

Les *Carmina* de Paulin de Nole et la Bible

Encodage en XML TEI et site internet

Alice Leflaëc

Juillet 2023

1 Genèse du projet

Le projet d'encodage en XML TEI des *Carmina* de Paulin de Nole et des réminiscences bibliques présentes dans les poèmes est né d'une discussion avec B. Bureau et I. Burri, les responsables « Humanités numériques » du projet [Poexégèse](#). Ce projet, dirigé par F. Stella, vise à collecter et encoder en XML TEI des textes de la poésie latine biblique de l'Antiquité tardive et du Moyen Âge. Les poèmes sont reliés à leur hypotexte biblique et aux commentaires bibliques et patristiques traitant de cet hypotexte. Si le site ne présente actuellement que des poètes médiévaux, le souhait des responsables du projet est de l'étendre à d'autres poètes. Comme nous avons effectué un relevé, plus complet que celui, déjà riche, de l'édition du *Corpus christianorum*, des réminiscences bibliques dans les *Carmina* de Paulin de Nole pour notre thèse, nous avons proposé de nous charger de l'encodage en XML TEI de ces poèmes.

L'objectif premier de cet encodage est de nourrir le site internet de Poexégèse. Toutefois, nous avons décidé de nous servir des documents encodés pour réaliser un site internet en vue d'un travail universitaire. Ce site vise à mettre en évidence les réminiscences bibliques présentes dans les *Carmina* en listant les références à côté des poèmes et en renvoyant au texte de la Vulgate pour les versets cités par Paulin de Nole.¹

2 Description du jeu de données

- Le texte latin des *Carmina*, sans l'apparat critique, et la numérotation des poèmes sont ceux de l'édition de Franz Dolveck au *Corpus christianorum* (Turnhout, Brepols, 2015). Nous avons complété le relevé des réminiscences bibliques effectué par F. Dolveck.
- Les versets bibliques sont cités dans la version de la Vulgate, tombée dans le domaine public, présente sur le site [BibleGateway](#).

Le corpus poétique de Paulin de Nole comporte deux ensembles :

- les *Carmina natalicia* : une série de quatorze poèmes de longueur variable, composés annuellement pour l'anniversaire de la mort de Félix de Nole, le saint auquel Paulin a décidé de consacrer sa vie.
- les *Carmina uaria* : la réunion de divers poèmes comme des lettres en vers, des paraphrases bibliques, un *proptempticon*, une consolation...

Pour l'instant, seuls les *Natalicia* ont été encodés en XML TEI. Les *carmina uaria* seront encodés ultérieurement et le site internet pourra être complété.

1. Nous sommes bien consciente que le site de Poexégèse permet de faire tout cela, mais 1. l'affichage est assez différent, 2. il fallait rendre un travail universitaire et 3. tous les passionnés de Paulin de Nole seront heureux d'avoir un site qui lui est consacré.

3 Choix d'encodage en XML TEI

3.1 Le document *Natalicia*

Comme l'encodage en XML TEI des *Carmina* de Paulin de Nole s'insère dans un projet existant, nous n'étions pas libre d'encoder à notre guise. Un manuel d'encodage a été rédigé pour le projet Poexégèse et nous avons respecté les consignes relatives à l'encodage des poèmes (cf. figures 1 et 2).

2. Préparer les textes des poètes

a. Présentation de l'encodage général

Les livres des poèmes ont été balisés dans des `<div1>` auquel on a ajouté un `@type` « Liber », un `@n` (= number) et un `@xml:id` (= un identifiant unique). Une balise `<p>` (= paragraphes) regroupe des balises `<seg>` qui servent, ici, à encoder les vers d'une même phrase. Elles possèdent un `@n` et un `@xml:id`. On y trouve des balises autofermantes `<lb/>` pour marquer les débuts de vers¹, qu'on complète avec un `@type` « versus » et un `@n`.

```
<div1 type="Liber" n="2" xml:id="Hild_2">
<p>
  <seg n="1" xml:id="Hild_2_1">
    <lb type="versus" n="1"/>Fert Amalechites quoniam Saul ipse peremit,
    <lb type="versus" n="2"/>Protinus in Siceleg David eum perimit.
  </seg>
  <seg n="2" xml:id="Hild_2_2">
    <lb type="versus" n="3"/>Ungitur a Iuda rex David versus in Ebron,
    <lb type="versus" n="4"/>Eius et imperio dux Ioab est equitum.
  </seg>
ETC...
```

Figure 5 Encodage général des poèmes

FIGURE 1 – Consignes Poexégèse : encodage général des poèmes

b. Écrire une référence

Ainsi, l'arborescence des textes étant déjà encodés, il reste à baliser les renvois aux passages bibliques cités. On part du principe qu'un segment renvoie au minimum à un passage biblique voire à plusieurs. Si un segment n'a aucun correspondant dans la Bible, le fait de ne pas réaliser l'encodage, décrit ci-dessous, suffit à faire comprendre qu'il s'agit d'un segment original sans source dans le texte biblique.

Les informations concernant les renvois doivent être balisées dans les `<seg>`. À l'intérieur de celles-ci et à leur commencement, on utilisera une balise `<listRef>`, qui contiendra une ou plusieurs balises `<ref/>` autofermantes. C'est ici qu'il faudra mentionner la/les référence(s) à l'aide d'un attribut `@cRef`.

