

**Cahier des clauses techniques particulières**  
**Objet : dématérialisation**  
**du *Glossarium mediae et infimae latinitatis* (le Du Cange)**  
**Annexe 1 : spécifications pour l'encodage en XML**  
**des lots de type dictionnaire**

Nota : Aucune utilisation ou diffusion du présent document, sous quelque forme que ce soit, hors des locaux de la société destinataire, n'est autorisée sans l'accord préalable de l'École nationale des Chartes.

**Table des matières**

Introduction.....	2
Observables.....	3
Balises.....	12
Schéma.....	21

## Introduction

---

La structuration XML demandée a pour objet de conserver la totalité des caractères, toute la mise en forme typographique visible, en laissant le plus de repères stables pour établir les fonctions lexicographiques des informations encodées. L'espace de noms est un sous-ensemble du standard [TEI P5](#)<sup>1</sup>, incluant notamment l'appareil d'encodage des dictionnaires imprimés<sup>2</sup>. Cette étape est un préalable à des traitements ultérieurs visant à former une édition électronique (nomenclature, liens, recherche avancée...). L'objectif visé explique le choix des noms d'éléments et certaines précautions de balisage. Mais l'étape décrite ici ne demande pas d'interprétation ou de connaissance du latin.

Le présent document est organisé selon les sections suivantes :

- **Observables** : présentation dans un ordre logique des éléments typographiques repérables, avec indication de l'élément XML correspondant.
- **Balises** : documentation du schéma XML dans l'ordre alphabétique des éléments, avec des exemples de balisage d'articles complets.
- **Schéma** : grammaire des balises XML en langage DTD.

Il est abondamment illustré d'images du dictionnaire et d'exemples XML. Ces images ne sont absolument pas significatives du niveau de qualité souhaité pour la numérisation ; par contre, les exemples de code sont normatifs.

---

1 TEI P5: Guidelines for Electronic Text Encoding and Interchange

<http://www.tei-c.org/release/doc/tei-p5-doc/html/>.

2 TEI P5: Guidelines for Electronic Text Encoding and Interchange, 9 Print Dictionaries

<http://www.tei-c.org/release/doc/tei-p5-doc/html/DI.html>.





## Observables

Cette section énumère les éléments typographiques observables qui doivent être balisés. L'ordre d'exposition de cette section est pensé comme une suite d'opérations en commençant par les plus automatisables, garantissant des bornes fiables pour ce qui suit. Cette procédure a été testée sur un échantillon de 5% du dictionnaire (240 pages). Le prestataire est évidemment libre de procéder comme il l'entend selon les outils de sa chaîne de traitement, pourvu que soit compris ce qui nous importe. Le problème critique concerne le balisage des citations qui paraissent en italique. Elles sont essentielles à la valorisation de ce dictionnaire, sans le bruit d'autres segments en italique.

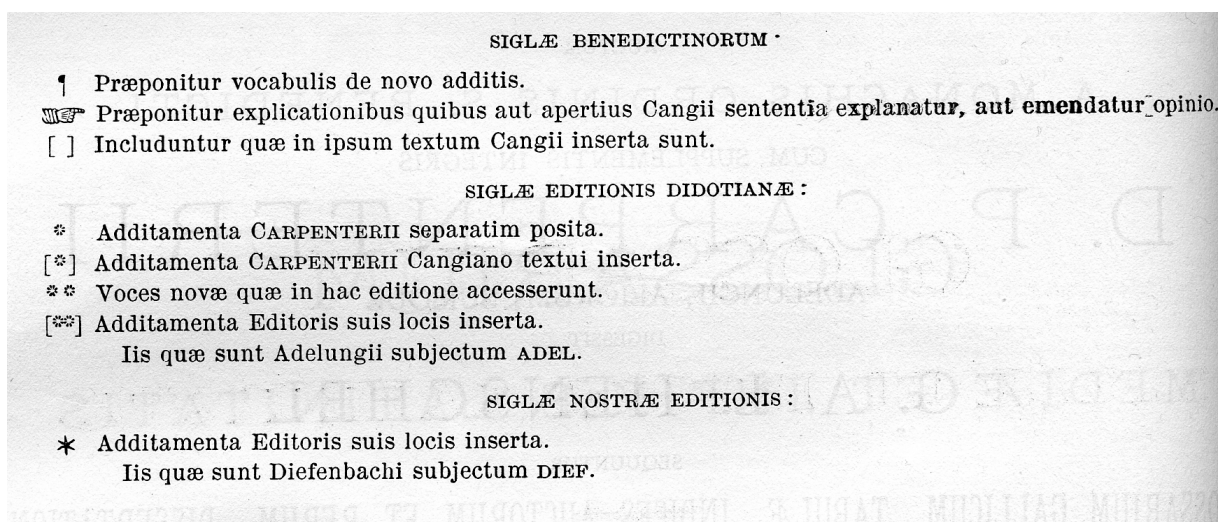
## Caractères

Les caractères sont à encoder selon le taux garanti par le prestataire, en Unicode-UTF8, sans entités. Le balisage ne retient pas les sauts de ligne de l'édition imprimée, il n'y a donc pas de césure des mots. La pose des balises ne doit pas affecter l'espacement du texte original, par exemple autour d'une virgule ou d'un point.

Certaines puces demandent une interprétation univoque dans un caractère Unicode, compatible avec les polices courantes.

	¶ (u00B6)		# (u0023)		* (u002A)		□ (u00A4)
---	-----------	---	-----------	---	-----------	---	-----------

Ces signes sont expliqués dans une légende en début de chaque tome (*siglæ*).



## Doutes

La majorité des erreurs risquent de concerner des mots ou des citations en grec ancien accentué. Le prestataire est invité à faire de son mieux, pourvu que les doutes soient balisés avec `<unclear>`.

Si la chaîne de traitement permet des réglages, nous préférons plus de doutes déclarés que des erreurs silencieuses.

## Capitales grasses

Les capitales grasses sont généralement des vedettes, indiquant le lemme traité par un article, qu'elles servent à délimiter. Cependant, il existe des variations possibles qui forcent à un traitement postérieur pour assurer la nomenclature. La forme doit être conservée à l'identique (en capitales), avec accentuation et ligatures si rencontrées (ex: Æ, Œ, Ç). Elle est balisée au moyen de l'élément `<hi rend="bold">`. Tout article doit contenir au moins une vedette (2 peut être possible) ; toute exception doit être signalée.

```
<entry>
  <dictScrap>* <hi
rend="bold">PRECATORIUS</hi>. Vide
[...] </dictScrap>
</entry>
```

A terme, la vedette doit identifier un article de manière unique, en traitant notamment le cas des homographes. Des homographes sont des mots ayant la même orthographe, dont le sens est distingué par des articles différents. Ils sont signalés par un numéro avant le lemme en majuscules grasses. On peut le comprendre dans l'image présentée ici : « Vide *Procuratio* 2. » renvoie à l'article « ¶ 2. **PROCURATIO** ». Cette remarque est destinée à faire comprendre le type de traitement dont le balisage typographique fera l'objet.

```
<entry>
  <dictScrap>¶ 4. <hi rend="bold">procurare</hi>, Providere...</dictScrap>
</entry>
<entry>
  <dictScrap>¶ 5. <hi rend="bold">procurare</hi>, Rei...</dictScrap>
</entry>
```

Tolérance : ~0%, tests de présence obligatoire pour un article, majuscules latines, absence de ponctuation (évaluation XPath).

