

## Comment convertir un PDF numérisé en document Word ?

Lorsque vous scannez des documents, vous avez parfois besoin de modifier certaines informations, voire de reprendre la totalité du contenu pour le transformer. Grâce à un **logiciel d'OCR** (Optical Character Recognition) ou, en français, ROC (Reconnaissance optique de caractères), vous pouvez reprendre le document numérisé sur Word sans avoir à le ressaisir entièrement vous-même et c'est ce que nous allons découvrir dans ce tutoriel, notamment avec le **[logiciel PDFelement 6 de Wondershare](#)**.

Il y a quelque temps, j'ai publié un [tutoriel sur la création de formulaires interactifs et l'extraction des données collectées via Excel avec le logiciel PDFelement 6](#). La création de formulaires n'est pas son seul atout, il possède également un plug-in pour la reconnaissance de caractères.

Dans un premier temps, je vais vous montrer le logiciel gratuit que j'utilise parfois, puis, dans un second temps, pourquoi, si vous avez souvent besoin d'effectuer cette tâche, je vous recommande le logiciel de Wondershare.

### **Le logiciel d'OCR que j'utilise en version gratuite : FreeOCR**

Lorsque je découvre une nouvelle recette de cuisine dans un livre ou magazine, je la numérise en vue de la ressaisir dans mon application Access que je me suis créée. Pour des raisons de droits d'auteur, je vais plutôt utiliser un texte d'Alphonse Daudet extrait du site [Textes libres](#), mais le principe est le même.

Une fois le document numérisé, je l'ouvre avec le [logiciel FreeOCR](#) en cliquant sur **Open PDF**




et en allant pointer le document numérisé par l'imprimante.



Je clique sur **OCR** > *OCR Current Page* et voici le résultat :

LA DERNIERE CLASSE  
RÉCIT D'UN PETIT ALSACIEN  
Ce matin-là, j'étais très en retard pour aller à l'école. et j'avais grand-peur d'être grondé. d'autant que M. Hamel nous avait dit qu'il nous interrogerait sur les participes. et je n'eil savais pas le premier mot. Un moment l'idée me vint de manquer la classe et de prendre ma course à travers champs.  
Le temps était si chaud. si clair on entendait les merles siffler à la lisière du bois et dans le pré Rippert. derrière la scierie. les Pmssiens qui faisaient l'exercice. Tout cela me tentait bien plus que la règle des participes : mais j'eus la force de résister. et je courus bien vite vers l'école.  
En passant devant la mairie. je vis qu'il y avait du monde arrêté près du petit grillage aux affiches.  
Depuis deux ans. c'est cle là que nous sont venues toutes les mauvaises nouvelles. les batailles perdues. les réquisitions, les ordres de la commandanture : et je pensai sans m'arrêter :  
" Qu'est-ce qu'il y a encore 7 " Alors, comme je traversais la place en courant. le forgeron Wachter. qui était là avec son apprenti en train de lire l'afficelle. me cria :  
Ne te dépêche pas tant, petit 1 tu y arriveras toujours assez tôt à ton école !"

Même si cela me fait gagner un peu de temps, il y a quand même un peu de travail de saisie et de correction à effectuer (lorsque ce sont des recettes de cuisine, il y a beaucoup plus de travail à effectuer du fait qu'ici j'ai copié le texte du site pour le coller dans Word et en sortir le PDF, donc c'est plus net que lorsque c'est numérisé). J'ai souvent des "dc" à la place de "de", des "ct" à la place de "et" ou des barres verticales (|) au lieu de L, I ou T.

Ensuite, je clique sur le bouton **Export text into Microsoft Word**  pour qu'il ouvre un nouveau document Word avec ce même texte.

Si je conserve l'arrière-plan noir du site et que je mets un peu d'ondulation du texte avec Photoshop, FreeOCR n'arrive déjà plus à ouvrir le PDF : pour lui, il est vide.

