

La prétraduction automatique, outil de productivité et d'évolution professionnelle

par Claude Bédard
Publié dans META, vol. 37, n° 4, décembre 1992, pp. 738-760

1 Introduction

Depuis deux décennies, on présente volontiers la traduction automatique (TA) comme le paradigme par excellence d'aide aux traducteurs. Pourtant, son utilisation ne s'est pas tellement répandue à ce jour. Trois causes, probablement, sont à invoquer à ce sujet.

- D'abord, il faut admettre que les performances des systèmes de TA ne sont pas encore convaincantes¹.
- Ensuite, il existe au départ une certaine résistance au changement de la part des traducteurs. À cette résistance originelle, on peut ajouter des réactions d'antipathie dues au point précédent.
- Une troisième cause, encore méconnue, réside selon moi dans une *erreur de perception* quant à l'intérêt que représente la TA pour les traducteurs.

Ce dernier point sera développé dans les pages qui suivent, et m'amènera à poser un nouveau paradigme de traitement : *la prétraduction automatique*.

2 Le paradigme «traduction automatique»

Pourquoi est-il si difficile de se faire une idée claire de *l'utilité réelle de la TA*? Une des raisons est le travail de désinformation auquel se livrent plus ou moins consciemment divers intervenants du milieu². Mais une autre raison réside dans l'absence d'analyse éclairante sur la question; nous allons tenter ici de combler cette lacune.

¹ Un indice concret de cet état de choses — si l'on était porté à en douter — est le récent engouement pour l'évaluation des systèmes de TA (démarche très complexe et encore mal définie). Si ces systèmes étaient réellement au point, la mesure d'évaluation adoptée serait le degré de satisfaction de l'utilisateur, autrement dit la rentabilité effectivement constatée. La question d'une évaluation «scientifique» ne se poserait pas — comme d'ailleurs pour tout ce qui, dans la pratique, fonctionne de toute évidence de façon satisfaisante.

² Le discours officiel des développeurs de systèmes de TA (et de certains de leurs clients, pour des raisons notamment d'image publique) réaffirme régulièrement le succès de la TA, souvent en contraste flagrant avec la réalité concrète.

2.1 TA : l'actif

Il s'agit d'abord de poser la bonne question : *Que peut attendre le traducteur professionnel, dans l'état actuel de la technologie, d'un système de TA?* Il faut aussi s'astreindre à une réponse concrète et honnête. Celle-ci peut se résumer ainsi :

- Le *rappel du vocabulaire* contenu dans un dictionnaire. Ce rappel a un double avantage : faire gagner du temps sur la consultation manuelle d'un lexique et assurer l'uniformité terminologique.

C'est certainement dans l'*uniformité* que réside un avantage incontestable de la TA. L'humain — c'est son point fort — est incomparablement plus souple que l'ordinateur. Mais cette grande souplesse, qui le fait réagir de façon variable à des situations variables, le fait réagir de même à des situations invariables. Autrement dit, la répétabilité de l'humain est douteuse. L'ordinateur, à cet égard, apporte au traducteur une aide réelle, parce que complémentaire à ses facultés.

- La *construction de phrases* LA (langue d'arrivée) qui sont le reflet plus ou moins littéral du texte LD (langue de départ).
- Une certaine protection contre les *fautes d'inattention* : le système transcrit fidèlement les chiffres et autres invariants; il ne traduit jamais *higher* par *inférieur*, *increase* par *diminuer*; il ne saute jamais de paragraphes.

Signalons que les textes les plus fastidieux et répétitifs (donc très faciles pour le traducteur) sont curieusement le siège privilégié de fautes d'inattention. On peut expliquer cela par le fait que la mémoire à court terme du traducteur «s'encrasse» à cause du peu de variété entre les segments successifs à traduire; le traducteur a du mal à bien «vider» cette mémoire après chaque segment.

- Un allègement de la tâche de *frappe du texte* traduit.

Voilà qui semble à première vue encourageant. Pourtant, tout traducteur d'expérience reste confusément sceptique : la traduction, dont la maîtrise exige tant d'années d'étude et de pratique, se ramène-t-elle donc à des opérations si simples?

2.2 Limites de l'actif

En fait, le constat suivant s'impose : un système de TA *ne touche que l'aspect trivial de la traduction* — et encore, avec un bonheur inégal.

Le tableau de la figure 1 dissèque les composantes du processus traductionnel et met en parallèle le degré de difficulté pour le traducteur humain et un système de TA.

Figure 1 — Composantes du processus traductionnel

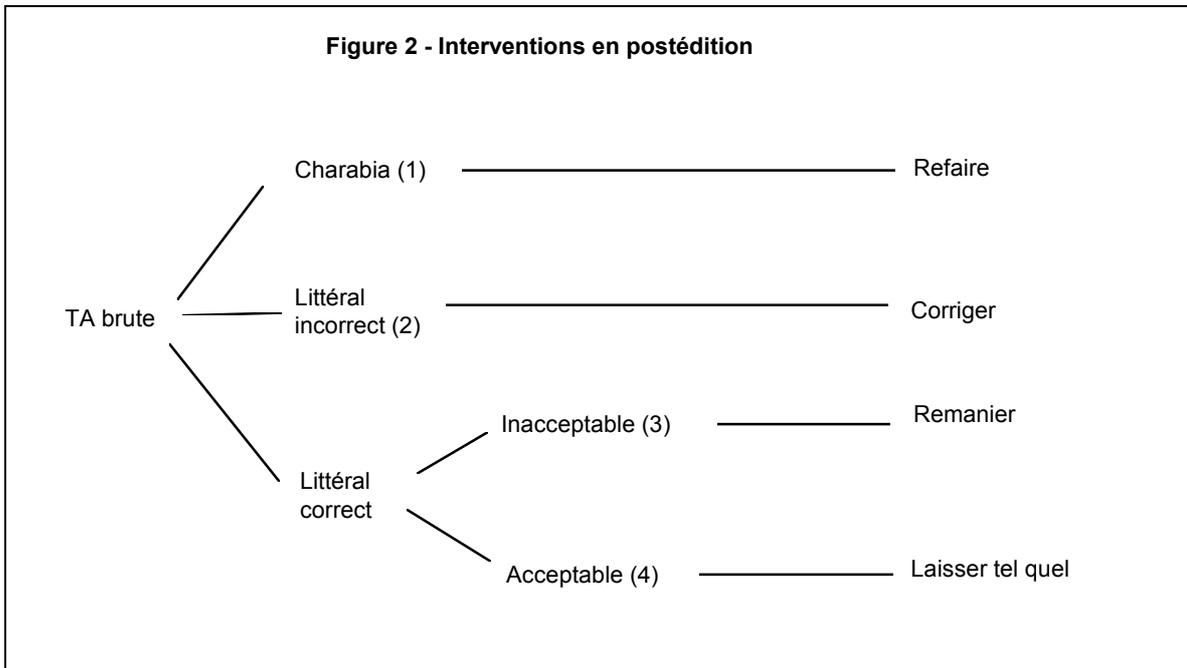
	...pour l'humain	...pour la machine
Transcription des invariants	Fastidieux	Facile
Rappel lexical	Fastidieux	Facile ¹
Choix lexicaux élémentaires	Facile	Difficile
Analyse syntaxique	Facile	Difficile
Formulation littérale correcte	Facile	Facile/Difficile ²
Choix lexicaux difficiles	Difficile	Impossible
Reformulation	Difficile	Impossible
Compensation des défauts	Difficile	Impossible
Vérification du message	Facile/Difficile	Impossible

- 1 La machine est sensible à la moindre variante d'écriture (ainsi qu'aux disjonctions).
 2 Possible uniquement si l'analyse a réussi.

- Côté lexical, le rappel des équivalents est facile pour la machine (bien qu'il reste sensible à la moindre variante d'écriture dans le texte LD). Tout va bien tant qu'on est en situation d'univocité; mais dès que se présente une situation de choix, la machine est mal équipée pour trancher intelligemment³.
- Côté syntaxique, la production d'une traduction littérale (proche de la LD, mais sans faute de grammaire en LA) n'est souvent pas une mince affaire pour la machine : l'analyse automatique de la LD est encore mal maîtrisée; quant à la génération en LA, en plus de la linguistique contrastive (information sur le passage d'une langue à l'autre) à mettre en oeuvre pour combler les besoins les plus élémentaires, elle dépend bien sûr du succès préalable de l'analyse.
- Toutes les composantes plus exigeantes (choix lexicaux difficiles, reformulations visant l'idiomaticité en LA, compensation des inexactitudes du texte LD, etc.) sont laissées à peu près intégralement au traducteur.
- Quant à la vérification du message, il s'agit d'une opération incontournable — tantôt simple, tantôt difficile pour l'humain; chose certaine, la machine n'est d'aucun secours en l'occurrence.

Étant donné la superficialité de l'intervention machine, la traduction brute trouve difficilement grâce aux yeux du postéditeur (voir la figure 2). On pourrait distinguer les quatre cas suivants :

³ Voir Alan Melby, «Lexical Transfer: A missing element in linguistic theories», dans Coling '86, pp. 104-106, et «Lexical Transfer: Between a Source Rock and a Hard Target», dans Coling '88, pp. 411-413.



