
4 – Codage du corps du texte

Comme indiqué ci-dessus, un document TEI simple comporte, au niveau textuel, les éléments suivants :

- `<front>` regroupe tous les éléments (en-têtes, page de titre, préfaces, dédicaces, etc.) situés avant le début du texte lui-même ;
- `<group>` regroupe plusieurs textes unitaires ou groupes de textes ;
- `<body>` regroupe le corps entier d'un texte unitaire seul, à l'exclusion de toute pièce liminaire ou annexe ;
- `<back>` regroupe toutes les annexes qui suivent le texte principal.

Les éléments spécifiques aux pièces liminaires et annexes sont décrits ci-après, dans la section 19 « Pièces liminaires et annexes ». Dans la présente section, nous présentons les éléments qui constituent le corps d'un texte.

4.1. Éléments servant à marquer les divisions de texte

Le corps d'un texte en prose peut avoir la forme d'une simple suite de paragraphes ; les paragraphes peuvent également être regroupés en chapitres, sections, subdivisions, etc. Dans le premier cas, chaque paragraphe est balisé au moyen de `<p>`. Dans le second cas, l'élément `<body>` peut être divisé soit en une série d'éléments `<div1>`, soit en une série d'éléments `<div>`, l'un ou l'autre pouvant se voir subdivisé à son tour, comme nous le verrons ci-dessous :

- `<p>` marque les paragraphes écrits en texte courant ;
- `<div>` contient une subdivision des pièces liminaires, du corps, ou des annexes d'un texte ;
- `<div1>` regroupe une subdivision de premier niveau des pièces liminaires, du corps, ou des annexes d'un texte (la plus grande si `<div0>` n'est pas employée, sinon la seconde en ordre de grandeur).

Lorsque des subdivisions structurales plus petites que `<div1>` sont nécessaires, un `<div1>` peut être divisé en éléments `<div2>`, et un `<div2>` en éléments encore plus petits `<div3>`, etc., jusqu'au niveau `<div7>`. En présence de plus de sept niveaux de division structurale, il sera nécessaire soit de modifier l'ensemble du balisage TEI pour lui permettre d'accepter `<div8>`, etc., soit d'employer un élément `<div>` non numéroté : un `<div>` peut être subdivisé en de plus petits éléments `<div>` (le niveau d'imbrication n'est pas limité).

Tous ces éléments sont dotés des trois attributs suivants :

- type* indique le nom conventionnel pour cette catégorie de division de texte. Typiquement, sa valeur sera *livre*, *chapitre*, *poème*, etc. D'autres valeurs sont possibles : *roupe* pour des groupes de poèmes, etc. traités comme une seule unité, ou bien *sonnet*, *speech*, et *song*. À noter que toute valeur fournie pour l'attribut *type* du premier `<div>`, `<div1>`, `<div2>`, etc. dans un texte est supposée s'appliquer pour tous les `<div>`, `<div1>` `<divn>` ultérieurs dans le même `<body>`. Cela implique qu'une valeur doit être donnée pour le premier élément de division de chaque type, ou lors de tout changement de valeur.
- id* spécifie un identifiant unique, spécifique à cette division, qui servira dans le cas des références croisées ou d'autres liens pointant vers cette division (par exemple un commentaire) comme c'est expliqué de façon plus détaillée à la section 8 «Liens et références croisées». Il est souvent utile de fournir un attribut *id* pour chaque unité structurale majeure d'un texte et de déterminer les valeurs de l'identifiant de façon systématique, par exemple en accolant un numéro de section à un court libellé rappelant le titre de l'œuvre en question, comme ceci est montré dans les exemples suivants.
- n* indique un nom court mnémonique ou un numéro pour la division, qui serviront à l'identifier de préférence à l'attribut *id*. S'il existe déjà une forme conventionnelle de référence ou d'abréviation pour les parties d'un ouvrage existant (tel que le modèle «livre / chapitre / vers» pour les citations bibliques), ladite forme sera enregistrée dans l'attribut *n*.

Les attributs *id* et *n*, sont de fait d'un intérêt si général qu'ils sont permis sur tout élément dans n'importe quelle DTD de la TEI : ce sont des attributs globaux. D'autres exemples d'attributs globaux définis dans les recommandations *TEI Lite* sont présentés dans la section 8.3 «Attributs de liaison».

La valeur de chaque attribut *id* doit être unique au sein d'un document. Il existe une façon simple de s'assurer que ceci est bien le cas : faire en sorte que la valeur de l'attribut reflète la structure hiérarchique du document.

