

Méthode A à Z pour créer un ebook

Coolmicro

avec l'aide de **Rémy**

Sommaire

Introduction	4
I LE SCAN.....	5
II L'OCR.....	9
1 – Les options de FineReader.....	9
2 – FineReader en action.....	12
III LE TRAITEMENT DE TEXTE.....	16
1 – Mise en forme et correction	16
1 ^{er} niveau de travail.....	16
2 ^e niveau de travail	23
3 ^e niveau de travail	25
2 – Ce qu'il faut faire et ne pas faire avec Word (ou un autre traitement de texte)	25
Les styles Titre 1,2, 3.....	26
À propos des sauts de page.....	26
À propos de sauts de paragraphe multiples	28
À propos des retraits et indentations	28
Autres erreurs à éviter	28
IV CONVERSION AUX FORMATS EBOOKS.....	29
1 – Format ePub.....	29
Méthode Amanuensis.....	29
Méthode Atlantis	29
Méthode Calibre	30
2 – Format Mobipocket / Kindle.....	31
3 – Format PDF	31
V ANNEXES : POUR ALLER PLUS LOIN.....	33
1 – La typographie.....	33
2 – La mise en page (MEP).....	34
3 – Utilisation des styles.....	34
4 – Les macros	34
5 – La relecture	35
CONCLUSION.....	36

Licence 36

Introduction

Cette méthode n'a surtout pas vocation à être universelle. Elle ne se veut qu'une description de la méthode que j'utilise, rodée sur des années de travail, mais utilisant mes outils, souvent payants et fermés (car, bien que supporter du logiciel libre et même utilisateur occasionnel de Linux, je n'ai pas trouvé dans le Libre les outils qui me satisfaisaient, à l'exception notable de 2 logiciels). Par ailleurs, j'ai certains principes concernant les ebooks et le travail informatique en général :

1. J'aime avoir un résultat de qualité, je suis assez exigeant, mais je ne suis pas non plus un maniaque du détail, il est hors de question de travailler 2 heures pour améliorer quelques trucs que personne ne verra dans un ebook. Une mise en page simple sera toujours plus facile à réaliser qu'une mise en page compliquée. Donc, ne comptez pas sur moi pour vous apprendre à mettre des lettrines ou des trucs de ce genre : je considère que cela n'est pas essentiel à la lecture. Par contre, la typographie du texte doit être correcte pour que la lecture soit agréable.
2. J'essaye de trouver à chaque fois la méthode la plus simple et la plus rapide et je choisis mes outils en fonction de cela.

I

LE SCAN

Tout commence par le scan, sauf si vous avez la chance de trouver un scan déjà prêt, auquel cas vous passez au chapitre II...

Le scan est une opération extrêmement simple (et fastidieuse) dont le résultat dépend avant tout de la qualité du capteur du scanner. En dehors des scanners spécialisés pour les livres (en général beaucoup trop chers, à l'exception des [Optibooks Plustek](#) qui sont loin d'être parfaits et du futur [Booksaver](#) pour lequel j'ai beaucoup d'espoirs), il y a 2 catégories de scanners à plat :

- * Les scanners personnels avec capteur CIS, qui donnent globalement des résultats médiocres et obligent à littéralement écraser le livre (et donc bousiller la reliure) pour obtenir un scan à peu près correct.
- * Les scanners avec capteur CCD, qui sont très nettement supérieurs aux précédents. Un capteur CCD est presque obligatoire pour quelqu'un qui scanne beaucoup de livres. Problème ; il n'y a presque plus de scanners personnels avec ces capteurs, et les prix s'envolent vite.

Les autres critères concernant les scanners sont sans importance, hormis bien sûr la vitesse, qui est un critère subjectif dépendant de la patience et de la masse de scans de chacun.

Je n'entrerai pas dans les réglages du scanner. Tout d'abord parce qu'ils varient d'un scanner à l'autre, et d'un logiciel à l'autre. Et ensuite parce que je les utilise peu personnellement, hormis les réglages de bases que je vais indiquer ci-après et qui sont **impératifs**.

1^{re} règle : Une fois qu'on a scanné un livre, avec les paramètres que je vais indiquer, il faut rescanner les illustrations, avec des paramètres différents.

Scan du livre : Il faut choisir l'option **noir et blanc**, parfois appelée également OCR ; il faut éviter absolument les options, couleurs, niveaux de gris, photos, texte/photos pour ne citer que quelques exemples. La meilleure résolution est **300 dpi** (moins, c'est périlleux, sauf si vous avez une grande expérience, plus, cela donnera des images énormes pour un résultat paradoxalement inférieur).

Scan des illustrations éventuelles : couleurs, ou niveaux de gris ou photos selon la nature de l'illustration ; 300 dpi minimum, mais il peut être intéressant dans certains cas de monter à 400 dpi : tout est affaire de test dans ce domaine, de rendu visuel.

2^e règle : Le mieux quand c'est possible, est de scanner la livre ouvert contre la vitre, c'est-à-dire 2 pages à la fois (tout dépend évidemment de la taille du livre et du scanner). Il est très important que **le livre soit toujours au même endroit de la vitre du scanner**, il est donc conseillé de caler le livre dans des coins, toujours le même. Par ailleurs, il faut que le livre soit bien plaqué contre la vitre, quitte à abimer un peu la reliure (avec un capteur CIS, c'est fondamental, et il faut écraser complètement la reliure), et **ne pas bouger le livre** pendant la numérisation d'une page (ou plutôt 2, si vous avez mis le livre à plat).

Parmi les réglages proposés par les scanners/logiciels de scan, il y a souvent le couple luminosité/contraste : là, il faut essayer, tâtonner. Augmenter le contraste est souvent dangereux, surtout sur les vieux livres, car cela fait ressortir les défauts, rousseurs, etc. qui seront ensuite traités comme des caractères par le logiciel d'OCR, augmentant ainsi le temps de correction. Il peut être intéressant d'augmenter un peu la luminosité pour diminuer la zone

noire correspondant à la reliure si vous ne pouvez pas écraser la reliure, et surtout si vous avez un capteur CIS. Tout est affaire de test : il faut d'abord bien étudier les 2 ou 3 premières pages scannées, voire même les tester en OCR si on n'a pas l'habitude, avant de se lancer dans le scan du livre complet.

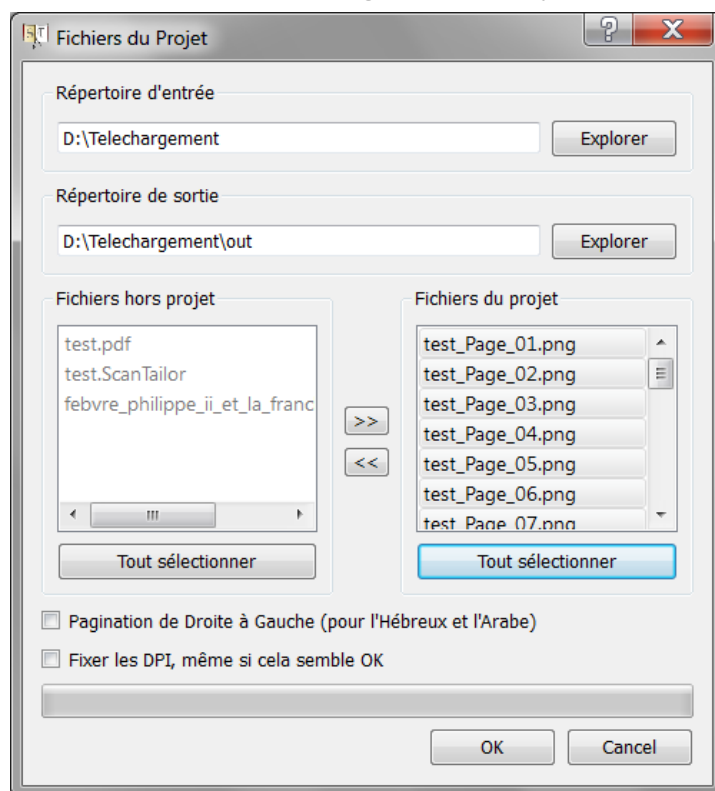
Parlons maintenant du **format de l'image** à choisir pour le scan. Ce choix n'est pas déterminé par l'OCR, tous les logiciels d'OCR acceptant la plupart des formats en entrée, mais par le logiciel que vous allez utiliser avant l'OCR pour améliorer vos images de scan :

* Si vous avez les outils nécessaires pour travailler les PDF (ce sont malheureusement des outils payants : Adobe Acrobat, pour les plus fortunés, mais aussi l'excellent Nuance Converter pro, beaucoup moins cher), cela s'avère un format très pratique, et c'est en général la solution que j'utilise...

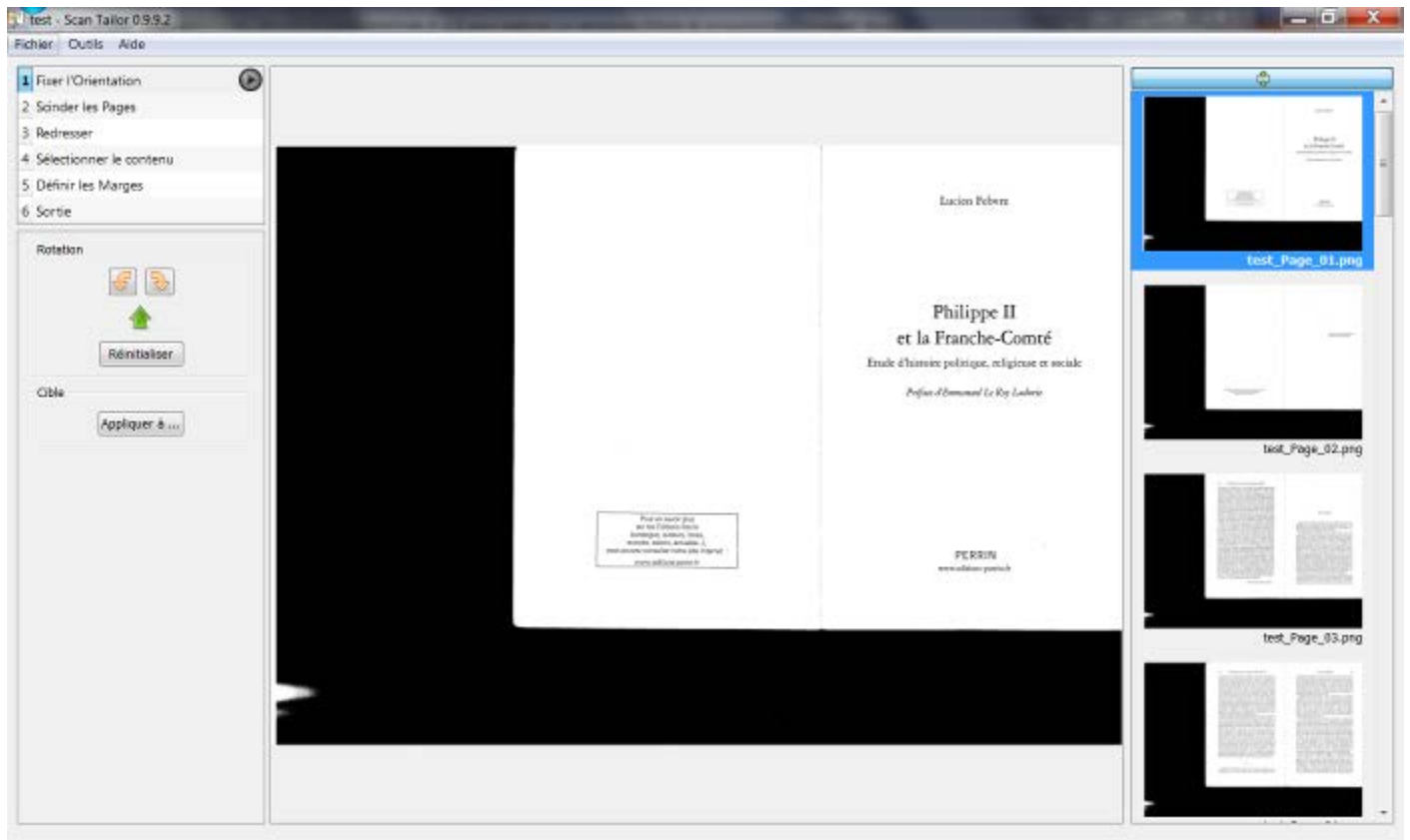
* Mais vous pouvez tout simplement choisir un format image classique, jpg, png ou tiff. Je vous conseille alors d'utiliser l'excellent [ScanTailor](#) (libre et gratuit) pour retravailler vos images.

Évidemment, il faut veiller dans le logiciel de scan au nommage des fichiers qui doivent impérativement être nommés de manière séquentielle, le nom des fichiers se terminant par 001, 002, 003, etc. pour que tout se passe bien lorsque vous allez retravailler les images. Par ailleurs, si vous n'avez pas suivi mes conseils concernant la place fixe du livre du livre sur la vitre lors du scan, vous êtes mal barré... En effet, le principal traitement qu'il faut faire subir aux images, ou au PDF, c'est un recadrage, et c'est plus facile quand le texte ne balade pas dans tous les coins.

Attardons-nous sur ScanTailor, puisqu'il est gratuit, et voyons ce qu'il nous propose :



1. Commençons par créer un nouveau projet. Indiquez le dossier d'entrée où sont stockées vos images de scan, le dossier de sortie pour les images modifiées, cliquer sur « Tout sélectionner » sur la fenêtre de droite où apparaissent les images, puis cliquer sur OK. Vous obtenez l'interface ci-dessous, avec sur la gauche les 6 étapes de traitement que je vais détailler :



2. Fixer l'orientation : Normalement, vous laissez tel que puisque votre scan est parfait... Dans le cas contraire, vous pourriez appliquer une rotation avec les 2 icônes. La cible « Appliquer à... » se retrouve dans toutes les étapes avec le choix d'appliquer la modification à : la page active / toutes les pages / cette page et les suivantes / les pages sélectionnées, etc. Si vous avez fait une ou des modifications, vous devez ensuite cliquer sur le bouton « Play » (même pictogramme que pour un lecteur de DVD) pour appliquer effectivement les changements.

3. Scinder les pages : Même si les logiciels d'OCR le font très bien, si vous avez scanné 2 pages à la fois, c'est bien de le faire dans ce logiciel. Vous avez le choix entre 3 modes de détection automatique et un mode manuel, appliqué à la page active / toutes les pages / cette page et les suivantes / les pages sélectionnées. Malheureusement, cette option fonctionne très mal si vous avez de grosses bandes noires comme dans l'image ci-dessus, ce qui arrive si vous avez scanné un livre d'une taille donnée, dans une taille automatique plus grande sur le scanner. Dans ce cas, choisissez l'icône de gauche sous « Mise en page », ce qui signifie que vous ne voulez pas scinder les pages : vous le ferez plus tard dans le logiciel d'OCR.

4. Redresser : Si le scan est bien fait, cette option est inutile, d'autant que la fonction de redressement est également incluse dans les logiciels d'OCR.

5. Sélectionner le contenu : Si vous avez appliqué mes précédents conseils, il suffit de laisser le mode automatique, et de cliquer sur « Play ». Mais, si le logiciel ne reconnaît pas le contenu et sélectionne des zones blanches ou noires, il vous sera simple de définir manuellement le masque du contenu si vous avez bien veillé à mettre le livre toujours au même endroit de la vitre.

6. Vous choisissez les marges (je prends 5 mm pour les 4), vous laissez coché « Faire correspondre la taille avec les autres pages », et vous cliquez sur le bouton « Play ».

