



**HAL**  
open science

## Encodage structurel et ré-éditorialisation

Thierry Pellé, Bernard Jacquemin, Eric Kergosien, Charlotte Tournier,  
Séverine Tarantino

► **To cite this version:**

Thierry Pellé, Bernard Jacquemin, Eric Kergosien, Charlotte Tournier, Séverine Tarantino. Encodage structurel et ré-éditorialisation. Vers une épistémè numérique. Actes du 19e Colloque international sur le Document Électronique (CiDE.19), Nov 2016, Athènes, Grèce. pp.121-132. hal-01407433

**HAL Id: hal-01407433**

**<https://hal.univ-lille.fr/hal-01407433v1>**

Submitted on 6 Dec 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Encodage structurel et ré-éditorialisation

## *On structural encoding and repurposing*

Thierry PELLÉ (1), Bernard JACQUEMIN (2), Éric KERGOSIEN (2)

Charlotte TOURNIER (1), Séverine TARANTINO (1)

(1) HALMA, Université de Lille (France)

(2) GERiCO, Université de Lille (France)

Thierry.Pelle@etu.univ-lille3.fr

{Bernard.Jacquemin, Eric.Kergosien

Charlotte.Balandraud, Severine.Tarantino}@univ-lille3.fr

### Résumé

Le projet pluridisciplinaire TALIE est dédié, dans sa partie « Textes », à la valorisation de la tradition des œuvres de l'Antiquité gréco-romaine telle qu'elle est représentée dans les fonds anciens ou patrimoniaux des bibliothèques de la région Nord-Pas de Calais-Picardie, sous la forme des manuscrits portant ces œuvres, mais aussi des éditions anciennes qui en ont été données, la plupart du temps avec des commentaires, ainsi que de leurs traductions. Dans cet article, nous présentons une approche semi-automatique pour l'édition numérique du commentaire monumental de l'œuvre de Virgile par Juan Luis de La Cerda. Ce commentaire est en train d'être redécouvert, mais il est difficilement exploitable dans son format d'origine : organisé en trois niveaux (sujet, explication littérale, notes érudites), riche de très nombreuses références à d'autres textes, il appelle une lecture dynamique que seule une édition numérique est à même de proposer. Concernant la phase d'annotation sémantique, nous nous appuyons sur le langage de balisage de contenu XML-TEI et nous décrivons les choix de marquage réalisés en accord avec les experts philologues participant au projet.

**Mots-clés.** Fonds patrimonial, La Cerda, Edition numérique, Indexation TEI.

### Abstract

The multidisciplinary project called TALIE is dedicated to enhancing the tradition of classical antiquity works as represented in heritage collections of libraries located in the Nord-Pas de Calais-Picardy region. These collections often contain manuscripts, older edition, comments and translations. In this article, we present a semi-automatic

approach to the digital edition of the monumental commentary on the Virgil work made by Juan Luis de La Cerda. This comment is being rediscovered, but it is difficult to exploit in its original format: organized in three levels (subject, literal explanation, erudite notes). With numerous references to other texts, a dynamic reading is required. and a digital edition is able to offer it. Concerning the semantic annotation stage, we rely on XML-TEI content markup language. Then, our choices of indexing made in accordance with the philologists experts are described.

**Keywords.** Heritage collection, La Cerda, Digital publishing, TEI indexing.

## 1 Introduction

Les problématiques liées à l'édition électronique de documents anciens, en particulier celle des manuscrits médiévaux et des documents imprimés qui précèdent l'âge industriel, sont un sujet de préoccupation qui naît avec l'avènement du micro-ordinateur dans les communautés de recherche en Sciences Humaines. Il s'agit d'une part de transcrire des contenus dans l'univers numérique pour leur faire bénéficier des outils d'analyse, de recherche et de diffusion actuels, et de l'autre de rendre compte de la forme tant typographique qu'éditoriale du document original, nécessaire à l'étude de ces types d'ouvrages. Dans ce contexte, le projet pluridisciplinaire TALIE « Traditions [Textes, Traces] de l'Antiquité à Lille et dans l'Eurorégion »<sup>1</sup> a parmi ses objectifs de concevoir et de développer un prototype d'édition numérique du commentaire de l'œuvre de Virgile par Juan Luis de La Cerda. Ce commentaire est une œuvre monumentale d'environ 2000 pages réparties en 3 tomes (Stevens, 1945 ; Laird, 2002). Publié entre 1607 (date d'obtention de l'imprimatur) et 1642, il porte sur la totalité de l'œuvre de Virgile : les *Bucoliques*, les *Géorgiques* (premier tome), ainsi que les douze chants de l'*Énéide* (tomes 2 et 3). La phase actuelle du projet se concentre sur le début du commentaire à la troisième *Géorgique* contenue dans le tome I dans l'édition sortie de l'imprimerie lyonnaise d'Horace Cardon (en 1619), conservée à la Bibliothèque d'Agglomération de Saint-Omer. Dans ce texte, La Cerda fait appel à de très nombreux commentateurs classiques (près de 600 auteurs sont ainsi référencés dans l'*elenchus*, le *syllabus* et la fin des prolégomènes). La plupart du temps explicites, ces références sont aussi régulièrement incomplètes par rapport aux usages d'aujourd'hui (pas de numéros de vers s'il s'agit d'une œuvre poétique, pas toujours le numéro de chapitre pour une œuvre en prose). Des paraphrases « masquent » certaines cita-

