L'édition de textes fragmentaires en TEI-XML

Stratégies d'encodage

Estelle DEBOUY
ARSCAN (UMR 7041)

Un certain nombre de textes n'est connu que par la tradition indirecte, c'est-à-dire qu'ils ont été transmis sous la forme de citations chez d'autres auteurs. Ces fragments peuvent se présenter sous la forme de citations précises ou bien sous la forme de simples allusions, ou encore de paraphrases plus ou moins éloignées du texte original. En les insérant dans son texte, un auteur les préserve de l'oubli tout en leur conférant d'autres significations[1]. C'est précisément cette double caractéristique que doit restituer une édition numérique de textes fragmentaires : l'éditeur doit s'attacher à représenter les citations des fragments afin d'être en mesure de les repérer, de les extraire et de fournir par la même occasion un inventaire précis des auteurs et des œuvres citées ; il doit également éditer le contexte dans lequel ces citations apparaissent, car c'est le contexte qui lui permettra de proposer une lecture interprétative et scientifique du texte.

L'objet de cet article est de rendre compte de la réflexion méthodologique qui a accompagné la mise en œuvre d'une telle édition numérique. Cette réflexion a été entamée dans le cadre du projet d'édition de textes latins *The Digital Latin Library* (Huskey et Cayless, 2020) ; elle a aussi été menée plus précisément dans le cadre de l'édition de corpus de textes anciens fragmentaires, comme *The Leipzig Open Fragmentary Texts Series* (LOFTS), qui s'insère dans l'*Open Philology Project*[2] et dont l'ambition est de numériser des éditions imprimées d'œuvres fragmentaires, de faire des liens vers les sources, mais également de produire directement des éditions numériques d'œuvres fragmentaires (Berti, 2018). À titre d'exemple, citons au sein du LOFTS le projet

[1] Sur le fragment défini comme une réalité textuelle, voir Daviet-Taylor et Gourmelen (2016).

[2] Le projet est une initiative de « The Humboldt Chair of Digital Humanities » de l'université de Leipzig ([http://sites.tufts.edu/perseusupdates/2013/04/04/the-open-philology-project-and-humboldt-chair-of-digital-humanities-at-leipzig/], consulté le 13 juillet 2021).

Digital Athenaeus[3] ou encore l'édition des fragments des historiens grecs dont la version en ligne, réalisée sous la direction de Monica Berti (2019), reproduit l'édition imprimée de Müller (1841-1872).

En m'appuyant sur ces travaux, je souhaite réfléchir à la façon de présenter en TEI les éléments incontournables dans une édition numérique de fragments : les références aux auteurs grâce auxquels les fragments nous sont parvenus (*apparatus testium*) ; les variantes des fragments présentées dans un apparat critique positif, c'est-à-dire qui fait explicitement figurer les leçons de chaque témoin préalablement déclaré dans les métadonnées (que ce soit une source manuscrite ou imprimée) ; et le contexte dans lequel les fragments apparaissent, car l'un des buts poursuivis dans une édition numérique de fragments est bien de dépasser les limites des corpus relevant de la culture de l'imprimé où les citations sont reproduites sous la forme d'extraits décontextualisés. Pour illustrer mon propos, je prendrai ici un exemple concret, celui des fragments latins d'atellanes. Après avoir présenté les stratégies d'encodage en TEI, je montrerai comment le modèle TEI adopté peut être produit automatiquement à partir d'une saisie en LuaLaTeX grâce au logiciel *ekdosis* (Alessi, 2020a).

1 Texte et contexte

Il ne reste que quelque trois cents vers d'atellanes, ces comédies latines écrites au I^er^ siècle av. J.-C. et transmises sous la forme de citations par le grammairien latin Nonius (IV^e^ siècle apr. J.-C.) qui, pour composer son dictionnaire (*De compendiosa doctrina*), a choisi des extraits d'auteurs servant à illustrer l'emploi d'un mot ou d'une locution rare. Comme l'œuvre d'Athénée citée supra, le traité de Nonius est donc une source précieuse de textes perdus qu'il faut pouvoir encoder de façon à les identifier, les analyser et les extraire. Le grammairien suit toujours la même méthode : il énonce le mot étudié, en explique la signification et cite des auteurs chez qui le terme apparaît. En voici un extrait où c'est le nom *senica* qui est étudié[4] :

SENICA significat senex. Pomponius Pictoribus (111):

páppus hicin Máedio habitat, séneca non sescúnciae?

idem Praecone Posteriore (131):

sét me exercet sénica nequam, néque illo quid faciám scio.