\* **PRECATORIUS**. Vide supra in *Precatoria* 2.

¶ 4. **PROCURARE**, Providere, suppeditare, Gall. *Procurer*. Tabular. Major. Monast. : *Religiosum satis opus... ecclesiarum ædificia construere, ... instituere ministros, victualiaque iisdem sufficientia Procurare*.

¶ 5. **PROCURARE**, Rei alicujus auctorem esse, quo sensu Galli *Procurer* dicunt. Litteræ Innocentii III. PP. ann. 1213. inter Instr. tom. 6. Gall. Christ. novæ edit. col. 57 : *Licet eidem (Arnaldo Archiep. Narbonensi) hominium feceris et fidelitatis præstiteris juramentum, nihilominus tamen, prout tibi placuit, Procurasti ut demolirentur muri et turres civitatis ipsius, etc. Vide Procuratio 2.*

## Articles

Les pages à baliser sont considérées comme une suite d'articles, rangés par ordre alphabétique des vedettes. Un article commence avec la ligne contenant une vedette et s'arrête avant la vedette suivante. Il contient un à plusieurs paragraphes. L'élément à utiliser est <entry>.

Tolérance : ~0%, vérifiable par validation DTD.

## Vers

Dans un article, on peut repérer des citations en vers, décrochées et en plus petits caractères (environ ~5% des articles). Il peut être utile de les baliser avant les paragraphes, car elles sont logiquement contenues dans un paragraphe. Ces vers partagent la fonction de citation définie plus bas (avec référence bibliographique adjacente). Chaque vers est inclus dans un élément <l>.

```
<dictScrap>[...] Le Roman de <hi rend="italic">Robert le Diable</hi> MS. :
<quote><l>L'Empereur est enmy la sale</l><l>U il ne Trepe, ne ne
bale.</l></quote> Vide <ref>Tripare</ref>.
```

Certains vers en distique sont présentés en alternance avec un alinéa. Cette information est importante à conserver, elle sera indiquée par 2 espaces insécables pour le vers décroché.

```
<quote>
  <l>Hancque sibi recipit, Patriotas convocat, ac de</l>
  <l> Urhevetanorum posse redemit eos.</l>
</quote>
```

Tolérance : ~0%, vérifiable par validation DTD.

## Paragraphes

Le texte du dictionnaire se divise en paragraphes avec alinéa. Certains paragraphes commencent avec une puce. Un paragraphe est toujours dans un article. Un paragraphe peut être coupé par un saut de page, ou par un saut de colonne, mais il se poursuit jusqu'à l'alinéa suivant. Il peut aussi être interrompu par une citation en vers, qui est considérée comme

**PIZZA**, Placenta, ex Italico *Pinza*. Charta ann. 1195. apud Ughellum tom. 7. pag. 1321: *Duas spallas porcorum, et sex Pizzas: et in Carnelevamine, unam gallinam et tres Pizzas,... unam gallinam, unum pullum, et 6. Pizzas, etc.*

Vitiis serm. pag. 628.

† **TREPARE**, verbum Ital. Jocari, saltare, tripudiare, Gall. *Danser*. Charta Tolosana ann. 1192: *Ipsium pratum et gravaria erant publica... causa spaciandi et Trepandi*. Le Roman de Robert le Diable MS. :

L'Emperere est enmy la sale  
U il ne Trepe, ne ne bale.

Vide *Tripare*.

\* Alias *Treper* et *Trepper*. Vitæ SS. MSS. ex Cod. 28. S. Vict. Paris. fol. 137.

**Thierricus Valliscolor** in Urbano IV. PP. :

Hancque sibi recipit, Patriotas convocat, ac de  
Urhevetanorum posse redemit eos.

**Occurrit præterea apud Silvestrum Gi-**

contenue dans le paragraphe. Tous les paragraphes doivent être balisés. L'élément à utiliser est `<dictScrap>`.

Tolérance : ~0%, vérifiable par validation DTD et évaluation XPath sur les puces.

### Paragraphes centrés

Joannes le Clerc 16. Novembris 1420.  
ad 6. Februarii 1424. qua die abdicavit.

Il existe quelques rares cas de paragraphes centrés qui demandent un attribut particulier pour retenir la disposition typographique

*Sub Carolo VII.*

Ludovicus Luxemburg. Episcop. Tervanensis litteris Henrici VI. Regis An-

@rend="center".

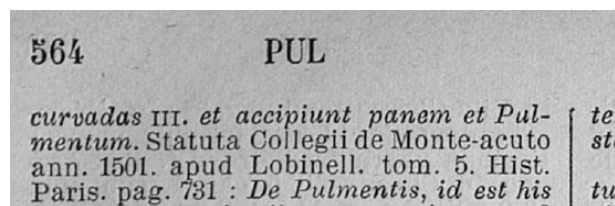
```
<dictScrap>Joannes le Clerc 16. Novembris 1420. ad 6. Februarii 1424. qua  
die abdicavit.</dictScrap>  
<dictScrap rend="center"><hi rend="italic">Sub Carolo VII.</hi></dictScrap>  
<dictScrap>Ludovicus Luxemburg. Episcop. Tervanensi litteris Henrici VI.  
Regis Angliæ...</dictScrap>
```

### Sauts de pages

Afin de conserver un lien avec les images des pages, le texte balisé indiquera les sauts de page au point où ils interviennent : entre deux articles, entre deux paragraphes, dans un paragraphe ; on évitera cependant qu'il coupe un mot, pour les rendus ultérieurs. L'élément à utiliser est l'élément vide `<pb/>`, dont l'attribut `@n` permet de conserver le numéro de page indiqué dans le titre courant de la page imprimée.

```
<dictScrap>...<quote>... <pb n="564"/>curvadas...
```

Tolérance : ~0%, série vérifiable par évaluation XPath.



### Sauts de colonnes

TRU	TRU	TRU	199
senses ann. 1351. e libro viridi fol. 53 : <i>Quod Ortolanus prædictus recipiat..... fimum totum stabulorum omnium..... et totam vindemiam Trullatam tempore vin- demiarum.</i> Litteræ ann. 1354. tom. 4. Ordinat. Reg. Franc. pag. 296. n. 15 : <i>Lidit habitant ne moorront . cuirront.</i>	riæ species, vel Jaculum. Stat. Mutin. ann. 1306. apud Murator. tom. 2. Antiq. Ital. med. ævi col. 484 : <i>Eligantur unus dominus et unus notarius pro qualibet porta, qui faciat parari trabuchos, sive manganos, balistas grossas, sagittamenta, Trulos, et alia necessaria.</i>	<i>Trumpam vocat.</i> [Vossius lib. 2. de Vi- tilis serm. cap. 18. vocem hanc deducit a veteri Germanico <i>Trompe</i> ; unde, in- quit, et Gallicum <i>Trompe</i> et Ital. <i>Tromba</i> , et Hispan. <i>Trompa</i> ; Belgæ usitatus <i>Trompet</i> : Angli et <i>Trump</i> et <i>Trompet</i> dicunt.]	

Les pages s'étalent sur trois colonnes. Un élément vide `<cb/>` est inséré à l'endroit où s'effectue le saut, selon les mêmes consignes que pour une page. Un saut de colonne est identifié par le numéro de page et une lettre d'ordre (a, b, c) pour les premières, deuxièmes et

troisièmes colonnes. Le commencement d'une page peut être doublement marqué, d'un saut de page et d'un saut de colonne.

```
<cb n="199a"/>, <cb n="199b"/>, <cb n="199c"/>
```

Tolérance : ~0%, série vérifiable par évaluation XPath.