- **Cas 1.** Il s'agit d'un non-sens complet (charabia) souvent dû à une analyse erronée du texte LD : le postéditeur doit tout refaire.
- **Cas 2.** La traduction est émaillée d'erreurs grammaticales, qu'il faut corriger (le mal est moindre s'il s'agit d'interventions ponctuelles).
- **Cas 3.** La traduction brute est réussie, mais n'en reste pas moins littérale — ce qui souvent n'est pas acceptable; le postéditeur doit alors remanier la traduction, souvent en profondeur. Considérons par exemple la traduction brute suivante, sans erreur de grammaire ni de sens; elle serait pourtant jugée inacceptable par la plupart des traducteurs.

Independence is essentially an attitude of mind, characterized by integrity, objectivity and an impartial approach to professional work.

L'indépendance est essentiellement une attitude d'esprit, caractérisée par l'intégrité, l'objectivité et une approche impartiale du travail professionnel.

La question de l'acceptabilité d'une traduction littérale est fort difficile à trancher. Elle est pourtant très importante, car le remaniement de phrases littérales représente souvent le plus gros du travail de postédition, et c'est de l'ampleur de ce travail que dépend la rentabilité de la TA. Cette question mériterait d'être débattue en détail⁴; je propose ici divers facteurs qui entrent en jeu quant à la décision de reformuler ou non un passage :

- La nature du texte : certains textes exigent un effort de style, alors que d'autres tolèrent un style quelconque.
- La compétence du traducteur; si par exemple la LA n'est pas la langue maternelle du traducteur, ce dernier sera beaucoup plus tolérant à des formulations douteuses.
- L'attitude du traducteur : elle va du laxisme au perfectionnisme (et elle peut être influencée par les circonstances, notamment les délais de travail et l'utilisation prévue de la traduction).
- Les normes de qualité en vigueur dans la profession, lesquelles peuvent varier d'un employeur, d'une région ou d'un pays à l'autre. Au Canada, par exemple, elles sont assez élevées, ce qui explique en partie le succès très limité de la TA dans ce pays.
- **Cas 4.** La traduction brute est acceptable. Ici, cependant, il ne faut pas surestimer le gain réalisé. En effet :
 - Il revient quand même au traducteur de prendre le temps de juger s'il accepte la traduction littérale proposée.
 - Ce qui lui est épargné alors n'est qu'un travail de bas niveau, qu'il aurait exécuté rapidement et sans effort.

Nous avons ici la base d'une réponse à notre question du début : *Un système de TA ne peut aider le traducteur que dans ce qui lui est déjà facile.* Or les opérations restantes sont passablement chronophages pour le traducteur. Par exemple, telle traduction littérale «qui ne sonne pas bien» peut prendre de longues minutes à améliorer; de même pour lever telle ou telle ambiguïté du texte LD, ou pour n'importe quelle autre difficulté de traduction. Rappelons deux lois bien connues des traducteurs de métier :

- La dernière tranche de 10 % de qualité peut englober 50 % des efforts du traducteur.
- Pour assurer la qualité, rien ne remplace jusqu'à un certain point la quantité de temps qu'on consacre à la traduction.

⁴ Souhaitons que se fasse un examen de conscience entre traducteurs sur ce qu'on pourrait appeler le «syndrome d'acharnement reformulationnel», qui nuit à la productivité des traducteurs dans les milieux où il sévit.

Il faut se l'avouer : les gains de productivité qu'on peut attendre d'un système de TA ne peuvent jouer que dans la proportion «facile» du texte; la partie difficile du texte, celle qui «bouffe du temps» au traducteur, reste entièrement à sa charge⁵. Voilà une constatation incontournable, qui devrait déjà nous inciter à modérer nos attentes face à la TA actuelle.

2.4 TA : le passif

Ayant examiné les gains possibles de la TA (l'actif) ainsi que ses limites, tournons-nous maintenant vers l'autre face de la médaille, qu'on pourrait appeler «le passif de la TA» : en effet, les gains de productivité attribuables à l'aide de la machine sont contrebalancés par des coûts sur d'autres plans.

- Il y a d'abord, bien entendu, le prix d'*acquisition* du système : le logiciel lui-même et, le cas échéant, du matériel informatique.
- Viennent ensuite les coûts d'*exploitation* (plus ou moins cachés) du système. Il y a bien sûr les tâches de maintenance et autres manipulations purement informatiques, qui peuvent être assez lourdes pour la plupart des gros systèmes. Il ne faut pas oublier non plus le temps consacré à la formation des utilisateurs. Mais surtout, un certain travail d'enrichissement des dictionnaires machine est nécessaire (selon la convivialité de l'interface et la quantité d'information réclamée par le système, ce travail est plus ou moins chronophage). À ce chapitre, il importe de distinguer les *entrées terminologiques*, lesquelles sont pleinement justifiées, et les *entrées linguistiques*, qui visent à améliorer les performances du système; ces dernières sont souvent les plus difficiles (donc chronophages) et ne font que combler les lacunes du dictionnaire machine ainsi que des mécanismes linguistiques du système.
- Enfin, la sortie brute du système de TA comporte une proportion variable de charabia (dû tantôt à des erreurs d'analyse, tantôt à une formulation boîteuse en LA). Non seulement ce charabia est-il une source d'inconfort mental pour le traducteur, mais il se trouve en fait à le ralentir dans son travail : il s'agit donc véritablement d'un «coût».

2.5 La TA se serait-elle trompée de public-cible?

Le paradigme de la TA est hérité de la période «naïve» où l'on visait une traduction brute d'une qualité suffisante pour consommation directe — par un public unilingue (du moins quant aux langues en cause). Depuis l'on a bien sûr déchanté; les concepteurs se sont tournés vers les traducteurs pour la postédition... sans se demander si la formule était réellement valable pour une personne *dont les compétences sont radicalement différentes*.

⁵ «...If this MT output were to be used only as a rough first translation, *the translator would still not be much closer to the type of translation produced by the human translation group than he would be if he had just the German source text. Someone must still analyze and understand the source text on all levels (not just morphologically and sometimes syntactically) before an adequate translation can be achieved. To state or imply that this job can be done by the computer and that MT output need only be revised like any human translation is at best misleading, although this is an argument which is frequently put forward by proponents of MT.*». M.J. Wormwood, «MT Project at University of Innsbruck», META, vol. 37-2, 1992, p. 314. (Les italiques sont de moi.)

On pourrait illustrer ce fait par l'analogie suivante. Passons un texte imprimé dans un lecteur optique. L'opération n'étant pas tout à fait réussie, la sortie électronique comporte en moyenne trois fautes par ligne de 60 caractères. Il reste à détecter et à corriger ces fautes manuellement afin d'obtenir un texte parfaitement transcrit.

- Soumettons cette sortie à une personne sans compétences particulières devant un clavier — qui tape, comme on dit, «avec trois doigts». Cette personne préférera certainement corriger cette sortie plutôt que de taper le texte en entier.
- Soumettons la même sortie à une copiste professionnelle — qui tape, disons, 80 mots/minute. Il y a de bonnes chances qu'elle préfère tout retaper. Pourtant, arithmétiquement, le texte était parfait à (3 fautes pour 60 caractères) 95 pour cent!

De cette image, on pourrait conclure ceci : il est souvent erroné de penser que *plus on en fait pour le professionnel, plus on l'aide*.

Autre analogie : on pourrait comparer une traduction brute à la surface miroitante d'un lac, sur laquelle on aurait l'impression de pouvoir marcher. Le système de TA, en effet, n'a fait que traiter la *surface* du texte (tous les mots sont traduits, les verbes conjugués, les adjectifs accordés...). Dès que le traducteur essaie de «marcher» sur cette pellicule linguistique, il s'enfonce — car la traduction comporte une dimension *profondeur* avec laquelle il est inéluctablement confronté. Cette illusion d'optique pourrait se résumer ainsi : «Ce n'est pas parce qu'on *automatise* un processus qu'on le *simplifie*.»

Toutefois, la traduction automatique, même de qualité douteuse, reste forcément une chose souhaitable pour l'unilingue, et il importe de poursuivre très activement les recherches dans ce sens. Par contre, il serait bon de commencer à douter de son degré d'efficacité pour les traducteurs eux-mêmes.

