

Quick start Typex

Qu'est-ce que Typex

Typex est un outil de correction de la typographie française, il corrige également quelques erreurs courantes d'OCR ou de mise en forme. C'est **une extension pour Calibre qui sera utilisée dans l'éditeur de Calibre**. Elle n'est pas compatible avec Sigil.

Dans l'éditeur, Typex corrige les fichiers au format epub. Il ne se substitue pas aux diverses vérifications qui ont pu être faites au long du processus de fabrication de l'epub (par exemple l'extension checkbook pour libreoffice), et dont vous n'avez pas forcément connaissance. Il n'est pas non plus rendu inutile par celles-ci. Il est le fruit de l'expérience concernant les erreurs qu'il est possible de rencontrer dans un epub, quelle que soit son origine.

C'est une interface pour un ensemble de règles (ou regex, expressions rationnelles) groupées par thème. Chaque règle est facultative, et peut être appliquée automatiquement en totalité (mode automatique) ou occurrence par occurrence (mode pas à pas).

Installer Typex

Si vous avez déjà installé une version précédente de Typex, il est prudent de la désinstaller au préalable.

Ne pas dézipper le fichier téléchargé. Dans Calibre, aller dans :

Préférences / Avancé / Extensions

Cliquer en bas à droite sur [*Charger une extension à partir d'un fichier*], retrouver l'extension dans les dossiers sur l'ordinateur, puis valider et relancer Calibre.

 **Typex s'installe dans Calibre et est mis en œuvre dans l'éditeur de calibre.**

Dans l'éditeur de Calibre, Typex apparaît dans le menu *Extensions*, et une icône figure dans la barre d'outils :



Si la barre d'outils des extensions n'apparaît pas, cliquer dans l'éditeur sur :

Aperçu / Barres d'outils / Barre d'outil des extensions.

On peut aussi ajouter une icône dans la barre d'outil principale ou ajouter une entrée pour Typex dans la barre de menu principale, pour cela aller dans les préférences de calibre, *Interface / Barres d'outils et menus*.

Supprimer Typex

Dans Calibre, aller dans :

Préférences / Avancé / Extensions / Extensions Outil d'édition de livres

Sélectionner Typex et cliquer en bas à droite sur [*Supprimer une extension*], puis valider.

Prérequis

Pour que toutes les règles de Typex fonctionnent correctement, il faut qu'**un dictionnaire**, par exemple celui de Grammalecte, soit installé dans l'éditeur de Calibre. En effet, quelques règles utilisent une fonction qui fait appel au dictionnaire. Si celui-ci n'est pas installé, Typex le signalera et refusera de démarrer.

Pour installer un dictionnaire, aller sur la page <https://grammalecte.net/#download>, cliquer sur « Anciennes versions » et charger « dictionnaire V x.x (sans grammalecte) ». Aller ensuite dans les préférences de l'éditeur de Calibre / propriétés de l'éditeur / organiser les dictionnaires, bouton Ajouter un dictionnaire.

Exécuter Typex



Pour tester Typex, il est fortement recommandé de lancer l'éditeur de Calibre sur une copie de l'epub.

Phase préalable

Typex exécute d'abord une phase préalable, qui est une phase nécessaire de mise en forme de l'epub afin que cette forme soit la plus lisible et prévisible possible. Elle délivre également quelques informations issues d'une analyse de l'epub. Durant cette phase, deux commandes de l'éditeur de Calibre sont mises à contribution :

- **Embellissement** des fichiers (après suppression des retours ligne).
- **Suppression des styles inutilisés**, sans supprimer les attributs de classe inutilisés ni fusionner les règles css identiques (les cases à cocher de la commande).

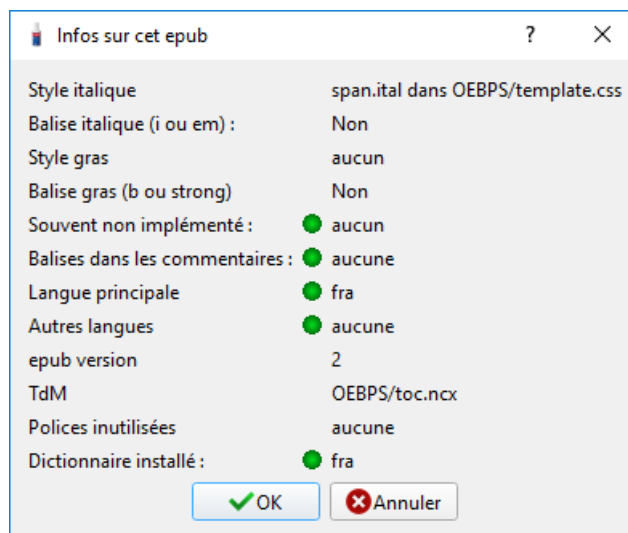
La mise à contribution de ces deux commandes est un préalable indispensable au bon déroulement, ensuite, des règles, notamment le remplacement des entités (voir ci-dessous).