## Chiffres romains

Les chiffres romains sont considérés comme en petites capitales. Afin de les distinguer des formes dans cette typographie, ils seront distingués au moyen de l'élément `<num>`. Les chiffres apparaissent généralement "pointés", sauf lorsqu'ils sont suivis d'une autre ponctuation (deux points, virgule, voire point), ce point doit être inclus dans l'élément.

```
<dictScrap>¶ <hi rend="smallcaps">Processus</hi>, Judicium, sententia.
Litteræ Edwardi <num>III.</num> Reg. in Chron. Angl. <hi
rend="italic">Otterbourne</hi> pag. 122 : <quote>Quorum <note>(Parium
Franciæ)</note> Processum factum in nostrum præjudicium, ipso jure nostra
fecit invalidum minor ætas</quote></dictScrap>
```

Tests sur les petites capitales contenant des chiffres romains.

## Petites capitales

Certains articles longs (~ 6%) comportent des paragraphes introduits par une forme en petites capitales. Ces formes sont à baliser dans un élément `<hi rend="smallcaps">`,

sans ponctuation, les grandes capitales sont inscrites en majuscules, les petites sont en minuscules. Elles peuvent avoir des statuts lexicographiques très différents. Dans l'exemple ci contre, CANATUM complète la forme **PRÆDIUM** pour constituer une vedette. PRÆDIA FORMENSIA est une sous-vedette destinée à une exploitation de type base de données.

```
<entry>
  <dictScrap>¶ <hi rend="bold">PRÆDIUM</hi> <hi
rend="smallcaps">Canatum</hi>, ...</dictScrap>
  <dictScrap>¶ <hi rend="smallcaps">Prædia Formensia</hi>, vel...</dictScrap>
</entry>
```

Tolérance : ~0%, ponctuation et segmentation vérifiable par évaluation XPath.

## Crochets

Les crochets sont employés dans l'édition papier pour encadrer des ajouts éditoriaux au sein d'un paragraphe (voir plus haut l'image *siglæ*). On utilisera donc la balise `<add>`. Ce signe est

assez fiable, la plupart des crochets ouverts sont fermés. En cas d'absence de crochet fermant, il suffit de poursuivre jusqu'à la fin du paragraphe. Un crochet fermant sans crochet ouvrant dans le paragraphe sera laissé sans balisage (3 repérés sur 240 pages). On propose de baliser les crochets avant les citations en italique, car ils assurent des repères stricts dans un paragraphe. Il ne doit pas y avoir de chevauchements.

```
<entry>
  <dictScrap>¶ <hi
rend="bold">PLACIDUM</hi>, Lis
intentata, controversia, litigatio.
Acta S. Lidani tom. 1. Julii pag.
346 : <quote>Devoratus extitit a lupis, de quo maximum Placidum
habuerunt.</quote><add>[* Vide in <ref>Placitum</ref>.]</add></dictScrap>
</entry>
```

**¶ PLACIDUM**, Lis intentata, controversia, litigatio. Acta S. Lidani tom. 1. Julii pag. 346 : *Devoratus extitit a lupis, de quo maximum Placidum habuerunt.* [\* Vide in *Placitum*.]

Tolérance : ~0%, tests sur crochets non balisés.

## Renvois

Environ 15% des segments en italique sont des renvois, c'est-à-dire un mot qui fait appel à un autre article dans le dictionnaire. Cette

information est une valeur ajoutée importante au texte. La vérification du lien n'est pas requise (elle sera effectuée par des traitements postérieurs sur toute la livraison). Les renvois reconnus sont balisés avec l'élément `<ref>`. Ce sont des formes en italique précédées de formules parfois normalisées, notamment : (Vide infra in|Vide infra|infra in|ut infra|infra|Vide supra in|Vide supra|supra in|ut supra|supra|Vide mox in|Vide mox|Vide in|Vide) {italique}. Il restera bien sûr d'autres formes en italique pouvant être des renvois ; leur reconnaissance est laissée à l'appréciation du prestataire. L'essentiel est ici de ne pas les confondre avec une citation. Dans l'exemple ci-dessus, « Vide supra » introduit un renvoi à « *Politia 2.* », la conjonction « et » introduit un autre renvoi à « *Pollere* », la citation s'arrête à « *loci, etc.* ». On notera que typographiquement, les chiffres ne sont jamais en italique, bien qu'ils puissent être logiquement rattachés à un tel segment. Les numéros d'homographes doivent être contenus dans le lien balisé, avec le point qui le suit généralement.

```
...dicti loci, etc.</quote> Vide supra <ref>Politia 2.</ref> et
<ref>Pollere</ref>.</dictScrap>
```

Tolérance : ~0%, non pas sur tous les renvois, mais sur ceux précédés par la liste contractuelle ci-dessus ; seront aussi vérifiés les citations et l'italique qui contiendrait ces motifs. Vérifiable par évaluation XPath.

**\* POLLITIA**, Regimen, administratio. Lit. ann. 1410. tom. 9. Ordinatio reg. Franc. pag. 559. art. 19 : *Ordinatio quod cura et sollicitudo, ... quæ pertinet... consulis de Lautrico circa regimen victualium et aliarum rerum, et circa alias Politias et regimina dicti loci, etc.* Vide supra *Politia 2.* et *Pollere*.



## Parenthèses

Les parenthèses encadrant des lettres droites au milieu d'un passage en italique sont des notes de l'auteur dans une citation. Elles sont balisées avec l'élément `<note>`. Les repérer avant peut assurer de segmenter correctement les passages cités. Il existe quelques cas de parenthèses en italique.

\* 4. **PROCURATIO**, *Officium procuratoris. Stat. synod. Guill. Duprat episc. Claromont. ann. 1537 : Aliquod seculare officium non exercent (sacerdotes) videlicet receptam laicorum, officium bajulatus, judicaturæ aut Procurationis in foro seculari. Vide Procurare 3.*

Dans l'exemple ci contre, la citation englobe la glose « (sacerdotes) », et s'arrête après le point, avant « Vide ».

```
<quote>Aliquod seculare officium non exercent <note>(sacerdotes)</note>
videlicet receptam laicorum, officium bajulatus, judicaturæ aut
Procurationis in foro seculari.</quote> Vide <ref
target="procurare3">Procurare 3.</ref>
```

Tolérance : ~0%, test sur la présence de parenthèses non balisées, adjacentes ou à l'intérieur, d'une citation ou d'un segment en italique (évaluation XPath).

## Citations

Les citations paraissent en italique, avec une référence bibliographique adjacente, (généralement introduite par deux points ":"). Une citation est encadrée par un élément `<quote>`, elle inclut la ponctuation terminale. Un point final ou d'interrogation fait partie de la citation.

\* **PUNEIRA**, *Quantum pugno continetur, Gall. Poignée. Charta Bern. vicecomit. Carcass. ann. 1082. ex Bibl. S. Germ. Prat. : Dono vobis in ipsam meam leudam de Limos tres Puneiras de sal, in unaquaque die mercatile. Vide Puniata.*

...Caracass. ann. 1082. ex Bibl. S. Germ. Prat. : `<quote>Dono Vobis in ipsam de sal, in unaquaque die mercatile.</quote>` Vide `<ref>Puniata</ref>`.

Tolérance : ~0% sur les segments italiques précédés de ':' (évaluation XPath) ; tout effort pour faire mieux sera apprécié.

