

Ce qu'il faut savoir sur les mémoires de traduction

par Claude Bédard

Publié dans Circuit, numéros 60 et 61 (été et automne 1998)

Encore peu répandus, les gestionnaires de mémoire de traduction feront demain partie de la pratique des traducteurs. Reste à savoir s'en servir à bon escient.

L'histoire remonte au début des années 80 avec le système TSS d'ALPS (plus tard Alpnet), véritable pionnier de ce genre d'outil, qui a longtemps fait cavalier seul avant d'être imité, depuis le début des années 90, par des logiciels comme *Translation Manager (TM/2)*, *Translator's Workbench (TWB)*, *Optimizer*, *Déjà Vu* et *Transit* (des sociétés IBM, Trados, Eurolang, Atril et Star respectivement). Après une longue éclipse commerciale, Alpnet a récemment mis en marché un module de TSS sous le nom de *JOUST*.

Tous ces logiciels, entre autres fonctionnalités, gèrent les phrases répétitives en s'appuyant sur une ressource appelée *mémoire de traduction (MT)*. Il s'agit d'un type spécial de *bitexte* (groupement de textes originaux et de leur traduction dont les phrases correspondantes sont reliées en parallèle) exploité par un mécanisme de *rappel automatique*.

Un *gestionnaire de mémoire de traduction (GMT)* a pour fonction essentielle de rechercher automatiquement, dans une MT, les phrases identiques ou semblables à celles du document à traduire. (En général il dispose aussi d'un outil de gestion terminologique, que nous ne décrivons pas dans le cadre de cet article.)

Cet article vise non pas à comparer les différents logiciels commerciaux, mais plutôt à faire comprendre l'intérêt et les limites de ce genre d'outil.

Qu'est-ce qu'un texte répétitif?

La question n'est pas aussi triviale qu'elle en a l'air. Précisons d'abord que pour un GMT, la répétitivité doit résider à *l'échelle de la phrase*, et non au niveau de la terminologie et d'expressions phraséologiques figées qui ne représentent que des portions de phrase.

Il ne faudrait pas croire que dès qu'il y a répétitivité à l'échelle de la phrase, l'usage d'un GMT s'impose. Ces logiciels, outre leur coût, imposent une certaine complexité d'utilisation qui ne se justifie pas dans toutes les situations de répétitivité. Ainsi, certains types de répétitivité sont gérables manuellement :

- Par exemple, s'il y a des *parallélismes* évidents (documents très structurés comme des contrats d'assurance, mise à jour peu complexe, etc.), on peut repérer les variantes au moyen d'un comparateur de versions, puis ne traduire que les variantes.
- Si les phrases répétitives sont *regroupées* de façon facilement repérable (une page d'introduction, des conditions commerciales standard, etc.), on a affaire en fait à une *répétitivité par blocs de texte*, et l'intérêt d'une gestion phrase à phrase est moins évident.
- Si la répétitivité porte sur un *nombre limité* de phrases différentes qui reviennent très souvent, on peut se constituer un *dictionnaire de phrases* et l'appliquer par prétraduction, ou même manuellement.

Par contre, si les phrases répétitives sont *variées* (beaucoup de phrases, peu d'occurrences par phrase) et *dispersées*, le traducteur ne sera guère capable de les gérer sans un effort assez considérable; un GMT est alors l'outil approprié.

Comment on crée une MT

Une MT se crée de deux façons :

- Soit à *mesure qu'on traduit* un document au moyen du GMT; chaque phrase, dès qu'elle est traduite et validée par le traducteur, est intégrée à la MT et est désormais disponible pour le reste du document.

- Soit *par construction*, au moyen des deux versions d'un document déjà traduit. Il s'agit alors d'apparier les phrases entre les versions. On peut soit procéder manuellement en traitement de texte et importer le résultat dans la MT, soit recourir à un logiciel interactif spécial appelé «aligneur».