En html, les caractères spéciaux peuvent être représentés par des entités, c'est-à-dire une séquence de 4 à 8 caractères commençant par « & » et se terminant par un point-virgule. L'embellissement remplace les entités par le caractère unicode correspondant sauf pour les 3 caractères <, > et &.

L'embellissement répare, aussi, automatiquement certaines erreurs de syntaxes.

L'extension effectue ensuite quelques vérifications et affiche l'écran suivant :



Ces caractéristiques de l'epub sont données à titre d'information.

La pastille verte, qui donne le feu vert, peut se transformer en pastille rouge, qui indique un point d'attention. Si le dictionnaire n'est pas installé, la pastille correspondante devient rouge et Typex refuse de continuer.

Style italique

Cette ligne n'est pas obligatoirement renseignée. Elle affiche le résultat d'une recherche sur l'ensemble des fichiers de styles (fichiers css) de styles de caractères ayant la propriété « *font-style: italic;* » sans être en gras, ou faisant appel à la police italique en cas de fichiers de police embarqués. Les styles de paragraphes mêlant attributs de paragraphes (alignement, retraits...) et attributs de caractères sont ignorés par cette recherche.

Balise italique (i ou em)

Ce sont les deux balises généralement utilisées pour marquer l'italique. Si un style italique a été trouvé et pas de balises *i* ou *em*, on peut en déduire que l'italique est marqué par une balise *span*. En effet, Typex a supprimé les styles inutilisés avant de faire ses recherches, les styles italiques trouvés sont forcément utilisés par une balise.

Style gras

Même chose que pour l'italique.

Balise gras (**b** ou **strong**)

Même chose que pour l'italique.

Souvent non implémenté (c'est-à-dire : souvent incompris par les liseuses)

Cette ligne affiche le résultat d'une recherche sur l'ensemble des fichiers de styles (fichiers css) de styles de caractères ayant la propriété « font-variant: small-caps; » ou « text-transform ». *small-caps* transforme en petites capitales des chaînes en minuscules, et *text-transform* met le texte en minuscules, en majuscules, ou met une majuscule à la 1re lettre de chaque mot suivant que lui est affectée la valeur uppercase, lowercase ou capitalize.

Balises dans les commentaires

Cette ligne affiche le résultat d'une recherche de balises à l'intérieur d'éventuelles lignes de commentaire dans les fichiers xhtml. Cette présence de balises dans une ligne de commentaire peut empêcher certaines regex de bien fonctionner.

Langue principale

Typex cherche la balise <dc:language> dans le fichier OPF et affiche son contenu. Cette balise est obligatoire, la ligne doit afficher quelque chose. Le résultat « fr » (ou « FR ») est important car, sinon, la vérification d'orthographe et certaines regex-fonctions de Typex qui utilisent le dictionnaire sont compromises.

Autres langues

Cette ligne est le résultat de :

- La recherche dans les en-têtes des fichiers xhtml d'un attribut xml:lang ayant une valeur différente de « fr » (ou « FR »). Si le fichier de couverture du livre a été créé par Calibre, il comporte un attribut xml:lang ayant la valeur « en ». Dans ce seul cas, cette valeur est ignorée par la recherche parce qu'elle n'est ni signifiante ni gênante.
- La recherche dans les fichiers xhtml de balises ayant une propriété lang différente de « fr » (ou « FR »). Il s'agit ici de passages du livre déclarés être dans une langue étrangère. Cette déclaration est facultative. Elle est cependant utile, car prise en compte par la vérification d'orthographe.

Epub version

2 ou 3, selon la version donnée par la balise package du fichier opf.