## Italique

Pour les mots en italique qui ne sont pas des citations ou des renvois. La typographie sera conservée avec l'élément `<hi rend="italic">`, en évitant d'inclure les ponctuations finales (contrairement à une citation), afin de faciliter des récupérations ultérieures. Ainsi, l'exemple montre un mot français « Gall. [français] *Poussin*. » qui pourra être caractérisé plus tard.

\* **PUGINUS**, Pullus. Gall. *Poussin*. Codex censualis Castellionis Dombensis ann. 1463 : *Debet unum Puginum et quartam partem alterius Pugini. Vide Pulcinus.*

```
<entry><dictScrap>... Gall. <hi rend="italic">Poussin</hi>. Codex...
```

Tests sur les segments italiques longs, pour vérifier les citations manquées.

## Autres variations typographiques

On peut observer des lettres en exposant, généralement dans des abréviations. Elles sont balisées conformément à la TEI avec `<hi`

`rend="superscript">`. L'élément `<hi>` peut être utilisé pour baliser des observables typographiques qui n'auraient pas été identifiés plus haut, en choisissant une valeur normalisée pour l'attribut `@rend`.

```
<dictScrap>* 2. <hi rend="bold">POTURA</hi>. Charta Joan. comit. Cabillon. ann. 1227. in Chartul. Campan. ex Cam. Comput. Paris. fol. 211. v<hi rend="superscript">o</hi>. col.
```

Des tests seront effectués pour mesurer les variations typographiques qui auraient mérité une modélisation, ou qui auraient pu trouver un balisage déjà défini plus haut.

## Grec ancien

Dans le texte on peut trouver des segments en alphabet autre que romain, généralement en grec ancien accentué (~8% des articles), et quelques très rares segments en hébreu.

La langue de ces segments doit pouvoir être identifiée par un attribut `@xml:lang`, porté par une balise déjà décrite (ex : `<quote>`) ou à défaut par un élément `<foreign>`.

```
<dictScrap><hi rend="bold">PLASMARE</hi>, Formare, ex Gr. <foreign xml:lang="grc">Πλασμαζειν</foreign>. Glossæ antiquæ MSS. : <quote>Finxit...
```

Tolérance : ~0%, vérifiable par expressions régulières

## Guillemets

L'éditeur Favre (voir l'histoire éditoriale au titre des puces) a introduit d'autres conventions dans ses ajouts (2% des articles). Les guillemets français, ouvrants et fermants, doivent être conservés, et encadrés d'un élément `<quote>`.

```
<quote>« Et hoc est quod dicit primus sapiens qui locutus est in libro <hi rend="italic">Preallegato</hi> de alphilaha. »</quote>
```

Tolérance : ~0%, vérifiable par expressions régulières.

\* 2. **POTURA**. Charta Joan. comit. Cabillon. ann. 1227. in Chartul. Campan. ex Cam. Comput. Paris. fol. 211. v<sup>o</sup>. col.

\* **PLASMARE**, Formare, ex Gr. Πλασμαζειν. Glossæ antiquæ MSS. : *Finxit, Plasmavit, composuit, formavit*. [Pactum matrimonii inter Herbertum et Gode-

\* **PREALLEGATUS**, [Déjà cité : « Et hoc est quod dicit primus sapiens qui locutus est in libro *Preallegato* de alphilaha. » (B. N. Ms. Lat. 10272. p. 9.)]

## Figures

Pour quelques dizaines d'articles sur environ 80 000, le balisage ne conservera pas toute l'information typographique. Un élément `<figure>` servira à signaler qu'il est nécessaire de recourir à l'image pour ne pas perdre une partie de l'information. L'usage de cet élément est évident pour les illustrations graphiques

inclues dans un paragraphe (très rare), il est aussi nécessaire pour signaler des surfaces où le texte est récupérable, mais pas sa disposition. L'encodage tentera de conserver une version linéaire du contenu qui soit lisible, le texte sera encadré pour signaler qu'il constitue une sorte de figure.

```
<!-- tome 1 p. 382 -->
<dictScrap><quote>Argent Fin est à 12. den. ob. d'Argent
le Roy.</hi>
<figure>Gros Tournois de S. Loys et les autres<lb/>
Argent signié en vesselles et mailles Tierces vieilles<lb/>
} sont à 12. den. d'Argent le Roy.<lb/>
Estatlins d'Angleterre et et d'Escoce<lb/>
Gambirosins de Philippe<lb/>
Chevaliers de Guillaume<lb/>
Passavants neufs<lb/>
Fors de Nevers de Philippe<lb/>
} sont à 11. den. ob. Argent le Roy.</figure></quote></dictScrap>
```

*Argent Fin est à 12. den. ob. d'Argent le Roy.*  
*Gros Tournois de S. Loys et les autres*  
*Argent signié en vesselles et mailles Tierces vieilles*  
*Estatlins d'Angleterre et et d'Escoce*  
*Gambirosins de Philippe*  
*Chevaliers de Guillaume*  
*Passavants neufs*  
*Fors de Nevers de Philippe.*

*sont à 12. den. d'Argent le Roy.*  
*sont à 11. den. ob. Argent le Roy.*

## Difficultés d'interprétation

Quelle que soit la précision de ces consignes, elles donnent un guide mais ne peuvent pas prévoir tous les cas que le balisage complet rencontrera. S'il n'est pas possible de prendre contact avec l'École nationale des chartes pour ajuster de nouvelles règles, toute initiative conforme à la TEI et avançant vers l'objectif évoqué ci-dessus sera appréciée. Pour des exceptions rares, un commentaire XML est attendu dans le fichier encodé afin de faciliter le repérage et les reprises manuelles par des spécialistes du texte concerné.

```
<!-- En cas d'exception en cours de balisage,
laisser un commentaire XML -->
```

## Balises

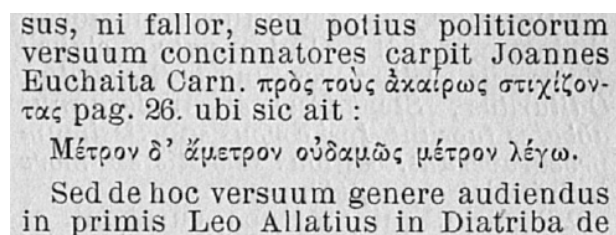
### Unicode

Les fichiers doivent être du XML valide, strictement en unicode/UTF-8, sans entités.

```
<?xml version="1.0" encoding="UTF-8"?>
```

### @xml:lang<sup>3</sup>

La langue par défaut de ce dictionnaire est le latin. Certains segments, ou des citations peuvent apparaître dans un alphabet autre que latin. Ces caractères doivent être balisés, l'élément portera un attribut @xml:lang, avec un code de langue ISO 639, conformément à la norme XML. La portée concerne tous les noeuds inclus. La reconnaissance des mots de langues européennes (ancien français, italien médiéval, etc.) n'est pas requise. L'élément <foreign> est utilisé pour porter l'attribut qui n'aurait pas été balisé par ailleurs.



```
<entry>...
```

```
<dictScrap>... Euchaïta Carn. <foreign xml:lang="grc">πρὸς τοὺς ἀκαίρως  
στιχίζοντα</foreign> pag. 26. ubi sic ait : <quote xml:lang="grc"><l>Μέτρον δ'  
ἄμετρον οὐδαμῶς μέτρον λέγω.</l></quote></dictScrap>
```

```
<dictScrap>Sed de hoc...</dictScrap>
```

```
</entry>
```

### <add><sup>4</sup>

<add> « addition » contient un segment de texte inséré dans un paragraphe, par un auteur identifié. Selon la légende en début de chaque tome (*siglæ*), ces portions sont délimitées par des crochets droits. Quand un crochet ouvrant n'est pas fermé, il se poursuit jusqu'à la fin du paragraphe. Quand un crochet fermant est trouvé sans commencement, il est laissé sans balise.

