

NUMÉRISER UN LIVRE ?

À propos d'une expérience en cours

Nous proposons d'expliquer ce qu'est la numérisation d'un livre et d'en montrer les divers problèmes en prenant comme cas d'école la numérisation, en cours, du *Manuel typographique* de Fournier.

Le *Manuel typographique*¹ de Pierre Simon Fournier (1764) fait partie du patrimoine typographique français, voire mondial. Il y a une quinzaine d'années, m'occupant de typographie numérique dans un centre de recherche en informatique², j'aurais aimé en consulter diverses pages de près. Hélas, les rares bibliothèques françaises qui disposent d'un exemplaire l'ont mis dans leur fonds ancien, consultable sur place, uniquement avec patte blanche et non photographiable ; la seule réédition fac-similé existant à ma connaissance était celle anglaise de Harry Carter (1930) mais épuisée depuis des années ; quant aux exemplaires originaux circulant encore, ils dépassaient de loin mes moyens³. Il s'est alors trouvé que j'ai pu disposer d'un exemplaire à la reliure cassée dont j'ai eu le droit de faire une photocopie⁴. J'ai alors imaginé, en collaboration avec un éditeur rennais, de rééditer cet ouvrage en fac-similé, accompagné du texte français moderne que j'ai saisi sur un système bureautique. James Mosley, bibliothécaire de la *St-Bride Library* à Londres, spécialiste de la typographie française des XVI^e-XIX^e siècles, avait même accepté d'en faire une introduction. Mais, pour diverses raisons, les choses en sont restées là pour moi ; James Mosley⁵ a alors réédité la version fac-similé de Carter en y rajoutant

(1) *Manuel typographique utile aux gens de lettres & à ceux qui exercent les différentes parties de l'art de l'imprimerie*, par Fournier le jeune, à Paris, chez Barbou, tome 1, 1764, et tome 2, 1766. Seuls deux autres ouvrages antérieurs au dix-neuvième siècle traitent du même sujet : les *Mechanik Exercices* de Moxon (Londres, 1683) dont il existe encore de nombreuses éditions fac-similé (d'autant que ce livre traite aussi de charpenterie, de faïencerie, de serrurerie, etc.) et *Kurzedoch Nützliche Anleitung von Form* (anonyme, Erfurt, 1740). Aucun des trois ouvrages n'est, à ma connaissance, numérisé.

(2) Ce travail avait démarré dans le cadre du projet européen Didot d'enseignement de la typographie que je dirigeais depuis l'Irisa.

(3) À titre indicatif, un exemplaire était en vente fin 2003 à la librairie Jammes à Paris pour 18 000 €.

(4) Merci encore au Musée de l'imprimerie et de la banque, à Lyon.

(5) *The Manuel Typographique of Pierre-Simon Fournier le jeune, together with Fournier on Type founding, an English Translation of the Text*, par Harry Carter, avec une introduction et des notes par James Mosley, 3 vol., Darmstadt, 1995. ISBN 3-88607-094-8.

introduction et notes, toutefois de façon plutôt confidentielle, cet ouvrage étant ainsi introuvable en France.

Il y a un an, voyant que le web est un lieu formidable de diffusion d'information, je me suis dit que c'était l'endroit idéal pour rendre accessible ce *Manuel* à tous. Bien que non terminée¹, cette expérience constitue déjà une étude de cas montrant les divers problèmes posés par un tel travail, tant sur le plan technique (de lecture et d'écriture) que culturel ou légal !

Image ?

Quand on parle de numériser un livre, la première idée qui vient en général en tête est celle d'image « scannée », de photo numérique. Il existe d'ailleurs déjà sur le web de très nombreux ouvrages numérisés, disponibles publiquement.

La BNF (Bibliothèque nationale de France) met ainsi sur son site Gallica² des milliers de livres imprimés sous forme image et on retrouve ainsi le même principe que celui des microfilms ou des photocopies : chaque page d'un livre est photographiée mais l'image obtenue n'est plus une matrice de grains de sels d'argent plus ou moins noircis ou colorés, mais une matrice de nombres représentant un code de couleur (par exemple 0 pour blanc, 50 pour rouge et 100 pour noir). Ces nombres sont parfois appelés « pixels » et l'image est dite « scannée ». Comme une image sortie d'un appareil photo numérique, on peut la transmettre sur Internet. Il suffit alors de se connecter sur ce site Gallica pour télécharger une ou plusieurs copies de pages, voire un ouvrage entier, puis en faire l'impression. Mais tout comme on a des photographies à plus ou moins gros grains, on a des images numériques à plus ou moins bonne définition³ ; et selon le nombre de codes choisis pour les couleurs (par exemple 10, 100 ou 1 000), on aura une image de plus ou moins bonne qualité. Finalement, la qualité de l'image d'un livre va dépendre énormément de la qualité du scanner et de la façon dont on l'a utilisé : cela va aller de la grossière image en noir et blanc à des images très piquées en couleurs. Mais hélas, la taille en mémoire de ces images va être proportionnelle à la qualité : meilleure est celle-ci, plus volumineuse sera l'image (et plus longtemps il faudra à Internet

(1) On a accès, parfois à l'aide d'un mot de passe (pour des problèmes de copyright non encore résolus), à une version de travail à l'url <http://www.irisa.fr/BiViTy/Fournier-Manuel.html>.