---

1. Projet coordonné par le laboratoire HALMA (Histoire, archéologie et littérature des mondes anciens) et mené en collaboration avec des chercheurs des laboratoires GERiiCO (Sciences de l'information et de la communication) et STL (linguistique) de l'Université Lille 3, financement Université de Lille – MESHS.

tions, amenant à se demander si La Cerda cite de mémoire avec quelques inexactitudes, ou s'il tire volontairement des expressions (altérées dans l'opération) de textes vers lesquels il s'est expressément tourné. Le prototype résultant du projet, principalement destiné aux experts (scientifiques, latinistes, etc.), doit tout d'abord permettre la consultation de ce commentaire, qui n'est pas aisée à partir des versions numérisées disponibles en ligne (à cause de son format et de sa structure); sa lecture n'est pas satisfaisante non plus sur une version imprimée mais pour une raison qui tient plus au texte lui-même et au fait que par rapport à ce que nous avons l'habitude de chercher et trouver dans un commentaire, des informations manquent (comme les références précises des œuvres mentionnées ou citées). Le prototype permet dans un second temps aux scientifiques de mener des analyses fines du commentaire. Dans cette perspective, ce projet s'inscrit dans la pensée de l'*épistémè* foucauldienne (Foucault, 1968), puisqu'il concourt à provoquer une rencontre que nous souhaitons fructueuse entre des savoirs appartenant à des cultures épistémiques apparentées, mais qui se distinguent l'une de l'autre à travers les âges : ainsi l'*épistémè* de la Renaissance, matérialisée par le commentaire de La Cerda, s'appuie d'abondance sur les textes antiques (*épistémè* classique) pour interpréter une œuvre désormais rendue accessible à nos contemporains (*épistémè* moderne). Nous pensons que cette rencontre contient en germe des savoirs nouveaux qui ne demandent qu'à émerger.

Concrètement, il s'agit de compléter le fac-similé image par une interface web de consultation permettant d'accéder tant au commentaire latin original de La Cerda qu'à sa traduction en français, le tout enrichi par un appareil critique et des liens externes vers les nombreuses ressources mentionnées par l'auteur et de nombreux index (noms, lieux, lemmes, citations...) accompagnés d'outils de recherche et d'extraction multicritères. Des solutions logicielles existent pour l'édition de documents anciens mais celles-ci ne permettent pas de prendre en compte les spécificités du type de commentaires sur lesquels nous travaillons. Pour la dématérialisation des documents, nous proposons une chaîne de traitements classique en 4 étapes : (1) encodage en mode image avec le soutien de l'ANRT<sup>2</sup>; (2) ajout d'une couche éditoriale « techniques » par application d'un OCR (*Finereader 12 / ABBY Engine 10*); (3) ajout de couches éditoriales « critiques » par annotation sémantique s'appuyant sur le modèle de description TEI<sup>3</sup>; (4) conception et développement d'une interface pour la visualisation et l'analyse de commentaires.

Dans cet article, nous présentons d'abord les premiers résultats des travaux menés sur un extrait du commentaire de la troisième *Géorgique*

---

2. Agence Nationale de Reproduction des Thèses (Lille) qui apporte son savoir faire dans la numérisation et l'OCRisation des documents anciens.

3. Manuel d'encodage XML-TEI Renaissance et temps modernes (Imprimés – Manuscrits) : <http://www.bvh.univ-tours.fr/XML-TEI/index.asp>.

pour les phases 1 à 3 allant de la numérisation jusqu'à l'annotation sémantique. Puis nous présentons un état des modèles de description existants pour l'annotation sémantique. Nous énumérons également un ensemble de solutions numériques pour la valorisation et l'analyse de documents anciens. Ensuite, nous détaillons notre méthodologie pour la dématérialisation du commentaire de La Cerda et nous précisons les choix faits pour l'annotation sémantique. La dernière section conclut nos premiers travaux et dresse les perspectives à court et moyen termes.