3 <<http://www.w3.org/TR/REC-xml/#sec-lang-tag>>

4 <<http://www.tei-c.org/release/doc/tei-p5-doc/html/ref-add.html>>

¶ PTOCHOTROPHIUM, Eadem notione, in lib. 2. Capit. cap. 29. [\*\* Julian. Nov. const. 7. tit. 1. (§ 32.)]: *Nulla sub Romana ditone constituta Ecclesia vel xenodochium, vel Ptochotrophium, etc.*

```
<dictScrap><hi rend="smallcaps">Ptochotrophium</hi>, Eadem notione, in lib.
2. Capit. cap. 29. <add>[** Julian. Nov. const. 7. tit. 1. (§ 32.)]</add> :
<quote>Nulla sub Romana ditone constituta Ecclesia vel xenodochium, vel
Ptochotrophium, etc.</quote></dictScrap>
```

## <cb><sup>5</sup>

<cb/> *column break* « saut de colonne » indique le début d'une colonne dans le texte balisé. L'attribut @n porte le numéro de page et une lettre minuscule pour chaque colonne, ce qui donne des valeurs comme 45a, 45b, 45c. Les consignes d'insertion dans le texte sont les mêmes que pour un saut de page <pb/>.

```
<cb n="{numéro de page}{a|b|c}" />
```

## <dictScrap><sup>6</sup>

<dictScrap> contient les paragraphes délimités dans la version imprimée par des alinéas. Les paragraphes sont toujours contenus dans un article, <entry> ; un article a toujours au moins un paragraphe.

¶ PROSATRIX, *Genitrix*. Gloss. Isidor.

```
<entry>
  <dictScrap>¶ <hi rend="bold">PROSATRIX</hi>, <hi
rend="italic">Genitrix</hi>. Gloss. Isidor.</dictScrap>
</entry>
```

## <entry><sup>7</sup>

<entry> balise un article, c'est-à-dire un ensemble de paragraphes (<dictScrap>), depuis un alinéa précédent une vedette jusqu'au suivant.

478

AUG

alisque expansis, cum hisce characteribus, FRIDERICUS.

## <figure><sup>8</sup>

<figure>, signale une surface imprimée où le balisage du texte perdra une information



[\*\* Exhibemus hunc nummum ex Gazophylacio Regio, summa cura delineat.

5 <<http://www.tei-c.org/release/doc/tei-p5-doc/html/ref>

6 <<http://www.tei-c.org/release/doc/tei-p5-doc/html/ref>

7 <<http://www.tei-c.org/release/doc/tei-p5-doc/html/ref>

visuelle, notamment des images, mais aussi des arrangements typographiques particuliers. Une interface de consultation saura ainsi inviter l'utilisateur à consulter l'image. Il s'agit d'un cas rare, rien n'a été repéré sur un dépouillement de 240 pages, une quinzaine d'articles ont pu être identifiés sur le reste de l'ouvrage.

```
<dictScrap>[...] alisque expansis, cul hisce characteribus,
FREDERICUS.<figure/></dictScrap>
<dictScrap><add>[*] Exhibemus hunc
nummum ex Gazophylacio Regio, summa
cura [...]</add></dictScrap>
```

```
<entry>
  <dictScrap><hi
rend="bold">ABRACADABRA</hi>,
Inscriptio amuleti apud Serenum
Sammonicum pro hemitritæo, <add>[seu
semi-tertianu febrî, Gallice <hi
rend="italic">Fièvre double-
tierce</hi>,</add> cap. 52.
Editionis Parisiensis apud Colinæum
1538 : <hi rend="italic">Remede
superstitieux</hi>.
```

```
  <quote>
    <l>Inscribes chartæ quod
dicitur Abracadabra</l>
    <l>Sæpius et subter repetis,
sed detrahe suma</l>
    <l>Et magis atque magis desint
elementa figuris</l>
    <l>Singula, quæ semper rapies,
et cætera figes,</l>
    <l>Donec in angustum redigatiir
lilera conum.</l>
    <l>His lino nexis collum redimire memento.</l>
  <figure>
    <l>A B R A C A D A B R A</l>
    <l> A B R A C A D A B R</l>
    <l>  A B R A C A D A B</l>
    <l>   A B R A C A D A</l>
    <l>    A B R A C A D</l>
    <l>     A B R A C A</l>
    <l>      A B R A C</l>
    <l>       A B R A</l>
    <l>        A B R</l>
    <l>         A B</l>
    <l>          A</l>
  </figure>
</quote>
Et addit:
```

**ABRACADABRA**, Inscriptio amuleti  
apud Serenum Sammonicum pro hemi-  
tritæo, [seu semi-tertiana febrî, Gallice  
*Fièvre double-tierce*,] cap. 52. Editionis  
Parisiensis apud Colinæum 1533: *Remede  
superstitieux*.

Inscribes chartæ quod dicitur Abracadabra  
Sæpius et subter repetis, sed detrahe suma  
Et magis atque magis desint elementa figuris  
Singula, quæ semper rapies, et cætera figes,  
Donec in angustum redigatur litera conum.  
His lino nexis collum redimire memento.

ABRACADABRA  
ABRACADABR  
ABRACADAB  
ABRACADA  
ABRACAD  
ABRACA  
ABRAC  
ABRA  
ABR  
AB  
A

Et addit :

Talia languentis conducunt vincula collo ;  
Lethales abigent, miranda potentia ! morbos.

Quæ quidem inscriptio Sammonicum  
Basilidis hæretici, qui sub Hadriano  
vixit sectatorem fuisse arguit, cui pro  
summo deo fuit nescio quis ABRASAX.

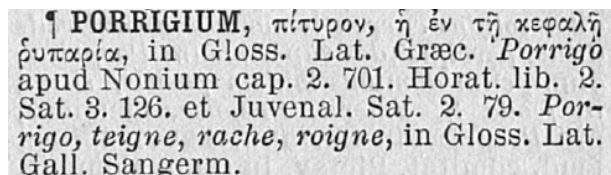
```

<quote>
  <l>Talia languentis conducunt vincula collo ;</l>
  <l>Lethales abigent, miranda potentia ! Morbos.</l>
</quote>
Quæ quidem inscriptio Sammonicum Basilidis hæretici, qui sub Hadriano
[...]</dictScrap>
</entry>

```

## <foreign><sup>9</sup>

<foreign> « étranger » encadre les segments de texte dans un alphabet autre que latin qui n'ont pas encore été balisés afin de porter un attribut @xml:lang. Le balisage évitera d'encadrer la ponctuation finale. Cet élément ne doit être employé que si aucun autre n'a pu convenir (<quote>, <ref>, <hi rend="italic">).



```

<entry><dictScrap>¶ <hi rend="bold">PORRIGIUM</hi>, <foreign
xml:lang="grc">πίτυρον, ἡ ἐν τῇ κεφαλῇ ῥυπαρία</foreign>, in Gloss. Lat. Græc.
<hi rend="italic">Porrigo</hi> apud Nonium cap. 2. 701. Horat. lib. 2. Sat.
3. 126. et Juvenal. Sat. 2. 79. <hi rend="italic">Porrigo</hi>, <hi
rend="italic">teigne</hi>, <hi rend="italic">rache</hi>, <hi
rend="italic">roigne</hi>, in Gloss. Lat. Gall.
Sangerm.</dictScrap></entry>

```

## <hi><sup>10</sup>

<hi>, *highlighted* « mis en valeur » : *marks a word or phrase as graphically distinct from the surrounding text, for reasons concerning which no claim is made*. Cette balise permet de conserver la mise en forme typographique observable de segments de texte, dans l'attente de pouvoir définir leur fonction lexicographique. Le type de rendu est indiqué par l'attribut @rend. L'échantillon observé permet de distinguer par ordre d'importance (au moins fonctionnelle, sinon en nombre) du gras (*bold*), des petites capitales (*smallcaps*), de l'italique (*italic*), des lettres en exposant (*superscript*). Il est autorisé d'ajouter une autre valeur normalisée si d'autres mises en formes apparaissaient en cours de balisage.