FIGURE 2 – Consignes Poexégèse : encodage des références bibliques

Ce respect des consignes explique pourquoi nous avons encodé les réminiscences bibliques non pas vers par vers (ce que nous aurions préféré faire) mais par rapport à des segments (`<seg>`), chaque segment correspondant à une phrase. Il explique aussi pourquoi les vers sont encodés avec la balise auto-fermante `<lb/>` et non avec la balise traditionnelle `<l>`.²

2. Le choix de cette balise, comme l'explique bien le manuel d'encodage, est d'éviter de possibles chevauchements avec la balise `<seg>`, la fin d'un vers ne coïncidant pas toujours avec la fin d'une phrase.

En vue de l’affichage HTML des poèmes, nous avons fini par créer une copie du fichier .xml des *Natalicia* : cela nous a offert une plus grande souplesse dans l’encodage, tout en gardant un fichier répondant aux consignes de Poexégèse. Ce dernier document sera bientôt remis aux responsables du projet. Nous avons ajouté, dans le second fichier, une balise `` pour chaque groupe de vers à l’intérieur d’une balise `<seg>` afin de faciliter la réalisation d’un affichage en vis-à-vis du texte poétique et des références bibliques. Nous avons également placé chaque balise de fermeture `</seg>` à la fin d’un vers même si cela coupait une phrase pour éviter certains problèmes d’affichage en HTML.

3.2 Le document Réminiscences bibliques

L’encodage des versets bibliques présents dans les *Carmina* est très simple :

- La balise `<head>` pour le titre de la page et le titre des livres bibliques.
- Une balise `<div xml:id="abréviation_livre_biblique">` pour chaque livre biblique.
- La balise `` pour chaque verset cité. La référence du verset est écrite en toutes lettres à l’intérieur d’une balise `<title>`.

4 Chaîne de traitement

Voici les grandes étapes du travail :

1. **Récupération des textes** : Le texte de chaque *Natalicium* de Paulin de Nole a été récupéré sur le site de la [Library of Latin Texts](#) et copié, dans un premier temps, en format .txt dans un éditeur de texte. Le texte biblique de la Vulgate été récupéré sur le site [BibleGateway](#) et copié, au fur et à mesure de l’avancée du travail, dans le fichier XML de travail.
2. **Mise en forme général du document .xml « *Natalicia* »** : Un script Python a permis de convertir les fichiers .txt des *Natalicia* en un unique document .xml où le contenu texte de chaque fichier *Natalicium* a été inséré à l’intérieur d’une balise `<div>` dont l’attribut `@xml:id` correspondait au titre du fichier .txt.
Une série de feuilles de transformation XSLT a ensuite permis de mettre en forme, selon les consignes du projet Poexégèse, le document XML contenant le texte des *Natalicia*. Grâce à ces feuilles, les transformations suivantes ont pu être réalisées :
 - Division des `<div>` en `<seg>` correspondant à des phrases et ajout des attributs `@n="place_dans_la_div"` et `@xml:id="xml:id_de_div+numéro_de_la_div"` pour les balises `<seg>`.
 - Pour chaque vers, ajout d’une balise `<lb/>` avec un attribut `@type="versus"` et un attribut `@n="place_dans_la_div"`.
 - Ajout d’une balise `<p>` englobant le contenu de chaque `<div>`.
 - Remise des `<div>` dans l’ordre, car la série des *Natalicia* était dans le désordre.
 - Correction de quelques éléments de ponctuation qui avaient dû être enlevés ou avaient disparu au cours des transformations. Certains éléments de ponctuation ont aussi été corrigés manuellement dans un éditeur de texte.
3. **Encodage manuel en XML TEI** : Les références bibliques dans le fichier contenant les *Natalicia* ont été encodées manuellement. Le fichier contenant les réminiscences bibliques a également été encodé manuellement.
4. **Réalisation d’un ODD** : Un ODD, commun aux deux documents XML a été réalisé avec l’outil [ROMA](#) (1^{ère} version).

5. **Feuilles de transformation XSLT pour la conversion en HTML :** À partir de cette étape, nous avons travaillé, comme nous l'avons dit précédemment, sur une copie du fichier contenant le texte des *Natalicia* afin de pouvoir modifier certaines balises au besoin. Deux feuilles de transformation XSLT, l'une pour les *Natalicia*, l'autre pour les réminiscences bibliques, ont permis de convertir les balises XML en balises HTML.
Remarque sur la numérotation des vers : En raison de quelques exceptions dans la numérotation de certains poèmes (cf. *Natalicia* 4, 7, 13) qui rendaient une automatisation du comptage de 5 en 5 plus difficile, nous avons fait le choix d'afficher la numérotation de chaque vers. Les particularités dans la numérotation sont signalées par un point d'exclamation.
6. **Rédaction d'une page index.html :** Une page d'accueil pour le site a été conçue dans un document HTML.
7. **Rédaction d'une feuille de style CSS :** l'aspect de chaque balise HTML a été défini dans une feuille de style commune à tous les documents.
8. **Création de pages .html :** En vue de la création du site internet, nous avons enregistré et renommé les outputs.html des feuilles de transformation XSLT. Cela nous a permis d'y insérer facilement la barre de navigation du site avec son menu déroulant, le pied de page et le bouton permettant de remonter rapidement en haut de la page.
9. **Dépôt du dossier sur GitHub et hébergement du site par GitHub pages :** Pour finir, nous avons hébergé le site internet avec GitHub pages.
Tout le dossier avec les fichiers XML, l'ODD, les feuilles de transformation XSLT pour la conversion en HTML, les pages HTML et le CSS a été déposé dans un répertoire sur GitHub.