(2) <http://gallica.bnf.fr>.

(3) En anglais *resolution*, d'où le mauvais emploi du faux ami « résolution ». En informatique, on mesure la définition en dpi (*dots per inch*, points par pouce, ce qui revient à compter en nombre de pixels par cm ; une résolution courante de scanners d'amateurs est 200 dpi) ; en photographie numérique on compte souvent le nombre de pixels dans une surface donnée, par exemple 3,4 millions de pixels pour une photo standard).

pour la transmettre)¹. Le premier choix à faire lorsque l'on veut numériser un livre va donc être celui de la qualité des images et en fait de l'usage que l'on veut en faire. Voici deux exemples extrêmes.

- Le site Gallica propose des images avec une définition très basse, ressemblant souvent à de mauvaises photocopies. Outre la rapidité de l'opération de numérisation de ces images, donc la possibilité de numériser plus de livres dans un temps donné, ce choix permet de stocker davantage d'images dans un volume donné (par exemple sur un disque dur) et surtout de les transmettre beaucoup plus rapidement sur Internet. La priorité a donc été donnée à la disponibilité des œuvres, au contenu des ouvrages et non à leur présentation. Il serait inconcevable de publier sous forme papier une copie numérisée d'un livre de Gallica. En revanche, un étudiant travaillant sur le *Traicte de la Grammaire francoise* de Robert Estienne (1569) peut disposer, après quelques petites minutes de téléchargement, d'une photocopie de l'ouvrage original (auquel il n'aurait accès qu'en salle de lecture et encore avec beaucoup de difficultés) qu'il pourra considérer comme son propre instrument de travail, qu'il pourra surligner, annoter, griffonner, etc. à sa guise.

- Le site des enluminures médiévales françaises² au contraire ne donne que des images de haute qualité, toutes en couleurs, très belles à regarder, que l'on peut copier chez soi, une par une et avec un délai parfois long. On pourrait, n'étaient des problèmes de copyright sur lesquels nous reviendrons plus loin, les publier comme un bon livre d'art.

Lorsque l'on numérise un livre ancien, il se passe aussi des phénomènes qui vont à leur tour réduire la qualité de l'image, indépendamment d'ailleurs de la définition. Un livre ancien est d'abord très fragile et sensible à la lumière et aux manipulations, il faut donc un scanner qui limite les temps d'exposition et la manipulation physique de l'ouvrage. Par ailleurs, le papier est souvent non lisse et l'encre peut avoir « foulé » (débordé du caractère en empattant la trace imprimée) ; le papier est mince et on peut voir sur un recto apparaître par transparence des lignes du verso ; le papier est friable et des trous peuvent montrer des mots de la page dessous ; le papier peut se tacher et des traces de doigts (notamment) peuvent apparaître ; la courbure de la reliure peut déformer les lignes du texte côté blanc de couture, etc. Bref, il y a beaucoup de « bruits » sur une image de livre. Des

(1) À titre indicatif, l'image numérisée d'une même page A4 peut avoir une taille de 22 000 octets à 75 dpi en noir et blanc et de 71 millions d'octets à 1 200 dpi en 256 couleurs. Ces nombres varient par ailleurs selon le mode de stockage et de compactage des données dans l'image (formats JPEG, GIF, etc.).

(2) Ministère de la Culture : <http://www.enluminures.culture.fr/>

programmes¹ savent désormais « nettoyer » ces pages, mais ils sont encore peu commercialisés.

... ou texte ?

Si on compare à nouveau ces deux œuvres, une différence est à bien préciser. La *Grammaire* d'Estienne est un « texte » imprimé alors que l'autre est surtout un ensemble d'illustrations. Pourquoi alors ne pas stocker ce texte comme une simple suite de lettres et signes et non comme une image ? C'est ce qu'ont fait plusieurs sites, typiquement l'ABU² qui offre ainsi de très nombreuses œuvres littéraires en mode texte. Toutefois, ceci ne peut être fait n'importe comment.