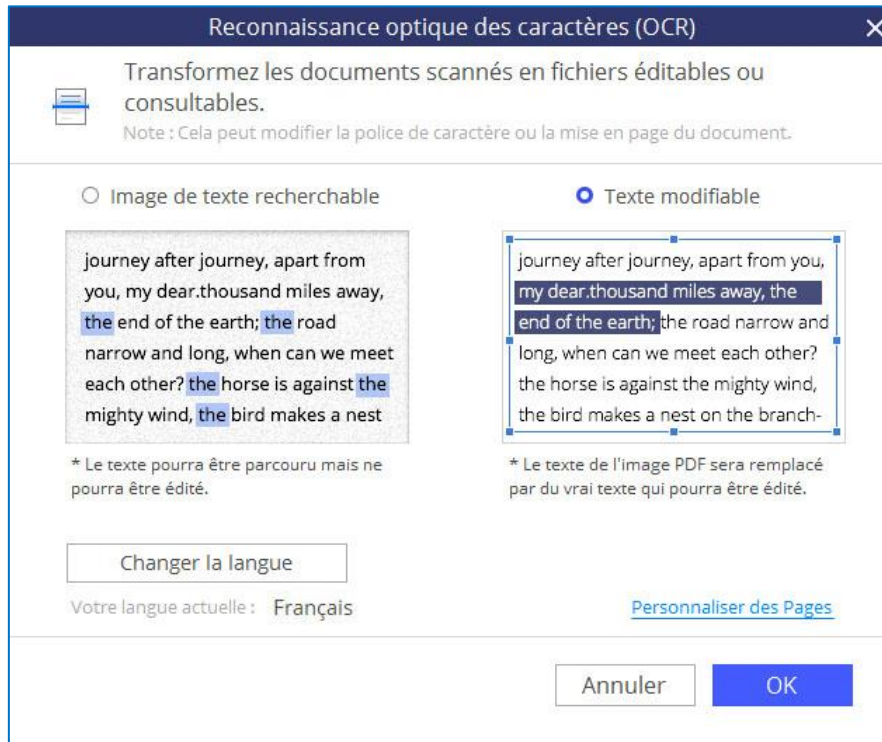
En résumé, il y a du travail et, si vous avez souvent besoin de modifier des documents scannés, la gratuité a ses limites 😊

### Le logiciel d'OCR PDFelement6 : payant, mais plus performant

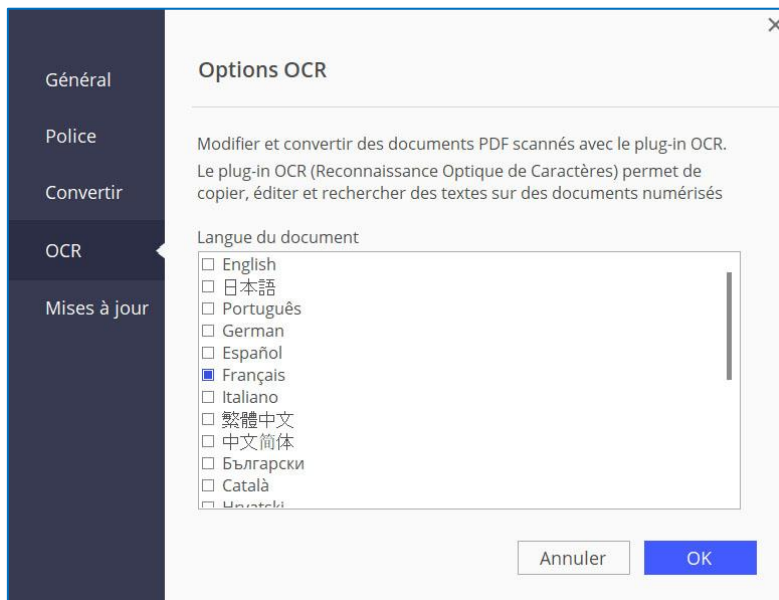
Ouvrez le logiciel PDFelement 6, cliquez sur **Ouvrir un fichier...** et pointez le document numérisé.



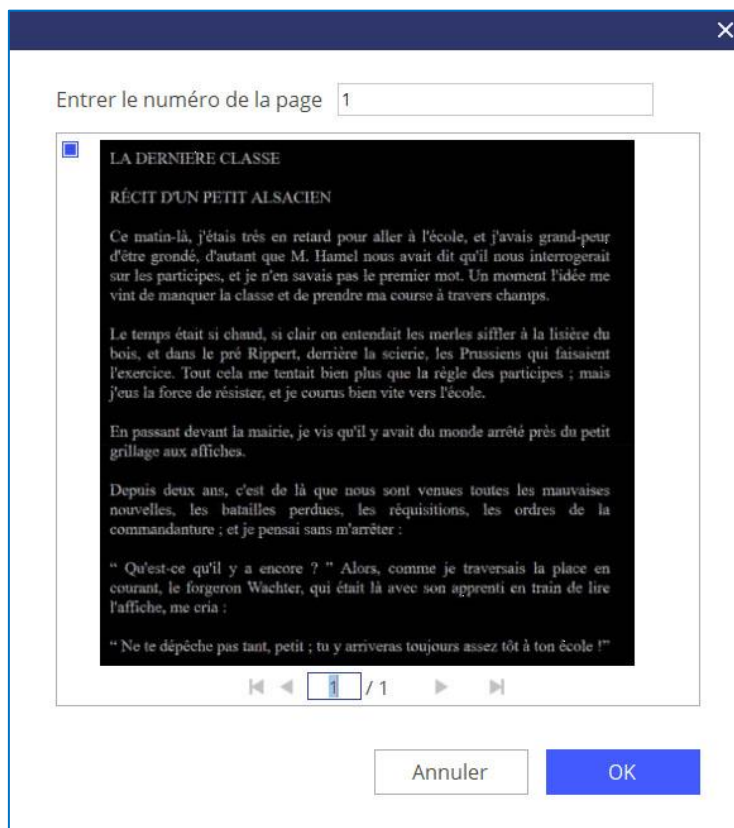
Puis, dans l'onglet **Modifier**, cliquez sur **OCR** :



