

# *Cahiers* **GUT** *enberg*

## ☞ LA TEI SIMPLIFIÉE : UNE INTRODUCTION AU CODAGE DES TEXTES ÉLECTRONIQUES EN VUE DE LEUR ÉCHANGE

☞ Lou BURNARD, C.M. SPERBERG-McQUEEN

*Cahiers GUTenberg*, n° 24 (1996), p. 23-151.

<[http://cahiers.gutenberg.eu.org/fitem?id=CG\\_1996\\_\\_24\\_23\\_0](http://cahiers.gutenberg.eu.org/fitem?id=CG_1996__24_23_0)>

© Association GUTenberg, 1996, tous droits réservés.

L'accès aux articles des *Cahiers GUTenberg*

(<http://cahiers.gutenberg.eu.org/>),

implique l'accord avec les conditions générales

d'utilisation (<http://cahiers.gutenberg.eu.org/legal.html>).

Toute utilisation commerciale ou impression systématique

est constitutive d'une infraction pénale. Toute copie ou impression

de ce fichier doit contenir la présente mention de copyright.



---

# La TEI simplifiée : une introduction au codage des textes électroniques en vue de leur échange

---

Lou BURNARD<sup>a</sup> et C.M. SPERBERG-MCQUEEN<sup>b</sup>

<sup>a</sup>*Oxford University  
Computing Services  
13 Banbury Road  
Oxford OX2 6NN, UK*

<sup>b</sup>*University of Illinois at Chicago:  
Computer Center  
Box 6998  
Chicago IL 60680, USA*

Traduction française de François ROLE\*  
*Ministère de l'Éducation nationale,  
de l'Enseignement supérieur et de la Recherche  
DISTNB – 1, rue Descartes, 75005 Paris  
email: role@distb.mesr.fr*

**Résumé.** Cet article est la traduction française de *TEI Lite : An Introduction to Text Encoding for Interchange*. La TEI Lite est une version simplifiée des « Recommandations » de la *Text Encoding Initiative* (TEI) qui s'adressent à tous ceux qui souhaitent échanger des informations stockées sous forme électronique. Elles mettent l'accent sur l'échange des données textuelles, notamment pour les sciences humaines et les études sur les textes littéraires, mais d'autres types de données comme les images et les sons sont également pris en compte.

**Abstract.** *This document is a French translation of the English document: "TEI Lite : An Introduction to Text Encoding for Interchange". These Text Encoding Initiative (TEI) Guidelines are addressed to anyone who wants to interchange information stored in an electronic form. They emphasize the interchange of textual information, but other forms of information such as images and sounds are also addressed. References of the genuine English documents are listed below.*

## Remerciements et copyright

Les *copyrights* de cet article sont ceux de la version originale. Cette traduction est publiée avec l'aimable autorisation des auteurs que le rédacteur de ces *Cahiers* tient à remercier ici.

---

\* Avec la collaboration et une mise en page de Jacques ANDRÉ, Irisa/Inria-Rennes. Le traducteur tient aussi à remercier Michael Jordan pour son aide.

Ce document est une introduction à la TEI *Text Encoding Initiative*, une DTD de SGML (voir section 22 «Références bibliographiques» pour une bibliographie plus complète sur SGML et sur la TEI), et en décrit un sous-ensemble connu sous le nom de *TEI Lite*. Ce que nous décrivons ici peut servir à coder une grande variété de phénomènes couramment rencontrés dans les textes pour en faciliter leur étude par des moyens électroniques et pour en faciliter l'échange entre humanistes utilisant des systèmes informatiques différents. Ce sous-ensemble est entièrement compatible avec la version intégrale de la TEI publiée à Chicago et à Oxford en mai 1994: «TEI P3, *Guidelines for Electronic Text Encoding and Interchange* ». Dans la suite de ce document, cette version intégrale est appelée «les *Recommandations*» ou parfois P3.

### Accès par Internet/WWW

- La version originale anglaise de ce document se trouve dans :  
<http://www-tei.uic.edu/orgs/tei/intros/teiu5.tei>  
<ftp://info.ox.ac.uk/pub/ota/TEI/doc/teiu5.tei>
  
- Une version HTML se trouve à :  
<http://www-tei.uic.edu/orgs/tei/intros/teiu5.html>  
<ftp://info.ox.ac.uk/ota/teilight/>
  
- La version intégrale de la DTD SGML décrite dans ces documents se trouve à :  
<http://www-tei.uic.edu/orgs/tei/p3/dtd/teilight.dtd>  
<ftp://ftp-tei.uic.edu/pub/tei/lite/teilight.dtd>  
<ftp://ota.ox.ac.uk/pub/ota/TEI/dtd/teilight.dtd>
  
- La présente version française de ce document sera bientôt dans  
<http://www.univ-rennes1.fr/pub/GUTenberg/publications>  
<http://distb.mesr.fr/norm/>

# TEI-Lite — sommaire

---

1	Introduction .....	27
2	Premiers exemples .....	30
	2.1 : Un court exemple .....	30
	2.2 : La TEI et le français .....	35
3	Structure d'un texte TEI .....	39
4	Codage du corps du texte .....	42
	4.1 : Éléments servant à marquer les divisions de texte .....	42
	4.2 : Têtes et fins de chapitre .....	46
	4.3 : Prose, vers et art dramatique .....	46
5	Numérotation des lignes et des pages .....	50
6	Marquage d'expressions mises en valeur .....	52
	6.1 : Changements des styles de caractères .....	52
	6.2 : Citations et éléments associés .....	54
	6.3 : Expressions ou mots étrangers .....	56
7	Notes .....	58
8	Références croisées et liens .....	60
	8.1 : Références croisées simples .....	60
	8.2 : Pointeurs étendus .....	63
	8.3 : Attributs de liaison .....	67
9	Interventions éditoriales .....	69
10	Omissions, effacements et ajouts .....	71
11	Noms, dates, chiffres et abréviations .....	75
	11.1 : Noms et chaînes de caractères de référence .....	75
	11.2 : Dates et heures .....	77
	11.3 : Nombres .....	78
	11.4 : Les abréviations et leur développement .....	79
	11.5 : Adresses .....	80
12	Listes .....	82

Suite: TSVP →

---

13	Citations bibliographiques .....	85
14	Tables .....	87
15	Figures et graphiques .....	89
16	Interprétation et analyse .....	92
	16.1 : Phrases orthographiques .....	92
	16.2 : Éléments d'interprétation polyvalents .....	93
17	Documentation technique .....	98
	17.1 : Éléments supplémentaires .....	98
	17.2 : Divisions générées automatiquement .....	101
	17.3 : Génération d'index .....	102
18	Jeux de caractères, signes diacritiques, etc. ...	104
19	Pièces liminaires et annexes .....	108
	19.1 : Pièces liminaires .....	108
	19.2 : Pièces annexes .....	112
20	La page de titre électronique .....	114
	20.1 : Description du fichier .....	115
	20.2 : Description relative au codage .....	120
	20.3 : Description de profil .....	125
	20.4 : Description de révision .....	127
21	Liste des éléments décrits .....	128
	21.1 : Attributs globaux .....	128
	21.2 : Éléments de la TEI Lite .....	128
22	Références bibliographiques .....	140
23	Index alphabétique .....	146

---

# 1 – Introduction

---

Les *Recommandations* de la TEI – *Text Encoding Initiative (TEI) Guidelines* – s’adressent à tous ceux qui souhaitent échanger des informations stockées sous forme électronique. Elles mettent l’accent sur l’échange des données textuelles mais d’autres types de données comme les images et les sons sont également pris en compte. Les *Recommandations* peuvent être appliquées aussi bien pour créer de nouvelles informations que pour échanger des informations existantes.

Les *Recommandations* fournissent le moyen de rendre explicites certaines caractéristiques d’un texte, de façon à faciliter le traitement de ce texte par des programmes informatiques pouvant s’exécuter sur des plate-formes différentes. Cette tâche d’explicitation est appelée «balisage» ou «codage». La représentation d’un texte sur un ordinateur met toujours en œuvre une forme de balisage ou une autre. La TEI tire son origine d’une part de l’anarchie qui règne dans la communauté scientifique en matière de format, et d’autre part du nombre croissant de traitements que les chercheurs opèrent sur les textes sous forme électronique.

Les *Recommandations* de la TEI s’appuient sur SGML (*Standard Generalized Markup Language*) pour définir leurs règles de codage<sup>1</sup>. SGML est une norme internationale de plus en plus utilisée par les entreprises spécialisées dans le traitement de l’information. Cette norme permet de définir les règles de codage en termes d’éléments, d’attributs et de règles régissant l’utilisation de ces éléments et attributs. L’utilisation que la TEI fait de SGML est ambitieuse, tant par sa complexité que par sa recherche d’universalité ; elle n’est cependant pas fondamentalement différente de tous les autres schémas de balisage fondés sur SGML. La preuve en est que tous les outils SGML généralistes sont capables de traiter des textes conformes à la TEI.

La TEI est soutenue par l’*Association for Computers and the Humanities*, l’*Association for Computational Linguistics* et l’*Association for Literary and Linguistic Computing*. Le projet a été en partie financé par le *National Endowment for the Humanities* américain, la DG XIII de la CEE, la fondation Andrew W. Mellon et le *Social Science and Humanities Research Council* du Canada. Les *Recommandations* ont été publiées en mai 1994, après six ans de travaux auxquels ont participé des chercheurs de toute nationalité et de toute discipline.

---

1. Voir la bibliographie sommaire donnée en section 22 «Références bibliographiques»

Au début de cette entreprise, les objectifs généraux de la TEI ont été définis par la résolution finale de la conférence préparatoire tenue au Vassar College de New York en novembre 1987. Cette résolution connue sous le nom de «Principes de Poughkeepsie» fut peu à peu précisée à travers une série de documents de travail. D'après ces documents les *Recommandations* devaient :

- être suffisamment précises pour représenter les propriétés des textes intéressants pour les chercheurs ;
- être simples, claires et concrètes ;
- être utilisables facilement par les chercheurs et ne pas nécessiter l'utilisation de logiciels spécifiques ;
- permettre une définition rigoureuse des textes en vue de traitements efficaces ;
- être modifiables par l'utilisateur ;
- respecter les normes en vigueur ou sur le point d'être adoptées.

Le monde de la recherche est large et divers. Pour que ces recommandations aient une large audience, il était important de s'assurer que :

1. les descriptions des caractéristiques fondamentales d'un texte puissent être facilement échangées ;
2. les descriptions spécialisées puissent être facilement ajoutées (ou supprimées) d'un texte ;
3. la même caractéristique puissent être encodée en parallèle de plusieurs façons ;
4. la richesse du balisage puisse être déterminée par l'utilisateur de la façon la plus simple possible ;
5. une documentation relative au texte et à la façon dont il a été codé soit fournie.

Ce document décrit un sous-ensemble utilisable des éléments définis par la TEI et fournit des recommandations en conformité avec les principes directeurs qui viennent d'être rappelés plus haut.

En effectuant un choix parmi les centaines d'éléments définis par la TEI, nous nous sommes efforcés d'identifier un ensemble initial contenant les éléments susceptibles de servir à n'importe quel utilisateur. L'expérience acquise en ne

---

travaillant qu'avec la «TEI Lite» sera d'une grande utilité pour comprendre la DTD TEI complète et pour identifier les parties de cette DTD qui sont nécessaires pour travailler sur un type de texte donné.

Nous pouvons résumer comme suit les objectifs que nous nous sommes assignés en définissant ce sous-ensemble :

- il doit comprendre la plupart des éléments qui appartiennent au *core tag set* de la TEI ; ce dernier contient en effet des éléments généraux susceptibles d'être utilisés pour tout type de texte et tout type de traitement ;
- il doit permettre de traiter correctement une large gamme de texte, avec un niveau de précision correspondant aux pratiques courantes des chercheurs ;
- il doit pouvoir être utilisé aussi bien pour produire de nouveaux documents que pour coder des documents existants ;
- il doit pouvoir fonctionner avec la plupart des logiciels SGML ;
- il doit être dérivé de la DTD TEI complète en utilisant les mécanismes d'extension décrits dans les *Recommandations* de la TEI ;
- il doit enfin respecter les contraintes énoncées ci-dessus de la façon la plus simple et la plus concise possible.

---

## 2 – Premiers exemples

---

*Ce chapitre a pour titre anglais « A Short Example » et correspond intégralement à la première section 2.1 « Un court exemple ». La section 2.2 présente un exemple en français qui n'apparaît pas dans l'original mais qu'il nous a paru intéressant d'ajouter pour montrer dès à présent quelques possibilités de la TEI en ce qui concerne les textes en français. Ce second exemple n'est donc qu'un complément au premier que le lecteur est supposé avoir étudié d'abord.*

{Note du traducteur}

### 2.1. Un court exemple

Nous commençons par un bref exemple, reproduit ci-dessous et destiné à illustrer ce qui se passe lorsqu'un passage de prose est saisi sur ordinateur par un opérateur n'ayant qu'une faible notion de ce que représente le balisage, ou du potentiel des textes électroniques. Dans un monde parfait, une telle sortie imprimée pourrait être générée par un lecteur optique très précis. Elle tente de rester fidèle à l'aspect du texte imprimé, en retenant les mêmes coupures de ligne que l'original, en insérant des blancs afin de représenter la disposition des titres originaux et des fins de page, et ainsi de suite. Lorsque certains caractères sont nécessaires mais indisponibles au clavier (tels que la lettre « a » accentuée dans le mot « faàl », ou le trait long), elle essaie de reproduire leur aspect.

#### CHAPTER 38

READER, I married him. A quiet wedding we had: he and I, the parson and clerk, were alone present. When we got back from church, I went into the kitchen of the manor-house, where Mary was cooking the dinner, and John cleaning the knives, and I said -- 'Mary, I have been married to Mr Rochester this morning.' The housekeeper and her husband were of that decent, phlegmatic order of people, to whom one may at any time safely communicate a remarkable piece of news without incurring the danger of having one's ears pierced by some shrill ejaculation and subsequently stunned by a torrent of wordy wonderment. Mary did look up, and she did stare at me; the ladle with which she

was basting a pair of chickens roasting at the fire, did for some three minutes hang suspended in air, and for the same space of time John's knives also had rest from the polishing process; but Mary, bending again over the roast, said only --

'Have you, miss? Well, for sure!'

A short time after she pursued, 'I seed you go out with the master, but I didn't know you were gone to church to be wed'; and she basted away. John, when I turned to him, was grinning from ear to ear. 'I telled Mary how it would be,' he said: 'I knew what Mr Edward' (John was an old servant, and had known his master when he was the cadet of the house, therefore he often gave him his Christian name) -- 'I knew what Mr Edward would do; and I was certain he would not wait long either: and h'es done right, for aught I know. I wish you joy, miss!' and he politely pulled his forelock.

'Thank you, John. Mr Rochester told me to give you and Mary this.'

I put into his hand a five-pound note. Without waiting to hear more, I left the kitchen. In passing the door of that sanctum some time after, I caught the words --

'Sh'ell happen do better for him nor ony o' t' grand ladies.' And again, 'If she ben't one o' th' handsomest, sh'es noan faa\l, and varry good-natured; and i' his een sh'es fair beautiful, onybody may see that.'

I wrote to Moor House and to Cambridge immediately, to say what I had done: fully explaining also why I had thus acted. Diana and

474

JANE EYRE

475

Mary approved the step unreservedly. Diana announced that she would just give me time to get over the honeymoon, and then she would come and see me.

'She had better not wait till then, Jane,' said Mr Rochester, when I read her letter to him: 'if she does, she will be too late, for our honey moon will shine our life long: its beams will only fade over your grave or mine.'

How St John received the news I don't know: he never answered the letter in which I communicated it: yet six months after he wrote to me, without, however, mentioning Mr Rochester's name or alluding to my marriage. His letter was then calm, and though very serious, kind. He has maintained a regular, though not very frequent correspond ence ever since: he hopes I am happy, and trusts I am not of those who live without God in the world, and only mind earthly things.

Cette transcription souffre d'un certain nombre de limitations :

- les numéros de page ainsi que les titres courants sont mêlés au texte de façon si inextricable qu'un logiciel aurait du mal à les démêler ;
- aucune distinction n'est faite entre l'apostrophe et les guillemets simples, d'où la difficulté de déterminer de façon sûre quels passages sont en discours direct ;
- la conservation des césures de l'original implique que des programmes de recherche peu élaborés ne trouveront pas les mots coupés ;
- la lettre accentuée dans « faàl » et le trait long ont été saisis selon des conventions ad hoc et ne seront donc traités correctement que si le transcritteur prend le soin de les mentionner dans la documentation ;
- les divisions de paragraphe sont marquées uniquement par l'emploi d'espaces blancs, et des fins de ligne forcées ont été introduites au bout de chaque ligne. Par conséquent, tout changement de taille de police lors de l'impression du texte engendrera des problèmes de mise en page.

Nous présentons maintenant le même passage avec un codage conforme aux recommandations de la TEI. Comme nous le verrons, ce codage pourrait être étendu de bien des façons, mais au minimum, le codage recommandé par la TEI permet d'effectuer les distinctions suivantes :

- les divisions de paragraphe sont maintenant marquées de façon explicite ;
- les apostrophes sont distinguées des guillemets ;
- les références d'entité sont employées pour la lettre accentuée et le trait long ;
- les divisions de page ont été marquées par un seul élément `<pb>` vide ;
- afin de simplifier la recherche et le traitement, la division en lignes de l'original n'a pas été retenue et des mots coupés par accident typographique en fin de ligne se trouvent réassemblés sans commentaire ; si la division en lignes de l'original avait un intérêt particulier, comme ce pourrait être le cas pour une édition rare, elle pourrait être notée facilement, bien que ce ne soit pas le cas ici ;
- afin de faciliter la correction des épreuves, une nouvelle ligne a été insérée en début de chaque paragraphe, mais l'alinéa n'a pas été conservé.

```
<pb n='474'>  
<div1 type=chapter n='38'>
```

<p>Reader, I married him. A quiet wedding we had: he and I, the parson and clerk, were alone present. When we got back from church, I went into the kitchen of the manor-house, where Mary was cooking the dinner, and John cleaning the knives, and I said &dash;

<p><q>Mary, I have been married to Mr Rochester this morning.</q> The housekeeper and her husband were of that decent, phlegmatic order of people, to whom one may at any time safely communicate a remarkable piece of news without incurring the danger of having one's ears pierced by some shrill ejaculation and subsequently stunned by a torrent of wordy wonderment. Mary did look up, and she did stare at me; the ladle with which she was basting a pair of chickens roasting at the fire, did for some three minutes hang suspended in air, and for the same space of time John's knives also had rest from the polishing process; but Mary, bending again over the roast, said only &dash;

<p><q>Have you, miss? Well, for sure!</q>

<p>A short time after she pursued, <q>I seed you go out with the master, but I didn't know you were gone to church to be wed</q>; and she basted away. John, when I turned to him, was grinning from ear to ear.

<q>I telled Mary how it would

be,</q> he said: <q>I knew what Mr Edward</q> (John was an old servant, and had known his master when he was the cadet of the house, therefore he often gave him his Christian name) &dash; <q>I knew what Mr Edward would do; and I was certain he would not wait long either: and h s done right, for aught I know. I wish you joy, miss!</q> and he politely pulled his forelock.

<p><q>Thank you, John. Mr Rochester told me to give you and Mary this.</q>

<p>I put into his hand a five-pound note. Without waiting to hear more, I left the kitchen. In passing the door of that sanctum some time after, I caught the words &dash;

<p><q>She'll happen do better for him nor any o' t' grand ladies.</q> And again, <q>If she ben't one o' th'

handsomest, sh'es noan fa&agrave;l, and varry good-natured;  
and i' his een sh'es fair beautiful, onybody may see  
that.</q>

<p>I wrote to Moor House and to Cambridge immediately, to  
say what I had done: fully explaining also why I had thus  
acted. Diana and <pb n='475'> Mary approved the step  
unreservedly. Diana announced that she would just give me  
time to get over the honeymoon, and then she would come and  
see me.

<p><q>She had better not wait till then, Jane,</q> said Mr  
Rochester, when I read her letter to him; <q>if she does,  
she will be too late, for our honeymoon will shine our life  
long: its beams will only fade over your grave or mine.</q>

<p>How St John received the news I don't know: he never  
answered the letter in which I communicated it: yet six  
months after he wrote to me, without, however, mentioning Mr  
Rochester's name or alluding to my marriage. His letter was  
then calm, and though very serious, kind. He has maintained  
a regular, though not very frequent correspondence ever  
since: he hopes I am happy, and trusts I am not of those who  
live without God in the world, and only mind earthly things.

La décision de se concentrer sur le texte de Brontë, plutôt que sur la façon dont il a été imprimé dans cette édition spécifique, témoigne d'un problème de codage fondamental: celui de la sélectivité. Un codage rend explicite seulement celles des caractéristiques du texte qui ont une importance pour l'encodeur. Il n'est pas difficile d'imaginer comment on pourrait étendre facilement le codage d'un passage même aussi court que celui-ci. Par exemple :

- une forme en anglais officiel pourrait être fournie pour les passages en dialecte ;
- des gloses ou des commentaires sous forme de notes de pied de page pourraient être ajoutés ;
- des pointeurs pourraient être ajoutés afin de lier certaines parties du texte entre elles ;
- des noms propres de divers types pourraient être distingués par rapport au texte environnant ;
- des données bibliographiques détaillées concernant la provenance du texte ainsi que son contexte pourraient être fournies dans un en-tête ;

- une analyse linguistique du passage en phrases, clauses, mots, etc. pourrait être fournie, chaque unité étant associée à des codes de catégorie appropriés ;
- le texte pourrait être segmenté dans des unités de discours ou de narration ;
- l’interprétation ou l’analyse systématique du texte pourraient être incluses dans le codage, avec des associations ou des alignements potentiellement complexes entre le texte et l’analyse, ou entre le texte et une ou plusieurs traductions de ce dernier ;
- certains passages dans le texte pourraient être liés à des images ou à du son stockés sur d’autres supports.

La méthode recommandée par la TEI pour faire tout ceci est décrite dans le reste de ce document. La TEI dans son ensemble fournit également maintes possibilités, dont nous ne citons que quelques exemples :

- l’analyse détaillée des composants de noms ;
- des méta-informations détaillées fournissant des données quant aux origines du texte et aux sujets qu’il aborde ;
- des informations sur l’historique des impressions, ou des variantes manuscrites telles qu’elles apparaissent dans une suite de versions particulières du texte.

Pour connaître les recommandations applicables dans ces différents cas, et pour découvrir de nombreuses autres possibilités, il convient de se reporter au texte complet des recommandations de la TEI.

## 2.2. La TEI et le français

Le premier exemple a mis l’accent sur la typologie du texte. Nous complétons cette introduction par un exemple en langue française montrant quelques possibilités de la TEI liées à notre langue. Cette fois partons du document original<sup>1</sup> de la figure 1.

Si l’on devait coder ce texte, on aurait quelques problèmes plus évidents en français qu’en anglais :

- l’emploi de lettres accentuées est très important ;

---

1. Il s’agit du premier chapitre de *Les belles images* de Simone de Beauvoir, Gallimard, 1960.

«C'est un mois d'octobre... exceptionnel», dit Gisèle Dufrène ; ils acquiescent, ils sourient, une chaleur d'été tombe du ciel gris-bleu — Qu'est-ce que les autres ont que je n'ai pas? — ils caressent leurs regards à l'image parfaite qu'ont reproduite *Plaisir de France* et *Votre Maison* : la ferme achetée pour une bouchée de pain — enfin, disons, de pain brioché — et aménagée par Jean-Charles au prix d'une tonne de caviar. (“ je n'en suis pas à un million près ”, a dit Gilbert), les roses contre les murs de pierre, les chrysanthèmes, les asters, les dahlias “ les plus beaux de toute l'Ile-de-France ”, dit Dominique ; le paravent et les fauteuils bleus et violet — c'est d'une audace! — tranchent sur le vert de la pelouse, la glace tinte dans les verres, Houdan baise la main de Dominique, très mince dans son pantalon noir et son chemisier éclatant, les cheveux pâles, mi-blonds, mi-blancs, de dos on lui donnerait trente ans. ...