9 s. a. *Cic. Tusc. II, 23*: tui mei volans: tum av. *Mu.* ‖ **11** qui—sunt *secl. On.*: qui et m. d. s. *edd.* ‖ **13** manducus: *corr. Mu.* ‖ casmillus *Quich.*: *fort.* camellus, *i. e.* cāmus *diminutive* ‖ **17** uti *Meinek.*: ut ‖ **19** hic cinaedus (?) *Mu.* ‖ Maedio *scripsi*: medio: medicus *Ribb.* ‖ sescuntiae L^1: -tia *BA* ‖ **22** Cradurius L^1 ‖ **24** victor ‖ **27** te medius: *corr. On.* ‖ tu et *On.*: tute.

Fig. 1 : Article *senica* dans le *De compendiosa doctrina*[5]

[3] L'œuvre d'Athénée, qui se présente comme une compilation en quinze livres de propos de table de savants grecs et romains qui foisonnent de citations d'auteurs anciens, s'avère être une bibliothèque inestimable d'œuvres perdues ([http://digitalathenaeus.org/], consulté le 13 juillet 2021).

[4] Nonius ne cite ici que deux exemples à titre d'illustrations, tous les deux tirés des pièces atellanes. Cela peut s'expliquer par le fait que le mot étudié est un hapax dans la littérature latine. Il s'agit donc d'une invention verbale de l'auteur d'atellanes Pomponius.

[5] Toutes les images sont produites par l'auteur de l'article.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

Observons maintenant l'entrée précédente (*manducones*) :

MANDVCONES, qui manduci dicti sunt, et MANDONES, edaces. Pomponius Pictoribus (112):
magnús manduco's, cámillus cantérius.
Lucilius (inc. 163):
atque ómnes mandonúm gulae.
Varro Bimarco (53):
mágna uti treméscat Roma et mágnae mandonúm gulae.
SENICA significat senex. Pomponius Pictoribus (111):
páppus hicin Máedio habitat, séneca non sescúnciae?
idem Praecone Posteriore (131):
sét me exercet sénica nequam, néque illo quid faciám scio.

FIG. 2 : Articles *manducones* et *senica* dans le *De compendiosa doctrina*

Pour illustrer ce nom, le compilateur a également recours à un vers d'atellanes. L'éditeur de la fin du XIX^{e} siècle, Lindsay (1903), a montré qu'il est tout à fait vraisemblable que l'entrée des mots dans le dictionnaire de Nonius corresponde à l'ordre dans lequel ils se présentaient dans ses sources[6]. On comprend donc que dans la pièce *Pictores*, le vers *Magnus* [...] *canterius* apparaît avant *Pappus hic* [...] *sescunciae*, contrairement à ce qu'on lit dans l'édition Ribbeck de 1898 dont la numérotation des vers figure entre parenthèses sur les figures 1 et 2. Le contexte est donc ici fort précieux puisqu'il permet d'ordonner les fragments[7]. Pour en prendre la mesure, il est nécessaire de ne pas se contenter de faire apparaître ces éléments de contextualisation dans l'apparat présentant les *testimonia*, mais bien de donner au lecteur accès au contexte dans son ensemble.

2 Stratégies d'encodage en TEI-XML

On l'a dit, au moins trois éléments incontournables doivent être encodés dans une édition critique de fragments :

1. Les références à l'auteur ou aux auteurs qui citent les fragments.
2. Les variantes.
3. Le contexte dont on vient de présenter l'intérêt à travers un exemple.

[6] Cette hypothèse est aujourd'hui désignée sous l'expression « loi de Lindsay » (Nonius Marcellus, 1901). Lindsay a montré que Nonius dressait des listes de citations qu'il copiait au fur et à mesure de ses lectures et dans lesquelles il puisait ensuite pour illustrer les mots qu'il étudiait dans son dictionnaire. D'après Lindsay, il en avait ainsi constitué quarante-trois.