Si la langue de votre document n'est pas le français, changez-la en cliquant sur **Changer la langue** afin d'avoir une meilleure reconnaissance des caractères :








Si vous ne souhaitez pas lancer la reconnaissance pour toutes les pages du document, cliquez sur **Personnaliser des Pages** (pour traiter les pages de 2 à 5 d'un document de 10 pages par exemple, saisissez 2-5) :



Deux choix de reconnaissance vous sont proposés :

1. Image de texte recherchable ;
2. Texte modifiable.

### OCR en mode Image de texte recherchable :

Cette option vous permet de reconnaître le texte pour pouvoir le parcourir et faire une recherche par exemple, mais aussi sélectionner tout ou partie d'une zone de texte pour la surligner , la souligner avec un trait  ou une vague , la barrer  ou lui ajouter un signe d'insertion . Pour cela, sélectionnez un mot ou groupe de mots et cliquez sur l'icône correspondante et, en re cliquant sur le mot ou groupe de mots, le panneau latéral **Propriétés** apparaît vous offrant la possibilité de changer les informations ou l'apparence données. Vous avez également la possibilité d'*Ajouter une note sur le texte*, d'*Ajouter un signet* ou de *Créer un lien* par un clic droit sur le document.

### OCR en mode Texte modifiable :

La seconde option vous permet de reconnaître le texte et de le rendre totalement modifiable.

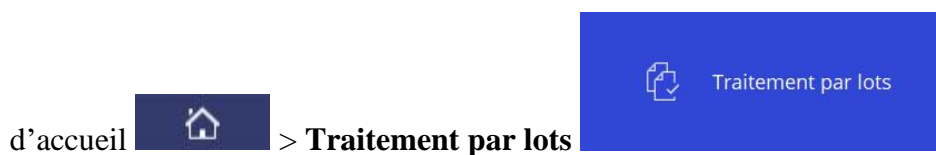
Grâce au panneau latéral des **Propriétés**, vous pouvez réaliser différentes modifications :



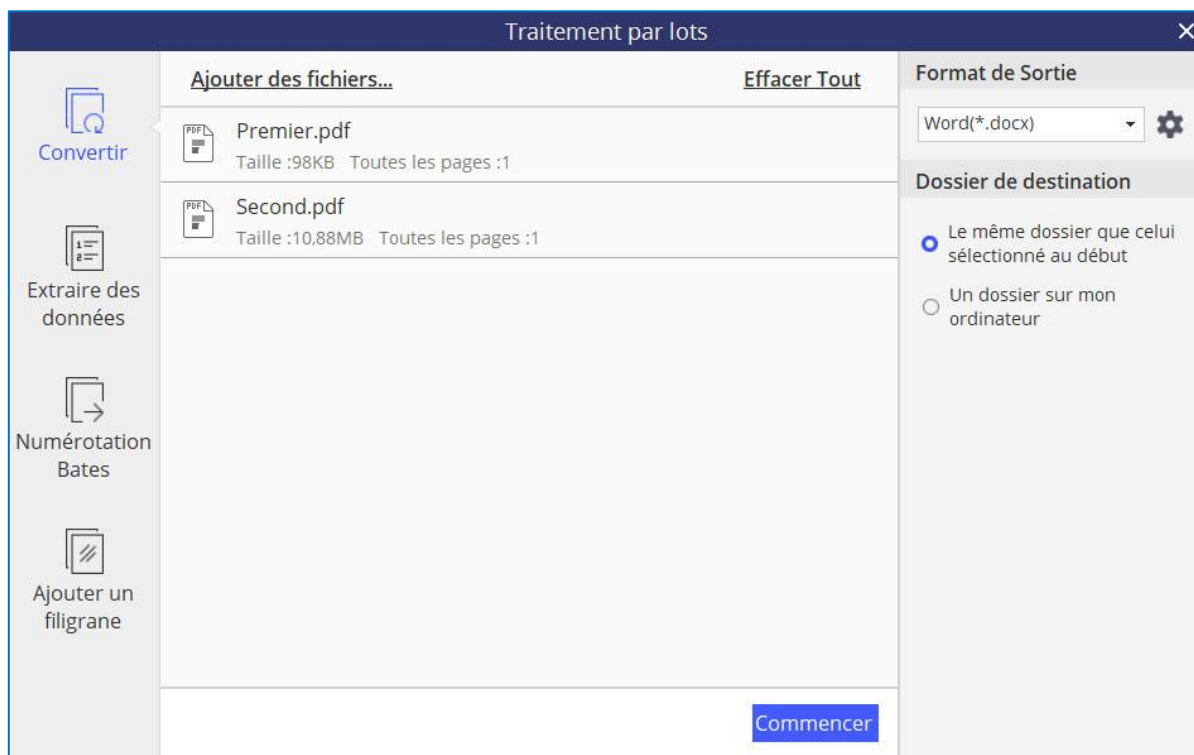
Via ce choix de reconnaissance, vous pouvez passer en *Mode ligne* pour modifier le texte ligne par ligne et déplacer certaines lignes ou en *Mode paragraphe* pour le modifier par paragraphe (bloc de texte) et déplacer certains paragraphes en cliquant simplement dessus.


