

Avec l'évolution de la technologie, le métier de la traduction a chaque jour de nouveaux défis à relever. Aujourd'hui, un des moindres n'est pas la traduction sur Internet et la capacité d'échanger des textes plus ou moins complexes sur le réseau des réseaux. Même s'il est peu probable que la livraison des traductions sur papier disparaisse, il ne faut pas se voiler la face: la part des échanges de textes sur support électronique ira croissant, avec tous les problèmes de compatibilité de formats que cela entraîne. Il nous a donc semblé important d'attirer l'attention des lecteurs d'Hieronymus sur ce problème et nous avons demandé à E. Ross, Docteur en sciences de l'information, de traiter ce sujet en deux articles. Le premier, que vous avez sous les yeux, brosse un tableau de la situation actuelle et prévisible à court terme. Le deuxième, prévu pour le prochain numéro de Hieronymus, traitera des normes disponibles en matière de formats d'échange de l'information. (JEF)

## **De l'imprimé à l'électronique: Quelles solutions techniques pour une traduction en transition ?**

Evelyne Ross  
Université de Technologie de Compiègne

De nombreux systèmes d'information sont mis en place dans le but d'atteindre un objectif en apparence assez simple : favoriser la circulation de l'information mono- et multilingue depuis les auteurs jusqu'aux lecteurs, afin que ces derniers puissent en profiter et satisfaire leurs besoins en information. Ce sont aujourd'hui les réseaux informatiques qui semblent dicter la voie à suivre en matière de systèmes d'information.

Il ne fait donc aucun doute que les traducteurs doivent s'adapter à ces nouvelles pratiques et à ces nouveaux outils.

Dans cet article, nous essaierons de montrer comment un traducteur peut s'y prendre pour exploiter les nouvelles technologies et rendre ainsi de meilleurs services à ses clients et aux lecteurs, tout en conservant des coûts de production suffisamment bas pour demeurer compétitif. Notre propos est de présenter brièvement les modèles d'édition électronique et les éléments à considérer pour les évaluer.

### **Les modèles techniques de publication**

Au cours des dernières années, les réseaux, en particulier Internet, nous ont permis de réaliser de nombreux projets de diffusion électronique d'informations. Plusieurs techniques nous permettent d'y parvenir, et presque toutes ont comme point commun l'utilisation des formats de documents popularisés par Internet, soit HTML et PDF<sup>1</sup>. Ces initiatives ont pour effet de constituer un immense réservoir d'informations en format électronique, soit le World Wide Web, source intéressante mais difficile à exploiter. Certains traducteurs embrassent les techniques habituelles : production et diffusion de HTML sur le Web ou encore production de document PDF à partir de leur chaîne de traitement traditionnelle et diffusion de ces documents sur la toile ou par Intranet.

---

<sup>1</sup> Le format PDF (Portable document format) est un format de document électronique développé par la compagnie Adobe. Un document PDF conserve l'allure originale de la forme imprimée du document (textes, graphiques, couleurs), peu importe la plate-forme utilisée.

Cependant, ces techniques sont-elles suffisantes pour assurer une édition de qualité de la traduction? Les quelques pages qui suivent sont consacrées à cette question et passent en revue certains critères importants.

## **1. Moyens de production**

Au cours des vingt ou trente dernières années, les traducteurs ont su profiter des développements technologiques dans le domaine de l'informatique. Ils l'ont utilisée dans la chaîne de production, à l'aide des techniques de publication assistée par ordinateur (PAO), que ce soit à l'aide de logiciels de traitement de texte, de graphisme ou de mise en page. La plupart des traducteurs techniques fonctionnent avec ces moyens de production. Cependant, cette intégration des nouvelles technologies n'est ni complète ni idéale. Même si ces chaînes de production utilisent en très grande partie l'informatique, leur finalité est de produire des documents imprimés et non des documents électroniques.

Nous pouvons tirer deux conclusions de ces méthodes en apparence contradictoires : presque toutes les traductions imprimées, même produites de façon traditionnelle, existent sous une forme électronique quelconque mais, pour certaines traductions imprimées, nous ne possédons pas de version finale et définitive en format électronique

## **2. Formats de diffusion et de conservation**

Tout document électronique sera représenté par un fichier informatique, dont le contenu exact sera déterminé par le format utilisé pour représenter l'information. Très souvent, les formats de documents sont associés à l'application qui produit le document, par exemple les formats Excel ou Word.

La question des formats est probablement la plus importante dans le cadre de la traduction. Elle a un impact majeur sur la production, la diffusion et sur la conservation des textes traduits.