[7] L'accès au contexte ne permet pas seulement d'ordonner les fragments : il permet aussi de confirmer une leçon donnée par un manuscrit ou une famille de manuscrits contre une autre leçon lue ailleurs. En effet, comme je l'ai montré dans ma communication présentée au sein du groupe de travail « Les éditions critiques numériques : langues, textes et fragments » qui s'est réuni le 11 mai 2021 lors du colloque Humanistica, dans le fragment tiré de l'atellane *Maccus et son jumeau* le texte de Nonius a permis de confirmer la leçon *abscondisti* qu'on lit dans la plupart des manuscrits contre *abscondidisti* qu'on lit dans le seul manuscrit H, car le grammairien s'intéresse à cette forme verbale particulière que le copiste du manuscrit H a corrigée par erreur.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

2.1 Les références aux auteurs qui citent les fragments

Ce point est parfaitement documenté dans les *Guidelines for Encoding Critical Editions for the Library of Digital Latin Texts*[8]. Samuel J. Huskey et Hugh Cayless expliquent en quoi un *apparatus testium* diffère d'un *apparatus fontium* : alors que ce dernier liste les références aux auteurs cités dans le texte de l'édition lui-même, le premier liste les références aux auteurs qui citent le texte édité. Pour reprendre l'exemple présenté plus haut, on trouvera dans l'*apparatus fontium* de l'édition du compilateur Nonius toutes les références aux auteurs chez qui il puise ses exemples pour illustrer son étude de la langue[9] ; et on trouvera dans l'*apparatus testium* de l'édition des atellanes les références aux auteurs grâce à qui les fragments nous ont été transmis (en l'occurrence presque 80% des fragments ont été transmis par Nonius).

Bien souvent, il est pertinent de faire apparaître ces références dans un appareil de notes dédié que l'on introduira ainsi : <note type="testium"> </note>. Il restera alors à indiquer dans la note le nom de l'auteur qui cite les fragments et l'œuvre où la référence apparaît grâce aux éléments <author> et <title> (dans <bibl>), sans oublier la référence précise au passage grâce à <biblScope>. Pour éviter de saisir de multiples fois une même référence identifiée par ailleurs dans la section bibliographique du fichier xml, on peut utiliser l'élément <bibl> avec l'attribut @corresp de la manière suivante :

```
<bibl corresp="#entreebib">
  <biblScope>xxx</biblScope>
</bibl>
```

où "#entreebib"désigne le @xml : id affecté à l'entrée référencée et xxx le numéro du chapitre, de la page, du paragraphe, etc.

Si l'on reprend notre exemple, on obtient :

```
<bibl corresp="Nonius">
  <biblScope>17</biblScope>
</bibl>
```

Cela évite de saisir les éléments suivants à chaque fois que la référence au même ouvrage de Nonius doit être introduite :

[8] Cf. le chapitre 10 « Apparatus Testium » ([https://digitallatin.github.io/guidelines/LDLT-Guidelines.html#testimonia], consulté le 13 juillet 2021).

[9] Ses sources sont les auteurs qu'il qualifie d'*honesti auctores*, c'est-à-dire les auteurs qui ont vécu avant Auguste, à l'exclusion des poètes nouveaux comme Catulle par exemple. Les auteurs d'atellanes y ont donc toute leur place.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

```
<bibl>
  <author>Nonius</author>
  <title>De compendiosa doctrina</title>
  <biblScope>17</biblScope>
</bibl>
```

2.2 Variantes du texte

Si trois méthodes sont proposées dans le module 12.2 de la TEI, la méthode *location-referenced*, la méthode *double-end-point-attached* et la méthode *parallel segmentation*, c'est cette dernière qui est traditionnellement retenue, car elle permet une saisie plus facile et plus précise. Appliquons-la à notre vers d'atellane :

```
<l n="1">
  Pappus
    <app>
    <lem wit="#E #F #H #L #P">hic in medio</lem>
    <rdg wit="#W">hic medio</rdg>
    <rdg source="#Ribbeck">hicin medicus</rdg>
    </app>
  habitat, seneca non sescunciae.
</l>
```

On voit ici que plusieurs distinctions ont été faites :

1. entre la leçon retenue, saisie avec l'élément <lem>, et les variantes rejetées par l'éditeur, saisies avec l'élément <rdg> (pour *reading*). Les *Guidelines for Electronic Text Encoding and Interchange* précisent bien que l'élément <lem> peut notamment être utilisé pour indiquer la préférence de l'éditeur, étant entendu que « the term *lemma* is used here in the text-critical sense of "the reading accepted as that of the original or of the base text"[10] ».
2. entre les différents témoins : l'éditeur peut citer les leçons de tel ou tel manuscrit (<wit> pour *witness*), présenter les conjectures ou les notes de tel ou tel éditeur moderne (<source>), ou encore prendre en compte ses propres conjectures, souvent introduites par *scripsi* ou *addidi* (<resp>).