## 2 État de l'art

Les problématiques liées à l'édition électronique de documents qui précèdent l'âge industriel sont un sujet de préoccupation qui naît avec l'avènement du micro-ordinateur dans les communautés de recherche en Humanités. Il s'agit en effet, d'une part, de transcrire des contenus dans l'univers numérique pour les faire bénéficier des outils d'analyse, de recherche et de diffusion actuels et, d'autre part, de rendre compte de la forme tant typographique qu'éditoriale du document original, nécessaire pour l'étude de ces ouvrages. Une réflexion menée dès 1987 au sein d'une communauté internationale d'acteurs de l'édition électronique et de l'ingénierie linguistique et textuelle, rejoints depuis par ceux des humanités numériques, a produit à partir de 1994 un schéma d'encodage et de description des textes qui s'appuie sur un langage de balisage des contenus pour rendre compte tant de la forme typographique et éditoriale du document originel que de la signification de ses contenus : la *Text Encoding Initiative* (TEI, Bauman et Catapano, 1999). Le langage de balisage XML, qui a remplacé en 2003 le langage SGML dans la dernière version de la TEI, présente divers avantages, en plus de sa capacité à étiqueter des contenus dans une structure conforme aux recommandations de la TEI. XML permet par exemple de définir des étiquettes supplémentaires susceptibles d'associer du sens aux bribes de texte marquées, ainsi que de faciliter la publication des textes sur le web soit directement, soit à travers une transformation XSL(T).

En dehors du XML-TEI, les années 90 ont produit d'autres formats de structuration du texte basés également sur un langage de balisage, comme DocBook, conçu pour formaliser la documentation techniques, EAD qui décrit les documents d'archives, ou HTML qui assure l'affichage des contenus sur le web. Cependant, aucun de ces formats ne peut rivaliser avec XML-TEI dans la description fine, à la fois physique et sémantique, des textes patrimoniaux et de leur contenu. En effet, HTML privilégie une mise en forme pour l'affichage et ne porte pas de sémantique (Rimelé, 2011), tandis que DocBook se contente de donner un sens généralement limité aux propriétés de mise en forme, sans permettre d'extension au-delà de l'étiquetage prévu (Walsh, 2010). EAD, restreint à 150

descripteurs, permet la description catalographique de fonds d'archives, mais en aucun cas l'encodage physique des documents eux-mêmes, ni surtout de leur contenu (Queroux, 2003).

Quant à la partie du commentaire de l'œuvre de Virgile par Juan Luis de La Cerda (le commentaire sur la troisième *Géorgique*) à partir de laquelle nous avons entrepris notre travail d'édition, la thèse espagnole de José Ortega Castejón (1990) en a assuré une édition moderne et même une diffusion électronique en PDF. Cependant, celle-ci reste difficilement exploitable en raison des erreurs relativement nombreuses qui émaillent le texte latin non traduit, et surtout à cause de l'incapacité du papier ou du PDF à assurer la mise en correspondance entre les trois facettes de la structure du document original – le texte de Virgile, l'*explicatio* et les *notæ* (*cf. infra*) – qui sont réparties dans trois parties successives dans la thèse, alors même que l'hypertexte est en mesure d'opérer facilement ces renvois.

Dans sa version électronique en PDF, J. Ortega fait se succéder l'*argumentum* de la séquence de vers, la citation des vers concernés par la première *explicatio*, l'*explicatio*, les *notæ* qui portent sur des mots apparaissant dans les vers concernés. L'auteur défend que les notes sont vraiment liées aux *explicationes*, ce qui ne nous a pas toujours semblé le cas. Il a donc proposé une structure à lui, impliquant une interprétation et différant de l'original; reste que c'est du PDF, ça se lit linéairement, cela morcelle aussi le texte et cela ne permet pas de va-et-vient entre le texte et son commentaire autant qu'on pourrait l'espérer (il faut imprimer, et en grande taille, pour s'y retrouver). Notre idée de départ était par ailleurs de mettre en valeur ce qui à nos yeux de modernes en particulier (et en particulier quand on est intéressé par les phénomènes et procédés de l'intertextualité) fait le cœur et l'intérêt dominant de ce commentaire : les notes (*notæ*). Inversement, l'*argumentum* ne nous paraît pas intéressant en tant que commentaire : nous assumons de le faire passer dans le texte de Virgile à la fin, parce qu'il sert fondamentalement et pratiquement à clarifier le découpage de ce dernier – en clair, à donner des sortes de titres ou de chapeaux introducteurs aux séquences de vers « découpées » pour l'édition.

Notre démarche propose de remédier à ces faiblesses en réalisant une édition numérique hypertextuelle des trois parties du texte et de leur traduction, en suivant d'abord le modèle *HyperDonat*<sup>4</sup> (Bureau *et al.*, 2016), tout en bénéficiant de l'expérience accumulée par l'édition numérique d'autres textes anciens : le *Tractatus de Piscibus* (projet Ichtya<sup>5</sup>, Jacquemard *et al.*, 2013), les *Bibliothèques Virtuelles Humanistes* (Demonet, 2004) et le portail *Renaissance d'Horace*<sup>6</sup>.

---

4. <http://hyperdonat.huma-num.fr>.

5. <https://www.unicaen.fr/puc/sources/depiscibus/recherche>.

6. <http://www.univ-paris3.fr/horace>.