7. Il ne reste plus qu'à cliquer sur la Sortie. Vous choisissez la résolution de sortie : restez en 300 DPI, le mode : rester en noir et blanc (avec la possibilité d'éclaircir ou d'épaissir, page par page, ou toutes les pages d'un coup). Et enfin vous pouvez tenter d'éliminer le bruit (faire un test sur 2 ou 3 pages d'abord...). Et quand vous avez fini, cliquez sur le bouton « Play »...

Un must...

II

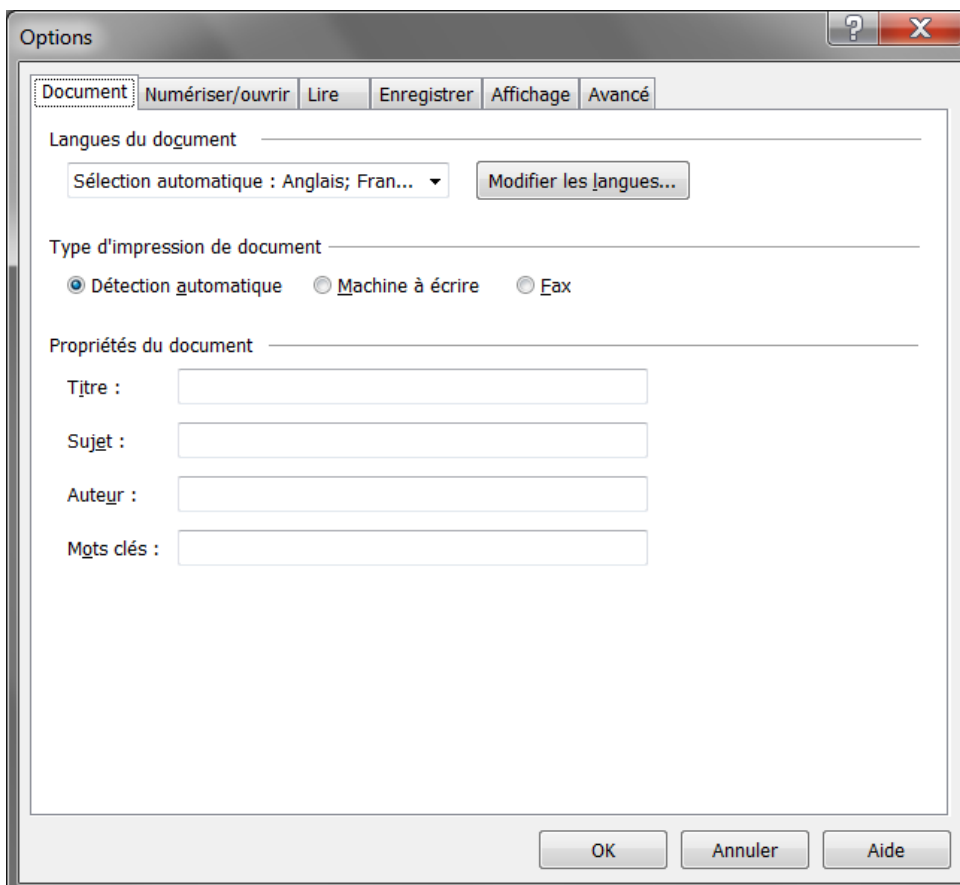
L'OCR

Plusieurs logiciels d'OCR existent, mais je dis clairement que j'en préfère un, par ailleurs utilisé par la très grande majorité de ceux qui œuvrent dans le domaine des ebooks, il s'agit de Abby FineReader. J'utilise actuellement la version 10, mais les versions 8 et 9 conviennent parfaitement. Je ne vous parlerai donc que de lui...

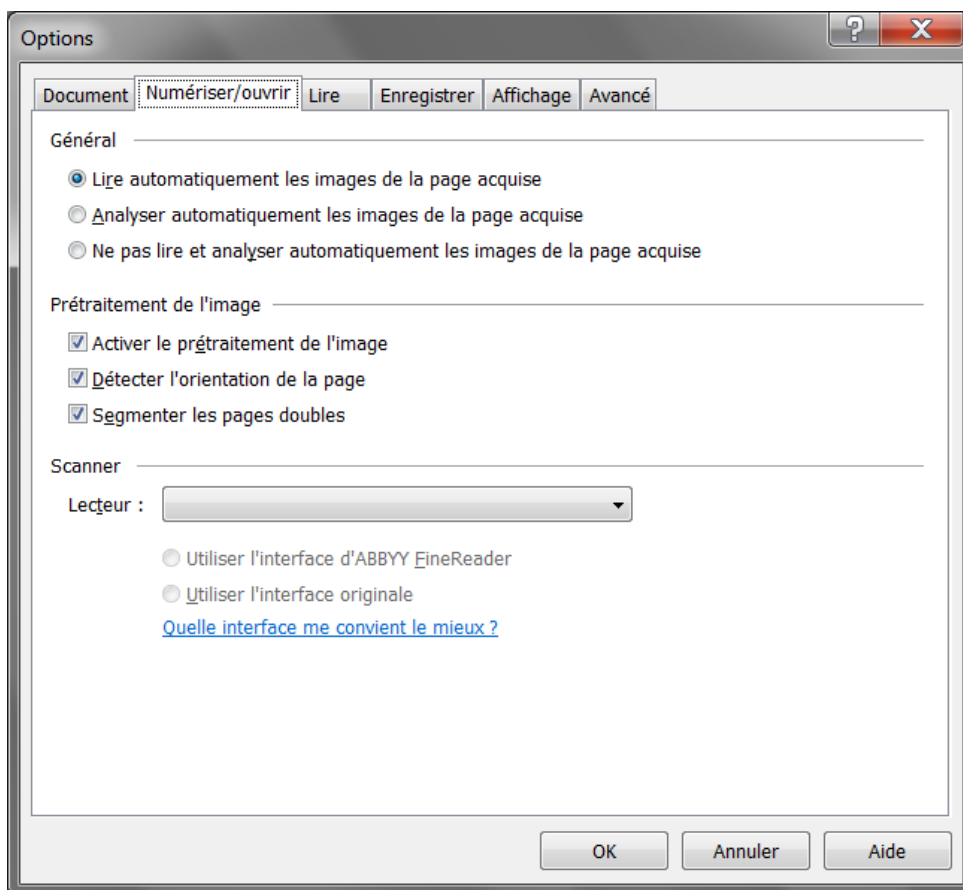
1 – Les options de FineReader

La première à chose est de comprendre les options de FineReader, d'appliquer les bonnes options, donc direction menu Outils / Options. Regardons tout de suite l'onglet « Avancé », le dernier. On voit qu'on peut enregistrer, ou charger les options à partir d'un fichier. Voici un lien vers le [fichier d'options](#) que j'utilise en général et que je vais décrire maintenant. Que vous utilisiez mon fichier d'option ou le vôtre, je vous conseille de procéder ainsi, car il est plus facile de changer éventuellement une ou deux options que de modifier à chaque fois tous les onglets.

Revenons au premier onglet ; je le laisse en général tel que :

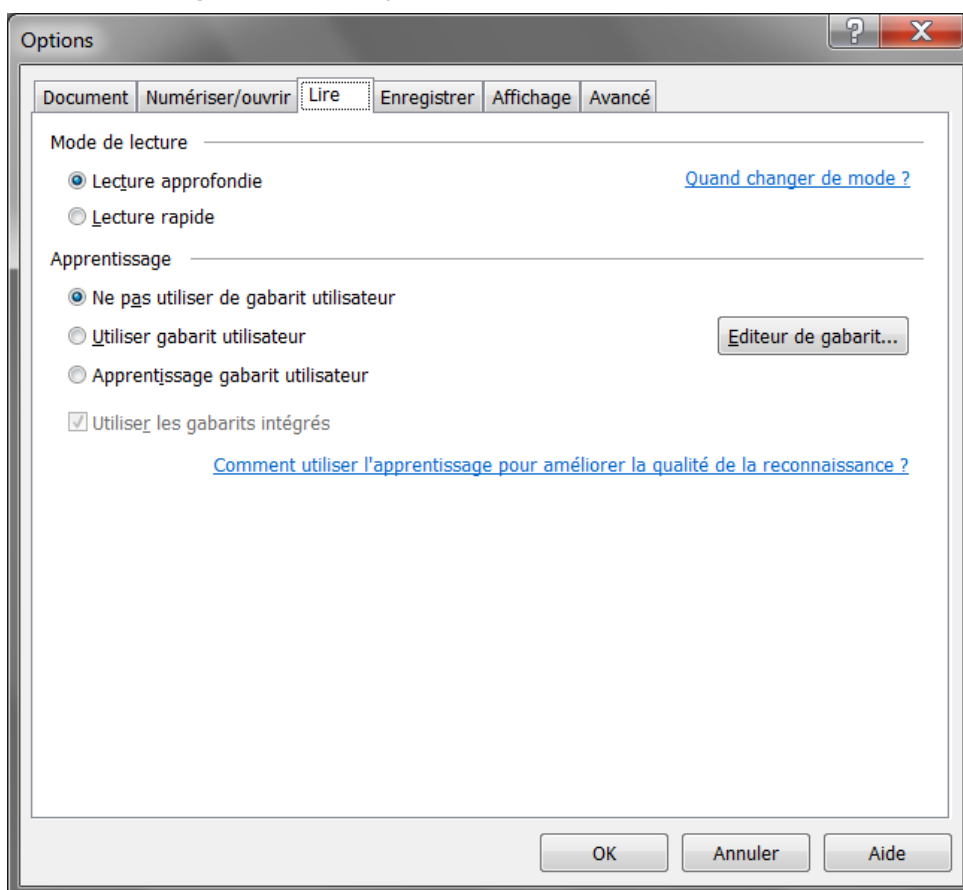


Deuxième onglet, « Numériser/Ouvrir » :



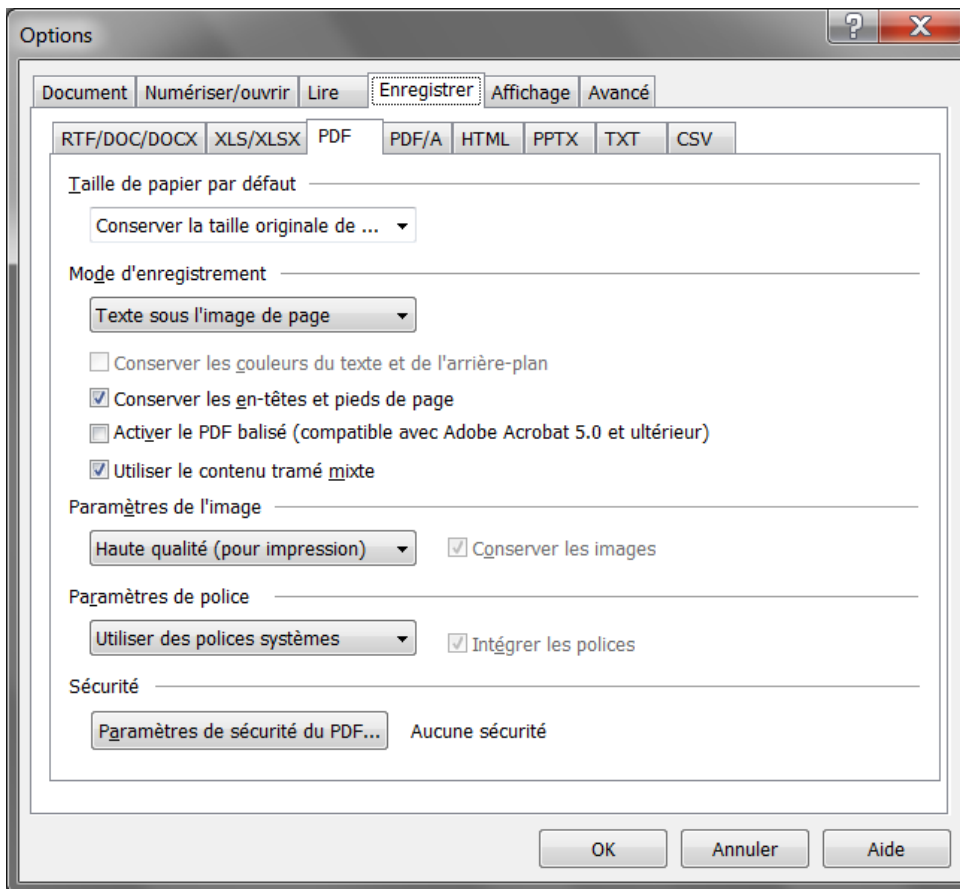
Bien sûr, je décoche la segmentation des pages doubles si je n'ai pas scanné les pages par deux, ou si j'ai fait la séparation dans un autre logiciel précédemment.

Troisième onglet, « Lire », je laisse tel que :



Je laisse ceux qui le désirent découvrir dans l'aide de FineReader l'intérêt des gabarits utilisateurs.

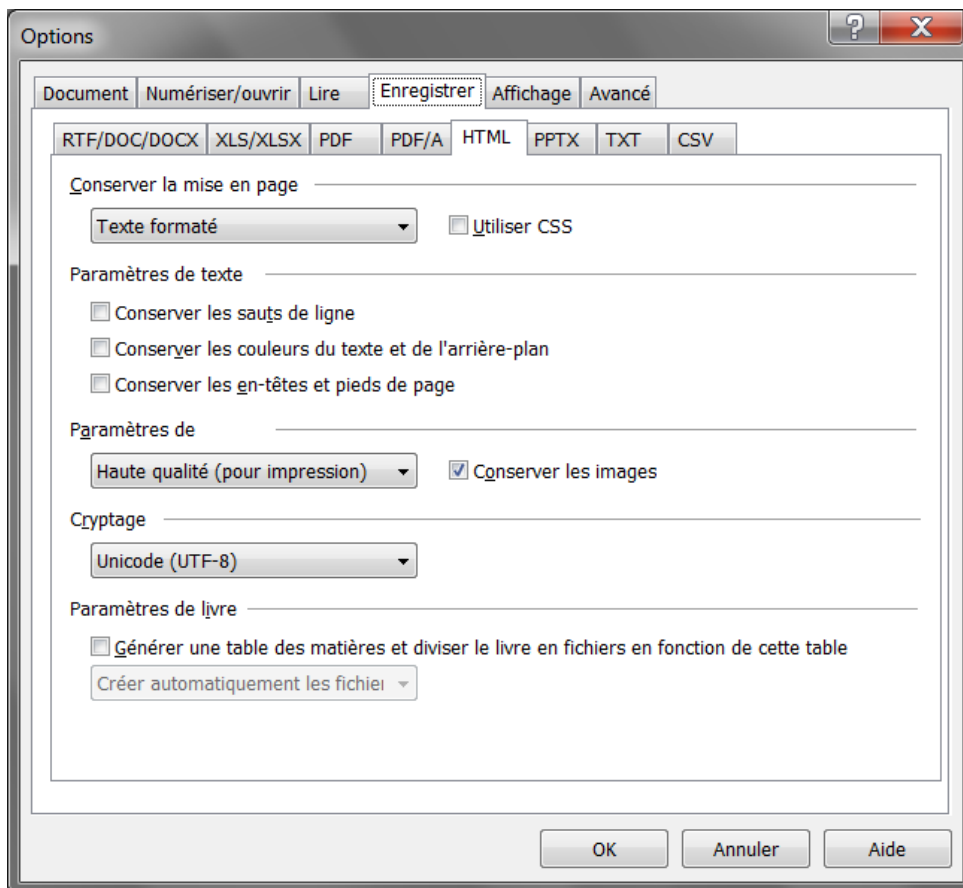
Quatrième onglet, « Enregistrer », le plus important, qui comporte de nombreux sous-onglets. Je ne vais en commenter que 2 et vous expliquer pourquoi.