Signalons que cet alignement n'est pas nécessairement chose aisée :

- Les frontières des phrases peuvent être ambiguës (par exemple le point peut être abrégatif).
- Les phrases ne sont pas nécessairement traduites une pour une (une phrase peut être traduite par deux phrases, et inversement).

Comment on utilise une MT

Une MT s'utilise de deux façons :

- En *mode interactif*. Le document à traduire est affiché à l'écran et l'utilisateur sélectionne les phrases une à une pour les traduire. Chaque fois qu'une nouvelle phrase est sélectionnée, le logiciel cherche dans la MT s'il existe une ou plusieurs phrases identiques ou semblables et les propose au traducteur dans une fenêtre séparée. Le traducteur décide alors de retenir la phrase (ou une des phrases), l'insère dans sa traduction et la modifie au besoin.
- En *mode automatique*. Le document à traduire est traité automatiquement par le logiciel, qui insère dans le document à traduire les phrases de la MT reconnues comme identiques à celles du document à traduire. Cette opération peut être commode si le nombre de phrases identiques est élevé, car elle évite d'avoir à les manipuler une à une (désigner la phrase, rappeler la traduction, valider la phrase).

Traitements analytiques

La plupart des GMT offrent des traitements automatiques basés sur l'analyse du texte à traduire.

- Extraire du document à traduire les phrases répétitives, qu'on peut alors traduire séparément et importer dans la MT (la chose est surtout utile si plusieurs traducteurs doivent utiliser la MT en parallèle).
- Extraire du document à traduire les phrases qui n'ont pas de correspondance dans la MT désignée. Ces phrases peuvent être envoyées à un système de traduction ou de prétraduction automatique, puis réinjectées dans le texte via la MT.
- Calculer le degré de répétitivité des phrases d'un texte ou d'un groupe de textes; ou encore, calculer le nombre de phrases semblables entre un texte ou un groupe de textes et la MT désignée.

Tolérance aux variantes

Le GMT doit pouvoir rappeler les phrases de la MT malgré des variantes mineures par rapport au texte en cours de traduction. Idéalement, dans le cas de variantes de nombres, le logiciel pourra même réinsérer automatiquement les nouveaux nombres aux bons endroits.

Certains logiciels permettent aussi de régler le seuil de similitude des phrases en fonction de ses préférences ou de la situation.

«Jamais plus vous ne traduirez...»

Le principe de la mémoire de traduction est séduisant. Comme le souligne volontiers la publicité de ces logiciels, «jamais plus vous ne traduirez deux fois la même phrase». Cette affirmation n'est pas en soi inexacte, mais la réalité n'est pas aussi simple.

D'abord, la traduction des phrases n'est pas indépendante de leur contexte; autrement dit, une même phrase ne se traduit pas toujours de la même façon.

- D'abord, une phrase peut se traduire de plusieurs façons dans deux dossiers de traduction différents (macrocontexte), à cause de préférences de vocabulaire ou de style.
- Mais encore et surtout, la traduction peut aussi varier selon le contexte immédiat (microcontexte). Par exemple, l'emploi d'un pronom qui renvoie à la phrase précédente.

Pour l'utilisateur d'un GMT, la tentation existe de traduire strictement phrase pour phrase et d'éviter systématiquement les renvois hors de la phrase... mais qu'advient-il de la tenue stylistique de la traduction?

Par ailleurs, il arrive assez souvent que deux phrases soient traduites par une seule phrase, et inversement. Si la phrase unique est en langue source, les deux phrases en langue cible sont rappelées automatiquement; mais dans le cas inverse, la recherche est a priori infructueuse.

La révision : tout se complique

Dans le meilleur des mondes possibles du point de vue d'un GMT, le traducteur produit du premier coup une traduction parfaite et finale de chacune des phrases. Or dans la réalité, on procède souvent en deux étapes : une traduction provisoire (ou à réviser) et la traduction finale.