## OCR en traitement par lots :

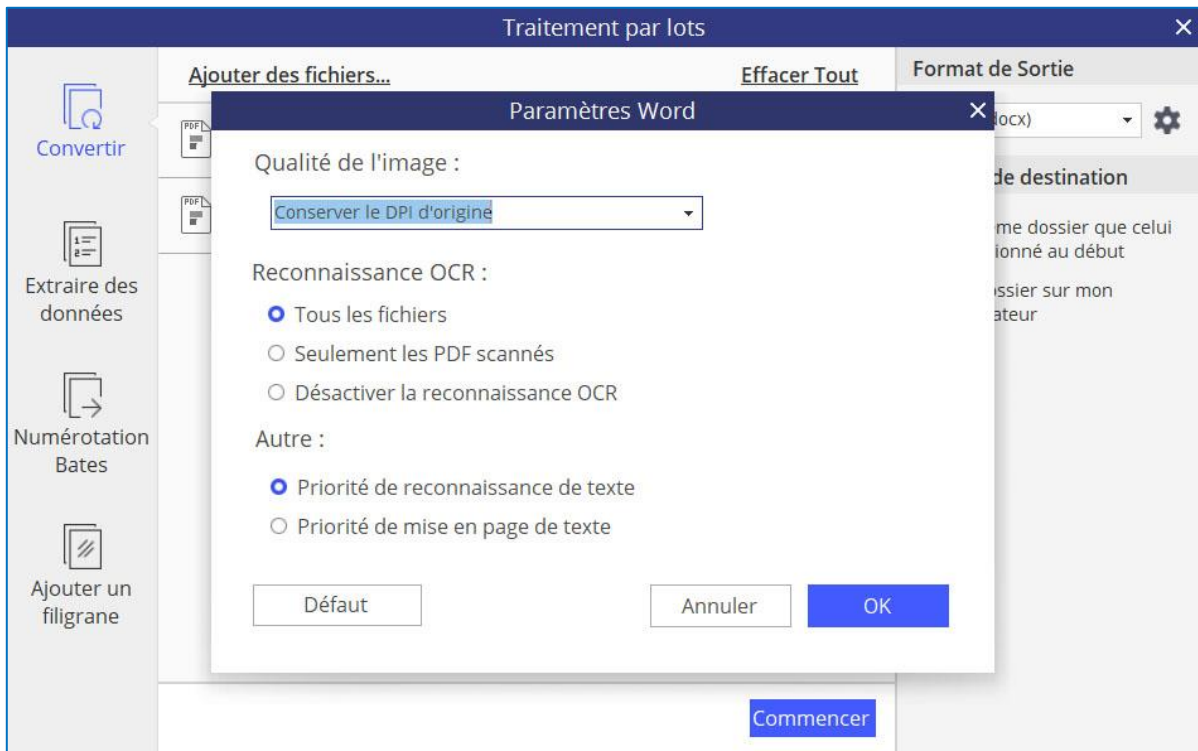
Vous pouvez également effectuer un traitement par lots pour lancer la reconnaissance de caractères sur plusieurs documents en retournant sur le menu