- Il faut avoir saisi le texte, ce qui peut être long (penser aux œuvres de Victor Hugo, Proust, Duhamel, etc.). On peut utiliser des systèmes de reconnaissance optique de caractères (OCR) qui, aujourd'hui, peuvent donner de bons résultats sur des textes bien imprimés avec des caractères courants³. Mais dans tous les cas, il faut aussi vérifier le texte⁴, ce qui est encore long.

- Il faut aussi coder les caractères de façon que le texte soit bien interprété par le « lecteur ». Il existe en effet plusieurs normes de codage des caractères, notamment Ascii (sans lettres accentuées), Latin-1 (avec toutes les lettres françaises sauf œ, Œ et Ÿ) et aujourd'hui Unicode⁵ (avec tous les caractères du monde), mais aussi des « standards privés » (comme ceux utilisés par Windows, IBM, MacIntosh, etc.). Aujourd'hui encore on trouve sur le web des textes mal codés qui donnent des résultats dépendants du navigateur utilisé⁶.

(1) Basés sur des techniques de traitement du signal notamment. On trouvera des références sur ces problèmes de numérisation dans B. Couâsson, P. D'Albera et H. Emptoz (éds), *Numérisation et patrimoine*, numéro spécial de *Document numérique* (Lavoisier éd.), vol. 7, n° 3-4, 2003 et sur le site <http://www.culture.gouv.fr:80/culture/nvt/numerisation/fr/technique/technique.htm>.

(2) ABU : la Bibliothèque Universelle : <http://abu.cnam.fr/>

(3) Notre *Grammaire* de Estienne donne encore de très mauvais résultats avec tous les OCR commercialisés testés, même ceux avec « apprentissage ». Nous y reviendrons plus bas.

(4) Même si un OCR travaille avec un taux d'erreur de un pour mille, ça veut quand même dire deux fautes en moyenne tous les 30×60 signes, c'est-à-dire deux fautes toutes les pages (de 30 lignes de 60 caractères). Même si les correcteurs orthographiques que l'on trouve de plus en plus sur nos ordinateurs signalent de mieux en mieux les fautes d'orthographe, leur connaissance de la grammaire est encore bien insuffisante pour éviter une relecture humaine.

(5) Sur les concepts de codage de caractères, Ascii, Latin-1 et Unicode, voir Jacques André et Henri Hudrisier (éd.), *Unicode, caractères du monde entier ?*, numéro spécial de la revue *Document Numérique* (éd. Lavoisier), volume 6, n° 3-4, 2002.

(6) Ce qui explique souvent ces ? (au lieu d'une apostrophe), voire ces Ā© (au lieu de ê) que l'on trouve notamment dans les courriers électroniques entre correspondants ne disposant pas du même système de codage.

• On peut aussi coder certains attributs typographiques¹, comme l'italique qui marque par exemple l'emphase ou les mots étrangers. On peut aussi inclure dans le texte des marques² pour indiquer certaines « structures » telles qu'un titre de section, un grand titre, une liste, une note de bas de page, etc., voire pour indiquer des concepts plus sémantiques comme « un nom de lieu », une « date », etc. Internet a répandu l'usage de langages structurés comme HTML et XML.

Ce codage textuel permet d'économiser beaucoup de place³ (donc de temps de transmission sur Internet). Il permet surtout de travailler sur lui comme sur du texte, par exemple pour y chercher si un mot est présent et où, ou bien toutes les occurrences d'une expression⁴.

En revanche, ce mode texte est peu utilisable seul dès que les images deviennent importantes (en taille ou par leur contenu) ou lorsque le texte lui-même a une forme graphique prépondérante : c'est le cas des *Calligrammes* d'Apollinaire, du *Coup de dés* de Mallarmé ou... des spécimens de caractères comme ceux constituant une bonne partie du *Manuel typographique* de Fournier.

Image et texte !

Que faire alors dans le cas de ce *Manuel* de Fournier ? Il comprend deux tomes d'environ 300 pages chacun. Le second comprend essentiellement des exemples de caractères dans divers corps et styles. Le mode texte n'a guère d'intérêt pour ce tome ; en revanche, il est important de bien voir ces caractères, surtout les petits, d'où bien sûr la présentation en mode image de haute définition. Le premier tome, lui, comprend aussi quelques planches, qui seront donc numérisées en mode image, mais il contient essentiellement du texte pour lequel le mode texte s'impose si on veut que les lecteurs puissent y faire des recherches, indexations, etc. Mais ce texte est émaillé de vignettes, bandeaux, etc., qu'il serait dommage de ne pas montrer. Par ailleurs la mise en page et la typographie de ce livre sont exemplaires et il serait dommage de les perdre par une édition moderne. On a donc décidé de faire

(1) On parle parfois d'« orthotypographie » car, comme l'orthographe, ces règles d'emploi relèvent plus de la langue, du sens, que de la présentation graphique (comme la justification, ou l'emploi de telle ou telle fonte).