Normalement, le réviseur apporte les changements dans le document même, puis répercute ces changements sur la MT (après chaque phrase, ou en bloc après chaque séance de travail). Tout va bien si la tranche révisée ne comporte pas de *répétitivité interne*. Mais si des phrases se répètent plus ou moins souvent, le dilemme suivant se pose :

- En apportant les changements au fil du document, on retrouve des phrases répétitives sans nécessairement les reconnaître; les changements apportés à ces phrases sont une perte de temps et ne seront pas uniformes.
- Une solution consiste à apporter les changements dans la MT elle-même (ce qui est souvent permis par le GMT); ensuite, on réapplique la MT sur le document pour mettre celui-ci à jour. Mais le travail se fait hors contexte, avec les risques évoqués plus haut.

Enfin, la situation se complique encore davantage si plusieurs traducteurs travaillent en parallèle à un même dossier de traduction. Comme on le voit, une MT peut être une invitation au chaos. Non seulement il importe d'avoir une procédure de travail rigoureuse, mais celle-ci doit s'appuyer sur une bonne dose d'ingéniosité et sur une excellente connaissance des mécanismes du GMT.

Compatibilité avec le formatage

Les GMT permettent de conserver le formatage des textes traduits. Deux stratégies très différentes sont pratiquées :

- Le document à traduire est *importé* dans le GMT; celui-ci doit filtrer le formatage, puis le reconstituer à l'exportation – avec les risques que cela comporte. La traduction se fait dans l'éditeur du GMT. C'est la méthode adoptée notamment par TM/2, et qui fonctionne très bien pour des documents dont le formatage est inscrit en clair (Bookmaster, HTML, Ventura, etc.).
- Dans l'autre cas, le document reste dans son traitement de texte d'origine, et c'est le GMT qui vient s'y arrimer par le biais de macrocommandes. Cette méthode a le double avantage de ne pas «affronter» le formatage et de laisser le traducteur dans un environnement qu'il connaît bien et où il dispose d'un vaste choix de commandes (y compris ses propres macrocommandes). Par contre, l'achat du logiciel de traitement de texte est nécessaire.

Consultation manuelle de la MT

À l'origine, la MT n'était pas consultable manuellement. De plus en plus, les GMT offrent une telle fonction, qui constitue un véritable *outil bitexte*.

- Le traducteur peut ainsi récupérer des phrases quand même assez semblables, mais pas suffisamment pour que le GMT puisse les repérer lui-même.
- Mais surtout, le traducteur a ainsi accès à un *réservoir de solutions ponctuelles de traduction*. Cet avantage peut devenir inestimable pour le traducteur en formation ou auquel on confie un dossier qui ne lui est pas familier. À la limite, il peut justifier l'utilisation d'un GMT même si la répétitivité des phrases est nulle!

La plupart des GMT affichent les phrases de la MT hors contexte, c'est-à-dire que la séquence naturelle des phrases dans les textes de référence n'est pas reconstituée.

Autres qualités à rechercher

- Si l'on s'aperçoit qu'il y a une faute dans la phrase à traduire, peut-on la modifier (et le GMT en tient-il compte)?
- Si l'on s'aperçoit que la phrase à traduire est mal découpée (point abrégatif interprété comme fin de phrase, par exemple), peut-on modifier le découpage?
- Quels sont les formats de documents traités par le logiciel?
- Peut-on modifier le contenu de la MT sans avoir à l'exporter, puis à le réimporter?
- Peut-on fusionner, extraire, combiner, prioriser, etc., les MT?

Acheter ou ne pas acheter?

À cause de leur prix (de 1000 à 5000 \$) et de leur utilisation plus ou moins complexe, les GMT ne pénètrent le marché que lentement. Pour bien des traducteurs, une telle acquisition est sans doute prématurée. Toutefois, il convient d'être aux aguets, car on aurait tort de se passer de cet outil si les circonstances le motivent. Et à la longue, gageons qu'un GMT trouvera de plus en plus d'utilisations au quotidien.