TdM (Table des matières)

Cet élément est obligatoire, aussi bien en Epub2 qu'en Epub3. En Epub V2, il s'agit d'un fichier de format spécifique ayant l'extension ncx. En Epub V3, c'est un fichier dit « de navigation » au format HTML5. Un ebook au format Epub3 peut comprendre à la fois un fichier de navigation et un fichier ncx pour une lecture sur des dispositifs de lecture exclusivement V2.

Polices inutilisées

Lorsque des fichiers de polices de caractères sont embarqués dans l'epub, il arrive que certains ne soient pas utilisés, notamment la police fournissant les glyphes à la fois italique et gras.



Pour supprimer une police dans l'éditeur de Calibre, il faudrait utiliser la commande de l'éditeur *Outils / Gérer les polices*, jamais en la supprimant sauvagement dans un gestionnaire d'archive comme *Winrar*. Supprimer un fichier de police grâce à cette commande implique que vous ne pourrez plus faire une modification de l'epub qui utilise ce fichier. Si, par exemple, vous conservez le fichier des glyphes droites et que vous supprimez le fichier des glyphes italiques, si vous appliquez un style italique absent précédemment de l'epub, le moteur d'affichage ne pourra pas utiliser le fichier supprimé et procèdera par déformation des glyphes droites. Pour cette raison, il est préférable de ne supprimer un fichier de police qu'à la toute fin de votre révision des styles, sauf si vous avez sauvegardé le fichier de police sur disque.

Si nécessaire, la phase préalable effectue également les opérations suivantes :

- Suppression des directives *ltr*, inutilement inscrites par certaines versions précédentes de *Writer2xhtml*.

- Remplacement des insécables fines par des insécables, afin de simplifier considérablement les règles (elles pourront être rétablies en phase finale de typex, si désiré).
- Remplacement des pseudo-fines Amanuensis par des insécables.
- Suppression de caractères invisibles (traits d'union conditionnels & espaces sans chasse).
- Remplacement d'espacements rares (espace cadratin etc.) par des espaces sécables.
- (Projet) Changement des balise *span* d'Amanuensis pour marquer l'italique, le gras ou la mise en exposant respectivement en balises *i*, *b* ou *sup*.
- (Projet) Création d'un style exposant en remplacement de celui d'Amanuensis.
- Suppression de *line-height: normal*; dans les fichiers css.

Suppression de *line-height: normal*;



La valeur *normal* de la propriété *line-height* est par défaut. L'indiquer dans un style donne un mauvais résultat sur certaines liseuses, et empêche d'utiliser et gérer la hauteur de ligne par défaut sur ces liseuses.



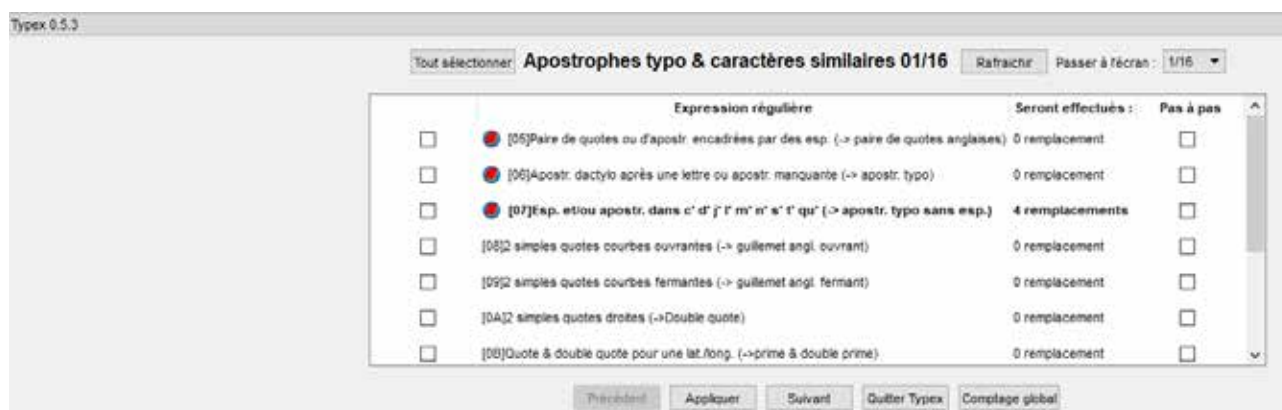
En raison de ces possibles modifications, si vous quittez Typex sans avoir exécuté une règle, l'extension peut quand même vous proposer de valider des changements.