Il faut toujours faire une sortie PDF qui va devenir votre « vrai scan », celui que vous utiliserez ensuite pour vos corrections / relectures. Les 2 premières options de cet onglet sont **fondamentales** :

* « Conserver la taille originale de l'image » vous permet d'avoir un PDF de la même taille que vos images, au lieu d'un PDF A4 avec votre image perdue au milieu de grands blancs.

* « Texte sous l'image de page » permet de conserver l'image de scan originale, nécessaire pour les corrections, mais de mettre au-dessous une couche de texte invisible avec le texte issu de l'OCR, texte qui sera brut ou corrigé selon que vous ferez le premier niveau de correction directement dans FineReader ou dans Word. Dans tous les cas, cela vous permettra de faire des recherches texte sur le PDF, ce qui est très pratique lors des corrections / relectures.



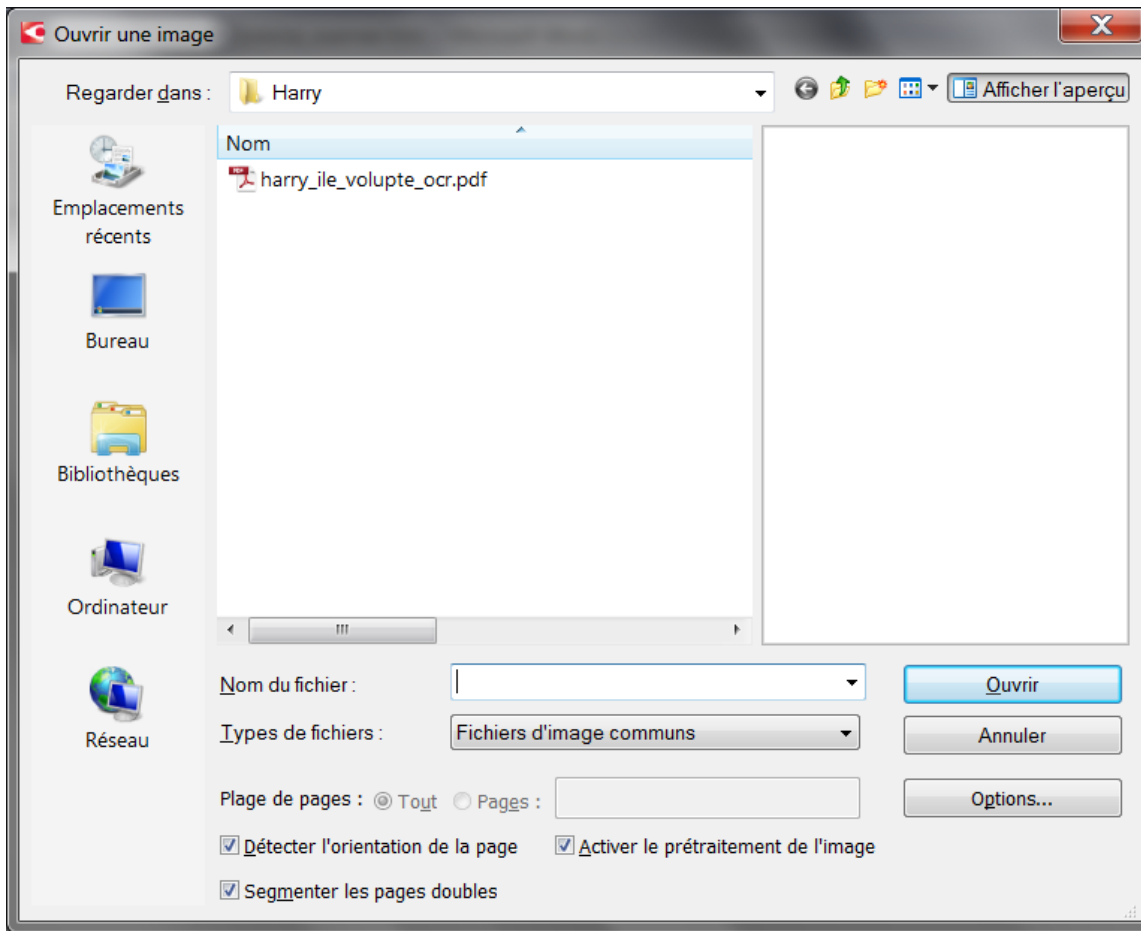
Vous remarquerez que j'ai choisi l'onglet HTML pour illustrer la sortie du texte lui-même, et non l'onglet RTF/DOC. Il y a une raison précise. Je considère que depuis les dernières versions, la sortie RTF/DOC est devenue une vraie catastrophe pour l'élaboration d'un ebook : création de dizaines de styles inutiles qu'il faut ensuite enlever, de marges délirantes, de sections pour simuler les sauts de pages du livre papier, toutes choses dont nous n'avons pas besoin. Le HTML nous fournit ce qui nous est nécessaire, et qui se résume à peu de chose : le flux de texte, le respect de l'italique. Même pour le HTML, je choisis les options « les plus simples », de façon à ce que mon flux de texte soit le plus pur possible, et que j'ai donc moins de travail ensuite. Nous ouvrirons ensuite le fichier HTML dans le traitement de texte, Word en ce qui me concerne et dans le cadre de ce tutoriel, et l'enregistrerons au format DOC.

2 – FineReader en action

Lorsque FineReader s'ouvre, il a déjà un document FineReader vide. Prenez l'habitude d'enregistrer immédiatement ce document vide – menu Fichier / Enregistrer le document FineReader. Le document FineReader est en fait un dossier (dossier caché pour la version 8), du même nom que celui que vous avez donné au document, et situé à l'endroit où vous avez enregistré le document ; il est donc facile de déplacer, sauvegarder ce document (pour une sauvegarde, nous vous conseillons de zipper le dossier, pas son contenu, le dossier lui-même)

Je n'aborderai pas la numérisation directe dans FineReader, pour la simple raison que je n'utilise pas un scan relié à un ordinateur. Mais cet aspect est très bien expliqué dans l'aide de FineReader. Je vais donc partir de l'hypothèse que le scan est déjà fait, par exemple avec le logiciel propre au scanner.

Cliquez sur l'icône ouvrir, et allez chercher votre PDF image ou toutes vos images (JPF, PNG ou TIFF) :



Vous voyez que vous pouvez cocher/décocher des éléments présents dans les options, tels que la segmentation des pages doubles par exemple.

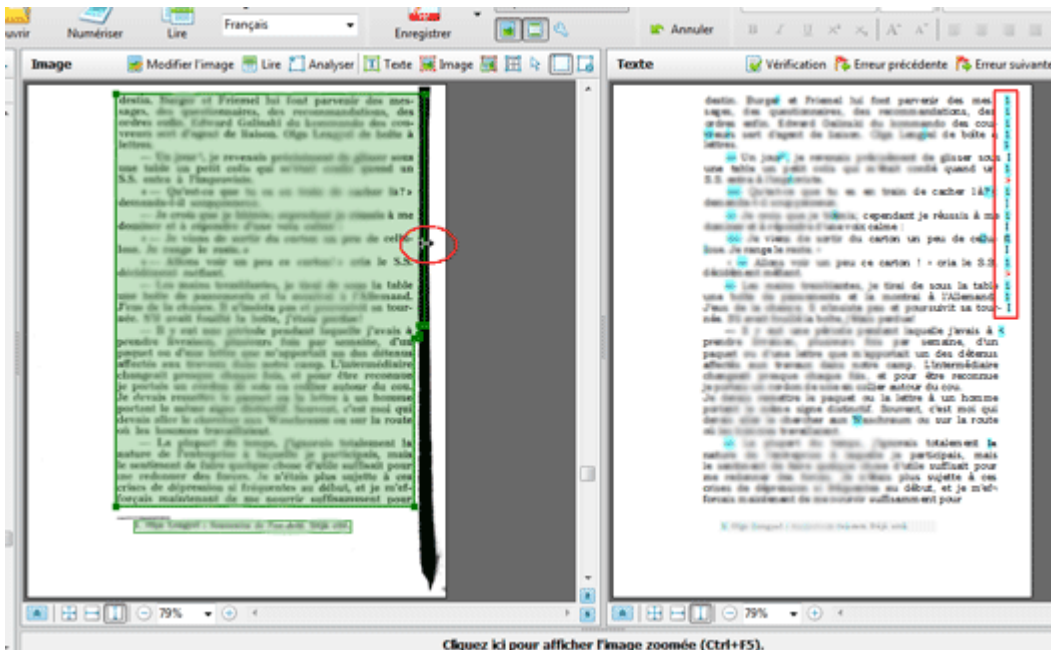
Il ne vous reste plus qu'à prendre un bon bouquin, aller faire un tour ou vous atteler à une autre tâche sur votre ordinateur. Car l'OCR, cela prend un certain temps... (rappel : dans les options, nous avons coché les cases pour que FineReader fasse simultanément l'ouverture, la lecture et la reconnaissance des images).

Une fois l'OCR terminé, vous avez 2 solutions :

- * Soit vous faites le premier niveau de correction du texte directement dans FineReader.
- * Soit vous faites tout de suite les sorties PDF et HTML, et faites ensuite la correction dans votre traitement de texte, Word au cas présent.

Personnellement, je préfère la seconde solution. Mais certains membres éminents de notre groupe (ELG), préfèrent la première solution. Tout est question d'interface. Tout dépend si vous maîtrisez bien votre traitement de texte ou si vous maîtrisez mieux l'interface FineReader. À vous de décider. Dans la suite de ce tutoriel, je partirai du principe que nous avons choisi la 2^e solution.

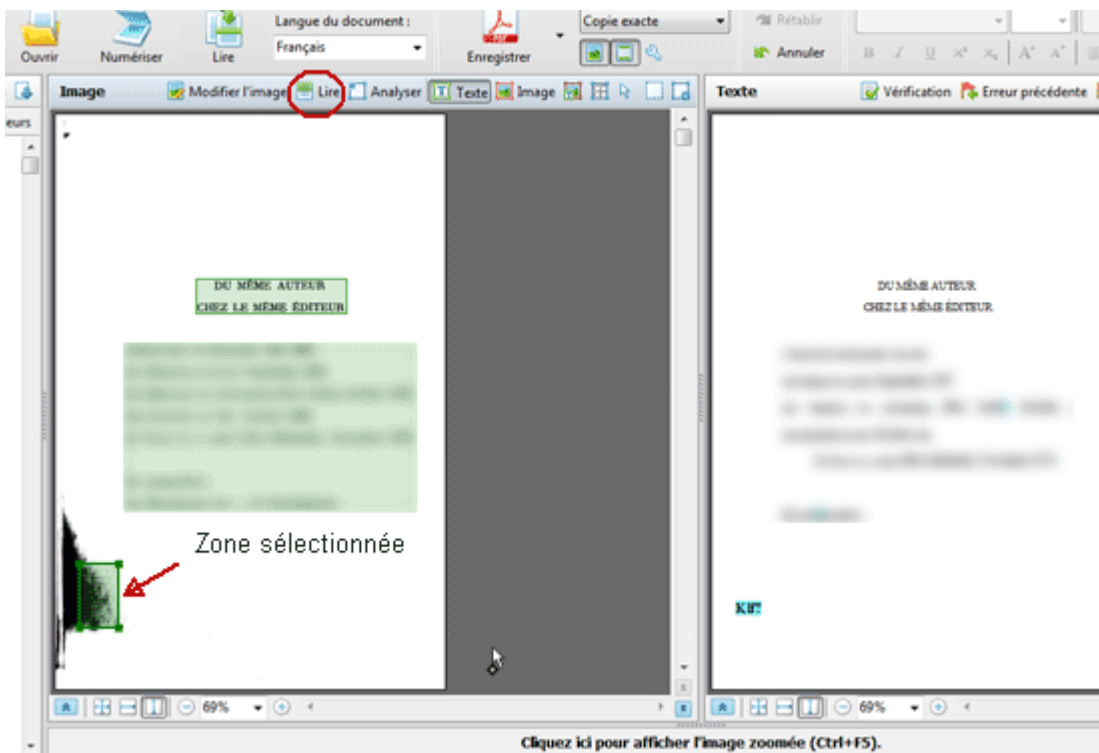
Que vous choisissiez l'une ou l'autre méthode, il est toujours utile de vérifier page par page les zones de reconnaissance qu'a sélectionné Finereader. Une zone trop large incluant l'ombre de la reliure provoquera des erreurs d'OCR qu'il est facile d'éviter : il suffit pour cela de redimensionner la zone de texte à l'aide de la petite flèche double qui apparaît lorsque l'on passe la souris sur une des bordures de cette zone.



Cliquez ici pour afficher l'image zoomée (Ctrl+F5).

Il peut aussi y avoir des zones indésirables qu'il suffit alors de supprimer purement et simplement en tapant « Suppr » après avoir sélectionné cette zone (elle apparaît alors avec des bordures plus épaisses et des angles matérialisés par un rectangle).

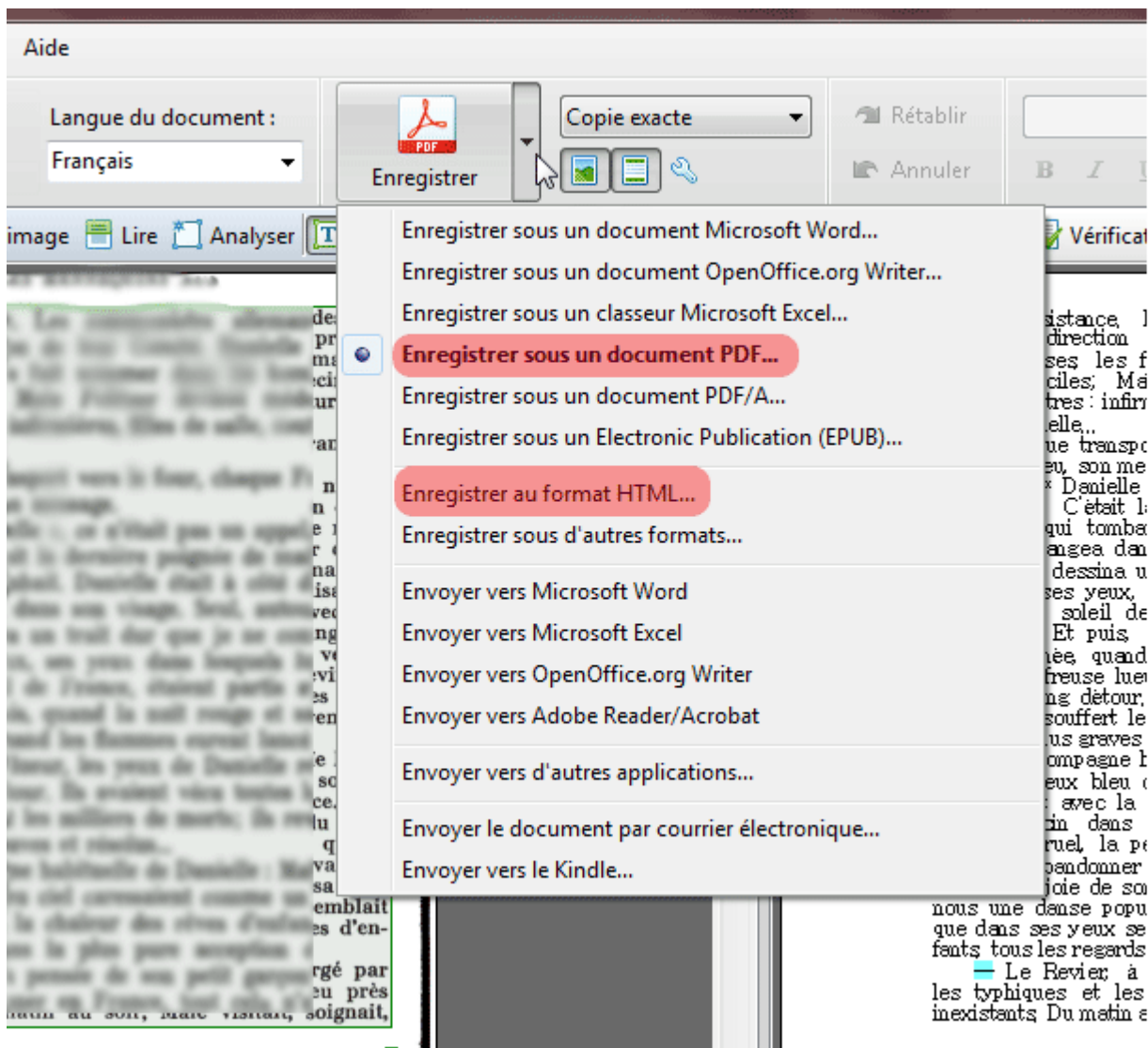
Rappel : les zones de texte sont matérialisées en vert, les zones d'image en rouge.