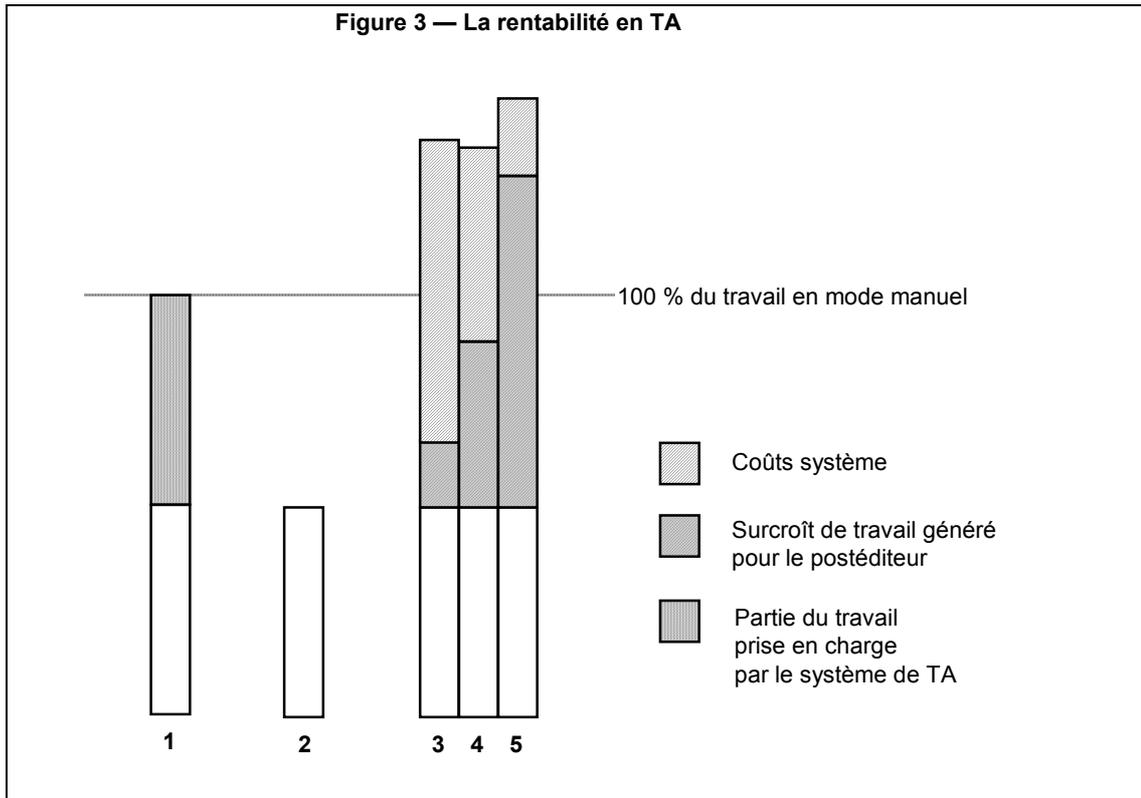
2.6 Une rentabilité douteuse

À l'instar de la question «La traduction est-elle possible?», la question «La TA (pour traducteurs) est-elle possible?» est devenue oiseuse : des postéditeurs sont bel et bien au travail depuis plus d'une décennie. Mais la vraie question, malheureusement, est la suivante : «La TA (pour traducteurs) est-elle *rentable?*»

De cette analyse de l'actif et du passif de la TA, il ressort que la «fenêtre de rentabilité» est assez étroite : d'une part les gains de productivité sont au départ limités, et d'autre part ils sont grugés (souvent en entier et même au-delà) par les coûts inhérents à la TA. En un mot, il semble que dans l'état actuel des choses, *la TA soit trop exigeante pour ce qu'elle peut rapporter.*

La figure 3 résume l'essentiel de ce raisonnement. La colonne 1 représente l'étendue du travail à effectuer en traduction manuelle, sans TA. Si l'on reprend l'idée que la TA se charge de la partie facile et qu'on chiffre celle-ci à 50 % du travail de traduction, cela permet de retrancher la moitié supérieure de la colonne 1. On obtient alors la colonne 2, qui représente le travail à effectuer par le traducteur qui utiliserait un système de TA; cette colonne représente la partie «incompressible» de son travail, dans lequel la machine est incapable de l'aider.

Face à ce gain de 50 % (l'actif de la TA), il faut maintenant déclarer le passif décrit plus haut, sous la forme d'une tâche de postédition alourdie par la correction de fautes archi-élémentaires et du charabia (hachures serrées) et de coûts système (hachures larges). Selon le système de TA employé (colonnes 3, 4 et 5), il est probable que le seuil de rentabilité (la barre des 100 %) sera dépassé par la combinaison de ces deux facteurs, dans des proportions variables selon que le système est coûteux et relativement performant (colonne 3), ou peu coûteux et moins performant (colonne 5).



Loin de moi la prétention d'affirmer (et encore moins de démontrer!) que la TA ne saurait être rentable. Ce qu'il importe simplement de mettre bien au clair, c'est que *le succès en TA est assujéti à de nombreuses contraintes* quant à l'application envisagée : volume de texte élevé, domaine plus ou moins restreint, syntaxe acceptable des textes LD, délai de mise en route suffisant, etc. Les textes qui répondent à toutes ces contraintes sont rares. Autrement dit, la marge d'applicabilité de la TA est mince et le recours à la TA comporte un *risque* assez élevé qui — avec raison — en décourage plus d'un.

C'est ce que j'appelle «le mur de l'applicabilité de la TA»; les briques de la non-rentabilité sont cimentées par les contraintes d'application trop sévères. Cette situation de piétinement risque de se prolonger, car l'évolution de la technologie actuelle ne laisse pas présager de meilleures conditions de rentabilité : une amélioration substantielle des performances des systèmes risque fort de s'accompagner d'une hausse des coûts, et l'on retrouve le cercle vicieux de l'actif et du passif de la TA — sous le plafond immuable de la partie incompressible du travail de traduction. Et pourtant, devant ce mur, les traducteurs sentent bien que devant les besoins de plus en plus pressants du marché, *il doit se passer quelque chose*.

N'aurions-nous pas intérêt à *penser différemment l'intervention technologique* dans le travail du traducteur? C'est de ce questionnement qu'est né le concept de prétraduction automatique.

3 La prétraduction automatique

Devant les limites assez largement reconnues de la TA, les chercheurs estiment volontiers qu'il faut revenir à la notion de THAM (traduction humaine assistée par la machine) en offrant au traducteur divers outils (notamment de consultation terminologique), tout en lui laissant le soin de créer lui-même sa traduction. C'est le concept, maintenant bien connu, du poste de travail du traducteur⁶.

Toutefois, ce faisant, on renonce en bloc aux avantages liés à la TA (voir en 1.1) — ce qui est quand même dommage car la transcription des invariants, l'assurance-oublis, les gains de frappe et le rappel systématique du vocabulaire sont loin d'être à dédaigner dans bien des situations. Ne pourrait-on pas trouver quelque part un compromis?

Dans la liste des avantages de la TA, si l'on y pense bien, celui qu'on sacrifierait le plus volontiers est la construction automatique de phrases en LA. En effet :

- Si l'on revient à la figure 1, on voit que cette fonction (qui cumule formulation littérale, analyse syntaxique et choix lexicaux élémentaires) correspond à ce qui est facile pour le traducteur humain (et encore hasardeux pour une machine). Dans l'hypothèse où le système de TA réussit cette tâche, il épargne au traducteur une opération que celui-ci aurait exécutée rapidement et sans grand effort.
- Par ailleurs c'est cette même fonction qui amène le risque de charabia, et donc d'alourdissement de la tâche du traducteur.

Comme on le voit, le gain net de productivité offert par cette fonction est loin d'être décisif; ne pourrait-on pas le sacrifier tout en conservant les autres avantages de la TA, c'est-à-dire le rappel lexical, la transcription des invariants et l'assurance-oublis? C'est possible... à condition d'oser envisager le concept de la PTA.

3.1 Principes de base

Une prétraduction est une *traduction partielle* du texte, par simple remplacement global d'une partie des mots du texte au moyen d'un dictionnaire d'équivalences. En voici un exemple :

⁶ Elliott Macklovitch, «An Off-The-Shelf Workstation for Translators», L'Actualité terminologique, vol. 3, n° 1, pp. 4-7.

Texte original

This subsection describes how unbalance is detected in the vibration spectrum, explains the different types of unbalance, and outlines the procedure for correcting unbalance. Methods of calculating the correction masses are described in subsections 5.2, 5.3 and 5.4, among which the vector diagram and the use of a programmable calculator and a balancing program.

Version prétraduite

This sous-section décrit how balourd est détecté dans le spectre_des_vibrations, explique les différents types de balourd, et outlines la méthode pour correcting balourd. méthodes de calculating les masses_d'équilibrage sont décrits dans sous-sections 5.2, 5.3 et 5.4, parmi which le diagramme_vectoriel et l' use d'une calculatrice_à_programme et d'un programme_d'équilibrage.

Cet échantillon de prétraduction, malgré son apparence échevelée, obéit à deux objectifs bien précis :

- Donner au traducteur *des éléments de vocabulaire utiles*, qui l'aideront à construire sa traduction. Cette utilité prend deux formes :
 - tantôt le rappel du vocabulaire spécialisé (consultation automatique du dictionnaire technique);
 - tantôt, simplement, un gain de frappe (vocabulaire général, invariants, etc.).
- Préserver la capacité du traducteur de *reconnaître instantanément le texte LD* en lisant le texte prétraduit. En effet, le traducteur qui lit l'exemple prétraduit *en pensant en anglais* n'a aucun mal à reconstituer la phrase d'origine. C'est sur cette découverte fondamentale que s'appuie la PTA.

L'équilibre entre ces deux objectifs quelque peu divergents est réalisé grâce à différentes règles de traitement :

- Un certain nombre seulement des mots du texte sont traduits; la priorité va évidemment au vocabulaire technique. *Tout mot dont la traduction risque de créer de la confusion est laissé en LD*. Quel intérêt y aurait-il à traduire des mots comme *since* (depuis que/puisque), ou encore des formes en -ing? La *couverture lexicale* (nombre mots traduits) est en fait variable au gré de l'utilisateur du système de PTA.

Parmi les mots dangereux, citons en particulier les mots homographiques (qui peuvent avoir plusieurs catégories grammaticales) : un système de PTA doit faire en sorte que de tels mots ne soient pas traduits si leur catégorie grammaticale est incertaine dans le contexte, car rien n'est plus déroutant que de voir surgir un verbe là où devrait y avoir un adjectif!

