

Consigne — Table des matières, sommaire et index automatiques

Insérer dans le .docx d'origine des champs Word qui se mettent à jour avec F9. Révision 5.

Nouveauté de la révision 3. Toute insertion d'une division automatique — sommaire, table des matières, index — doit sécuriser ses **deux** frontières, amont et aval. Ces champs ne se déploient qu'au F9 : tant qu'ils ne sont pas générés, ils n'occupent qu'une ligne, et la division qui les suit paraît bien placée alors qu'elle se collera à eux une fois le champ développé. Le nouveau piège C (§2) énonce la règle ; les §3.4, §3.5, §4 et §6 en tirent les conséquences.

Nouveauté de la révision 4. Le §3.4 intègre l'expérience d'un sommaire posé puis réduit : recopie du cache d'une table native pour obtenir un sommaire peuplé dès l'injection ; réglage de la profondeur par le commutateur \o "1-N" ; décalage de pagination propre à l'insertion en tête ; et, en amont, l'audit des styles de titre dont dépend toute table.

Nouveauté de la révision 5. Le §3.5 précise la portée de la règle des pages impaires : demandée par l'auteur, elle s'applique à toutes les divisions de niveau 1 (Titre1), non à la seule division insérée. Correction au §3.6 : updateFields se place à sa position dans la séquence du schéma (avant footnotePr), non juste après la balise settings, ce qui invaliderait le fichier.

À lire avant d'agir

Lire toute la consigne avant de commencer : les choix de marquage dépendent des sections « pages impaires » et « vérification ».

C'est une boîte à outils. Mais une distinction commande tout le reste, et c'est l'enseignement principal de la révision 2 : l'index s'injecte en XML et fonctionne ; le sommaire et la table des matières, eux, gagnent à être posés nativement dans Word, car un champ de table injecté en XML ne se rafraîchit pas de façon fiable sous Word, même valide, même marqué « à recalculer » et même vidé de son cache.

À cet enseignement s'en ajoute un second, transversal à toute insertion : **insérer une division ne sécurise que la frontière amont**. La division qui suit l'élément inséré reste, elle, sans rupture, et vient se coller. Cela vaut pour le sommaire, la table des matières et les index indifféremment. Voir le piège C (§2).

Voie par défaut : éditer directement le vrai .docx d'origine (section 3). La voie Markdown (annexe A) n'est qu'un secours si l'original est perdu.

Les pièges (section 2) et la vérification (section 4) s'appliquent toujours, quelle que soit la voie.

1. Vérifier le fichier et obtenir le vrai original

Vérifier la nature du fichier : la commande file. « Microsoft Word 2007+ » signale un vrai OOXML, c'est l'idéal ; « Unicode text » signale du Markdown.

Piège d'ingestion. Un .docx déposé dans un Projet est aplati en Markdown pour l'indexation : la copie du Projet perd la typographie. L'auteur a normalement l'original ; le faire joindre directement à la conversation, et non via le Projet, préserve l'OOXML réel. Vérifier ensuite avec file qu'on a bien un vrai .docx.

Vrai .docx obtenu : section 3, voie directe. Original introuvable : annexe A, voie de secours.

2. Les pièges (toujours valables)

Piège A. Du texte masqué (vanish) sur les marques XE donne un index vide : le terme bascule dans une clé secondaire. Ne jamais masquer les runs d'un champ XE ; ils doivent être nus. Word n'affiche de toute façon pas les codes de champ, le corps reste propre sans masquage.

Piège B. Le couple unpack / pack de la skill fusionne les runs de même format, et un champ ainsi fusionné n'est plus reconnu par Word. Un champ XE, INDEX ou TOC est fait de trois runs séparés : début, code, fin. Éditer le XML avec un vrai parseur (lxml) ou par édition directe du zip, sans fusionner les runs.

Piège C — le collage de la division aval (vécu). Insérer une division automatique entre deux divisions existantes ne règle que la frontière amont : on pose un saut **avant** l'élément inséré, mais la frontière **aval** reste sans rupture, et la division suivante se colle. Le cas typique : on injecte un index entre la conclusion et la bibliographie ; la bibliographie se retrouve collée à la fin de l'index, en milieu de page.

