

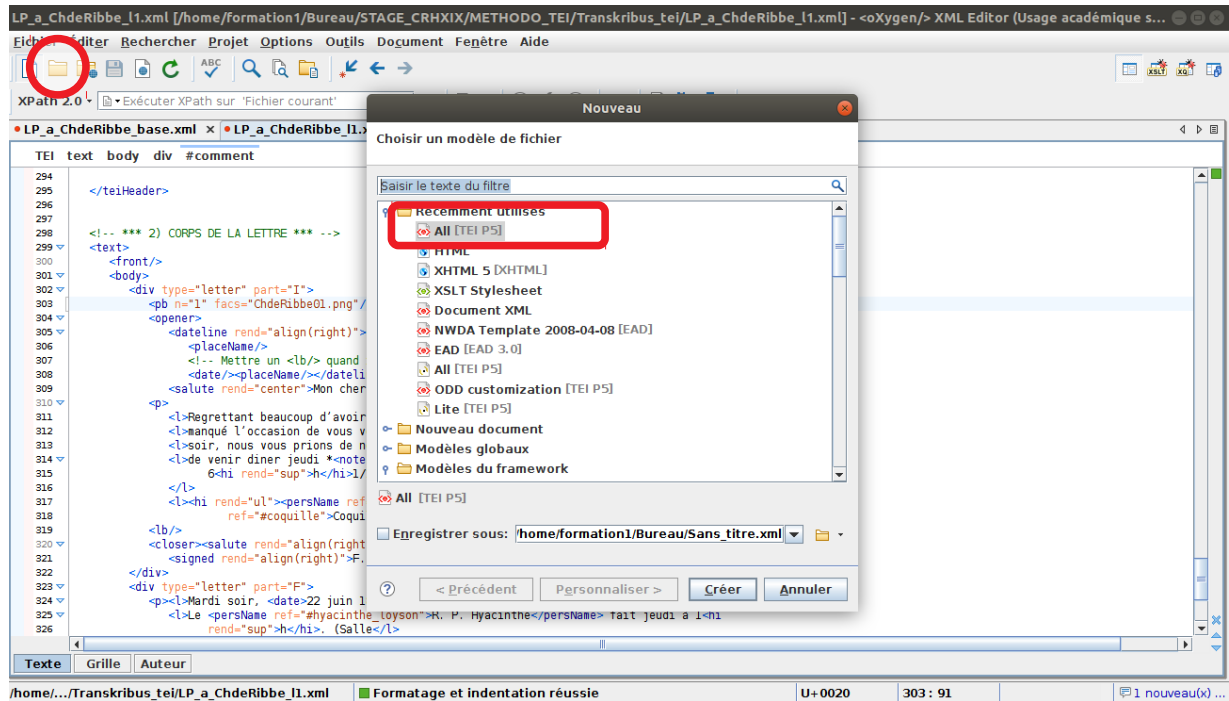
Tutoriel pour l'encodage XML

En complément à l'ODD mais ne s'y substitue pas

Lucie Slavik
Stage mai-juillet 2020
CRHXIX

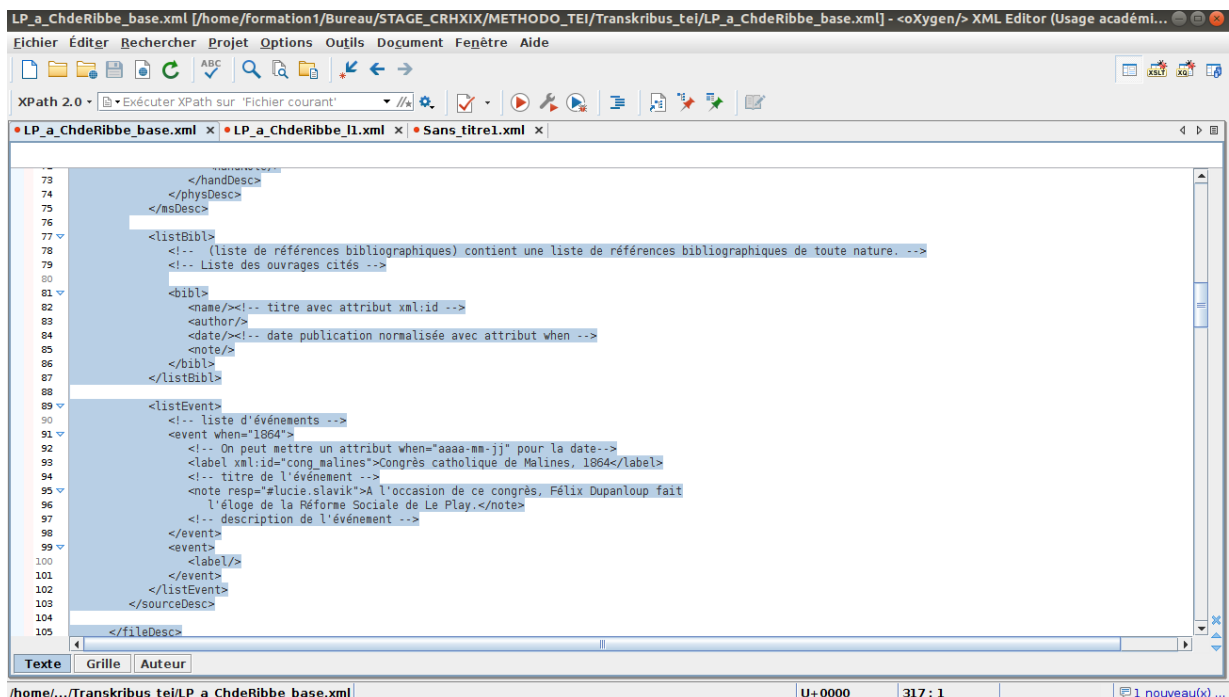
1) Créer un fichier XML

- aller sur l'icône dossier.
- Sélectionner le modèle de fichier ALL [TEI P5]

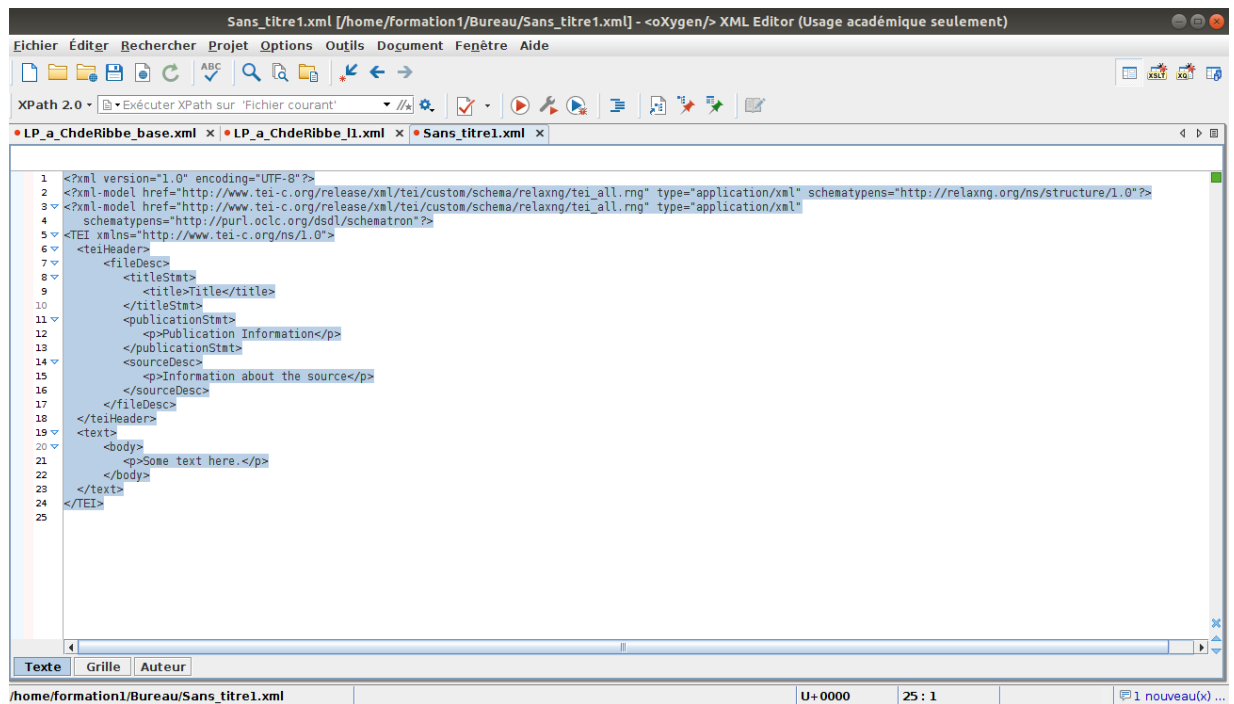


- si une base commune pour le fonds en question a déjà été faite, la copier et la coller dans le fichier nouvellement créé.

Pour cela, aller sur le fichier de base (LP_a_ChdeRibbe_base.xml) et faire ctrl + A pour tout sélectionner puis ctrl + C pour le copier.



Retourner sur le fichier nouvellement créé (Sans_titre1.xml), tout sélectionner (Ctrl + A) et copier la base (ctrl + V).