- Les équivalents de plusieurs mots sont «bloqués» par des traits de soulignement; le traducteur voit tout de suite que ces chaînes de mots sont bel et bien un terme, et non une juxtaposition fortuite.
- Aucun mot du texte LD n'est déplacé. L'ordre des mots, en effet, est très indicatif; le moindre changement risquerait de déstabiliser la lecture par le traducteur.
- Pour la même raison également, aucun mot n'est créé ni supprimé à la prétraduction.

3.2 Après la prétraduction : la «ptaduction»

Le premier moment de stupeur passé, il est remarquable de constater qu'une prétraduction est fort exploitable — c'est l'avis d'une bonne majorité des traducteurs qui ont fait l'expérience de la chose. Autant l'aide de la TA pour les traducteurs a quelque chose d'illusoire, autant la gêne que peut présenter à première vue la PTA n'est qu'apparente.

3.2.1 Ptaduction et postédition

Le traducteur, donc, affiche le texte prétraduit à l'écran de son micro-ordinateur et le manipule jusqu'à créer sa propre traduction. Cette opération ne saurait être qualifiée de «postédition», car ce mot implique qu'on corrige une traduction. Or une prétraduction n'est que le texte LD sur lequel on a plaqué des éléments LA; c'est l'étape préliminaire à la traduction proprement dite, qui est ici du ressort exclusif du traducteur. Devant le besoin d'un terme pour désigner le travail de traduction effectué à partir d'un texte prétraduit, j'ai forgé le mot «ptaduction» (télescopage entre PTA et traduction).

Il semble bien que la ptaduction soit plus alléchante pour les traducteurs que la postédition. À ce que j'ai déjà écrit à ce sujet⁷, j'ajouterai ici deux observations intéressantes :

- La figure 1 présente trois barres verticales à la droite du tableau : ces barres correspondent aux champs d'intervention de la PTA (barre A) et de la TA (barre B) ainsi qu'au domaine d'expertise du traducteur (barre C). Avec la TA, il y a chevauchement (pour ne pas dire brouillage) entre l'intervention de la machine et celle de l'humain. Par contraste, la PTA s'arrête prudemment au seuil de l'expertise du traducteur; cela rend la jonction plus harmonieuse et plus productive.

⁷ Claude Bédard, «La prétraduction automatique (PTA) : un pas en arrière dans la bonne direction?», Actes du colloque Les industries de la langue — Perspectives des années 1990, Montréal, 1991, pp. 435-448

- Autre facteur d'inconfort en postédition : d'une phrase à l'autre, le postéditeur fait face à quelque chose de différent (traduction réussie, traduction non idiomatique, traduction manquée à cause d'erreurs d'analyse, etc.). Avec la PTA, au contraire, *le ptaducteur a toujours affaire à la même chose* (le texte original «décoré» lexicalement); il n'a pas à choisir, après évaluation de la phrase, quel processus mental mettre en branle. Ce petit détail compte pour beaucoup, dans la pratique, à rendre «coulant» le travail de ptaduction.

Il est certain que la ptaduction demande (en principe) plus de changements à l'écran que la postédition. Mais leur quantité ne doit pas masquer le fait qu'ils se font plus rapidement — d'un seul jet, alors qu'en postédition le travail des doigts est remplacé par le travail mental (encore une illusion qui s'ajoute à celles déjà signalées; la PTA semble décidément avoir le don — quand cela serait son unique mérite — de révéler toutes sortes de failles dans les idées reçues en TA).

3.2.2 Le gain de frappe

Le gain de frappe varie largement; selon la couverture lexicale du système, la simplicité du style et la proportion d'invariants. Il faut créer les constructions de phrase appropriées, mais il peut arriver souvent que la PTA, par effet de symétrie entre les deux langues, se trouve à produire une traduction presque complète.

Dans_le_cas_d'un réseau_c.c. connecté to un réseau_c.a. through
un onduleur_à semi-conducteurs , le réseau_c.c. ne_doit_pas être
mis_à_la_terre_à_moins_que there_is un transformateur_d'isolement
dans le circuit.

Par hasard aussi, la reformulation par le ptaducteur peut rendre utilisable telle quelle une erreur (par exemple un verbe mal conjugué).

En somme, la prétraduction vise à rendre justice à la composante manipulative — qui existe bel et bien — en traduction, mais sans pousser l'intervention trop loin.

3.2.3 La transparence et ses limites

La laideur du texte prétraduit, il est remarquable de le constater, est invisible au ptaducteur, dont l'esprit s'évade instantanément dans le texte LD et le texte LA, sans s'attarder à la littéralité de la prétraduction. Une phrase comme celle-ci en est un bon exemple.

Care doit être pris lorsque installing câbles_chauffants below la surface à être heated.

Cela dit, il faut en toute honnêteté signaler certains facteurs limitatifs à la transparence.

Le mieux est souvent l'ennemi du bien; en PTA, il faut se méfier de l'application fortuite de certaines chaînes. Par exemple, en voulant traduire l'expression *there shall be no (breakage, failure, excessive wear...)* par *on ne doit constater aucun*, on risque d'obtenir :

On_ne_doit_constater_aucun robinets_d'arrêt situé entre le limiteur_de pression et le pressure-imposing élément.

Autre cas, où *required to comply with* est traduit par *tenu de satisfaire à*. Ici, chose plus grave, la prétraduction masque un double sens, ce qui peut induire une grave erreur à la traduction.

If a time-delay fuse is required to comply with clause 6.11.2...

Si un fusible temporisé est tenu_de_satisfaire_à l'article 6.11.2... [ici : est exigé conformément à...]

De tels abus montrent qu'un certain doigté (ou du moins une certaine retenue) est nécessaire en PTA. Cela dit, il existe certaines limites *inhérentes* à la transparence en PTA. Le principal facteur est celui des chaînes bloquées, lesquelles renferment souvent des inversions. Dans certaines situations, ces inversions produisent une confusion locale dans l'esprit du traducteur non averti. Exemple :

This requirement also applies to self-contained *coin, banknote, or credit mechanisms*.

This exigence aussi s'applique to self_contained *pièce_de_monnaie, billet_de_banque, ou mécanismes_de_crédit*.

Le changement de séquence *A, B or C D => A, B ou D_C* est source de confusion. En effet, les entités A, B et C (*coin, banknote et credit*) sont toutes trois reliées à l'entité D (*mechanism*); l'inversion de C et D dans l'équivalent *mécanismes de crédit* vient brouiller la lecture.

Par contre, un mauvais accord sujet-verbe ne nuit pas à la compréhension, car la contiguïté est une règle très simple et facile à garder à l'esprit.

The check valves required in this application shall be of an approved type.

Les clapets_de non-retour required dans this application doit être d'un approuvé type.

Outre la dimension structurale évoquée ci-dessus, il y a la dimension lexicale. Ici, il existe une marge de réglage de la prétraduction selon qu'on veut donner priorité au gain de frappe ou à la transparence. Dans les exemples suivants, le traducteur pourra tolérer une traduction parfois erronée, dans la mesure où :

- il bénéficie d'un gain de frappe dans la plupart des occurrences;
- l'équivalence est bien univoque, et révèle tout de suite son origine.

Si le normal éclairage on this plancher... [floor = étage]

Check that it est dans [in = dans] bon contact avec...

...par [by = en] protecting le matériau avec une couche_de_peinture.

3.2.4 La synopticit 

La PTA permet au traducteur de travailler «en synoptique», c'est- -dire que son regard ne quitte pas le texte pr traduit — qu'il lit d'abord comme texte LD avant de la transformer progressivement en sa traduction. La PTA se trouve   recrer quelque peu les conditions de la dict e (fort appr ci es des traducteurs),   quelques diff rences pr s :

- D'abord, un texte pr traduit n'est pas aussi instantan ment transparent qu'un texte enti rement en LD; cela appara t surtout lorsqu'on veut balayer de fa on al atoire le contexte suivant le passage   traduire.
- En revanche, le traducteur dispose imm diatement de la version traduite; il n'a pas   se reporter en arri re sur la bande magn tique, op ration souvent malcommode. Autre d tail, il peut retravailler tout de suite son premier jet, sans avoir   attendre qu'il revienne de la transcription.

Dans la pratique, le traducteur produit un premier jet en synoptique (sur un paragraphe   la fois, par exemple), puis confronte   loisir les versions LD et LA pour bien contr ler le sens et raffiner sa traduction.

3.3 Une approche souple et tol rante

Lorsqu'on compare la PTA et la TA, on est frapp  par la grande *tol rance* de la PTA   diff rents  gards et (c'est presque un corollaire) la *marge d'initiative* qu'elle laisse   l'utilisateur.

3.3.1 Tolérance à la formule de travail

Nous avons vu que la couverture lexicale en PTA est variable. Elle l'est, en fait, au point où l'on peut abandonner toute idée de gain de frappe et se concentrer sur l'aspect «rappel du vocabulaire» : seul le vocabulaire technique est alors traduit, et il devient alors tout à fait possible de *dicter* à partir de la prétraduction — qui n'est alors rien d'autre que le texte LD «décoré» çà et là des équivalents de vocabulaire technique. Les termes prétraduits sont alors affichés (ou imprimés) en caractères gras, pour être repérables plus facilement.

This **sous-section** describes how **balourd** is detected in the **spectre_des_vibrations**, explains the different types of **balourd**, and outlines the method for correcting **balourd**. Methods of calculating the **masses_d'équilibrage** are described in **sous-sections** 5.2, 5.3 and 5.4, among which the **diagramme_vectoriel** and the use of a **calculatrice_à_programme** and a **programme_d'équilibrage**.