Typex ne modifie jamais le fichier epub sur le disque dur. La validation des changements modifie le fichier temporaire affiché dans l'éditeur de Calibre. Pour enregistrer les changements sur le disque, il faudra utiliser la commande de l'éditeur *Fichier / Enregistrer*, ou cliquer sur « Sauvegarder et quitter » en quittant l'éditeur.

Parcours des règles, section par section

Après validation de la phase préalable, Typex parcourt une centaine de règles en 18 sections (à ce jour) :





Pour chaque règle, Typex propose deux cases à cocher. La case de gauche déclenche une exécution automatique, la case de droite une exécution pas à pas. Dans le premier cas, Typex remplace automatiquement l'élément sélectionné. Dans le dernier cas, Typex parcourt les fichiers d'élément sélectionné en élément sélectionné, en permettant à l'utilisateur de juger de l'opportunité du remplacement. Sélectionner « pas à pas » coche automatiquement la case de gauche.

Après avoir appliqué la sélection, Typex revient à la même section. Il n'est donc nullement nécessaire de cocher d'un coup toutes les règles dites positives, c'est-à-dire dont le nombre de remplacements est supérieur à 0. Celles-ci sont affichées en gras afin de mieux les visualiser.

Lorsque qu'une icône est affichée à gauche du nom de la regex cela signifie qu'un commentaire est disponible pour cette regex. Celui-ci explicite le fonctionnement de cette regex et peut être visualisé en survolant le titre de la regex avec la souris. Par défaut, cette icône représente un livre. Lorsqu'elle représente une étoile, cela signifie que la règle possède un prérequis parmi les règles précédentes. Le commentaire précise le n° des règles en prérequis, c'est-à-dire qu'il faut avoir activées au préalable (sauf, bien sûr, si leur comptage donne un résultat nul).

On peut cocher à la fois des règles en mode automatique et des règles en mode pas à pas. Typex exécute d'abord les règles automatiques puis les règles pas à pas. Lorsque plusieurs règles sont cochées en mode pas à pas, elles sont exécutées dans l'ordre, l'une après l'autre. On parcourt le texte de sélection en sélection pour chaque règle cochée, et non toutes règles cochées confondues.

 Appliquer une règle présente un risque de faux positifs, c'est-à-dire de chaînes sélectionnées alors qu'elles ne le devraient pas. En caricaturant, si un auteur décide d'écrire un livre sans la ponctuation « point », la règle qui regarde les ponctuations en fin de paragraphe va présenter des faux positifs. Le risque est quasi-nul pour certaines règles. Pour d'autres règles, il est utile de sélectionner des chaînes qui sont des cas classiques d'erreurs, sans qu'il soit possible de les distinguer par une expression régulière de cas corrects (par exemple, pardessus et par-dessus). Lorsque le risque de faux positifs est important, le « pas à pas » est forcé (obligatoirement coché).

 Inversement, deux règles sont en mode automatique obligatoire : E4 et E5. Prenons l'exemple de la règle E4 : Séparer deux mots collés.

La regex va de mot en mot dans l'epub, sélectionne le mot et le passe à la fonction qui regarde si le mot est validé par le dictionnaire. Si le mot est dans le dictionnaire, la fonction remplace le mot par lui-même et on passe au mot suivant. Sinon, la fonction coupe le mot en deux et vérifie si les deux nouveaux mots y figurent. La fonction déplace progressivement la frontière entre les deux mots décollés en re-testant à chaque fois leur présence dans le dictionnaire.

Avec ce fonctionnement, aller en pas à pas consisterait à aller de mot en mot sur tout l'epub ! Pour éventuellement n'avoir au bout du compte aucun mot décollé. La regex ne peut pas vérifier la présence du mot dans le dictionnaire, seule la fonction le peut. C'est la raison pour laquelle il faut exécuter cette règle en mode automatique, puis [Voir ce qui a changé], et valider ou annuler le résultat.

L'écran 04 présente une variante : une règle est une variante de celle qui la précède. Voir le *Guide* pour les détails.

L'écran 07 présente un choix. On peut choisir d'exécuter les deux règles du choix 1, ou bien les deux règles du choix 2 (ou aucune, ou une seule), mais pas une règle du choix 1 et une règle du choix 2.