FIGURE 1 – Un exemple de texte français à coder

- mais si l'on saisit en utilisant un codage comme ISO 8859-1 (Latin1), on sera quand-même obligé de coder certains caractères en utilisant des entités SGML, par exemple «œ» ne fait pas partie de ce codage et devra donc être codé &oumlig; ;
- alors que les textes anglais peuvent provoquer des confusions entre l'apostrophe et le guillemet simple, les textes français sont souvent composés avec toute une artillerie de guillemets divers (ce texte de Simone de Beauvoir utilise les guillemets français «et» et ceux anglais “ et ”);
- la typographie française nécessite divers types d'espaces (par exemple avant le point-virgule on met une espace fine, avant le deux-points une espace justifiante, etc.) ou de tirets (par exemple « — ») qu'une saisie simple aurait du mal à distinguer ;

Un texte TEI peut tenir compte de tous ces éléments: ce texte pourrait être codé comme suit :

```
<p> <q rend=frdquo> C'est un mois d'octobre &hellip; exceptionnel
</q>, dit Gis&egrave;le Dufr&egrave;ne;
ils acquiescent, ils sourient, une chaleur d'&eacute;t&eacute;
tombe du ciel gris-bleu&mdash;
<q>Qu'est-ce que les autres ont que je n'ai pas? </q>
&mdash; ils caressent leurs regards &aacute;l'image parfaite
```

qu'ont reproduite <title>Plaisir de France</title> et <title>Votre Maison</title>: la ferme achetée pour une bouche de pain – enfin, disons, de pain brioche – et amenée par Jean-Charles au prix d'une tonne de caviar. (<q rend=endq>je n'en suis pas un million près</q>, a dit Gilbert), les roses contre les murs de pierre, les chrysanthèmes, les asters, les dahlias <q rend=endq> les plus beaux de toute l'Ile-de-France</q>, dit Dominique; le paravent et les fauteuils bleus et violet – <q>c'est d'une audace!</q> – tranchent sur le vert de la pelouse, la glace tinte dans les verres, Houdan baise la main de Dominique, trépane mince dans son pantalon noir et son chemisier éclatant, les cheveux pécunies, mi-blonds, mi-blancs, de dos on lui donnerait trente ans.

...

Dans ce texte, on a trois types de citations: celles entre guillemets français, celles entre guillemets anglais et celles sans guillemets (par exemple <q>c'est d'une audace!</q>). Pour les deux premiers types, on a délimité la citation par le couple de balises <q> et </q> et au moyen de l'attribut *rend* on a mémorisé la nature des guillemets utilisés dans le texte imprimé (frdquo pour français double quote et enddquo pour anglais double quote) de façon à pouvoir les restituer si besoin est. De même, nous supposons que c'est lors de l'édition/impression du document que l'on doit se soucier de la présence d'espaces avant les points-virgules, deux-points, etc., pas au moment de la saisie ni du traitement. Aussi codons-nous «dit Dominique;» et non «dit Dominique\_;».

Dans les documents SGML, les caractères accentués doivent être notés à l'aide d'«entités» (voir section 18 «Jeux de caractères, signes diacritiques, etc.») qui prévoient, par exemple, que «é» soit codé «&eacute;». Le but de ce mécanisme est d'obtenir un document composé uniquement de caractères appartenant au jeu ISO 646-IRV (ASCII) de base, donc très facile à transmettre.

Cependant, pour des raisons de lisibilité, nous avons employé les caractères ISO 8859-1 (Latin-1) pour coder les lettres des exemples utilisant le français dans la suite de cette traduction. En réalité, il est techniquement possible d'échanger des documents SGML de cette façon, mais au prix d'une modification de la «déclaration SGML» qui ne peut être effectuée que par un utilisateur expérimenté et qui rend le document moins interchangeable.

Comme il a été dit plus haut<sup>2</sup>, travailler directement sur le codage TEI n'est pas chose naturelle. On écrira donc, dans les exemples qui suivent, des textes

2. Voir l'article de Jacques André sur les balises, ci-dessus.

lisibles (quitte à imaginer qu'il faille passer un programme de remplacement automatique sur ces textes), comme :

la ferme achetée pour une bouchée de pain &mdash;  
enfin, disons, de pain brioché &mdash; et aménagée par  
Jean-Charles au prix d'une tonne de caviar.

### *Balises françaises*

Un autre aspect de l'utilisation du français pour la TEI est celui de la francisation des balises et attributs.

La TEI prévoit des mécanismes pour paramétrer le nom des balises et donc, si on le souhaite, utiliser des balises en français de son choix. Il y a là un compromis entre le confort d'écriture et de lisibilité pour le créateur du document et l'échangeabilité: les textes avec des balises « localisées » ne peuvent évidemment plus être immédiatement interprétés par d'autres. À noter que les tentatives pour franciser les langages de programmation (Basic, Pascal, etc.) n'ont jamais eu de succès, sans doute pour la raison ci-dessus... Mais enfin, avec la TEI c'est possible si on le souhaite.

Notre attitude, ici, a été de garder le nom anglais des balises dans le codage SGML des exemples, mais de proposer en cas d'ambiguïté, un équivalent français dans le texte de l'article.

Enfin, en ce qui concerne les valeurs d'attribut, nous avons parfois utilisé des chaînes « françaises » à la place des chaînes « anglaises » suggérées dans l'original, du moins lorsque la DTD le permettait (par exemple quand le type de l'attribut était CDATA).

L'ensemble de ces noms, anglais et français, se retrouve dans l'index final.

## 3 – Structure d'un texte TEI

Tout texte conforme à la TEI comporte

1. un en-tête TEI (balisé comme un élément `<teiHeader>`,
2. la transcription du texte lui-même (balisé comme un élément `<text>`).

L'en-tête TEI contient des informations analogues à celles que l'on trouve sur la page de titre d'un texte imprimé. Il contient jusqu'à quatre parties :

1. une description bibliographique du texte électronique ;
2. une description de la manière dont il a été codé ;
3. une description non-bibliographique du texte (le «profil» du texte) ;
4. un historique de révision.

L'en-tête est décrit avec plus de détails dans la section 20 «Page de titre électronique».

Un texte TEI peut être unitaire (une œuvre isolée) ou composite (un recueil d'œuvres, comme une anthologie). Dans un cas comme dans l'autre, le texte peut éventuellement comporter des pièces liminaires ou des annexes. Entre les deux se trouve le corps du texte qui, dans le cas d'un texte composite, peut comporter des groupes, chacun contenant encore des groupes ou des textes.

Un texte unitaire sera codé à peu près comme suit<sup>1</sup> :

```
<TEI.2>
  <teiHeader> [informations contenues dans l'en-tête TEI ]
  </teiHeader>,
  <text>
    <front>[ textes préliminaires... ] </front>,
    <body>[ corps du texte... ] </body>
    <back> [annexes... ] </back>
  </text>
</TEI.2>
```

1. Dès cet exemple, on remarquera que les diverses lignes de ce texte-source sont «indentées», c'est-à-dire qu'elles sont en retrait les unes par rapport aux autres, ce qui exhibe la structure sous-jacente ; ce n'est bien sûr pas une obligation, mais une bonne façon de coder les textes-sources ! {NdT}

Un texte composite peut également comporter des pièces liminaires et des annexes. Entre les deux, on trouve éventuellement un ou plusieurs groupes de textes, chaque groupe ayant éventuellement ses propres pièces liminaires et annexes. Un texte composite sera donc codé à peu près comme suit :

```
<TEI.2>
  <teiHeader> [ informations contenues dans l'en-tête
                du texte composite ] </teiHeader>
  <text>
    <front> [ textes liminaires du texte composite ]
    </front>
    <group>
      <text>
        <front>[textes liminaires du premier texte ]
        </front>
        <body> [corps du premier texte ] </body>
        <back>[ annexes du premier texte ] </back>
      </text>
      <text>
        <front>[ textes liminaires du deuxième texte ]
        </front>
        <body> [ corps du deuxième texte ] </body>
        <back> [ annexes du deuxième texte ] </back>
      </text>
      [ autres textes ou groupes de textes ici ]
    </group>
    <back> [ annexes du texte composite ] </back>
  </text>
</TEI.2>
```

Il est également possible de définir un ensemble de textes TEI, chacun avec son propre en-tête. Un tel recueil est appelé «corpus TEI», et peut lui-même avoir un en-tête :

```
<teiCorpus>
  <teiHeader> [ information contenue dans l'en-tête du corpus ]
  </teiHeader>
  <TEI.2>
    <teiHeader>[ information contenue dans l'en-tête
                du premier texte ] </teiHeader>
    <text> [ premier texte du corpus ] </text>
  </TEI.2>
  <TEI.2>
    <teiHeader>[ information contenue dans l'en-tête du
```

---

```
                deuxième  texte ] </teiHeader>
    <text> [deuxième texte du corpus ] </text>
</TEI.2>
</TEICorpus>
```

Cependant, il n'est pas possible de créer un ensemble de corpus, c'est-à-dire plusieurs éléments *<teiCorpus>* combinés les uns avec les autres et traités comme un objet unique. Ceci représente une des limites de la version actuelle des recommandations de la TEI.

Dans la suite de ce document, nous discutons surtout de structures de texte simples. Dans chaque cas, la présentation consiste en une courte liste des éléments TEI pertinents, avec une brève définition de chacun, suivies de la définition des attributs spécifiques à cet élément. Dans la plupart des cas, de brefs exemples sont également fournis.

## 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  $\langle div1 \rangle$  sont nécessaires, un  $\langle div1 \rangle$  peut être divisé en éléments  $\langle div2 \rangle$ , et un  $\langle div2 \rangle$  en éléments encore plus petits  $\langle div3 \rangle$ , etc., jusqu’au niveau  $\langle div7 \rangle$ . 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  $\langle div8 \rangle$ , etc., soit d’employer un élément  $\langle div \rangle$  non numéroté : un  $\langle div \rangle$  peut être subdivisé en de plus petits éléments  $\langle div \rangle$  (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  $\langle div \rangle$ ,  $\langle div1 \rangle$ ,  $\langle div2 \rangle$ , etc. dans un texte est supposée s’appliquer pour tous les  $\langle div \rangle$ ,  $\langle div1 \rangle$   $\langle divn \rangle$  ultérieurs dans le même  $\langle body \rangle$ . 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 mnémorique 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 transcritteur 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'oeil 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>

```

## 5 – Numérotation des lignes et des pages

---

Les coupures de lignes et de pages peuvent être marquées par les éléments vides suivants :

- `<pb>` marque la limite entre une page d'un texte et la suivante, dans un système de référence normalisé ;
- `<lb>` marque le début d'une nouvelle ligne (typographique) dans une édition ou version donnée d'un texte.

Ces éléments marquent un endroit précis dans le texte et non une étendue de texte. L'attribut global *n* devrait être employé pour fournir le numéro de la page ou de la ligne commençant à la balise. En outre, ces deux éléments partagent l'attribut suivant :

- ed* indique l'édition ou la version du texte dans laquelle la fin de page est située à cet endroit.

Lorsqu'on travaille à partir d'un original paginé, il est souvent utile d'enregistrer sa pagination, ne serait-ce que pour faciliter les corrections ultérieures. L'enregistrement des coupures de ligne peut être utile pour la même raison ; le traitement de la division (césure) des mots en fin de ligne dans les originaux imprimés nécessitera donc une certaine considération.

Si la pagination est indiquée pour plus d'une édition, chaque édition peut être identifiée au moyen de l'attribut *ed*, en fournissant autant de balises que nécessaire. Par exemple, dans le passage suivant nous indiquons l'emplacement des fins de page dans deux éditions différentes (ED1 et ED2) :

```
<p>I wrote to Moor House and to Cambridge immediately, to
say what I had done: fully explaining also why I had thus
acted. Diana and <pb ed=ED1 n='475'> Mary approved the
step unreservedly. Diana announced that she would
<pb ed=ED2 n='485'>just give me time to get over the
honeymoon, and then she would come and see me.
```

---

Les éléments `<pb>` et `<lb>` sont des cas particuliers d'une classe générale d'éléments de type « jalons » qui servent à marquer des points de référence dans un texte. La *TEI Lite* inclut également un élément générique `<milestone>` qui n'est pas limité aux cas spéciaux mais peut marquer toutes sortes de point de référence : par exemple une fin de colonne, le début d'un nouveau type de section qui n'est pas balisé autrement, etc. Cet élément a la description et les attributs suivants :

`<milestone>`

marque la limite entre des sections d'un texte, comme indiqué par des changements dans un système de référence standard. Les attributs sont :

`ed` indique l'édition ou la version à laquelle le jalon s'applique ;

`unit` indique le type de section qui change à partir de cet endroit.

Les noms employés pour désigner les types d'unité ainsi que les éditions désignés par ces attributs `ed` et `unit` peuvent être choisis librement, mais il convient de les expliciter dans l'en-tête.

L'élément `<milestone>` peut être employé pour remplacer les autres attributs, ou ces derniers peuvent être employés ensemble ; toutefois, il convient d'éviter de les mélanger arbitrairement.

## 6 – Marquage d’expressions mises en valeur

---

### 6.1. Changements des styles de caractères ou alternances typographiques

Par expressions ou mots mis en valeur, nous entendons ceux qui sont distingués visuellement du reste du texte, notamment par un changement de police (ou fonte), de style d’écriture ou de couleur d’encre, dans le but d’attirer l’attention du lecteur.

L’attribut global *rend* peut être attaché à tout élément et employé autant que nécessaire pour préciser les détails de la mise en valeur retenue. Par exemple, un titre en gras pourrait être balisé par `head rend='gras'` et un titre en italique par `head rend='italique'`.

Il n’est pas toujours possible ni souhaitable d’interpréter les raisons de telles alternances typographiques dans un texte. Dans ces cas, l’élément `<hi>` peut être employé pour marquer une séquence de texte mise en évidence, sans formuler d’interprétation quant à son statut.

`<hi>` marque un mot ou une expression comme étant graphiquement distinct du texte environnant, sans donner d’indication quant aux raisons de cette mise en valeur.

Dans l’exemple suivant, l’emploi d’un type de caractère distinct pour le sous-titre et pour le nom inclus est enregistré mais ne fait l’objet d’aucune interprétation :

```
<hi rend=gothic>And this Inventure further witnesseth</hi>
that the said <hi rend=italic>Walter Shandy</hi>, merchant,
in consideration of the said intended marriage...
```

Alternativement, là où les raisons de la mise en valeur peuvent être identifiées de façon sûre, d’autres éléments plus spécifiques sont disponibles :

- 
- <*emph*> marque des mots ou des expressions qui sont accentués ou mis en valeur pour des effets rhétoriques ou linguistiques ;
- <*foreign*> identifie un mot ou une expression comme appartenant à une langue autre que celle du texte qui l'entoure ;
- <*mentioned*> marque des mots ou des expressions mentionnés, mais non employés ;
- <*term*> contient un mot unique, plusieurs mots ou une désignation symbolique considérée comme un terme technique ;
- <*title*> contient le titre d'une œuvre, que ce soit un article, un livre, un journal ou une série, y compris tous les sous-titres ou titres alternatifs ; les attributs possibles sont :
- level* indique s'il s'agit d'un titre d'article, de livre, de journal, de série ou d'un texte non publié : les valeurs autorisées sont :  
 m pour le titre monographique (livre, collection, ou autre ouvrage publié comme une œuvre distincte, comportant des volumes uniques de travaux multivolumes) ;  
 s (titre de série) ;  
 j (titre de journal) ;  
 u pour le titre d'un document non publié (y compris des dissertations et des thèses, à moins qu'elles n'aient été publiées par un éditeur commercial) ;  
 a pour le titre analytique (article, poème ou autre œuvre publiée comme une partie d'un ouvrage plus important).
- type* classe le titre selon une typologie adéquate ; des valeurs possibles sont :  
 abbreviated, main, subordinate (pour des sous-titres et des titres de parties),  
 parallel (pour des titres alternatifs, souvent dans une autre langue, sous lesquels l'œuvre est aussi connue).

Dans certains textes, certains éléments (notamment des citations et des gloses) peuvent être mis en évidence soit par une mise en valeur typographique, soit par l'utilisation de guillemets. Dans l'un ou l'autre cas, les éléments <*q*> et <*gloss*> (comme nous le verrons dans la section suivante) doivent être employés. Si la présentation physique doit être enregistrée, il convient d'utiliser l'attribut global *rend*.

Pour illustrer ces types d'éléments, examinons la phrase suivante :

On the one hand the *Nibelungenlied* is associated with the new rise of romance of twelfth-century France, the romans d'antiquité, the romances of Chrétien de Troyes, and the German adaptations of these works by Heinrich van Veldeke, Hartmann von Aue, and Wolfram von Eschenbach.

Si l'on choisit d'interpréter les raisons de la mise en valeur typographique, la phrase pourrait avoir l'aspect suivant :

On the one hand the <title>Nibelungenlied</title> is associated with the new rise of romance of twelfth-century France, the <foreign>romans d'antiquité</foreign>, the romances of Chrétien de Troyes, ...

Si l'on choisit de décrire uniquement l'aspect de l'original, le résultat pourrait être le suivant :

On the one hand the <hi rend=italic>Nibelungenlied</hi> is associated with the new rise of romance of twelfth-century France, the <hi rend=italic>romans d'antiquité</hi>, the romances of Chrétien de Troyes, ...

## 6.2. Citations et éléments associés

Tout comme les changements de famille de caractères, les guillemets sont conventionnellement employés pour signaler certains éléments apparaissant dans un texte, le cas le plus fréquent étant la citation. Toutefois, dans la mesure du possible, nous recommandons que l'élément logique sous-jacent soit balisé, plutôt que de se contenter d'enregistrer que des guillemets apparaissent dans le texte, ceci en employant les éléments suivants :

<q> contient une citation ou une citation apparente – la représentation d'une parole ou d'une pensée, balisée pour bien indiquer qu'il s'agit d'une citation (qu'il s'agisse réellement ou non d'une citation) ; dans un texte narratif, les mots sont habituellement ceux d'un personnage ou d'un locuteur ; dans des dictionnaires, <q> peut être employé pour marquer des exemples d'usage arrangés ou réels ; parmi les attributs possibles citons :

- type* peut être employé pour indiquer si le passage cité est prononcé ou simplement pensé, ou pour le caractériser plus finement : des valeurs possibles sont :  
spoken (pour la représentation de discours directs, marquée habituellement par des guillemets) et  
thought (pour la représentation de pensées, par exemple monologue intérieur) ;
- who* identifie le locuteur dans le cas d'un passage de discours direct ;

<mentioned>

marque des mots ou des expressions mentionnés, mais non employés ;

<soCalled>

contient un mot ou une expression vis-à-vis desquels l'auteur ou le narrateur décline toute responsabilité, ce qui est en général indiqué par l'utilisation des guillemets ou de l'italique ;

<gloss>

marque un mot ou une expression qui fournit une glose ou une définition pour un autre mot ou une autre expression ; parmi les attributs possibles citons :

*target* identifie l'expression ou le mot associé.

Voici un exemple simple de citation :

```
Few dictionary makers are likely to forget
Dr. Johnson's description of the lexicographer
as <q>a harmless drudge.</q>
```

Pour noter la façon dont une citation a été imprimée (par exemple, sur la même ligne ou dans un pavé typographique distinct), l'attribut *rend* devrait être employé. Ceci sert également pour indiquer le type de guillemets employés.

Le discours direct interrompu par un narrateur peut être représenté simplement en terminant la citation et en la recommençant après l'interruption, comme dans l'exemple suivant :

```
<p><q>Who-e debel you?</q> &mdash; he at last said &mdash;
<q>you no speak-e, damme, I kill-e.</q> And so saying, the
lighted tomahawk began flourishing about me in the dark.
```

S'il est nécessaire de faire comprendre que les deux éléments `<q>` se rapportent au même discours, les attributs de liaison *next* et *prev* peuvent être utilisés, comme cela est décrit dans la section 8 «Attributs de liaison».

Les citations peuvent être accompagnées d'une référence à la source ou au locuteur, au moyen de l'attribut *who*, que la source soit donnée dans le texte ou non, comme dans l'exemple suivant :

```
<q who=Wilson>Spaulding, he came down into the office just  
this day eight weeks with this very paper in his hand, and he  
says:&mdash;<q who=Spaulding>I wish to the Lord, Mr. Wilson,  
that I was a red-headed man.</q></q>
```

Cet exemple montre aussi comment les citations peuvent être imbriquées au sein d'autres citations : un locuteur (Wilson) cite un autre locuteur (Spaulding).

Le créateur du texte électronique doit décider si les guillemets seront remplacés par les balises ou si les balises seront ajoutées et les guillemets retenus. Si les guillemets sont ôtés du texte, l'attribut *rend* peut être employé pour enregistrer la façon dont ils étaient rendus dans la copie.

Comme dans le cas des mises en valeur, il n'est pas toujours possible ni souhaitable d'interpréter la fonction des guillemets dans un texte de cette façon. Dans de tels cas, la balise `<hi rend=quoted>` pourrait être employée pour marquer un texte cité, sans formuler d'interprétation quant à son statut.

### 6.3. Expressions ou mots étrangers

Les mots ou expressions n'appartenant pas à la langue principale du texte, peuvent être balisés comme tels de deux façons. Si le mot ou l'expression est déjà balisé pour une raison quelconque, l'élément indiqué devrait comporter une valeur pour l'attribut global *lang* indiquant la langue utilisée. Lorsqu'il n'y a aucun élément applicable, l'élément `<foreign>` peut être employé en utilisant de nouveau l'attribut *lang*. Par exemple :

```
John has real <foreign lang=fra>savoir-faire</foreign>.
```

```
Have you read <title lang=deu>Die Dreigroschenoper</title>?
```

```
<mentioned lang=fra>Savoir-faire</mentioned> is French  
for know-how.
```

---

The court issued a writ of <term lang=lat>mandamus</term>.

Savoir-faire se dit <mentioned lang=eng>know-how</mentioned>  
en anglais!

Dans ce dernier exemple, nous supposons que la langue principale du texte est le français.

Comme le montrent ces exemples, l'élément <foreign> ne devrait pas être employé pour baliser des mots étrangers si un autre élément plus spécifique tel que <title>, <mentioned> ou <terme> peut s'appliquer. L'attribut global *lang* peut être attaché à tout élément, pour montrer qu'il est écrit dans une autre langue que celle du texte dans lequel il se trouve.

## 7 – Notes

---

Toutes les notes, qu'elles soient imprimées comme des notes de bas de pages, des annotations de fin d'ouvrage ou de chapitre, des notes marginales, ou sous une autre forme, devraient être marquées au moyen du même élément :

<note> contient une note ou une annotation. Parmi les attributs citons :

<i>type</i>	décrit le type de la note ;
<i>resp</i>	indique le responsable de l'annotation: l'auteur, le rédacteur, le traducteur, etc. ; la valeur pourrait être <i>author</i> , <i>editor</i> , etc., ou les initiales de la personne qui a ajouté l'annotation ;
<i>place</i>	indique l'emplacement de l'annotation dans le texte source ; des valeurs possibles sont : <i>inline</i> (sur la ligne), <i>interlinear</i> (interlinéaire), <i>left</i> (gauche), <i>right</i> (droit), <i>foot</i> (pied de page) et <i>end</i> (fin), valeurs qui indiquent respectivement que les notes paraissent comme des paragraphes marqués dans le corps du texte, entre les lignes, dans la marge gauche ou droite, en bas de la page, ou à la fin du chapitre ou du volume ;
<i>target</i>	indique le point d'attachement d'une note, ou le début du passage auquel la note est attachée ;
<i>targetEnd</i>	indique la fin du passage auquel la note est attachée, si la note n'est pas insérée dans le texte à cet endroit ;
<i>anchored</i>	indique si la copie donne l'endroit exact de référence pour la note.

Dans la mesure du possible, le corps d'une note devrait être inséré dans le texte au même endroit que celui de son identifiant ou marqueur. Ceci peut ne pas être possible, par exemple dans le cas de notes marginales qu'il peut être difficile d'ancrer à un emplacement exact. Pour la simplicité, il convient de placer les notes marginales avant le paragraphe (ou autre élément) correspondant. Les notes peuvent également être placées dans une division séparée

du texte (comme les notes de fin le sont dans des livres imprimés) et liées à la portion pertinente du texte au moyen de leur attribut *target*.

Si nécessaire, l'attribut *n* peut être employé pour donner le numéro ou l'identifiant d'une note. L'attribut *resp* devrait être employé de façon systématique pour distinguer les notes d'auteur et d'éditeur, si l'œuvre comporte ces deux sortes de note ; autrement, l'en-tête TEI devrait indiquer leur type.

Voir le second exemple

Voici quelques exemples :

```

Collections are ensembles of distinct
entities or objects of any sort.
<note place=foot n=1>
We explain below why we use the uncommon term
<mentioned>collection</mentioned>
instead of the expected
<mentioned>set</mentioned>.
Our usage corresponds to the <mentioned>aggregate</mentioned>
of many mathematical writings and to the sense of
<mentioned>class</mentioned> found
in older logical writings.
</note>
The elements ...