Cliquez ici pour afficher l'image zoomée (Ctrl+F5).

IMPORTANT : après avoir modifié une zone de reconnaissance il faut relire la page en cliquant sur le bouton « Lire » juste au-dessus de l'image pour enregistrer les modifications ou avec le raccourci Ctrl+R (et non pas sur le gros bouton « Lire » dans la barre de menu qui lit toutes les pages du document – Ctrl+Maj+R).

Pour enregistrer les sorties d'OCR, cliquez sur la flèche à côté de l'icône « Enregistrer » puis choisissez le format voulu, soit « Enregistrer sous un document PDF » puis « Enregistrer au format html ».



Maintenant que vous avez vos 2 fichiers de travail au format html et PDF, vous pouvez passer à l'étape suivante sous Word.

III

LE TRAITEMENT DE TEXTE

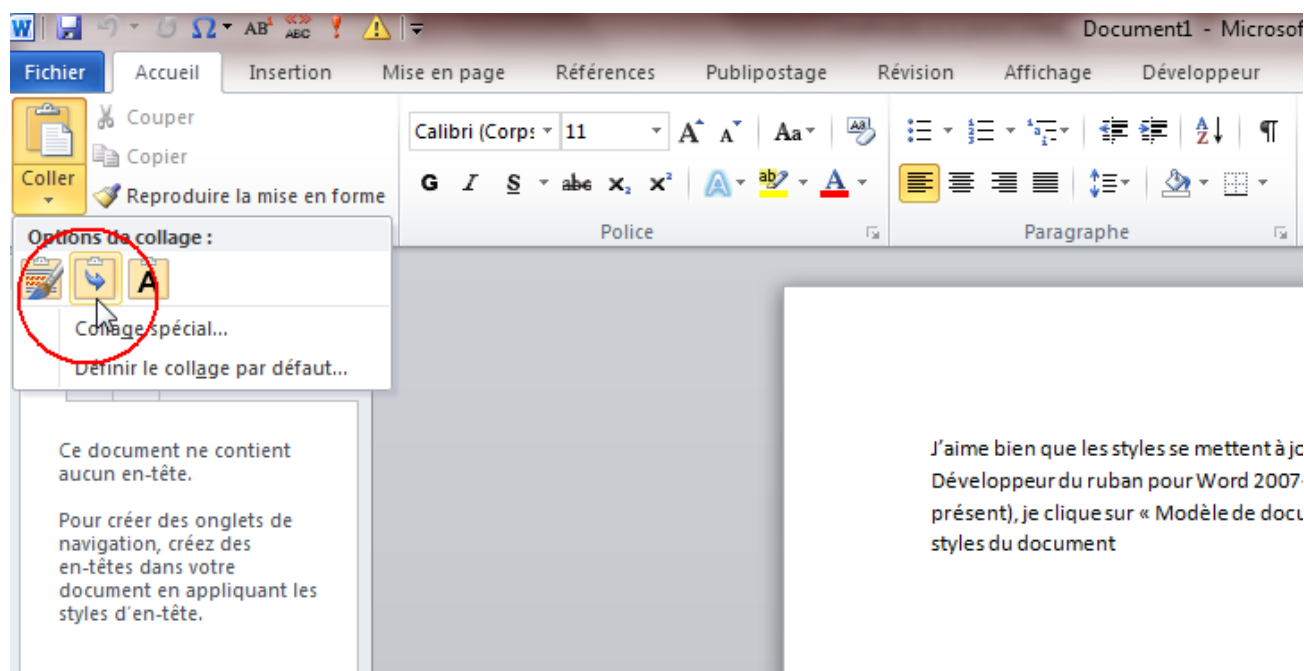
1 – Mise en forme et correction

1^{er} niveau de travail

Préalable : Pour appliquer ma méthode, il est nécessaire de télécharger un modèle, [epub_tuto.dot](#) (il est zippé, dézippez-le et copiez-le dans le dossier où se trouvent vos modèles).

Voilà, nous avons notre HTML, nous lançons Word, et nous allons chercher notre document HTML pour l'ouvrir. Nous sélectionnons tout le texte (Ctrl+A) et le copions (Ctrl+C).

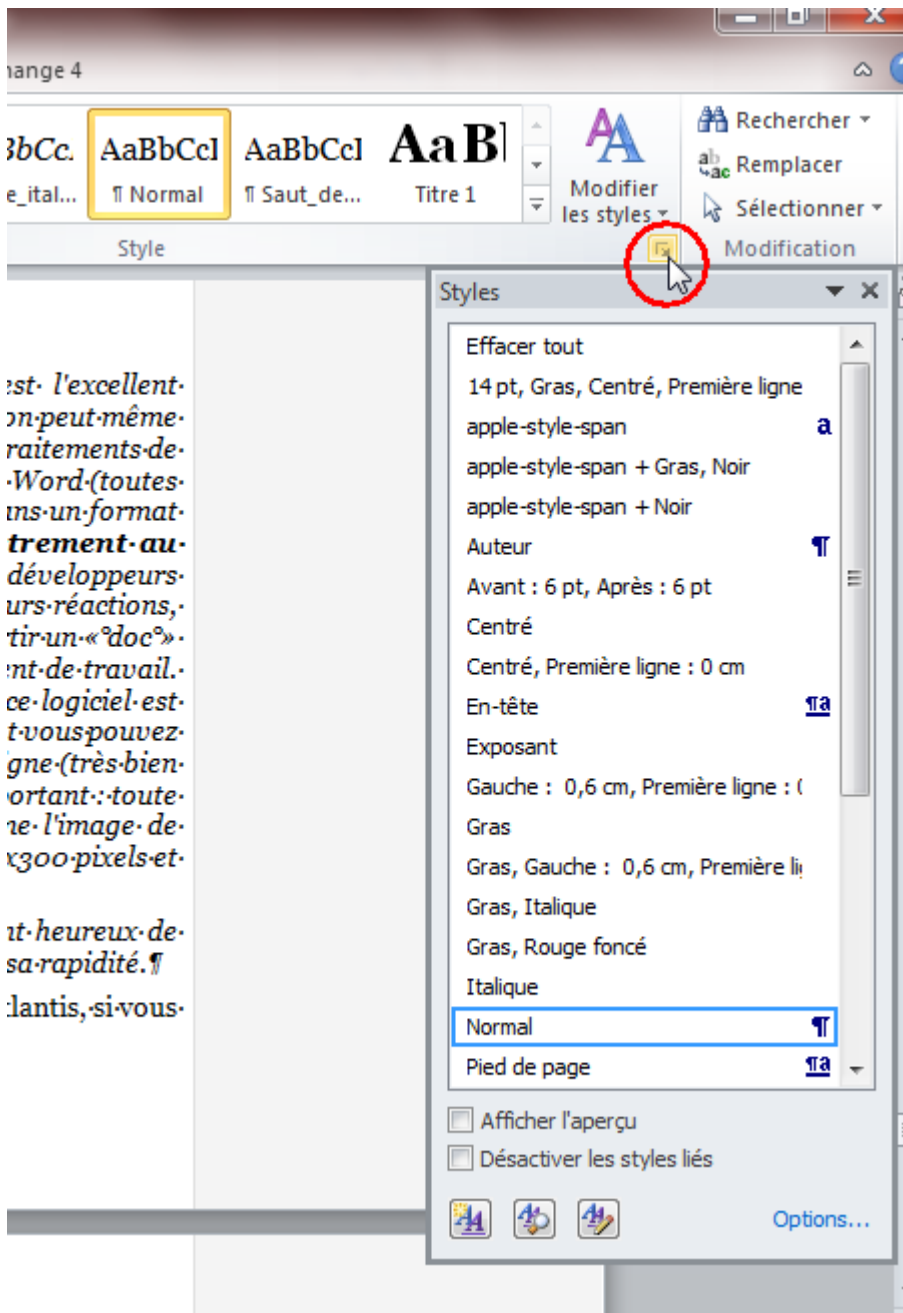
Puis menu Fichier / Nouveau / Mes Modèles et nous créons un document basé sur le modèle epub_tuto.dot. Nous collons maintenant tout ce que nous avons copié en **fusionnant la mise en forme** (voir ci-dessous) et nous enregistrons notre doc au format Word 2003 (pour ceux qui ont Word 2007 ou 2010, je vous déconseille d'utiliser le nouveau format docx si vous devez travailler en collaboration avec d'autres qui n'ont pas forcément, eux, cette version).



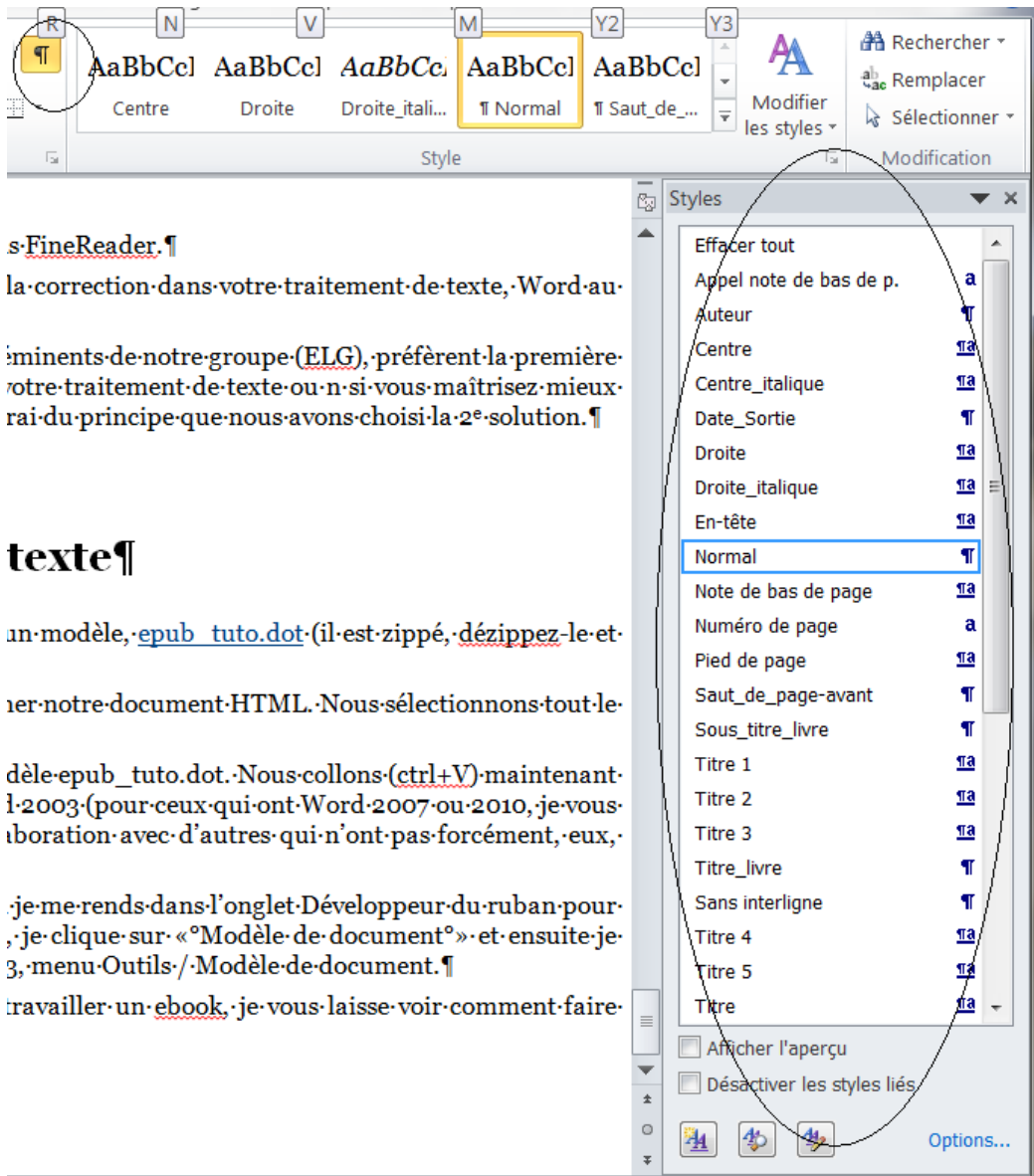
J'aime bien que les styles se mettent à jour à partir de mon modèle, aussi je me rends dans l'onglet Développeur du ruban pour Word 2007-2010 (le faire apparaître avec les options s'il n'est pas présent), je clique sur « Modèle de document » et ensuite je coche « Mise à jour automatique des styles du document ». Pour Word 2003, menu Outils / Modèle de document.

S'il y a des images dans notre doc, à nouveau Ctrl+A pour tout sélectionner, et ensuite Ctrl+Maj+F9 pour convertir les images qui sont encore extérieures au document (comme dans une page HTML) **en images intégrées au document**.