### **2.1 Conservation**

Puisque les documents électroniques monolingues et multilingues doivent être stockés dans un format donné, il est nécessaire de s'interroger sur les critères à utiliser dans le choix d'un format de document électronique, en ayant pour objectif la conservation à long terme. D'abord, il faut que le format soit capable de représenter correctement l'information contenue dans le document. Par exemple, s'il y a du texte et des images ou des graphiques, il est nécessaire d'utiliser un format qui permette d'intégrer à la fois des informations textuelles et graphiques. De nos jours, ce n'est plus un véritable problème, car la plupart des formats de document permettent d'intégrer différents types d'information et sont à proprement parler des formats de documents « multimédia ».

Nous voulons également un format qui puisse être « lu » par une application et ce aussi longtemps que nous le souhaitons. C'est à ce stade-ci que les difficultés se présentent habituellement ; les conversions d'un format à l'autre (par exemple WordPerfect 5.0 à WordPerfect 8) ne sont pas une solution à ce problème puisqu'il y a souvent des pertes d'informations, des changements dans la présentation et d'autres manifestations indésirables.

Heureusement, il existe une façon élégante de résoudre ce problème: l'utilisation concomitante d'un format de représentation de l'information très simple et très universel et de techniques rendant ces traductions « lisibles par l'humain ». Un document structuré représenté à l'aide de XML est bel et bien un exemple d'un tel document.

En effet, on peut représenter un document XML à l'aide de caractères faisant partie du jeu de caractères ASCII. Concrètement, le fichier produit sera un pur fichier ASCII, c'est-à-dire le type de fichiers le plus universel que l'on trouve dans le monde informatique (Techniquement, les documents XML sont stockés en Unicode<sup>2</sup>, mais on peut les réduire à un simple ASCII sans perte d'informations). Il y a de fortes chances qu'il subsistera à l'avenir, et pendant longtemps, des plates-formes informatiques et des applications qui nous permettront de « voir » un fichier ASCII. De plus, un document structuré utilisant XML contient de l'information du genre <titre>Introduction</titre>. Même sans application particulière, même sans connaissances informatiques ou de XML, il est assez facile de s'imaginer que le mot « Introduction » constitue ici un titre, et non le nom du traducteur.

## **2.2 Diffusion**

Les formats de diffusion de documents électroniques sont multiples et variés. Toutefois, il y a présentement une forte convergence vers deux formats associés au Web : HTML et PDF. Le format PDF est particulièrement bien adapté à la représentation exacte de documents imprimés dans un format facilement diffusable sur le Web, car le logiciel de consultation des documents PDF est gratuit, disponible en plusieurs langues, et bien connu des traducteurs. De plus, il est facile de produire des documents PDF à partir de n'importe quelle application informatique. Il est donc tout à fait normal d'utiliser ce format à des fins d'impression à distance et sur demande, ce qui répond à de nombreux besoins pour la traduction puisque les lecteurs peuvent n'avoir besoin que d'une partie du texte.

L'autre format de convergence est évidemment le format HTML, né avec le Web et popularisé par l'évolution du réseau. Presque toutes les applications documentaires peuvent maintenant lire ou produire des documents HTML. Lorsque l'on associe HTML avec le langage JavaScript et les feuilles de style CSS (Cascading style sheet, voir <http://www.w3.org/Style/css/>) et que l'on obtient ainsi du Dynamic HTML (DHTML), il est possible de créer de véritables interfaces de consultation et non de simples documents électroniques. HTML est donc un excellent format de diffusion; malheureusement, il est pauvre en matière de gestion ou de production de documents. De surcroît, pour produire un document HTML de qualité, il faut travailler à partir d'une source d'informations très riche, sous peine que le travail ne doive être fait manuellement et ne soit fastidieux.

Ces deux remarques nous amènent à conclure que le format HTML s'avère un format de diffusion à privilégier, pour autant que nous utilisions un autre format de gestion et que nous puissions produire facilement des documents HTML de qualité. Ajoutons que les formats de PAO ne remplissent ni l'un ni l'autre de ces conditions, mais que les normes SGML et XML, elles, satisfont ces besoins.

## **2.3 Format de production ou de gestion**

---

<sup>2</sup> La norme Unicode est un jeu de caractères étalé sur 16 bits, ce qui permet la représentation d'un maximum de 65 536 caractères différents. Elle correspond au premier plan à 2 octets d'une norme plus universelle basée sur 4 octets, à savoir la norme ISO 10646.

Au moment de la production d'un texte multilingue, nous devons travailler avec un format qui nous permette d'atteindre tous les objectifs fixés dans les paragraphes précédents, et ce de façon efficace. En résumé, nous recherchons un format d'encodage de l'information qui nous permette de répondre à nos besoins, soit :

- Manipuler aisément les documents pour effectuer toutes les activités de production (gestion, workflow, diffusion, etc.) ;
- Permettre l'exploitation de toutes les possibilités qu'offrent les documents électroniques (multimédia, hypertexte, génération dynamique de contenu, recherche plein texte, données complémentaires/supplémentaires, etc.) ;
- Produire des documents électroniques dans d'autres formats (par exemple HTML), et ce en exploitant toutes les possibilités de ces formats ;
- Permettre la diffusion sur différents supports (cédérom, DVD, réseaux, etc.) y compris le support imprimé à l'aide d'applications de PAO ;
- Conserver à long terme et dans des conditions optimales l'information et sa structure, afin d'en assurer la pérennité.

