

Ébauche de guide d'encodage (corpus des correspondances)

M2 ENC – février 2013 (v2)

Cette ébauche de guide d'encodage s'inspire des préconisations énoncées dans France Martineau, Yves Charles Morin, Caroline Fauchon *et al.*, *Protocole de codage TEI (Modéliser le changement : les voies du français)*, 2010, Université d'Ottawa, révision 1.11. adaptées pour les besoins du cours d'édition TEI de l'ENC.

1. L'entête (teiHeader) des lettres

1.1 La description du fichier (fileDesc)

1.1.1 Le titre et les responsabilités (titleStmt)

Titre

Reprise du modèle du gabarit fourni pour l'élément <title>. Il est suggéré de remplacer le sous-élément desc contenant l'intitulé de la lettre par un élément **<title type=letter>** et de faire figurer l'idno dans le même élément <title>.

```
<title>
  <title type="letter">Justine, épouse d'Entremont, écrit à Guillaume d'Entremont, son
beau-frère</title>
  <title type="corpus">Corpus Fran Argyle</title>
  <desc>
    <desc n="type1">Historique</desc>
    <desc n="type2">correspondance personnelle</desc>
  </desc>
  <date type="date-creation_doc_original">18680807</date>
  <date type="siècle">19</date>
  <location>
    <region n="pays">Canada</region>
    <region n="province/etat">Nouvelle-Écosse</region>
    <region n="Argyle"/>
    <region n="autre"/>
  </location>
  <idno
type="Cote_GRTC">FNEARG-H-LET_Argyle_EntJu18680807_IM_CM_F_20130930</idno>
</title>
```

Responsabilités

Reprise du modèle du gabarit fourni.

<editor> L'institution responsable de l'édition. Prédéterminé.

<principal> Le nom du chercheur responsable du projet. Prédéterminé.

<funder> L'institution responsable du financement. Prédéterminé.

Puis autant d'éléments **respStmt** que nécessaire.

1.1.2 Édition (editionStmt)

Ajout d'un élément editionStmt reprenant les informations qui figuraient auparavant dans un sous-élément bibl de encodingDesc :

<editionStmt>

<edition>Première édition, réalisée dans le cadre du projet **<orgName>**Le français à la mesure d'un continent: un patrimoine en partage (CRSH/GTRC)**</orgName>** en **<date when="2014">**2014**</date>** **</edition>**
</editionStmt>

1.1.3 Publication (publicationStmt)

Reprise du modèle du gabarit.

Est-il nécessaire de redoubler publisher/distributor ?

- **<publisher>** Nom de l'institution responsable de la publication ou diffusion d'un document TEI, l'éditeur (coordonnées du MCVF) ;

- **<distributor>** Nom de l'institution responsable de la diffusion du document TEI (coordonnées du MCVF) ;

/!\ Bien songer à reprendre la description de la licence figurant dans le gabarit.

<availability status="restricted">

<p>(c) Le français à la mesure d'un continent: un patrimoine en partage (CRSH/GTRC), dirigé par F. Martineau.**</p>**

<p>Les Corpus FRAN demeurent la propriété du projet Le français à la mesure d'un continent: un patrimoine en partage (GTRC/CRSH). Les références à ces Corpus devraient apparaître ainsi: Corpus FRAN, Le français à la mesure d'un continent: un patrimoine en partage (F. Martineau (dir.), U. d'Ottawa). Tous les documents du Corpus FRAN ne peuvent être divulgués à de tierces parties sans l'autorisation préalable de la directrice du Projet.**</p>**

<p>Pour tout autre usage, contacter : **<persName>**France Martineau**</persName>**,
<email>fmartin@uottawa.ca**</email>****</p>**

</availability>

1.1.4 La description des sources (sourceDesc)

Reprise du modèle de gabarit.