9 <<http://www.tei-c.org/release/doc/tei-p5-doc/html/ref-foreign.html>>

10 <<http://www.tei-c.org/release/doc/tei-p5-doc/html/ref-hi.html>>

**CAPITALES GRASSES** @rend="bold" — Le balisage contenu de cet élément doit permettre d'établir la nomenclature. Aussi ne doit-on y trouver qu'un mot ou une expression, à l'exclusion des puces, des ponctuations de séparation, et des numéros pointés d'homographes.

\* **PULVERISARE, PŪLVERIZARE**,  
[Gall. nettoyer, enlever la poussière.  
«..... Habui ad volvendum et *Pulverisandum* omnia blada graneriorum. »  
)Art. Histor. de la Gironde t. 12. p. 190.)]

```
<entry>
  <dictScrap>⌘ <hi rend="bold">PULVERISARE</hi>, <hi
rend="bold">PULVERIZARE</hi>, <add>[Gall. nettoyer, enlever la poussière.
<quote>«..... Habui ad volvendum et <hi rend="italic">Pulverisandum</hi>
omnia blada graneriorum. »</quote> (Art. Histor. de la Gironde t. 12. p.
190.)]</add></dictScrap>
</entry>
```

**PETITES CAPITALES** @rend="smallcaps" — Les petites capitales repèrent des formes importantes dans la structure de l'article, certaines pouvant devenir des sous-vedettes. Selon les tests effectués, il n'y a pas de raison que cette balise contienne de la ponctuation.

† **TRICAPITINUS**, Triceps, tria habens  
capita. Gloss. Lat. Græc. *Tricapitinus*,  
Τριφνής, τρικέφαλος, ibidem.  
† **TRICAPITUS**, Idem, tom. 1. SS. Julii  
pag. 439.

```
<entry>
  <dictScrap>† <hi rend="bold">TRICAPITINUS</hi>, Triceps, tria habens
capita. Gloss. Lat. Græc. <hi rend="italic">Tricapitinus</hi>, <foreign
xml:lang="grc">Τριφνής</foreign>, <foreign
xml:lang="grc">τρικέφαλος</foreign>, ibidem.</dictScrap>
  <dictScrap>† <hi rend="smallcaps">Tricapitus</hi>, Idem, tom. 1 SS. Julii
pag. 439.</dictScrap>
</entry>
```

† **POLLINARE CRIBRUM**, γυριστήριον  
κόσκινον, in Gloss. Lat. Gr. Cod. Reg.  
*Pollinarium*. Vide *Polentrudium*.

```
<entry>
  <dictScrap>† <hi rend="bold">POLLINARE</hi> <hi
rend="smallcaps">Cribrum</hi>, <foreign xml:lang="grc">γυριστήριον
κόσκινον</foreign>, in Gloss. Lat. Gr. Cod. Reg. <hi
rend="italic">Pollinarium</hi>. Vide <ref>Polentrudium</ref>.</dictScrap>
</entry>
```

*Italique* @rend="italic" — Pour les segments italiques qui ne sont pas des renvois ou des citations, le balisage typographique en conservera l'apparence. Cependant, de nombreuses formes ont une fonction lexicographique qui pourra être définie plus tard, aussi, l'élément ne doit pas contenir de ponctuation finale (point ou virgule).



¶ **FUX.** pro *Faux*, Ostium fluminis,  
Gall. *Embouchure*, in Annal. Genuens.  
ann. 1204. 1214. 1242.

<entry>

<dictScrap>¶ <hi rend="bold">FUX</hi>. pro <hi rend="italic">Faux</hi>,  
Ostium fluminis, Gall. <hi rend="italic">Embouchure</hi>, in Annal.  
Genuens. ann. 1204. 1314. 1242.</dictScrap>

</entry>

¶ **PONTONAGIUM**, PONTONARIA, PON-  
TONATICUM, etc. Vide *Pontaticum* et *Pon-*  
*tones*.

<entry>

<dictScrap>¶ <hi rend="bold">PONTONAGIUM</hi>, <hi  
rend="smallcaps">Pontonaria</hi>, <hi rend="smallcaps">Pontonaticum</hi>,  
etc. Vide <ref>Pontaticum</ref> et <ref>Pontones</ref>.</dictScrap>

</entry>

<l><sup>11</sup>

<l>, *line* « ligne » contient un vers dans une citation (<quote>).

tionem recesserunt, uti observat Terentianus Maurus :

Sed qui pedestres fabulas socco premunt  
Ut quæ loquuntur sumpta de vita putes,  
Vitiant iam bon tractibus spondaicis,  
Et in secundo, et cæteris æque locis :  
Fidemque fictis dum procurant fabulis,  
In metro peccant, arte, non inscitia,  
Ne sint sonora verba consuetudinis,  
Paulumque rursus a solutis differant.

His consentanea scribit Hephestion  
pag. 17. comicos nempe ita versus suos

<dictScrap>...observationem recesserunt, uti observat Terentianus Maurus :  
<quote>

<l>Sed qui pedestres fabulas socco premunt</l>

<l>Ut quæ loquuntur sumpta de vita putes,</l>

<l>Vitiant iam bon tractibus spondaicis,</l>

<l>Et in secundo, et cæteris æque locis :</l>

<l>Fidemque fictis dum procurant fabulis,</l>

<l>In metro peccant, arte, non inscitia,</l>

<l>Ne sint sonora verba consuetudinis,</l>

<l>Paulumque rursus a solutis differant.</l>

</quote> His consentanea scribit Hephestion pag. 17. comicos nempe ita  
versus suos exarasse, ...</dictScrap>

<sup>11</sup> <<http://www.tei-c.org/release/doc/tei-p5-doc/html/ref-l.html>>

<lb><sup>12</sup>

<lb> *line break* « saut de ligne » est proposé pour de très rares cas où un saut de ligne ne signale pas un changement de paragraphe, le commencement d'une citation, ou un vers dans une citation. Il est rappelé que les sauts de lignes et césures de l'édition originale ne sont pas conservés.

<note><sup>13</sup>

<note> est ici employé pour l'usage très restreint d'isoler les segments de texte droit dans une citation en italique, apparaissant généralement entre parenthèses.