Ce piège est **général** : il concerne toute demande d'insertion — **sommaire, table des matières et index**. La raison est commune : ce sont des **champs** qui ne se déploient qu'au F9. Tant que le champ n'est pas généré, il n'occupe qu'une ligne de réserve, la division suivante paraît bien placée ; une fois le champ développé (sommaire ou table sur plusieurs pages, index sur deux à quatre pages), cette division remonte se coller à sa dernière ligne.

Deux corollaires de méthode. D'abord : **ne jamais juger la séparation sur le brouillon non généré** ; raisonner sur la mise en page post-F9. Ensuite : **le PDF de LibreOffice ne la révèle pas non plus de façon fiable**, car il gonfle l'index (un renvoi par ligne, ≈ six pages) ; la division suivante peut alors tomber en haut d'une page par hasard, alors que sous Word le champ se compacte, finit en milieu de page et la colle. La séparation ne se garantit donc que par un **saut structurel**, pas par lecture d'un rendu.

Règle. Toute insertion d'une division impose de (re)poser un saut **avant la division qui suit**, pas seulement avant la division insérée. Voir le « comment » au §3.5 et les contrôles aux §4 et §6.

3. Voie directe — éditer le vrai .docx (par défaut)

3.1 Principe

Le fichier a déjà ses styles, son thème, souvent un pied de page numéroté et parfois des tables. On n'ajoute que ce qui manque. On ne touche ni aux styles, ni au thème, ni aux notes. Résultat : typographie préservée à l'identique (styles.xml et theme1.xml restent inchangés).

Outillage : charger le zip, parser word/document.xml, insérer les éléments, réécrire le zip avec [Content_Types].xml en premier. Jamais unpack/pack.

3.2 Index (concepts et noms) — par injection XML

L'index est le cas où l'injection XML est la bonne méthode et la seule praticable à grande échelle.

Marque XE : trois runs nus, ajoutés à la fin du paragraphe de la première occurrence du terme.

```
<w:r><w:fldChar w:fldCharType="begin"/></w:r><w:r><w:instrText xml:space="preserve"> XE  
"Concept" </w:instrText></w:r><w:r><w:fldChar w:fldCharType="end"/></w:r>
```

Sous-entrée : un deux-points dans le libellé (« Invisibilisation:matérielle » donne l'entrée « Invisibilisation », sous-entrée « matérielle »). Catégories d'index : un identificateur \f "x" sur chaque marque.

```
<w:instrText xml:space="preserve"> XE "Bourdieu, P." \f "a" </w:instrText>
```

Champ INDEX : \c "2" pour deux colonnes, \h "A" pour les séparateurs de lettres, \z "1036" pour le français. Un index par catégorie en ajoutant \f "x" au champ.

```
<w:p><w:r><w:fldChar w:fldCharType="begin" w:dirty="true"/></w:r><w:r><w:instrText  
xml:space="preserve"> INDEX \f "a" \h "A" \c "2" \z "1036"  
</w:instrText></w:r><w:r><w:fldChar w:fldCharType="separate"/></w:r><w:r><w:t  
xml:space="preserve">[Index : F9 pour générer]</w:t></w:r><w:r><w:fldChar  
w:fldCharType="end"/></w:r></w:p>
```

Cloisonnement : un index sans \f ramasse les marques non taguées. Pour des index totalement séparés, donner un \f à chaque marque. Un index de concepts laissé sans \f et un index de noms en \f "a" restent indépendants, ce qui suffit en pratique.

Détails de marquage qui ont coûté des erreurs : chercher les sigles en mot entier, sinon PPE remonte dans « développe » ; tenir compte des apostrophes typographiques (droite et courbe) dans les motifs ; pour l'index des noms, rechercher les patronymes en respectant la casse, sinon des mots courants font des faux positifs (« Super » contre « super »).

Rappel transversal (piège C). L'index est une grande division : lui appliquer les règles de pagination du §3.5, et les appliquer **aussi à la division qui le suit** (souvent la bibliographie), faute de quoi celle-ci se colle à l'index une fois généré.

3.3 Index des noms (index nominum) — méthode

Le construire à partir des citations réelles du texte, en style auteur-date, et non d'une liste d'auteurs établie à la main. Le corps cite souvent bien plus d'auteurs qu'une telle liste : extraire les noms des citations du texte, puis les croiser avec la bibliographie.