### 3 Contribution

Le commentaire de l'œuvre de Virgile par Juan Luis de La Cerda présente une structure très particulière qu'il s'agit d'appréhender afin d'en comprendre la complexité pour en rendre compte dans l'interface de consultation finale. Cette structure s'appuie en partie sur les subdivisions traditionnelles thématiques de taille variable des poèmes virgiliens. Nous présentons en figure 1 à titre d'illustration un extrait exemplaire de cette structure. Il s'agit des 9 premiers vers de la troisième *Géorgique*, accompagnés de leur synopsis (*argumentum*), de quatre explications (*explicationes*) et des trois premières notes (*notæ*).

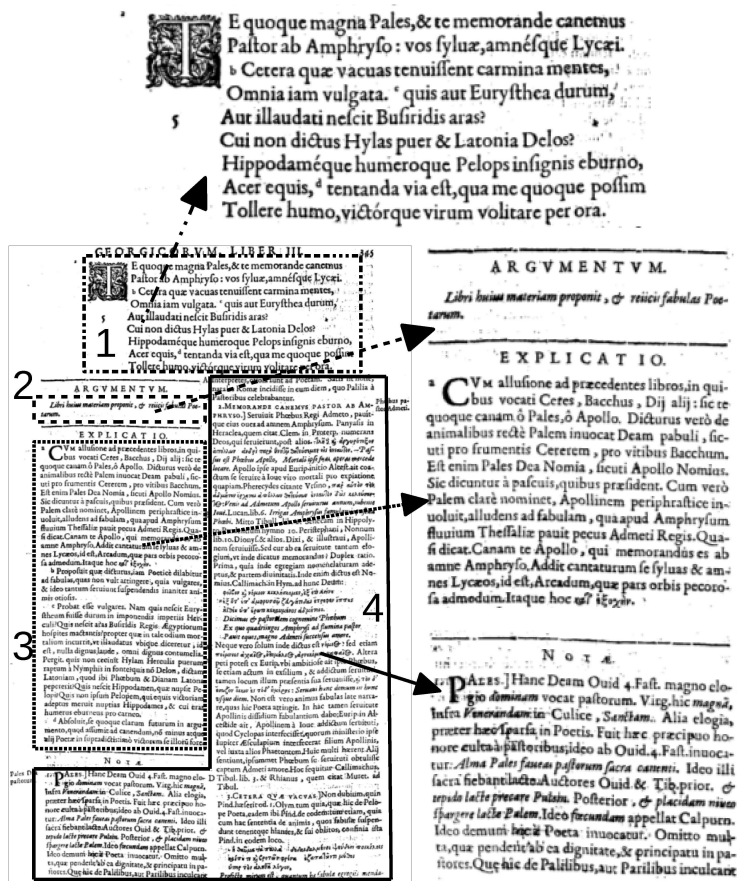


FIGURE 1 – Structure éditoriale chez La Cerda (1 - Texte original de Virgile ; 2 - Synopsis de la subdivision affichée en 1 ; 3 - Explications ; 4 - Notes).

### 3.1 Organisation du document original

La page 345 présentée sur la figure 1 permet de matérialiser la structure éditoriale appliquée à chaque subdivision. Les notes marginales sont utilisées comme entrée dans l'index *cum rerum et verborum tum auctorum* présent en fin d'ouvrage<sup>7</sup>.

Toutes les subdivisions comportent d'abord un extrait (numéroté 1 dans la figure 1) du poème de Virgile commenté, sous forme d'un groupe de vers consécutifs numérotés classiquement de cinq en cinq vers, flanqués d'indices alphabétiques placés en exposants qui servent de renvoi aux explications (*explicationes*, cf. *infra*).

Suit alors le commentaire avec sa structure à trois niveaux d'analyse :

- le synopsis (*argumentum*) expose brièvement le thème du passage commenté (numéroté 2 dans la figure 1);
- l'explication (*explicatio*) apporte un commentaire sur un segment plus ou moins conséquent de la subdivision traitée. Chaque explication, associée à un indice alphabétique, porte sur la portion de texte marquée à l'initiale par le même indice, et s'étendant jusqu'à l'indice suivant (ici l'indice « a » porte sur le segment « *Te quoque [...] Lycæi* »);
- la note (*nota*) fournit un commentaire plus ponctuel sur un ou plusieurs items du texte original. Chaque note est numérotée (ici « 1 ») et mentionne l'item (souvent lemmatisé) commenté (ici « *Pales* ») avant de produire son information. Il s'agit donc du niveau de commentaire le plus fin, qui offre notamment des références bibliographiques<sup>8</sup>, des citations en latin, et des citations en grec le plus souvent traduites en latin.

### 3.2 Du document original à son édition moderne électronique

Dans un premier temps, nous nous sommes servi du balisage TEI proposé par le projet *HyperDonat* puis celui-ci a progressivement été adapté aux spécificités de l'œuvre et aux objectifs éditoriaux. Les éléments TEI ont été choisis en collaboration avec les experts philologues participant au projet TALIE, et en fonction du sens porté par ces éléments.