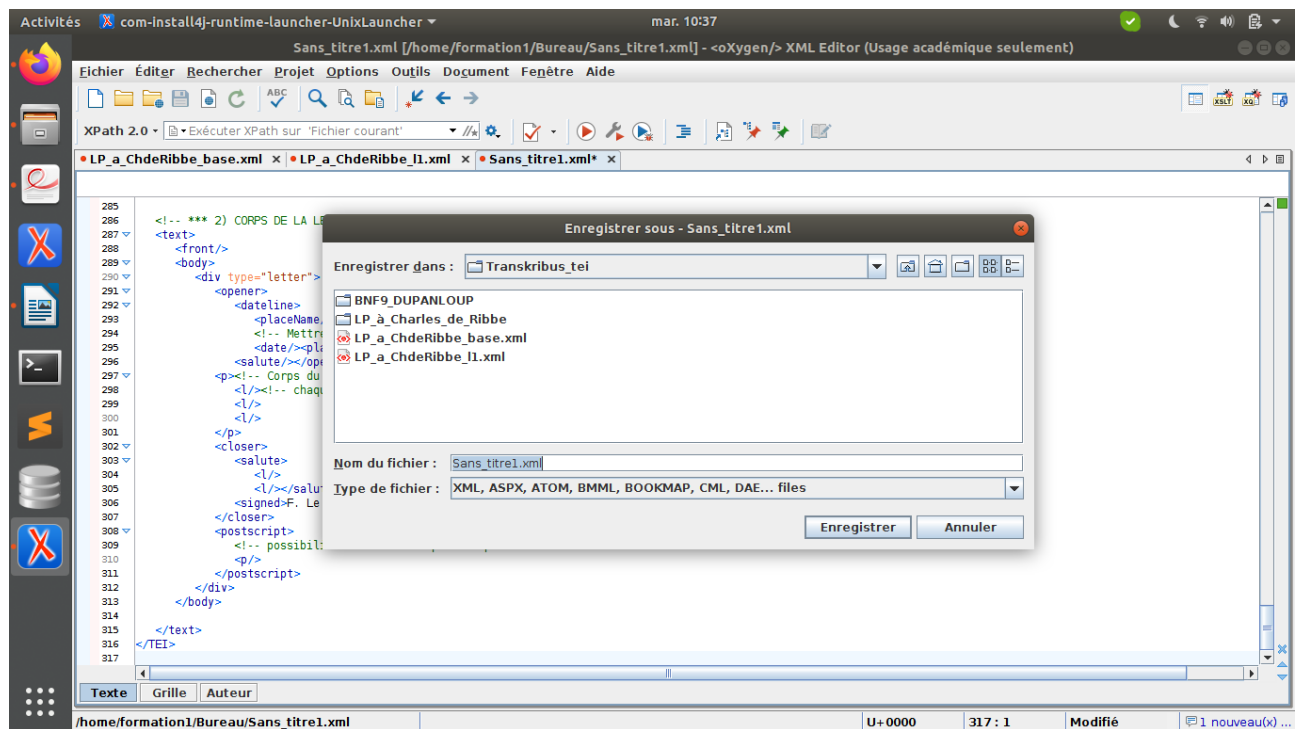


Enregistrer : Ctrl + S

Une fenêtre s'ouvre.

Renommer le fichier selon sa numérotation : LP_a_ChdeRibbe_l2.xml (l2 = lettre 2)

Cliquer sur enregistrer.



2) Copier la transcription dans le fichier XML nouvellement créé

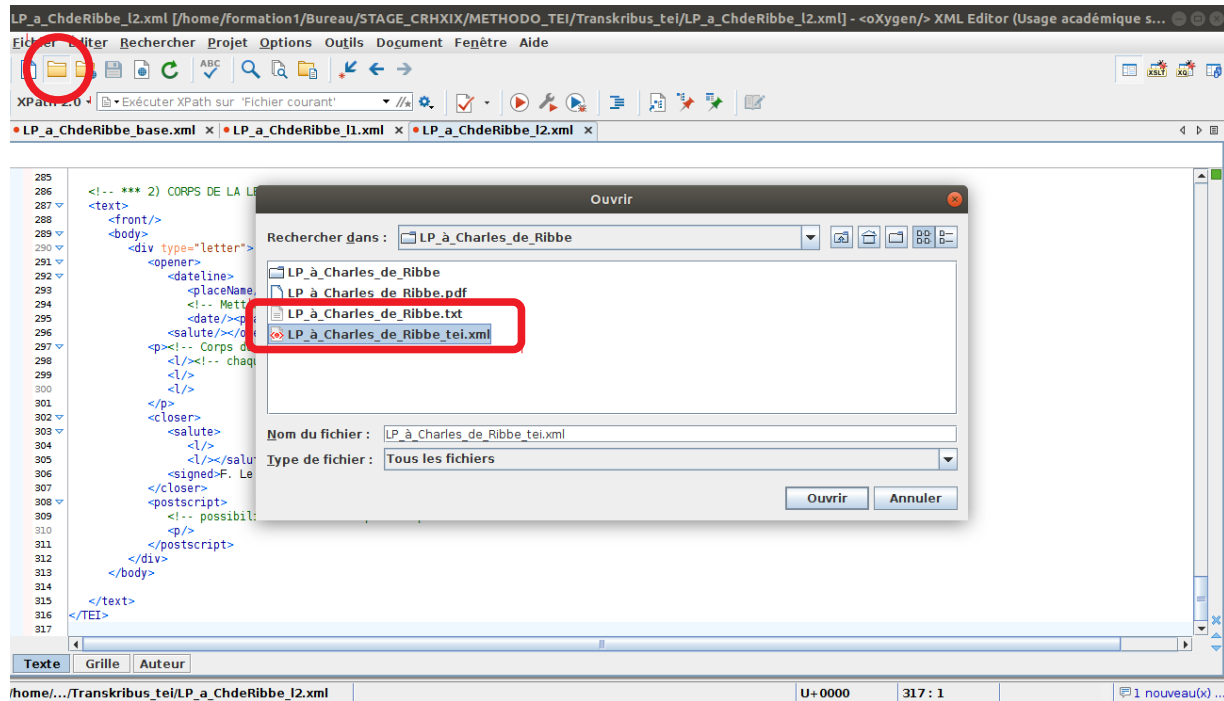
Aller dans le dossier où sont les transcriptions et numérisations des manuscrits exportés de Transkribus.