Personnes seulement. Exclure la bibliographie, redondante puisqu'elle est déjà classée par nom, et écarter les auteurs collectifs et les institutions.

Libellés au format de la bibliographie du livre. Si elle est en APA 7, c'est « Nom, Initiales. », les initiales étant reprises de la bibliographie elle-même.

Signaler à l'auteur les noms cités dans le texte mais absents de la bibliographie : ce sont peut-être des références manquantes. Lui signaler aussi les institutions écartées, et le laisser trancher les cas limites.

Titres visibles : styler les titres « Index... » avec le style de titre du livre (Titre1) pour qu'ils figurent dans la navigation et dans le sommaire.

3.4 Table des matières et sommaire — de préférence natifs

Un sommaire ne vaut que ce que valent les styles de titre (vécu). Le champ ne fait que refléter les paragraphes stylés en titre. Avant d'insérer un sommaire ou une table — et après toute correction de structure —, auditer les styles et corriger deux défauts symétriques qui, sinon, polluent ou tronquent la table. Premier défaut : des paragraphes de prose ou des puces stylés par erreur en titre remontent en faux intertitres (phrases entières, lignes « - ... ») ; les rétrograder en style de corps, les puces en liste réelle, et ôter leurs signets _Toc parasites. Second défaut : de vrais sous-titres laissés en style de corps (des « 13.2.3 ... » saisis en texte courant) sont absents de la table ; les promouvoir au bon niveau, retirer le numéro littéral, leur poser un signet _Toc neuf. Repérage : signaler les titres anormalement longs ou commençant par « - », et les sous-titres numérotés restés en corps. Après correction, synchroniser les deux tables : ôter du cache les entrées des titres rétrogradés (appariées par leur signet _Toc) et insérer celles des titres promus (signet neuf, numéro de page provisoire) dans l'ordre du document, puis F9.

Le constat, répété en pratique : un champ TOC injecté en XML ne se rafraîchit pas de façon fiable sous Word. S'il conserve des entrées en cache, Word se contente de mettre à jour les numéros de page sans reconstruire la table ; vidé de son cache et marqué « à recalculer », il peut malgré tout rester inerte.

Recommandation : ne pas fabriquer le sommaire ni la table des matières par injection. Deux cas se présentent.

- Le document a déjà un sommaire ou une table natifs : les conserver tels quels ; l'auteur les rafraîchit sous Word par Ctrl+A puis F9, en choisissant « mettre à jour toute la table ».
- Il en manque un : le faire poser par l'auteur via Références puis Table des matières. Sommaire condensé : niveaux 1-2. Table détaillée : niveaux 1-3. Il sera alors pleinement actif.

Si l'on doit tout de même injecter un champ, faute d'accès à Word : donner un champ propre, sans entrées en cache, et prévenir l'auteur qu'une réinsertion native restera peut-être nécessaire. Vider le cache est nécessaire, pas suffisant. Marquer dirty le fldChar de début de tout champ existant que l'on veut voir se reconstruire.

```
<w:p><w:r><w:fldChar w:fldCharType="begin" w:dirty="true"/></w:r><w:r><w:instrText  
xml:space="preserve"> TOC \o "1-2" \h \z \u </w:instrText></w:r><w:r><w:fldChar
```

```
w:fldCharType="separate"/></w:r><w:r><w:t xml:space="preserve">[Sommaire : Ctrl+A puis F9]</w:t></w:r><w:r><w:fldChar w:fldCharType="end"/></w:r></w:p>
```

Sommaire injecté mais peuplé — par recopie d’une table native (vécu). Quand le document possède déjà une table native (souvent la table détaillée de fin), ne pas injecter un champ vide : recopier ses entrées en cache aux niveaux voulus — TM1 seul pour un sommaire à un niveau, TM1 et TM2 pour deux — dans un nouveau champ TOC posé sous le titre « Sommaire ». Les copies réutilisent tels quels les signets _Toc des titres (plusieurs renvois vers un même signet sont licites) ; retirer leurs w14:paraId et w14:textId pour éviter les identifiants en double. Le sommaire est ainsi lisible dès l’injection, et non réduit à une ligne d’attente. Marquer malgré tout dirty le fldChar de début, et conserver le repli natif du §5 : on a vérifié la validité, le rendu LibreOffice et l’équilibre statique, non que « mettre à jour toute la table » reconstruise un TOC injecté sous Word.