<p>
<note place=margin>Voir le second exemple</note>
Si nécessaire, l'attribut ...
devrait indiquer leur type.
</p>

```

---

## 8 – Références croisées et liens

---

Il est possible de coder les références croisées, ou des liens explicites entre des points différents d'un même document SGML, au moyen des éléments décrits en section 8.1 «Références croisées simples». Les références ou liens vers des éléments situés dans un autre document SGML, ou vers des parties de documents non-SGML, peuvent être codés au moyen des pointeurs étendus TEI décrits en section 8.2 «Pointeurs/indicateurs étendus». Les liens implicites (telle l'association entre deux textes parallèles ou celle entre un texte et son interprétation) peuvent être codés au moyen des attributs de liaison expliqués en section 8.3 «Attributs de liaison».

### 8.1. Références croisées simples

Une référence croisée reliant deux endroits situés dans le même document peut être codée au moyen de l'un ou l'autre des éléments suivants :

- <ref>* référence à un autre emplacement dans le document courant, en termes d'un ou plusieurs éléments identifiables, le tout pouvant éventuellement être modifié par un commentaire ou du texte supplémentaire ;
- <ptr>* pointeur vers un autre emplacement dans le document courant, en termes d'un ou plusieurs éléments identifiables.

Ces éléments partagent les attributs suivants :

- target* indique la destination du pointeur en termes d'un ou plusieurs identifiables SGML ;
- type* catégorise le pointeur d'une manière quelconque, au moyen de tout ensemble de catégories approprié ;
- targType* indique le ou les types d'éléments désignés par ce pointeur ;
- crDate* indique la date de création du pointeur ;
- resp* indique le créateur du pointeur.

La différence entre ces deux éléments est que `<ptr>` est un élément vide qui marque simplement un endroit à partir duquel un lien est à créer, tandis que `<ref>` peut contenir également du texte – typiquement le texte de la référence croisée elle-même. L'élément `<ptr>` serait plutôt employé pour une référence croisée signalée par des dispositifs non-verbaux tels qu'un symbole ou une icône, ou dans un texte électronique par un bouton. Il sert également dans des systèmes de production de document, où le logiciel de formatage a la possibilité de générer la forme verbale correcte de la référence croisée.

Les deux formes présentées ci-après, par exemple, sont logiquement équivalentes (en supposant que nous avons documenté ailleurs la forme verbale exacte des références croisées représentées par les éléments `<ptr>`):

Voir notamment `<ref target=SEC12>la section 12, page 34</ref>`.

Voir notamment `<ptr target=SEC12>`.

la valeur de l'attribut *target* doit être un identifiant SGML dans le document SGML courant. Ceci implique que le passage ou l'expression pointé doit porter un identifiant, et donc être balisé comme un élément d'un certain type. Dans l'exemple suivant, la référence croisée pointe vers un élément `<div1>` :

```
...
voir notamment <ptr target=SEC12>.
...
<div1 id=SEC12><head>En ce qui concerne les identificateurs
...
```

Puisque l'attribut *id* est global, tout élément dans un document peut être pointé de la même façon. Dans l'exemple suivant, un paragraphe a été pourvu d'un identifiant pour qu'on puisse pointer sur lui :

```
...
on en reparlera dans <ref target=pspec>la section sur
les liens</ref>
...
<p id=pspec>Les liens peuvent être composés de n'importe
quel type d'éléments
...
```

Il est possible d'utiliser l'attribut *targetType* pour préciser que l'élément pointé doit être d'un type particulier, comme dans l'exemple qui suit.

...

```
on en reparlera dans <ref target=dspec targType='div1 div2'>
la section sur les liens</ref>
```

Cette référence devrait échouer si l'élément portant l'identifiant *dspec* n'est ni un *<div1>* ni un *<div2>*. À noter cependant que cette vérification ne peut être effectuée par un analyseur syntaxique SGML seul puisque ce dernier ne peut que vérifier l'existence d'un élément *dspec* donné.

L'attribut *type* peut être employé pour catégoriser le lien représenté par le pointeur, par tout moyen approprié. Les attributs *resp* et *crDate* peuvent également être utilisés pour identifier la personne ou l'entité responsable de la création du lien, ainsi que la date de création de ce dernier, comme dans l'exemple suivant :

```
on en reparlera dans
<ref type=xref resp=auto crDate=950521 target=dspec
  targtype='div1 div2'>
la section sur les liens</ref>
```

Selon toute vraisemblance, ces attributs seront les plus utiles dans des systèmes hypertextes contenant de nombreux pointeurs ayant de multiples fonctions et créés par toute sorte de moyens.

Parfois, la cible d'une référence croisée ne correspond à aucune caractéristique particulière d'un texte, et ne peut donc pas être balisée comme étant d'un certain type d'élément. Si la cible désirée est simplement un endroit dans le document courant, la façon la plus facile de le marquer est d'introduire un élément *<anchor>* à l'endroit correspondant. Si la cible est une séquence de mots non balisée autrement, l'élément *<seg>* (segment) peut être introduit pour la marquer. Ces deux éléments sont décrits comme suit :

*<anchor>*

indique un emplacement ou un endroit sur lequel on peut pointer dans un document ;

*<seg>* identifie un passage ou un segment de texte au sein d'un document, sur lequel on peut pointer ; exemple d'attribut :

*type* catégorise le segment.

Dans l'exemple (imaginaire) suivant, des éléments *<ref>* ont été employés pour représenter des points dans le texte devant être attachés d'une certaine

manière à d'autres parties du texte, un point précis dans le premier cas et une séquence de mots dans le second :

Si je reviens sur `<ref target=ABCD>`le passage sur lequel je me suis endormi`</ref>`, je note que `<ref target=EFGH>`trois mots`</ref>` ont été entourés de rouge par un précédant lecteur.

Ce codage nécessite que des éléments ayant les identifiants spécifiés (ABCD et EFGH dans cet exemple) existent quelque part ailleurs dans le document courant. En supposant qu'aucun élément n'existe déjà pour porter ces identifiants, les éléments `<anchor>` et `<seg>` (segment) peuvent être employés :

```
.... <anchor type=bookmark id='ABCD'> ....
....<seg type=target id='EFGH'> ... </seg> ...
```

L'attribut *type* devrait être employé (comme ci-dessus) pour préciser les différents rôles de ces éléments généraux dans un texte. D'autres applications sont présentées ci-après en section 8.3 «Enchaînement d'attributs».

## 8.2. Pointeurs étendus

Les éléments `<ptr>` et `<ref>` servent uniquement pour des références croisées ou des liens dont les cibles se trouvent dans le même document SGML que leur source. En outre, ils ne peuvent indiquer que des éléments SGML. Les éléments présentés dans la présente section ne sont pas limités de cette façon.

`<xptr>` définit un pointeur vers un autre emplacement dans le document courant ou dans un document externe ;

`<xref>` définit un pointeur vers un autre emplacement dans le document courant ou dans un document externe, modifié éventuellement par un commentaire ou un texte supplémentaire.

En plus des attributs applicables aux pointeurs présentés dans la section 8.1 «Références croisées simples» ci-dessus, ces éléments partagent les attributs supplémentaires suivants, dont le but est de spécifier la cible de la référence croisée ou du lien, au lieu de l'attribut *target* :

*doc* indique le document dans lequel l'emplacement requis doit être trouvé ; par défaut, c'est le document courant ;

- from* indique le début de la destination du pointeur, sous forme d'une expression dans la syntaxe du pointeur étendu TEI : par défaut, c'est l'ensemble du document indiqué par l'attribut *doc* ;
- to* indique le point final de la destination du pointeur, sous forme d'une expression dans la syntaxe du pointeur étendu TEI ; cet attribut ne peut être spécifié que si l'attribut *from* l'a été.

Une spécification complète du langage formel employé pour exprimer la cible des pointeurs étendus TEI dépasserait les limites de ce document ; nous nous bornerons donc à indiquer seulement quelques-unes des caractéristiques les plus généralement utiles. Pour plus de détails, il convient de consulter la version complète des recommandations de la TEI.

Un élément `<xptr>` (ou `<xref>`) peut indiquer l'ensemble d'un autre document simplement en donnant un nom d'entité comme valeur de l'attribut *doc*, comme dans cet exemple :

```
voir <xref doc=P3>The TEI Guidelines, passim</xref>
```

Cet exemple suppose qu'une entité système ou publique avec le nom *P3* a été déclarée. Cette déclaration peut être placée dans le fichier d'extension *litemods.ent* ou effectuée par une autre méthode spécifique au logiciel auteur SGML employé – voir section 15 «Chiffres et Graphiques».

L'attribut *from* est employé pour préciser un emplacement donné au sein du document spécifié par l'attribut *doc*. La spécification utilise un langage spécial, appelé syntaxe du pointeur étendu TEI, dont seulement quelques détails sont donnés ici. Dans ce langage, les emplacements sont définis comme une série de pas (*steps*), chacun identifiant une certaine partie du document, souvent en termes des emplacements identifiés par le pas précédent. Par exemple, on pointera vers la troisième phrase du deuxième paragraphe du chapitre deux en sélectionnant le chapitre deux dans le premier pas, le deuxième paragraphe dans le deuxième pas, et la troisième phrase dans le dernier pas. Un pas peut être défini en termes de concepts SGML (tels que *parent*, *descendant*, *précédant*, etc.) ou, plus vaguement, en termes de structure de texte, de positions de mots ou de caractères. On peut également utiliser une notation étrangère (non-SGML), ou préciser un emplacement dans un graphique en termes de son système de coordonnées.

Les attributs *from* et *to* utilisent la même notation. Chacun indique une certaine portion du document cible ; le pointeur étendu dans son ensemble indique la section qui commence au début de *from* et qui se termine à la fin de *to*.

La première étape du chemin vers un emplacement sera souvent de spécifier l'identifiant d'un certain élément dans le document cible, comme dans cet exemple :

```
<xptr doc=P3 from='id (SA) '>
```

Ceci sélectionne l'ensemble de l'élément qui porte l'identifiant *SA* à l'intérieur de l'entité *P3*. Si une cible plus fine est nécessaire, d'autres pas peuvent suivre. Les mots clés suivants sont disponibles afin de permettre la sélection d'autres éléments en termes de leur relation avec un élément identifié lors du pas précédent :

*child* éléments contenus par cet élément ;

*ancestor* éléments englobant cet élément, directement ou indirectement ;

*previous* éléments ayant le même parent que cet élément mais situés avant dans le document ;

*next* éléments ayant le même parent que cet élément mais situés après dans le document ;

*preceding*  
éléments dans le document qui commencent avant cet élément, indépendamment de leurs parents ;

*following*  
éléments dans le document qui commencent après cet élément, indépendamment de leurs parents.

Chacun de ces mots clés implique un ensemble particulier d'éléments (un ensemble d'enfants, un ensemble d'ancêtres, un ensemble de descendants, etc.). Afin de préciser lequel des éléments d'un ensemble est pointé, le mot clé peut éventuellement être suivi d'une liste entre parenthèses contenant :

- un nombre positif ou négatif, indiquant lequel des nombreux éléments éventuellement trouvés est le bon (+1 indiquant le premier élément rencontré, commençant à l'emplacement actuel, et -1 indiquant le dernier), ou le mot clé *all* (tous) indiquant que tous les éléments dans l'ensemble doivent être pointés ;
- un identifiant générique, indiquant le type d'élément requis, ou une étoile *\** indiquant que n'importe quel type d'élément fait l'affaire ;

- un ensemble de noms et de valeurs d'attribut, indiquant que l'élément sélectionné devrait avoir des attributs ayant les noms et les valeurs spécifiés, s'il en existe.

En reprenant l'exemple ci-dessus, la référence suivante sélectionne le troisième élément `<p>` contenu directement par tout élément ayant l'identifiant `SA`:

```
<xptr doc=P3 from='id (SA) child (3 p)'\>
```

De la même manière, en supposant que l'entité `P3` est en fait une référence à la version SGML des recommandations de la TEI, alors la référence suivante sélectionne la section 14.2.2 de cette publication qui se trouve être celle où la syntaxe du pointeur étendu est formellement définie :

```
Pour plus de détails, voir  
<ref doc=P3 from='id (SA) child (2 div2) child (2 div3)'\>  
  TEI Extended pointer syntax definition  
</ref>
```

Normalement, l'étendue d'une référence croisée est suffisamment définie par l'attribut `from`. Pour certains documents cependant, il peut être plus commode de définir une plage de début et une plage de fin. Comme indiqué ci-dessus, l'attribut `to` est fourni dans ce but. Par exemple :

```
<xptr doc=P1 from='id (xyz)' to='id (abc)'\>
```

est un pointeur étendu dont la cible est la séquence qui commence au début de l'un quelconque des éléments du document `P1` ayant l'identifiant `XYZ`, et se terminant à la fin de n'importe quel élément du même document ayant l'identifiant `ABC`. Tout élément se trouvant dans l'intervalle est également inclus, indépendamment de sa structure. Le pointeur est erroné si la fin de `ABC` précède le début de `XYZ`.

L'utilisation de cette syntaxe permet la construction facile de spécifications très complexes. Par exemple, la référence suivante sélectionne l'élément `<head>` le plus proche ayant un attribut `lang` avec la valeur `LAT`, et apparaissant avant le début de l'élément ayant l'identifiant `SA` :

```
<xptr doc=P3 from='id (SA) preceding (1 head lang lat)'\>
```

Si aucune valeur n'est fournie pour l'attribut *doc*, on suppose qu'il s'agit du document courant. Ainsi, les références suivantes sont sémantiquement équivalentes. Les deux indiquent l'élément ayant l'identifiant *X1* dans le document courant :

```
<ptr target=X1>
<xptr from='id (X1) '>
```

### 8.3. Attributs de liaison

Les attributs de liaison spécifiques suivants ont été définis pour chaque élément de la DTD *TEI Lite* :

- ana* relie un élément à son interprétation ;
- corresp* relie un élément à un ou plusieurs autres éléments correspondants ;
- next* relie un élément au prochain élément dans un agrégat ;
- prev* relie un élément à l'élément précédent dans un agrégat.

L'attribut *ana* (analyse) est destiné à être utilisé lorsqu'un ensemble d'interprétations ou d'analyses abstraites a été défini quelque part dans un document, comme nous le verrons de façon plus détaillée en section 16 « Interprétation et analyse. » Par exemple, une analyse linguistique de la phrase « John aime Nancy » pourrait être codé comme suit :

```
<seg type=sentence ana=SVO>
  <seg type=lex ana=NP1>John</seg>
  <seg type=lex ana=VVI>aime</seg>
  <seg type=lex ana=NP1>Nancy</seg>
</seg>
```

Ce codage implique l'existence, ailleurs dans le document, d'éléments ayant les identifiants *SVO*, *NP1*, et *VVI* où la signification de ces codes particuliers est expliquée. On notera l'emploi de l'élément *<seg>* (segment) pour marquer des composants particuliers de l'analyse, différenciés par l'attribut *type*.

L'attribut *corresp* (correspondant) est un moyen simple de représenter une certaine forme de correspondance entre deux éléments dans un texte. Par exemple, dans un texte plurilingue, il peut être employé pour relier les diverses traductions d'un même passage, comme dans l'exemple suivant :

```
<seg lang=FRA id=FR1 corresp=EN1>Jean aime Nancy</seg>
<seg lang=ENG id=EN1 corresp=FR1>John loves Nancy</seg>
```

Le même mécanisme peut être employé à beaucoup d'autres fins. Dans l'exemple suivant, il est employé pour représenter des correspondances anaphoriques entre *the show* et *Shirley*, et entre *NBC* et *network*:

```
<p><title id=shirley>Shirley</title>, which made  
its Friday night debut only a month ago, was  
not listed on <name id=nbc>NBC</name>'s new schedule,  
although <seg id=network corresp=nbc>the network</seg>  
says <seg id=show corresp=shirley>the show</seg>  
still is being considered.
```

Les attributs *next* et *previous* sont un moyen simple de relier les composants d'un élément discontinu, comme dans l'exemple suivant:

```
<q id=Q1a next=Q1b>Who-e debel you?</q>  
&mdash he at last said &mdash  
<q id=Q1b prev=Q1a>you no speak-e,  
damme, I kill-e.</q> And so saying,  
the lighted tomahawk began flourishing  
about me in the dark.
```

## 9 – Interventions éditoriales

---

Le codage d'un texte électronique a beaucoup de points communs avec l'édition d'un manuscrit ou d'un texte destiné à être imprimé. Dans les deux cas, un éditeur consciencieux peut vouloir enregistrer l'état originel de la source ainsi que toutes les corrections éditoriales ou les modifications qui y ont été apportées. Les éléments présentés dans cette section et la suivante fournissent quelques ressources permettant de répondre à ces besoins. Le couple d'éléments suivant peut être employé pour marquer une «correction», c'est-à-dire des changements éditoriaux introduits aux endroits où le rédacteur croit rencontrer une erreur dans l'original :

*<corr>* contient la forme correcte d'un passage apparemment erroné dans la copie ; parmi les attributs, citons :

- sic* donne la forme originelle de l'erreur supposé dans la copie ;
- resp* identifie l'éditeur ou le transcripteur ayant proposé la correction contenue par l'élément *<corr>* ;
- cert* signifie le degré de certitude attribué à la correction contenue par l'élément *<corr>* ;

*<sic>* contient le texte reproduit bien qu'apparemment incorrect ou inexact ; parmi les attributs possibles, citons :

- corr* donne une correction pour l'erreur apparente dans la copie ;
- resp* identifie le rédacteur ou le transcripteur ayant suggéré la correction ;
- cert* signifie le degré de certitude attribuée à la correction.

Le couple d'éléments suivant peut être employé pour marquer la normalisation, c'est-à-dire des changements éditoriaux introduits pour des raisons de cohérence ou pour transcrire le texte sous une forme plus lisible pour un lecteur moderne :

*<orig>* contient la forme originelle d'une lecture, pour laquelle une forme régularisée est donnée dans une valeur d'attribut ; parmi les attributs possibles, citons :

- reg* donne une forme régularisée (normalisée) du texte ;

- resp* identifie la personne responsable de la régularisation du mot ou de l'expression ;
- <reg> contient une lecture qui a été régularisée ou normalisée dans un certain sens ; parmi les attributs possibles, citons :
- orig* donne la forme non régularisée du texte telle qu'elle apparaîtrait dans l'original ;
- resp* identifie la personne responsable de la régularisation du mot ou de l'expression.

Par exemple, la lecture :

*... for his nose was as sharp as a pen and a' table of green feelds*

comporte selon Gifford

1. la substitution erronée de *table* pour *babbled*, et
2. les orthographes non-standard *a'* et *feelds* pour *he* et *fields*.

Donc, la conjecture de Gifford pourrait être codée ainsi :

```
... for his nose was as sharp as a pen and  
<reg sic="a'">he</reg>  
<corr sic='table' ed=Gifford>babbl'd</corr> of green  
<reg sic='feelds'>fields</reg>
```

## 10 – Omissions, effacements et ajouts

Outre la correction ou la normalisation des mots et des expressions, les rédacteurs et les transpositeurs peuvent aussi ajouter du texte dans des passages lacunaires, ôter du texte, ou encore transcrire du texte effacé ou biffé dans l'original. En outre, un texte donné peut être particulièrement difficile à transcrire car difficile à déchiffrer dans la page. Les éléments suivants peuvent être employés pour enregistrer de tels phénomènes :

*<add>* contient des lettres, mots ou expressions insérés dans le texte par un auteur, scribe, annotateur ou correcteur ; parmi les attributs possibles, citons :

*place* si l'ajout est écrit dans la copie, il indique l'emplacement du texte supplémentaire. Des exemples de valeurs possibles sont *inline* (sur la ligne), *supralinear* (au dessus de la ligne), *infralinear* (sous la ligne), *left* (dans la marge gauche), *right* (dans la marge droite), *top* (au dessus), *bottom* (en dessous), etc.

*<gap>* indique un endroit où du texte a été omis dans une transcription, que ce soit pour des raisons éditoriales décrites dans l'en-tête TEL, (dans la section décrivant les principes de sélection), ou parce que l'original est illisible ou inaudible ; parmi les attributs possibles, citons :

*desc* donne une description du texte omis ;

*resp* indique l'éditeur scientifique, le transpositeur ou l'encodeur ayant pris la décision de ne pas fournir une transcription du texte, d'où l'application de la balise *<gap>*.

*<del>* contient une lettre, un mot ou un passage effacé, marqué comme effacé, sinon indiqué comme étant superflu ou erroné dans la copie par un auteur, scribe, annotateur ou correcteur ; parmi les attributs possibles, citons :

*type* classe le type d'effacement selon une typologie appropriée ;

*status* peut être employé pour indiquer des effacements erronés, par exemple des raturages comportant trop de texte ou pas assez de texte ;

*hand* désigne la main de l'agent qui a effectué l'effacement.

<*unclear*>

contient un mot, une expression ou un passage qui ne peut pas être transcrit avec certitude parce qu'il est illisible ou inaudible dans la source ; parmi les attributs possibles, citons :

*reason* indique la raison pour laquelle le passage est difficile à transcrire ;

*resp* indique la personne responsable de la transcription de la lettre, du mot ou du passage contenu dans l'élément <*unclear*>.

Ces éléments peuvent être employés pour enregistrer des changements effectués par un éditeur, par le transcripteur ou (dans la source manuscrite) par l'auteur ou le scribe. Par exemple, si la source d'un texte électronique est :

Les éléments suivants sont proposés pour  
pour noter des interventions du transcripteur.

alors il pourrait être souhaitable de corriger l'erreur évidente, tout en enregistrant l'effacement du deuxième *pour* superflu, ainsi :

Les éléments suivants sont proposés pour  
<del hand=LB>pour</del> noter des interventions du transcripteur.

La valeur d'attribut LB sur l'attribut *hand* indique que «LB» a corrigé la duplication du mot *pour*.

Si la source était :

Les éléments suivants proposés pour  
pour noter des interventions du transcripteur.

(si le verbe avait été oublié par inadvertance), alors le texte lui-même pourrait se lire :

Les éléments suivants <add hand=LB>sont</add> proposés pour  
<del hand=LB>pour</del> noter des interventions du transcripteur.

Ces éléments ne sont pas limités aux changements faits par un éditeur; ils peuvent aussi être employés pour enregistrer des changements dus à l’auteur dans des manuscrits. Un manuscrit<sup>1</sup> dans lequel Proust a d’abord écrit «Bientôt la lumière sous la porte s’éteint et il retombe dans l’obscurité.» puis biffé «il retombe» et inséré en-dessous «tout rentre» pourrait être codé ainsi :

```
Bientôt la lumière sous la porte s’éteint et
<del hand=PROUST type=overstrike>il retombe</del>
<add hand=PROUST place=infralinear> tout rentre</add>
dans l’obscurité.
```

De la même manière, les éléments `<unclear>` et `<gap>` peuvent être employés ensemble pour indiquer l’omission d’un passage illisible; l’exemple suivant montre aussi l’emploi de `<add>` pour une correction conjecturale :

```
One hundred & twenty good regulars joined to me
<unclear><gap reason='indecipherable'></unclear>
& instantly, would aid me signally <add hand=ed>in?</add>
an enterprise against Wilmington.
```

L’élément `<del>` identifie les passages qui sont transcrits dans le texte électronique bien qu’ils soient marqués comme étant effacés, tandis que `<gap>` marque l’emplacement d’un passage qui est omis du texte électronique, qu’il soit lisible ou non. Un corpus contenant des passages en plusieurs langues, par exemple, pourrait omettre des citations longues dans des langues étrangères :

```
<p> ... An example of a list appearing in a fief ledger of
<name type=place>Koldinghus</name> <date>1611/12</date>
is given below. It shows cash income from a sale of
honey.</p>
<q><gap desc='quotation from ledger'
reason='in Danish'></q>
<p>A description of the overall structure of the account is
once again ... </p>
```

D’autres corpus (notamment ceux qui ont été créés avant l’emploi généralisé du lecteur optique) omettent systématiquement les schémas et les parties mathématiques :

---

1. Cité par Almuth Grésillon, *Éléments de critique génétique*, PUF, 1994 (p. 132–133).

<p>En avant du puzzle ...

Un des classeurs est ouvert sur une page en partie  
couverte d'équations transcrites d'une écriture  
fine et serrée:

<gap desc='demonstration morphisme' reason='maths'>  
</p>

# 11 – Noms, dates, chiffres et abréviations

---

La TEI définit des éléments pour un grand nombre de types de données spéciales que l'on peut rencontrer presque partout dans des textes de toutes sortes. Ces types de données peuvent être d'un intérêt particulier dans tout un éventail de disciplines. Ils se réfèrent tous à des objets externes au texte lui-même (noms de personnes et de lieux, chiffres, dates). Ils posent toutefois des problèmes particuliers à beaucoup d'applications de traitement du langage naturel (NLP), à cause des formes variées sous lesquelles ils peuvent apparaître dans les textes. Les éléments décrits dans le présent chapitre, en rendant ces types de données explicites, facilitent le traitement des textes qui les contiennent.

## 11.1. Noms et chaînes de caractère de référence

Une « chaîne de référence » (*referring string*) est une expression qui se réfère à une personne, un endroit, un objet donné, etc. Deux éléments sont fournis pour marquer de telles chaînes :

`<rs>` contient une chaîne de référence ou un nom général ; parmi les attributs possibles, citons :

*type* indique plus spécifiquement l'objet auquel la chaîne se réfère. Des exemples de valeurs sont : *personne*, *endroit*, *navire*, *élément*, etc.

`<name>`

contient un nom propre ou une proposition substantive ; parmi les attributs possibles, citons :

*type* indique le type d'objet qui est nommé par l'expression.

L'attribut *type* est employé pour distinguer (par exemple) entre des noms de personnes, d'endroits ou d'organisations, dans les cas où cela est possible :