1.2 La description des pratiques d'encodage (encodingDesc)

1.2.1 La description du projet (projectDesc)

Reprise de la description standardisée présente dans le gabarit :

```
<projectDesc>
  <p>Le français à la mesure d'un continent: un patrimoine en partage (GTRC/CRSH)
    <lb/>Resp.:<persName>France Martineau</persName>,Département de français,
Université d'Ottawa
  <address>
    <addrLine>Université d'Ottawa, département de français</addrLine>
    <addrLine>60 rue Université, pièce 229b</addrLine>
    <addrLine>Ottawa (Ontario)</addrLine>
    <addrLine>K1N 6N5</addrLine>
    <addrLine>Tél : 613-562-5800, poste 1104</addrLine>
    <addrLine>Fax : 613-562-5981</addrLine>
    <addrLine><email>fmartin@uottawa.ca</email></addrLine>
    <addrLine>http://www.continent.uottawa.ca/</addrLine>
  </address></p>
</projectDesc>
```

1.2.2 Déclaration des principes d'encodage (editorialDecl)

Renvoi possible à ce document et au guide d'encodage qui l'inspire. Si, à terme, cet élément sera normalisé à l'échelle du corpus, néanmoins, pour les besoins de l'exercice d'une part et surtout pour les besoins d'harmonisation ultérieure, il est demandé à chacun d'y décrire en détail les éventuels choix faits et mis en œuvre dans l'encodage du corps du document, quitte à répéter certaines informations. Pour ce faire, il est recommandé d'utiliser les éléments spécialisés proposés par les *Guidelines* :

- **<normalization>** (normalisation) indique l'étendue de la régularisation effectuée sur le texte source dans sa conversion vers sa forme électronique.
- **<hyphenation>** : « (césurage) résume la façon dont les traits d'union sécants de fin de ligne d'un texte source ont été traités dans sa version encodée. »
- **<interpretation>** : « (Interprétation) décrit le champ d'application de toute information analytique ou interprétative ajoutée à la transcription du texte. »

complété si besoin par de simples paragraphes pour des informations qui n'entreraient pas dans ces catégories (avec <p>).

1.3 Les éléments de contexte (profileDesc)

Dans la description des éléments de contexte figurent les informations importantes qui ne font pas partie des éléments bibliographiques usuels, à savoir notamment les informations d'ordre prosopographique, sur les langues du texte, ou les éléments thématiques.

1.3.1 Les éléments prosopographiques (particDesc)

`<particDesc>` (« participation description ») permet de décrire l'auteur de la lettre, en donnant (reprise des éléments du gabarit)

- nom, sexe (*sex*), métier (*occupation*), lieu et date de naissance (*birth*) ;
- connaissances linguistiques (*langKnowledge*) ;
- éducation suivie et niveau atteint (*education*).

Il est proposé de compléter le `profileDesc` par l'ajout d'informations sur les langues utilisées dans la lettre et en reprenant le système de mots-clés du *Protocole de codage TEI*..

1.3.2 Les langues utilisées dans la lettre (langUsage)

Autant d'éléments language que de langues présentes dans la lettre, en précisant l'étendue de la présence des langues (@usage contenant un pourcentage) et en utilisant le code linguistique (@ident).

```
<langUsage>
  <language ident="fr-CA" usage="99">français du Canada</language>
  <language ident="en-CA" usage="1">anglais (quelques mots)</language>
</langUsage>
```

1.3.3 Mots-clés

Reprise du système décrit dans France Martineau *et al.*, *Protocole de codage TEI*..., « 4.3.4 La classification du document `<textClass>` », p. 20.

```
<textClass>
  <keywords scheme="GTRC">
    <list>
      <item n="domaine">Non littéraire</item>
      <item n="genre">Correspondance</item>
      <item n="docRegion1">Canada</item>
      <item n="docRegion2">Nouvelle-Écosse</item>
      <item n="docRegion3">Argyle</item>
    </list>
  </keywords>
</textClass>
```