À titre d'exemple, l'ouvrage *Wealth of Nations* de Smith contenait dans sa première édition cinq tomes, chacun étant partagé en chapitres, certains chapitres étant eux mêmes subdivisés en parties. Nous pouvons définir les valeurs de l'attribut *id* pour cette structure comme suit :

```
<div1 id=WN1 n='I' type='book'>
  <div2 id=WN101 n='I.1' type='chapter'>
    ...
  </div2>
  <div2 id=WN102 n='I.2' type='chapter'>
    ...
  </div2>
  ...
  <div2 id=WN110 n='I.10' type='chapter'>
    <div3 id=WN1101 n='I.10.1' type='part'>
      ...
    </div3>
    <div3 id=WN1102 n='I.10.2' type='part'>
      ...
    </div3>
  </div2>
  ...
</div1>
<div1 id=WN2 n='II' type='book'>
  ....
</div1>
...
```

Pour prendre un exemple français, les *Mémoires d'outre-tombe* de Chateaubriand sont divisées en parties, ces dernières étant elles mêmes subdivisées en livres et en chapitres. Nous pouvons définir les valeurs de l'attribut *id* pour cette structure comme suit :

```
<div1 id=MOT1 n='I' type='partie'>
  <div2 id=MOT11 n='I.1' type='livre'>
    ...
  </div2>
  <div3 id=MOT111 n='I.1.1' type='chapitre'>
    ...
  </div3>
  <div2 id=MOT1 n='I.2' type='livre'>
```

```

...
</div2>
.....
<div2 id=MOT110 n='I.10' type='livre'>
  <div 3 id=MOT1101 n='I.10.1' type=chapitre>
    ...
  </div3>
  <div3 id=MOT1102 n='I.10.2' type=chapitre>
    ...
  </div3>
</div2>
.....
</div1>
<div1 id=MOT2 n='II' type='partie'>
  .....
</div1>
.....

```

Les attributs *id* et *n* peuvent être traités au moyen d'un système de numérotation différent : ceci est souvent utile lorsqu'un schéma canonique de référence est employé mais que ce dernier ne concorde pas avec la structure de l'œuvre. Par exemple, dans un roman partagé en livres, chacun contenant des chapitres, ces derniers étant numérotés séquentiellement dans la totalité de l'œuvre plutôt qu'à l'intérieur de chaque livre, on pourrait employer une représentation proche de ce qui suit :

```

<div1 id=TS01 n='1' type='Volume'>
  <div2 id=TS011 n='1' type='Chapter'>
    ...
  <div2 id=TS012 n='2'>
    ...
</div1>
<div1 id=TS02 n='2' type='Volume'>
  <div2 id=TS021 n='3' type='Chapter'>
    ...
  <div2 id=TS022 n='4'>
    ...
</div1>

```

Dans cet exemple, l'œuvre comporte deux volumes, chacun contenant deux chapitres. Les chapitres sont numérotés conventionnellement de 1 à 4, mais les valeurs spécifiées pour l'attribut *id* leur permettent d'être considérés comme s'ils étaient également numérotés par 1.1, 1.2, 2.1, 2.2.

4.2. Têtes et fins de chapitre

Chaque `<div>`, `<div1>`, `<div2>`, etc. peut commencer par un titre ou une tête de chapitre et se terminer (moins couramment) par une formule de fermeture tel que «Fin du Chapitre Premier». De tels cas peuvent être transcrits par les éléments suivants :

`<front>`

contient toute forme de titre, par exemple, le titre d'une section, ou l'en-tête d'une liste ou d'un glossaire ;

`<trailer>`

contient un formule de fin de chapitre ou un pied de page apparaissant à la fin d'une division de texte.

Certains autres éléments éventuellement nécessaires au début ou à la fin d'une division de texte sont présentés dans la section 19.1.2 «Pièces liminaires».

La décision d'inclure ou non les titres et formules dans une transcription appartient au transcripateur individuel. Lorsqu'un titre est tout à fait régulier (par exemple «Chapitre 1») ou a été utilisé comme valeur d'attribut (par exemple `<div1 type='chapitre' n=1>`), il peut être omis ; lorsqu'il contient du texte qui n'apparaît pas ailleurs, il devrait toujours être inclus. Par exemple, le début de *Under the Greenwood Tree* de Hardy pourrait être codé comme suit :

```
<div1 id=UGT1 n='Winter' type='Part'>
<div2 id=UGT11 n='1' type='Chapter'>
<head>Mellstock-Lane</head>
<p>To dwellers in a wood almost every species of tree ...
```

Pour prendre un exemple français du même genre, le début du sixième livre de *Notre-Dame de Paris* de Victor Hugo pourrait être codé comme suit :

```
<div1 id=NDP6 n='VI' type='livre'>
<div2 id=NDP61 n='1' type='chapitre'>
<head>Coup d'œil impartial sur l'ancienne
magistrature</head>
<p>C'était un fort heureux personnage...
```

4.3. Prose, vers et art dramatique

Comme nous l'avons vu plus haut, les paragraphes constituant une division textuelle devraient être balisés à l'aide de la balise `<p>`. Par exemple :

```
<body>
<p>I fully appreciate Gen. Pope's splendid achievements
with their invaluable results; but you must know that
Major Generalships in the Regular Army, are not as
plenty as blackberries.
</p>
</body>
```

Un exemple français équivalent serait :

```
<body>
<p>Nous apprécions beaucoup les résultats qui
ont été obtenus et nous en tiendrons compte.
Il faut toutefois garder en mémoire...
</p>
</body>
```