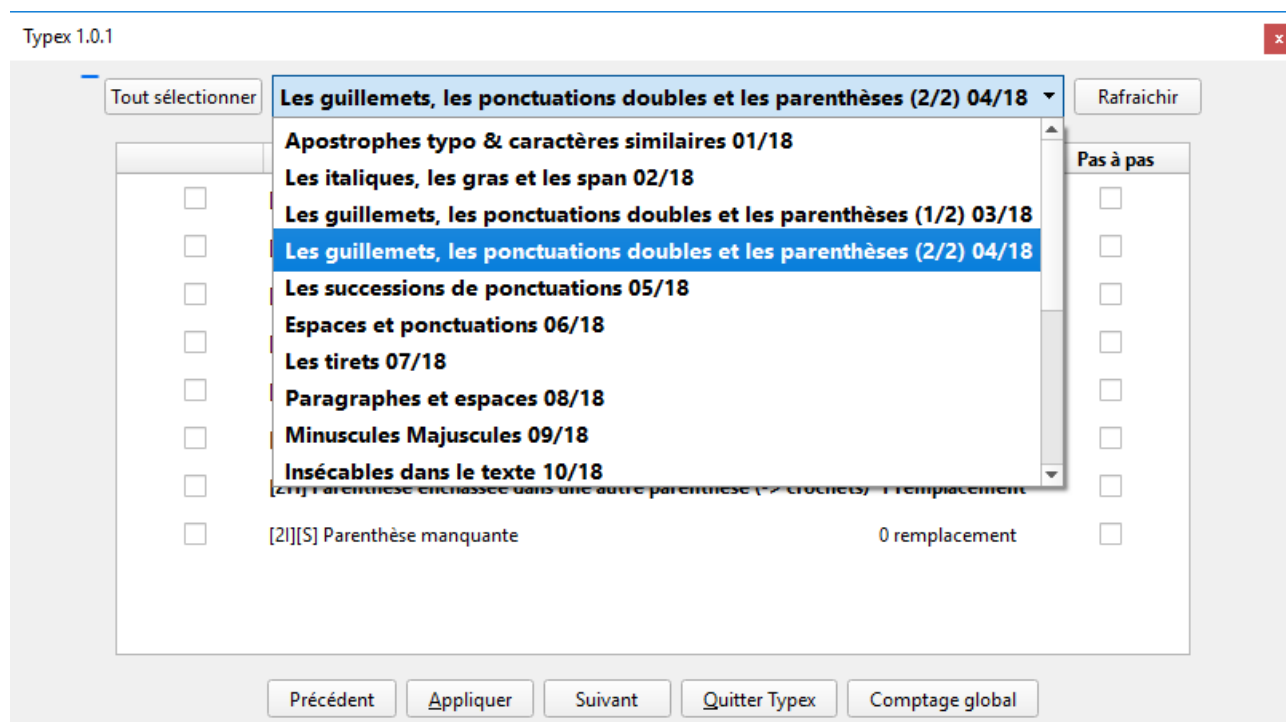
En haut à gauche, figure le numéro de version de Typex.

[**Tout sélectionner**] sélectionne toutes les règles positives.

nn/16 : numéro de la section.

[**Rafraichir**] recalcule les prévisions de remplacement. L'exécution de certaines règles peut modifier les prévisions d'autres règles, en révélant des remplacements masqués. Ce rafraichissement est effectué automatiquement après chaque exécution de regex.

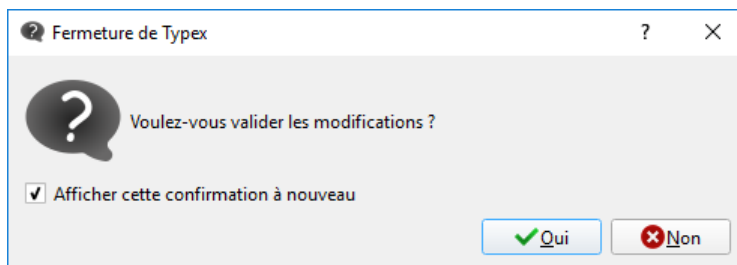
[**Intitulé de la section**] est déroulant, ce qui permet d'accéder directement à la section voulue sans avoir à faire défiler les écrans :




En cliquant sur [**Précédent**], Typex affichera l'écran précédent recalculé.

[**Appliquer**] exécute dans l'ordre les regex **de la présente section**. On peut sélectionner toutes les règles positives comme indiqué plus haut, ou bien les exécuter en plusieurs groupes, ou encore une par une. Typex n'exécute que les règles cochées dans cette section, pas celles des sections précédentes. Penser à appliquer les règles désirées avant de passer à la section suivante.