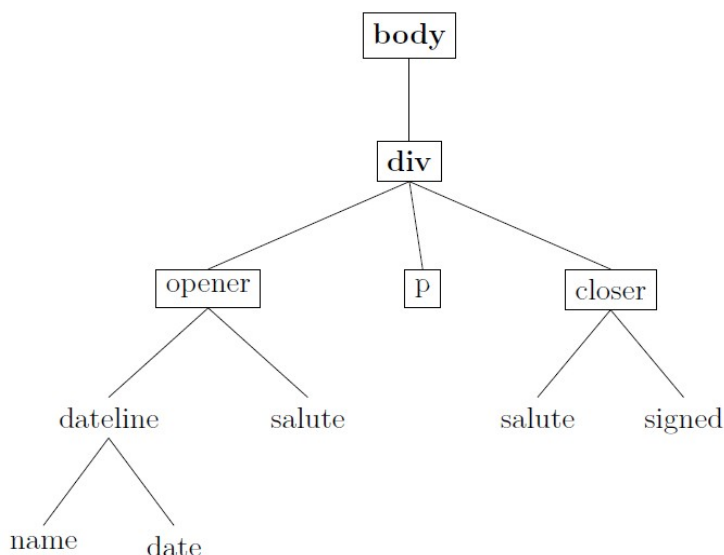
1.4 La liste des révisions du fichier (revisionDesc)

Songer à ajouter une entrée (**change**) pour votre travail d'encodage, en utilisant les attributs de date formalisée (@when) et le renvoi à l'@xml:id du `<respStmt>` (attribut @who).

2. Le corps du document

2.1 Structure logique des lettres

Au sein d'une division *div*, structure tripartite entre



opener, pour les mentions de début de la lettre, **dont dateline** pour la date et lieu, **éventuellement byline** pour l'auteur, **salute** pour les salutations ;

p pour le corps de la lettre

closer, pour les mentions finales dont **salute**, **dateline**, et **signed** pour la signature.

add pour les mentions « hors teneur » considérées comme des additions ? (impossible d'utiliser **nota**, qui figure dans le schéma de

l'ENC, mais pas dans les *Guidelines*) ; pour ces mentions, on pourrait en préciser par des attributs le type (@type) et le lieu (@place), en restant pour ce dernier attribut dans le cadre des valeurs suggérées par les *Guidelines* (below, bottom, etc., voir att.placement¹). N.B. : il est impossible de préciser, avec un attribut @hand, la main lorsque celle-ci diffère de la main principale, car la valeur d'@hand doit être un pointeur vers un élément handNote du header, qui ne peut être utilisé en TEI-LITE.

N.B. : il est souhaitable de préciser l'orientation du texte de ces mentions, quand elle diffère de celle du texte principal, car le scripteur, faute d'espace, a tendance à commettre plus d'erreurs dans ces segments.

2.2 Régularisations et normalisations

2.2.1 Corrections de la transcription

La difficulté à distinguer ce qui relève d'une normalisation graphique d'une correction d'« erreur » nous invite à utiliser pour tous les cas de normalisation reg et orig, et à conserver sic et corr, temporairement, pour marquer les corrections que nous aurions nous mêmes à apporter à la transcription fournie.

Ex. « quoique les affaires de mr wrigth ne soient pas [fiais] »

<choice>

<sic>[fiais]</sic>

<corr>finis</corr>

1 <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.placement.html>.

</choice>

Utiliser att.responsibility pour préciser l'origine et le degré de certitude de la correction

<corr cert="high" resp="#Untel">finis</corr>

N.B. : @resp contient un pointeur, qu'il faut donc spécifier dans le header

<titleStmt>

<!-- [...] -->

<respStmt>

<resp>Encodé par</resp>

<name xml:id="Untel">Untel</name>

</respStmt>

</titleStmt>

2.2.2 Normalisation des graphies

Utilisation de reg et orig

(...)

<choice>

<reg>embrasse</reg>

<orig>ambrasse</orig>

</choice>

les petits enfants

2.2.3 Abréviations

17

<choice>

<abbr>oc<hi rend="sup">bre</hi></abbr>

<expn>octobre</expn>

</choice>

1822

N.B. : Il est a priori impossible de marquer les lettres ajoutées lors de la résolution de l'abréviation en TEI Lite (voici ce que ça aurait donné à titre d'exemple)

17

<choice>

<abbr>oc<hi rend="sup">bre</hi></abbr>

<expn>oc<ex>to</ex>bre</expn>

</choice>

1822

N.B. : il est souhaitable de développer systématiquement les abréviations les plus courantes, même si elles se rapprochant de l'usage moderne (ex. « Mr » pour « Monsieur »), pour faciliter l'interrogation du corpus via PhiloLogic.