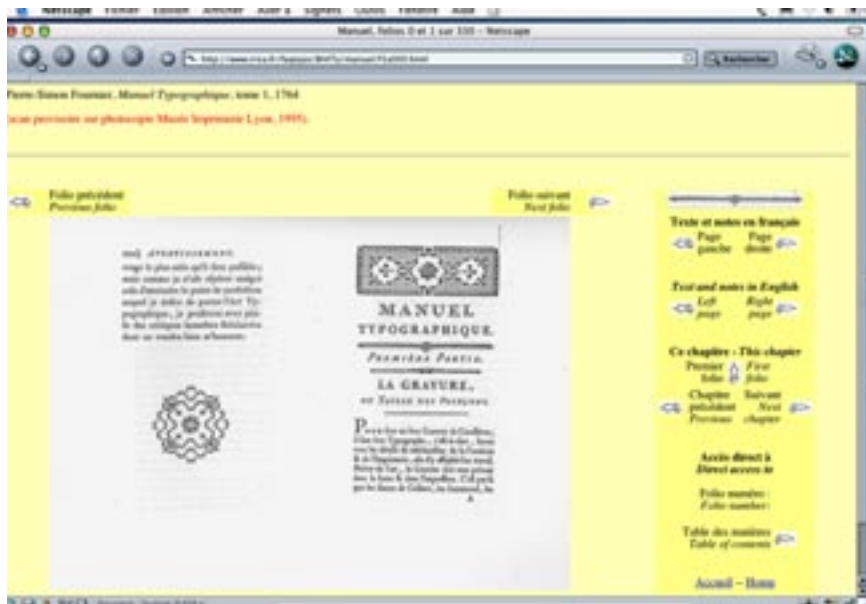
(2) En typographie, on appelle ces marques des balises depuis le début du XX^e siècle. Les langages de balisage comme SGML, HTML, XML n'ont fait que reprendre le concept et le terme.

(3) Une page A4 tapée à la machine à écrire (ou par ordinateur) prend environ 500 000 octets en mode image à 200 dpi, mais seulement 7 500 octets en mode texte.

(4) Ce sont ces fonctions de recherche qu'utilisent les navigateurs du web. Elles ne fonctionnent que sur des fichiers de type « texte », pas sur les fichiers en mode image. Du moins aujourd'hui, car il est probable que d'ici peu on saura faire passer un OCR sur une image de texte, obtenir le contenu en mode texte et sur ce texte faire les recherches désirées. Ceci nécessite alors d'avoir des images numérisées de très bonne qualité.

pour ce premier tome une édition fac-similé (donc en mode image) et ce en livre ouvert¹, édition doublée d'une version en mode texte.

Mais comme ces pages de texte sont malgré tout lourdes à charger en haute définition, on a décidé de proposer au lecteur deux jeux d'images, l'un de haute définition et l'autre de définition plus basse, mais d'affichage rapide, plutôt pour le feuilletage².



Feuilletage et passage d'un mode à l'autre

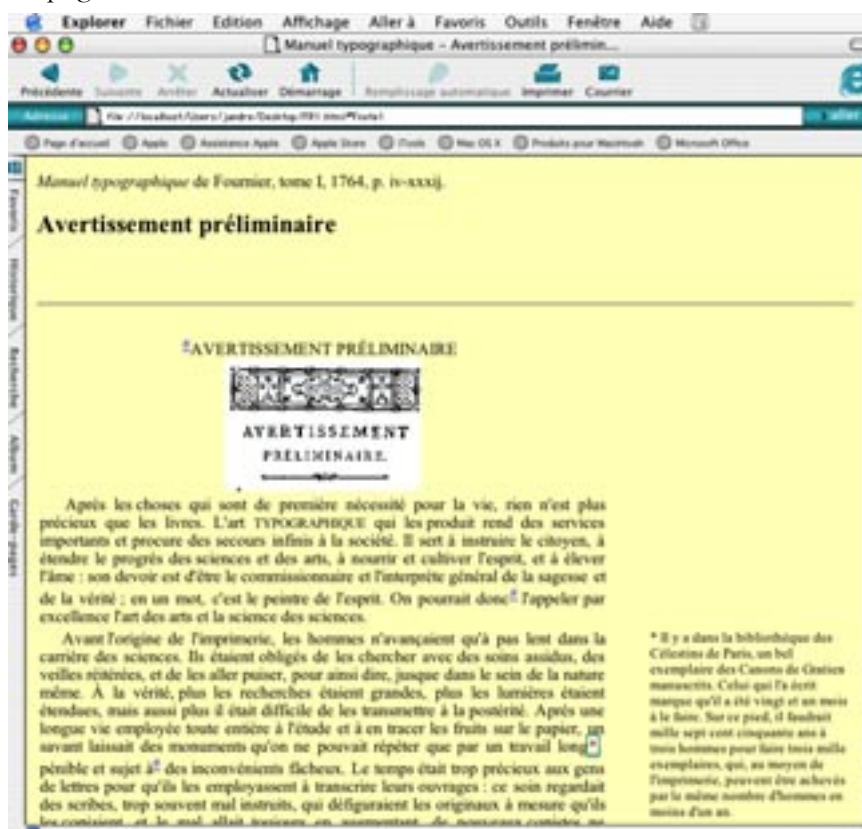
Lorsque l'on se connecte sur le site, on atteint la première image du tome 1. Deux flèches, en fait des « mains » dessinées par Fournier, permettent, lorsque l'on clique dessus, d'atteindre l'image suivante (ou la précédente) comme lorsque l'on feuillet³ un livre (figure 1 ci-dessus).