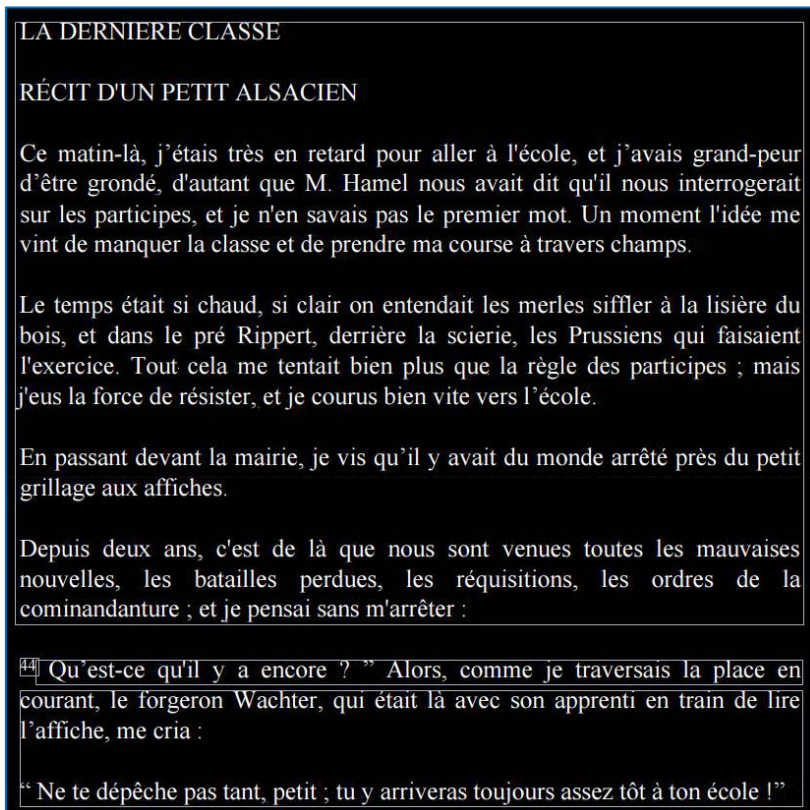
Dans la fenêtre, cliquez sur le + pour ajouter vos documents à traiter et modifier le **Dossier de destination** si *Le même dossier que celui sélectionné au début* ne vous convient pas :



Puis, cliquez sur la roue crantée à droite  et, dans la fenêtre qui s'affiche, choisissez la **Reconnaissance OCR** pour *Tous les fichiers* et la *Priorité de reconnaissance de texte* pour mieux reconnaître le texte avant de cliquer sur **Ok**, puis **Commencer** :



Une fois le traitement effectué, cliquez sur **Terminé** et PDFelement vous aura généré autant de documents Word que de documents numérisés ajoutés :



### OCR vers Word :

Grâce à la fonctionnalité [OCR vers Word](#), vous obtiendrez le même type de résultat, mais directement dans Word. Pour cela ouvrez votre document numérisé avec PDFelement et, dans



l'onglet **Accueil**, cliquez sur **Vers Word** [Vers Word](#). Choisissez l'emplacement du document Word qui sera généré et, une fois la conversion effectuée, cliquez sur **Terminer** : le PDF est converti.

Quel que soit le choix de reconnaissance choisi, le résultat est meilleur avec [PDFelement 6](#) d'après mes propres tests.



### Quelques conseils pour une meilleure reconnaissance des caractères :

Afin d'obtenir le meilleur résultat possible avec un logiciel d'OCR, il est recommandé que vos documents scannés soient :

- Le plus droits et à plat possible, c'est-à-dire que si vous numérisez un document de travers ou un livre ouvert qui donne un côté bombé, le résultat risque de ne pas être optimal ;
- De préférence avec une écriture noire sur blanc ou au moins de couleur unie ;
- Rédigés avec une police classique et en **Style de police Normal** (une police exotique ou en italique peut ne pas être bien reconnue).

*Wondershare propose une version d'essai de 30 jours pour tester leur logiciel, ce qui vous laissera largement le temps de tester la performance de cet outil pour vos documents numérisés. Passé ce délai, vous pourrez toujours l'utiliser, mais avec des [options limitées](#).*

*Article sponsorisé : la marque [Wondershare](#) m'a permis de tester le [logiciel PDFelement](#) afin d'écrire ce tutoriel en échange d'une compensation financière.*

*Tutoriel réalisé avec Word 2016 et PDFelement 6 Professional*

[Voir la version vidéo de cet article](#)