Diverses balises sont disponibles pour coder la structure des textes poétiques ou dramatiques (les pièces de théâtre, les films, etc.) :

`<l>` contient une seule ligne de vers, éventuellement incomplète ; parmi les attributs s'appliquant à cet élément citons :

part indique si la ligne est un vers complet ; les valeurs permises sont :

- Y si la ligne n'est pas un vers complet ;
- N si la ligne est complète, ou si aucune supposition n'est faite quant à son intégralité ;
- F pour la partie finale d'une ligne incomplète ;
- I pour la partie initiale d'une ligne incomplète ;
- M pour une partie médiane d'une ligne incomplète ;

`<lg>` contient un groupe de lignes de vers formant un tout, par exemple une strophe, un refrain, un paragraphe de vers, etc. ;

`<sp>` contient une réplique dans un texte théâtral, ou un passage présenté comme tel dans un texte de vers ou un prose ; parmi les attributs de cet élément, on trouve :

who identifie le locuteur, en donnant son attribut *id* ;

`<speaker>` contient une forme spéciale de titre ou d'étiquette, donnant le nom d'un ou de plusieurs locuteurs dans un texte ou un fragment de texte ;

<stage>

contient toute sorte d'indication scénique au sein d'un texte ou d'un fragment de texte dramatique ; parmi les attributs citons :

type indique le type d'indication scénique ; les valeurs suggérées sont : entrée, sortie, mise en scène, diction, etc.

Les exemples suivants représentent le début de textes poétiques dans lequel les strophes et les lignes de vers sont balisés :

```
<lg n=I>
<l>I Sing the progresse of a
  deathlesse soule,</l>
<l>Whom Fate, with God made,
  but doth not controule,</l>
<l>Plac'd in most shapes; all times
  before the law</l>
<l>Yoak'd us, and when, and since,
  in this I sing.</l>
<l>And the great world to his aged evening;</l>
<l>From infant morne, through manly noone I draw.</l>
<l>What the gold Chaldee, of silver Persian saw,</l>
<l>Greeke brass, or Roman iron, is in this one;</l>
<l>A worke t'out weare Seths pillars, bricke and stone,</l>
<l>And (holy writs excepted) made to yeeld to none,</l>
</lg>
```

```
<lg n=I>
<l>Les sanglots longs</l>
<l>Des violons</l> <l>De l'automne</l>
<l>Blessent mon coeur</l>
<l>D'une langueur</l><l>Monotone.</l>
</lg>
```

À noter que l'élément <l> marque les lignes de vers et non les lignes typographiques : le découpage typographique des premières lignes ci-dessus n'a donc pas été rendu de manière explicite par ce codage et pourrait être perdu. L'élément <lb> décrit en section 5 « Numérotation des lignes et des pages » peut être employé pour marquer des lignes typographiques si besoin est.

Quelquefois, surtout dans des textes dramatiques, les lignes de vers sont coupées entre les personnages. La façon la plus facile de coder ce phénomène est d'employer l'attribut *part* pour indiquer que les lignes ainsi fragmentées sont incomplètes, comme dans cet exemple :

```

<div1 type = 'Act' n='I'><head>ACT I</head>
<div2 type = 'Scene' n='1'><head>SCENE I</head>
<stage rend=italic>
    Enter Barnardo and Francisco, two Sentinels,
    at several doors</stage>
<sp><speaker>Barn<l part=Y>Who's there?
<sp><speaker>Fran<l>Nay, answer me. Stand and unfold yourself.
<sp><speaker>Barn<l part=i>Long live the King!
<sp><speaker>Fran<l part=m>Barnardo?
<sp><speaker>Barn<l part=f>He.
<sp><speaker>Fran<l>You come most carefully upon your hour.

```

Le même mécanisme peut être appliqué à des strophes partagées entre deux locuteurs :

```

<sp><speaker>First voice</speaker>
<lg type=stanza part=I>
<l>But why drives on that ship so fast
<l>Withouten wave or wind?
</lg>
<sp><speaker>Second Voice</speaker>
<lg part=F>
<l>The air is cut away before.
<l>And closes from behind.
</lg>

```

L'exemple suivant indique comment coder le dialogue dans une œuvre de prose comme s'il s'agissait d'une pièce. Il illustre également l'emploi de l'attribut *who* pour porter un code identifiant le locuteur de la partie de dialogue en question :

```

<sp who=OPI><speaker>The reverend Doctor Opimiam</speaker>
<p>I do not think I have named a single unrepresentable fish.
<sp who=GRM><speaker>Mr Gryll</speaker>
<p>Bream, Doctor: there is not much to be said for bream.
<sp who=OPI><speaker>The Reverend Doctor Opimiam</speaker>
<p>On the contrary, sir, I think there is much to be said
for him. In the first place...
<p>Fish, Miss Gryll -- I could discourse to you on fish by
the hour: but for the present I will forbear...
</sp>

```