```
<q>My dear <rs type=person>Mr. Bennet</rs>, </q>
said his lady to him one day, <q>have you heard
```

that `<rs type=place>Netherfield Park</rs>` is let  
at last?</q>

It being one of the principles of the  
`<rs type=organization>Circumlocution Office</rs>` never,  
on any account whatsoever, to give a straightforward answer,  
`<rs type=person>Mr Barnacle</rs>` said, <q>Possibly.</q>

Comme le montrent les exemples suivants, l'élément `<rs>` peut être employé pour toute référence à une personne, un endroit, etc., qui n'est pas forcément un nom propre ou une proposition substantive :

<q>My dear `<rs type=person>Mr. Bennet</rs>`,</q>  
said `<rs type=person>his lady</rs>` to him  
one day...

<q>Peu après son installation `<rs type=lieu>rue Simon-  
Crubelier</rs>`, `<rs type=personne>Maurice Réol</rs>`, qui  
était `<rs type=metier>aide-rédacteur</rs>` à la  
`<rs type=organisation>CAMP&A</rs>` ...

L'élément `<name>`, au contraire, est prévu pour le cas spécial des chaînes de référence contenant uniquement des noms propres ; il peut être utilisé de la même façon que l'élément `<rs>`, ou imbriqué au sein de celui-ci si une chaîne de référence contient un mélange de noms communs et de noms propres.

Le simple fait de baliser un objet en tant que «nom» ne suffit généralement pas pour permettre le traitement automatique des noms de personnes afin d'obtenir les formes canoniques généralement requises à des fins de référence. Un nom tel qu'il apparaît dans le texte peut être orthographié de façon incohérente, ou être partiel ou flou. Qui plus est, des particules de noms tel que *van* ou *de la* peuvent ou non être incluses dans la forme de référence d'un nom. Ceci dépend de la langue et du pays de celui qui porte le nom en question.

Les attributs suivants sont également disponibles pour ces éléments et pour des éléments similaires, afin de surmonter ces difficultés :

*key*      fournit un autre identifiant pour l'objet nommé, telle qu'une clé d'un enregistrement de base de données ;

*reg*      donne une forme régularisée ou normalisée du nom utilisé.

L'attribut *key* peut être un moyen utile pour rassembler toutes les références se rapportant à la même personne ou au même emplacement éparpillés à travers un document :

```
<q>My dear <rs type=person key=BENM1>Mr. Bennet</rs>,
  </q> said <rs type=person key=BENM2>his lady</rs>
  to him one day, <q>have you heard that
  <rs type=place key=NETP1>Netherfield Park</rs>
  is let at last?</q>
```

Cette utilisation devrait être distinguée du cas de l'attribut *reg* (régularisation), qui permet de marquer la forme standard d'une chaîne de référence, comme ci-dessous :

```
<name type=person key=WADLM1 reg='de la Mare, Walter'>
  Walter de la Mare
</name>
was born at
<name key=Ch1 type=place>Charlton</name>, in
<name key=KT1 type=county>Kent</name>, in 1873.
```

On peut aussi baliser de façon plus détaillée les composants de noms propres, en utilisant le jeu de balises supplémentaires traitant les noms et les dates.

## 11.2. Dates et heures

Les balises suivantes permettent un codage plus détaillé des dates et de l'heure :

*<date>* contient une date dans n'importe quel format; parmi les attributs possibles, citons :

*calendar* indique le système ou le calendrier auquel la date se rattache;

*value* donne la valeur de la date sous une forme standard, habituellement aaaa-mm-jj;

*<time>* contient une expression définissant une heure du jour dans n'importe quel format; parmi les attributs possibles, citons :

*value* donne la valeur de l'heure sous une forme standard.

L'attribut *value* indique une forme normalisée pour la date ou l'heure, au moyen d'un format reconnu tel que celui qui est prescrit par la norme ISO 8601. Les dates ou les heures partielles (par exemple « 1990 », « septembre 1990 », « autour de midi ») peuvent habituellement être exprimées en omettant simplement une partie de la valeur donnée; ou bien, les dates ou les heures imprécises (par exemple « début août », « entre dix et douze heures ») peuvent être exprimées comme une plage de dates ou d'heures. Si l'une ou l'autre extrémité de la plage d'heure ou de date est connue avec certitude (par exemple, « avant 1230 », « quelques jours après Hallowe'en »), l'attribut *exact* peut être employé pour le préciser.

Exemples :

```
<date value='1980-02-21'>21 Feb 1980</date>
<date value='1990'>1990</date>
<date value='1990-09'>September 1990</date>
```

```
Given on the <date value='1977-06-12'>Twelfth Day of June
in the Year of Our Lord One Thousand Nine Hundred and
Seventy-seven of the Republic the Two Hundredth and first
and of the University the Eighty-Sixth.</date>
```

```
<l>specially when it's nine below zero
<l>and <time value='15:00'>three o'clock in the afternoon</time>
```

```
<p>C'était une belle matinée de la <date value='1323-11'>fin
novembre</date> ...
```

### 11.3. Nombres

Les nombres peuvent être écrits en lettres ou en chiffres (vingt et un, XXI et 21) et leur présentation dépend de la langue (par exemple *5th* en anglais devient 5. en grec; *123,456.78* en anglais équivaut à 123.456,78 en français<sup>1</sup>. Dans des applications de traitement du langage naturel ou de traduction automatique, il est souvent utile de les différencier par rapport à d'autres parties

---

1. Toutefois les codes typographiques français recommandent aujourd'hui d'écrire plutôt 123 456,78 {NdT}.

plus « lexicales » du texte. Dans d'autres applications, la capacité d'enregistrer une valeur numérique en utilisant une notation standard est importante. L'élément `<num>` fournit cette possibilité :

`<num>` contient un chiffre, écrit dans n'importe quel format ; attributs possibles :

<i>type</i>	indique le type de valeur numérique ; les valeurs suggérées comprennent : <i>fraction</i> , <i>ordinal</i> (pour des chiffres ordinaux, par exemple « vingt et unième », pourcentage, et <i>cardinal</i> (un nombre absolu, par exemple « 21 », « 21,5 », etc.) ;
<i>value</i>	fournit la valeur du nombre dans un format dépendant de l'application.

Par exemple :

```
<num value='33'>xxxiii</num>
<num type=cardinal value='21'>twenty-one</num>
<num type=percentage value='10'>ten percent</num>
<num type=percentage value='10'>10%</num>
<num type=ordinal value='5'>5th</num>
```

## 11.4. Les abréviations et leur développement

De même que les noms, les dates et les nombres les abréviations peuvent être transcrites telles quelles ou sous une forme développée ; elles peuvent être soit non-balisées, soit codées au moyen de l'élément suivant :

`<abbr>` contient une abréviation de tout genre ; parmi les attributs possibles, citons :

<i>expan</i>	donne le développement de l'abréviation ;
<i>type</i>	permet au codeur de classer l'abréviation selon une typologie adéquate ; exemples : <i>contraction</i> , <i>suspension</i> , <i>brevigraph</i> , <i>superscription</i> ou <i>acronym</i> ; l'attribut <i>type</i> peut aussi recevoir une valeur telle que <i>titre</i> (pour des titres d'adresse), <i>géographique</i> , <i>organisation</i> , etc., décrivant la nature de l'objet auquel on se réfère.

L'élément `<abbr>` est utile pour distinguer les éléments semi-lexicaux tels que des acronymes ou des termes de jargon :

```
We can sum up the above discussion as follows: the identity
of a <abbr>CC</abbr> is defined by that calibration of values
which motivates the elements of its <abbr>GSP</abbr>;
```

```
Every manufacturer of <abbr>3GL</abbr> or <abbr>4GL</abbr>
languages is currently nailing on <abbr>OOP</abbr> extensions
```

```
Le brevet <abbr>BBR</abbr> de justification des lignes a été
repris par la <abbr>CII</abbr> ...
```

L'attribut *type* peut être employé pour distinguer des types d'abréviation selon leur fonction, et l'attribut *expan* peut être employé pour fournir un développement :

```
Le <name><abbr type=titre expan='Docteur'>Dr.</abbr>
<abbr type=initiale expan='Jean'>J.</abbr>
Transen</name>
est le <abbr type=fonction expan='Président directeur
général'>PdG</abbr> de la <abbr type=acronyme expan='Compagnie
des Compteurs Informatisés'>C2I</abbr>.
```

Cet élément est aussi particulièrement utile lorsqu'on transcrit des documents manuscrits contenant de fréquentes abréviations.

## 11.5. Adresses

L'élément `<address>` est employé pour noter une adresse postale de n'importe quelle sorte. Il contient un ou plusieurs éléments `<addrLine>`, un pour chaque ligne de l'adresse.

`<address>`

contient une adresse postale ou d'un autre type, par exemple l'adresse d'un éditeur, d'une organisation ou d'une personne ;

`<addrLine>`

contient une ligne d'une adresse postale ou autre.

Voici un exemple simple :

```
<address>
<addrLine>Rédaction des Cahiers GUTenberg</addrLine>
<addrLine>Irisa/Inria, campus de Beaulieu</addrLine>
<addrLine>F-35042 Rennes cedex</addrLine>
<addrLine>France</addrLine>
</address>
```

Les parties individuelles d'une adresse peuvent être subdivisées au moyen de l'élément `<name>` expliqué ci-dessus en section 11.1 «Noms et chaînes de référence».

```
<address>
<addrLine>Rédaction des Cahiers GUTenberg</addrLine>
<addrLine>Irisa/Inria, campus de Beaulieu</addrLine>
<addrLine><name type=code>F-35042</name>
    <name type=city>Rennes cedex</name>
<addrLine><name type=country>France</name></addrLine>
</address>
```

## 12 – Listes

---

L'élément `<list>` est utilisé pour marquer toutes sortes de listes. Une liste est une séquence d'items, qui peuvent être ordonnés, non ordonnés ou présentés sous forme de glossaire. Chaque élément peut être précédé d'une étiquette (dans une liste de type glossaire, cette étiquette est le terme défini) :

`<list>` contient toute séquence d'éléments organisée comme une liste ; attribut(s) possible(s) :

*type* décrit la forme de la liste ; les valeurs suggérées sont :

- `ordered`, `bulleted` (respectivement pour des listes marquées à l'aide de lettres ou de chiffres, ou bien à l'aide de puces typographiques<sup>1</sup>),
- `gloss` (pour des listes comprenant un ensemble de termes techniques, chacun marqué par un élément `<label>` et accompagné d'une glose ou d'une définition marquée comme un `<item>`),
- `simple` (pour des listes avec des éléments non marqués par des chiffres ou des puces) ;

`<item>` contient un composant d'une liste ;

`<label>` contient l'étiquette associée à un élément de la liste ; dans des glossaires, marque le terme défini.

Les éléments de la liste sont balisés par `<item>`. Le premier `<item>` peut, en option, être précédé par un élément `<header>`, qui fournit un en-tête à la liste. La numérotation d'une liste peut être omise (si elle est reconstituable), ce qui est indiqué au moyen de l'attribut `n` sur chaque entrée de la liste ou (rarement) balisé comme un contenu au moyen de l'élément `<label>`. Les listes suivantes sont par conséquent toutes équivalentes :

---

1. Les codes typographiques français recommandent plutôt l'emploi du tiret – que la puce • [NdT].

```
<list>
<head>Une petite liste :</head>
<item>premier élément de la liste ;</item>
<item>second élément ;</item>
<item>dernier élément.</item>
</list>
```

```
<list>
<head>Une petite liste :</head>
<item n=1>premier élément de la liste ;</item>
<item n=2>second élément ;</item>
<item n=3>dernier élément.</item>
</list>
```

```
<list>
<head>Une petite liste :</head>
<label>1</label><item>premier élément de la liste ;</item>
<label>2</label><item>second élément ;</item>
<label>3</label><item>dernier élément.</item>
</list>
```

Les styles ne devraient pas être mélangés dans la même liste.

Un tableau simple à deux colonnes peut être traité comme une liste de type glossaire, balisée `<list type=gloss>`. Ici, chaque élément comprend un terme et une glose, marqués par `<label>` et `<item>` respectivement. Ceux-ci correspondent aux éléments `<term>` et `<gloss>` que l'on peut rencontrer n'importe où dans un texte de prose.

```
<list type=gloss>
<head>Vocabulary</head>
<label lang=enm>nu</label> <item>now</item>
<label lang=enm>lhude</label> <item>loudly</item>
<label lang=enm>bloweth</label> <item>blooms</item>
<label lang=enm>med</label> <item>meadow</item>
<label lang=enm>wude</label> <item>wood</item>
<label lang=enm>awe</label> <item>ewe</item>
<label lang=enm>lhouth</label> <item>lows</item>
<label lang=enm>sterteth</label> <item>bounds, frisks</item>
<label lang=enm>verteth</label> <item lang=lat>pedit</item>
<label lang=enm>murie</label> <item>merrily</item>
<label lang=enm>swik</label> <item>cease</item>
<label lang=enm>naver</label> <item>never</item>
</list>
```

Lorsque la structure interne d'un élément de la liste est plus complexe, il peut être préférable de considérer la liste comme une table, entité pour laquelle un balisage spécial est défini dans un jeu de balises TEI supplémentaire.

Les éléments d'une liste peuvent bien sûr contenir eux-mêmes des listes de toute sorte jusqu'au niveau de profondeur requis. Ici, par exemple, une liste de type glossaire contient deux éléments, chacun d'entre eux étant lui-même une liste simple :

```
<list type=gloss><label>EVIL</label>
<item><list type=simple>
  <item>I am cast upon a horrible desolate island, void
    of all hope of recovery.</item>
  <item>I am singled out and separated as it were from
    all the world to be miserable.</item>
  <item>I am divided from mankind &mdash a solitaire; one
    banished from human society.</item>
</list> <!-- end of first nested list --></item>
<label>GOOD</label>
<item><list type=simple>
  <item>But I am alive; and not drowned, as all my
    ship's company were.</item>
  <item>But I am singled out, too, from all the ship's
    crew, to be spared from death...</item>
  <item>But I am not starved, and perishing on a barren
    place, affording no sustenances....</item>
</list><!-- end of second nested list --></item>
</list><!-- end of glossary list -->
```

Une liste ne doit pas nécessairement être affichée sous forme de liste. Par exemple :

```
On those remote pages it is written that animals are
divided into <list rend='runon'><item n='a'>those that
belong to the Emperor,<item n='b'> embalmed ones, <item n='c'>
those that are trained, <item n='d'> suckling pigs, <item n='e'>
mermaids, <item n='f'> fabulous ones, <item n='g'> stray
dogs, <item n='h'> those that are included in this
classification, <item n='i'> those that tremble as if they
were mad, <item n='j'> innumerable ones, <item n='k'> those
drawn with a very fine camel's-hair brush, <item n='l'>
others, <item n='m'> those that have just broken a flower
vase, <item n='n'> those that resemble flies from a
distance.</list>
```

Les listes de références bibliographiques devraient être balisées au moyen de l'élément `<listBibl>` décrit dans la section suivante.

## 13 – Citations bibliographiques

Il est souvent utile de distinguer les citations bibliographiques aux endroits où elles apparaissent dans les textes qui sont transcrits à des fins scientifiques, ne serait-ce que pour s'assurer de leur formatage correct lors de l'impression. L'élément `<bibl>` est fourni dans ce but :

`<bibl>` contient une citation bibliographique structurée de façon lâche, dans laquelle les sous-éléments peuvent ou non être balisés explicitement.

Lorsque les composants d'une référence bibliographique doivent être distingués, les éléments suivants peuvent être employés selon les besoins. Il est généralement utile de marquer au moins les parties (tels les titres d'articles, de livres ou de journaux) qui nécessitent un formatage spécial. Les autres sont prévus pour les cas où on attache de l'importance à ces détails.

`<author>`

dans une référence bibliographique, contient le nom de l'auteur (qu'il s'agisse d'une personne physique ou d'une collectivité) ; ceci constitue la mention de responsabilité principale de chaque unité bibliographique ;

`<biblScope>`

définit l'étendue d'une référence bibliographique, par exemple, sous forme d'une liste de numéros de page ou d'un nom de subdivision d'un ouvrage plus important ;

`<date>` contient une date dans n'importe quel format ;

`<editor>`

mention de responsabilité secondaire pour une unité bibliographique, par exemple le nom d'une (ou de plusieurs) personnes, institutions ou organisations agissant en tant que rédacteur, compilateur, traducteur, etc. ; parmi les attributs possibles, citons :

`role` indique la nature de la responsabilité intellectuelle ; valeurs possibles : `translator`, `compiler`, `illustrator`, etc. ; la valeur par défaut est `editor` ;

<imprint>

regroupe des informations relatives à la publication ou la diffusion d'une unité bibliographique ;

<publisher>

fournit le nom de l'organisation responsable de la publication ou de la diffusion d'une unité bibliographique ;

<pubPlace>

contient le nom du lieu de publication d'une unité bibliographique ;

<series>

contient des informations sur la série dans laquelle un livre ou une autre unité bibliographique a paru ;

<title> contient le titre d'une œuvre (article, livre, journal, périodique), y compris tous les titres ou sous-titres alternatifs ; parmi les attributs possibles, citons :

*type* catégorise le titre d'une certaine façon, par exemple comme un titre main ou subordinate, etc.

*level* indique la classe ou le niveau bibliographique du titre ; les valeurs permises sont décrites dans la section 6.1 « Alternances typographiques ».

Par exemple, l'annotation éditoriale suivante

*He was a member of Parliament for Warwickshire in 1445, and died March 14, 1470 (according to Kittredge, Harvard Studies 5. 88ff).*

pourrait être transcrite comme suit :

```
He was a member of Parliament for Warwickshire in 1445, and
died March 14, 1470 (according to <bibl><author>Kittredge
</author>, <title>Harvard Studies</title> <biblScope>5.
88ff</biblScope></bibl>).
```

Dans le cas des listes de citations bibliographiques, l'élément <listBibl> devrait être utilisé ; il peut contenir un ensemble d'éléments <bibl>. Pour un exemple, se reporter à la liste donnée en section 22 « Références ».

## 14 – Tables

---

Les tableaux représentent un défi considérable pour tout système de traitement de texte, mais des tableaux simples, au moins, apparaissent dans un si grand nombre de textes que même le jeu de balises simplifié présenté ici doit comporter un balisage pour les tableaux. Les éléments suivants sont donc proposés :

`<table>` contient le texte affiché sous forme tabulaire, en lignes et en colonnes ; parmi les attributs possibles, citons :

`rows` indique le nombre de lignes dans la table ;

`cols` indique le nombre de colonnes dans chaque ligne de la table ;

`<row>` contient une ligne d'une table ; parmi les attributs possibles, citons :

`role` indique le type d'information contenu dans les cellules de cette ligne ; les valeurs suggérées sont `label` pour des étiquettes ou des informations descriptives, et `data` pour les données proprement dites ;

`<cell>` contient une cellule de table ; parmi les attributs possibles, citons :

`role` indique le type d'information contenu dans la cellule ; les valeurs suggérées sont `label` pour des étiquettes ou des informations descriptives et `data` pour les données proprement dites ;

`cols` indique le nombre de colonnes occupées par cette cellule ;

`rows` indique le nombre de lignes occupées par cette cellule.

Par exemple, Defoe utilise des tables de mortalité comme celle qui suit dans son *Journal of the Plague Year*, pour montrer la montée et la descente de l'épidémie :

```
<p>It was indeed coming on amain, for the burials that
same week were in the next adjoining parishes thus:&mdash;
<table rows=5 cols=4>
<row role='data'>
```

```
<cell role='label'>St. Leonard's, Shoreditch</cell>
  <cell>64</cell> <cell>84</cell> <cell>119</cell></row>
<cell role='label'>St. Botolph's, Bishopsgate</row>
  <cell>65</cell> <cell>105</cell> <cell>116</cell></row>
<cell role='label'>St. Giles's, Cripplegate</row>
  <cell>213</cell> <cell>421</cell> <cell>554</cell></row>
</table>
```

```
<p>This shutting up of houses was at first counted a very cruel
and unchristian method, and the poor people so confined made
bitter lamentations.... </p>
```

De même, Perce dans *La vie mode d'emploi* utilise-t-il des tables pour montrer des «jeux» :

```
<p>Il résolvait avec une facilité déconcertante :
les d'un-mot-à-l'autre
<table rows=5 cols=3>
<row><cell>VIN</cell><cell>HOMME</cell><cell>POÈME</cell></row>
<row><cell>VAN</cell><cell>GOMME</cell><cell>POÈTE</cell></row>
<row><cell>VAU</cell><cell>GEMME</cell><cell>PRÊTE</cell></row>
<row><cell>EAU</cell><cell>FEMME</cell><cell>PROTE</cell></row>
<row><cell>  </cell><cell>    </cell><cell>PROSE</cell></row>
</table>
les problèmes mathématiques ...
</p>
```

## 15 – Figures et graphiques

---

Tous les composants d'un document ne sont pas nécessairement textuels. Le texte le plus simple contient souvent des diagrammes ou des illustrations, sans parler des documents dans lesquels images et textes sont inextricablement mêlées, ou de produits électroniques dans lesquels les deux sont complémentaires.

Le codeur peut simplement enregistrer la présence d'un graphique dans un texte, éventuellement avec une description brève de son contenu, en utilisant les éléments décrits dans cette section. Les mêmes éléments peuvent aussi être employés pour insérer des versions numérisés du graphique au sein du document électronique.

`<figure>`

marque l'endroit où un graphique doit être inséré dans un document ; parmi les attributs possibles, citons :

*entity* nom d'une entité système prédéfinie contenant une version numérisée du graphique à insérer ;

`<figDesc>`

contient une description textuelle de l'aspect ou du contenu d'un graphique, lorsqu'une image est documentée sans être visualisée.

Toute information textuelle accompagnant le graphique, tel qu'un titre ou une légende, peut être incluse dans l'élément `<figure>`, à l'intérieur d'un élément `<head>` et d'un ou plusieurs éléments `<p>`. Il en va de même pour tout texte figurant dans le graphique lui-même. Il est fortement recommandé d'inclure une description en texte libre de l'image, sous forme d'un élément `<figDesc>`. Cette description sera utilisée pour les applications incapables d'afficher le graphique et permettra de rendre le document accessible aux lecteurs malvoyants (de tels textes ne sont pas normalement considérés comme faisant partie du document lui-même).

Dans leur utilisation la plus simple, ces éléments servent à marquer l'emplacement d'un graphique, comme dans l'exemple suivant :

```
<pb n=412>
<figure><figure>
<pb n=413>
```

À noter que la balise de fin ne peut être omise, bien que l'élément n'ait aucun contenu. Plus habituellement, un graphique aura au moins un titre servant à l'identifier, ce dernier étant codé au moyen de l'élément `<head>`. De plus, il est souvent utile d'inclure une brève description de l'image, comme dans l'exemple suivant :

```
<figure>
  <head>Mr Fezziwig's Ball</head>
  <figdesc>A Cruikshank engraving showing Mr Fezziwig
    leading a group of revellers.</figdesc>
</figure>
```

Lorsqu'une version numérisée du graphique est disponible, il est nettement préférable de l'insérer à l'endroit approprié du document. Les éléments graphiques tels que les images sont généralement stockés dans des entités (des fichiers) distinctes de ceux qui contiennent le texte du document et au moyen d'une notation différente (format de stockage). La DTD de TEI Lite supporte des graphismes codés au moyen des normes CGM, TIFF et JPEG, sous les noms `cgm`, `tiff`, et `jpeg` dans la notation SGML. Néanmoins, d'autres notations peuvent être employées, pourvu qu'une déclaration de notation appropriée soit ajoutée dans la DTD. Pour plus d'informations concernant la déclaration d'une NOTATION en SGML, se reporter au chapitre traitant des tableaux, formules et graphismes dans la TEI P3, ou dans tout ouvrage de référence traitant de SGML.

Quel que soit le format employé pour coder l'image, celle-ci peut être insérée dans le document par le même procédé. La première phase consiste à déclarer une entité SGML d'un type particulier, précisant le nom de l'entité elle-même, un identifiant externe (tel qu'un nom de fichier), et la notation employée. Par exemple, en supposant que l'image numérisée «Le Bal de Mr Fezziwig» soit au format TIFF, dans le fichier `fezzi.tff`, une déclaration d'entité semblable à celle qui est présentée ci-dessous serait nécessaire :

```
<!ENTITY fezziPic SYSTEM "fezzi.tff" NDATA tiff>
```

Toutes les déclarations de ce type doivent être traitées avant le document SGML lui-même ; avec la DTD de la TEI Lite, ceci est possible en incluant ces déclarations dans un fichier nommé `litedecls.ent` ou un fichier quelconque associé à l'identifiant public suivant :

```
-//TEI U5-1995//DTD TEI Lite 1.0 Extensions//EN
```

---

Une fois que cette déclaration est effective, l'insertion de l'image numérisée à l'emplacement voulu dans le document ne nécessite que l'insertion d'une valeur pour l'attribut *entity* de l'élément `<figure>` :