Sélectionner soit le format TXT (texte brut), soit le format XML selon que l'on est plus à l'aise avec l'un ou l'autre.

L'avantage du TXT est sa simplicité. On ne se perd pas dans les informations. Mais il est donc très incomplet.

L'avantage du XML est son pré-balisage. Beaucoup d'informations sont à supprimer, mais c'est peut-être quand-même plus rapide. Je choisis le XML pour ce tutoriel.

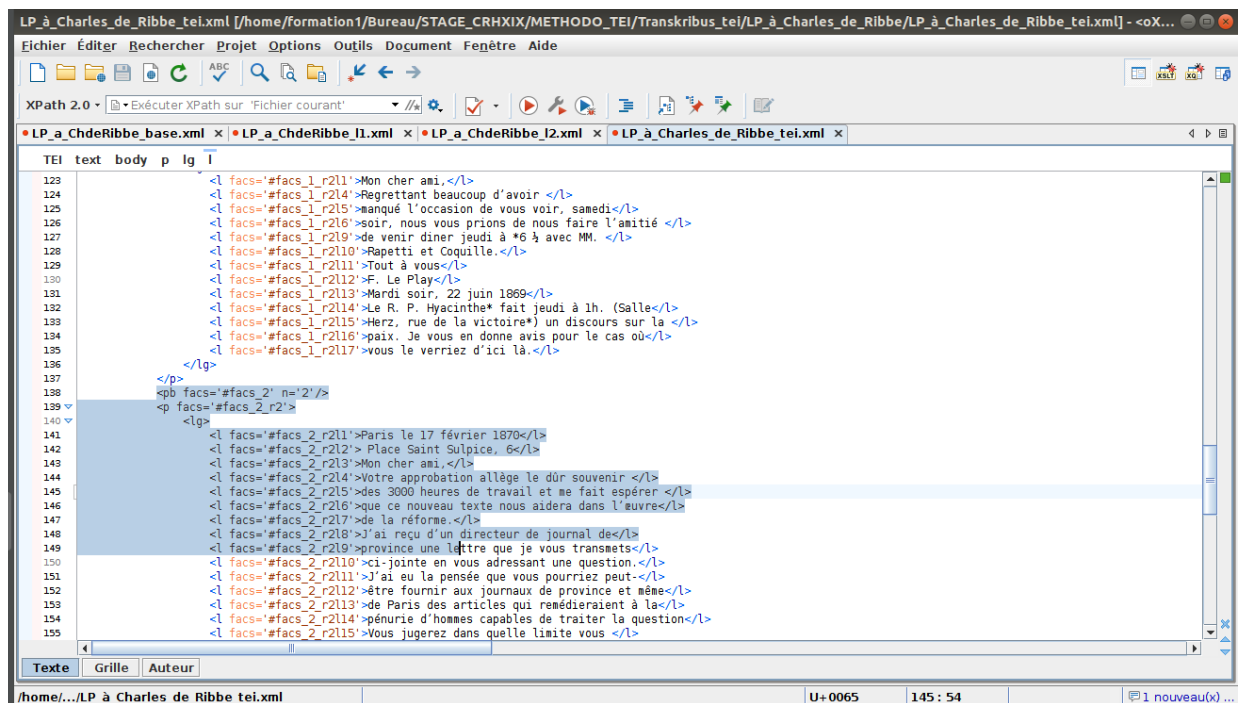
Cliquer sur OUVRIIR.