[**Quitter Typex**] quitte le logiciel. Si une modification a été faite (y compris dans la phase préalable), vous avez le message de fermeture suivant :



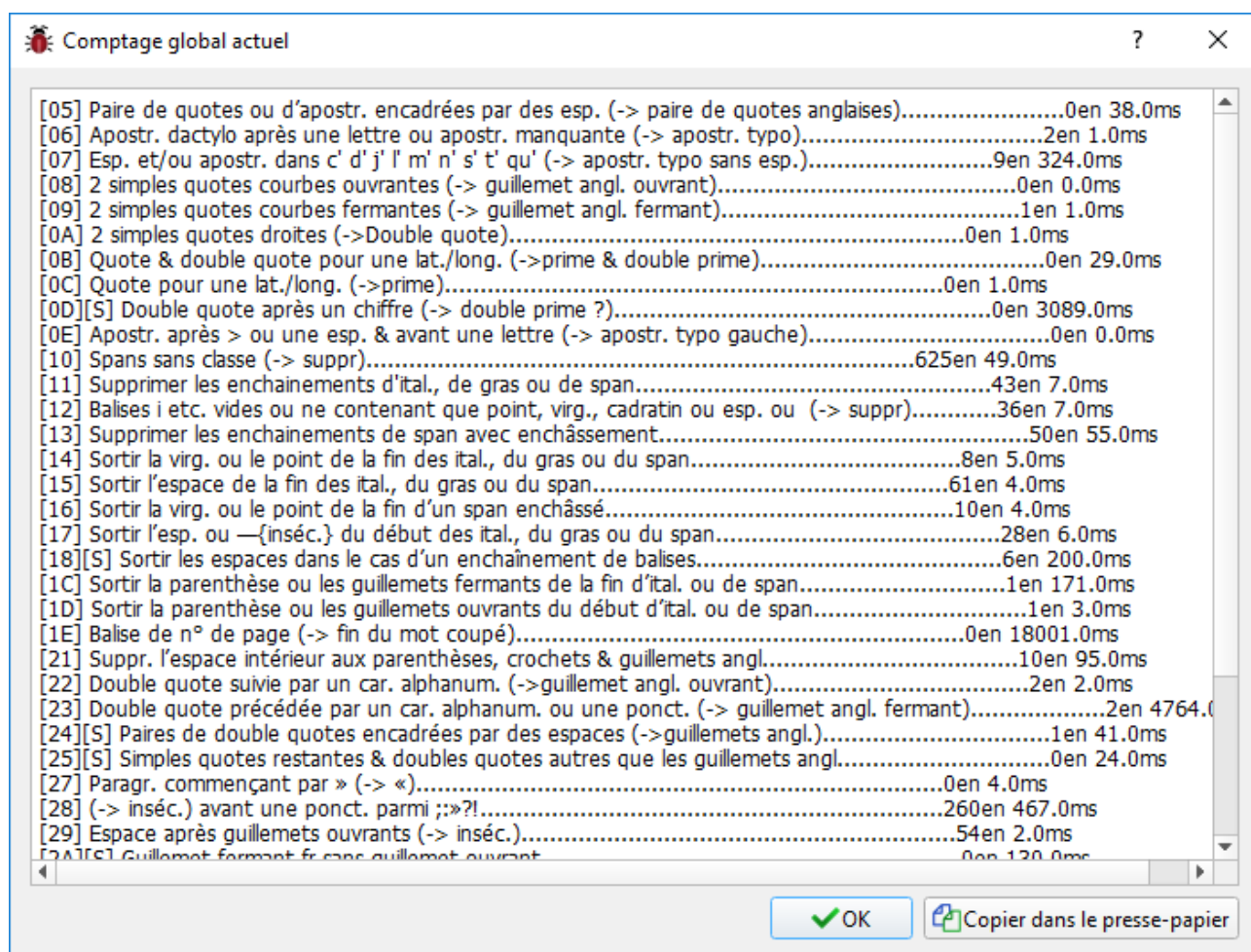
Cette validation consiste à pérenniser les modifications de Typex dans le fichier de travail affiché dans l'éditeur de Calibre. Ensuite, pour sauvegarder sur disque, utiliser la commande **de l'éditeur** *Fichier / Sauvegarder*, ou cliquer sur « Sauvegarder et quitter » en quittant l'éditeur.