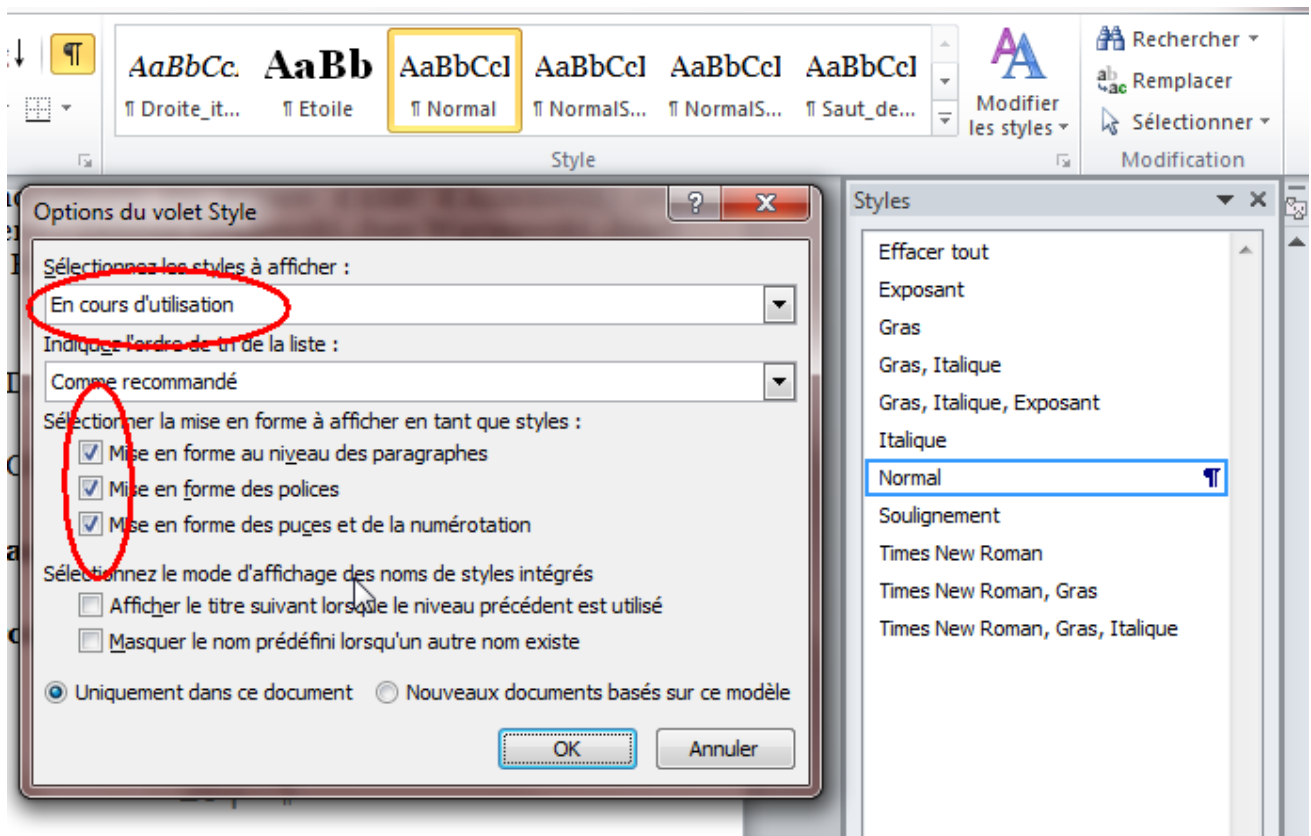
Je fais également apparaître la fenêtre des styles – indispensable pour travailler un ebook, en cliquant ici dans Word 2010 :



Et je mets **les marques de format visibles** (en haut à gauche de l'image ci-dessous) => **important**.

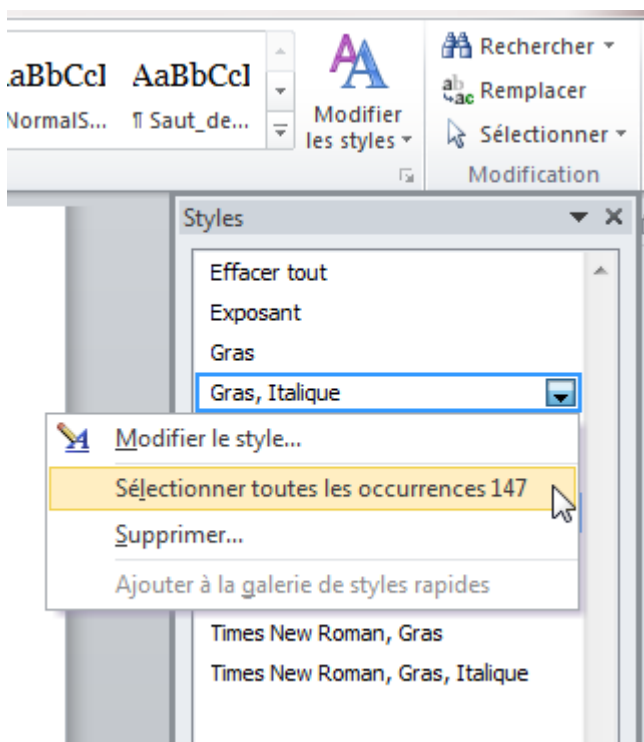


Je clique ensuite sur Options en bas de la fenêtre des styles, et j'obtiens la fenêtre suivante que je modifie comme ci-dessous (j'ai entouré ce qui est important) :



Dans la fenêtre des styles apparaissent alors les styles utilisés dans votre document. On distingue les styles de paragraphe, dont le nom est suivi du signe de saut de paragraphe (pied-de-mouche), et les styles de caractère, dont le nom est suivi par un « a » ou rien.

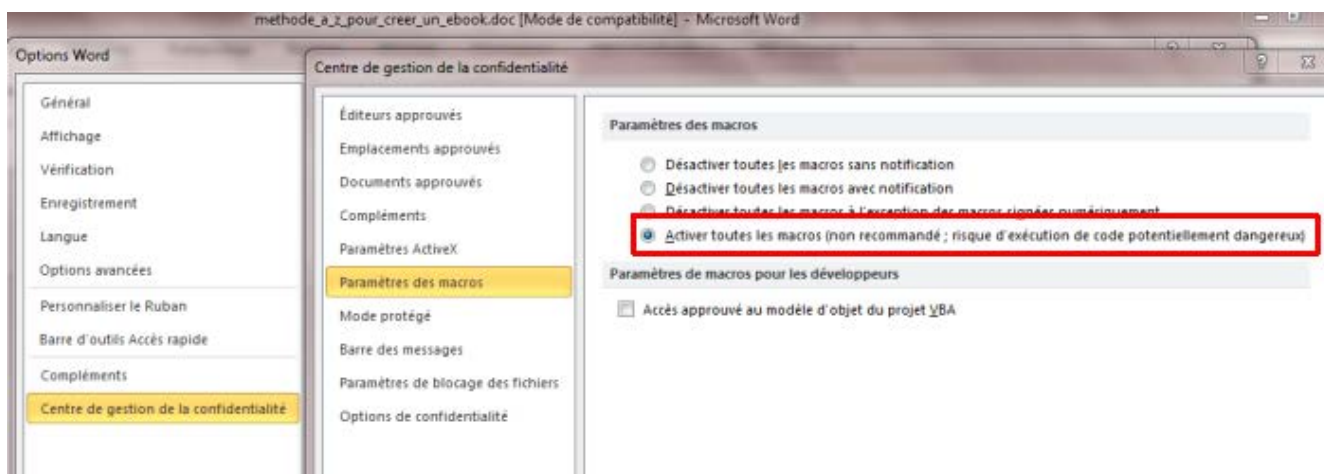
Si vous avez bien suivi la procédure décrite ci-dessus vous ne devriez avoir qu'un seul style de paragraphe : le style Normal. Les autres sont des mises en forme de caractère, certaines utiles comme les italiques et parfois les exposants, d'autres qu'il faudra supprimer car elles sont dues à des erreurs d'OCR (gras, soulignement, autre police,...). Pour ce faire, cliquez sur la flèche à côté du style à modifier :



Puis choisissez « Sélectionner toutes les occurrences ». Dans l'onglet Accueil supprimez alors la mise en forme non souhaitée (gras, soulignement). Les autres polices utilisées sont généralement des signes spéciaux non rendus en Georgia et peuvent très souvent être supprimés purement et simplement, mais ça ne coûte rien de vérifier quand même en les sélectionnant et en les supprimant une à une.

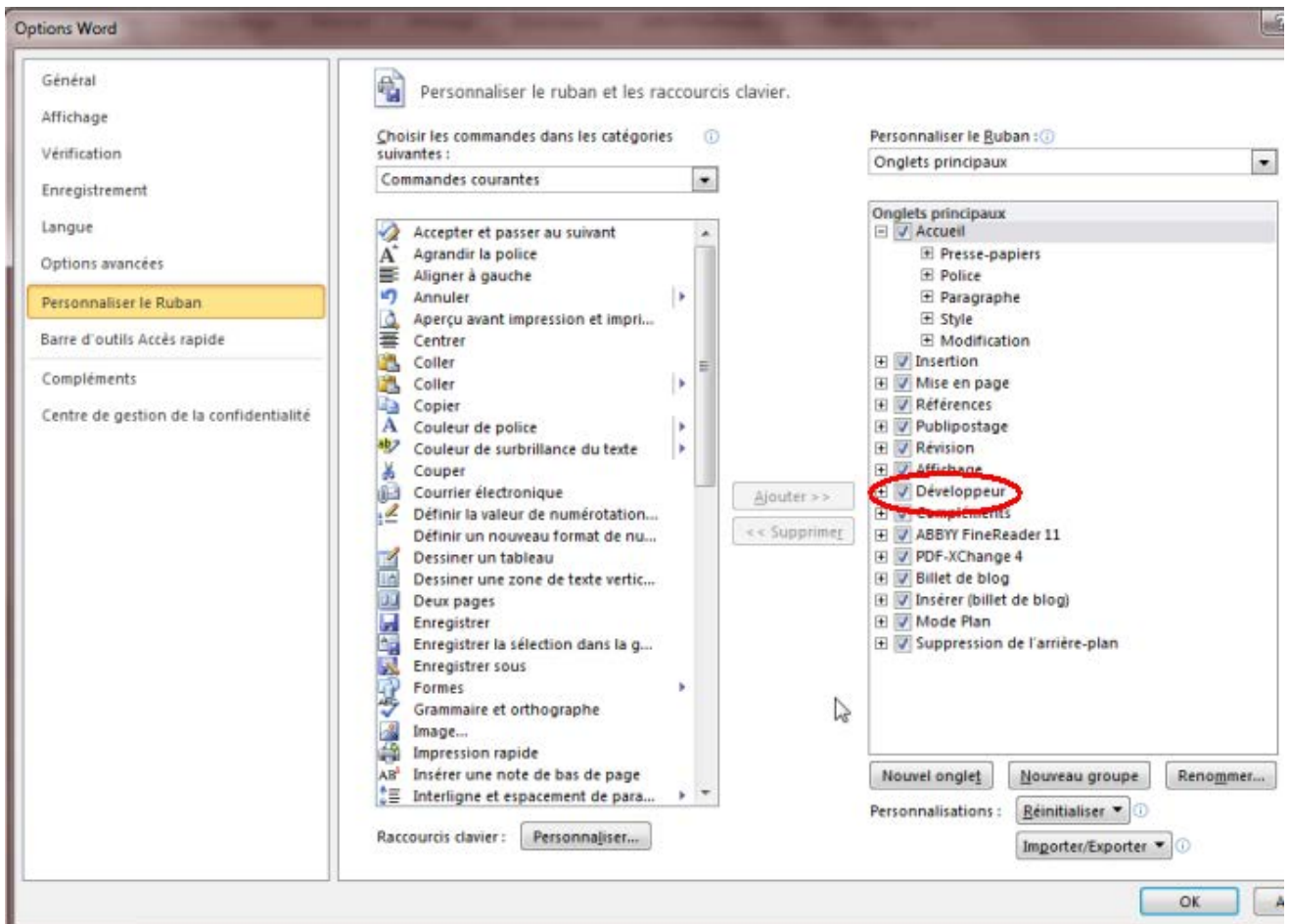
Voilà, maintenant il ne vous reste plus que le style Normal, avec pour seule mise en forme l'italique et éventuellement l'exposant. Ce style Normal est par défaut paramétré pour gérer la césure des mots. Certains préféreront – pour la commodité de lecture, mais surtout pour mieux détecter les mots coupés par l'OCR – supprimer ce paramétrage : pour cela modifiez le style, dans Format / Paragraphe onglet Enchaînements, décochez « Ne pas couper les mots » (ne pas oublier de rétablir la césure sur le document Word définitif, notamment si vous voulez générer un PDF à partir de ce doc)

Nous allons maintenant exécuter la macro de mise en forme, qui se nomme Typo, présente dans le modèle. Problème, si vous avez les dernières versions de Word, il faut que vous ayez configuré votre logiciel de manière à pouvoir exécuter les macros. Dans Word 2010 cela se trouve dans Fichier / Options / Centre de gestion de la confidentialité (si vous ne savez pas le faire pour une autre version, voir les forums ou les groupes de discussion pour résoudre ce problème).