2.2.4 Régularisation des coupures de mots

Séparation de mots unis dans la source
ou même enville

```
<w rend="aggl">en</w><w>ville</w>
```

Séparation de mots donnant lieu à une élision

Leun en vers L'autre

```
<w rend="elision">L</w><w>eun</w>
```

Union de mots séparés dans la source

```
<w>en vers</w> L'autre
```

Cas particulier de l'hyphénation

mademois- // elle

```
<w>mademois<pc type="hyphen">-</pc><lb/>elle</w>
```

ou, plus simple : mademois<lb rend="hyphen"/>elle

2.2.5 Représenter les allographes

cas du s rond et du s long

```
<choice>
  <reg>s</reg>
  <orig>f</orig>
</choice>
```

Si on veut éviter des problèmes d'affichage, dus à la présence ou non du caractère dans la fonte utilisée, on peut aussi utiliser l'entité hex. XML

```
<choice>
  <reg>s</reg>
  <orig>&#383;</orig>
</choice>
```

Une simplification est possible par le biais d'entités :

```
<!DOCTYPE TEI [
  <!ENTITY s-long '<choice> <reg>s</reg> <orig>&#383;</orig> </choice>' >
]>
```

Et dans le corps du texte : « je fuis » > « je &s-long;uis »

(N.B. : emploi de <g> impossible, car pas inclus dans la TEI-Lite)

Point important : comment traiter certaines lettres transcrites comme des majuscules mais qui paraissent peut-être être plutôt des allographes (ex. r droit et ʀ rond) ? Il est préférable, dans ces cas, d'utiliser des entités en les traitant comme des allographes, quitte à adapter le traitement par la suite.

2.2.6 Autres usages typographiques

Majuscules

Outre l'emploi possible de <reg> et <orig>, il doit être possible de ne pas encoder cette normalisation et de la faire ensuite par expressions régulières (en tenant compte des éléments *name* et de la ponctuation).

Ponctuation

Pour distinguer ponctuation moderne et ancienne, des éléments **pc** typés (@type avec une valeur comme 'supplied') peuvent être utilisés.

<pc type="supplied">-</pc>

2.3 Représentation de la source

2.3.1 Changements de feuillets et de ligne

<pb n="1"/> (changement de page) et <lb/> (changement de ligne).

2.3.2 Passage difficilement lisibles, dégâts sur la source

Difficulté de lecture : <unclear> (@reason)

pas [fiais] : pas <unclear reason="illegible">fiais</unclear>

<gap> si un passage n'est pas lisible du tout (en précisant l'étendue, @quantity, @unit,...)

N.B. : <damage> ne semble pas disponible en TEI-Lite

2.3.3 Interventions dans la source

Mot raturés dans la source par le scribe

Et votre papa arangeras-vous pourvoiras

Et votre papa arangeras vous pourvoiras

[si on rencontre des cas où plusieurs mains sont à l'œuvre, il faudra préciser la main avec un attribut @hand]

Mots ajoutés dans la source par le scribe

Mr

j'ai vu Soulligni : j'ai vu <add place="above">Mr</add> Soulligni

N.B. : il existe une possibilité de grouper les deux quand cela est nécessaire, mais ce n'est pas possible en TEI-Lite

```
<subst>
  <del>arangeras</del>
  <add>vous pourvoiras</add>
</subst>
```

2.4 Autres

2.4.1 Mots étrangers employés dans le texte

À distinguer par l'élément <foreign> et un attribut @xml:lang (cf. 2e séance).

2.4.2 Noms de personnes et de lieux

Nom : `<name type="place" xml:id="Hull1">Hull</name>`,
mon cher `<name type="person" xml:id="Benjamin1">Benjamin</name>`
Date : `<date when="1822-10-17"> 17 octobre 1822 </date>`