```
<figure entity=fezziPic>
<head>Mr Fezziwig's Ball</head>
<figdesc>A Cruikshank engraving showing Mr Fezziwig leading
  a group of revellers.</figdesc>
</figure>
```

## 16 – Interprétation et analyse

On dit souvent que le balisage constitue une forme d'interprétation ou d'analyse. Tandis qu'il est certainement difficile, voire impossible, de distinguer de façon certaine entre les informations « objectives » et « subjectives » d'une manière universelle, il reste vrai que les jugements concernant les dernières sont plus sujets à controverse que les premières. De nombreux chercheurs préfèrent donc présenter de telles interprétations uniquement lorsqu'il est possible d'informer le lecteur qu'il s'agit de parties plus sujettes à caution que le reste du balisage. Cette section décrit certains des éléments fournis par la TEI pour répondre à ce besoin.

### 16.1. Phrases orthographiques

Typiquement, l'interprétation porte sur l'ensemble d'un texte, sans prendre particulièrement en compte les autres unités structurales. Un préliminaire utile avant toute interprétation intensive consiste donc à segmenter le texte dans des unités discrètes et identifiables, chacune pouvant ensuite recevoir une étiquette servant de « référence canonique ». Afin de faciliter ce genre d'application, ces unités ne doivent pas se croiser, ni être imbriquées les unes dans les autres. Il est possible de les représenter de façon commode en utilisant l'élément suivant :

<s> identifie un *s-unit* au sein d'un document, dans le but d'établir un système de référence s'appliquant au texte entier ; parmi les attributs possibles, citons :

*type* qui catégorise l'unité (par exemple comme étant déclaratif interrogatif, etc.).

Comme le nom suggère, l'élément <s> est le plus souvent employé (au moins dans les applications linguistiques) pour marquer des phrases orthographiques, c'est-à-dire des unités définies par des caractéristiques orthographiques telles que la ponctuation. Par exemple, le passage pris dans *Jane Eyre* cité dans le premier chapitre pourrait être divisé en *s-units* comme suit :

```

<pb n='474'>
<div1 type=chapter n='38'>
<p><s n=001>Reader, I married him.</s>
<s n=002>A quiet wedding we had:</s>
<s n=003>he and I, the parson and clerk, were
    alone present.</s>
<s n=004>When we got back from church, I went
into the kitchen of the manor-house, where Mary
was cooking the dinner,
and John cleaning the knives, and I said &dash;</s>
<p><q><s n=005>Mary, I have been married
to Mr Rochester this morning.</s></q>...}

```

Les balises de fin utilisées ci-dessus ne sont pas strictement nécessaires, car les éléments `<s>` ne peuvent pas s’emboîter : le début d’un élément `<s>` implique que le précédent est terminé. Quand les *s-units* sont balisés comme indiqué ci-dessus, il est conseillé de baliser le texte tout entier, afin que chaque mot du texte analysé soit contenu dans exactement un élément `<s>`, dont l’identifiant peut alors servir de référence unique. Si les identifiants employés sont uniques au sein du document, alors l’attribut *id* pourrait être utilisé de préférence à l’attribut *n* employé dans l’exemple ci-dessus.

## 16.2. Éléments d’interprétation polyvalents

Un élément de segmentation plus polyvalent, le `<seg>` a déjà à été présenté : il sert à identifier les cibles de références croisées et de liens hypertexte qui ne disposent pas par ailleurs d’éléments d’identification – se reporter à la section 8 « Liens et références croisées ». Cet élément identifie une quelconque expression de type « phrase » à laquelle le codeur peut attribuer un type spécifié par l’utilisateur, ainsi qu’un identifiant unique ; il peut ainsi servir à baliser les caractéristiques textuelles pour lesquelles rien n’a été prévu dans la version publiée des recommandations de la TEI.

Par exemple, les Recommandations ne fournissent aucun élément `<apostrophe>` pour marquer les parties d’un texte littéraire où le narrateur s’adresse au lecteur (ou à l’auditeur) directement. Une approche possible serait de considérer ces cas comme des exemples de l’élément `<q>`, distingués des autres par une valeur appropriée de l’attribut *who*. Cependant, une solution peut-être plus simple (et certainement plus générique) serait d’employer l’élément `<seg>` comme suit :

```

<div1 type=chapter n='38'>
<p><seg type='apostrophe'>Reader, I married him.</seg>
A quiet wedding we had:...

```

L'attribut *type* de l'élément `<seg>` peut prendre n'importe quelle valeur, donc peut servir à enregistrer les phénomènes du genre «expression» de toute sorte ; il est souhaitable d'enregistrer les valeurs employées ainsi que leur signification dans l'en-tête.

Un élément `<seg>` d'un type donné (à la différence de l'élément `<s>` auquel il ressemble superficiellement) peut s'emboîter à l'intérieur d'un élément `<seg>` du même type ou non. Cela permet la représentation de structures relativement complexes ; certains exemples ont été donnés dans la section 8.3 «Attributs de liaison» ci-dessus. Cependant, puisqu'il doit respecter la contrainte imposée par SGML qui est que les éléments doivent être emboîtés correctement et ne doivent pas se chevaucher, cet élément ne permet pas d'associer une interprétation à des segments arbitraires d'un texte éventuellement sans tenir compte de la hiérarchie du document. Il nécessite également que l'interprétation elle-même soit représentée par une valeur unique codée au sein de l'attribut *type*.

Aucune de ces contraintes ne s'applique à l'élément `<interp>`, qui comporte de puissantes caractéristiques permettant de coder d'une manière assez simple des interprétations relativement complexes.

`<interp>`

fournit une annotation interprétative qui peut être reliée à un passage de texte. Les attributs sont :

- |               |  |
|---------------|--|
| <i>valeur</i> | identifie le phénomène spécifique en cours d'annotation ;  |
| <i>resp</i>   | identifie le responsable à l'origine de l'interprétation ;   |
| <i>type</i>   | indique quelle sorte de phénomène est notée dans ce passage ; des valeurs possibles de cet attribut sont :<br>image, caractère, thème, allusion,<br>ou le nom d'un type particulier de discours dont des exemples<br>sont identifiés ; |
| <i>inst</i>   | pointe vers des exemples de l'analyse ou de l'interprétation<br>représentée par l'élément courant.   |

`<InterpGrp>`

regroupe un ensemble de balises `<interp>`.

Ces éléments permettent au codeur de préciser à la fois la classe d'une interprétation, et l'instance particulière de cette classe impliquée par l'interprétation. Ainsi, tandis qu'avec `<seg>` nous pouvons dire simplement que quelque chose est une apostrophe, avec `<interp>` nous pouvons dire qu'il s'agit d'une instance (apostrophe) d'une plus grande classe (de figures rhétoriques).

De plus, `<interp>` est un élément vide qui doit être relié au passage auquel il s'applique, soit au moyen de l'attribut *ana* présenté dans la section 8.3 «Attributs de liaison», ci-dessus, soit au moyen de son propre attribut *inst*. Cela signifie que toute sorte d'analyse peut être représentée, sans que l'on soit obligé de respecter la hiérarchie SGML du document ; cette approche facilite également le regroupement d'analyses d'un type particulier. Un élément spécifique `<interpGrp>` est fourni à cette fin.

Par exemple, supposons que l'on désire baliser des caractéristiques textuelles variées telles des thèmes, des sujets ou des figures rhétoriques, ainsi que les emplacements des diverses scènes d'un récit. Par exemple, diverses parties de notre passage spécimen puisé dans *Jane Eyre* pourraient être associées à des figures rhétoriques comme l'apostrophe, l'hyperbole et la métaphore, accompagnées de références à des églises, des serviteurs, la cuisine, la poste ou des lunes de miel, et illustrées par des scènes à l'église, dans la cuisine, ou dans un endroit non spécifié (le salon?).

Ces interprétations pourraient être placées n'importe où dans l'élément `<texte>` ; il est cependant souhaitable de les inclure au même endroit (par exemple dans une section séparée des textes liminaires ou annexes), comme dans l'exemple suivant :

```
<back>
<div1 type='Interpretations'>
<interp id='fig-apos'  resp='LB, MSM'
      type='figure of speech' value='apostrophe'>
<interp id='fig-hyp'   resp='LB, MSM'
      type='figure of speech' value='hyperbole'>
<!--... -->
<interp id='set-church' resp='LB, MSM'
      type='setting' value='church'>
<!--... -->
<interp id='ref-church' resp='LB, MSM'
      type='reference' value='church'>
<interp id='ref-serv'   resp='LB, MSM'
      type='reference' value='servants'>
<!--... -->
</p></div>
```

La redondance évidente de ce type de codage peut être considérablement réduite en utilisant l'élément `<interp>` pour regrouper tous ces éléments `<interp>` qui partagent des valeurs d'attribut communes, comme suit :

```

<back>
<div1 type='Interpretations'>
<interpGrp type='figure of speech' resp='LB, MSM'>
<interp id='fig-apos' value='apostrophe'>
<interp id='fig-hyp' value='hyperbole'>
<interp id='fig-meta' value='metaphor'>
<!--... -->
</interpGrp>
<interpGrp type='scene-setting' resp='LB, MSM'>
<interp id='set-church' value='church'>
<interp id='set-kitch' value='kitchen'>
<interp id='set-unspec' value='unspecified'>
<!--... -->
</interpGrp>
<interpGrp type='reference' resp='LB, MSM'>
<interp id='ref-church' value='church'>
<interp id='ref-serv' value='servants'>
<interp id='ref-cook' value='cooking'>
<!--... -->
</interpGrp>
</p></div>

```

Une fois que ces éléments d'interprétation ont été définis, ils peuvent être reliés aux parties du texte auquel ils s'appliquent, d'une ou deux manières. L'attribut *ana* peut servir sur tout élément approprié :

```

<div1 type=chapter n='38'>
<p id='P38.1' ana='set-church set-kitch'>
<s id=P38.1.1 ana='fig-apos'>Reader, I married him.</s>
...

```

Notons dans cet exemple que puisque le paragraphe a deux scènes (dans l'église et dans la cuisine), les identifiants des deux sont fournies.

Alternativement, les éléments *<interp>* peuvent pointer sur toutes les parties du texte auquel ils s'appliquent, en utilisant leur attribut *inst* :

```

<interp id='fig-apos' type='figure of speech' resp='LB, MSM'
value='apostrophe' inst='P38.1.1'>
<!--... -->
<interp id='set-church' type='scene-setting' value='church'
inst='P38.1' resp='LB, MSM'>
<interp id='set-kitchen' type='scene-setting' value='kitchen'
inst='P38.1' resp='LB, MSM'>
<!--... -->

```

---

L'élément *<interp>* n'est pas limité à un quelconque type d'analyse ; l'analyse littéraire montrée ci-dessus n'est qu'une possibilité et l'on pourrait aussi bien employer *<interp>* afin de donner une analyse linguistique des parties du discours. Par exemple, la phrase d'exemple donnée dans la section 8.3 « Attributs de liaison » suppose une analyse linguistique qui pourrait être représentée comme suit :

```
<interp id=NP1 type=pos value='noun phrase, singular'>
<interp id=VV1 type=pos value='inflected verb,
      present-tense singular'>
...
```

---

## 17 – Documentation technique

---

Bien que ce document soit axé sur l'utilisation de la TEI pour le codage de documents existant déjà sous forme électronique, la même approche peut également servir pour le codage de nouveaux documents. Dans la préparation de nouveaux documents (tels que celui-ci), SGML a beaucoup d'attraits : la structure du document peut être représentée clairement et le même texte électronique peut être réutilisé pour maintes fins (par exemple pour produire à la fois des versions en ligne – en hypertexte ou consultables – et des versions imprimées bien formatées et mises en pages, le tout à partir d'une source SGML commune).

Pour faciliter ceci, un nombre limité d'éléments supplémentaires sont inclus dans la TEI Lite en tant qu'extensions du DTD principal de la TEI ; ils serviront à marquer les caractéristiques particulières de documents techniques en général et des documents SGML en particulier.

### 17.1. Éléments supplémentaires pour les documents techniques

Les éléments suivants peuvent servir pour marquer les caractéristiques particulières de documents techniques :

- `<eg>` regroupe un exemple court de notation technique, par exemple un code fragmenté ou un échantillon de codage SGML ;
- `<code>` regroupe un fragment code écrit dans une certain langage formel (souvent un langage de programmation) ;
- `<ident>` contient un identifiant d'un certain type, par exemple un nom variable ou le nom d'un élément ou d'un attribut SGML ;
- `<gi>` contient un type spécial d'identifiant : un identifiant générique SGML, ou un nom d'élément ;
- `<kw>` contient un mot clé dans une certaine langue formelle ;

<formula>

contient une formule chimique ou mathématique, présentée facultativement dans une quelconque notation non-SGML ; parmi les attributs possibles, citons :

*notation* précise la notation employée pour représenter le corps de la formule ; la valeur par défaut est `tex`, signifiant que la formule est représentée au moyen du système de composition  $\text{T}_{\text{E}}\text{X}$ .

Les exemples suivantes indiquent une utilisation possible de ces éléments, à savoir, le codage d'un passage faisant partie d'un cours de présentation du langage de programmation Fortran :

<p>Il est de tradition de présenter un langage de programmation à l'aide d'un premier exemple comme :

```
<eg>
  CHAR*20 GRTG
  GRTG = 'BONJOUR TOUT LE MONDE'
  PRINT *, GRTG
  END
```

</eg></p>

<p>Dans cet exemple, on commence par déclarer la variable <ident>GRTG</ident>, dans la ligne <kw>CHAR\*20 GRTG</kw>, qui identifie <ident>GRTG</ident> comme formée de 20 octets de type <kw>CHAR</kw>. On affecte alors à cette variable la valeur <mentioned>BONJOUR TOUT LE MONDE</mentioned>. Suivent alors l'ordre d'impression <kw>PRINT</kw> et l'instruction finale <kw>END</kw>.

Une application de formatage ayant à traiter un tel texte pourrait être programmée de façon à formater convenablement des exemples (par exemple, en conservant les coupures de ligne, ou en employant une police distinctive). Parallèlement, l'emploi de balises telles que <ident> et <kw> facilite grandement la création d'un index utile.

L'élément <formula> devrait servir à entourer une formule chimique ou mathématique présentée au sein du texte comme étant un passage distinct. Puisque les formules comportent généralement une grande variété de caractéristiques typographiques spéciales qui ne figurent pas ailleurs dans le texte courant, il sera habituellement nécessaire de présenter le corps de la formule dans une notation spécialisée. La notation employée devrait être spécifiée par l'attribut *notation*, comme dans l'exemple suivant :

```
<formula notation=tex>
  \(\text{E} = mc^{\{2\}}\)
</formula>
```

La notation *tex* est pré-définie pour la DTD de la TEI Lite ; d'autres notations peuvent être employées si besoin est, mais elles doivent être définies au préalable au moyen d'une mention du type «notation» dans la DTD.

Presque toutes les séquences de caractères sont admises dans le corps d'un élément *<formula>*, du moins du point de vue d'une application capable de traiter des données SGML. Les données sont transférées sans modification par l'analyseur syntaxique à toute application associé à la notation spécifiée. La seule exception à cette règle est que l'analyseur syntaxique reconnaîtra tout objet qui ressemble au début d'une fin-de-balise SGML, c'est-à-dire le caractère «plus petit que» (<) suivi aussitôt par une barre oblique (/) et un caractère alphabétique. En traitant l'exemple imaginaire ci-dessous, un analyseur syntaxique SGML générerait toute une suite de messages d'erreurs.

```
<formula notation=tex>
  \ (E = mc^{2} </a\ )
</formula>
```

Heureusement , la séquence «</ » est plutôt rare dans la majorité des notations mathématiques utilisées actuellement. Néanmoins, si elle apparaît, il est nécessaire de prendre des dispositions qui dépassent le cadre de ce document (pour plus d'informations, se reporter au texte complet des recommandations de la TEI).

Ce problème existe sous une forme plus aiguë lorsque le codage SGML lui-même est le sujet de discussion à l'intérieur d'un document technique, lui-même codé en SGML. Dans un tel document, il est évidemment essentiel de faire une distinction entre le balisage SGML contenu dans les exemples, et celui qui est employé pour le balisage du document lui-même ; dans ces textes, en effet, l'emploi de balises de fin est très vraisemblable. La solution la plus générale consiste à marquer le corps de chaque exemple SGML comme contenant des données qui ne doivent pas être balayées par l'analyseur syntaxique pour des fins de balisage SGML. Ceci est fait en l'entourant d'une structure SGML spéciale appelée *CDATA marked section*, comme dans l'exemple suivant :

```
<p>Une liste devrait être codée comme suit :
<eg><![ CDATA [
  <list>
    <item>Premier élément de la liste</item>
    <item>Second élément</item>
  </list>
]]>
</eg>
```

Les éléments *<gi>list</gi>* sont composés d'une série d'éléments *<gi>item</gi>*.

L'élément `<list>` employé dans l'exemple ci-dessus ne sera pas considéré comme faisant partie du document proprement dit, parce qu'il est inséré à l'intérieur d'une section balisée (qui commence avec la mention spéciale de balisage `<! [ CDATA [, et qui se termine avec ] ]>`).

À noter également l'utilisation de l'élément `<gi>` pour baliser les références à des noms d'éléments SGML (ou des *identifiants génériques*) au sein du corps du texte.

## 17.2. Divisions générées automatiquement par les outils bureautiques

La plupart des systèmes modernes de production de documents sont capables de générer automatiquement des sections entières telles qu'une table de matières ou un index. La TEI Lite fournit un élément pour marquer l'emplacement d'une section générée de cette façon.

`<divGen>`

indique l'emplacement prévu pour une division textuelle générée automatiquement par une application de traitement de texte ; parmi les attributs possibles, citons :

*type* précise le type de la division de texte prévue (par exemple, un index, table de matières etc.) ; exemples : `index` (un index doit être généré et inséré à ce point), `toc` (une table des matières), `figlist` (une liste de figures) et `tablist` (une liste de tables).

L'élément `<divGen>` peut être placé partout où un élément de division serait autorisé, comme dans l'exemple suivant :

```
<front>
<titlePage>... </titlePage>
<divGen type=toc>
<div type='Preface'><head>Preface</head>... </div>
</front>
<body>... </body>
<back>
<div1><head>Appendix</head>... </div1>
<divGen type=index n='Index'>
</back>
```

Cet exemple montre également l'emploi de l'attribut *type* pour distinguer les différentes sortes de division à générer : dans le premier cas une table des matières (*toc*) et dans le second un index.

Lorsqu'une table des matières ou un index existant doit être codé (plutôt que généré) pour une raison quelconque, il convient d'employer l'élément `<liste>` présenté dans la section 12 «Listes».

### 17.3. Génération d'index

Tandis que la génération d'une table des matières à partir d'un document correctement balisé se passe généralement sans problèmes pour un processeur automatique, la production d'un index de bonne qualité nécessitera dans bien des cas un balisage plus réfléchi. Il peut ne pas être suffisant de produire une simple liste de toutes les parties balisées d'une certaine façon, bien que le fait d'extraire (par exemple) toutes les occurrences d'éléments tels que *<terme>* ou *<nom>* soit souvent un bon point de départ pour un index.

La DTD de la TEI fournit une balise *<index>* spéciale qui peut servir pour indiquer à la fois les parties du document qui devrait figurer dans l'index, et la façon dont l'indexage devrait être fait.

*<index>*

marque un emplacement à indexer pour une certaine raison ; parmi les attributs possibles, citons :

- level1* donne la forme principale de l'entrée d'index ;
- level2* donne la forme du second niveau, s'il existe ;
- level3* donne la forme du troisième niveau, s'il existe ;
- level4* donne la forme du quatrième niveau, s'il existe ;
- index* indique à quel index (entre plusieurs) appartient l'entrée d'index.

Par exemple, le deuxième paragraphe de la présente section pourrait présenter le balisage suivant :

...

```
La DTD de la TEI fournit une balise <gi>index</gi> tag
<index level1='indexing'>
<index level1='index (tag)' level2='use in index generation'>
spéciale qui peut servir ...
```

L'élément *<index>* peut également servir pour fournir une forme d'information analytique ou interprétative. Par exemple, dans une étude d'Ovide, on pourrait vouloir enregistrer toutes les références du poète concernant les différents personnages, pour des besoins d'étude stylistique comparative. Dans les lignes suivantes des *Métamorphoses*, une telle étude enregistrerait les références du poète à Jupiter (comme *deus*, *se* et en tant que sujet de *confiteor* [sous

la forme inflectionnelle 227]), à Jupiter en guise de taureau (*imago tauri fallacis* et sujet de *teneo*), et ainsi de suite<sup>1</sup>. L'exemple a été légèrement simplifié.

```
<l n=3.001>iamque deus posita fallacis imagine tauri
<l n=3.002>se confessus erat Dictaeaque rura tenebat
```

Cet objectif pourrait être atteint au moyen de l'élément `<note>` présentée dans la section 7 «Notes» ou au moyen de l'élément `<interp>` présenté dans la section 16 «Interprétation et Analyse». Ici nous montrons le moyen d'obtenir le même résultat avec l'élément `<index>`.

Nous supposons que l'objet doit générer plus d'un index : un pour des noms de divinités (appelé *dn*), un autre pour des références onomastiques (appelé *on*), un troisième pour les références pronominales (appelées *pr*), et ainsi de suite. Une façon d'y parvenir est indiquée ci-dessous :

```
<l n=3.001>iamque deus posita fallacis imagine tauri
  <index index="dn" level1="Iuppiter" level2="deus">
  <index index="on" level1="Iuppiter (taurus)"
    level2="imago tauri fallacis"></l>
<l n=3.002>se confessus erat Dictaeaque rura tenebat
  <index index="pr" level1="Iuppiter" level2="se">
  <index index="v" level1="Iuppiter" level2="confiteor
    (v227)">
  <index index="mons" level1="Dicte" level2="rura Dictaea">
  <index index="regio" level1="Creta" level2="rura Dictaea">
  <index index="v" level1="Iuppiter (taurus)"
    level2="teneo (v9)"></l>
```

Pour chaque élément `<index>` ci-dessus, une entrée sera générée dans l'index approprié, en employant comme mot principal la valeur de l'attribut *level1* et comme mot clé secondaire celui de l'attribut *level2* qui contient le mot cité sous sa forme nominative. La référence elle-même sera prise dans le contexte où figure l'élément `<index>`, c'est-à-dire dans le cas présent, l'identifiant de l'élément `<l>` qui le contient.

---

1. L'analyse est empruntée, avec autorisation, à Willard McCarty et Burton Wright, *An Analytical Onomasticon to the Metamorphoses of Ovid* (Princeton: Princeton University Press, à paraître).

## 18 – Jeux de caractères, signes diacritiques, etc.

Pour ceux qui travaillent avec des formes classiques des langues européennes, les recommandations de la TEI concernant l'emploi des jeux de caractères sont simples. Pour un emploi local, il suffit d'utiliser tout jeu de caractères supporté par la machine et le logiciel. Au cas où le logiciel rend difficile la saisie directe des caractères spéciaux au clavier, il est possible de définir des conventions propres de saisie (par exemple, représenter les lettres accentuées en dactylographiant l'accent approprié aussitôt après la lettre, ou en utilisant des séquences spéciales qui n'ont que peu de chance d'apparaître dans le texte normal, tel que «aE» pour «ä»). Des fonctions de recherche et de remplacement globales peuvent être ensuite utilisées pour transformer ces raccourcis en des caractères corrects<sup>1</sup>. Si l'on doit employer des écritures non latines et qu'il existe un système de translitération normalisé dans le domaine particulier (par exemple, pour le grec ancien, le code beta du *Thesaurus Linguae Græcæ*), il faut utiliser cette norme. Toute translitération employée devrait être réversible (ce qui exclut un nombre surprenant de schémas employés communément dans l'écriture normale), et son utilité sera plus grande si elle ne nécessite aucune ligature spéciale ni lien ni signe diacritique (ce qui exclut un nombre surprenant des schémas restants...).

Pour l'échange de fichiers entre des systèmes, seules les références d'entité SGML sont à employer pour remplacer tout caractère ne figurant pas dans la liste de caractères ci-dessous (les caractères de cette liste sont ceux qui peuvent être échangés sans perte d'informations entre la plupart des systèmes):

```
a b c d e f g h i j k l m n o p q r s t u v w x y z
A B C D E F G H I J K L M N O P Q R. S T U V W X Y Z
0 1 2 3 4 5 6 7 8 9
" % & ' ( ) * + , - . / : ; < = ? _ (space)
```

Cette liste exclut les caractères suivants