---

7. Les lettres « A » à « D » entre les deux colonnes n'ont pas d'utilité connue.

8. La première référence donnée est *Ovid 4 Fast* pour le quatrième livre des *Fastes* d'Ovide. La référence complète que nous attendons aujourd'hui et que notre édition va donner (par un survol de la citation, dans une fenêtre *pop-up*) est Ovide, *Fastes*, IV, 476 (qui peut s'abrégéer d'après la pratique anglo-saxonne en Ov., F. 4.476; l'usage français donnerait plutôt Ov., *Fast.*, IV, 476).



### 3.2.1 Organisation en fichiers

Ce sont donc deux textes latins qui sont entrelacés, exception faite de quelques citations grecques : le poème de Virgile et le commentaire de La Cerda. Contrairement à la structure physique du document original, le poème et son commentaire sont scindés en deux fichiers distincts étiquetés selon le balisage sémantique TEI. En outre, dans la perspective du projet TALIE, chacun des textes est associé à sa traduction française. L'édition finale, et en conséquence l'interface de publication fait donc intervenir quatre fichiers dans lesquels les textes latins de Virgile et de la Cerda se répondent tout en étant connectés à leur traduction respective, qui ne sont pas directement liées entre elles. La figure 2 illustre les interconnexions entre ces quatre fichiers, tout en soulignant l'absence de lien direct entre les fichiers de traductions.

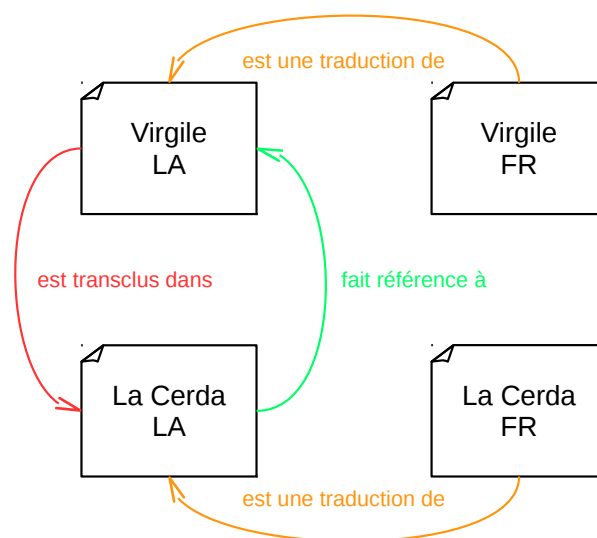


FIGURE 2 – Relations entre les fichiers de texte du projet TALIE.

### 3.2.2 Projection de la structure du document source

Intéressons-nous maintenant à la façon dont la structure du document source présentée précédemment section 3.1 a été transcrite en TEI.

- Le poème virgilien des *Géorgiques* est composé de 4 livres ou « chants ». Ces livres ont été découpés par La Cerda en 32 « chapitres » thématiques regroupant une suite de vers, qui sont constitués d'items (« mots »). La hiérarchie TEI est la projection de cette structure (figure 3).
- La traduction française du texte de Virgile, établie chapitre par

```
<div1 type="œuvre">
  <div2 type="livre" n="1">
    <div3 type="chapitre" n="1">
      <l> [niveau du vers]
      <w> [niveau de l'item latin]
```

FIGURE 3 – Structure TEI du texte virgilien originel.

chapitre, ne peut suivre cette structure qu'au niveau de l'œuvre, du livre et du chapitre. En TEI, chaque élément de traduction traduisant une phrase est enchâssé dans un paragraphe (figure 4).

```
<div1 type="œuvre">
  <div2 type="livre" n="1">
    <div3 type="chapitre" n="1">
      <p> [niveau de la phrase traduite en français]
```

FIGURE 4 – Structure TEI de la traduction du poème.

- Le commentaire latin et sa traduction répondent pour leur part à une structure identique, compatible avec celles du texte latin et de sa traduction : le commentaire est ainsi attaché à l'œuvre divisée en livres eux même-découpés en chapitres. La partie *Argumentum* du commentaire est identifiée au niveau du chapitre sur lequel il porte. La partie *explicatio* comme la partie *notæ* doivent s'exprimer à un niveau spécifique inférieur au chapitre qui permettra de les distinguer (figure 5).

```
<div1 type="œuvre">
  <div2 type="livre" n="1">
    <div3 type="chapitre" n="1">
      <div4 type="argumentum">...</div4>
      <div4 type="commentary">
        <seg type="explicatio" n="a">...</seg>
        <seg type="nota" n="1">...</seg>
```

FIGURE 5 – Structure du commentaire en latin et en français.