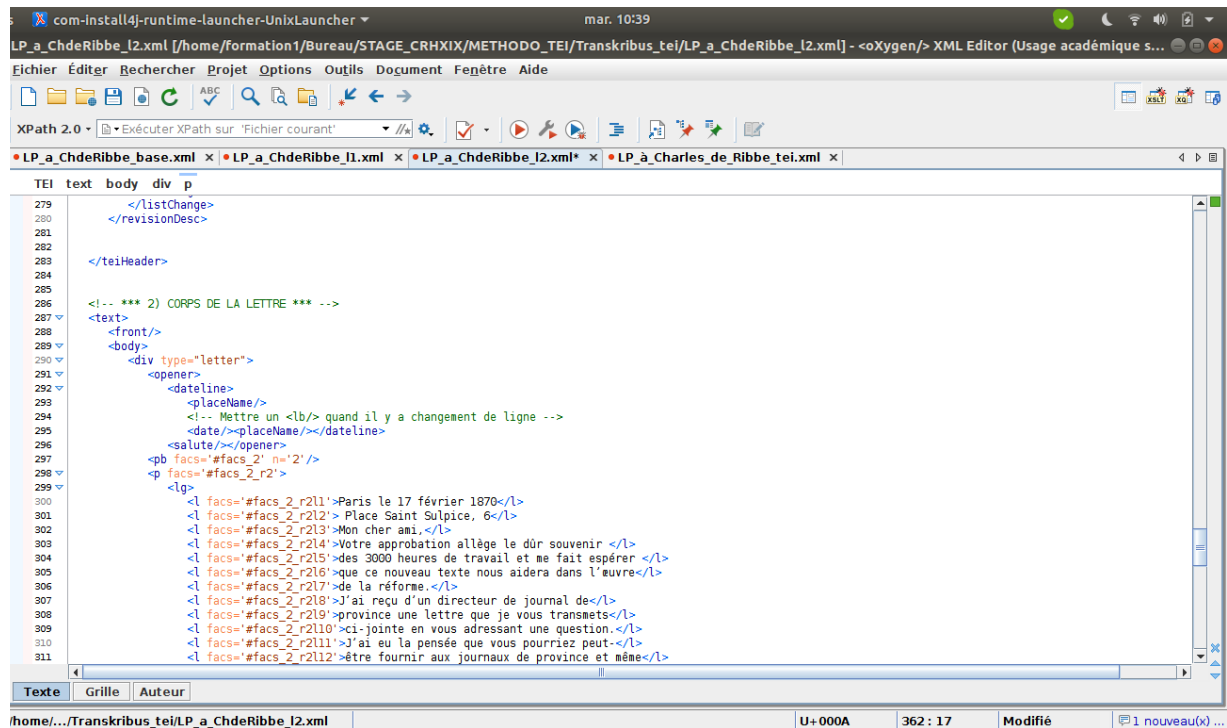
Sélectionner dans le XML la partie qui nous intéresse. Ici, c'est la lettre 2.

Ne pas s'encombrer du reste.

Copier : ctrl C



Coller dans notre fichier : ctrl V
 Soit sous la <div>
 Soit sous l' <opener/>



3) Nettoyer le fichier XML et encoder correctement selon nos besoins.

a) Nettoyer

```
<date/><placeName/></dateline>
<salute/></opener>
<pb facets='#facets_2' n='2' />
<p facets='#facets_2_r2'>
  <lg>
    <l facets='#facets_2_r2l1'>Paris le 17 février 1870</l>
    <l facets='#facets_2_r2l2'>Place Saint Sulpice, 6</l>
```

Mettre les bonnes informations (les bons attributs) dans la balise <pb>
 Les remettre dans l'ordre.

n= « 2a »

C'est la deuxième lettre, première page.

Facs : lien vers le manuscrit numérisé.

Copier le nom : ChdeRibbe2A.png

(Il faut préalablement veiller à un bon nommage des manuscrits. Il faudra le faire peu à peu, avec une logique de projet. Pour ma part, ce que j'ai fait est encore une ébauche).



```
<text>
  <front/>
  <body>
    <div type="letter">
      <pb n="2a" facets="ChdeRibbe02A.png" />
      <opener>
        <dateline>
          <placeName>Paris</placeName>le
```