\* **PROVISIA** et PROVISIO, Concessio in compensationem rei controversæ, qua ne quis patiaturs damnum, provide-tur. Charta ann. 1427. inter Instr. tom. 11. Gall. Christ. col. 53 : *Laborabit* (archiepiscopus) *obtinere Provisiam per medium translationis vel permutationis, de jure quod prætendit... Et si, pendente dicta continuatione processus, Provisionem seu recompensationem non obtineret, etc.*

<entry>

<dictScrap>\* <hi rend="bold">PROVISIA</hi> et <hi rend="smallcaps">Provisio</hi>, Concessio in compensationem rei controversæ, qua ne quis patiaturs damnum, providetur. Charta ann. 1427. inter Instr. tom. 11. Gall. Christ. col. 53 : <quote>Laborabit <note>(archiepiscopus)</note> obtinere Provisiam per medium translationis vel permutationis, de jure quod prætendit... Et si, pendente dicta continuatione processus, Provisionem seu recompensationem non obtineret, etc.</quote></dictScrap>

</entry>

<num><sup>14</sup>

<num> *number*, nombre, est requis au stade du balisage typographique pour les chiffres romains, afin d'éviter des confusions avec les formes en petites capitales, sera ensuite étendu aux chiffres arabes, pour faciliter la récupération des références bibliographiques.

\* **OSTADEGARI, OSTEDEGARI, Idem ac Ostagiare, Hostagiare**, liberare sub conditione obsides dandi. Stat. Bonon. ann. 1250-67. tom. III. pag. 316 : *Et quod dicti*

<entry>

12 <<http://www.tei-c.org/release/doc/tei-p5-doc/html/ref-lb.html>>

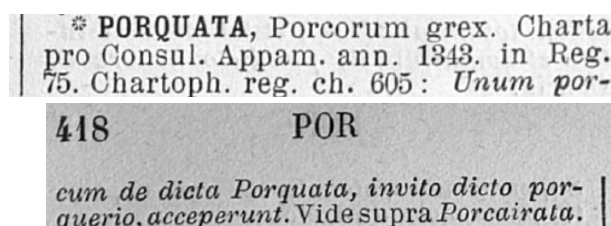
13 <<http://www.tei-c.org/release/doc/tei-p5-doc/html/ref-note.html>>

14 <<http://www.tei-c.org/release/doc/tei-p5-doc/html/ref-num.html>>

```
<dictScrap>* <hi rend="bold">OSTADEGARI</hi>, <hi
rend="smallcaps">Ostedegari</hi>, Idem ac <hi rend="italic">Ostajiare</hi>,
<hi rend="italic">Hostagiare</hi>, liberare sub conditione obsides dandi.
Stat. Bonon. ann. 1250-67. tom. <num>III.</num> pag. 316 : <quote>Et quod
dicti...</quote></dictScrap>
</entry>
```

## <pb/><sup>15</sup>

<pb/> *page break* « saut de page » marque un saut de page à l'endroit d'un changement de page. La précision n'est pas au caractère mais au mot près, afin d'éviter les césures. L'attribut @n porte le numéro de la page tel qu'indiqué dans le titre courant.



```
<entry>
<dictScrap>* <hi rend="bold">PORQUATA</hi>, Porcorum grex. Charta pro
Consul. Appam. ann. 1343. in Reg. 75. Chartoph. reg. ch. 605. : <quote>Unum
porcum<pb n="418"/> de dicta Porquata, invito dicto porquerio,
acceperunt.</quote> Vide supra <ref>Porcairaata</ref>.</dictScrap>
</entry>
```

## <quote><sup>16</sup>

<quote> contient une citation, repérable par l'italique et une référence bibliographique adjacente. Cette référence bibliographique est en style normal, souvent avec des abréviations comme *pag.* ou *vol.*, ainsi que des chiffres. Elle introduit la citation par ":" lorsqu'elle est avant.

\* **POSATORIUS FUSTIS**, Cui quis innititur. Form. Mss. ex Cod. reg. 7657. fol. 40. v° : *Dictam talem cum uno Posatorio fuste..... diversis ictibus percussit, etc.*

```
<entry>
<dictScrap>* <hi rend="bold">POSATORIUS</hi> <hi
rend="smallcaps">fustis</hi>, Cui quis innititur. Form. Mss. ex Cod. reg.
7657. fol. 40. v<hi rend="superscript">o</hi> : <quote>Dictam talem cum uno
Posatorio fuste..... diversis ictibus percussit, etc.</quote></dictScrap>
</entry>
```

<sup>15</sup> <<http://www.tei-c.org/release/doc/tei-p5-doc/html/ref-pb.html>>

<sup>16</sup> <<http://www.tei-c.org/release/doc/tei-p5-doc/html/ref-quote.html>>

PRIMICERIUS SÂCRI CUBICULI, in Notitia Imperii pag. 3 : *Primicerius totius officii. Primicerius scrinii Canonum, tabulariorum, aureæ massæ, auri ad responsum, vestiarii sacri, argenti, annularenensis, a pecuniis*, in eadem Notitia pag. 49. ubi consulendus Pancirolus.

```
<dictScrap><hi rend="smallcaps">Primicerius Sacri Cubiculi</hi>, in
Notitia Imperii pag. 3 : <quote>Primicerius totius officii.</quote>
<quote>Primicerius scrinii Canonum, tabulariorum, aureæ massæ, auri ad
responsum, vestiarii sacri, argenti, annularenensis, a pecuniis</quote>, in
eadem Notitia pag. 49. ubi consulendus Pancirolus.</dictScrap>
```

### <ref><sup>17</sup>

<ref>, « référence », indique un lien vers une vedette, dont la cible sera vérifiée ultérieurement. Il se repère par un préfixe de type « Vide... ». Le contenu doit permettre de retrouver la vedette pointée. Par défaut, il ne doit pas contenir de ponctuation de séparation (sauf chiffre pointé d'homographe).

\* **PROSEQUI** CAMPUM, Duellum peragere. Vide supra in *Campus* 3.

```
<entry>
  <dictScrap>* <hi rend="bold">PROSEQUI</hi> <hi
rend="smallcaps">Campum</hi>, Duellum peragere. Vide supra in <ref>Campus
3.</ref></dictScrap>
</entry>
```

### <unclear><sup>18</sup>

<unclear> « incertain », encadre des caractères dont la reconnaissance n'est pas sûre.

<sup>17</sup> <<http://www.tei-c.org/release/doc/tei-p5-doc/html/ref-ref.html>>

<sup>18</sup> <<http://www.tei-c.org/release/doc/tei-p5-doc/html/ref-unclear.html>>