[10] *Cf.* le chapitre 12.1.2 « Readings » (https://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html#TCAPLL), consulté le 13 juillet 2021.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

2.3 Présentation du contexte

L'éditeur peut vouloir présenter une édition de fragments afin de donner au lecteur la possibilité de consulter le contexte dans la marge, en regard des fragments eux-mêmes. Cette première stratégie est pertinente quand les fragments sont cités par un grand nombre d'auteurs. C'est celle qui a été adoptée, par exemple, pour l'édition des fragments de l'historien Néanthe de Cyzique dans le projet DFHG (*Digital Fragmenta Historicorum Graecorum*[11].

Une autre stratégie consiste au contraire à présenter une édition de la compilation où apparaissent les fragments et à donner la possibilité au lecteur de les extraire. On peut citer ici, comme exemple de compilation scientifique, l'édition du *De Piscibus* publiée par le pôle Document numérique de Caen[12].

Un encodage possible appliqué, dans chacun des deux cas, à l'exemple d'atellane présenté plus haut donnerait :

1. Pour l'édition des fragments et la présentation du contexte en regard :

```
<div xml : id="div-context*1"> Seneca significat senex. Pomponius
  Pictoribus.</div>
   <div xml : id="div-edition*1">
    <lg>
     <l>
      <note type="testium"target="#right(Nonius.17)">cf.
       <bibl>
        <author>Nonius</author>
         <title>De compendiosa doctrina></title>
        <biblScope>17</biblScope>
       </bibl>
      </note>
     <anchor xml : id="Nonius.17"/>
    Pappus […] sescunciae.
   </l>
  </lg>
</div>
```

[11] Voir http://www.dfhg-project.org/DFHG/index.php?volume=Volumen\%20tertium\#, consulté le 13 juillet 2021.

[12] Le texte est édité par Catherine Jacquemard, Brigitte Gauvin et Marie-Agnès Lucas-Avenel, avec la collaboration de Caroline Février et Françoise Lecocq (https://www.unicaen.fr/puc/sources/depiscibus/consult/hortus_fr/FR.hs.4.4, consulté le 13 juillet 2021).

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

On comprend que l'auteur grâce à qui les fragments sont parvenus est Nonius (*apparatus testium*), que le passage où apparaît le fragment se trouve p. 17 du *De compendiosa doctrina* et que le compilateur cite le vers d'atellane à l'occasion de l'étude du mot *seneca*.

2. Pour l'édition de la compilation (où sont cités les fragments) et le repérage des fragments :

```
<div xml : id="div-edition*1">
  <p>
   Seneca significat senex.
   <note type="fontium"target="#right(Pomponius155)">
     <bibl>
       <author>Pomponius</author>
       <title key="Pictores">Pictoribus</title>
       <biblScope>155</biblScope>
     </bibl>
   </note>
   <anchor xml : id="Pomponius155"/>
  </p>
  <seg type="atellanes">
   <l>Pappus [...] sescunciae.</l>
  </seg>
</div>
```

On comprend ici que le compilateur Nonius, qui étudie le mot *seneca*, cite un vers d'atellane de Pomponius (*apparatus fontium*), tiré de la pièce *Pictores*, afin d'illustrer son propos.

3 Inversion de stratégie : pour un encodage premier en LaTeX

Les rares projets d'éditions critiques numériques[13] reposent sur l'utilisation de l'encodage XML-TEI qui permet de transformer le texte et son apparat en une base de données complète consultable par le lecteur en fonction de ses besoins. Comme on vient de le voir, la TEI fournit en effet des éléments normalisés très précis pour la

[13] En effet rares sont les éditions critiques numériques car, dans la plupart des cas, le texte est présenté sans préface qui étudie l'histoire du texte, sans bibliographie et surtout sans apparat critique. Or, comme l'écrit Frédéric Duval (2017), « la place de la "critique" s'est affirmée comme un enjeu important des *digital scholarly editions* afin d'asseoir leur statut académique à côté des éditions critiques traditionnelles où l'adjectif "critique" implique une prise en compte de la tradition textuelle, qu'il s'agisse d'une édition reconstructionniste ou d'une édition conservatrice ».