Aucun traducteur ne peut raisonnablement se dire rebuté par une telle méthode; à l'extrême limite, il peut se servir du texte prétraduit comme d'une simple référence, parallèlement à la version LD intégrale. Cette souplesse est très appréciée s'il faut former une équipe de projet dont certains membres préfèrent travailler en traitement de texte et d'autres au dictaphone.

3.3.2 Tolérance à l'entrée

La hantise des systèmes de TA est la «qualité» du texte LD (erreurs de frappe ou de grammaire) — terme ambigu et volontiers employé de façon malhonnête pour désigner des phénomènes tout à fait licites (longueur des phrases, syntaxe inhabituelle, style télégraphique...). Or les textes qui à première vue se prêteraient bien à un traitement automatisé (en particulier les manuels techniques) se révèlent plus souvent qu'autrement d'une syntaxe rébarbative à un traitement linguistique conventionnel. En voici un exemple, parmi bien d'autres⁸ :

This window contains an HP-UX shell (either a Bourne shell or C-shell, depending on the value of the SHELL environment variable; for details, see the "Concepts" section of the "Using Commands" chapter).

En PTA, ce facteur a peu d'importance puisque l'utilité de la sortie ne dépend pas du succès d'une analyse linguistique, et que dans le doute le système s'abstient. Les quelques erreurs qui surviennent malgré tout sont assez rares et de portée très limitée.

⁸ Exemple cité par Masaru Tomita, «"Linguistic" Sentences and "Real" Sentences», dans Coling Budapest, Budapest, 1988, p. 448.

Cette tolérance est bien illustrée par l'exemple (authentique) suivant, où le texte LD (à cause d'une mauvaise conversion électronique) était émaillé d'espaces manquants :

Moteurs sont requis to beguarded par l' utilisation d'un boîtier ou par locating le electricalequipment out_of normal reach , to enlever la possibilité ofdangerous contact avec pièces_sous_tension_à_découvert . Commutator brushes andthe like sont situé inside un moteur end bracket et sont notconductively connecté to the supply circuit.

Le résultat reste malgré tout fort exploitable; le seul inconvénient est que certains termes n'ont pas été traduits à cause des variantes d'écriture ainsi créées.

3.3.3 Tolérance à la sortie

La TA a des exigences de sortie très strictes. Ainsi, la traduction brute doit :

- être intégrale (tous les mots du texte doivent figurer au dictionnaire);
- ne pas comporter de charabia (l'analyse doit être parfaitement réussie), car la postédition deviendrait alors instantanément contre-productive;
- comporter le moins possible de «mal-dit» (d'où le besoin d'une foule d'entrées «linguistiques» au dictionnaire).

En un mot, *la TA doit tout réussir*.

La PTA, par contraste, a *un mandat beaucoup moins exigeant*; elle se définit au fond comme étant «toute annotation plaquée sur un texte, dont le résultat est jugé utile par le traducteur». Elle a fort peu de choses à «réussir» : en cas de doute, on ne traduit pas. S'il y a erreur, les conséquences sont très limitées car les remaniements syntaxiques dans la version prétraduite sont réduits au minimum.

Je m'étonne toujours, même après quelques années de pratique, de constater à quel point la prétraduction demande peu de chose au système. Reportons-nous à l'exemple cité en 3.1 :

- Le mot *méthodes* en début de phrase n'a pas de majuscule. C'est une exception voulue, car un nom pluriel en début de phrase demande normalement un article en français; s'il était traduit avec une majuscule, cela ajouterait aux manipulations par le traducteur.
- Par ailleurs, à quoi bon s'acharner à prétraduire des mots comme *how*, *which*, et même *this*? On n'économise pas grand-chose au traducteur — et l'on prend des risques — en le faisant. Par exemple, traduire correctement *among which* par *parmi lesquelles* demande à un système de TA beaucoup d'efforts... dont le traducteur annulera probablement l'effet en remplaçant le tout par *notamment*.

- Dans l'exemple cité en 3.3.4, on remarquera que certains passages entre unités de substitution (p. ex. MAIS WILL MELT AS IT MOVES TO) restent pratiquement en friche; cela n'enlève rien à l'utilité de la prétraduction. En fait, si l'on tentait de bien prétraduire ces passages, on obtiendrait probablement un rendement décroissant pour l'ensemble de l'opération.

Il faut le répéter, en matière d'aide informatique au traducteur, le mieux (tentation constante des développeurs) est souvent l'ennemi du bien.

3.3.4 Variété des stratégies

Un autre avantage de la PTA, c'est qu'elle se prête à des stratégies de travail variées; c'est là un point important, car les situations à traiter dans la pratique sont elles-mêmes fort variées.

- C'est ainsi que nous avons vu que la «PTA pour dictée» diffère sensiblement de la «PTA pour traitement de texte» (à beaucoup plus large couverture). On peut donc jouer considérablement sur la *couverture lexicale*.
- En outre, la *longueur des unités de substitution* est très variable. Tel texte peut se prétraduire à coups de chaînes de deux ou trois mots, réparties assez uniformément. Dans l'exemple suivant, les chaînes sont plus longues :

AS IT MOVES TO LE_SUD_DU CAP_BONAVISTA. ON_PREVOIT_PEU_DE_CHANGEMENT
 AU_NORD_DU CAP_FREHEL ET DANS LE_SUD_DE LA_BAIE_NOTRE_DAME.
 DESINTEGRATION_RAPIDE_DES_GLACES.

Enfin, certaines applications peuvent consister à traduire en bloc des phrases ou de longs segments de phrase, et à laisser intraduits les passages intermédiaires⁹.

- Enfin, on peut jouer sur le plan des *étapes successives*. Par exemple, au lieu de travailler au maximum le dictionnaire de prétraduction, on peut laisser un certain nombre de choix pour l'étape de la ptaduction : le traducteur a alors la ressource de lancer des remplacements globaux sur le texte prétraduit. (Il a la même ressource en TA, mais le texte a été tellement remanié par le système que le résultat en est moins prévisible. En outre, l'existence de chaînes «bloquées» permet aux remplacements de s'appliquer de façon plus sélective.)

En somme, on voit que la PTA laisse beaucoup de place à l'inventivité du traducteur face à des situations variées. Elle permet notamment d'exploiter de façon sélective ce qui est automatisable dans un texte donné; la TA, par contraste, n'offre qu'une seule formule d'automatisation et l'applique indifféremment sur toutes les composantes du textes — y compris celles où elle fait plus de tort que de bien.

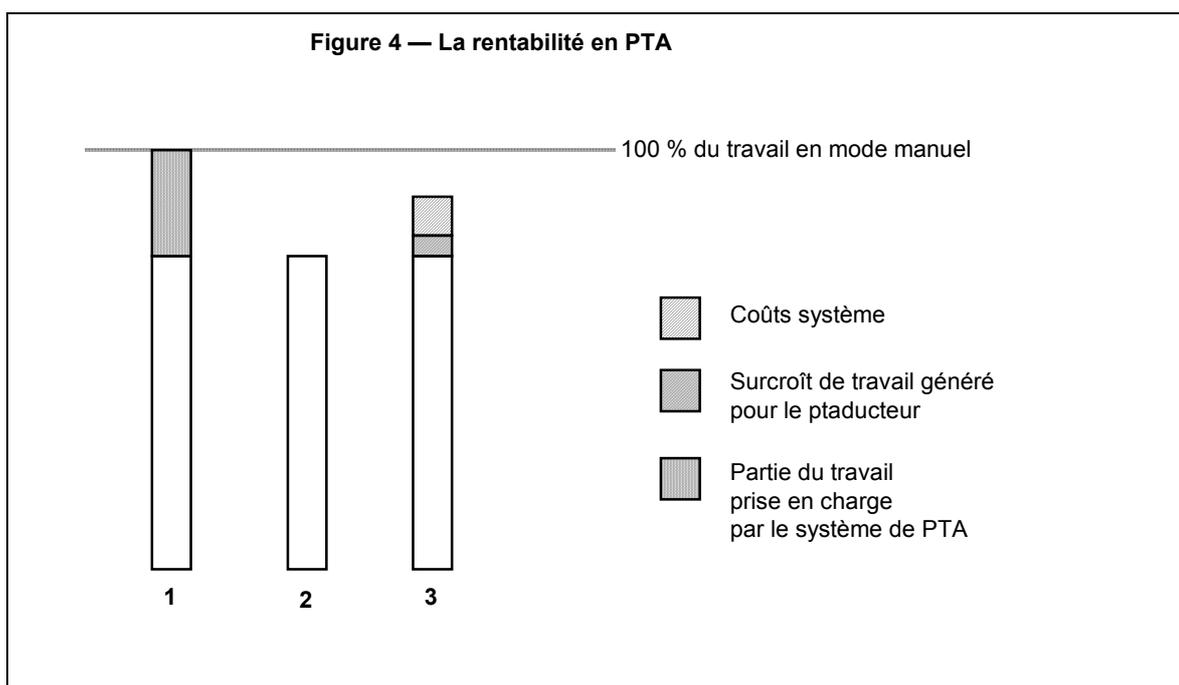
⁹ C'est effectivement la stratégie adoptée par le système Général TAO, développé par John Chandioix pour La Confédération-Vie.