 Typex crée un point de contrôle (en utilisant le mécanisme prévu par Calibre) lors de son initialisation. Annuler les modifications consiste à revenir à ce point de contrôle, c'est à dire à l'état pré-typex, toutes les modifications effectuées manuellement en marge de Typex et durant son application **seront perdues**.

Ne pas valider est toutefois pratique si on estime qu'on a appliqué indûment des règles de Typex et qu'on désire recommencer à zéro.

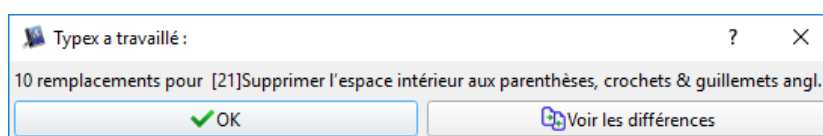
Vous pouvez aussi recourir à la commande de l'éditeur *Fichier / Sauvegarder une copie*, qui permet notamment de suivre l'évolution de l'epub. Dans cet objectif, il faut garder l'epub original intact et ne plus faire *Fichier / Sauvegarder*.

[**Comptage global**] affiche la fenêtre suivante :



Remplacement automatique

En cochant des cases de gauche, puis en cliquant sur *[Appliquer]*, la fenêtre suivante s'affiche :



Elle affiche règle par règle le nombre de remplacements pour la ou les règles sélectionnées. Ici, une seule règle.

Il est recommandé de choisir **[Voir les différences]**. Dans ce cas, Typex affiche côte à côte le code source avant le changement (côté gauche) et après (côté droit). Chaque ligne changée est affichée sur fond bleu entre 2 lignes de contexte. Les caractères modifiés sont affichés en rose côté gauche (ci-dessous, l'espace avant la parenthèse) et les caractères ajoutés, en vert.



En bas à droite, vous pouvez [Annuler les changements] ou bien [Fermer], qui valide les changements (mais ne sauvegarde pas sur disque, voir plus haut, « Phase préalable »).

Si vous constatez que la plupart des modifications sont souhaitables, à quelques exceptions près, il serait dommage d'annuler. Avec [Annuler les changements], c'est tout ou rien : on accepte tous les changements ou aucun.



Pour bien comprendre la méthode proposée ci-dessous, il est important de remarquer que cette fenêtre des différences figure dans la barre des tâches, et qu'il est donc facile de passer de cette fenêtre à celle de l'éditeur et réciproquement.



Voici la façon de procéder pour **annuler les seules modifications indésirables** :

Clic droit sur le paragraphe modifié à tort (côté du texte modifié)
Choisir *Ouvrir le fichier dans l'éditeur*

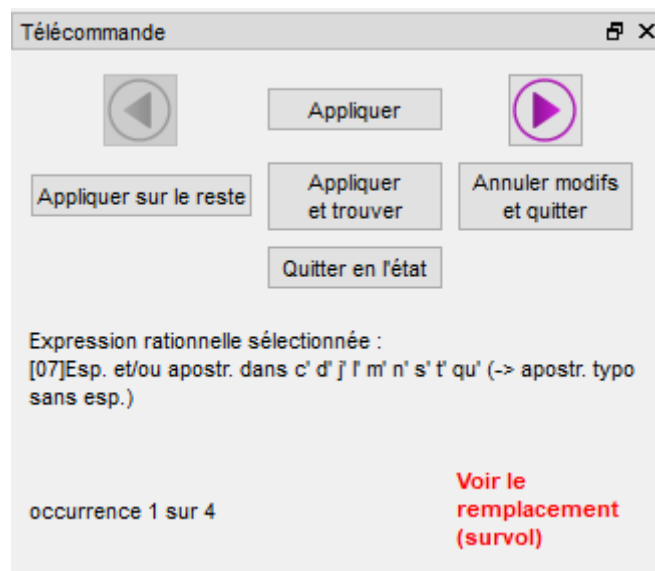
Vous retrouvez la fenêtre de l'éditeur, positionnée sur le paragraphe modifié. Vous pouvez alors annuler manuellement la modification indésirable. Si la fenêtre de Typex est gênante, déplacez-la. Retournez dans la fenêtre des différences et procédez de même pour la modification indésirable suivante. Quand vous avez terminé, cliquez sur *Fermer*. **Toutes les modifications restent effectives, sauf celles que vous avez annulées manuellement.**

Autrement dit, au moment où vous avez la fenêtre qui affiche les changements, **l'éditeur affiche les fichiers déjà modifiés** (alors que vous n'avez pas encore validé). Quand vous cliquez sur *Fermer*, il ne se passe rien de plus. En revanche, si vous cliquez sur [Annuler les changements], l'éditeur restaure les fichiers dans leur état précédent. C'est d'ailleurs cette sauvegarde qui a permis de comparer puis d'afficher les différences.