Nettoyer en enlevant tous les attributs qui ont été ajoutés aux l. Ils ne nous servent pas.

```
<!-- Mettre un <lb/> quand il y a changement de l:
<placeName>Paris</placeName>le <!-- Mettre un <lb/> quand il y a changement de l:
<date>17 février 1870</date><placeName/><lb/><placeName>Place Saint Sulpice,
6</placeName></dateline>
<salute/></opener>

<salute>Mon cher ami,</salute>
<l>Votre approbation allège le dūr souvenir </l>
<l>des 3000 heures de travail et me fait espérer </l>
<l facs="#facs_2_r2l6">que ce nouveau texte nous aidera dans l'œuvre</l>
<l facs="#facs_2_r2l7">de la réforme.</l>
<l facs="#facs_2_r2l8">J'ai reçu d'un directeur de journal de</l>
<l facs="#facs_2_r2l9">province une lettre que je vous transmets</l>
<l facs="#facs_2_r2l10">ci-jointe en vous adressant une question.</l>
<l facs="#facs_2_r2l11">J'ai eu la pensée que vous pourriez peut-</l>
<l facs="#facs_2_r2l12">être fournir aux journaux de province et même</l>
<l facs="#facs_2_r2l13">de Paris des articles qui remédieraient à la</l>
<l facs="#facs_2_r2l14">pénurie d'hommes capables de traiter la question</l>
<l facs="#facs_2_r2l15">Vous jugerez dans quelle limite vous </l>
<l facs="#facs_2_r2l16">pourriez exercer ce genre de propagande. Mais</l>

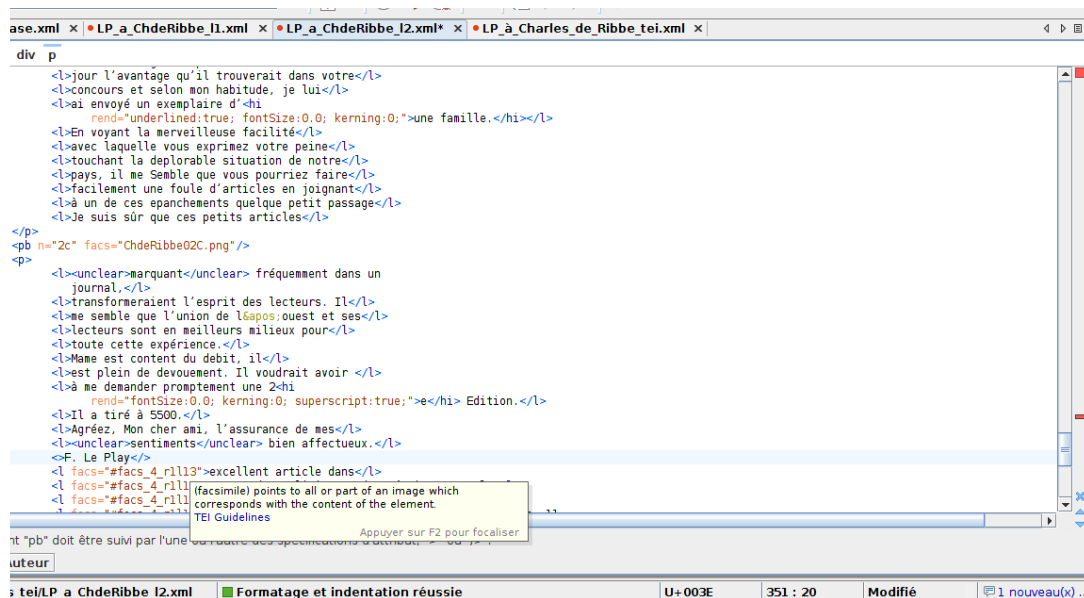
<pb facs="#facs_3" n="3"/>
<p facs="#facs_3_r1">
  <lg>
    <l facs="#facs_3_r1l1">alors même que vous devriez vous</l>
    <l facs="#facs_3_r1l2">borner à votre region, je vous prierais</l>
```

b) Encodage

Enlever les <lg>.

Regrouper les <l> dans des balises <p> à chaque paragraphe.

Pour cela, sélectionner tous les <l> qui nous intéressent. Faire ctrl + E, taper p, cliquer sur entrée, et le tour est joué : désormais les <l> en question sont entouré d'une balise <p> qui marque le paragraphe.



Pour changer le nom des balises, aller à l'intérieur de la balise, enlever par exemple <l>, et écrire <signed> à la place (pour la balise qui entoure la signature de Le Play). Ainsi, la balise fermante est modifiée en même temps que la balise ouvrante.

```
<l>Mame est content du debit, il</l>
<l>est plein de devouement. Il voudrait avoir </l>
<l>à me demander promptement une 2<hi
  rend="fontSize:0.0; kerning:0; superscript:true;">e</hi> Edition.</l>
<l>Il a tiré à 5500.</l>
<l>Agréez, Mon cher ami, l'assurance de mes</l>
<l><unclear>sentiments</unclear> bien affectueux.</l>
<signed>F. Le Play</signed>
<l facs="#facs_4_r1l13">excellent article dans</l>
<l facs="#facs_4_r1l14">Les Etudes religieuses des Jésuites, par le</l>
<l facs="#facs_4_r1l15">R. P. Clair-N°. 26. paru le 15 février.</l>
```

Affiner l'encodage :

Relever chaque nom de lieu, de personnage, etc. Tout ce que l'on doit indexer.

On renseigne le nom en question dans l'index.

```
<settingDesc>
  <listPlace>
    <!-- Index des noms de lieux -->
    <place xml:id="paris">
      <placeName>Paris</placeName>
      <country>France</country>
      <!-- Voir si c'est utile de mentionner le pays -->
      <note/>
      <!-- On peut renseigner quelque chose ici sur le lien entre Le Play et Paris, le nombre d'années qu'il y a passées ou autre -->
    </place>

    <place xml:id="stDom">
```

Puis on fait pointer le texte vers l'index avec l'attribut ref et le #

```
<pb n="2a" facs="ChdeRibbe02A.png" />
<opener>
  <dateline>
    <placeName ref="#p">Paris</placeName>le <!-- Mettre un <lb/> quand il y a changement de ligne -->
    <date>17 février 1870</date> <!-- #paris -->
    <salute/></opener>

  <salute>Mon cher ami,</salute>
  <p><!-- Votre approbation a été reçue le 17 février 1870 -->
    <!-- des 3000 heures de travail -->
    <!-- que ce nouveau texte -->
    <!-- de la réforme. -->
  </p>
  <p>
    <!-- J'ai reçu d'un directeur de journal de -->
    <!-- province une lettre que je vous transmets -->
    <!-- ci-jointe en vous adressant une question. -->
    <!-- J'ai eu la pensée que vous pourriez peut-être -->
  </p>
```