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

description des sources manuscrites et pour l'encodage de l'apparat critique. Le texte de l'édition étant encodé en XML-TEI, se pose alors la question de la visualisation des données car, au-delà d'une consultation en ligne, on peut vouloir lire l'édition annotée dans une version imprimée. Dans le cas de ces éditions numériques encodées en XML-TEI, le format XML est ensuite transformé en format LaTeX pour produire *in fine* un fichier au format PDF. Mais l'apparat critique généré à partir de ces données brutes se présentera comme un *apparat de données* et non comme un *apparat rédigé,* à moins d'appliquer, lors de la transformation, des algorithmes spécifiques qui restent à écrire, à supposer que les éditeurs soient disposés à les utiliser. À ce titre, l'approche méthodologique adoptée par le « Pôle Document numérique » de l'université de Caen est intéressante[14] : l'éditeur saisit toujours son texte en XML-TEI, mais il a la possibilité d'utiliser l'attribut @cause pour introduire un certain nombre d'expressions latines dans l'apparat critique imprimé[15]. Cependant le nombre des expressions disponibles dans le schéma est limité et l'éditeur n'est pas libre de fixer l'endroit où elles apparaissent. Enfin, cette méthode ne permet pas de fusionner dans l'apparat critique imprimé plusieurs unités critiques correspondant à une structure XML dans laquelle on trouve des éléments <app> imbriqués.

Marina Buzzoni (2016) a d'ailleurs bien souligné les insuffisances de la TEI pour l'encodage d'apparats dynamiques : « Encoding a long ("verbose") apparatus entry is most of the times extremely time consuming, very tricky and often frustrating, especially when information needs to be suppressed because of the lack of elements and/or attributes, or even because it is not always possible to break up the information into consistent base-units in order to keep all witnesses synchronised ». Or, il est important que l'éditeur puisse non seulement saisir les variantes du texte mais encore, s'il le souhaite, présenter au lecteur les arguments qui lui ont permis de prendre telle ou telle décision.

C'est ce que permet le logiciel *ekdosis*[16] développé par Robert Alessi qui a fait le choix d'une approche différente par rapport aux projets mentionnés plus haut : l'éditeur saisit tous les éléments constitutifs de l'édition dans un fichier source LaTeX et demande ensuite une sortie aussi bien vers un fichier PDF que vers un fichier XML-TEI qui pourra être l'objet de requêtes[17]. Cette démarche repose sur le constat que LaTeX[18], contrairement à XML, est un langage naturellement conçu pour être utilisé dans la composition de documents complexes, tels les éditions critiques où les notes de bas de page et autres types d'annotations peuvent être particulièrement abondantes. Cette approche permet à l'éditeur d'écrire l'apparat critique sous la forme d'un paragraphe

[14] Le Pôle Document numérique de Caen propose aux éditeurs de saisir leur texte avec XML Mind (http://www.xmlmind.com/xmleditor/, consulté le 13 juillet 2021).

[15] *Cf.* la méthodologie d'encodage présentée dans le manuel disponible en ligne (https://www.unicaen.fr/recherche/mrsh/sites/default/files/public/document_numerique/manuel_apparat.pdf, consulté le 13 juillet 2021). L'utilisation de cet attribut est conforme aux recommandations qu'on lit dans le chapitre 12.1.2 Readings du *Guidelines* de la TEI.

[16] Le logiciel et sa documentation sont disponibles sur le site du CTAN (https://www.ctan.org/pkg/ekdosis, consulté le 13 juillet 2021).

[17] Pour une présentation plus détaillée du logiciel *ekdosis*, voir Alessi (2020b).