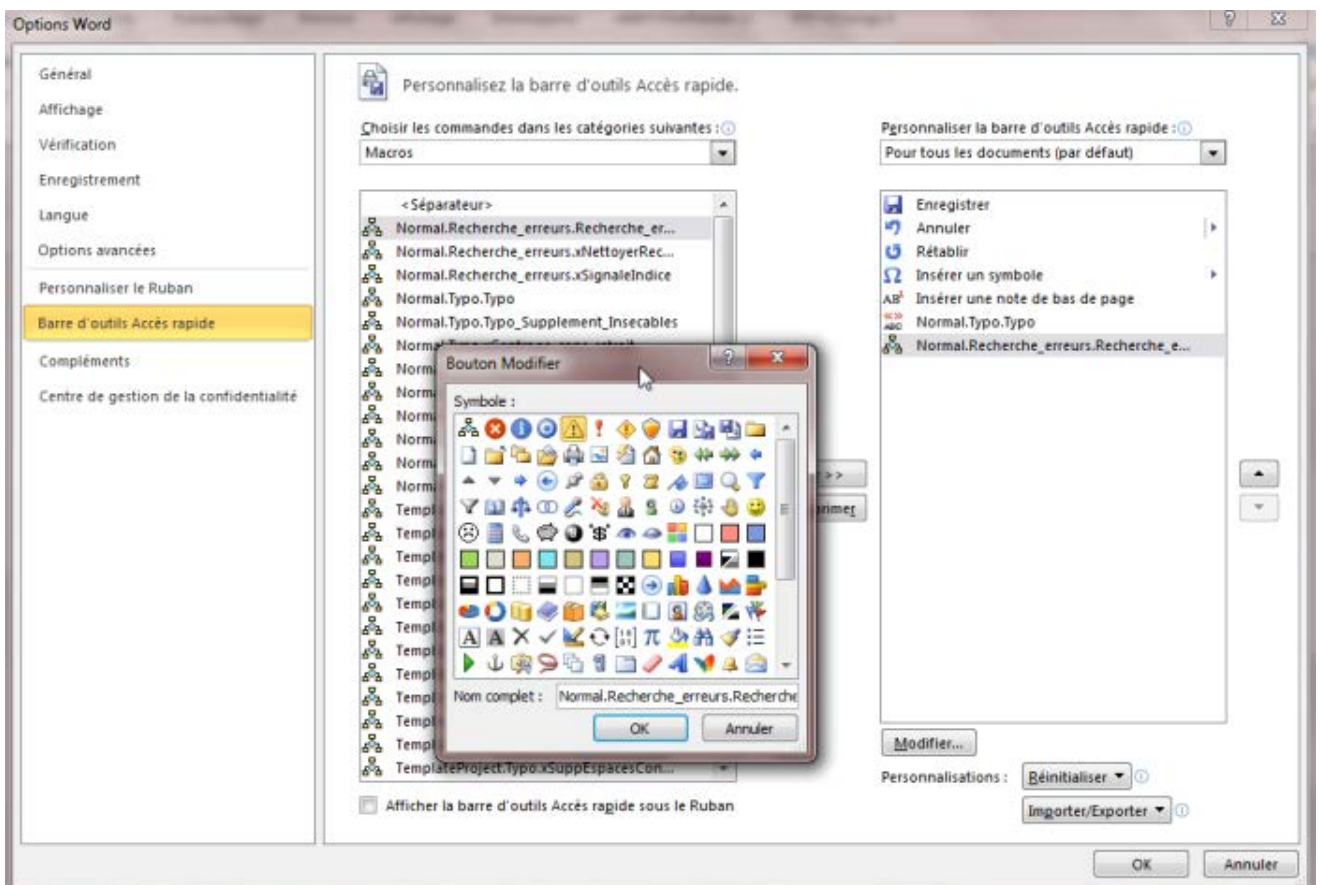
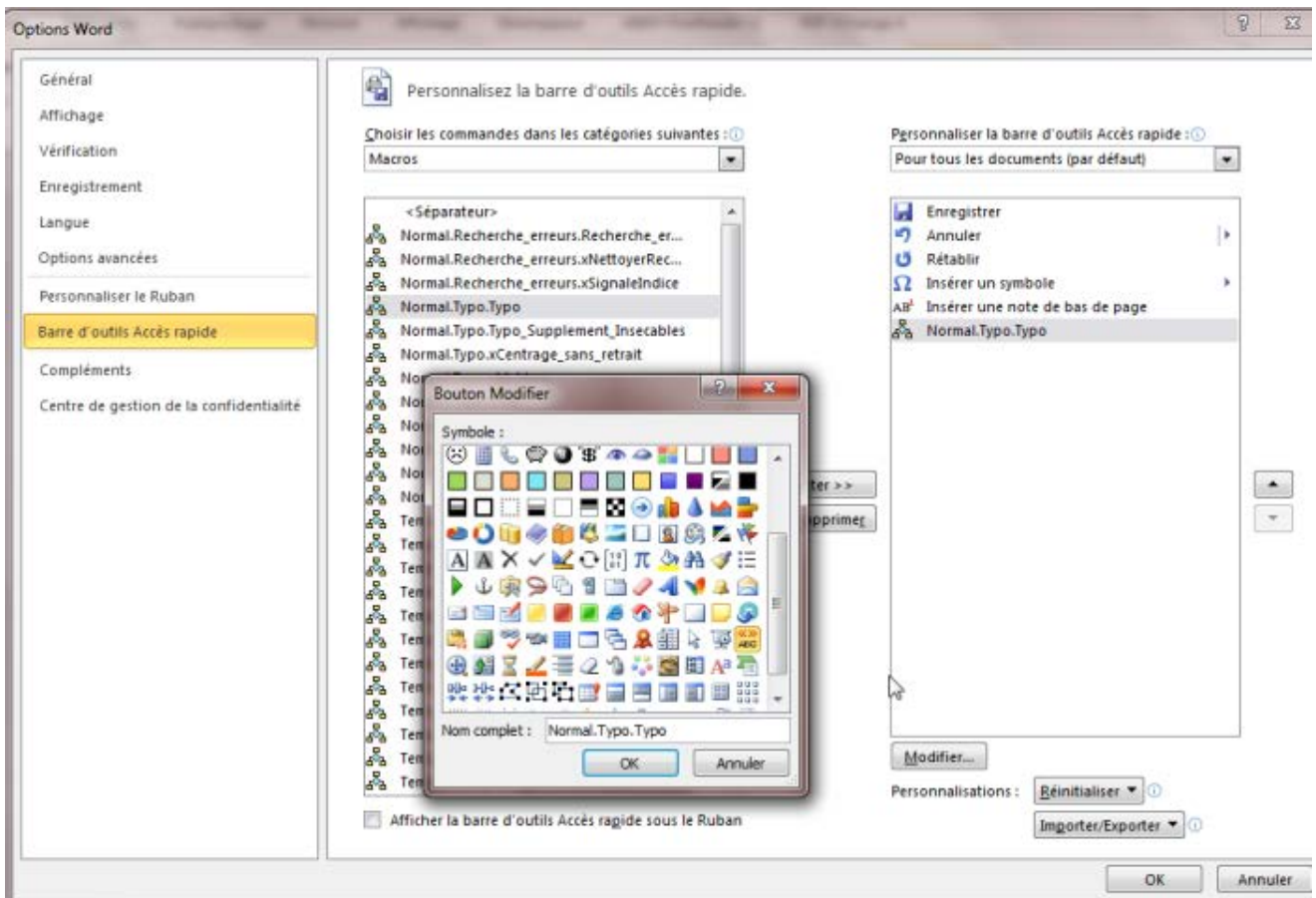
Pour accéder aux macros, 2 solutions :

a) Soit activer l'onglet Développeur dans les options en le cochant :

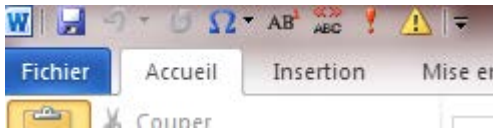


Ensuite onglet « Développeur » du ruban / Macros pour Word 2007-2010 ou menu Outils / Macros pour Word 2003, vous choisissez la macro Typo (typo.typo), et vous cliquez sur Exécuter.

b) Soit vous pouvez placer les macros dans la barre d'accès rapide (tout à fait en haut à gauche de la fenêtre)



Vous pouvez même les personnaliser en leur attribuant une petite icône (vous remarquerez que j'ai rajouté aussi dans cette barre l'insertion des caractères spéciaux et des notes de bas de page, bien pratiques lors de la relecture).



Il vous suffit alors de cliquer sur l'icône pour lancer la macro Typo.

Vous allez voir que votre texte va devenir beaucoup plus beau sur le plan typographique...

Et voilà, les travaux préliminaires sont terminés, et il faut maintenant s'attaquer au plus long : **comparer mot à mot, phrase à phrase, le doc et le PDF pour corriger les erreurs**. Là, pas de technique particulière, il suffit d'avoir de bons yeux. Bien évidemment, si vous avez fait le choix de faire cette correction mot à mot dans FineReader, alors vous ne la faites pas dans Word...

2^e niveau de travail

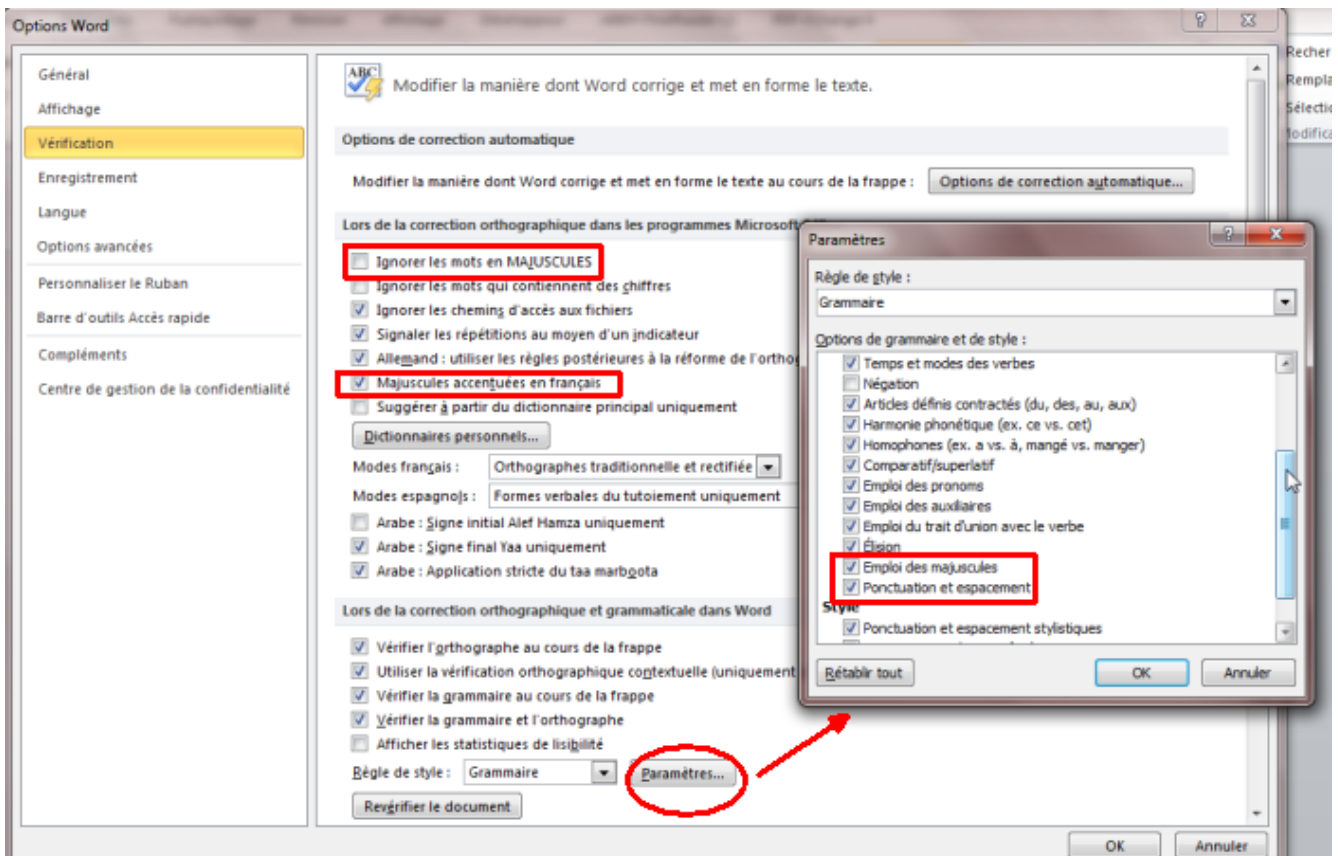
Nous avons laissé passer quelques jours, quelques mois (selon le temps de chacun), et le 1^{er} niveau de correction est terminé. Nous avons même trouvé quelqu'un qui accepte de relire le texte pour trouver les dernières fautes. Oui, mais avant, il nous reste deux types de tâches :

- Améliorer la correction par des moyens informatiques
- Finaliser la mise en forme

Améliorer la correction par des moyens informatiques

Lancer une **vérification orthographique et grammaticale** pour le document. Cette étape vous permettra de détecter les mots mal orthographiés mais aussi de mettre des majuscules accentuées et de détecter des sauts de paragraphes inopportuns.

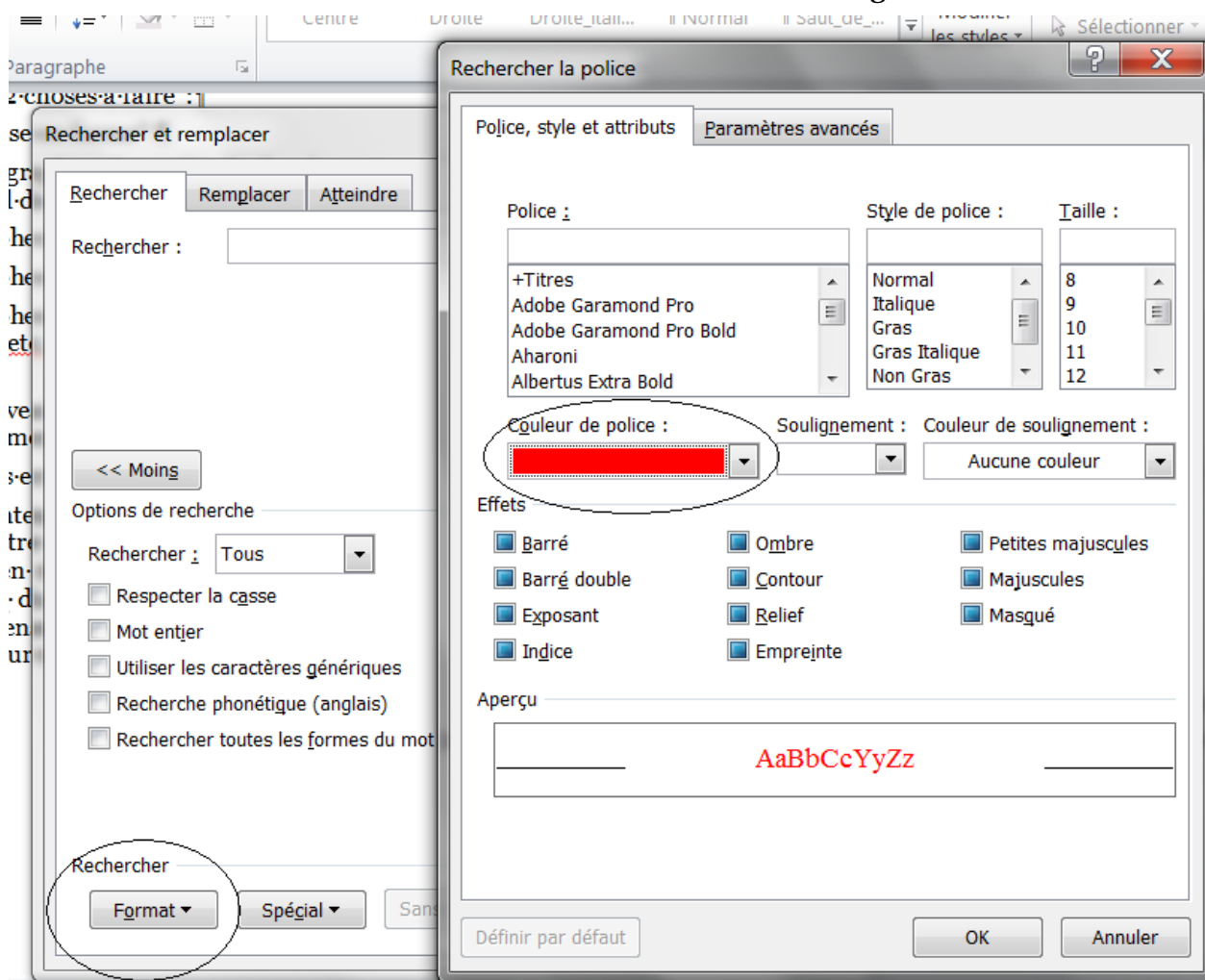
Il faut choisir les bonnes options. Tout se passe dans l'onglet Fichier / Options, j'ai entouré les points importants :



Attention, 80 % minimum des propositions de corrections proposées par Word ne sont pas pertinentes, il ne faut donc pas les accepter. Mais les 20 % restant valent le coup... Et pour savoir celles qui sont valables, c'est très simple, consultez à chaque fois le scan... Et n'oubliez pas une chose : il y a peu d'erreurs dans les éditions papier, généralement, il y a donc peu d'erreurs dans le scan...