On repère l'emplacement de la date dans la lettre. Après l'avoir mise dans la balise correspondante, copier la date et renseigner dans le teiHeader la date d'envoi de la lettre.

```
LP_a_ChdeRibbe_base.xml x LP_a_ChdeRibbe_IL.xml x LP_a_ChdeRibbe_IJ.xml x LP_a_Charles_de_Ribbe_tei.xml x
TEI text body div opener dateline date
287 <!-- *** 2) CORPS DE LA LETTRE *** -->
288 <text>
289 <front>
290 <body>
291 <div type="letter">
292 <pb n="2a" facs="ChdeRibbe02A.png" />
293 <opener>
294 <dateline>
295 <placeName ref="#paris">Paris</placeName>le <!-- Mettre un <lb/> quand il y a changement de ligne -->
296 <date>17 février 1870</date></placeName></lb/><placeName>Place Saint Sulpice,
297 6</placeName>
298 <salute/>
299 <salute>Mon
300 <p><!-- Votre approbation a été reçue le 17 février 1870 -->
301 <!-- des 3000 heures de travail -->
302 <!-- que ce nouveau texte -->
303 <!-- de la réforme. -->
304 </p>
305 <p>
306 <!-- J'ai reçu d'un directeur de journal de -->
307 <!-- province une lettre que je vous transmets -->
308 <!-- ci-jointe en vous adressant une question. -->
309 <!-- J'ai eu la pensée que vous pourriez peut-être -->
310 <!-- fournir aux journaux de province et même -->
```

ON COPIE

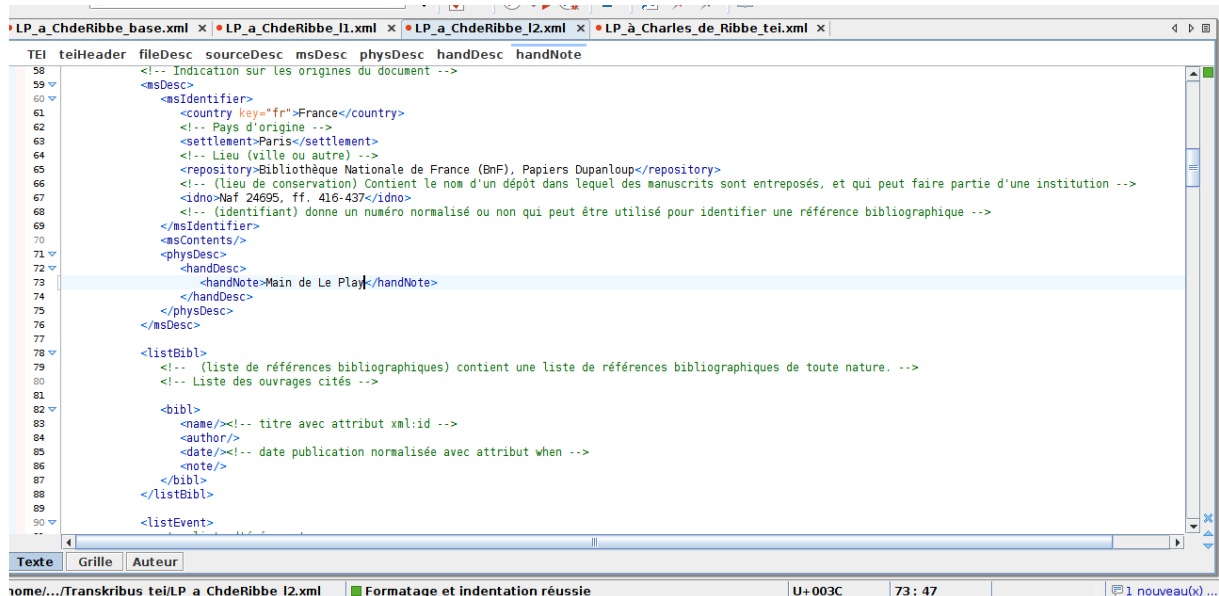
```

114
115 <correspDesc>
116 <!-- Description de l'action liée à la correspondance -->
117 <correspAction type="sent">
118 <!-- On indique qui envoie la lettre -->
119 <!-- Description normalisée et renvoi vers la fiche de data.bnf -->
120 <persName key="Le Play, Frédéric (1806-1882)"
121 ref="https://data.bnf.fr/fr/11911744/frederic_le_play/">Frédéric Le
122 Play</persName>
123 <settlement>Paris</settlement>
124 <date when="1870-02-17">17 février 1870</date>
125 <!-- Renseigner la date avec un attribut when -->
126 </correspAction>
127
128 <correspAction type="received">
129 <!-- On indique qui reçoit la lettre -->
130 <persName key="de Ribbe, Charles (1827-1899)"
131 ref="https://data.bnf.fr/fr/11921782/charles_de_ribbe/">Charles de
132 Ribbe</persName>
133 <settlement/>
134 <!-- L'indiquer comme non obligatoire dans le schéma -->
135 <date/>
136 </correspAction>
137 <correspContext>
138 <ref type="previous"/>
139 <ref type="next"/>
140 </correspContext>
141

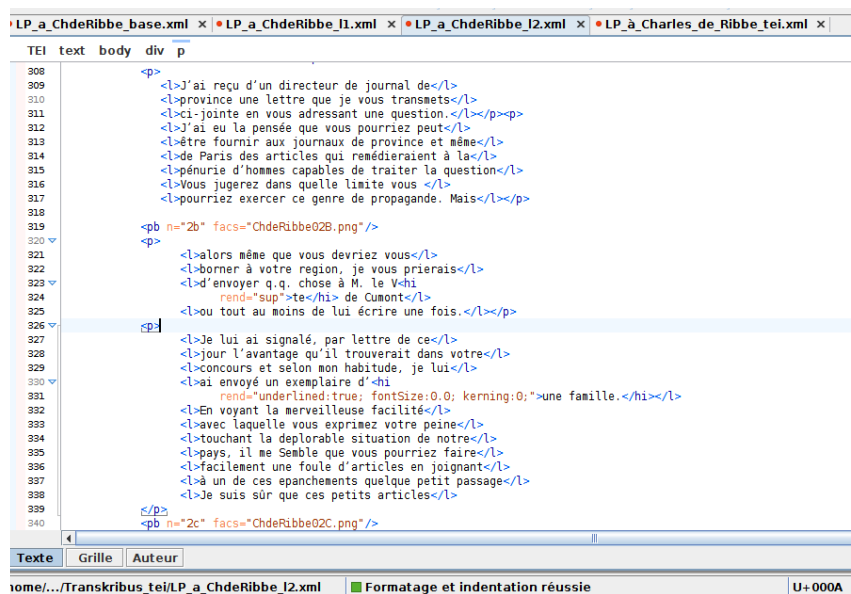
```