La méthode : remplacement automatique \leftrightarrow voir les différences \leftrightarrow annuler une ou plusieurs modifications est efficace règle par règle parce qu'on sait ce que la règle fait. En appliquant plusieurs règles, il faut démêler changement par changement les différentes règles appliquées. Au bout du compte, il n'est pas certain qu'on gagne du temps en appliquant plusieurs règles. Si beaucoup de changements semblent inadéquats, il sera probablement plus intéressant d'annuler tout et de reprendre avec Typex la règle en pas à pas.

Remplacement pas à pas

En cochant des cases à droite, puis en cliquant sur [Appliquer], la télécommande s'ouvre :



Elle peut être « dockée » (quand une fenêtre flottante devient un panneau à l'intérieur d'une fenêtre parente) en bas à gauche de l'éditeur de Calibre. Elle permet de naviguer d'occurrence en occurrence dans l'epub.

L'expression en cours de traitement, parmi celles qui ont été sélectionnées, figure en bas. L'occurrence en cours est indiquée en-dessous, avec le nombre total d'occurrences (ici : 1 sur 4).

À droite en survolant [Voir le remplacement (survol)] avec la souris vous pourrez visualiser le remplacement que la regex se propose d'effectuer.

Les flèches permettent de naviguer d'une chaîne sélectionnée à l'autre. Elles permettent donc de sauter une occurrence, sans appliquer le remplacement.

[Appliquer] applique le remplacement à la chaîne sélectionnée.



Quelques règles remplacent la chaîne sélectionnée par elle-même. Dans ce cas, la règle a pour objectif d'attirer l'attention sur un problème, sans pouvoir le résoudre automatiquement. Par exemple, la règle *21, parenthèse manquante* : Elle sélectionne une chaîne dans laquelle figure une parenthèse isolée, mais ne peut pas deviner où il faut ajouter la parenthèse manquante. **Lorsque la chaîne de remplacement est différente de la chaîne sélectionnée, [Voir le remplacement (survol)] s'affiche en rouge** mais reste en noir si les chaînes sont identiques

[Appliquer et trouver] applique le remplacement à la chaîne sélectionnée et va à l'occurrence suivante.

[Appliquer sur le reste] applique le remplacement à toutes les occurrences suivantes, sauf sur les occurrences que vous avez sautées avec les flèches.

[Annuler mods et quitter] annule les modifications déjà faites dans ce module et quitte la télécommande.

[Quitter en l'état] prend en compte les modifications déjà faites et quitte la télécommande.



Si vous avez sauté certaines occurrences sans appliquer le changement et que vous changez d'avis, vous pouvez ainsi sortir de la télécommande puis appliquer à nouveau la même règle.

En quittant la télécommande, la fenêtre principale d'affichage de la section en cours réapparaît.

Comparer les fichiers epub

En phase de test du logiciel, travailler sur une copie de l'epub d'origine, puis suivre son évolution est possible avec la commande de l'éditeur *Fichier / Sauvegarder une copie*. La commande *Fichier / Comparer à un autre livre* permet la comparaison des versions. Cependant, si votre epub nécessite la suppression de retours ligne à l'intérieur d'un paragraphe, un embellissement, la suppression de fines ou de pseudo-fines, cette comparaison devient vite fastidieuse.



Une façon de faire pourrait être de lancer Typex, de le quitter sans exécuter de règle, et de sauvegarder dans l'éditeur. En repartant de ce fichier, les modifications de la phase préalable ne viendront plus polluer votre comparaison. Sinon, il peut être prudent de simplement embellir l'epub

et de le sauvegarder avant de lancer Typex sur l'epub obtenu.

Par ailleurs, lorsqu'il y a beaucoup de modifications, la comparaison de fichiers peut être aussi une façon efficace de communiquer à l'intérieur d'une équipe d'ebookers.