### 3.2.3 Transcription TEI de la structure et des contenus

**Interconnexion des contenus** Au-delà des liens établis entre les quatre fichiers de texte et commentaires numérisés et traduits (*cf.* figure 2), l'am-

bition du projet TALIE est d'offrir une interconnexion fine entre les contenus précis de ces fichiers qui se répondent, entre segments de poèmes et commentaires d'une part, entre passages latins et traductions françaises de l'autre. Pour ce faire, chaque élément TEI est dans un premier temps doté d'un attribut d'identification `xml:id`. Puis au moyen d'éléments `<span>` et `<link>`, on explicite les interconnexions entre les bribes du texte latin et les commentaires (*explicationes* et *notæ*) dont elles font l'objet, ainsi qu'entre les groupes de vers latins et leur traduction. Par exemple la mise en correspondance de l'*explicatio* (a) est encodée `<span from="#w1" to="#w15" corresp="#..." />` où « #... » est l'identifiant de l'*explicatio*.

**Détail de l'encodage structurel** L'un des avantages de l'encodage TEI (Burnard, 2016) est de proposer un choix étendu de balises permettant d'exprimer le point de vue adopté par l'éditeur scientifique sur l'œuvre éditée. Nous avons écarté l'approche quasi-diplomatique au profit d'un encodage privilégiant la mise en évidence des liens entre les textes. Par exemple (figure 6), l'encodage des *explicationes* et des *notæ* enchâsse les commentaires de La Cerda dans un segment (bloc de texte) typé (`@type`) mis en correspondance avec le texte commenté au moyen d'un *XPointer* (`@target`).

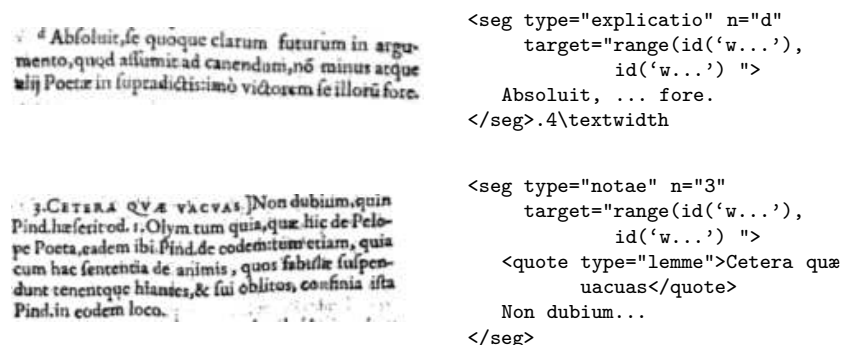


FIGURE 6 – Encodage des commentaires latins.

**Cas des références bibliographiques d'une note** Un autre avantage de TEI est de permettre l'annotation sémantique du texte au sein même de celui-ci, autorisant ainsi une lecture et des traitements à plusieurs niveaux. Par exemple, dans la portion du commentaire portant sur la déesse Pales « *Hanc Deam Ovid 4. Fast. magno elogio dominam vocat pastorum* », dont l'original est représenté figure 7, La Cerda fait référence à Ovide qui nomme la déesse « patronne » des bergers au livre IV des *Fastes* (vers 476).

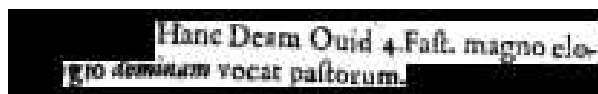


FIGURE 7 – Référence bibliographique d'une *nota*.

L'encodage de cette phrase (*cf.* figure 8) fait apparaître à un premier niveau la présence d'une référence bibliographique, puis précise successivement l'auteur, le titre de l'œuvre et le livre cité (élément `biblScope`). Dans un second temps, le terme d'Ovide cité « *dominam* » est non seulement indiqué comme tel, mais aussi rattaché à la source bibliographique (`@corresp` faisant référence à `@xml:id`). La référence bibliographique est également complétée (vers 476 des *Fastes*) dans un formalisme contemporain (`@cRef`), élément essentiel dans le travail d'encodage.

```
Hanc Deam
<bibl xml:id="#xxx">
  <author>Ouid</author> <biblScope>4.</biblScope>
  <title>Faft.</title>
</bibl>
magno elogio
<cit corresp="#xxx"> <cRef="Ovide, Fastes, IV, 476"/>
  <quote>dominam</quote>
</cit>
vocat paStorum.
```

FIGURE 8 – Référence bibliographique d'une *nota* encodée en TEI.

Si cet encodage permet la mise en cohérence des différents composants du texte, enrichissant ainsi celui-ci sémantiquement, il permet également d'envisager la création automatique d'entrées d'index (biographiques, bibliographiques, de citations), toutes ces entrées étant interconnectées par des liens hypertextes de par leur construction même.

Les trois premiers chapitres de la troisième *Géorgique* ont ainsi été marqués selon les différents choix d'indexation TEI, choix réalisés avec les experts du projet travaillant sur l'analyse des commentaires de La Cerda. Le travail en cours consiste à mettre en place un démonstrateur permettant (1) de valoriser ce travail d'indexation, et (2) de naviguer de façon dynamique au sein du commentaire.