Les documents structurés constituent la meilleure façon de répondre efficacement à l'ensemble de ces critères. A l'opposé, les formats associés à la PAO n'offrent pas la même souplesse ni la même puissance, car ils contiennent de l'information en fonction d'un seul et unique support.

#### **2.4 Exploitation du format électronique**

Passer de l'imprimé à l'électronique constitue un changement qui va bien au-delà du mode traditionnel de diffusion d'une traduction. Les formats électroniques permettent en effet de représenter plusieurs types d'information que l'on peut retrouver dans une traduction papier. Une traduction en transition vers l'électronique intégrera peu à peu ces types d'information et son éditeur devra mettre en place les outils nécessaires pour y parvenir.

L'information statique est la plus évidente, mais aussi la seule qui puisse véritablement être représentée sur une feuille de papier. Il s'agit de textes ou d'images qui, une fois imprimés ou stockés dans le document traduit, ne changeront pas. La plupart des documents traduits qui existent ne contiennent que de l'information statique, car ils ont été produits d'abord et avant tout pour un support qui ne permet que ce genre d'information: le papier.

L'information dynamique est celle qui « bouge », qui s'anime. Ces animations ne sont pas contrôlées par les lecteurs / utilisateurs, mais plutôt par l'importance qu'elles accordent au contrôle par le lecteur. Par exemple, dans une traduction technique, la simulation d'une tâche de maintenance en trois dimensions, avec la possibilité pour le lecteur de manipuler une partie du produit dans tous les sens pour la voir sous tous ses aspects, constitue de l'information fortement interactive. La publication d'un algorithme, auquel l'ingénieur ou le développeur peuvent fournir des valeurs de départ et vérifier les résultats de façon instantanée ou presque, est un autre exemple d'information dynamique.

L'information active permet aux lecteurs / utilisateurs d'agir sur le contenu du document ou encore sur l'environnement de consultation. Les liens hypertextuels font partie de cette catégorie, de même que les formulaires interactifs. On peut imaginer qu'un sondage publié dans un article scientifique pourra être mis à jour dynamiquement par les lecteurs qui pourront ainsi faire connaître leur opinion à partir du document.

Un des problèmes majeurs pour le traducteur et, partant, pour le lecteur, est l'absence ou la surabondance de normes pour certains types d'informations. Dans le cas du texte et des images simples, la situation est assez facile à maîtriser de par l'omniprésence du format HTML et de ses formats d'image associés, GIF et JPEG. Mais au-delà de ces quelques formats, la situation devient difficile, car le support n'est habituellement pas inclus dans les navigateurs courants, ce qui oblige le lecteur à installer des modules externes ou des applications supplémentaires afin de pouvoir consulter certaines parties des documents. Toutefois, il est évident que le lecteur ne va installer un tel module que s'il en a besoin, au moins occasionnellement, ou si l'information manquante est très importante pour lui. Dans le cas contraire, il ira voir ailleurs ou il s'en passera.

### **3. Le défi à relever**

Le défi technologique que le traducteur doit relever est double :

- Premièrement, il s'agit de trouver des formats adéquats pour chaque type d'information susceptible de se présenter et trouver un format de base qui puisse lier tous ces types d'information et servir de « ciment » aux différentes parties du document électronique. Dans le cas du format de base, un modèle de traitement centré sur XML peut s'avérer suffisant. En effet, cette norme permet d'intégrer des parties de document en différents formats et d'assembler un document ayant des composantes textuelles, iconographiques, vidéo et sonores, ainsi que des algorithmes, n'est plus un réel problème. De plus, en utilisant HTML comme principal format de diffusion pour la consultation électronique de l'information, il est possible d'assembler des documents très complexes, avec des composantes très variées, sauvegardés dans des formats divers.
- La deuxième partie du défi cause plus de problèmes : quel(s) format(s) utiliser pour différents types d'information ? Un grand effort de normalisation reste à faire dans ce domaine, avant que les navigateurs habituels puissent présenter tous les types d'informations sur toutes les plates-formes, même si des pistes encourageantes commencent à s'ouvrir à nous, telles la mise en place d'une norme de l'industrie pour les images vectorielles ou l'utilisation croissante du langage de programmation Java pour les applications dynamiques.

Pour conclure, bien que provisoirement, il nous paraît essentiel que le traducteur cherche quelles sont les normes les plus utiles au type de documents qu'il traduit, qu'il les adopte et qu'il apprenne à s'en servir efficacement, pour éviter de perdre le contrôle du document qu'il a produit.