[18] LaTeX est un interpréteur de haut niveau reposant sur une collection de macro-commandes intégrant toutes les fonctions TeX. Pour compléter les fonctions de LaTeX, l'utilisateur peut charger des packages tiers. Depuis le 2 février 2020, le format de LaTeX contient la couche de programmation de la toute dernière version, appelée LaTeX3.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

continu, librement rédigé en latin, tout en encodant strictement les données dans le fichier de sortie XML-TEI. Le logiciel *ekdosis* permet donc de :

- rédiger tout un appareil critique complexe. Les textes édités peuvent être présentés avec plusieurs étages d'apparats, un *apparatus criticus* (qui liste les variantes), un *apparatus fontium* (qui liste les références au texte édité), un *apparatus testium* (qui liste les *testimonia*), ainsi que toutes sortes de notes complémentaires au texte lui-même.
- produire un fichier XML-TEI : *ekdosis*, à partir de la seule saisie en LaTeX, peut générer à la fois un PDF et un fichier XML-TEI pour toutes sortes de requêtes sur les données.
- présenter des éditions critiques multilingues avec alignement (ou non) du texte et de la traduction, voire des traductions. Mais *ekdosis* peut aussi être adapté aux modules de la TEI pour l'encodage de textes spécifiques tels que des transcriptions.

Afin de mieux prendre la mesure des différences de stratégie d'encodage entre l'approche qui consiste à partir du XML-TEI avant d'exporter les données en PDF *via* LaTeX et l'approche inverse adoptée dans le logiciel *ekdosis*, examinons maintenant plus en détail à la fois comment se présente le fichier source LaTeX saisi par l'éditeur et quels sont les résultats des exports TEI et PDF automatiquement produits par *ekdosis*.

3.1 Édition des fragments et contexte en regard

Tout fichier LaTeX est composé de deux parties, le préambule et le corps du document lui-même : dans le préambule, on trouve les réglages et les instructions générales concernant l'ensemble du document, comme la taille et le choix des polices, le choix des langues, les éléments du titre du document, ainsi que les extensions que l'on souhaite charger nommées *packages*. Voici la ligne qui doit figurer dans le préambule pour charger le package *ekdosis* :

```
\usepackage[teiexport=tidy]{ekdosis}
```

où l'option [teiexport=tidy] permet de produire automatiquement le fichier PDF mais aussi le fichier XML-TEI. Il convient ensuite de déclarer dans le préambule plusieurs éléments qui apparaîtront dans l'édition, à savoir les apparats, les témoins ou familles de témoins, et les diverses sources utilisées :

1. La commande \DeclareApparatus permet de paramétrer les apparats qui se présenteront sous la forme de deux étages de notes critiques, l'*apparatus testium* (ou *testimonia*) et l'apparat critique.
2. La commande \DeclareWitness permet de déclarer les témoins. En voici un exemple : \DeclareWitness{H}{\emph{H}}{\emph{Harleianus}(Mus. Brit.2719)}

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

[origDate={s. \textsc{ix-x}}] où le premier argument correspond à l'identifiant du témoin qui sera utilisé comme xml : id, le deuxième argument est le signe qui sera utilisé dans l'apparat critique, et le dernier argument présente une courte description nécessaire à la construction du *conspectus siglorum* qui apparaîtra au début de l'édition. Ce qui apparaît entre crochets droits à la fin de la commande est optionnel.

3. La commande \DeclareSource permet de déclarer les conjectures lues dans une édition moderne.

Une fois ces éléments définis dans le préambule, l'éditeur peut saisir l'édition elle-même dans le corps du document. Il utilise alors les environnements context et edition. L'environnement alignment, quant à lui, permet d'indiquer à LaTeX qu'il faut placer le texte et son contexte en regard. Comme on le voit dans l'exemple présenté ici, l'environnement alignment du fichier LaTeX admet plusieurs options : l'option tcols définit le nombre total de colonnes de texte à aligner et l'option lcols définit le nombre total de colonnes à imprimer sur la page de gauche. La valeur par défaut est 1 pour une présentation du contexte et de la traduction sur deux pages en vis-à-vis. Dans notre exemple, la valeur choisie est 2 pour une présentation sur une même page.

Comme on peut le voir ci-dessous, l'encodage utilisé pour présenter les variantes rappelle la syntaxe du XML-TEI. Voici, à gauche, le fichier source LaTeX et à droite, l'export TEI généré automatiquement par *ekdosis* :

\begin{alignment}[tcols=2, lcols=2,

texts=context ; edition]

\begin{context}

Seneca significat senex. Pomponius Pictoribus.