```
! # $ [ \ ] ^ _ ` { } | ~
```

qui, dans bien des cas et au grand mécontentement des utilisateurs non avertis, ne survivent pas aux transferts au-delà des frontières nationales ou à travers les réseaux longue distance<sup>2</sup>.

1. C'est ce que nous faisons dans la version française de ce document où nous avons utilisé les caractères accentués é è ê ã etc. au lieu de leur codage TEI.

2. Ces caractères font en fait partie de l'IVR (*International Reference Version*) du codage Ascii/Iso 646. Voir le *Cahier GUTenberg* n° 20 sur ces problèmes de codage.

Par contre, pour un simple transfert entre Mac et PC, ces caractères pourront peut-être être échangés sans dommage.

Afin d'assurer la transmission correcte à travers des réseaux hétérogènes, des références d'entité doivent être employées pour tous les caractères latins accentués et étendus, pour tous les caractères non latins, et enfin pour tous les symboles ne figurant pas sur un clavier d'ordinateur conventionnel.

Si on le désire, on peut employer ses propres noms d'entité SGML dans des fichiers conforme à la TEI, à condition de fournir des mentions standard d'entité SGML à leur place ; mais les noms standard, (quoique longs ou compliqués) ont l'avantage de la clarté ; ces noms sont parlants pour tout locuteur anglophone<sup>3</sup> qui peut reconnaître qu'il s'agit d'un nom de caractère, souvent même sans recours à une liste. Notons que ce n'est pas le cas de beaucoup d'autres méthodes employées pour représenter des caractères accentués.

Les noms d'entité requis pour les caractères présentés ci-dessus comme peu sûrs, et pour les caractères accentués de certaines langues majeures de l'Europe occidentale, sont donnés ci-dessous. Les listes des jeux d'entité publics ainsi que leur contenu sont disponibles dans tout ouvrage de référence traitant de SGML : les noms donnés ci-dessous sont extraits des jeux d'entité publics ISO, sont largement employés et sont donc recommandés.

Lorsqu'un caractère ne paraît pas dans les jeux d'entité public, on peut désirer générer un nom, au moyen des mêmes conventions de nommage employées dans les jeux d'entité publics ISO, comme ici :

#### digrammes

créer les noms d'entité pour les digrammes en adjoignant la chaîne *lig* aux lettres formant le digramme ; si une forme capitalisée est nécessaire, les deux lettres sont rendues en majuscules (rappelons que la casse est habituellement significative dans des noms d'entité) ; exemple : *aelig* (æ), *AElig* (Æ) *szlig* (ß) ;

#### signes diacritiques et accents

créer les noms d'entité pour des lettres accentuées dans la plupart des langues européennes occidentales en adjoignant une des chaînes suivantes à la lettre qui porte l'accent, celle-ci pouvant être en majuscules ou en minuscules ;

umlaut employer *uml* pour umlaut ou trema<sup>4</sup> : par exemple *auml* (ä), *Auml* (Ä), *euml* (ë), *iuml* (ï), *ouml* (ö), *Ouml* (Ö), *uuml* (ü), *Uuml* (Ü) ;

3. Il est donc très important que le français soit aussi accessible «naturellement» grâce justement à ces mentions d'entités.

4. Notons toutefois que ces deux symboles ne sont pas, typographiquement parlant, équivalents : le *umlaut* allemand est plus bas, plus proche, de la voyelle que le trema français {NdT}.

acute employer *acute* pour l'accent tonique ou aigu : par exemple *aacute* (á), *eacute* (é), *Eacute* (É), *iacute* (í), *oacute* (ó), *uacute* (ú) ;

grave employer *grave* pour l'accent grave : par exemple *agrave* (à), *egrave* (è), *igrave* (î), *ograve* (ó), *ugrave* (ù) ;.

circumflex

employer *circ* pour circonflexe : par exemple *acirc* (â), *ecirc* (ê), *Ecirc* (Ê), *icirc* (î), *ocirc* (ô), *ucirc* (û) ;

tilde employer *tilde* pour tilde : par exemple *atilde* (ã), *Atilde* (Ã), *ntilde* (ñ), *Ntilde* (Ñ), *otilde* (õ), *Otilde* (Ö) ;

consonnes

les noms d'entité suivants sont recommandés pour certaines consonnes spéciales utilisées dans les langues de l'Europe de l'ouest :

*ccedil* (ç), *Ccedil* (Ç),

*eth* (eth bas de casse : le d croisé «ð» anglo-saxon ou islandais), *ETH* majuscule : «Ð» ,

*thorn* (thorn minuscule : «þ»), *THORN* (thorn majuscule : «Þ»),

*szlig* (ligature s-z allemande ou *esszett* : ß) ;

signes de ponctuation

les noms d'entité suivants sont recommandés pour certains signes de ponctuation communément rencontrés :

*ldquo* (*left double quotation mark* guillemet double gauche anglais : «“»),

*rdquo* (*right double quotation mark* guillemet double droit anglais : «”»),

*mdash* (*one-em dash* – tiret d'un cadratin «—»),

*hellip* (*horizontal ellipsis* – points de suspension horizontaux «... »),

*rsquo* (*right single quote* – signe anglais de citation droite «' ») ;

voir également la liste des «caractères dangereux» juste ci-après et la figure 1 des caractères français ;

caractères «dangereux»

les caractères présentés ci-dessus (page 104) comme dangereux pour la transmission sur des réseaux internationaux académiques et publics peuvent être représentés par les entités suivantes :

*excl* (!), *num* (#), *dollar* (\$),

*lsqb* (*left square bracket* – crochet gauche [),

*bsol* (*back-slanted solidus* – barre de fraction inverse : \),

*rsqb* (*right square bracket* – crochet droite ]),

*circ* (*circumflex* – circonflexe, ^),

*lquo* (*left single quotation mark* – fin de citation gauche),

*grave* (accent grave), *lcub* (*left curly bracket* – accolade gauche, {),

*rcub* (*right curly bracket* – accolade droite, }),

*verbar* (*vertical bar* – barre verticale, |)

*tilde* (~).

En résumé, pour le français<sup>5</sup>, les codages utiles à connaître (mais rappelons le, le codeur ne devrait normalement pas s’en soucier) sont les suivants :

FIGURE 1 – Codage des caractères français

à	&agrave;	À	&Agrave;
â	&acirc;	Â	&Acirc;
é	&eacute;	É	&Eacute;
è	&egrave;	È	&Egrave;
ê	&ecirc;	Ê	&Ecirc;
ë	&euuml;	Ë	&Euml;
î	&icirc;	Î	&Icirc;
ï	&iuml;	Ï	&Iuml;
ô	&ocirc;	Ô	&Ocirc;
ù	&ugrave;	Ù	&Ugrave;
û	&ucirc;	Û	&Ucirc;
ü	&uuml;	Ü	&Uuml;
ç	&ccdel;	Ç	&Ccdel;
æ	&aelig;	Æ	&AElig;
œ	&oelig;	Œ	&OElig;
«	&ldaq;	»	&rdaq;
–	&ndash;	—	&mdash;

5. D’après le *Lexique des règles typographiques en usage à l’Imprimerie nationale*, Imprimerie nationale, Paris, 1990 (p. 102). Notons toutefois que les caractères «œ Œ» et «æ Æ» ne sont pas des ligatures, facultatives, mais de vrais caractères : voir *Cahier GUTenberg* n° 22 à ce sujet [NdT].

## 19 – Pièces liminaires et annexes

### 19.1. Pièces liminaires

Pour de nombreuses applications, particulièrement dans le cas des textes anciens, les pages liminaires telles les pages de titre, les épîtres introductrices etc., peuvent fournir un complément très utile d'informations sociales ou linguistiques. Le document P3 donne un ensemble de recommandations pour distinguer les éléments textuels rencontrés le plus souvent dans les liminaires. Nous décrivons ces éléments ci-dessous.

#### 19.1.1. Page de titre

Le début d'une page de titre devrait être marqué au moyen de l'élément *titlePage*. Tout texte contenu sur la page devrait être transcrit et balisé au moyen de l'un des éléments figurant dans la liste suivante :

*<titlePage>*

contient la page de titre d'un texte, située à l'intérieur des textes liminaires ou annexes ;

*<docTitle>*

contient le titre d'un document, y compris tous ses constituants, tel que présenté sur une page de titre ; doit être partagé en éléments *<titlePart>* ;

*<titlePart>*

contient une subdivision ou division du titre d'une œuvre, comme indiqué sur une page de titre ; également employé pour les fragments «flottants» de la page de titre ne faisant pas partie du titre de document (paternité de l'ouvrage, etc.) ; parmi les attributs, on trouve :

*type* précise le rôle de division du titre ; les valeurs suggérées sont :

- main* (intitulé principal),
- sub* (sous-titre),
- desc* (une paraphrase descriptive de l'œuvre incluse dans le titre)
- alt* (titre alternatif) ;

*<byline>*

regroupe la mention de responsabilité principale d'une œuvre donnée, tel que reproduite sur la page de titre ou au début ou à la fin de l'ouvrage ;

*<docAuthor>*

contient le nom de l'auteur du document, tel que présenté sur la page de titre (souvent mais pas toujours contenu dans un *<byline>*);

*<docDate>*

contient la date du document, telle que présentée (habituellement) sur la page de titre ;

*<docEdition>*

contient une mention d'édition, telle que présentée sur une page de titre d'un document ;

*<docImprint>*

contient la mention d'impression (lieu et date d'édition, nom d'éditeur), telle que donnée (habituellement) au bas de la page de titre ;

*<epigraph>*

contient une citation, anonyme ou attribuée, apparaissant au début d'une section ou chapitre, ou sur une page de titre.

Les distinctions de polices de caractère devraient être marquées avec l'attribut *rend* quand cela est nécessaire, comme ceci est décrit ci-dessus. La description très détaillée de l'interlettrage et du dimensionnement employé dans des titres ornementaux, n'est pas encore fournie par la TEI. Les changements de langue devraient être marqués par l'emploi approprié de l'attribut *lang* ou de l'élément *<foreign>*, selon les cas. Les noms, partout où ils paraissent, devraient être balisés au moyen de l'élément *<name>*, comme par ailleurs.

Deux pages de titre sont reproduites comme exemples :

*<titlePage>*

```

<docTitle><titlePart type=main>
  Histoire du Roi de Bohême</titlePart>
<titlePart type='sub'>et de ses sept châteaux
</titlePart></docTitle>
<titlePart>Pastiche.</titlePart>
<byline>Par <docAuthor>Charles Nodier</docAuthor>
</byline>
<epigraph>
  <q>O imitatores, servum pecus!
```

```
<bibl>Horat., Epist. I. XIX, 19.</bibl>
</epigraph>
<docImprint><name>PARIS</name>,
  <name>Delangle Frères</name>
  Éditeurs-libraires,
  <name>Place de la Bourse</name>
</docImprint>
<docDate>MDCCCXXX</docDate>
</titlePage>
```

```
<titlePage>
  <docTitle><titlePart type=main>
    Lives of the Queens of England, from the Norman
    Conquest;</titlePart>
  <titlePart type='sub'>with anecdotes of their courts.
</titlePart></docTitle>
  <titlePart>Now first published from Official Records
    and other authentic documents private as well as
    public.</titlePart>
  <docEdition>New edition, with corrections and
    additions</docEdition>
  <byline>By <docAuthor>Agnes Strickland</docAuthor>
</byline>
  <epigraph>
    <q>The treasures of antiquity laid up in old
      historic rolls, I opened.</q>
    <bibl>BEAUMONT</bibl>
  </epigraph>
  <docImprint>Philadelphia: Blanchard and Lea
  </docImprint>
  <docDate>1860</docDate>
</titlePage>
```

### 19.1.2. *Pièces liminaires*

Les blocs majeurs de texte à l'intérieur des pièces liminaires devraient être marqués comme des éléments *<div>* ou *<div1>*; les valeurs suggérées ci-après pour l'attribut *type* peuvent servir pour distinguer divers types communs de textes liminaires :

foreword

un texte adressé au lecteur, par l'auteur, le rédacteur ou l'éditeur, éventuellement sous forme d'une lettre ;

*preface*

un texte adressé au lecteur, par l’auteur, le rédacteur ou l’éditeur, éventuellement sous forme d’une lettre ;

*dedicace*

un texte (souvent une lettre) adressée à quelqu’un d’autre que le lecteur, et dans lequel l’auteur (typiquement) recommande l’œuvre en question à l’attention de la personne concernée ;

*abstract*

un argument de prose récapitulant le contenu de l’œuvre ;

*ack* les remerciements ;*contents*

une table des matières (typiquement, ceci devrait être balisé en tant que *<list>*) ;

*frontispiece*

un frontispice graphique, éventuellement assorti d’un texte.

Comme pour toute division de texte, celles situées dans les pièces liminaires peuvent contenir des éléments structuraux de bas niveau ou des éléments non structuraux, tel que décrits ailleurs. Dans la plupart des cas, elles commenceront avec un titre ou un intitulé d’un certain type, qui devrait être balisé au moyen de l’élément *<front>*. Les épîtres contiendront les éléments supplémentaires suivants :

*<salute>*

contient une salutation ou un salut préfixés à un avant-propos, une épître dédicatoire ou à toute autre division d’un texte, ou la salutation dans la clôture d’une lettre, préface, etc. ;

*<signed>*

contient la salutation finale, etc., ajoutée à la suite d’un avant-propos, d’une épître dédicatoire, ou de toute autre division d’un texte ;

*<byline>*

contient la mention de responsabilité principale donnée pour une œuvre sur sa page de titre, ou au début ou à la fin de l’œuvre ;

*<dateline>*

contient une brève description du lieu et de la date de création d’une lettre, d’un article de journal, ou de toute autre œuvre, qu’elle soit placée au début ou à la fin du document ;

<argument>

liste formelle ou description en texte libre des sujets abordés dans une subdivision d'un texte ;

<cit>

citation prise dans un autre document, accompagnée d'une référence bibliographique indiquant la source ;

<opener>

regroupe la date, la signature, la salutation ainsi que d'autres expressions similaires perçues comme un groupe préliminaire au début d'une division, surtout lorsqu'il s'agit d'une lettre ;

<closer>

regroupe la date, la signature, la salutation ainsi que les expressions similaires perçus comme un groupe préliminaire au début d'une division, surtout lorsqu'il s'agit d'une lettre.

Les épîtres qui paraissent ailleurs dans un texte, bien sûr, contiendront ces mêmes éléments.

À titre d'exemple, la dédicace qui apparaît au début du *Comus* de Milton devrait être balisée comme suit :

```
<div type='dedication'>
<head>To the Right Honourable <name>JOHN Lord Viscount
BRACLY</name>, Son and Heir apparent to the Earl of
Bridgewater, &amp;c.</head>
<salute>MY LORD,</salute>

<p>THis <hi>Poem</hi>, which receiv'd its first occasion of
Birth from your Self, and others of your Noble Family...
and as in this representation your attendant
<name>Thyrsis</name>, so now in all reall expression
<closer>
<salute>Your faithfull, and most humble servant</salute>
<signed><name>H. LAWES.</name></signed>
</closer>
</div>
```

## 19.2. Pièces annexes

### 19.2.1. Divisions structurelles des annexes

À cause des variations dans la pratique éditoriale, les annexes peuvent contenir virtuellement n'importe lequel des éléments présentés ci-dessus en tant

que pièces liminaires, et ces mêmes éléments devraient être employés dans ces cas. En outre, les annexes peuvent contenir les types de texte suivants, dans un élément `<back>`. Comme pour les divisions structurales du corps, ceux-ci devraient être balisés comme des éléments `<div>` ou `<div1>`, et distingués entre eux par l'attribut `type`, avec les valeurs suggérées suivantes :

appendix

un appendice ;

glossary

une liste de mots et de définitions, typiquement sous forme d'une liste de `type=gloss` ;

notes une série de `<note>` ;

bibliography

une série de références bibliographiques, typiquement sous forme d'un élément dédié à la représentation des listes bibliographiques `<listBibl>`, dont les composants sont des éléments individuels `<bibl>` ;

index

une série d'entrées d'index, éventuellement représentées comme une liste structurée ou une liste de type glossaire, avec une introduction optionnelle `<head>` et éventuellement des paragraphes de textes introductifs ou finaux – la TEI P3 définit d'autres éléments spécialisés pour générer des indices dans la production de documents ; ils sont décrits ci-dessus à la section 17.3 « Génération d'index » ;

colophon

description à la fin du livre mentionnant où, quand, et par qui il a été imprimé ; dans les livres modernes il donne souvent les détails de production et identifie les polices utilisées.

## 20 – Page de titre électronique

---

Chaque texte TEI a un en-tête qui fournit des informations analogues à celles que l'on trouve sur une page de titre imprimée. L'en-tête est introduite par l'élément `<teiHeader>` et comporte quatre parties majeures :

`<fileDesc>`

contient une description bibliographique complète d'un fichier électronique ;

`<encodingDesc>`

documente le rapport entre un texte électronique et la source ou les sources dont il dérive ;

`<profileDesc>`

fournit une description détaillée des aspects non bibliographiques d'un texte, spécifiquement les langues et les sous-langues employées, les circonstances de sa production, les participants, et leur environnement ;

`<revisionDesc>`

résume l'historique des révisions du fichier.

Un corpus ou une collection de textes, qui partagent beaucoup de caractéristiques, peuvent avoir un en-tête pour le corpus et des en-têtes individuels pour chaque composant du corpus. Dans ce cas l'attribut *type* indique le type d'en-tête.

```
<teiHeader type=corpus>
```

introduit l'en-tête pour les informations concernant le corpus.

Certains des éléments de l'en-tête contiennent du texte libre, codé sous forme d'un ou plusieurs `<p>`. D'autres sont groupés :

- les éléments dont les noms se terminent par *Stmt* (pour *statement*) englobent habituellement un groupe d'éléments représentant certaines informations structurées ;

- les éléments dont les noms se terminent par *Decl* (pour *declaration*) comportent des informations sur les règles de codage spécifiques mises en œuvre dans le document ;
- les éléments dont les noms se terminent par *Desc* (pour *description*) contiennent une description en texte libre.

## 20.1. Description du fichier

L'élément `<fileDesc>` est obligatoire. Il contient une description bibliographique complète du fichier, à l'aide des éléments suivants :

`<titleStmt>`

regroupe des informations concernant le titre d'une œuvre, ainsi que les parties responsables de son contenu intellectuel ;

`<editionStmt>`

regroupe des informations relatives à une édition donnée d'un texte ;

`<extent>`

décrit la taille approximative du texte électronique tel que stocké sur un certain type de support, spécifié dans n'importe quelles unités communément employées ;

`<publicationStmt>`

regroupe des informations concernant la publication ou la diffusion d'un texte de type électronique ou autre ;

`<seriesStmt>`

regroupe des informations concernant la collection ou série, s'il y en a une, à laquelle une publication appartient ;

`<notesStmt>`

regroupe toute annotation comportant des informations venant s'ajouter à celles qui sont enregistrées dans d'autres parties de la description bibliographique ;

`<sourceDesc>`

comprend une description bibliographique du ou des textes originaux à partir desquels un texte électronique a été dérivé ou généré.

Un en-tête minimal aura la structure suivante :

```
<teiHeader>
  <fileDesc>
    <titleStmt>... </titleStmt>
    <publicationStmt>... <publicationStmt>
    <sourceDesc>... <sourceDesc>
  </fileDesc>
</teiHeader>
```

### 20.1.1. *Mention du titre*

Un élément `<titleStmt>` peut contenir les éléments suivants :

`<title>` contient le titre d'une œuvre, que ce soit un article, un livre, un journal ou une série, y compris tous les sous-titres ou titres alternatifs ;

`<author>`  
dans une référence bibliographique, contient le nom du ou des auteur(s) d'une œuvre, que ce soit une personne ou une institution ; équivaut à la mention de responsabilité principale pour toute unité bibliographique ;

`<sponsor>`  
précise le nom d'une institution ou organisation de parrainage ;

`<funder>`  
précise le nom d'un individu, d'une institution, ou d'une organisation assurant le financement d'un projet ou d'un texte ;

`<principal>`  
fournit le nom du chercheur principal responsable de la création d'un texte électronique ;

`<respStmt>`  
fournit une mention de responsabilité nommant un responsable quant au contenu intellectuel d'un texte, une édition, un enregistrement ou une série, lorsque les éléments spécialisés prévus pour les auteurs, les rédacteurs, etc. ne suffisent pas ou ne s'appliquent pas.

Il est souhaitable que le titre permette de distinguer le fichier électronique du texte source, par exemple :

```
[title of source]: a machine readable transcription
[title of source]: electronic edition
A machine readable version of: [title of source]
```

L'élément `<respStmt>` regroupe les sous-composants suivants :

`<resp>` contient une expression décrivant la nature de la responsabilité intellectuelle d'une personne ;

`<name>`  
contient un nom propre ou une expression substantive.

Exemple :

```
<titleStmt>
  <title>Maximes</title>
  <author>La Rochefoucauld, (1613--1680).
  <respStmt><resp>Texte établi, avec introduction ...</resp>
  <name>Jean Truchet</name></respStmt>
</titleStmt>
```

### 20.1.2. Mention d'édition

L'élément `<editionStmt>` regroupe des informations relatives à une édition d'un texte (l'expression *édition* étant employée avec son sens usuel en bibliographie) et peut inclure les éléments suivants :

`<edition>`  
décrit les particularités d'une édition d'un texte ;

`<respStmt>`  
fournit une mention de responsabilité nommant un responsable pour le contenu intellectuel d'un texte, une édition, un enregistrement ou une série, lorsque les éléments spécialisés pour les auteurs, les rédacteurs etc., ne suffisent pas ou ne s'appliquent pas.

Exemple :

```
<editionStmt>
  <edition n=U2>Troisième version, revue et corrigée
  <date>1987</date>
  </edition>
</editionStmt>
```

La détermination exacte de ce qui représente une nouvelle édition d'un texte électronique est laissée à l'appréciation du codeur.

### 20.1.3. *Mention relative à la taille du fichier*

La mention `<extent>` décrit la taille approximative d'un fichier.

Exemple :

```
<extent>4500 bytes</extent>
```

### 20.1.4. *Mention de publication*

L'élément `<publicationStmt>` est obligatoire. Il peut contenir une simple description en texte libre ou des groupes composés à partir des éléments ci-dessus :

`<publisher>`

fournit le nom de l'organisation responsable pour la publication ou la diffusion d'un article bibliographique ;

`<distributor>`

fournit le nom d'une personne ou d'un organisme responsable de la diffusion d'un texte ;

`<authority>`

fournit le nom d'une personne ou d'un autre organisme responsable de la fourniture d'un fichier électronique, quand cette personne ou cet organisme n'est ni un éditeur ni un distributeur.

Au moins un de ces trois éléments doit être présent, à moins que la totalité de la mention de publication ne soit en texte libre. Ces éléments peuvent à leur tour contenir les éléments suivants :

`<pubPlace>`

contient le nom de l'endroit où une unité bibliographique a été publiée ;

`<address>`

contient l'adresse (postale ou d'autre type) d'un éditeur, d'une organisation, d'un individu, etc. ;

`<idno>` fournit tout chiffre standard ou non standard employé pour identifier une unité bibliographique ; parmi les attributs, citons :

*type* catégorise le nombre, par exemple en tant qu'ISBN ou tout autre numéro d'identification normalisé ;

<availability>

fournit des informations concernant la disponibilité d'un texte, par exemple présence de restrictions sur son emploi ou sa diffusion ; son statut au plan du copyright, etc. ; parmi les attributs possibles, citons :

*status* fournit un code identifiant la disponibilité actuelle du texte ; exemples : *restricted*, *unknown* et *free*.

<date> contient une date dans n'importe quel format.

Exemple (Céline, *Romans, iv*, La Pléiade) :

```
<publicationStmt>
  <publisher>Éditions Gallimard</publisher>
  <pubPlace>Paris</pubPlace> <date>1993</date>
  <idno type=ISBN> 2-07-011336-1</idno>
  <idno type=numero edition>64107</idno>
  <idno type=numero impression>I3-1903</idno>
  <idno type=depot legal>octobre 1993</idno>
  <availability>Copyright: Éditions Gallimard,
    Féerie pour une autre fois I, 1952 ;
    Féerie pour une autre fois II, 1954 ;
    Entretiens avec le professeur Y, 1955 ;
    Appendices, text, préface et apparat critique, 1993.
  </availability>
</publicationStmt>
```

#### 20.1.5. Mentions de série et de notes

L'élément <seriesStmt> regroupe les informations concernant la collection ou série, si elle existe, à laquelle une publication appartient. Il peut contenir <title>, <idno>, ou des éléments <respStmt>.

L'élément <notesStmt>, s'il est employé, contient un ou plusieurs éléments <note> qui contiennent une note ou annotation. Certaines des informations que l'on trouve dans la zone des notes dans les descriptions bibliographiques conventionnelles sont représentées par des éléments spécifiques dans la TEI.

#### 20.1.6. Description de la source

L'élément <sourceDesc> est un élément obligatoire qui enregistre des détails concernant la ou les sources à partir desquelles le fichier électronique a été dérivé. Il peut contenir du texte libre ou une citation bibliographique, au moyen d'un ou plusieurs des éléments suivants :

`<bibl>` contient une citation bibliographique structurée de façon très lâche, dont les composants peuvent être balisés explicitement ou non ;

`<biblFull>`  
contient une citation bibliographique complètement structurée, dans laquelle tous les composants de la description d'un fichier selon la TEI sont présents ;

`<listBibl>`  
contient une liste de citations bibliographiques de toute nature.

Exemples :

```
<sourceDesc>
  <bibl>The first folio of Shakespeare, prepared by
    Charlton Hinman (The Norton Facsimile, 1968)</bibl>
</sourceDesc>
```

```
<sourceDesc>
  <scriptStmt id=CNN12>
    <bibl><author>CNN Network News
      <title>News headlines
      <date>12 Jun 1989
    </bibl>
  </scriptStmt>
</sourceDesc>
```

## 20.2. Description relative au codage

L'élément `<encodingDesc>` précise les méthodes et les principes éditoriaux qui ont régi la transcription du texte. Son emploi est fortement recommandé. Il peut être une description en texte libre, ou se composer d'éléments pris dans la liste suivante :

`<projectDesc>`  
contient une description détaillée de la raison pour laquelle un fichier électronique a été créé ainsi que toute autre information pertinente concernant le procédé par lequel le fichier a été constitué ;

`<samplingDecl>`  
regroupe une description en texte libre des critères et des méthodes employés pour sélectionner des textes lors de la création d'un corpus ou d'un recueil ;

*<editorialDecl>*

fournit des détails des pratiques et des principes éditoriaux appliqués pendant le codage d'un texte ;

*<tagsDecl>*

fournit des informations détaillées concernant le balisage appliqué à un document SGML ;

*<refsDecl>*

précise la manière dont les références canoniques sont construites pour ce texte ;

*<classDecl>*

regroupe une ou plusieurs taxonomies définissant l'ensemble des codes de classification employés ailleurs dans le texte.

### 20.2.1. *Description du projet et des critères de sélection*

Exemples d'utilisation des éléments *<projectDesc>* et *<samplingDesc>* :