3.3.5 Une solution rapide et légère

Une des grandes qualités de la PTA est de se prêter à une mise en route rapide. Une argumentation qu'on entend souvent en TA est la nécessité d'investir à long terme et d'attendre patiemment les résultats. C'est ce que j'appelle le «syndrome de la galère» — c'est-à-dire qu'on s'embarque dans une entreprise longue, hasardeuse et coûteuse. Cette attitude a deux inconvénients :

- Plus l'investissement est lourd, plus il est difficile à récupérer, car nous avons vu que les gains de productivité escomptables sont au fond limités.
- Il est fort possible qu'une fois «les résultats» atteints, les circonstances ou les besoins auront changé, de telle sorte que ces résultats ne seront plus vraiment exploitables.

Face à cette attitude, la PTA propose des gains limités (tels, en réalité, qu'ils sont concevables) mais réels, et à court terme. Si l'on reprend le graphique de la figure 3 dans le contexte de la PTA, on obtient plutôt l'image suivante :



Avec au départ un gain plus modeste (et plus réaliste) de, disons, 25 %, la PTA réussit à minimiser les coûts système ainsi que le brouillage à la ptaduction de manière à dégager une marge de rentabilité mince, mais réelle. En somme, on reconnaît le proverbe «Un tiens vaut mieux que deux tu l'auras».

3.4 Quelques observations

Pour conclure cette description de la PTA, certaines observations plus générales permettront de mieux situer la PTA sur l'échiquier de la TAO.

- En PTA, le texte n'est pas l'objet d'un *traitement linguistique*, mais de *manipulations purement pragmatiques*, dictées par ce que le traducteur considère lui être utile dans la situation donnée.
- Les efforts d'un utilisateur de système de TA visant à améliorer la qualité linguistique de la sortie sont souvent faussés par le «filtre» que constituent les grammaires linguistiques du système. En PTA, le lien est beaucoup plus direct entre ce que l'utilisateur fait et ce qu'il obtient : ses initiatives sont récompensées directement, ce qui l'encourage dans la voie de l'inventivité.
- La version prétraduite n'est pas jugée en tant qu'objet linguistique, mais en tant que facteur d'efficacité et de confort pour le traducteur. La PTA conteste donc la priorité traditionnellement accordée aux critères linguistiques dans la mesure du degré d'aide au traducteur.
- Corollaire du point précédent : contrairement à la TA, l'intérêt de la PTA réside dans ce que la machine *ne fait pas*. Cette affirmation a de quoi surprendre ceux qui considèrent que la solution pour le traducteur réside dans la maximisation de l'intervention technologique. En fait, la PTA met moins l'accent sur la *technologie* que sur la *méthodologie* — c'est-à-dire sur une compréhension plus fine des processus mentaux du traducteur professionnels et sur l'évolution de ses méthodes de travail.

4 Intérêt et limites de la fonction «rappel»

La fonction «rappel» étant la base de l'intérêt de la PTA, il convient d'examiner plus en détail la question de son efficacité. En fait, l'intérêt d'un tel examen déborde radicalement le sujet de la PTA, et rejoint la question de la gestion terminologique active¹⁰, ou la terminologie de projet. Les concepts établis ici seront d'ailleurs utiles dans la section suivante.

J'aimerais établir ici une distinction fonctionnelle importante entre la *recherche*, la *consultation* et le *rappel*.

- La *recherche* implique une variété de sources, terminologiques ou bibliographiques (fiches ou textes); des décisions sont nécessaires pour la navigation dans les diverses sources.
- La *consultation* s'applique à une source terminologique (par exemple la banque de données TERMIUM), dont on doit déterminer si le contenu s'applique à la situation.

¹⁰ Claude Bédard, «Machine PreTranslation : A Further Step in Terminology Management», The ATA Chronicle, janvier 1991 pp. 19-20.

- Le *rappel*, enfin, est une consultation univoque d'une source d'équivalences, c'est-à-dire dont on est certain que le résultat s'applique à la situation. Le rappel peut être systématique (la consultation se fait automatiquement) ou simplement à la demande (le traducteur doit amorcer lui-même l'opération pour chaque terme).

4.1 Utilité du rappel systématique

La PTA, par l'action de son dictionnaire automatique, opère un *rappel systématique* du vocabulaire pertinent au texte. Cette forme de rappel est particulièrement utile dans trois situations:

- Pour les débutants, ou encore les traducteurs qui connaissent mal le vocabulaire du domaine. On se familiarise beaucoup plus vite avec le vocabulaire du domaine.
- Pour les pigistes. On sait à quel point il est difficile d'encadrer les pigistes (leur faire employer le vocabulaire, la phraséologie du client...). En leur fournissant un texte prétraduit, le donneur d'ouvrage résout largement ce problème — et peut même penser à négocier le tarif dans la mesure où le texte prétraduit rend le pigiste plus productif.
- Pour le travail en équipe, la PTA est très commode, car elle assure l'uniformité terminologique de base. Souvent, dans un projet de traduction particulier, bien des termes ont un équivalent spécial; les traducteurs expérimentés ont déjà leurs propres équivalences en mémoire et ne pensent pas à consulter le lexique du projet.
- En général, même pour un traducteur habitué à un vocabulaire particulier, une prétraduction est un bon aide-mémoire. Elle permet d'absorber les changements (évolution temporelle) de vocabulaire, et de compenser pour chaque client d'un même domaine.

Cela dit, le rappel systématique reste foncièrement contestable par le traducteur : celui-ci doit en tenir compte, mais ne pas se dérober pour autant à son jugement professionnel. La PTA vise à informer le traducteur, non à l'infantiliser.

4.2 Les textes à vocabulaire «compact»

La PTA s'adresse aux textes «à vocabulaire» (j'inclus dans ce terme tous les éléments répétitifs, y compris la phraséologie). On reconnaît ce genre de texte par différentes caractéristiques :

- La *densité* terminologique, c'est-à-dire la proportion des mots du texte qui font partie du vocabulaire. Un texte de 8000 mots dont 4000 mots de vocabulaire technique aura une densité terminologique de 50 %.

- La *répétitivité* du vocabulaire. Mieux vaut un petit nombre de termes répétés souvent qu'un grand nombre de termes répétés rarement : l'effort d'établissement du dictionnaire machine en rapport au gain à la traduction en dépend. On peut exprimer cette notion numériquement par le nombre moyen d'occurrences des termes dans le texte. Dans un texte dont les occurrences sont le plus souvent uniques, la répétitivité sera voisine de 1; à l'autre extrême, elle pourra atteindre des valeurs de 5, voire 10 ou 20.
- La *stabilité graphique* du vocabulaire en LD. Dans certains cas, il y a instabilité à cause de multiples variantes d'écriture des termes en LD, à tel point que le texte peut être inexploitable.
- La *stabilité sémantique* du vocabulaire en LD. Il arrive que l'*emploi* des termes dans le texte LD soit confus ou instable sémantiquement : l'occurrence d'un terme déclenche (ou du moins le devrait) une réaction de méfiance de la part du traducteur. Cet aspect est difficile à quantifier.
- L'*univocité* du vocabulaire (ou son inverse, la *polysémie*). Le vocabulaire de spécialité n'est que partiellement univoque¹¹; la polysémie y est courante. Si *transistor* en anglais se traduit invariablement par *transistor* en français, le mot *power* (même dans le domaine restreint de l'électricité) peut se rendre par *puissance, courant, énergie, alimentation, force motrice, etc.*, selon le contexte; dans le même domaine, *insulation* se rend soit par *isolation, isolant* ou *isolement*. La PTA, qui propose les équivalents à même le texte, se prête mal aux choix multiples; un fort pourcentage d'univocité (rapport entre les termes univoques à l'ensemble des termes) est nécessaire.

On pourrait exprimer l'ensemble de ces caractéristiques en disant que tel vocabulaire est «compact».

Les deux premiers points n'ont qu'une incidence quantitative sur le temps de préparation du dictionnaire machine en rapport au volume de texte à traiter, et se passent de commentaires. Les trois autres points, par contre, soulèvent des questions méthodologiques intéressantes.

4.2.1 En cas d'instabilité graphique

On peut agir jusqu'à un certain point sur l'instabilité graphique. Rappelons ici quelques-unes de ses manifestations :

- La présence ou l'absence de trait d'union (*single-pole switch / single pole switch*).
- Le soudage ou non-soudage de deux mots (*fuse holder / fuseholder*).

¹¹ Pour un exposé complet sur les diverses lacunes du vocabulaire technique, voir Claude Bédard, *La traduction technique : Principes et pratique*, Montréal, Liguatex, 1986, pp. 8-22.

- L'emploi aléatoire des terminaisons de modifieurs (*regulator device / regulating device / regulation device*).
- L'emploi d'éléments variables, souvent numériques (*in row X of column Y*), à l'intérieur de chaînes répétitives.
- La présence ou l'absence de certaines ponctuations d'une occurrence à l'autre.
- Les fautes de frappe ou d'écriture (*affect / effect, renumeration / remuneration*).
- Les variantes orthographiques entre l'anglais britannique, canadien et américain (*color / colour, analyze / analyse*), qui peuvent se manifester d'un texte à l'autre, voire à l'intérieur d'un même texte.
- L'emploi du tiret (simple ou double, collé ou non).
- La disjonction des termes. Ce phénomène se produit dans le cas d'une construction avec coordonnant (and, or). Par exemple, l'occurrence *step and touch voltages* cache en fait deux termes : *step voltage* (dont la reconnaissance est masquée dans ce contexte) et *touch voltage*.
- Les variantes aléatoires dans l'écriture de chaînes en principe répétitives. Certains textes sont écrits par copiage fidèle, par traitement de texte, de segments répétitifs; dans d'autres cas, ces segments sont reformulés à chaque occurrence, avec un potentiel variatoire quasi infini.