\end{context}

\begin{edition}

\begin{ekdverse}

\note[type=testium, labelb=Nonius.17,

lem=Seneca – sescunciae]

{*cf.* \icite[17]{Nonius}}

Pappus

\app{

\lem[wit={E, F, H, L, P}]{hic in medio}

\rdg[wit={W}]{hic medio}

\rdg[source={Ribbeck}]{hicin medicus}}

<**div xml : id="div-context*1"**>Seneca significat senex. Pomponius Pictoribus.

</**div**>

<**div xml : id="div-edition*1"**>

<note type="testium"target="#right (Nonius.17)">*cf.*

<bibl corresp="#Nonius">

<biblScope>17</biblScope>

</bibl>

</note>

<anchor xml : id="Nonius.17"/>

<lg>

<l>Pappus

<**app**>

<**lem** wit="#E #F #H #L #P">hic in medio </**lem**>

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

habitat, seneca non sescunciae.	<**rdg** wit="#W">hic medio</**rdg**>
\end{ekdverse}	<**rdg** source="#Ribbeck">hicin medicus</**rdg**>
\end{edition}	</**app**>
\end{alignment}	habitat, seneca non sescunciae.</l>
	</lg>
	/div>

Le fichier PDF, également généré automatiquement par *ekdosis*, présente donc le fragment et son contexte en regard, puis les deux étages de notes critiques.

Seneca significat senex. Pomponius Pictoribus.

Pappus hic in medio habitat, seneca non sescunciae. 1

Test. : 1 Seneca– sescunciae] cf. Nonius Marcellus, *De compendiosa doctrina*, 17.

1 hic in medio *E F H L P*] hic medio *W* hicin medicus Ribb.

Fig. 3 : Édition numérique du fragment de Pomponius avec son contexte en regard après export PDF avec *ekdosis*

3.2 Édition de la compilation et repérage des fragments

On vient de voir que la première méthode pour présenter une édition de fragments consiste à éditer les fragments pris isolément et à faire figurer le contexte dans lequel ils apparaissent en regard de chacun d'eux. Dans ce cas, c'est la référence au témoin qui a permis de les conserver – ici le compilateur Nonius – qui figure dans le premier appareil de notes (\note[type=testium] dans le fichier LaTeX et <note type="testium"></note> en TEI). Une autre méthode consiste à éditer le texte du compilateur lui-même et à isoler les fragments qui sont l'objet de l'édition : dans ce second cas, ce sont les environnements ekdpar et segment qui seront utilisés dans *ekdosis*. Cette fois, c'est la référence aux sources utilisées par le compilateur – en l'occurrence les comédies atellanes – qui figure dans le premier appareil de notes (\note[type=fontium] dans le fichier LaTeX et <note type="fontium"></note> en TEI).

\begin{ekosis}	<div xml : id="div-edition*1">
\begin{ekdpar}	<**p**>
\note[type=fontium, labelb=Pomponius.155,	<note type="fontium"
lem=Pappus – sescunciae]	target="#right(Pomponius155)">
{*cf.* \icite[155]{Pomponius}}	<bibl corresp="#Pomponius">
Seneca significat senex.	<biblScope>155</biblScope>
\ekdauthor{Pomponius}	</bibl>

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

```
\ekdtitle[Pictores]{Pictoribus} :
\end{ekdpar}
\begin{segment}
\begin{ekdverse}
Pappus
\app{
\lem[wit={E, F, H, L, P}]{hic in medio}
\rdg[wit={W}]{hic medio}
\rdg[source={Ribbeck}]{hicin medicus}}
habitat, seneca non sescunciae.\\!
\end{ekdverse}
\end{segment}
\end{ekdosis}
```

```
</note>
<anchor xml : id="Pomponius155"/>
Seneca significat senex.
<author>Pomponius</author>
<title key="Pictores">Pictoribus</title>
</p>
<seg type="atellanes"xml : id="edition11">
<lg>
<l>Pappus
<app>
<lem wit="#E #F #H #L #P">hic
in medio</lem>
<rdg wit="#W">hic medio</rdg>
<rdg source="#Ribbeck">hicin medicus</rdg>
</app>
habitat, seneca non sescunciae.</l>
</lg>
</seg>
</div>
```

Comme on le voit dans le PDF produit (*cf.* la figure 4 ci-dessous), le contexte fait partie intégrante de l'édition, ce qui permet de mieux mesurer à quel titre les fragments étudiés nous sont parvenus. Plus qu'une édition des auteurs d'atellanes Pomponius et Novius, c'est une édition en contexte que permet ce type d'approche méthodologique.