Recherche des erreurs résiduelles que l'œil ne voit pas

Nous allons maintenant exécuter la macro Recherche_erreurs. Que fait-elle ? Elle se contente de mettre en rouge du texte susceptible de contenir une coquille. La majorité du texte en rouge ne contient pas de coquille, heureusement. Mais, en général, une partie du texte en rouge contient des coquilles qui ont échappé au correcteur. Il suffit ensuite de se mettre au début du texte, et de faire une recherche sur le texte de couleur rouge :

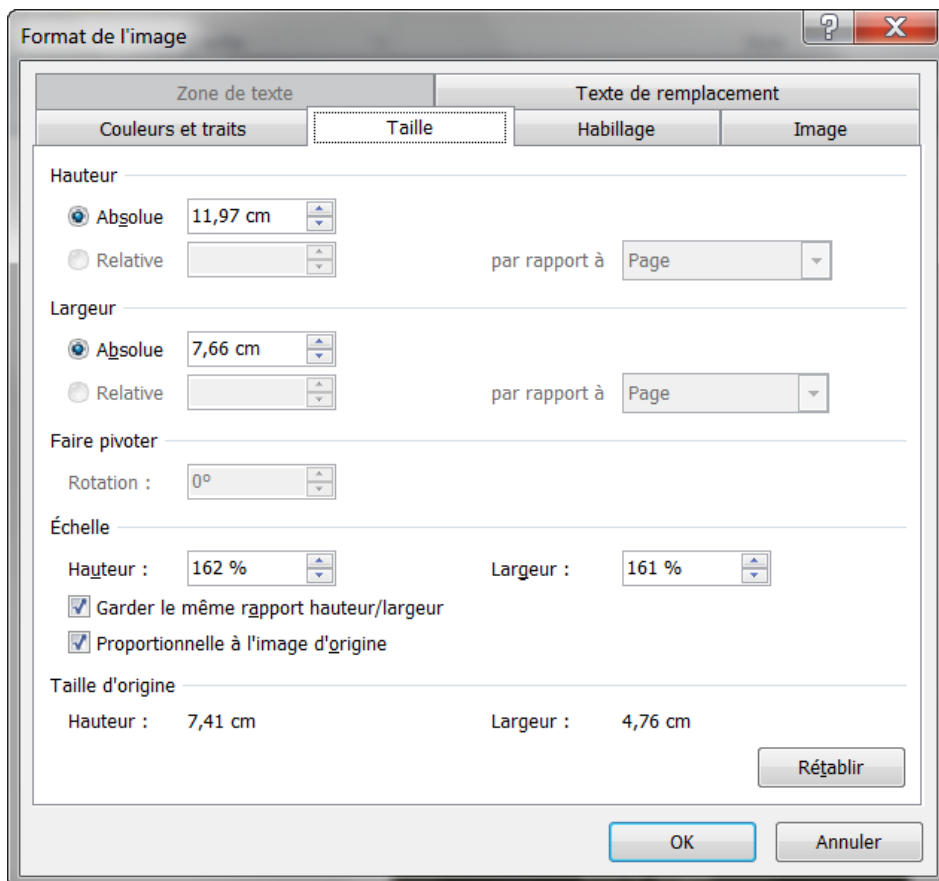


Il faut de 30 mn à 1h30 pour vérifier tout le rouge dans un livre, selon le nombre de pages et l'état du texte.

Évidemment, une fois que c'est terminé, vous enlevez tout le rouge... (Rechercher : police rouge / Remplacer tout : police couleur automatique)

Finaliser la mise en forme

* Mettre un paragraphe vide en tête du doc, le centrer en veillant à ce que tous les retraits soit égaux à 0 (autrement, ce n'est pas un vrai centrage...), et y coller une belle image, suffisamment grande et rectangulaire dans le sens de la hauteur (je vous conseille au moins 11 cm de haut), qui servira plus tard de couverture (onglet Insertion / Image). Vous pouvez vérifier / modifier la taille de l'image en faisant un clic droit sur l'image, Format de l'image / Onglet Taille :



- * Sur le paragraphe suivant, mettre le nom de l'auteur et appliquer le style Auteur.
- * Sur le paragraphe suivant, mettre le titre du livre et appliquer le style Titre_livre
- * Sur le paragraphe suivant, mettre des renseignements tels que date de parution, nom du traducteur, etc, éventuellement séparés par des sauts de ligne, et appliquer le style Date_Sortie
- * Exécutez à nouveau la macro Typo (il y a toujours des corrections faites qui ont modifié certains éléments de la typo)
- * Je vous conseille enfin de refaire un petit coup de Ctrl+A, et Ctrl+Maj+F9 pour éliminer les signets ou liens hypertexte qui se glissent subrepticement sous le texte et foutent ensuite le bordel pour la conversion au format epub.

3^e niveau de travail

Pour que la qualité finale soit bonne, il est fortement conseillé d'effectuer maintenant une relecture. L'idéal est de trouver un œil neuf, quelqu'un qui fera la relecture pour vous. Si ce n'est pas possible, vous la faites vous-même.

2 – Ce qu'il faut faire et ne pas faire avec Word (ou un autre traitement de texte)

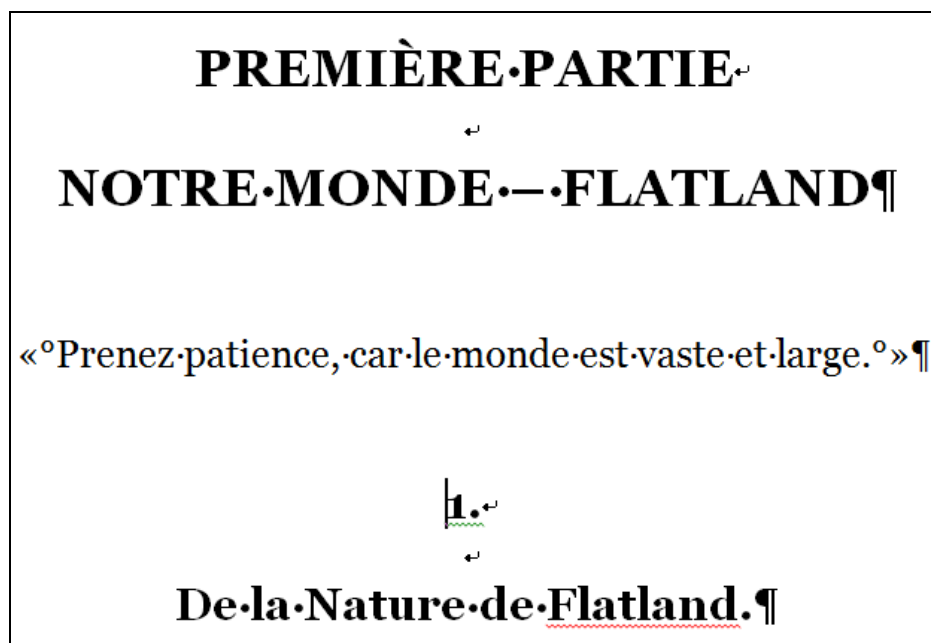
Avant d'aborder la question des conversions en différents formats, nous allons faire une petite liste de ce qu'il faut faire et ne pas faire.

Les styles Titre 1,2, 3...

Les chapitres, parties, doivent être impérativement structurés en leur appliquant les styles préexistants Titre 1, Titre 2, etc de façon à ce que les TDM (tables des matières) soient générées automatiquement dans les différents formats.

Personnellement, s'il n'y a que des chapitres, je leur applique le style Titre 1 ; s'il y a des parties et des chapitres dans le livre, j'applique le Titre 1 aux parties et le Titre 2 aux chapitres.

Si votre titre de partie ou de chapitre est sur 2 ou 3 lignes, comme dans l'exemple ci-dessous, vous ne devez surtout pas mettre un saut de paragraphe entre les lignes, il faut un **saut de ligne** seulement : touche **maj + entrée** (et non touche entrée).

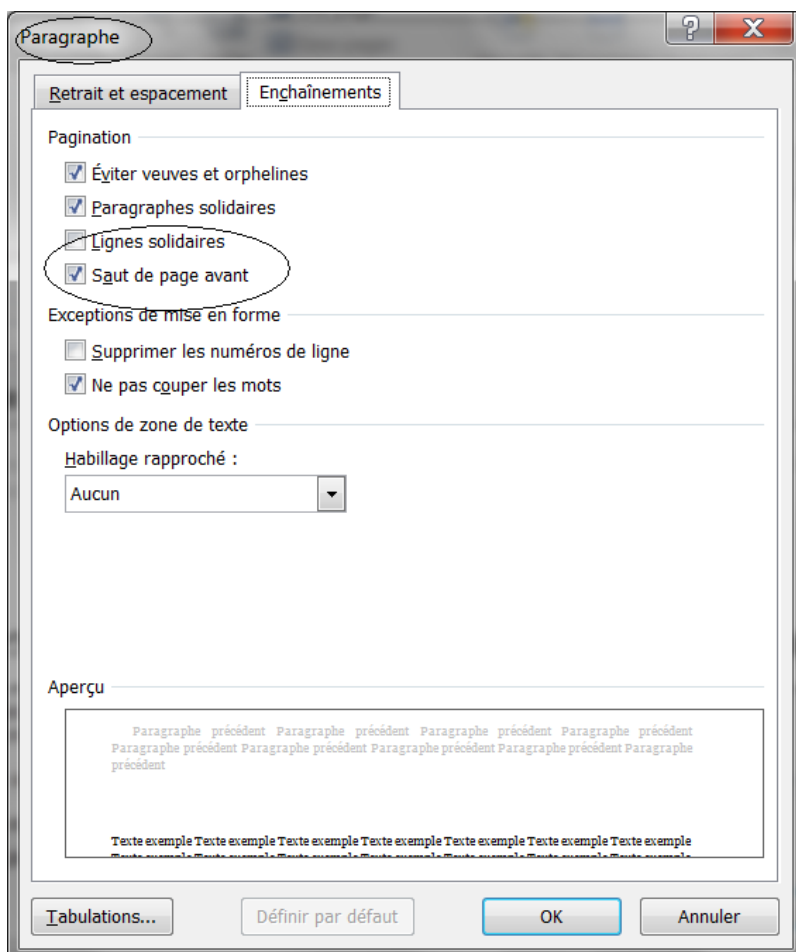
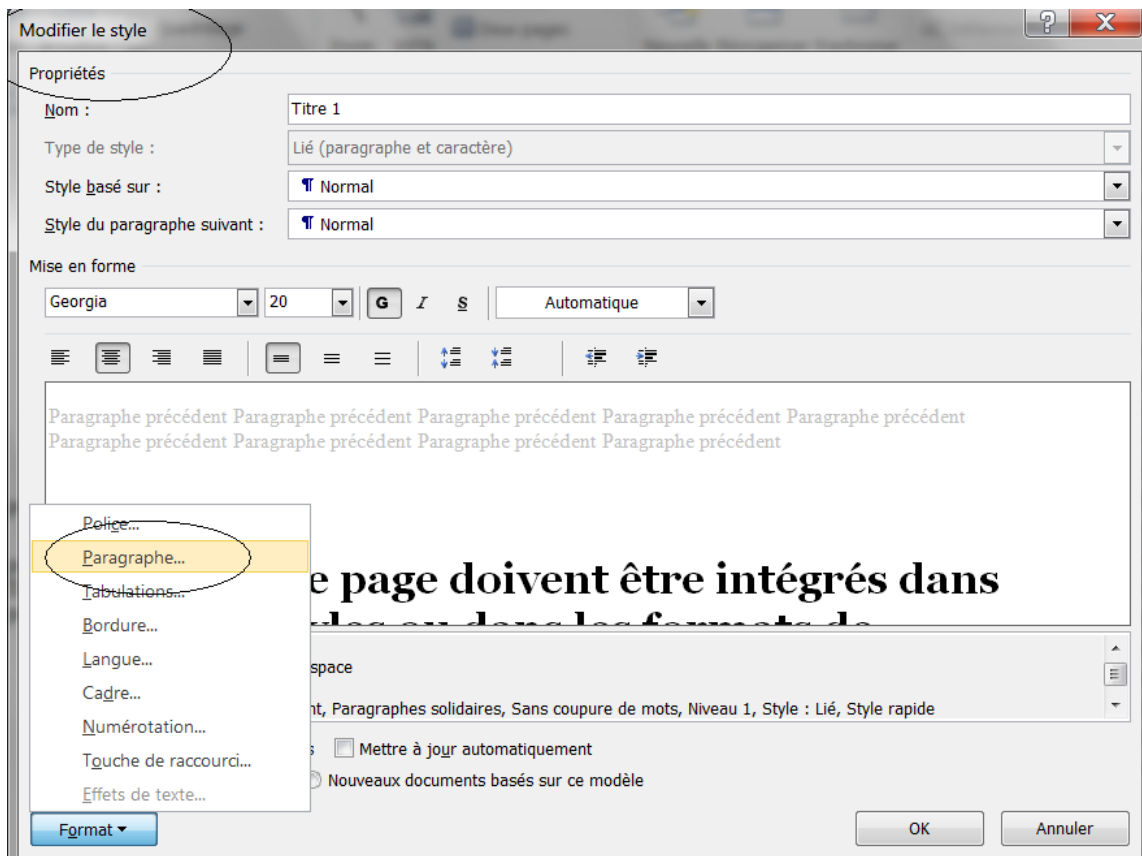


Il est inutile d'insérer une table des matières dans le doc Word de base, qui servira à la conversion vers les formats epub, Mobipocket : la table des matières n'apparaît pas dans le texte lui-même pour ces formats. Par contre, nous ferons ensuite un doc spécial pour le format PDF, et là, nous insérerons la table des matières.

À propos des sauts de page

Essayez de ne **jamais** insérer de sauts de page **manuellement**, c'est une source d'ennuis infinis pour corriger une mise en page.

Les sauts de page doivent être intégrés dans les styles ou dans les formats de paragraphes : deuxième onglet du format paragraphe ; pour les styles, en particulier Titre 1 qui comporte presque toujours un saut de page, vous modifiez le format paragraphe du style. Ainsi, les sauts de page deviennent visibles dans la fenêtre des styles et vous pouvez les modifier d'un clic.



À propos de sauts de paragraphe multiples

Au lieu de mettre 4 ou 5 paragraphes vides pour indiquer un grand espacement entre deux paragraphes de texte, il faut modifier l'espacement avant ou après du format de paragraphe (Format paragraphe / 1^{er} onglet) ou du style.

De plus veillez à être cohérent, à toujours mettre le même espacement avant ou après pour un type de paragraphe donné.

À propos des retraits et indentations

Format paragraphe / 1^{er} onglet

Comme vous le verrez, le modèle que je vous propose utilise une indentation (retrait de 1^{re} ligne) modeste de 0,6 cm. De même si j'applique des retraits à gauche pour décaler un paragraphe, ils sont en général de 0,6 ou 1,2 cm, au grand maximum de 1,8 cm, de façon à ce que le **cumul** avec l'indentation **ne soit jamais supérieur à 2,4 cm**. Il y a une raison précise à cela, qui concerne le rendu de l'ebook, par exemple au format epub, sur de petits écrans (nombreux sont ceux qui lisent sur un écran de téléphone de 3 – 4 pouces, le rendu n'est pas le même que sur un ordinateur).

Dans certains livres, le décalage horizontal des paragraphes est très important. Dans ce cas, j'utilise **les alignements centré ou droite** pour transposer la mise en page originale.

De même, je vois souvent des docs qui comportent des paragraphes avec des espaces avant ou après de 300, 400 pts. Tout cela pour imiter les livres papier... Ce type d'espacement est catastrophique pour les formats redimensionnables, epub ou mobi, pensez-toujours que votre ebook doit être lisible sur un iPhone (pour prendre un exemple). Dans cette optique, l'espacement maximum devrait être de 60 pts.

Autres erreurs à éviter

N'oublions pas que nous faisons un document en vue de le transformer en epub. Certaines pratiques courantes en traitement de texte ne passeront pas toujours bien dans ce format.

Il faut donc éviter d'utiliser les **tabulations**, **listes à puces** et les **petites majuscules**. De même ne pas laisser de paragraphe vide avant un saut de page.

De même, il ne faut généralement pas insérer de **signets** (sauf à savoir exactement pourquoi on le fait). Si vous en avez insérés par erreur, ce qui arrive, nous avons pu le constater, il suffit d'utiliser la commande « Insertion / Signets », et de les supprimer avec le bouton « Supprimer ».

Voilà, je pense avoir dit l'essentiel. Si vous voulez voir un **exemple de doc** prêt à être converti au format epub, et qui obéit aux règles ci-dessus énoncés : [richepin chanson des gueux epub.doc](#) (le doc est zippé). Le PDF image correspondant, pour voir la mise en page originale : [richepin chanson des gueux ocr.pdf](#) (7 Mo).