## 4 Discussion

Le choix du corpus se justifie par le fait qu'il représente un passage très important dans l'histoire de la littérature latine et dans l'œuvre de Virgile. En effet, Virgile y célèbre son triomphe en tant que poète de Rome (qui a su acclimater la poésie grecque en Italie et la faire sienne), et il y annonce métaphoriquement la construction d'une œuvre épique qui est sans doute l'*Énéide*.

Dans le travail de valorisation de cette œuvre, la traduction en français mentionnée section 3.2.1 est un point important. De façon générale, les spécialistes qui éditent des œuvres antiques ou des états de leur transmission (les éditions du xvi<sup>e</sup> siècle par exemple) ne les traduisent pas. Si nous prenons les exemples déjà abordés des *Bibliothèques Virtuelles Humanistes* ou du portail *Renaissance d'Horace*, il s'agit fondamentalement de rendre les textes accessibles, puis éventuellement d'en donner des éditions modernisées. Dans notre cas, nous tenons à traduire, en vertu de la conviction que, même si ce sont des textes faciles à comprendre en apparence, on n'en cerne précisément la logique ou l'intérêt qu'en les traduisant. La traduction fait ainsi partie d'une démarche qui vise à rendre l'Antiquité et le patrimoine plus accessibles ou moins « étrangers » ; mais elle a aussi une valeur scientifique et peut concerner elle-même les spécialistes dans la mesure où elle contribue à une meilleure compréhension et à une meilleure connaissance du texte édité. De fait, la question de la multiplicité des publics que vise le projet TALIE dans son ensemble reste une vraie question en ce qui concerne cette partie d'édition numérique : en entreprenant cette dernière, nous avons essayé de tenir compte non seulement de la « paire » grand public / spécialistes, mais aussi de différents types de spécialistes ; par exemple, en pratiquant au début un encodage qui restitue le plus possible les caractères physiques, typographiques du texte de La Cerda, nous avions en tête les historiens du livre ou professionnels des bibliothèques ; puis nous y avons renoncé en considérant qu'ils constitueraient sans doute un public minime et surtout, qu'il leur suffirait de considérer le fac-similé que nous avons décidé de joindre à notre édition. Ce que nous cherchons aujourd'hui, c'est fondamentalement à rendre cette œuvre lisible : peut-être n'est-ce pas son édition numérique qui la rendra attractive pour le « grand public » – et de fait, au sein de TALIE, nous prévoyons, en pensant spécialement à un public élargi, des expositions numériques qui pointent précisément des aspects remarquables des ouvrages que nous avons choisi de mettre en avant. Nous voudrions cependant que des lecteurs plus avertis puissent désormais en faire « usage » : il ne s'agit pas alors que de latinistes, il peut s'agir de personnes intéressées par l'histoire de la réception de Virgile ou par l'histoire des commentaires ou par l'enseignement jésuite au début du xvii<sup>e</sup> siècle en Espagne. Et l'on notera d'ailleurs que même les plus spécialistes doivent trouver une plus grande lisibilité, sinon la

lisibilité de l'œuvre tout court, dans une telle édition : s'il est vrai qu'une des richesses essentielles du commentaire de La Cerda réside dans son « hyper-intertextualité » (le fait que le commentateur, pour chaque passage de Virgile qu'il commente, fournisse une grande quantité de textes comparables, non seulement des modèles de Virgile mais des auteurs qui l'ont imité), les citations et surtout les références dans lesquelles elle réside nécessitent d'être traitées – elles sont elles-mêmes au cœur de la « rééditorialisation » – pour que le texte produise son effet sur le lecteur (même érudit, donc) d'aujourd'hui, qui, en tout état de cause, n'a pas la même mémoire des textes que La Cerda<sup>9</sup>. Rendre ce texte accessible implique bien ainsi de le rendre accessible pour tous, y compris les latinistes spécialistes de Virgile.

Dans cette première expérimentation, nous proposons un modèle de document pour la transmission d'un texte antique accompagné de ses commentaires savants, capable de rendre compte des choix éditoriaux érudits de la Renaissance tout en apportant les avantages notamment typographiques et hypertextuels de notre époque. Il s'agit maintenant de mettre ce modèle à l'épreuve d'un texte plus conséquent, plus ambitieux – à savoir l'œuvre maîtresse de Virgile, l'*Énéide* – afin d'identifier et de remédier à ses faiblesses ainsi que pour affiner et développer ses fonctionnalités, au meilleur bénéfice des usagers. Dans une même optique de valorisation et de confrontation du modèle à différentes réalités, nous souhaiterions associer notre édition de La Cerda à la dynamique d'autres projets de diffusion de commentaires de textes antiques commentés, comme par exemple le *Vergil Project*<sup>10</sup> qui cherche à associer au texte virgilien l'ensemble des commentaires anciens et modernes proposés par la tradition.