1 Seneca significat senex. Pomponius Pictoribus :
Pappus hic in medio habitat, seneca non sescunciae. at-1

Sources : 1 Pappus – sescunciae] cf. Pomponius, *Atellanae*, v. 155.

at-1 hic in medio *E F H L P*] hic medio *W* hicin medicus Ribb. habitat *codd.*] inhabitat Frass. sescunciae *W*] sescuntiae *F H L* sescuntia *E P*

Fig. 4 : Édition numérique du *De compendiosa doctrina* avec le fragment de Pomponius en contexte après export PDF avec *ekdosis*

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

4 Conclusion

Cette étude a permis de présenter, à partir d'un exemple commenté, plusieurs stratégies possibles pour encoder une édition critique numérique des textes fragmentaires : directement en XML-TEI ou bien en (Lua)LaTeX. Le langage XML-TEI ayant naturellement vocation à être transformé en HTML, le but est de produire des éditions consultées en ligne sous la forme d'un texte continu dans lequel les variantes ne sont pas consignées dans un apparat critique mais présentées sous la forme de courtes annotations associées de façon indépendante au mot du texte sur lequel elles portent. D'un autre côté, la saisie en LaTeX repose sur la culture du livre imprimé dans lequel l'éditeur rédige de façon continue un apparat critique sous la forme d'un paragraphe indépendant placé au bas de chaque page et détaché des mots sur lesquels portent les variantes. Dans cette approche, la production d'un fichier XML est secondaire et consiste à extraire de ce paragraphe les variantes dépouillées d'annotations (ainsi que toutes les données qui auront été structurées) pour les soumettre à une machine à des fins d'exploration et d'analyse.

C'est cette seconde méthode que j'ai adoptée pour l'édition des fragments d'atellanes : dans un premier temps, j'ai donc saisi avec *ekdosis* tous les fragments dans un fichier LaTeX et, sur le modèle de la dernière version imprimée de l'édition de ces fragments latins (*cf.* Frassinetti 1967), j'ai fait figurer deux appareils de notes, l'un présentant les *testimonia* (essentiellement constitués des références au compilateur Nonius) et l'autre présentant les variantes textuelles. Le logiciel a généré automatiquement un fichier PDF pour la lecture et l'impression, et un fichier XML-TEI pour l'analyse. Dans un second temps, j'ai regretté de ne pas avoir fait apparaître à part entière le contexte dans lequel les fragments avaient été cités. C'est pourquoi j'ai repris le fichier source LaTeX que j'ai modifié de façon à intégrer le texte de Nonius ; dans cette dernière version de mon édition, le premier appareil de notes présente donc les références aux auteurs d'atellanes comme sources du compilateur (*cf.* le modèle présenté p. 10).

Pour le moment, l'encodage s'est concentré sur l'édition critique à proprement parler : il est donc possible, par exemple, de reconstituer tout le texte d'un manuscrit donné ou encore de dresser des listes de fautes afin de déterminer des familles de manuscrits. Il resterait à s'intéresser au texte des fragments eux-mêmes (langue riche en hapax, formes rares, noms propres, etc.), ainsi qu'au contexte (encodage des mots étudiés par le compilateur, et éventuellement encodage des noms des autres auteurs cités comme sources au même titre que les auteurs d'atellanes). Enfin se posera la question de la visualisation de l'édition en ligne : comme de nombreux outils existent déjà[19] et que d'autres sont en cours de développement[20], il conviendra d'étudier quel est le plus adapté au projet.

[19] Citons notamment MaX, un Moteur d'affichage XML développé par le pôle Document numérique de l'université de Caen [https://www.unicaen.fr/recherche/mrsh/document_numerique/outils/max] ou encore EVT [http://evt.labcd.unipi.it/] (ressources consultées le 13 juillet 2021).

[20] À partir de textes en sanskrit, voir : https://github.com/radardenker/sanskrit-editing-suite (consulté le 13 juillet 2021).

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352