Profondeur du sommaire — un seul réglage. Le nombre de niveaux tient au commutateur \o "1-N" : \o "1-1" ne garde que le niveau 1 (parties, chapitres, introduction, conclusion, index, bibliographie), \o "1-2" ajoute les sous-sections, \o "1-3" donne le détail. Pour réduire un sommaire déjà posé — cas vécu : un 1-2 trop long, ramené à 1-1 —, changer le code du champ et élaguer du cache les entrées des niveaux supprimés (ôter tous les TM2 pour passer de 1-2 à 1-1). Sous Word, l’auteur fait de même par clic droit, Modifier le champ, Table des matières, « Afficher les niveaux ».

Décalage de pagination (insertion en tête). Un sommaire inséré en tête occupe plusieurs pages une fois généré et décale tous les numéros suivants. Conséquence : les numéros en cache du sommaire recopié comme ceux de la table de fin deviennent faux, et toute entrée nouvellement créée porte un numéro provisoire. Le F9 « mettre à jour toute la table » doit donc être passé sur les deux tables, pas seulement sur le sommaire.

Frontière aval du sommaire et de la table (piège C). Sommaire et table des matières sont eux aussi des champs qui se déploient au F9 : la division qui les suit (introduction, première partie, ou — pour la table de fin — la quatrième de couverture ou rien) doit redémarrer sur une nouvelle page. Qu’on les pose en natif ou qu’on les injecte, vérifier que la division suivante porte bien un saut avant elle ; ne pas se fier au brouillon, où le champ non généré n’occupe qu’une ligne. Même remarque pour le sommaire placé en tête : la page de titre et le sommaire doivent être séparés par un saut (cf. §3.5, page impaire).

3.5 Sections — pages impaires, et frontières des divisions insérées

En édition, une grande division démarre sur une page impaire (recto), au moyen d’un saut de section « page impaire ».

Portée — « les pages impaires » visent toutes les divisions, pas seulement celle qu’on insère (vécu).

Quand l’auteur demande d’appliquer la règle éditoriale des pages impaires, l’appliquer à toutes les divisions de niveau 1 du livre — introduction, chaque partie, chaque chapitre, conclusion, index, bibliographie, table des matières — et non à la seule division qu’on vient d’insérer. Repérage : ce sont les titres de niveau 1 (Titre1). Un pageBreakBefore sur le style de titre ne suffit pas : il ouvre une nouvelle page, pas forcément un recto. Méthode : devant chaque division, un saut de section oddPage dans un paragraphe vide dédié — sémantiquement, le sectPr qui ferme la section d’une division porte oddPage et la fait commencer en recto —, et le sectPr final du corps en oddPage pour la dernière division ; soit environ un saut par division. Cas vécu : un livre n’appliquait la règle nulle part (un seul flux de section plus pageBreakBefore, sans pied de page) ; l’« appliquer » a voulu dire introduire une section oddPage devant chacune des divisions de niveau 1. À signaler à l’auteur : des versos blancs apparaissent là où une division finissait déjà en recto, et chaque partie occupe sa propre page de garde si le premier chapitre est lui aussi mis en recto.

Le saut va dans un paragraphe vide dédié. Ne jamais logger le sectPr dans un paragraphe qui porte déjà un champ ou un titre : cela provoque une réparation ou un plantage. Mettre le sectPr dans le paragraphe de fin du sommaire, là où se referme le champ TOC, a fait planter le rendu.

```
<w:p><w:pPr><w:sectPr> [copie : footerReference, pgSz, pgMar, cols...] <w:type w:val="oddPage"/> [...] </w:sectPr></w:pPr></w:p>
```

Cloner le sectPr existant du document, qui porte déjà le pied de page et la géométrie, et y ajouter le type oddPage. Ordre des enfants : headerReference et footerReference d'abord, puis le type, puis pgSz et pgMar ; sinon Word réclame une réparation.