## Schéma

---

Les éléments XML décrits se combinent selon un sous-ensemble de TEI, dont on trouve ci-dessous la grammaire dans le langage DTD. Le code en est abondamment commenté à destination d'outils de documentation automatique. Il est normatif des documents XML attendus de la prestation. Un fichier non conforme à ce modèle sera refusé. On remarquera que la DTD est rédigée dans un ordre adressé au lecteur humain, isolant 4 sections. Seule la première, « balisage typographique », concerne la prestation. La seconde section est laissée à titre informatif et pratique ; elle résultera d'une couche de tests et de post-traitements appliquée par l'Ecole des chartes. La troisième section recueille des idées de traitements possibles dont la possibilité ou l'utilité scientifique n'a pas été validée. La quatrième section est un en-tête TEI minimal pour que le document soit conforme à ce schéma ; elle sera vide et inscrite dans le modèle de document XML à livrer.

```
<!ENTITY % lang '(fr|it|oc|grc|en|es|de|nl|heb|xx)' >
<!ENTITY % breaks "pb | cb ">
<!ELEMENT entry (dictScrap | %breaks;)+>
<!ELEMENT dictScrap (#PCDATA | form | add | ref | quote | hi | foreign |
figure | unclear | num | abbr | xr | cit | lb | %breaks; )*>
<!ATTLIST dictScrap
    rend ( center | left | right ) #IMPLIED
>
<!ELEMENT pb EMPTY>
<!ATTLIST pb
    n CDATA #REQUIRED
>
<!ELEMENT cb EMPTY>
<!ATTLIST cb
    n CDATA #REQUIRED
>
<!ELEMENT add (#PCDATA | %breaks; | ref | quote | hi | foreign | figure |
unclear | num | cit )*>
<!ELEMENT ref (#PCDATA) >
<!ELEMENT note (#PCDATA| hi | unclear | %breaks; )*>
<!ELEMENT quote (#PCDATA | l | note | oVar | hi | add | figure | unclear |
lb | num | %breaks; )* >
<!ELEMENT l (#PCDATA| note | num | hi | unclear | %breaks; )*>
<!ELEMENT num (#PCDATA)>
<!ELEMENT hi (#PCDATA)>
<!ATTLIST hi
    rend ( bold | smallcaps | italic | superscript ) #REQUIRED
>
<!ELEMENT foreign (#PCDATA | unclear )*>
```

```

<!ATTLIST foreign
  xml:lang %lang; #REQUIRED
>
<!ELEMENT figure (#PCDATA | 1 | 1b )*>
<!ELEMENT unclear (#PCDATA)>

<!-- <h2 id="bloc-lexico">Enrichissement lexicographique.</h2>

Les informations balisées qui suivent ne sont pas requises à l'étape du
balisage typographique.
Elles sont fournies à titre informatif, définissant un programme
d'enrichissement lexicographique
de la source.
-->
<!-- L'établissement des vedettes est une opération indispensable à
l'exploitation informatique du
dictionnaire. En général, les formes en capitales grasses sont des
vedettes.
Il faut aussi tenir compte du numéro d'homographe, qui n'a pas d'apparence
typographique particulière,
bien qu'il fasse logiquement partie de la vedette.
De plus, il faut parfois leur ajouter
des formes en petites capitales lorsqu'elles ne sont pas séparées par une
virgule ou un point.
Les autres formes en petites capitales sont importantes, mais peuvent avoir
une signification très variable.
Il faut parfois leur ajouter des segments en italique qui ne sont pas
séparés d'une ponctuation.
Un post-traitement appliquera un élément <form> par dessus les
variations typographiques
repérérées avec l'élément <hi>, afin de conserver l'apparence de
l'original imprimé,
tout en fournissant le moyen informatique d'établir une nomenclature et des
identifiants avec les
numéros d'homographes.
Le niveau d'une vedette est distingué par l'attribut &#64;type. -->
<!ELEMENT form (#PCDATA|hi)*>
<!--
@attr type Le type "lemma" est requis pour la vedette principale.
      Les autres valeurs proviennent de la documentation
TEI et peuvent peut-être aider à spécifier
      les formes en petites capitales.
@attr n    L'attribut &#64;n conserve le numéro d'homographe. Il sera
établi automatiquement
      après balisage des vedettes, pour vérifier la
cohérence de la nomenclature.
-->
<!ATTLIST form
  type (lemma|variant|compound|derivative|inflected) #IMPLIED
  n (1|2|3|4|5|6|7|8|9|10|11|12|13|14|15|16|17|18|19|20) #IMPLIED
>
<!-- L'indication de responsabilité selon la légende siglæ,

```

```

        sera appliquée automatiquement sur les puces. -->
<!ENTITY % attribution '(ducange|benedictin|carpentier|didot|favre|enc) '>
<!--
@attr xml:id Identifiant d'article construit sur la vedette et numéro
d'homographe, obligatoire après traitement.
@attr resp Attribution d'un article selon la légende siglæ. -->
<!ATTLIST entry
    xml:id ID #IMPLIED
    resp %attribution; #IMPLIED
>
<!-- @attr resp Attribution d'un paragraphe selon la légende siglæ. -->
<!ATTLIST dictScrap
    resp %attribution; #IMPLIED
>
<!-- @attr resp Attribution d'un ajout selon la légende siglæ. -->
<!ATTLIST add
    resp %attribution; #IMPLIED
>
<!--
@attr cRef Lien mort, cible absente après post-traitement.
@attr target Lien vérifié, cible correspondant à une vedette numérisée.
@attr xml:lang La déclaration de langue des renvois est en réserve pour les
mots en ancien français liable
au glossaire de l'annexe. -->
<!ATTLIST ref
    cRef CDATA #IMPLIED
    target CDATA #IMPLIED
    xml:lang %lang; #IMPLIED
>
<!-- Dans les citation de Favre, un mot en italique est un rappel de la
vedette. -->
<!ELEMENT oVar (#PCDATA)>
<!-- Les sauts de ligne sont apparus utiles dans certains blocs très longs.
-->
<!ELEMENT lb EMPTY>

<!-- <h2 id="bloc-experimental">Post-traitements non programmés.</h2>

Les déclarations qui suivent sont des idées d'enrichissement dont la
possibilité
et l'utilité n'a pas encore été assurée.

-->

<!-- La langue des citations (latin ou ancien français) pourrait être
obtenue par post-traitement avec une implémentation N-Grams. -->
<!ATTLIST quote
    xml:lang %lang; #IMPLIED
>
<!-- Les références bibliographiques pourraient être encadrées afin
de diminuer le bruit de la recherche inverse sur les gloses,

```

et faire partie d'une recherche sur les citations.  
 Le processus de traitement n'est pas encore assuré,  
 le balisage des nombres pointés et des abréviations courantes  
 pourrait y aider. -->

```
<!ELEMENT bibl ( #PCDATA | abbr | num | hi | %breaks; )*>
<!-- Balise les abréviations dans une référence bibliographique. -->
<!ELEMENT abbr (#PCDATA)>
<!-- Les références bibliographiques encadrées
pourront être rattachées au texte cité et enchassées dans un élément
<cit>.
On y trouvera un connecteur ":" "in" et des possibles sauts de page ou de
colonne. -->
<!ELEMENT cit (#PCDATA | bibl | quote | %breaks; )* >
<!-- Typage possible des renvois et des mots étrangers selon des
introduceurs, comme "a" ou "ex" pour l'étymologie.
-->
<!ELEMENT xr (#PCDATA | ref | foreign )*>
<!-- @attr type Type de relation à la vedette pour un renvoi ou un mot
étranger selon une liste normalisée en TEI.
-->
<!ATTLIST xr
  type (syn|etym|cf) #IMPLIED
>

<!--
<h2 id="bloc-tei">Déclarations minimales pour un document TEI conforme</h2>
```

Le cartouche d'en-tête TEI sera défini strictement lorsque le schéma XML  
 sera complètement établi. Il contiendra notamment une description  
 structurée des observables de l'édition Favre auxquels se rapportent  
 chaque élément.

```
-->
<!-- -->
<!ELEMENT TEI (teiHeader,text)>
<!ELEMENT teiHeader (fileDesc)>
<!ELEMENT fileDesc (titleStmt,publicationStmt,sourceDesc)>
<!ELEMENT titleStmt (title)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT publicationStmt (p)>
<!ELEMENT sourceDesc (p)>
<!ELEMENT p (#PCDATA)>
<!ELEMENT text (body)>
<!ELEMENT body ( entry | %breaks;)+>
```