(1) En effet, la mise en page des livres est conçue globalement pour les deux pages, paire et impaire, d'un livre ouvert. Comme par ailleurs il se trouve que les écrans d'ordinateur sont à l'italienne (terme correct pour ce que les informaticiens appellent « en mode paysage »), ça permet finalement d'utiliser mieux l'espace de l'écran que si on ne montrait qu'une page à la fois tout en respectant donc la tradition de l'édition des livres papier. Le même principe se trouve déjà pour l'édition sur CD du *Manuale tipografico* de Giambattista Bodoni (1818) dans la collection *Digital rare books* des éditions Octavo aux USA.

(2) À ce jour, fin décembre 2003, seul le premier jeu d'images à 200 dpi est disponible sur le web. En effet la réalisation du second nécessite un gros travail (« nettoyage ») et j'attends d'avoir quelques retonnées sur le reste de l'œuvre pour engager l'investissement nécessaire.

(3) Le feuilletage électronique consiste encore à prendre les pages les unes après les autres ce qui ne correspond pas à la réalité. Des recherches sont en cours pour produire le bon effet, par exemple en utilisant le concept de *fish-eye* de Xerox-Parc : <http://pages.cpsc.ucalgary.ca/~sheelagh/personal/pubs/2001/carpendaleuist01.pdf>

Ce mécanisme (ancre, la base des documents sur le web) est très facile à installer. Ce même mécanisme permet également, pour chacune des deux pages affichées dans une image, de cliquer sur une flèche et de voir apparaître la même page en mode texte. Mais, pour permettre une lecture continue¹, ces pages de texte sont en fait mises en page comme si elles étaient imprimées sur un rouleau (le *volumen* des œuvres anciennes avant l'invention du papier plié et du livre-codex) que l'on peut faire défiler sur l'écran de haut en bas ou en arrière. Les quelques notes de bas de page posent toutefois un problème puisqu'il n'y a plus de pages...



Une solution² est alors de les mettre dans la marge (exemple en figure 2), comme les gloses du Moyen Âge ! On a marqué³ dans ce

(1) Sans avoir à cliquer pour faire venir les pages une à une comme les images. Notons au passage la possibilité d'imprimer ce texte comme un livre (version pdf).

(2) Une autre solution, plus habituelle, est de regrouper ces notes en fin de chapitre, voire en fin de tome ou dans un fichier à part, et de les faire apparaître en cliquant sur la marque d'appel. Mais certains lecteurs préfèrent voir à la fois la note et le texte correspondant, d'où notre méthode.

(3) Avec le signe « # » qui est l'abréviation médiévale de *numerus* et qui n'a rien à voir avec quelque dièse ! Voir par exemple ce signe après « donc » à la fin du premier paragraphe de la figure 2 ci-dessus.

« rouleau » les changements de page du livre original et il suffit alors au lecteur d'y cliquer pour voir apparaître la page correspondante au texte. Diverses autres flèches permettent aussi des accès directs à un chapitre donné, de changer de tome, d'aller à une page donnée, ou d'utiliser l'index de Fournier pour aller à la page voulue¹. Toutes ces possibilités sont regroupées, dans une sorte de tableau de commandes, à côté de l'image de la double page en cours de lecture (voir figure 1). Enfin, on a gardé dans le texte certaines images telles que les têtes de chapitres, des figures du texte, etc.

Traduction, texte moderne ou conforme ?