IV

CONVERSION AUX FORMATS EBOOKS

1 – Format ePub

Normalement votre doc est maintenant prêt à être converti au format epub. Il existe plusieurs méthodes pour cela, j'en citerai quatre :

* Méthode Amanuensis : [Amanuensis](#) permet très facilement, à partir d'un fichier .odt de créer un epub léger, de qualité et qui gère bien les images, les tableaux et les lettrines. Pour découvrir toutes ses possibilités le mieux est de lire le manuel très complet inclus dans l'archive (sous-dossier Manual). C'est la méthode que je considère la meilleure, que j'utilise au quotidien pour le site Ebooks libres et gratuits.

* Méthode Atlantis : La plus rapide et la plus simple pour les débutants en informatique, même si elle est moins bonne qu'Amanuensis, notamment pour les images. Son gros inconvénient est qu'elle ne gère pas du tout les tableaux.

* OpenOffice – Libre Office : Il existe plusieurs extensions qui permettent de convertir un doc/odt en epub, que vous pouvez utiliser, même si vous avez travaillé votre doc sous Word. L'extension [writer2xhtml](#) permet d'obtenir d'excellents résultats, très proches des 2 premières méthodes. Vous trouverez sur [le forum MobileRead](#) un sujet très complet avec un tutorial rédigé par Roger64, très fan de cette solution.

* Méthode Calibre : Elle a de nombreux adeptes, même si je considère personnellement que c'est la moins bonne des 4 méthodes. Je reprendrai ici un tutoriel fait par quelqu'un d'autre.

Méthode Amanuensis

Une fois terminé votre travail sous Word fermez votre fichier. Ouvrez-le avec [OpenOffice](#) ou plutôt [Libre Office](#) qui est plus régulièrement mis à jour, puis enregistrez-le au format texte ODF (.odt).

Faites ensuite glisser le document odt obtenu sur le programme Amanuensis ou son raccourci sur le bureau. Vous pouvez aussi faire un clic droit sur votre document puis choisir Envoyer vers... Amanuensis, si vous avez créé un raccourci vers Amanuensis dans le dossier Envoyer vers (SendTo) : c'est expliqué dans la manuel Amanuensis.

Je vous conseille de lire attentivement le manuel d'Amanuensis et notamment ce qui concerne le fichier d'options AMANUENSIS.cfg ; si vous voulez connaître les options que j'utilise, voici [mon fichier de configuration](#)

Méthode Atlantis

Voyons ce je dis sur ce logiciel sur notre site :

Une dernière solution, particulièrement adaptée pour les débutants, est le traitement de texte [Atlantis](#) ; rappelons que c'est un traitement de texte léger (on peut même l'utiliser sur clé USB), rapide, et très facile d'accès, avec lequel on peut ouvrir les documents aux formats RTF et Word (toutes versions). Une autre qualité est qu'il est capable d'enregistrer les documents dans un format RTF extrêmement compact. Et surtout, il possède une fonction d'enregistrement au format epub tout à fait performante. Atlantis est une bonne solution pour convertir un « doc » en un ebook au format epub, d'un clic, qu'il s'agisse d'un livre ou d'un document de travail. Il faut néanmoins signaler un gros défaut : il ne sait pas gérer les tableaux... Le prix de ce logiciel est d'environ 30 euros TTC, un coût raisonnable au regard de ses qualités (et vous pouvez le tester gratuitement pendant un mois). Voici le [lien](#) vers la partie de l'aide en ligne (très bien faite) d'Atlantis, consacrée à la conversion au format epub.

Nota : Atlantis est un logiciel Windows seulement. Mais les linuxiens seront heureux de savoir qu'Atlantis fonctionne parfaitement sous Linux grâce à Wine, et conserve sa rapidité.

Il n'y a donc en fait aucune explication à donner pour faire votre epub avec Atlantis, si vous avez suivi mes conseils et utilisé mon modèle :

Menu Fichier / Enregistrements spéciaux / Enregistrer comme ebook (Ctrl+maj+E). Donner un nom au fichier. Ensuite écrire les métadonnées : Titre, Auteur (sous la forme « Jules Verne »), etc. Et c'est tout...

En dehors du problème des tableaux, signalons qu'Atlantis ne sait pas gérer en pourcentage les images rectangulaires en largeur et qu'il est difficile d'éditer l'epub produit avec Sigil si on veut retoucher tel ou tel aspect de l'epub du fait d'une gestion atypique des styles de titre (Titre1, Titre2, etc) et de la table des matières.

Méthode Calibre

[Calibre](#) est un excellent logiciel, libre et gratuit, pour gérer sa bibliothèque d'ebooks et pour faire de nombreuses conversions, de et vers les formats ebooks.

La méthode ci-après est reprise à partir d'un forum :

- * Ouvrir le .doc
- * Supprimer en première page les informations de couverture (image de couverture, auteur, titre, etc) et les pieds de pages
- * Sauvegarder le doc au format HTML – **Page web filtrée** (important, ne pas choisir le HTML « normal »)
- * Importer le HTML dans Calibre (Nouveau livre...). Calibre va le transformer en zip...
- * Modifier les données Méta (Ajouter la couverture...) et faire la conversion conformément aux screenshot ci-dessous.
- * Dans l'onglet « Table des matières », indiquer :
Niveau 1 TDM : //h : h1
Niveau 2 TDM : //h : h2
- * Dans l'onglet « Détection de la structure », pour « marque de Chapitre », il faut choisir « none » au lieu de « pagebreak » afin de coller au plus près du .doc initial, et mettre à blanc le champ « Insérer un saut de page avant (expression XPath) ».
- * Dans « Préférence ->conversion -> présentation », cocher la case « conserver les ligatures » pour que les œ soient conservés.
- * Compatibilité Booken (Opus Orizon) pour éviter les polices bloquées : Dans « Option commune ->détection de la structure » cocher la case « Supprimer l'entête » et dans

- * Logiciels payants tels que Adobe Acrobat ou Nuance converter pro
- * Traitements de texte tels qu'OpenOffice et Libre Office qui possèdent une fonction d'export PDF tout à fait performante ou Word 2007 – 2010 ([lien pour le Add-in Microsoft Save as PDF Word 2007](#) – déjà intégré dans « Enregistrer sous » pour Word 2010)
- * Imprimantes virtuelles gratuites telles que PDF creator (par contre, au contraire des autres types de solution, les hyperliens de la table des matières et les signets correspondant aux Titre 1 – Titre 2... sont perdus).

Bref, les solutions ne manquent pas.

Vous pouvez tout à fait convertir votre doc tel que au format PDF. Mais, le format PDF étant « figé », non redimensionnable, il est préférable de créer des modèles Word correspondant à l'usage que vous voulez avoir pour votre PDF : A4 pour être lu sur ordinateur, A5 ou A6 pour être lu sur liseuse ou tablette.

Par exemple, pour un PDF A4, je vais créer un modèle avec un style Normal 16 pt, espacement avant 12 pt, espacement après 12 pt, les autres styles étant adaptés en conséquence.). Après, il ne me reste plus qu'à ouvrir un nouveau document basé sur ce modèle, et à copier-coller la totalité du contenu de mon document de base (Ctrl+A – Ctrl+C – Ctrl+V) dans le nouveau document. Je vais aussi générer dans Word une table des matières au début du document, après la couverture (voir l'aide pour cela, cela varie d'une version à l'autre, mais c'est simple à faire.)

Pour un PDF A5, je vais faire un modèle basé sur une taille de page A5 avec des marges assez faibles (maxi 1 cm). Mon style Normal aura une taille minimum de 18 pt, mais en gardant un espacement avant et après raisonnable, 5 ou 6 pt.

Cela dit, c'est à vous de voir, de tester, pour faire les PDF qui vous plaisent...

V

ANNEXES : POUR ALLER PLUS LOIN

1 – La typographie

On ne peut réaliser un ebook correct sans respecter les règles typographiques. L'utilisation des macros Typo pour la mise en forme et Recherche erreurs pour détecter les anomalies vous permettra déjà d'obtenir un document correct mais vous devrez lors de la lecture vérifier vous-même le respect de ces règles. Rappelons les principales :

- Utilisation des majuscules accentuées ;
- Tirets cadratins (—, Ctrl + Alt + – du pavé numérique) suivis d'une espace insécable (Ctrl + Maj + Espace) à l'ouverture des dialogues ;
- Tirets semi-cadratins (–, Ctrl + – du pavé numérique) pour les incises et les énumérations ;

(Nota : le groupe Ebooks libres et gratuits a fait le choix de n'utiliser que des tirets semi-cadratins, de « supprimer » l'usage des cadratins, considérant que ces derniers ne sont pas particulièrement adaptés et esthétiques pour l'usage ebook)

- Sigles sans points entre les lettres (SNCF) ou avec point mais sans espace (S.N.C.F.)
- Utilisation des espaces insécables (voir liens) ;

Ce sujet étant trop vaste pour être traité ici, je vous renvoie vers votre moteur de recherche favori pour trouver les nombreux sites internet sur ce sujet. En voici quelques-uns :

Typographie en général :

- [Topo de typo à l'usage des nuls](#)
- [Petites leçons de typographie](#) et [FAQ typo](#) de Jacques André.
- <http://fr.wikipedia.org/wiki/Typographie>
- <http://www.cuy.be/orthotypo/orthotypo0.htm>
- <http://membres.multimania.fr/clo7/expression/code.htm>
- http://btb.termiumplus.gc.ca/tpv2guides/guides/redac/index-fra.html?lang=fra&lettr=chap_catlog&page=../srchparbychap

Italique :

- [De l'emploi de l'italique](#)

Majuscules :

- [De l'emploi des majuscules](#)

Espaces insécables :

- http://www.druide.com/points_de_langue_13.html

Espacement :

- http://66.46.185.79/bdl/gabarit_bdl.asp?id=2039

2 – La mise en page (MEP)

« Faire la mise en page » signifie mettre en forme votre document de manière à ce qu'il respecte l'esprit de la mise en forme originale du livre voulue par l'auteur, tout en étant adaptée à l'aspect électronique de l'ebook, qui doit pouvoir être lu sur un écran 3 pouces, ou 23 pouces... Cela consistera donc à :

– marquer les parties, chapitres et sous-chapitres du livre à l'aide des styles de titre (Titre 1, Titre 2,...). Ceux-ci constitueront alors automatiquement la table des matières (TDM) de notre epub ;

– respecter les « aérations » du texte en rajoutant un paragraphe vide aux mêmes endroits que l'auteur (un, pas plusieurs...) ;

– identifier et mettre en forme les citations, chansons et poèmes ;

– insérer les notes de bas de page comme dans le livre (en utilisant l'insertion de notes de bas de page prévue dans Word ou LibreOffice) ;

– rajouter les illustrations et tableaux éventuels, si nécessaire ;

– ajouter une couverture en première page.

Pour cela il est rigoureusement proscrit de procéder au coup par coup. Il faut utiliser les styles. Vous trouverez la quasi-totalité des styles utiles dans mon modèle de document, libre à vous ensuite de les modifier pour les adapter à vos besoins.

3 – Utilisation des styles

Ce point peut paraître ardu, pourtant une fois maîtrisé on constate que c'est la manière la plus facile et la plus rapide pour mettre en forme un texte. De plus les modifications ultérieures seront ainsi grandement facilitées.

Voici quelques liens qui vous aideront dans ce domaine :

<http://heureuxoli.developpez.com/office/word/styles/#L4-D>

<http://www.tutoriel-video.com/styles-et-mise-en-forme-de-vos-documents-word-2007/>

<http://www.siteduzero.com/tutoriel-3-417014-utilisation-des-styles.html>

4 – Les macros

Qu'est-ce qu'une macro ? C'est tout simplement un petit programme contenu dans un document, comme le modèle que j'utilise, qui effectue une suite d'instructions.

La macro Typo met en forme le document afin qu'il respecte au mieux les règles typographiques (espacement, tirets, ponctuations, espaces insécables, guillemets, abréviations).

La macro Recherche erreur identifie en les mettant en caractères rouges des erreurs potentielles : fin de phrase anormale, tirets non conventionnels, mots potentiellement douteux à vérifier).

Elles ont pour but de vous faciliter le travail mais ne dispensent pas de la correction en comparant phrase à phrase avec le texte original et de la relecture. Pour les utiliser il faut les « lancer » (voir Améliorer la correction par des moyens informatiques). Elles vous indiquent que leur travail est terminé par l'apparition d'un petit message : « les corrections sont terminées ».

Ces macros sont régulièrement mises à jour afin de les améliorer, aussi je vous invite à toujours utiliser la dernière version.

5 – La relecture

Quand on parle de relecture, il s'agit évidemment d'une DEUXIÈME lecture du texte. La première aura bien évidemment déjà été faite par vous-même. Mais malgré toute l'attention que vous y porterez, il pourra subsister quelques erreurs qu'un deuxième œil pourra mieux détecter.

Vous devrez donc fournir à votre relecteur votre document finalisé ainsi que le pdf image/texte issu de Finereader. Il pourra ainsi faire une recherche sur un mot et visualiser en même temps le scan original. Son rôle sera de vous signaler toute coquille, défaut de mise en page, oubli par rapport à l'édition originale. Mais aussi faire part de ses doutes ou suggestions.

Comment rendre son travail ? Il existe plusieurs méthodes, chacune ayant ses adeptes. En tout état de cause, ce sera au numériseur de choisir la façon dont il veut récupérer les corrections :

— La modification directe du texte en mode révision : onglet Révision, cliquer sur Suivi des modifications. Les modifications apparaissent alors en rouge et on peut les faire défiler, les accepter ou refuser. Il est ainsi très rapide de consolider un document en vue de publication. **Ebooks libres et gratuits impose une limite aux corrections directes en mode révision : elles doivent toujours être exactement conformes au scan.**

— Les commentaires : dans ce cas le relecteur ne modifie rien mais signale ses observations dans un commentaire sur la phrase litigieuse. Les commentaires peuvent aussi être utilisés en mode révision. **Ebooks libres et gratuits utilise les commentaires, en complément des corrections en mode révision, dans 2 cas :**

- **Lorsque le relecteur estime qu'il y a une coquille dans l'édition papier, c'est-à-dire dans le scan : il ne doit pas faire la correction, dans ce cas, mais expliquer dans un commentaire quelle correction il propose.**
- **Et lorsqu'il a un commentaire quelconque à effectuer sur un élément quelconque de l'ebook.**

— Le fichier de correction : certains préfèrent ne pas altérer le document original et recevoir un fichier texte dans lequel sont répertoriés les passages litigieux avec leurs corrections. **Cette dernière méthode, à l'ancienne et peu productive, n'est jamais utilisée par Ebooks libres et gratuits.**

CONCLUSION

J'ai rédigé ce guide, presque « contraint et forcé » tant on me demandait de partager mon expérience, et d'une traite, prenant à peine le temps de le relire.

C'est dire s'il est perfectible...

C'est pour cette raison que parmi les formats disponibles pour ce guide, il y a le doc que je suis en train de finir d'écrire. Si vous avez des corrections, suggestions, etc, c'est simple : vous prenez le doc, vous activez le « suivi des modifications » du mode Révision, et vous corrigez, améliorez, etc. Je pourrai ainsi voir de manière les corrections / ajouts / modifications proposées.

Licence

Il n'y a pas de licence, ce texte est totalement et définitivement libre. Vous pouvez vous l'attribuer, le découper, le torturer, le vendre, l'imprimer et l'encadrer sur le mur de votre salon, bref, vous en faites ce que vous voulez, je m'en lave les mains...