## 5 Conclusion et perspectives

Le projet TALIE vise à réaliser une édition moderne et numérique du commentaire de l'œuvre virgilienne de Juan Luis de La Cerda. Au-delà de l'intérêt scientifique et littéraire de ces contenus pour les férus de littérature latine, le document lui-même présente des particularités édi-

---

9. Nous voudrions ainsi que, pour chaque lemme commenté, ce lecteur puisse reconstituer la bibliothèque mobilisée par La Cerda et y accéder instantanément. Chaque référence que nous encodons doit pointer (par le biais d'une balise <target>) vers une URI permettant d'accéder au texte cité présenté dans son contexte : les liens ainsi créés doivent renvoyer vers les pages adéquates de *Perseus* ou du *Thesaurus Linguae Graecae (TLG)* pour les œuvres grecques, du *Packard Humanities Institute Classical Latin Texts (PHI)* ou de la *Latin Library* pour les œuvres latines. L'idée est de pouvoir donner accès à l'original des textes anciens cités dans l'édition la plus usuelle aujourd'hui. En ce qui concerne les œuvres plus tardives, du XVI<sup>e</sup> siècle en particulier, on pourrait renvoyer vers des sites qui en présentent une version numérisée ou une réédition numérique, ou bien donner accès à des PDF lorsqu'on en dispose.

10. <http://vergil.classics.upenn.edu/>.

toriales et de composition qui le rendent intéressant pour une édition numérique s'appuyant sur les modèles de description physique et sémantique récents. Dans cet article, nous avons présenté les premiers résultats de notre chaîne de traitements proposée pour la dématérialisation de documents anciens. Après avoir décrit les étapes d'encodage en mode image et d'ajout de la couche éditoriale technique par application d'un OCR, nous détaillons nos choix d'annotation sémantique en nous appuyant sur le langage d'indexation XML-TEI. Les trois premiers chapitres de la troisième *Géorgique* ont ainsi été marqués selon les différents choix d'indexation TEI réalisés en accord avec les experts philologues participant au projet.

Nous travaillons actuellement à la conception et au développement d'une interface pour la visualisation et l'analyse du commentaire. Cette interface, dynamique, est développée en langage Javascript et s'appuie sur l'ensemble de l'annotation XML-TEI réalisée en amont.

## 6 Bibliographie

- Bauman, S., Catapano, T. (1999). TEI and the Encoding of the Physical Structure of Books. *Computers and the Humanities*, vol. 33, num. 1, 113-127.
- Bureau, B., Nicolas C., Pinche, A. (2016). *Encoding (inter)textual insertions in Latin "grammatical commentary"*. Université Jean Moulin, Lyon. Disponible à : <https://hal.archives-ouvertes.fr/hal-01281266/>.
- Burnard, L. (2015). *Qu'est-ce que la Text Encoding Initiative?* OpenEdition Press, Marseille. Disponible à : <http://books.openedition.org/oep/1297> (généré le 08 juin 2016).
- Crozat, S. (2012). Chaînes éditoriales et rééditorialisation de contenus numériques. In *Le document numérique à l'heure du web de données*, Calderan, L. (Dir.), ADBS éditions, Paris.
- Démonet, M.-L. (2004). La Renaissance numérique : les Bibliothèques Virtuelles Humanistes. In *Digital technology and philological disciplines*, Bozzi, A., Cignoni, L., Lebrave, J.-L. (Dirs), Istituti editoriali e poligrafici internazionali, Pisa, Italie, 175-185. (Linguistica computazionale, 20-21).
- Foucault, M. (1969). *L'archéologie du savoir*. Paris, Gallimard (réimpr. 2008).
- Jacquemard, C., Gauvin, B., Lucas-Avenel, M.-C. (Éds) (2013). *Hortus sanitatis : Livre IV, Les poissons*. Presses universitaires de Caen, Caen. Disponible à : <https://www.unicaen.fr/puc/sources/depiscibus/accueil>.
- Laird, A. (2002). Juan Luis de la Cerda and the predicament of commentary. In *The classical commentary histories, practices, theory*, Gibson, R. K., Shuttleworth Kraus, C. (Dirs), Brill, Leiden.

Pierazzo, E. (2011). A rationale of digital documentary editions. *Literary and Linguistic Computing*, vol. 26, num. 4, 463-477.

Queyroux, F. (2003). EAD, la description archivistique encodée. In *La numérisation des textes et des images : techniques et réalisations*. CEGES, Villeneuve d'Ascq, 75-89. (Collection UL3 Travaux et recherches).

Rimelé, R. (2011). *HTML 5. Une référence pour le développeur web*. Eyrolles, Paris.

Stevens J., S. J. (1945). Un humaniste espagnol. Le Père Juan de la Cerda, S.J. (1558-1643), commentateur de Virgile. *Les Études Classiques*, vol. 13, 210-221.

Walsh, N. (2010). *DocBook 5 : The Definitive Guide*. O'Reilly Media, Beijing, Sebastopol.