Sémantique : le type décrit le début de la section que ce sectPr termine. oddPage sur le sectPr qui ferme une division fait commencer cette division en page impaire. Mettre aussi la section finale, c'est-à-dire le sectPr du corps, en oddPage pour la dernière division.

Couvrir toutes les divisions que l'on veut en recto, frontispice compris. Piège vécu : si la frontière entre la page de titre et le sommaire reste un simple saut de page, le sommaire tombe en page paire (page 2). Pour le mettre en recto, cette frontière doit être, elle aussi, un saut de section « page impaire ».

Toute division — insérée ou déjà présente — relève de ces règles (pour une insertion, sa frontière aval aussi : piège C). Concrètement, à chaque insertion :

- La **division insérée** démarre sur une nouvelle page. Tirage simple : un w:pageBreakBefore sur son titre. Édition (recto) : un saut de section « page impaire » dans un paragraphe vide dédié, placé **avant** son titre.
- La **division qui suit** (typiquement la bibliographie après un index, l'introduction après un sommaire) redémarre elle aussi sur une nouvelle page — page impaire en édition. Lui poser le même type de saut, dans un paragraphe vide dédié **placé entre la fin du champ inséré et le titre suivant**, jamais dans le paragraphe du titre ni dans celui qui porte le champ.
- Le saut « page impaire » se loge toujours dans un paragraphe vide à part ; on clone le sectPr existant et on lui ajoute oddPage ; on met aussi le sectPr final du corps en oddPage pour la dernière division.

Rappel de cohérence : un w:pageBreakBefore sur le titre d'une division ne protège que sa frontière amont. Il ne dispense jamais de traiter la frontière aval de la division précédente lorsqu'on vient d'insérer un champ entre les deux.

3.6 Densité et recalcul

Défaut : une marque par chapitre, première occurrence dans chaque bloc de titre de niveau 1, avec un plafond de six à huit renvois par entrée.

Exclure du marquage les titres, les tableaux, la bibliographie et la table des matières. En lxml, n'itérer que les paragraphes enfants directs du corps : ceux des tableaux ne le sont pas. S'arrêter au titre de la bibliographie.

Mesurer le nombre de marques ; viser deux à quatre pages d'index sous Word ; proposer à l'auteur d'ajouter ou de retirer des termes plutôt que de trancher seul.

Enfin, insérer updateFields à vrai dans settings.xml, à sa position dans la séquence du schéma — typiquement juste avant footnotePr, et non au début juste après la balise settings, ce qui casse la validation —, pour demander le recalcul à l'ouverture.

4. Vérifier réellement

La conversion en PDF ne reconstruit pas les champs, et le mode service UNO est bloqué dans le bac à sable. Utiliser une macro Basic « un coup » qui charge le document, boucle sur getDocumentIndexes().update() (ce qui couvre tables et index), exporte en PDF puis se termine elle-même par StarDesktop.terminate(). Le premier lancement initialise le profil et n'exécute pas la macro : relancer une seconde fois.

```
timeout 180 soffice --headless --invisible --nologo --norestore \ -
```

```
env:UserInstallation=file:///home/claude/lo_macro \
```

```
"vnd.sun.star.script:Standard.Module1.UpdateAndExport?  
language=Basic&location=application"
```

Lire le PDF et contrôler le nombre de champs attendus, puis que chacun est peuplé, avec des entrées et de vrais numéros, et non le texte d'attente.

Limites importantes de cette vérification, à ne pas ignorer.

- LibreOffice ignore le type oddPage à l'export PDF : la parité des pages ne s'y vérifie pas, les sections sortent en pages consécutives, et les versos blancs voulus n'y apparaissent pas. Ne contrôler la parité que sur la structure, et la confirmer sous Word.
- Que LibreOffice reconnaisse un champ TOC ne garantit pas qu'il se rafraîchira sous Word : la reconnaissance par LibreOffice n'est pas l'activation sous Word (voir 3.4).
- **Frontière aval (piège C).** Ni le brouillon non généré ni le PDF de LibreOffice (champ gonflé) ne montrent fiablement la séparation entre un champ inséré et la division qui le suit. Vérifier donc **sur la structure** qu'un saut existe avant le titre de la division insérée ET avant le titre de la division suivante. À défaut, sous Word, le champ se compacte et la division suivante se colle.