Éditer un texte français imprimé entre la fin de la Renaissance et le XVIII^e siècle nécessite de le traduire en... français d'aujourd'hui. Fournier écrit par exemple (sans tenir compte des choix typographiques comme les s longs ou les ligatures *st* et *ct* chères à la collection de la Pléiade chez Gallimard) : *Plusieurs savans & artistes, comme Lucas Paciolo, Albert Dure, J.B. Palatin, Pierre le Bé maître écrivain, & beaucoup d'autres, ont donné différens traités sur la forme et la figure des lettres, plutôt pour...* Aujourd'hui, on écrirait *savants, différens, plutôt*, etc. mais aussi *Pacioli, Dürer, Palatino*. C'est le style que nous avons adopté, car il permet de faire des recherches² de mots avec l'orthographe fixe d'aujourd'hui.

Mais ce faisant, on perd toute possibilité de travailler électroniquement³ sur le texte original, avec sa typographie et son français d'alors. Sauf si on fait une autre version conforme au texte imprimé. Mais on se heurte à un problème de fontes, de caractères, plus précisément à un problème de normalisation. Nous avons signalé plus haut que le standard Unicode permet de coder tous les caractères du monde, mais pour Unicode un caractère est une entité abstraite et ce standard ne tient alors pas compte des variations graphiques, telles que le s long, ni donc des ligatures⁴ comme *ct*. Or, chez Fournier on trouve parfois des coupures de mots comme *carac-ctère*, parfois *cara-ctère*, plus rarement *carac-tère*. Un érudit pourra avoir envie d'étudier l'emploi de cette ligature *ct* (ou de nombreuses autres choses équivalentes), mais il n'est donc pas possible de mettre sur le web une version codée dans une

(1) Étant bien entendu que, dans ce mode texte, le lecteur peut également utiliser les fonctions de recherche de son navigateur ou d'un logiciel externe, comme Google, pour y trouver des mots.

(2) À condition de disposer d'un outil de recherche qui, lui, n'efface pas les accents comme c'est souvent le cas ! Une recherche récente dans l'*Encyclopédie* de Diderot sur CD sur le nom « Dürer » m'a donné comme réponse les innombrables occurrences du verbe « durer » !

(3) Du moins tant que l'on ne dispose pas d'OCR capable de reconnaître efficacement les caractères anciens ce qui permettra alors de travailler à partir des images en haute définition.

(4) Pour être honnête, disons qu'Unicode tient compte de certaines ligatures, mais uniquement de celles existant dans des codages plus anciens.

norme puisqu'il n'y en a pas aujourd'hui qui tienne compte de ces détails typographiques¹. Finalement on se rend compte que les spécialistes des manuscrits médiévaux sont bien plus en avance que ceux des incunables ou des livres des XVI^e et XVII^e siècles car, le besoin étant sans doute plus flagrant, il existe diverses tentatives de codage² des abréviations, etc.



Édition critique

Telle qu'illustrée par les figures 1 et 2, cette première version électronique du *Manuel typographique* de Fournier va évidemment intéresser de nombreux spécialistes et amateurs de typographie. Mais, il lui manque quelques explications... Qui est ce Le Bé que Fournier cite avec Dürer et Pacioli ? Comment se sert-on d'une équerre pour mesurer la justification des matrices ?, etc. Or, nous l'avons déjà signalé, Harry Carter avait publié vers 1930 un fac-similé « commenté » de ce *Manuel*

(1) Ceci nous a donc amené à étudier une façon de normaliser le codage des caractères effectivement utilisés dans les livres anciens. En particulier, avec des collègues canadiens, nous étudions les caractères inventés par les grammairiens du XVI^e siècle pour lesquels un amendement à Unicode devrait être proposé prochainement.

(2) Citons par exemple le codage utilisé pour *Lancelot* ou *Le Chevalier de la Charette* de Chrétien de Troyes (XII^e siècle) : <http://www.mshs.univ-poitiers.fr/cescm/lancelot/> et le projet MUIF (*Medieval Unicode Font Initiative*) <http://www.engl.virginia.edu/OE/junicode/junicode.html>.

(en anglais) et il y a quelques années à peine, James Mosley a repris cette édition critique en la complétant de nouveaux commentaires (toujours en anglais). Il est donc important d'en faire profiter tous les chercheurs. Outre la traduction française de ces commentaires, il a fallu aussi les mettre en page de façon à ne pas passer son temps à cliquer pour lire tel ou tel d'entre eux. Pour le moment, nous utilisons une mise en page en colonnes (une pour les propres notes de Fournier, une pour celles de Carter et une pour celles de Mosley), la difficulté étant de régler des textes « parallèles » de longueur différente en utilisant par exemple des gloses en forme de L comme au Moyen Âge (voir figure 3 page précédente). Cette nouvelle « édition » est bien sûr dédoublée en une version française et une anglaise avec les liens entre elles.