Cette liste (non exhaustive) des facteurs d'instabilité graphique est assez impressionnante, et s'applique d'ailleurs autant à la TA qu'à la PTA, les deux opérations étant basées sur un appariement de formes (pattern matching) littéral.

Le remède le plus simple consiste à consigner dans le dictionnaire d'équivalences du système les variantes d'écriture. Toutefois, cette opération peut devenir fort onéreuse, car l'instabilité graphique a souvent un effet multiplicateur (par exemple la variante *fuse holder / fuseholder* peut se retrouver dans des dizaines de termes ou d'expressions); il faut alors rechercher d'autres procédés.

- La *prédiction automatique* consiste à constituer un dictionnaire qui s'applique sur le texte LD avant l'étape de la PTA, de manière à éliminer un certain nombre de variantes.
- Le *marquage d'un mot représentatif*. Si un passage assez long (toute une phrase, par exemple) est répétitif mais de formulation variable, on peut marquer une courte chaîne de mots très représentatifs au moyen d'un symbole qui renvoie à un lexique de référence (voir 4.2.2), lequel cite la phrase complète telle qu'elle doit être traduite. Exemple :

Version prétraduite

definitions_are_generally placé dans section 0. Cependant, un mot ou term, ou son derivative, applying only to one section est définie dans that section.

Traduction de référence

La plupart des définitions sont regroupées à la section 0; toutefois, les termes ou dérivés qui s'appliquent à une seule section sont définis à l'intérieur de la section en cause.

La double oblique inverse invite à chercher une entrée comportant les mots *definitions are generally* dans le lexique de référence. Comme on peut le voir, la lutte aux variantes peut prendre plusieurs visages.

4.2.2 En cas de polysémie ou d'instabilité sémantique

Quant à la *polysémie*, elle se manifeste plus ou moins selon qu'on a affaire à deux situations assez différentes.

- Beaucoup de textes se réfèrent à l'ensemble du vocabulaire disponible dans un ou plusieurs domaines (par exemple, un article de revue sur le développement de la robotique au Canada), sans convention particulière; la porte est alors ouverte à tous les flottements dans l'usage — et ils sont nombreux, même dans les domaines techniques⁵. Dans ce cas, il est plus difficile de prévoir quel sera l'emploi du vocabulaire, et donc de proposer à l'avance des équivalents adéquats. On pourrait dire que le vocabulaire de ces textes est «ouvert».
- D'autres textes, au contraire, se réfèrent à un corps de vocabulaire assez discipliné : par exemple le manuel des politiques internes d'une société emploie la terminologie privilégiée par cette société; ou encore un manuel d'utilisateur de logiciel emploie un vocabulaire moulé sur le réseau notionnel du logiciel en question. On pourrait qualifier ce vocabulaire de «fermé». Signalons aussi un détail important : il s'agit d'un *vocabulaire imposé*, c'est-à-dire que le traducteur, quelles que soient ses propres préférences ou habitudes quant aux équivalents à employer, doit s'astreindre à des emplois auxquels il n'est peut-être pas habitué. Ce genre de situation est en nette croissance.

Le *rappel* terminologique (donc la PTA) se prête mal aux textes à *vocabulaire ouvert*, pour plusieurs raisons :

- Nous avons vu que l'emploi du vocabulaire est assez imprévisible dans ce cas; le traducteur a davantage besoin de consulter librement des sources terminologiques assez complètes (lexiques variés, banques de terminologie, etc.), et d'estimer par lui-même à quel point les

équivalents proposés conviennent à la situation. En somme, il ne s'agit pas tant ici de *rappel* que de *consultation* terminologique, avec le travail de *décision* que cela implique souvent.

- Le traducteur qui a l'expérience du domaine connaît spontanément une grande partie du vocabulaire employé; il peut en outre laisser libre cours à ses préférences terminologiques.
- Enfin, une situation de vocabulaire ouvert n'offre pas souvent une répétitivité et une densité de vocabulaire intéressantes.

Ce sont donc les textes à vocabulaire fermé qui sont le plus favorables à la PTA, car la notion de *rappel* prime largement sur celle de *consultation*.

Cela dit, la polysémie a toujours une certaine incidence dans les textes à vocabulaire fermé. Comment traiter ce phénomène en PTA, qui se limite à une *annotation à plat* (sans fenêtre séparée; voir en 5.4)? La marge de manoeuvre à cet égard est mince, mais elle existe. On peut imaginer trois procédés :

- Énumérer les équivalents, en les séparant d'un symbole spécial (*isolation//isolement*). Bien entendu, ce procédé est lourd si le terme est fréquent; en outre, il est difficilement tolérable au-delà de deux équivalents. C'est pourquoi l'on préférera souvent des solutions plus légères, comme les suivantes.
- Citer l'équivalent le plus fréquent, en le décorant d'un symbole spécial (*isolement\|*). Ce symbole signifie que le terme est polysémique et qu'un complément d'information se trouve dans un lexique spécialement créé à cet effet. Ce lexique regroupe tous les cas de polysémie; il indique les équivalents possibles et explique comment choisir entre eux. Les entrées de ce lexique ont comme vedette le mot cité dans la prétraduction.
- Citer non pas l'équivalent, mais le terme LD, décoré du même symbole (*insulation\|*), avec le même système de renvoi. Ce procédé est préférable au précédent si les équivalents concurrents sont fortement contrastés et qu'on risque de perdre la transparence à la lecture. Par exemple, dans le cas d'une équivalence double comme *service => service* ou *branchement*, on ne veut évidemment pas voir apparaître le mot *branchement* dans un contexte comme «To provide a safe and satisfactory level of service...».

On le voit ici, la fonction annotation est assez puissante et offre un large terrain d'initiative à l'utilisateur, mais exige en même temps du doigté. Il s'agit donc de trouver le meilleur compromis possible entre la facilité de manipulation du texte et l'information vitale. Le traducteur ne doit pas être lassé par des annotations trop lourdes. N'oublions pas qu'il a toujours la possibilité de commander un remplacement global afin d'abolir ou de simplifier telle ou telle annotation (*isolation//isolement* remplacé par *isolation\|*).

5 La PTA par rapport aux autres formules de travail

La PTA a-t-elle sa place parmi les technologies et méthodes actuellement offertes au traducteur? La réponse réside largement dans les pages qui précèdent, mais il est utile de présenter ici la synthèse des grandes comparaisons qu'on peut vouloir faire.

5.1 Par rapport à la TA

À ce qui a déjà été dit en 3.3, ajoutons le commentaire suivant. Dans un récent article¹² constatant les limites traditionnelles de la TA, le chercheur Kenneth Church propose les caractéristiques souhaitables d'une «niche favorable» en TA. Ces critères ont ceci de frappant qu'*ils vont comme un gant à la PTA* — une formule dont l'auteur ignorait pourtant l'existence. Examinons ces critères un à un :

«it should set reasonable expectations»

La PTA vise simplement à aider le traducteur à produire son premier jet, avec gain de frappe et de rappel terminologique — rien de plus (voir la figure 4, colonne 3).

«it should make sense economically»

La PTA ne met pas en jeu une technologie onéreuse; son principe est au contraire de limiter le coût du traitement machine (étant donné les gains de productivité limités qu'on peut attendre du traducteur, dont le travail est en partie «incompressible»), afin de dégager une marge de rentabilité modeste, mais réelle.

«it should be attractive to intended users»

La ptaduction, selon mon expérience et celle d'autres collègues, est une activité nettement plus attrayante que la postédition. En outre, n'oublions pas la ressource de la «prétraduction légère», pour dictée.

*«it should exploit the strengths of the machine
and not compete with the strengths of the human»*

Comme nous l'avons vu en 3.2.1, la PTA n'empiète pas sur le domaine d'intervention du traducteur — contrairement à ce qui se passe en TA, avec les résultats désagréables qu'on connaît. La PTA a la lucidité de concéder d'entrée de jeu au traducteur l'exclusivité pour ce qui est de la partie «intelligente» de sa tâche.

*«it should be clear to the users
what the system can and cannot do»*

¹² Kenneth Church, «Good Applications for Crummy Machine Translation» (à paraître).

Les règles de fonctionnement de la PTA sont très simples et faciles à comprendre; lorsqu'il demande quelque chose au système, l'utilisateur obtient directement l'effet recherché étant donné la nature manipulatoire de la PTA.

*«it should encourage the field to move forward
toward a sensible long-term goal»*

Étant donné son seuil d'applicabilité très bas, la PTA est une occasion rêvée pour inciter les traducteurs à améliorer leurs méthodes de travail et à mettre leur inventivité au service du développement de la technologie qui les concerne (voir la section 6).

5.2 Par rapport à la traduction au dictaphone

Le concurrent le plus redoutable à la formule PTA reste bel et bien le traducteur d'expérience armé d'un dictaphone.