ON RENSEIGNE DANS LE
TEIHEADER LA DATE EN
LA NORMALISANT AVEC
L'ATTRIBUT WHEN.

Renseigner, toujours dans le teiHeader, quelle main a écrit la lettre dans le <physDesc>



A chaque changement de page de manuscrit, indiquer <pb> avec les mêmes attributs spécifiés plus haut, en adaptant leur valeur au contexte.



```
TEI text body div p
308 <p>
309 <l>J'ai reçu d'un directeur de journal de</l>
310 <l>province une lettre que je vous transmets</l>
311 <l>ci-jointe en vous adressant une question.</l></p><p>
312 <l>J'ai eu la pensée que vous pourriez peut</l>
313 <l>être fournir aux journaux de province et même</l>
314 <l>de Paris des articles qui remédieraient à la</l>
315 <l>pénurie d'hommes capables de traiter la question</l>
316 <l>Vous jugerez dans quelle limite vous </l>
317 <l>pourriez exercer ce genre de propagande. Mais</l></p>
318
319 <pb n="2b" facs="ChdeRibbe02B.png"/>
320 <p>
321 <l>alors même que vous devriez vous</l>
322 <l>borner à votre région, je vous prierais</l>
323 <l>d'envoyer q. q. chose à M. le V<hi>
324 <rend="sup">te</hi> de Cumont</l>
325 <l>ou tout au moins de lui écrire une fois.</l></p>
326 <p>
327 <l>Je lui ai signalé, par lettre de ce</l>
328 <l>jour l'avantage qu'il trouverait dans votre</l>
329 <l>concours et selon mon habitude, je lui</l>
330 <l>ai envoyé un exemplaire d'<hi>
331 <rend="underlined:true, fontSize:0.0; kerning:0;">une famille.</hi></l>
332 <l>En voyant la merveilleuse facilité</l>
333 <l>avec laquelle vous exprimez votre peine</l>
334 <l>touchant la déplorable situation de notre</l>
335 <l>pays, il me semble que vous pourriez faire</l>
336 <l>facilement une foule d'articles en joignant</l>
337 <l>à un de ces épanchements quelque petit passage</l>
338 <l>Je suis sûr que ces petits articles</l>
339
340 </p>
341 <pb n="2c" facs="ChdeRibbe02C.png"/>
```

Se référer à l'ODD pour l'encodage complet.

Ce tutoriel a pour but de compléter l'ODD mais il ne se substitue aucunement à une lecture attentive de l'ODD pour l'encodage des fichiers.

Pour mieux comprendre la logique interne du fichier XML, se référer aux TEIGUIDELINES.
<https://tei-c.org/guidelines/p5/>

Pour relier les fichiers XML à l'ODD, voir le tutoriel [Tuto_TEI_ODD.pdf](#) dans le dossier ODD.