Artéfact d'affichage à ne pas « corriger » : LibreOffice empile les renvois d'index, un par ligne, et gonfle l'index à environ six pages ; sous Word le champ se compacte en deux à quatre pages. Le fichier cible étant Word, c'est le rendu Word qui fait foi.

Repli quand LibreOffice est indisponible, ce qui arrive : vérification statique. Recharger le fichier avec un analyseur strict ; contrôler l'équilibre des fldChar, soit autant de débuts que de fins ; vérifier la présence des champs voulus, marqués « à recalculer » ; vérifier des sauts de section complets et bien placés, **y compris celui de la frontière aval de chaque champ inséré**. C'est aussi, faute de Word, le seul contrôle possible de la règle des pages impaires.

5. Activation côté utilisateur

Sous Word : accepter la mise à jour des champs à l'ouverture, ou faire Ctrl+A puis F9. Pour une table, choisir « mettre à jour toute la table ».

Si le sommaire reste inerte malgré tout, le supprimer et le réinsérer par Références puis Table des matières. C'est le moyen sûr, et il vaut mieux le prévoir d'emblée que d'insister sur le champ injecté.

6. Check-list

- ☐ Fichier vérifié avec file ; vrai .docx obtenu, sinon annexe A.
- ☐ Édition par lxml ou zip direct, jamais unpack/pack ; marques XE sans vanish, en trois runs.
- ☐ Index concepts et noms injectés : XE plus INDEX (\c "2", \f par catégorie) ; titres en style de titre.
- ☐ Index des noms bâti sur les citations réelles du texte ; personnes seulement ; libellés au format de la bibliographie ; noms hors bibliographie signalés à l'auteur.
- ☐ Sommaire et table des matières : natifs de préférence ; sinon champ propre marqué « à recalculer », avec réinsertion native si besoin.
- ☐ **Frontière aval (piège C) :** pour toute insertion — sommaire, table, index — la division insérée ET la division qui suit redémarrent chacune sur une nouvelle page (page impaire en édition). Vérifier que l'insertion n'a pas collé la division suivante une fois le champ généré.
- ☐ Sauts « page impaire » dans des paragraphes vides dédiés, jamais dans un titre ni dans un paragraphe portant un champ ; sectPr clonés de l'existant ; sectPr final du corps en oddPage ; frontispice compris. Règle des pages impaires demandée = appliquée à toutes les divisions de niveau 1 (Titre1), non à la seule division insérée.
- ☐ updateFields à vrai ; styles, thème et notes inchangés.

- ☐ Vérification : champs peuplés lus dans le PDF de la macro ; parité des pages et activation des tables contrôlées sur la structure et sous Word, pas dans le PDF.

Annexe A — Voie de secours (original perdu) : via Markdown

À n'utiliser que si le vrai .docx est introuvable. La copie du Projet est du Markdown ; on la convertit. La typographie d'origine n'y est pas, le Markdown ne portant aucune police, taille ni couleur : on obtient le gabarit par défaut de pandoc, sauf à fournir une charte via --reference-doc.

Convertir avec pandoc -f markdown+raw_attribute -t docx, l'option raw_attribute étant indispensable pour que les blocs openxml passent.

Dans le Markdown : insérer les marques XE selon les modèles de la section 3 ; couper l'ancienne table des matières figée, souvent deux titres et des liens PAGeref ; insérer les champs en blocs openxml ; retirer les titres vides parasites.

Retouches par zip direct : updateFields à vrai ; sauts de page par chapitre dans le style de titre 1 si voulu ; **et, pour chaque champ inséré, le saut de frontière aval (piège C).**

Reconstruire le pied de page, absent du Markdown : créer word/footer1.xml avec un champ PAGE, déclarer l'Override dans [Content_Types].xml, la Relationship dans document.xml.rels, et la référence de pied de page dans les sectPr.

Héritage OOXML : une section sans référence de pied de page hérite de la précédente, jamais de la suivante. Mettre la référence dans la première section, ou dans toutes, pas seulement la dernière, sinon seul le dernier bloc est numéroté.

Les sections 2 (pièges) et 4 (vérification) s'appliquent à l'identique. Dès que le vrai .docx est disponible, préférer la voie directe.