup de peine. Les propositions que nous avons déjà faites (enrichissement du dictionnaire des expressions et amélioration des algorithmes de reconnaissance de ces expressions) seraient valables ici pour réduire le nombre des erreurs commises par l'ordinateur dans ce genre de situations.

## Résultats en chiffres

Pour être en mesure d'évaluer scientifiquement l'importance des phénomènes traductologiques étudiés, il convient de recourir à quelques chiffres. L'observation de notre corpus, constitué d'une soixantaine de phrases, nous a permis de relever en tout 309 erreurs de traduction. Les problèmes dus à l'homonymie et à la polysémie représentent 170 erreurs, ce qui correspond à un taux de 55,01 %. Cela reflète la grande importance de ces deux phénomènes comme sources d'ambiguïté donnant du fil à retordre aux concepteurs de systèmes de traduction automatique. La polysémie produit plus d'erreurs que l'homonymie, mais l'écart entre les deux n'est pas énorme: 54,70 contre 45,30 %. On peut donc en déduire qu'il existe un certain équilibre entre ces deux sources d'erreurs.

Cependant, du point de vue des conséquences des erreurs commises par la machine, la différence entre les deux phénomènes étudiés est très importante. En effet, dans la totalité des cas, la polysémie ne produit qu'une *impropriété lexicale*, c'est-à-dire qu'elle mène à l'emploi d'un lexème (ou d'une expression) qui ne constitue pas le terme propre qui s'impose dans le contexte en question. Quant à l'homonymie, elle mène systématiquement à une *infidélité à l'information*, c'est-à-dire à une déformation nette de l'information véhiculée par le texte de départ, considéré dans la situation de communication à laquelle il appartient. La grande différence est donc dans la gravité de l'erreur: la polysémie produit plus d'erreurs, mais celles-ci sont nettement moins graves. Une autre différence vient corroborer ce qui vient d'être signalé: c'est que l'homonymie conduit, dans certains cas, à d'autres erreurs que la polysémie ne produit pas d'habitude; ces erreurs consistent dans des *incorrections* et des *non-conformités à l'usage*. Les premières produisent des formes grammaticales incorrectes, alors que les secondes conduisent à des tournures qui ne sont pas conformes aux habitudes linguistiques des usagers de la langue d'arrivée. Les *incorrections* commises par la machine représentent 18,18 % des erreurs dues à l'homonymie, et les *non-conformités à l'usage* constituent 16,88 % de ces erreurs. Donc, on peut dire que l'homonymie perturbe plus l'ordinateur et le pousse à commettre des erreurs linguistiques qui s'ajoutent au problème principal : l'*infidélité à l'information*.

## Conclusion

Ayant pris connaissance des questions traitées dans ce travail et des principaux résultats de la recherche, le lecteur est en mesure de conclure que le fond de la comparaison entre le comportement du traducteur humain et celui de l'ordinateur doté d'un système de traduction automatique ne consiste pas dans le fait de confronter les erreurs de l'un à celles de l'autre, mais dans la question de savoir comment améliorer la performance de l'ordinateur pour que celui-ci s'approche du niveau du traducteur humain. Les solutions que nous avons proposées au cours de notre présentation de la recherche relèvent de trois sources d'informations distinctes: la *langue*, le *discours* et l'*univers extra-linguistique*. Les connaissances touchant à la *langue* (lexique, syntaxe, morphologie, etc.) sont moins difficiles à intégrer dans un système de traduction automatique que celles qui relèvent du *discours*, car celui-ci étant une production individuelle et inédite, il se prête à toutes sortes de créations personnelles et de structurations sémantiques imprévisibles hors situation; ceci complique énormément la tâche de l'ordinateur. La considération des connaissances qui relèvent de l'*univers extra-linguistique* est encore plus compliquée, car ces connaissances sont tellement énormes et tellement variables qu'il serait pratiquement impossible de les intégrer dans un système de traduction automatique, sans parler de la difficulté de doter l'ordinateur de la capacité de les mobiliser au bon moment pour lever telle ou telle ambiguïté.

Pour faire face aux problèmes d'homonymie et de polysémie, l'ordinateur devrait donc être capable non seulement de mobiliser ses connaissances linguistiques de base, mais aussi de relever dans le micro-contexte ou le contexte les éléments pertinents qui permettent de lever toute ambiguïté. Si tous ces