Copyright

Fournier est mort depuis plus de 70 ans (et même plus de 230 ans), son œuvre est donc dans le domaine public et on peut la mettre sans problème sur le web ! Il n'en va sans doute pas de même des images scannées sur l'exemplaire du Musée de l'imprimerie de Lyon, et un consensus devra sûrement être trouvé. En revanche, les textes critiques de Harry Carter et surtout de James Mosley, eux, sont toujours couverts de droits et il faudra bien en obtenir les droits avant de mettre cette version sur le web.

En fait ceci relève du concept nouveau de *copyleft*, jeu de mot sur *copyright*, permettant l'auto-édition des scientifiques et le partage des connaissances organisé selon un cadre juridique de la circulation des textes des scientifiques et de la publication scientifique sur le web. Les choses avancent assez lentement dans ce domaine, mais je tenais à signaler que ce problème est bien d'actualité sur le web où la libre circulation de l'information prévaut.

Hypertexte

Mais que va-t-il se passer si à son tour un autre spécialiste a envie de commenter tel ou tel point, ou si encore un autre spécialiste tient à apporter des précisions et encore un autre à contredire tel avis ? Nos quatre colonnes ne suffiront plus et il faudra prévoir des procédures pour un travail collectif beaucoup plus important, dynamique, etc. En faire un « hypertexte » comme par exemple *HyperNietsche*¹. Mais il s'agit alors d'un travail bien plus considérable que ne nécessite peut-être pas

(1) Hypertexte sur Nietzsche, avec toute son œuvre, ses manuscrits, les travaux sur Nietzsche, etc., considéré un peu comme l'archétype des hypertextes littéraires notamment de ceux utilisant le *copyleft* dont nous avons parlé. Voir Paolo d'Iorio, *HyperNietzsche*, Presses Universitaires de France, 2000 et <http://www.hypernietzsche.org/>.

l'œuvre de Fournier et ce qu'il en reste, et qu'il sera de toute façon toujours possible d'entreprendre une fois le *Manuel* numérisé...

Ingénierie

On l'a compris, cette version électronique du *Manuel* de Fournier comprendra en fait des choses très diverses :

- Une version en mode image, en basse définition
- Une version en mode image, en haute définition
- Une version en mode texte en français moderne
- Une version en mode texte, mais en français original avec les caractères d'époque
 - Une version en mode texte en français moderne avec la traduction française des notes de Harry Carter et de James Mosley
 - Une version en mode texte comprenant la traduction anglaise de Harry Carter, ses notes et celles de James Mosley.

Toutes ces versions seront spécialement conçues pour la consultation sur le web et seront donc « truffées » de liens permettant de passer de l'une à l'autre. De plus, on prévoit aussi :

- Une version permettant une édition soignée sur papier, en français
- Une autre en anglais.

Ces diverses versions évoluent chacune à un rythme différent mais ont toujours des troncs communs. Pas question donc de réécrire toutes les données, ce serait une cause d'erreurs et par ailleurs la correction d'une erreur dans une partie pourrait ne pas être faite dans les autres. Pas question non plus de numéroter les diverses choses qu'il faut énumérer (telles que les numéros de page, etc.) : il faut pouvoir dire « suivant » (par exemple pour obtenir la page suivant celle en cours de lecture). Derrière la saisie du texte (et des images), il doit y avoir toute une machinerie qui gère ces aspects de partage d'informations et de navigation d'un endroit à l'autre. Lorsque j'avais commencé ce travail en vue d'une « simple » édition papier, j'avais utilisé un formateur de texte, LaTeX, peu connu du grand public mais permettant de faire de l'édition de très haute qualité et surtout de façon très paramétrée (en fait c'est plus un programme informatique qu'un système de traitement de texte). Lorsque j'ai repris ce travail pour le web, j'ai tout naturellement réutilisé ce système que je pouvais paramétrer à ma guise pour programmer toutes ces versions et tous ces numéros de façon quasi automatique. Je dispose donc d'une version globale qui, au travers de divers filtres, me donne les versions désirées. Modifier cette version d'en haut revient à modifier toutes les versions finales. Certes, il existe

sur le marché des produits de conception de pages web, mais je n'en ai pas trouvé qui permettent cette liberté de travail (même si certains d'entre eux sont malgré tout plus « conviviaux » que le vieux LaTeX qu'il faut programmer...). En fait, pour moi, tout se passe comme si j'écrivais en XML (un langage de description de documents disons plus propre que HTML qui a été, et est encore, très utilisé pour la grande majorité des pages web) et il sera toujours facile de convertir mon texte final en XML.