```
<encodingDesc>
  <projectDesc>Texte saisi d'après le fond du Musée d'Art Brut
    de Lausanne, Suisse, juillet 1996.
  </projectDesc>
</encodingDesc>

<encodingDesc>
  <samplingDecl>Samples of 2000 words taken from the
    beginning of the text
  </samplingDecl>
</encodingDesc>
```

### 20.2.2. *Mentions éditoriales*

L'élément *<editorialDecl>* regroupe une description en prose des pratiques employées lors du codage du texte. Typiquement, cette description devrait couvrir des sujets tels que ceux qui sont indiqués ci-après. Pour chacun de ces thèmes, il conviendrait de créer un paragraphe séparé :

correction

comment et dans quelles circonstances les corrections ont été apportées au texte ;

normalisation

le degré de régularisation ou de normalisation apporté à la source originale ;

**citation** le traitement réservé aux marques de citation dans l'original – ont-ils été retenus ou remplacés par des références d'entité? les guillemets ouvrants et fermants sont-ils différenciés, etc.;

**division des mots**

la façon dont les tirets (surtout ceux en fin de ligne) ont été traités dans l'original: ont-ils été retenus ou remplacés par des références d'entité?

**segmentation**

la façon dont le texte a été segmenté, par exemple en phrases, unités de ton, strates graphémiques, etc.;

**interprétation**

les informations interprétatives ou analytiques éventuellement ajoutées au texte.

Exemple :

```
<editorialDecl>
  <p> L'analyse du récit oral de la section 4 a été
    faite à la main et n'a pas été vérifiée ;
  <p> La transcription a été vérifiée à l'aide du
    « speller » d'Unix, mais avec le dictionnaire de
    l'Université de Paris-7 ;
  <p> Tous les mots américains sont écrits en suivant
    l'orthographe du Webster's 9th Collegiate dictionary.
  <p> Ici, nous utilisons le codage Iso-8859 (Latin1).
</editorialDecl>
```

### 20.2.3. *Déclarations relatives au balisage, au système de référence et à la classification utilisés*

L'élément `<tagsDecl>` est employé pour fournir des informations détaillées concernant les balises SGML paraissant effectivement à l'intérieur d'un texte. Il peut contenir une simple liste d'éléments à employer, avec un compteur pour chacun, au moyen des éléments suivants :

`<tagUsage>`

fournit des informations concernant l'usage d'un élément spécifique au sein de l'élément `<text>` le plus à l'extérieur d'un document conforme à la TEI; parmi les attributs citons :

*gi* le nom (identifiant générique) de l'élément indiqué par la balise;

*occurs* précise le nombre d'occurrences de cet élément au sein du texte.

L'élément `<rendition>` est employé pour documenter les manières différentes dont les éléments sont rendus dans le texte source.

`<rendition>`

fournit une information concernant l'interprétation prévue d'un ou plusieurs éléments ;

`<tagUsage>`

fournit une information concernant l'usage d'un élément spécifique à l'intérieur d'un élément `<text>` ; parmi les attributs possibles, citons :

*occurs* précise le nombre d'occurrences de cet élément au sein du texte ;

*ident* précise le nombre d'occurrences de cet élément au sein du texte qui portent une valeur distincte pour l'attribut global *id* ;

*render* précise l'identifiant d'un élément `<rendition>` qui définit comment cet élément doit être rendu.

Par exemple :

```
<tagsDecl>
  <tagUsage gi=text occurs=1>
  <tagUsage gi=body occurs=1>
  <tagUsage gi=p occurs=12>
  <tagUsage gi=hi occurs=6>
</tagsDecl>
```

Cette déclaration (imaginaire) de balisage conviendrait pour un texte contenant douze paragraphes dans son corps, parmi lesquels six éléments `<hi>` ont été marqués. À noter que si l'élément `<tagsDecl>` est employé, il doit contenir un élément `<tagUsage>` pour chaque élément balisé dans l'élément de texte associé.

L'élément `<refsDecl>` est employé pour documenter le fonctionnement d'un éventuel système de référence canonique incorporé dans le codage. Dans sa forme la plus simple, il consiste en une description en texte libre.

Exemple :

```
<refsDecl>
  <p>L'attribut N de chaque DV1 ou DV2 contient
    la référence canonique de chaque sous-division sous
    la forme XX.yyy où XX est le le numéro du tome, en
    nombres romaines, et yyy le numéro de section, en
    chiffres arabes.
</refsDecl>
```

L'élément `<classDecl>` regroupe les définitions pour, ou les sources de, tous les plans de classification descriptifs employés par d'autres parties de l'entête. Au moins un de ces plans doit être fourni, codé au moyen des éléments suivants :

```
<taxonomy>
  définit une typologie employée pour classer des textes soit implicite-
  ment, au moyen d'une citation bibliographique, soit explicitement
  au moyen d'une taxonomie structurée ;

<bibl>
  contient une citation bibliographique à structure informelle, dont les
  sous-composants peuvent être balisés explicitement ou pas ;

<category>
  regroupe une catégorie descriptive individuelle, éventuellement conte-
  nue dans une catégorie de niveau supérieur, à l'intérieur d'une taxo-
  nomie définie par l'utilisateur ;

<catDesc>
  décrit une certaine catégorie à l'intérieur d'une taxonomie ou d'une
  typologie de texte, sous forme d'une description brève en prose.
```

Dans le cas le plus simple, la taxonomie peut être définie par un référence bibliographique, comme l'exemple suivant :

```
<classDecl>
  <taxonomy id='LCSH'>
    <bibl>Library of Congress Subject Headings
    </bibl>
  </taxonomy>
</classDecl>
```

Alternativement, ou en outre, le codeur peut définir un plan de classification spécifique, comme dans l'exemple ci-après :

```

<taxonomy id=B>
  <bibl>Brown Corpus</bibl>
  <category id=B.A><catDesc>Press Reportage
    <category id=B.A1><catDesc>Daily</category>
    <category id=B.A2><catDesc>Sunday</category>
    <category id=B.A3><catDesc>National</category>
    <category id=B.A4><catDesc>Provincial</category>
    <category id=B.A5><catDesc>Political</category>
    <category id=B.A6><catDesc>Sports</category>
    ...
  </category>
  <category id=B.D><catDesc>Religion
    <category id=B.D1><catDesc>Books</category>
    <category id=B.D2><catDesc>Periodicals and
      tracts</category>
  </category>
  ...
</taxonomy>

```

Le lien entre un texte particulier et une catégorie à l'intérieur d'une telle taxonomie est réalisé au moyen de l'élément *<catRef>* incorporé dans l'élément *<textClass>*, comme décrit plus en détail ci-dessous.

### 20.3. Description de profil

L'élément *<profileDesc>* permet d'enregistrer dans un cadre unique des informations caractérisant divers aspects descriptifs d'un texte. Il a trois composants facultatifs :

- <creation>*  
contient des informations concernant la création d'un texte ;
- <langUsage>*  
décrit les langues, les sous-langues, les registres, les dialectes etc. représentés à l'intérieur d'un texte ;
- <textClass>*  
contient des informations décrivant la nature ou le sujet d'un texte en termes d'un système standard de classification, d'un thesaurus, etc.

Exemples :

```
<creation>
  <date value='1992-08'>August 1992</date>
  <name type=place>Taos, New Mexico</name>
</creation>
```

L'élément `<textClass>` classe un texte par référence au(x) système(s) défini(s) par l'élément `<classDecl>`. Il contient un ou plusieurs des éléments suivants :

`<keywords>`

contient une liste de mots clés ou de phrases clés identifiant le sujet ou la nature d'un texte ; parmi les attributs possibles, citons :

*scheme* identifie le vocabulaire contrôlé au sein duquel l'ensemble de mots clés concerné est défini ;

`<classCode>`

regroupe le code de classification employé pour ce texte dans un système standard de classification ; parmi les attributs possibles, citons :

*scheme* identifie le système de classification ou la taxonomie employés ;

`<catRef>`

précise une ou plusieurs catégories définies à l'intérieur d'une certaine taxonomie ou typologie de texte ; parmi les attributs possibles, citons :

*target* identifie les catégories concernées.

L'élément `<keywords>` regroupe une liste de mots clés ou d'expressions clés servant à identifier le sujet ou la nature d'un texte. L'attribut *scheme* relie ceux-ci au système de classification défini dans l'élément `<taxonomy>`.

```
<textClass>
  <keywords scheme=LCSH>
    <list>
      <item>Littérature française -- Histoire et critique --
        Bases de données.</item>
      <item>Littérature française -- Histoire et critique --
        Théorie, etc.</item>
      <item>Langue française -- Style -- Bases de
        données.</item>
    </list>
  </keywords>
</textClass>
```

## 20.4. Description de révision

L'élément `<revisionDesc>` fournit un journal de modifications dans lequel chaque modification apportée à un texte peut être enregistrée. Le journal peut être enregistré sous la forme d'une séquence d'éléments `<change>` dont chacun contient :

`<date>` contient une date dans n'importe quel format ;

`<respStmt>`

fournit une mention de responsabilité indiquant une personne responsable du contenu intellectuel d'un texte, d'une édition, d'un enregistrement ou d'une série, lorsque les éléments spécialisés pour définir les auteurs, les rédacteurs etc. ne suffisent pas ou ne s'appliquent pas ;

`<item>` contient un composant de liste.

Exemple :

```
<revisionDesc>
  <change><date>6/3/91:</date>
    <respStmt><name>JT</name><resp>ed.</resp>
    </respStmt>
    <item>Modifié le format de fichier</item>
  <change><date>5/25/90:</date>
    <respSmt><name>JT</name><resp>ed.</resp>
    <item>Traité les modids de Jean Sort</item>
</revisionDesc>
```

## 21 – Liste des éléments décrits

---

### 21.1. Attributs globaux

Tous les éléments employés dans la définition de type de document préconisée par la TEI Lite comportent les attributs globaux suivants (on trouvera, section 23 «Index», un index de ces termes par ordre alphabétique, renvoyant à leurs définitions) :

<i>ana</i>	relie un élément à son interprétation ;
<i>corresp</i>	relie un élément à un ou plusieurs autres éléments correspondants ;
<i>id</i>	identifiant unique d'un élément ; doit commencer avec une lettre, peut contenir des lettres, des chiffres, des tirets ou des points ;
<i>lang</i>	langue employée pour le texte dans cet élément ; si non précisée, la langue est supposée être la même que celle qui est employée dans le contexte immédiat ;
<i>n</i>	nom ou nombre de cet élément ; peut comporter toute chaîne de caractères ; souvent employé pour enregistrer des systèmes de référence traditionnels ;
<i>next</i>	relie un élément au prochain élément dans un agrégat ;
<i>prev</i>	relie un élément à l'élément précédent dans un agrégat ;
<i>rend</i>	aspect physique de l'élément dans la copie : <i>italic</i> , <i>roman</i> , <i>display block</i> , etc. ; la valeur peut être toute chaîne de caractères.

### 21.2. Éléments de la TEI Lite

La liste suivante donne tous les éléments définis pour la DTD, avec, dans chaque cas, une courte description (on trouvera, section 23 «index», un index de ces termes, anglais ou français, par ordre alphabétique renvoyant à leurs définitions) :

<*abbr*> contient une abréviation de n'importe quel type ; le développement peut être donné dans l'attribut *expn* ;

- 
- <add>* contient des lettres, des mots, ou des expressions insérées dans le texte par un auteur, scribe, annotateur, ou correcteur ;
- <address>*  
contient une adresse postale ou d'un autre type, par exemple celle d'un éditeur, d'une organisation, ou d'un individu ;
- <addrLine>*  
contient une ligne d'une adresse postale ou d'un autre type ;
- <anchor>*  
précise un emplacement ou un endroit à l'intérieur d'un document afin de permettre de pointer sur cet emplacement ;
- <argument>*  
une liste formelle ou une description en texte libre des sujets traités par une subdivision d'un texte ;
- <author>*  
dans un référence bibliographique, contient le nom de l'auteur ou des auteurs, soit personnel soit corporatif, d'un ouvrage ; constitue la mention de responsabilité principale pour toute unité bibliographique ;
- <authority>*  
fournit le nom d'une personne ou d'une autre agence responsable de la fourniture d'un fichier électronique, autre qu'un éditeur ou un distributeur ;
- <availability>*  
fournit des informations concernant la disponibilité d'un texte, par exemple toute restriction sur son emploi ou diffusion, son statut en matière de droits d'auteur, etc. ;
- <back>* contient tout appendice, etc., suivant la partie principale d'un texte ;
- <bibl>* contient une citation bibliographique structurée de façon informelle, et dont les sous-champs peuvent ou non être balisés explicitement ;
- <biblFull>*  
contient une citation bibliographique structurée de façon rigoureuse et contenant tous les composants de la description du fichier TEI ;
- <biblScope>*  
définit l'étendue d'une référence bibliographique, sous la forme d'une liste de numéros de page, ou d'un nom de chapitre ;

- <body>* contient le corps entier d'un texte unitaire unique, à l'exclusion de toute pièce liminaire ou annexe ;
- <byline>*  
contient la mention de responsabilité principale d'un ouvrage, telle qu'elle apparaît sur sa page de titre ou au début ou à la fin de cet ouvrage ;
- <catDesc>*  
décrit une certaine catégorie à l'intérieur d'une taxonomie ou typologie de texte, sous forme d'une courte description en texte libre ;
- <category>*  
contient une catégorie descriptive individuelle, éventuellement contenue dans une catégorie de niveau supérieur, le tout dans le cadre d'une taxonomie définie par l'utilisateur ;
- <catRef>*  
spécifie une ou plusieurs catégories définies à l'intérieur d'une certaine taxonomie ou typologie de texte ;
- <cell>* contient une cellule d'une table ;
- <cit>* citation empruntée à un autre document, accompagnée d'une référence bibliographique indiquant son origine ;
- <classCode>*  
contient le code de classification employé pour ce texte dans un système de classification normalisé et identifié par l'attribut *scheme* ;
- <classDecl>*  
contient un ou plusieurs taxonomies définissant tout code de classification employé ailleurs dans le texte ;
- <closer>*  
regroupe la date, la signature, la salutation et toute formule similaire pouvant apparaître en tant que groupe final à la fin d'une division, surtout dans le cas d'une lettre ;
- <code>* contient un court fragment de code dans un langage formel donné (souvent un langage de programmation) ;
- <corr>* contient la forme correcte d'un passage apparemment erroné dans la copie ;
- <creation>*  
contient des informations concernant la création d'un texte ;

- 
- <date>* contient une date dans n'importe quel format, avec la valeur normalisée dans l'attribut *value* ;
- <dateline>*  
contient une description brève de l'endroit, de la date et de l'heure de production d'une lettre, d'un article de journal ou de toute autre œuvre, s'y trouvant préfixé ou suffixé comme une sorte de titre ou bloc de fin ;
- <del>* contient une lettre, un mot ou un passage effacé, marqué comme effacé, ou désigné comme superflu ou parasite dans la copie, par un auteur, scribe, annotateur ou correcteur ;
- <distributor>*  
fournit le nom d'une personne ou d'une institution responsable de la diffusion d'un texte ;
- <div>* contient une subdivision des parties liminaires, du corps ou des annexes d'un texte ;
- <div1>...<div7>*  
contient une subdivision de niveau un à sept des parties liminaires du corps ou des annexes d'un texte ;
- <divGen>*  
indique l'emplacement où doit paraître une division textuelle générée automatiquement par une application de traitement texte; l'attribut *type* précise s'il s'agit d'un index, d'une table des matières, ou de quelque chose d'autre ;
- <docAuthor>*  
contient le nom de l'auteur du document, tel qu'il est indiqué sur la page de titre (souvent mais pas toujours contenu dans un élément *<byline>*) ;
- <docDate>*  
contient la date du document, telle qu'elle est donnée (habituellement) sur la page de titre ;
- <docEdition>*  
contient une mention d'édition telle quelle est présentée sur la page de titre d'un document ;
- <docImprint>*  
contient la mention d'impression (le lieu et la date de publication, le nom de l'éditeur), telle qu'elle apparaît (habituellement) dans le pied d'une page de titre ;

**<docTitle>**

contient le titre d'un document, y compris tous ses constituants, tels qu'ils apparaissent sur une page de titre ; doit être partagé en éléments *<titlePart>* ;

**<edition>**

décrit les particularités d'une édition donnée d'un texte ;

**<editionStmnt>**

contient des informations relatives à une édition d'un texte ;

**<editor>**

mention secondaire de responsabilité pour une unité bibliographique, par exemple le nom d'un individu, institution ou organisation, (ou de plusieurs de ceux-ci), agissant en tant que rédacteur, compilateur, traducteur, etc.

**<editorialDecl>**

fournit des détails sur les pratiques et les principes éditoriaux appliqués pendant le codage d'un texte ;

**<eg>**

contient un court et unique exemple d'un sujet technique en cours de discussion, par exemple, un fragment de code ou un échantillon du codage SGML ;

**<emph>**

marque des mots ou expressions qui sont accentuées ou mise en évidence pour produire un effet rhétorique ou linguistique ;

**<encodingDesc>**

documente le rapport entre un texte électronique et la ou les source(s) dont il a été dérivé ;

**<epigraph>**

contient une citation, anonyme ou attribuée, figurant au début d'une section ou d'un chapitre, ou dans une page de titre ;

**<extent>**

décrit la taille approximative du texte électronique tel que stocké sur un certain type de support ; cette taille peut être spécifiée dans n'importe quelle unité communément utilisée ;

**<figure>**

marque l'emplacement où un graphique doit être inséré dans un document ; les attributs peuvent servir pour indiquer une entité SGML contenant l'image elle-même (dans une quelconque notation non-SGML) ; les paragraphes à l'intérieur de l'élément *<figure>* peuvent servir pour transcrire les légendes ;

- 
- <fileDesc>** contient une description bibliographique complète d'un fichier électronique ;
- <foreign>** identifie un mot ou une expression écrits dans une langue différente de celle du texte avoisinant ;
- <formula>** contient une formule chimique ou mathématique, éventuellement présentée dans une quelconque notation non-SGML. L'attribut *notation* est employé pour nommer la notation non-SGML utilisée pour transcrire la formule ;
- <front>** contient tout texte liminaire (en-tête, page de titre, préfaces, dédicaces, etc.) se trouvant avant le début du texte proprement dit ;
- <funder>** précise le nom d'un individu, d'une institution, ou d'une organisation responsable du résultat d'un projet ou d'un document ;
- <gap>** indique un endroit où du texte a été omis dans une transcription, soit pour des raisons éditoriales décrites dans l'en tête de la TEI et ayant trait à la méthode d'échantillonnage, soit parce que le texte est illisible ou inaudible ;
- <gi>** contient un type d'identifiant spécial : un identifiant générique SGML, ou nom d'élément ;
- <gloss>** marque un mot ou une expression qui fournit une glose ou une définition pour un autre mot ou une autre expression ;
- <group>** contient plusieurs textes unitaires ou groupes de textes ;
- <head>** contient tout type de titre, comme par exemple, le titre d'une section, ou l'en-tête d'une liste ou d'un glossaire ;
- <hi>** marque un mot ou une expression comme étant graphiquement distincte du texte avoisinant, sans aucune interprétation quant à la raison de cette mise en valeur ;
- <ident>** contient un identifiant d'un certain type, par exemple, un variable de nom ou le nom d'un élément ou attribut SGML ;

- <*idno*> fournit tout numéro, normalisé ou non, employé pour identifier une unité bibliographique; le système ou la norme mis en œuvre sont identifiés par l'attribut *type*;
- <*imprint*>  
regroupe des information relatives à la publication ou la diffusion d'un unité bibliographique;
- <*index*>  
marque un passage devant être indexé pour une raison donnée. Des attributs sont employés pour donner la forme principale ainsi que les formes des deuxième, troisième et quatrième niveaux qui doivent être inclus dans l'index indiqué;
- <*interp*>  
fournit une annotation interprétative qui peut être reliée à un passage de texte; parmi les attributs citons: *resp*, *type* et *valeur*;
- <*InterpGrp*>  
recueille l'ensemble des balises <*interp*>;
- <*item*> contient un seul composant de liste;
- <*keywords*>  
regroupe une liste de mots ou d'expressions servant à identifier le sujet ou la nature d'un texte; si ces mots clés proviennent d'un vocabulaire contrôlé, ce dernier peut être identifié au moyen de l'attribut *scheme*;
- <*kw*> contient un mot clé noté dans un certain langage formel;
- <*l*> contient une ligne de vers unique (éventuellement incomplète);
- <*label*> contient l'étiquette associée à un élément d'une liste; dans les glossaires, marque le terme qui est défini;
- <*langUsage*>  
décrit les langues, les sous-langues, les registres, les dialectes, etc., représentés à l'intérieur un texte;
- <*lb*> marque le début d'une nouvelle ligne (au sens typographique) dans une certaine édition ou version d'un texte;
- <*lg*> contient un groupe de lignes en vers, constituant une unité formelle; par exemple, une strophe, un refrain, un paragraphe de vers, etc.;

- 
- <*list*> contient toute séquence d'éléments organisés sous forme de liste, celle-ci pouvant être soit numérotée, soit identifiée par des puces ou tirets ;
- <*listBibl*>  
contient une liste de citations bibliographiques de toute sorte ;
- <*mentioned*>  
marque des mots ou expressions mentionnés mais non employés ;
- <*milestone*>  
marque la limite entre des sections d'un texte, comme cela peut être indiqué par des changements dans un système de référence standard ; parmi les attributs possibles citons : *ed* (édition), *unit* (page), etc., et *n* (nouvelle valeur) ;
- <*name*>  
contient un nom propre ou une expression substantive ; les attributs peuvent indiquer son type, donner une forme normalisée, ou l'associer à une chose ou à un individu spécifique, au moyen d'identifiants uniques ;
- <*note*> contient une note ou une annotation, avec des attributs pour indiquer le type, l'emplacement et la source de la note ;
- <*notesStmt*>  
regroupe toute les notes fournissant des informations sur le texte, en complément à celles enregistrées ailleurs dans la description bibliographique ;
- <*num*> contient un nombre, écrit sous une forme quelconque, avec une valeur normalisée dans l'attribut *valeur* ;
- <*opener*>  
contient la date, la signature, la salutation et les autres expressions similaires susceptibles d'apparaître au début d'une division, surtout dans une lettre ;
- <*orig*> contient la forme originale d'une lecture, pour laquelle une forme régularisée peut être donnée dans l'attribut *reg* ;
- <*p*> marque les paragraphes écrits en prose ;
- <*pb*> (*pb=page break*) marque la limite entre une page d'un texte et la suivante, dans un système de référence normalisé ;

- <principal>  
fournit le nom du chercheur principal responsable de la création d'un texte électronique ;
- <profileDesc>  
fournit une description détaillée des aspects non bibliographiques d'un texte, spécifiquement les langues et les sous-langues employées, les circonstances de sa production, les participants, et leur environnement ;
- <projectDesc>  
décrit dans le détail l'objectif ou but pour lequel un fichier électronique a été codé, ainsi que toute autre information pertinente concernant le procédé par lequel il a été constitué ou recueilli ;
- <ptr> un pointeur vers un autre emplacement dans le document courant, en termes d'un ou plusieurs éléments identifiables ;
- <publicationStmt>  
regroupe des informations concernant la publication ou la diffusion d'un texte qu'il soit électronique ou non ;
- <publisher>  
fournit le nom de l'organisation responsable de la publication ou de la diffusion d'une unité bibliographique ;
- <pubPlace>  
contient le nom du lieu où une unité bibliographique a été publiée ;
- <q> contient une citation ou quelque chose qui s'apparente à une citation ;
- <ref> une référence à un autre emplacement dans le document courant, en termes d'un ou plusieurs éléments identifiables, modifiée éventuellement par du texte ou des commentaires additionnels ;
- <refsDecl>  
précise comment les références canoniques sont construites pour ce texte ;
- <reg> contient une lecture qui a été régularisée ou normalisée d'une certaine manière; la lecture originale peut être donnée dans l'attribut *orig* ;
- <rendition>  
fournit l'information concernant la présentation prévue pour un ou plusieurs éléments ;

- 
- <resp> contient une expression décrivant la nature de la responsabilité intellectuelle d'une personne ;
- <respStmt>  
fournit une mention de responsabilité indiquant une personne responsable du contenu intellectuel d'un texte, d'une édition, d'un enregistrement ou d'une série, lorsque les éléments spécialisés pour définir les auteurs, les rédacteurs etc. ne suffisent pas ou ne s'appliquent pas ;
- <revisionDesc>  
résume l'historique de révision d'un fichier.
- <row> contient une rangée d'une table ;
- <rs> contient un nom ou une chaîne de référence d'usage général ; les attributs peuvent indiquer son type, donner une forme normalisée, ou l'associer avec une chose ou un individu spécifique, au moyen d'un identifiant unique ;
- <s> identifie un *s-unit* à l'intérieur d'un document, afin d'établir un simple plan de référence canonique s'appliquant au texte entier ;
- <salute>  
contient une salutation placée au début d'un avant-propos, d'une épître dédicatoire ou d'une autre division d'un texte, ou la salutation apparaissant à la fin d'une lettre ou d'une préface, etc. ;
- <samplingDecl>  
contient une description en texte libre des critères ainsi que des méthodes employés pour le prélèvement des textes lors de la création d'un corpus ou d'un recueil ;
- <seg> identifie un passage ou un segment de texte à l'intérieur un document afin d'en permettre le pointage; l'attribut *type* catégorise le segment ;
- <series>  
contient une information concernant la série dans laquelle un livre ou une autre unité bibliographique a paru ;
- <seriesStmt>  
regroupe des informations concernant la série, si elle existe, à laquelle une publication appartient ;
- <sic> contient du texte reproduit tel quel bien qu'apparemment incorrect ou inexact ;

<signed>

contient la salutation finale, etc., apposée à un avant-propos, une épître dédicatoire, ou toute autre division textuelle ;

<soCalled>

contient un mot ou une expression vis-à-vis desquels l'auteur ou le narrateur décline toute responsabilité, ce qui est en général indiqué par l'utilisation des guillemets ou de l'italique ;

<sourceDesc>

fournit une description bibliographique du texte original à partir duquel un texte électronique a été dérivé ou généré ;

<sp>

contient un discours individuel dans un texte théâtral, ou un passage présenté comme tel dans un texte en vers ou en prose ; l'attribut *who* sert à identifier le personnage qui parle ;

<speaker>

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

<sponsor>

précise le nom d'une institution ou organisation de parrainage ;

<stage>

contient toute sorte d'indication scénique à l'intérieur d'une pièce ou d'un fragment de pièce de théâtre ;

<table>

contient du texte affiché sous forme tabulaire en rangées et en colonnes ;

<tagsDecl>

fournit une information détaillée concernant le balisage appliqué à un document SGML ;

<tagUsage>

fournit des informations concernant l'usage d'un élément spécifique au sein de l'élément <text> le plus extérieur d'un document conforme à la TEI ;

<taxonomy>

définit une typologie employée pour classer des textes soit implicitement, au moyen d'une citation bibliographique, soit explicitement au moyen d'une taxonomie structurée ;

<term>

contient un mot unique, plusieurs mots ou une désignation symbolique considérée comme un terme technique ;

- 
- <textClass>* regroupe des informations décrivant la nature ou le sujet d'un texte en termes d'un plan standard de classification, d'un thesaurus, etc. ;
- <time>* contient une expression définissant l'heure du jour, dans n'importe quel format, avec une valeur normalisée dans l'attribut *value* ;
- <title>* contient le titre d'une œuvre, que ce soit un article, livre, journal, ou une série, y compris tout sous-titre ou titre alternatif ;
- <titlePage>* contient la page de titre d'un texte, tel que figurant à l'intérieur des parties liminaires ou des annexes ;
- <titlePart>* contient une subdivision ou division du titre d'une œuvre, comme indiqué sur une page de titre; également employé pour les fragments « flottants » de la page de titre ne faisant pas partie du titre de document, comme par exemple le nom de l'auteur ;
- <titleStmt>* regroupe des informations concernant le titre d'une œuvre, ainsi que les personnes responsables de son contenu intellectuel ;
- <trailer>* contient un bas de page ou un titre final paraissant à la fin d'une division de texte ;
- <unclear>* contient un mot, une expression, ou un passage qui ne peut pas être transcrit avec certitude car illisible ou inaudible dans la source ;
- <xptr>* définit un pointeur vers un autre endroit du document courant ou d'un document externe ;
- <xref>* définit un pointeur vers un autre endroit du document courant ou d'un document externe, modifié éventuellement par un commentaire ou un texte supplémentaire ;

---

## 22 – Références bibliographiques

---

Cette section contient une liste de références bibliographiques concernant les œuvres traitant de SGML et les sujets connexes<sup>1</sup> ; elle sert également à montrer l'emploi de l'élément `<bibl>` présenté dans la section 13 « Citations bibliographiques » ci-dessus.

On trouvera donc la bibliographie elle-même suivie du texte TEI.

### Bibliographie

- [1] ALA (American Library Association), *ALA-LC Romanization Tables: Transliteration Schemes for Non-Roman Scripts*, approved by the Library of Congress and the American Library Association, tables compiled and edited by Randall K. Barry. Washington: Library of Congress, 1991.
- [2] ANSI (American National Standards Institute), *ANSI X3.4-1986. American National Standard for Information Systems — Coded Character Sets — 7-bit American National Standard Code for Information Interchange (7-bit ASCII)*. [New York]: ANSI, 1986.
- [3] BARNARD, David, *et al.*, «SGML-Based Markup for Literary Texts», *Computers and the Humanities*, n° 22 (1988), p. 265-76.
- [4] BARRON, David, «Why use SGML?», *Electronic Publishing Origination, Dissemination and Design*, n° 2.1 (April 1989), p. 3-24.
- [5] COOMBS, James H., Allen H. RENEAR, and Steven J. DEROSE, «Markup Systems and the Future of Scholarly Text Processing», *Communications of the ACM* n° 30.11 (November 1987), p. 933-947.
- [6] COVER, Robin C. (ed.), *et al.*, *A Bibliography on Structured Text: Technical Report 90-281* Queen's University, Kingston, Ont., June 1990. Une version tenue à jour de cette bibliographie se trouve à <http://www.sil.org/sgml/sgml.html>
- [7] GOLDFARB, Charles F., *The SGML Handbook*, Oxford: Clarendon Press, 1990.

---

1. À la liste anglaise initiale, nous avons ajouté quelques références françaises {NdT}.

- 
- [8] GOOSSENS Michel, «Introduction pratique à SGML», *Cahiers GUTenberg*, n° 19, janvier 1995, p. 27–58.
- [9] VAN HERWIJNEN, Eric, *Practical SGML*, Kluwer Academic Publishers, 1990; 2d ed. 1994. Version française : *SGML pratique*, International Thomson Publishing France, 1995.
- [10] ISO (International Organization for Standardization). *ISO 8859-1: 1987 (E). Information processing — 8-bit Single-Byte Coded Graphic Character Sets — Part 1: Latin Alphabet No. 1. (Traitement de l'information — Jeux de caractères graphiques codés sur un seul octet — Partie 1: Alphabet latin no 1.)* First edition — 1987-02-15. [Geneva]: International Organization for Standardization, 1987.
- [11] ISO (International Organization for Standardization). *ISO 8879-1986 (E). Information processing — Text and Office Systems — Standard Generalized Markup Language (SGML)*. First edition — 1986-10-15. [Geneva]: International Organization for Standardization, 1986.
- [12] ISO (International Organization for Standardization). *ISO 8879:1986 /A1: 1988 (E). Information processing — Text and Office Systems — Standard Generalized Markup Language (SGML), Amendment 1*. Published 1988-07-01. [Geneva]: International Organization for Standardization, 1988.
- [13] ISO (International Organization for Standardization). *ISO/TR 9573-1988(E). Information processing—SGML support facilities—Techniques for using SGML*. Final text of 1988-09-12.
- [14] ISO (International Organization for Standardization), and IEC (International Electrotechnical Commission). *ISO/IEC 10646-1: 1993. Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane*. [Geneva]: International Organization for Standardization, 1993.
- [15] ISO (International Organization for Standardization), and IEC (International Electrotechnical Commission). *ISO/IEC 10744: 1992. Information Technology — Hypermedia/Time-based Structuring Language (HyTime)*. [Geneva]: International Organization for Standardization, 1992.
- [16] LANGENDOEN, D. Terence, and Gary F. SIMONS, «A Rationale for the TEI Recommendations for Feature-Structure Markup», *Computers and the Humanities* (1995; in press).
- [17] ROLE, François, «La norme SGML : pour décrire la structure logique des documents», *Documentalistes – Sciences de l'information*, n° 2.8, (juillet-août 1991), p. 140–145.

- [18] VIGNAUD Dominique, *L'édition structurée des documents*, Editions du cercle de la librairie (1990).
- [19] WARMER, J., and S. VAN EGMOND, «The implementation of the Amsterdam SGML parser», *Electronic Publishing Origination, Dissemination and Design* n° 2.2 (July 1989), p. 65-90.

## Codage TEI de la bibliographie