éléments s'avèrent insuffisants, une recherche de connaissances pertinentes dans le bagage cognitif sera nécessaire. Une telle capacité de *raisonnement* exigerait une *intelligence* semblable à celle de l'être humain, ce qui n'est pas encore à la portée des technologies actuelles. La grande différence entre le traducteur humain et l'ordinateur réside dans le fait que le premier possède une intelligence qui lui permet de gérer toutes situations *imprévisibles*, alors que le second ne peut traiter avec succès que les cas *prévisibles*. A notre connaissance, aucun système actuel de traduction automatique ne donne entière satisfaction, parce que les textes à traduire ne sont jamais entièrement prévisibles. Il faut tout de même reconnaître que beaucoup de progrès ont été réalisés jusqu'à présent dans ce domaine. Il convient de signaler, par ailleurs, que les chercheurs d'aujourd'hui parlent de la traduction automatique avec un certain *réalisme* ou, au mieux, avec un *optimisme modéré*. L'ancien directeur scientifique de Systran International affirme que «la traduction humaine génère un produit cher de haute qualité, utilisable par exemple en édition. La traduction automatique a pour ambition de transmettre une information relativement globale vite et à bon marché.» (Trabulsi S., 1989, p. 18). Les qualités de la traduction automatique résident donc non dans la *qualité* du produit, mais dans la *rapidité* et le *coût réduit*. Certains même ont abandonné l'idée de la traduction automatique pour s'orienter vers la *traduction assistée par ordinateur*, plus particulièrement vers la création de *systèmes d'aide à la traduction* (Makhlovitch E., 1993; Isabelle P., 1992), c'est-à-dire des outils informatiques (dictionnaires, encyclopédies, banques de données terminologiques, mémoires de traduction, etc.) mis à la disposition du traducteur dans sa *station de travail* pour l'aider à améliorer son rendement, en quantité et en qualité. Donc, les orientations actuelles tendent moins à remplacer l'homme par la machine qu'à mettre les technologies informatiques au service du traducteur humain.

## Notes

1. Ces deux notions linguistiques, qui ne sont pas faciles à distinguer, seront expliquées plus loin.
2. Dans l'emploi générativiste du terme, les "morphèmes" couvrent à la fois les monèmes lexicaux et les monèmes grammaticaux.
3. Tous les mots soulignés dans les exemples ne le sont pas dans le corpus; nous les soulignons simplement pour que le lecteur puisse les retrouver plus facilement dans le contexte.
4. Le sigle ALPAC représente «the Automatic Language Processing Advisory Committee» (voir bibliographie).
5. Peter Toma est le concepteur du système SYSTRAN.

## Bibliographie

Abbou, A. (éd.), 1989. *La traduction assistée par ordinateur, Actes du séminaire international sur la TAO et dossiers complémentaires*, Paris, 17-18 mars 1988, Paris, DAICADIF-Observatoire des industries de la langue, 234 pages.

ALPAC, 1966. *Language and Machines. Computers in Translation and Linguistics*, Report by the Automatic Language Processing Advisory Committee (ALPAC), Division of Behavioral

Sciences, National Academy of Sciences, National Research Council, publication 1416, Washington.

Anis, J., 1994. «Ordinateur et traduction : survol d'un demi-siècle», *Langages*, N° 116, p. 111-122.

Bouillon, P.; Clas, A. (éd.), 1993. *La traductique : études et recherches de traduction par ordinateur*, Montréal, les Presses de l'Université de Montréal, 507 pages.

Delavenay, E., 1959. *La Machine à traduire*, coll. « Que sais-je ? », n° 834, Paris, P.U.F. 128 pages.

Dubois, J. et al., 1971. *Dictionnaire du français contemporain*, Paris, Larousse, 1260 pages.

Dubois, J. et al., 1973. *Dictionnaire de linguistique*, Paris, Larousse, 516 pages.

Ducrot, J. M., 1989. «Le Système TITUS IV: Système de traduction automatique et simultanée en quatre langues», in : ABBOU A. (éd.), 1989. *La traduction assistée par ordinateur, Actes du séminaire international sur la TAO et dossiers complémentaires*, Paris, 17-18 mars 1988, Paris, DAICADIF-Observatoire des industries de la langue, p. 55-67.

Isabelle, P., 1993. «Les corpus bilingues: une nouvelle ressource pour le traducteur», in: BOUILLON P.; CLAS A. (éd.), *La traductique : études et recherches de traduction par ordinateur*, Montréal, les Presses de l'Université de Montréal, p. 288-306.

Jacmin, L., 1993. «Classification générale des systèmes de traduction automatique», in: BOUILLON P.; CLAS A. (éd.), *La traductique : études et recherches de traduction par ordinateur*, Montréal, les Presses de l'Université de Montréal, p. 43-63.

Loffler-Laurian, A.-M., 1996. *La traduction automatique*, Villeneuve d'Ascq, Presses universitaires du Septentrion, 156 pages.

Makhlovitch, E., 1993. «Le poste de travail du traducteur (PTT), ou les aides à la traduction», in : Bouillon P.; Clas A. (éd.), *La traductique : études et recherches de traduction par ordinateur*, Montréal, les Presses de l'Université de Montréal, p. 281-287.

Martinet, A., 1967. «Syntagme et syntème», *La Linguistique*, fasc.2, p. 1-14.

Mounin, G. (éd.), 1974. *Dictionnaire de la linguistique*, Paris, P.U.F., 340 pages.

Nagao, M., 1983. «La traduction automatique». *La Recherche*, vol. 14, N° 150, p. 1530-1541.

Nirenburg, S., 1993. «L'interlangue et le traitement du sens dans les systèmes de traduction automatique», in : Bouillon P.; Clas A. (éd.), *La traductique : études et recherches de traduction par ordinateur*, Montréal, les Presses de l'Université de Montréal, p. 91-108.

Trabulsi, S., 1989. «Le Système SYSTRAN», in : Abbou A. (éd.), 1989 - *La traduction assistée par ordinateur, Actes du séminaire international sur la TAO et dossiers complémentaires*, Paris, 17-18 mars 1988, Paris, DAICADIF-Observatoire des industries de la langue, p. 15-19.