En revanche, j'ai tenu à ce que les fichiers mis sur le web suivent impérativement les normes internationales (ce qui est rarement le cas d'autres textes qui en général ne respectent que les usages d'un fabricant) et soient dépouillés de tout folklore des internautes (pas de lumière qui flashe, pas de son pour un texte qui n'en a nul besoin, très peu de couleurs). Finalement, je me suis rendu compte qu'on n'avait pas besoin de « feuille de style » élaborée pour un tel travail, la maquette standard de HTML suffisant à faire des choses très lisibles !

Pérennité

Quiconque a utilisé depuis quelques années Internet sait qu'il n'y a rien de plus volatile qu'un site... Que de fois se connecte-t-on à une adresse qui nous renvoie la célèbre « erreur 404 » (adresse inconnue, c'est-à-dire site disparu dans ce cas) !

Pour le moment, cette page électronique est sur le site de mon laboratoire d'où il n'a aucune raison de bouger... tant que je serai là pour vérifier qu'un administrateur n'aura pas décidé que finalement il n'est pas directement lié à l'activité de ce laboratoire aujourd'hui (c'est vrai maintenant), qu'il prend trop de place, que ceci ou cela.

Il faut donc que je lui trouve un site pérenne, d'où je sois garanti qu'il ne disparaisse pas et qu'il reste accessible¹. Heureusement, de bonnes pistes du côté d'instituts du livre ou de centres de typographie s'ouvrent.

CD, page Internet ou livre électronique ?

Nous n'avons parlé ici que d'une version du *Manuel typographique* sur le web. Mais sa conception « globale » permettrait aussi d'en faire une version pour CD ou sous forme de livre électronique. Pour terminer, montrons quelques différences.

(1) Il ne faut pas croire qu'une version électronique est une source de sécurité. Ainsi j'avais sauvegardé ma transcription du texte de Fournier sur un système de bureautique. Plusieurs années après, les disquettes étaient sans doute bien conservées, mais il n'existait plus de lecteur. Heureusement un collègue qui avait encore de vagues souvenirs a pu en extraire une partie, mais pas tout !

Un CD (ou DVD) est un objet que l'on achète en général, qui est physiquement chez soi (on le possède, on le voit sur son étagère, etc.), qui se suffit à lui-même (à condition bien sûr de disposer d'un ordinateur « récent »). Son mode de consultation préféré relève plutôt du codex, du livre relié, de la page papier mais avec les possibilités de liens (cliquer pour faire apparaître une autre partie du CD, mais aussi une page du web). Pour le Fournier, on gardera les images et pour le texte on prendra celui « mis en page » sous forme pdf.

Un document sur le web en revanche n'est pas physiquement chez le lecteur. Celui-ci en effet fait venir d'un site extérieur les pages qu'il veut consulter. Pour éviter les nombreux transferts que demanderait un chapitre de livre par exemple, on a pris l'habitude d'envoyer celui-ci d'une seule traite, sans divisions en pages : on est ainsi retourné au mode de lecture antérieur au Moyen Âge, celui du *volumen*, du rouleau que l'on fait défiler derrière son écran. Ces transferts peuvent toutefois être longs, surtout pour des images, et ne conviennent alors pas toujours à des lecteurs ne disposant pas des toutes dernières nouveautés (par exemple des étudiants ou des personnes vivant en dehors des grands centres urbains notamment dans certains pays, voire certaines régions, hors des circuits modernes de télécommunications). En revanche, ces documents disponibles sur le web sont plus fréquemment « libres » (de droits), gratuits. Et de plus, ils peuvent être mis à jour en continu.

Enfin, le livre électronique¹ relève un peu des deux autres méthodes : comme pour les CD, le texte est en quelque sorte fermé, mais il est téléchargé depuis une bibliothèque (on n'est plus propriétaire de l'œuvre, mais locataire). En revanche, le livre électronique se lit sur un ordinateur de poche muni de hautes qualités de lisibilité (écran sans brillance et technologiquement très précis, fait pour la lecture continue de plusieurs heures) et d'ergonomie (fonctions de navigation notamment). Pour le Fournier, il suffira de suivre les normes d'écriture en cours.

Jacques André

Jacques André est directeur de recherche à l'Irisa (Institut de recherche en informatique et automatique) à Rennes où il s'occupe de typographie et de documents numériques depuis plus de vingt ans. Il est par ailleurs rédacteur en chef de revues comme Document numérique et les Cahiers Gutenberg.

(1) Sur ce concept de livre électronique, voir Claire Bélisle (sous la direction de), *Le Livre électronique*, Éditions de l'ENSSIB, Lyon, 2004 (sous presse).