```
<listBibl>
```

```
<bibl>ALA (American Library Association). <title>ALA-LC  
Romanization Tables: Transliteration Schemes for Non-Roman  
Scripts</title>, approved by the Library of Congress and the  
American Library Association, tables compiled and edited by  
Randall K. Barry. Washington: Library of Congress, 1991.  
</bibl>
```

```
<bibl>ANSI (American National Standards Institute). <title>ANSI  
X3.4-1986. American National Standard for Information Systems  
--- Coded Character Sets --- 7-bit American National Standard  
Code for Information Interchange (7-bit ASCII).</title>  
[New York]: ANSI, 1986.  
</bibl>
```

```
<bibl>  
<author>Barnard, David, et al.</author>  
<title level=a>SGML-Based Markup for Literary Texts.</title>  
<title>Computers and the Humanities</title>  
<biblScope>22 (1988): 265-76.</biblScope>  
</bibl>
```

```
<bibl>  
  <author>Barron, David</author>  
  <title level=a>Why use SGML?</title>  
  <title>Electronic Publishing  
    Origination, Dissemination and Design</title>  
  <biblScope>2.1 (April 1989): 3-24.</biblScope>  
</bibl>
```

```
<bibl>  
<author>Coombs, James H., Allen H. Renear, and Steven J.  
DeRose.</author> <title level=a>Markup Systems and the Future  
of Scholarly Text Processing.</title> <title>Communications  
of the ACM</title>
```

<biblScope>30.11 (November 1987): 933-947.</biblScope>  
</bibl>

<bibl>  
 <editor>Cover, Robin C., et al.</editor>  
 <title>A Bibliography on Structured Text:  
 Technical Report 90-281</title>  
 <publisher>Queen's University,</publisher>  
 <pubPlace>Kingston, Ont.</pubPlace>  
 <date>June 1990</date>  
 <note place=inline>Une version tenue à jour de cette  
 bibliographie se trouve à  
 <code><http://www.sil.org/sgml/sgml.html></code>.  
 </note>  
 </bibl>

<bibl>  
 <author>Goossens, Michel.</author>  
 <title level=a>Introduction pratique à SGML.</title>  
 <title> Cahiers GUTenberg.</title>  
 <biblScope>19 (Janvier 1995): 27-58.</biblScope>  
 </bibl>

<bibl>Goldfarb, Charles F., <title>The SGML Handbook.</title>  
 Oxford: Clarendon Press, 1990.</bibl>

<bibl>  
 <author>van Herwijnen, Eric.</author>  
 <title>Practical SGML.</title>  
 <publisher>Kluwer Academic Publishers</publisher>  
 <date>1990; 2d ed. 1994.</date>  
 <note place=inline> Version française~: SGML pratique,  
 International Thomson Publishing France, 1995.  
 </note>  
 </bibl>

<bibl>ISO (International Organization for Standardization).  
 <title>ISO 8859-1: 1987 (E). Information processing ---  
 8-bit Single-Byte Coded Graphic Character Sets --- Part 1:  
 Latin Alphabet No. 1.</title> (<title>Traitement de  
 l'information --- Jeux de caractères graphiques  
 codés sur un seul octet --- Partie 1: Alphabet latin  
 no 1.</title>) First edition --- 1987-02-15.  
 [Geneva]: International Organization for Standardization,  
 1987.  
 </bibl>

<bibl>ISO (International Organization for Standardization).

<title>ISO 8879-1986 (E). Information processing --- Text and Office Systems --- Standard Generalized Markup Language (SGML).</title> First edition --- 1986-10-15. [Geneva]: International Organization for Standardization, 1986.</bibl>

<bibl>ISO (International Organization for Standardization).<title>ISO 8879:1986 / A1:1988 (E). Information processing --- Text and Office Systems --- Standard Generalized Markup Language (SGML), Amendment 1.</title> Published 1988-07-01. [Geneva]: International Organization for Standardization, 1988.</bibl>

<bibl>ISO (International Organization for Standardization).<title>ISO/TR 9573-1988(E). Information processing---SGML support facilities---Techniques for using SGML.</title> Final text of 1988-09-12.</bibl>

<bibl>ISO (International Organization for Standardization), and IEC (International Electrotechnical Commission). <title>ISO/IEC 10646-1: 1993. Information technology --- Universal Multiple-Octet Coded Character Set (UCS) --- Part 1: Architecture and Basic Multilingual Plane.</title> [Geneva]: International Organization for Standardization, 1993.</bibl>

<bibl>ISO (International Organization for Standardization), and IEC (International Electrotechnical Commission).<title>ISO/IEC 10744: 1992. Information Technology --- Hypermedia/Time-based Structuring Language (HyTime).</title> [Geneva]: International Organization for Standardization, 1992.</bibl>

<bibl>Langendoen, D. Terence, and Gary F. Simons.<title level=a>A Rationale for the TEI Recommendations for Feature-Structure Markup.</title><title>Computers and the Humanities</title> (1995; in press).</bibl>

<bibl><author>Vignaud, Dominique.</author>

---

```
<title>L'édition structurée des documents.</title>
<publisher>Editions du cercle de la librairie.</publisher>
<date>1990.</date>
</bibl>

<bibl>
  <author>Warmer, J., and S. van Egmond</author>
  <title level=a>The implementation of the Amsterdam
    SGML parser.</title>
  <title>Electronic Publishing
    Origination, Dissemination and Design</title>
  <biblScope>2.2 (July 1989): 65-90.</biblScope>
</bibl>

</listBibl>
```

## 23 – Index

---

- <abbr>, 79, 128
- abbreviated, 53
- abréviations, 79
- accents, 104
- acronym, 79
- acronymes, 80
- acute, 106
- <add>, 71, 129
- <address>, 80, 118, 129
- <addrLine>, 80, 129
- adresse http de ce document, 24
- adresses postales, 80
- ajouts, 71
- allusion, 94
- alt, 108
- alternances typographiques, 52
- ana, 67, 128
- analyse, 92
- ancestor, 65
- <anchor>, 62, 129
- anchored, 58
- annexes, 112
- annotations, 58
- anthologie, 39
- <argument>, 112, 129
- art dramatique, 46
- article, 82
- attributs de liaison, 67
- attributs globaux, 128
- <author>, 116, 129
- author, 58
- <authority>, 118, 129
- <availability>, 119, 129
  
- <back>, 42, 129
- <bibl>, 119, 124, 129
- <biblFull>, 120, 129
- bibliographie sur la TEI, 140
- <biblScope>, 129
- <body>, 42, 130
- bottom, 71
- brevigraph, 79
- bulleted, 82
- <byline>, 109, 111, 130
  
- calendar, 77
- caractère, 94
- caractères, 104
- caractères «dangereux», 106
- caractères français, 35, 107
- cardinal, 79
- <catDesc>, 124, 130
- <category>, 124, 130
- <catRef>, 126, 130
- <cell>, 87, 130
- cert, 69
- césure des mots, 50
- cgm, 90
- chaîne de référence, 75
- chapitre, 43
- child, 65
- circumflex, 106
- <cit>, 112, 130
- citations, 54
- <classCode>, 126, 130
- <classDecl>, 121, 130
- <closer>, 112, 130
- codage, 120
- <code>, 98, 130
- collection, 119
- cols, 87
- composant d'une liste, 82

- consonnes, 106  
 contraction, 79  
 corps du texte, 42  
 corpus TEI, 40  
 <corr>, 69, 130  
 corr, 69  
 correction, 69  
 corresp, 67, 128  
 crDate, 60  
 <creation>, 125, 130
- data, 87  
 <date>, 77, 119, 127, 131  
 <dateline>, 111, 131  
 dates, 77  
 déclaratif, 92  
 <del>, 71, 131  
 desc, 71  
 desc, 108  
 diacritiques, 104  
 diagrammes, 89  
 dialogue, 49  
 diction, 48  
 digrammes, 105  
 display block, 128  
 <distributor>, 118, 131  
 <div>, 42, 131  
 <div1>, 42  
 <div1>...<div7>, 131  
 <divGen>, 101, 131  
 division des mots, 50  
 divisions, 101  
 divisions de texte, 42  
 doc, 63  
 <docAuthor>, 109, 131  
 <docDate>, 109, 131  
 <docEdition>, 109, 131  
 <docImprint>, 109, 131  
 <docTitle>, 108, 132  
 documentation technique, 98
- ed, 50, 51
- <edition>, 117, 132  
 édition, 69, 117  
 <editionStmt>, 115, 132  
 <editor>, 132  
 editor, 58  
 <editorialDecl>, 121, 132  
 effacements, 71  
 effet linguistique, 53  
 effet rhétorique, 53  
 <eg>, 98, 132  
 élément, 75  
 éléments de la TEI, 128  
 éléments associés, 54  
 <emph>, 52, 132  
 emphase, 52  
 en-tête, 39  
 <encodingDesc>, 114, 132  
 end, 58  
 endroit, 75  
 ensembles, 65  
 entity, 89  
 entrée, 48  
 <epigraph>, 109, 132  
 épître, 111  
 erreur, 69  
 étiquette, 82  
 expan, 79  
 expressions étrangères, 56  
 expressions mises en valeur, 52  
 <extent>, 115, 132
- F, 47  
 <figDesc>, 89  
 figlist, 101  
 <figure>, 89, 132  
 figures, 89  
 figures rhétoriques, 94  
 <fileDesc>, 114, 133  
 fin de chapitre, 46  
 fin de ligne, 50  
 fin de page, 50  
 following, 65

- fontes, 52  
foot, 58  
<foreign>, 53, 56, 133  
<formula>, 99, 133  
formule, 99  
fraction, 79  
français, 35, 107  
free, 119  
from, 64  
<front>, 42, 46, 133  
<funder>, 116, 133  
  
<gap>, 71, 133  
génération d'index, 102  
géographique, 79  
<gi>, 98, 133  
gi, 122  
glose, 55  
<gloss>, 55, 133  
gloss, 82  
glossaire, 82  
graphiques, 89  
gras, 52  
grave, 106  
<group>, 42, 133  
guillemets, 54, 106  
  
hand, 72  
<head>, 133  
heures, 77  
<hi>, 52, 133  
  
I, 47  
id, 43, 128  
<ident>, 98, 133  
ident, 123  
<idno>, 118, 134  
illustrations, 89  
image, 94  
images, 89  
<imprint>, 134  
index, 102  
<index>, 102, 134  
  
index, 102  
index, 101  
indication scénique, 48  
infralinear, 71  
inline, 58, 71  
inst, 94  
interlinear, 58  
<interp>, 94, 134  
<InterpGrp>, 94, 134  
interprétation, 92  
interrogatif, 92  
interventions éditoriales, 69  
italic, 128  
italique, 52  
<item>, 82, 127, 134  
  
jalons, 51  
jargon, 80  
jeux de caractères, 104  
jpeg, 90  
  
key, 76  
<keywords>, 126, 134  
<kw>, 98, 134  
  
<l>, 47, 134  
<label>, 82, 134  
label, 87  
lang, 56, 128  
langue étrangère, 53  
<langUsage>, 125, 134  
<lb>, 50, 134  
left, 58, 71  
lettre accentuée, 104  
level, 53  
level1, 102  
level2, 102  
level3, 102  
level4, 102  
<lg>, 47, 134  
liens, 60  
lignes, 50  
liminaires (pièces), 108

- line break*, 50  
 <list>, 82, 135  
 <listBibl>, 120, 135  
 listes, 82  
 livre, 43
- M, 47  
 main, 53, 108  
 <mentioned>, 53, 55, 135  
 <milestone>, 51, 135  
 mise en scène, 48  
 mises en valeur, 52  
 mots clés, 126  
 mots étrangers, 56
- N, 47  
 n, 43, 128  
 <name>, 75, 117, 135  
 navire, 75  
 next, 65, 67, 128  
 nombres, 78  
 noms, 75  
 normalisation, 69  
 notation, 99  
 <note>, 58, 135  
 notes, 58, 119  
 notes marginales, 58  
 <notesStmt>, 115, 135  
 <num>, 79, 135  
 numérotation, 45  
 numérotation des lignes, 50  
 numérotation des pages, 50
- occurs*, 123  
 œuvre, 39  
 omissions, 71  
 <opener>, 112, 135  
 ordered, 82  
 ordinal, 79  
 organisation, 79  
 <orig>, 69, 135  
 orig, 70
- <p>, 42, 135  
*page break*, 50  
 page de titre, 108  
 page de titre électronique, 114  
 pages, 50  
 pagination, 50  
 paragraphes, 42  
 parallel, 53  
*part*, 47  
 pas, 64  
 <pb>, 50, 135  
 personne, 75  
 phrases orthographiques, 92  
 pièces annexes, 112  
 pièces liminaires, 108  
*place*, 58, 71  
 poème, 43  
 pointeur, 60  
 pointeurs étendus, 63  
 polices, 52  
 ponctuation, 92  
 pourcentage, 79  
*preceding*, 65  
 présentation physique, 53  
*prev*, 67, 128  
*previous*, 65  
 <principal>, 116, 136  
 profil, 125  
 <profileDesc>, 114, 136  
 <projectDesc>, 120, 136  
 prose, 46  
 <ptr>, 60, 136  
 publication, 118  
 <publicationStmt>, 115, 136  
 <publisher>, 118, 136  
 <pubPlace>, 118, 136
- <q>, 54, 136
- reason*, 72  
 recueil, 39  
 <ref>, 60, 136  
 références bibliographiques, 140

- références croisées, 60  
*referring string*, 75  
 <*refsDecl*>, 121, 136  
 <*reg*>, 70, 136  
*reg*, 69, 76  
*rend*, 52, 128  
*render*, 123  
 <*rendition*>, 123, 136  
 réplique, 47  
 <*resp*>, 117, 137  
*resp*, 58, 60, 69–72, 94  
 <*respStmt*>, 116, 117, 127, 137  
 restricted, 119  
 révision, 127  
 <*revisionDesc*>, 114, 137  
*right*, 58, 71  
*role*, 87  
 roman, 128  
 roupe, 43  
 <*row*>, 87, 137  
*rows*, 87  
 <*rs*>, 75, 137  
  
 <*s*>, 92, 137  
*s-unit*, 92  
 <*salute*>, 111, 137  
 <*samplingDecl*>, 120, 137  
*scheme*, 126  
 <*seg*>, 62, 137  
 <*series*>, 137  
 séries, 119  
 <*seriesStmt*>, 115, 137  
 SGML, 27, 140  
 <*sic*>, 69, 137  
*sic*, 69  
 <*signed*>, 111, 138  
 signes de ponctuation, 106  
 signes diacritiques et accents, 105  
 simple, 82  
 <*soCalled*>, 55, 138  
 song, 43  
 sonnet, 43  
  
 sortie, 48  
 source, 119  
 <*sourceDesc*>, 115, 138  
 sous-titre, 53  
 <*sp*>, 47, 138  
 <*speaker*>, 47, 138  
 speech, 43  
 spoken, 55  
 <*sponsor*>, 116, 138  
 <*stage*>, 48, 138  
*status*, 72, 119  
*step*, 64  
 structure d'un texte, 39  
 styles de caractères, 52  
 sub, 108  
 subdivisions, 42  
 subordinate, 53  
 superscription, 79  
 supralinear, 71  
 suspension, 79  
  
 <*table*>, 87, 138  
 table de matières, 101  
 tableau simple, 83  
 tableaux, 87  
 tables, 87  
 tablist, 101  
 <*tagsDecl*>, 121, 138  
 <*tagUsage*>, 122, 123, 138  
*target*, 55, 58, 60, 126  
*targetEnd*, 58  
*targType*, 60  
 <*taxonomy*>, 124, 138  
 <*teiCorpus*>, 40  
 <*teiHeader*>, 39  
 <*term*>, 53, 138  
 tête de chapitre, 46  
 T<sub>E</sub>X, 99  
 tex, 99  
 <*text*>, 39  
 <*textClass*>, 125, 139  
 texte français, 35

---

texte TEI, 39  
texte unitaire, 39  
thème, 94  
thought, 55  
tiff, 90  
tilde, 106  
<time>, 77, 139  
<title>, 53, 116, 139  
<titlePage>, 108, 139  
<titlePart>, 108, 139  
<titleStmt>, 115, 139  
titre, 53, 116  
titre, 79  
to, 64  
toc, 101  
top, 71  
<trailer>, 46, 139  
transcription du texte, 120  
trema, 105  
trou, 71  
type, 43, 48, 53, 54, 58, 60, 62, 71, 75, 79,  
82, 92, 94, 101, 108, 118  
typographie, 52  
typographie française, 35  
  
umlaut, 105  
<unclear>, 72, 139  
unit, 51  
unknown, 119  
  
valeur, 94  
value, 77, 79  
vers, 46  
vide, 71  
  
who, 47, 55  
  
<xptr>, 63, 139  
<xref>, 63, 139  
  
y, 47