Cette réalité de base est souvent masquée par les chiffres de gains de productivité annoncés pour des projets de TA, car on y adopte souvent le raisonnement trompeur suivant : on compare la productivité obtenue avec le système (sur un texte favorable, donc répétitif et assez facile) à la productivité générale du service sur l'ensemble des textes qu'il a à traiter (dont chacun sait que la difficulté moyenne est sensiblement plus élevée) — ce qui donne invariablement des chiffres encourageants pour la TA. En fait, si l'on comparait la productivité des postéditeurs à celle de traducteurs travaillant normalement *sur le même texte*, les résultats seraient fort différents. Ce genre de comparaison masque le fait que dans les situations qui se prêtent à la TA (ou à la PTA), la traduction manuelle reste fort performante.

- En situation de *vocabulaire ouvert*, la PTA n'est guère concurrentielle, car il y a peu à attendre du rappel terminologique et du gain de frappe.
- Par contre, si le rappel joue un rôle important, la PTA est certainement concurrentielle — d'autant plus que la prétraduction légère pour dictée permet de combiner au besoin les avantages des deux méthodes.

Ouvrons ici une petite parenthèse. On pose parfois la question de la «pollution mentale» causée par la PTA (à cause du travail sur un texte «franglais»). Cette pollution me semble faible, en tout cas par rapport à la TA. En traduction, on ne fait que penser plus ou moins simultanément dans les deux langues — ce que le traducteur fait déjà de toutes façons, par exemple quand il dicte. La «fausse traduction» déguisée en vraie traduction fournie par un système de TA me semble plus pernicieuse à cet égard. Enfin, il faudra bien de toutes façons regarder en face et accepter le défi de la pollution en TAO — tout comme on l'accepte dans la vie de tous les jours. Il est difficilement concevable pour les traducteurs de relever les défis d'un avenir qui s'annonce très chargé sans accepter certaines mutations professionnelles. La PTA me semble d'ailleurs un lieu

passablement «écologique» à cet égard, et où le rapport entre les gains potentiels et la pollution possible est assez favorable.

5.3 Par rapport au rappel sur demande

En quoi la PTA est-elle plus efficace que le simple emploi d'outils de consultation terminologique, comme les grandes banques de données (TERMIUM, BTQ) ou les gestionnaires de bases de données terminologiques comme PROTERM, ÉDIBASE ou TERMEX?

En situation de *vocabulaire ouvert*, le rappel sur demande qu'offrent ces outils est mieux indiqué, car il s'agit alors davantage d'un processus de *consultation* (avec décision).

En situation de *vocabulaire fermé*, le rappel systématique a l'avantage essentiel de mieux encadrer le traducteur en attirant son attention sur l'existence des équivalences préétablies. S'il doit lui-même penser à consulter le lexique, le traducteur se trouve confronté au cercle vicieux suivant :

- Tantôt il consulte sans rien trouver, d'où perte de temps et diminution du réflexe de consulter.
- Tantôt il néglige de consulter, soit parce qu'il est loin de se douter qu'il existe une entrée dans le lexique, soit par crainte de perdre son temps (surtout après un certain nombre d'interrogations infructueuses).

Enfin, le rappel sur demande n'offre aucun gain de frappe au traducteur.

5.4 Par rapport au rappel multifenêtres

La PTA opère une annotation linéaire, unidimensionnelle, à même le texte; cette annotation est limitée par des considérations d'encombrement visuel — et c'est ce qui amène à prévoir un système de renvois pour consultation à la demande (voir 4.2.2).

Les systèmes comme TSS (mode AutoTerm), UNITRAN et IBM Translation Tool combinent l'avantage d'un rappel systématique et d'un affichage multiple (une fenêtre pour le texte LD, une autre pour la version traduite et une autre encore pour le rappel terminologique). L'emploi d'une fenêtre séparée permet d'afficher plus d'information, notamment des choix multiples.

Si le vocabulaire est monosémique, la PTA reste très compétitive. Par contre, moins le vocabulaire est monosémique, plus de tels systèmes sont avantageux par rapport à la PTA, à cause de leur capacité d'affichage supérieure. Mais ils ont aussi certains désavantages :

- Leur complexité technologique et les coûts inhérents.
- Le fait que chaque utilisateur doit disposer d'un système : le coût augmente donc à mesure qu'on étend l'utilisation du système.

- Un certain manque de souplesse. Le multifenêtres limite la quantité de texte affichable; en outre, les déplacements dans le texte peuvent être limités. L'utilisateur est enchaîné au système — alors que la PTA permet le travail en traitement de texte ordinaire, voire la dictée à partir d'une version papier. Cet inconvénient devient encore plus sensible si l'on pense à employer du personnel extérieur (pigistes).
- L'environnement est fermé, et laisse donc peu de place à l'inventivité du traducteur et à l'adaptation aux situations spéciales.
- De tels systèmes n'offrent pas un gain de frappe aussi poussé que la PTA.

Remarquons encore ici que les arguments avancés sont d'ordre éminemment pratique. C'est sur ce plan — le répétera-t-on jamais assez — que doit se faire actuellement la réflexion sur le sujet des outils informatiques des traducteurs.

6 Perspectives

Tout traducteur maîtrisant bien le traitement de texte est capable, au moyen de macrocommandes, d'obtenir un système de PTA rustique. D'ailleurs, répétons-le, la PTA est au départ une *méthode* de traduction plutôt qu'une *technologie*.

Toutefois, cette méthode a évidemment besoin d'un certain soutien technologique. C'est pourquoi, j'ai voulu élaborer un système relativement puissant, très paramétrable et conçu pour maximiser le confort et l'efficacité d'utilisation.

6.1 Une réalisation concrète : ATAO

Ce système¹³, nommé ATAO (L'Atelier traductique), comprend les modules suivants :

- Des outils de dépouillement, qui aident à constituer rapidement un dictionnaire à partir du ou des textes à traiter.
- Un concordancier, appelé KWSIEVE (développé par Michel Laliberté), qui donne accès à de multiples contextes à partir des listages de dépouillement.
- Une base de données (conçue à partir du logiciel PROTERM, développée par Jean-Louis Ronvaux et François Rousse), qui gère le dictionnaire de PTA au moyen des mêmes fiches qui reçoivent l'information terminologique usuelle aux fins de consultation manuelle : *une seule maintenance est nécessaire*. Autrement dit, on peut créer un dictionnaire de PTA à

¹³ Dont la réalisation doit beaucoup au langage de programmation GramR, créé par John Chandiooux.

même un fichier terminologique usuel. C'est un point de départ dans l'intégration tant souhaitable des fichiers terminologiques et des dictionnaires machine.

- Un module de préédition, paramétrable par l'utilisateur.
- Un moteur de prétraduction, qui effectue notamment la lemmatisation (réduction des mots à leur forme de base non fléchie) et une résolution partielle de l'homographie, plus un certain nombre de flexions sur le texte de sortie.
- Un pavé de commandes de navigation permettant différentes actions sur le texte (par exemple des inversions de mots), qui optimise l'ergonomie du travail de ptaduction.

Ce système, opérationnel depuis quelques années, m'a permis de faire la démonstration quotidienne de la faisabilité et des avantages de la PTA.

6.2 Une invitation aux praticiens

La PTA voit le jour dans un désert quasi total pour ce qui est de l'étude de la démarche de travail des traducteurs professionnels. Il n'est pas étonnant que ses principes de base apparaissent révolutionnaires — voire iconoclastes. En fait, les réactions d'étonnement ou d'incrédulité qu'on peut avoir face à la PTA ne font que révéler la minceur de notre perception de ce qu'est l'acte de traduire.

Il ne faut pas s'attendre à ce que la PTA soulève beaucoup d'intérêt chez les linguistes, partisans traditionnels de la TA : non seulement ils n'y trouveront pas matière à alimenter les activités propres à leur discipline (l'intervention linguistique en PTA étant en effet minimale), mais ils sont en outre mal placés pour apprécier les avantages de la PTA, qui sont essentiellement d'ordre *opérationnel* (gains de frappe, gestion terminologique active, rationalisation du travail dans les projets de traduction¹⁴, etc.).

Les traducteurs professionnels, en revanche, ont tout intérêt à s'intéresser à la PTA et sont bien placés pour en apprécier les avantages. Pour eux, elle offre à la fois la perspective de gains de productivité (modérés mais réels) et aussi une excellente occasion d'entreprendre une démarche de systématisation de leur méthodes de travail (la productivité ne découle pas nécessairement du fait de *travailler plus vite*, mais de la capacité de *travailler mieux*), laquelle les rendra en retour mieux préparés à faire usage de la technologie disponible, quelle qu'elle soit¹⁵.

¹⁴ Voir The production team of Rocky Mountain Translators, «The Team Approach to Large-Volume Documentation Translations» et «The Team Approach Applied to Glossary Management», dans *The ATA Chronicle*, nov.-déc. 1991, pp. 10-11 et janvier 1992, pp. 5-7, respectivement.

¹⁵ Voir le vivifiant article de Victor Loewen, «Language Specialists in the Year 2000 : Preparing for Machine Translation», paru dans *L'Actualité terminologique*, vol. 25, n° 1, 1992, pp. 6-9, dont le passage suivant (p. 7) résume bien ma pensée : «It is, of course, possible for us to refuse to acknowledge and incorporate technological change. While progress is made in machine translation, we can go on providing translation services for the next ten years using today's technology. But unless we get involved, the quality of machine translation and the affordability of the systems will follow a track that favours developers, not translators. The end result will be that our profession will be shaped from without, not from within.»