



**HAL**  
open science

## TEI-encoding of text reuses in the BIBLINDEX Project

Elysabeth Hue-Gay, Laurence Mellerin, Emmanuelle Morlock

► **To cite this version:**

Elysabeth Hue-Gay, Laurence Mellerin, Emmanuelle Morlock. TEI-encoding of text reuses in the BIBLINDEX Project. *Journal of Data Mining and Digital Humanities*, 2017, Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages, 10.46298/jdmdh.3989 . halshs-01543050v2

**HAL Id: halshs-01543050**

**<https://shs.hal.science/halshs-01543050v2>**

Submitted on 7 Oct 2017

**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.

## TEI-encoding of text reuses in the BIBLINDEX Project

Elysabeth Hue-Gay<sup>1</sup>, Laurence Mellerin<sup>2\*</sup>, Emmanuelle Morlock<sup>3</sup>

CNRS/Université de Lyon, UMR 5189 Histoire et Sources des Mondes Antiques, France

\*Corresponding author: [laurence.mellerin@mom.fr](mailto:laurence.mellerin@mom.fr)

### Abstract

This paper discusses markup strategies for the identification and description of text reuses in a corpus of patristic texts related to the BIBLINDEX Project, an online index of biblical references in Early Christian Literature and the Middle Ages. In addition to the development of a database that can be queried by canonical biblical or patristic references, a sample corpus of patristic texts has been encoded following the guidelines of the TEI (Text Encoding Initiative), in order to provide direct access to quoted and quoting text passages to the users of the <https://www.biblindex.info> platform.

### Keywords

TEI, Text Encoding Initiative, text markup, Bible, Greek, Latin, Septuagint, Vulgata, patristics, BIBLINDEX, Biblia Patristica, quotations, text reuses, Sources Chrétiennes, Bernard of Clairvaux

### INTRODUCTION

BIBLINDEX [[http1](#)] is a project led by the French Institut des Sources Chrétiennes [[http2](#)], part of the HiSoMA research center in Lyon, with the support of the LIRIS (Laboratoire d'Informatique en Image et Systèmes d'Information) and the LIG (Laboratoire d'Informatique de Grenoble), two computer science labs in Rhône-Alpes region (France). It aims to create an exhaustive online index of biblical quotations and allusions in Early Christian Literature, both Western and Eastern texts, which should eventually cover the whole of Late Antiquity and the early Middle Ages. After introducing the BIBLINDEX Project, this paper describes a new initiative to expand the data available by the direct encoding of biblical text reuse in patristic texts using a TEI encoding strategy that can account for the concepts and categories theoretically defined by the BIBLINDEX Project council. We present the way our TEI encoding choices were defined in order to give a closer account of the text reuses phenomenon, bearing in mind that these choices have to be compliant with the data already stored in the database and with the printed edition prepared at the same time. Firstly, the issue of text reuse delimitation will be addressed – that is to say both the delimitation of the biblical text segment to which the text reuse refers and the delimitation of the text segment considered as a reuse in the patristic work. Secondly, the issue of their characterization will be studied. In doing this, we extend a methodological reflection from [Mellerin, 2013]<sup>1</sup>.

---

<sup>1</sup> We are very thankful to Catherine J. Smith (Department of Theology and Religion, Birmingham) for her careful revision. All remaining shortcomings are, as usual, with the authors.

## 1. Description

BIBINDEX currently offers a comprehensive inventory of quotations and allusions providing bibliographical information. It includes a broad sweep of references ranging from verbatim quotations to loose allusions. Each record consists of a series of numbers indicating the chapter and verse of the biblical text, its location in the patristic work and the corresponding page and line numbers in the reference edition. The web interface lists the authors included in the database in a drop-down menu. The search facilities return all the entries for a biblical verse or text passage, which can be further restricted by date, geographical region and author or work. Bibliographical details for the edition as well as the page and line number of the reference are provided for each online entry and can be used to filter the maximal list provided in the search results.

## 2. The available corpus

The project is based on the resumption of *Biblia Patristica*, a collection of printed indexes of biblical citations in the writings of Greek and Latin Church Fathers from the first centuries, published by a CNRS team, the CADP (Centre for Patristics Analysis and Documentation) in Strasbourg, between 1965 and 2000 [Allenbach, 1967] [Junod, 1995]. Sources Chrétiennes has been entrusted with the CADP database and further unpublished data.

The seven published volumes of [*Biblia Patristica*, 1975-2000], along with a supplementary volume of biblical references in Philo of Alexandria, cover the first three centuries, along with part of the fourth:

- volume 1, which appeared in 1975, covers the beginnings of extracanonical Christian literature up to Clement of Alexandria and Tertullian;
- volume 2 indexes the literature of the third century, apart from Origen,
- to whom volume 3 is exclusively devoted;
- coverage of the fourth century begins with volume 4, which includes the writings of Eusebius of Caesarea, Cyril of Jerusalem, and Epiphanius of Salamis;
- volume 5 covers Basil of Caesarea, Gregory of Nazianzus, Gregory of Nyssa, and Amphilochius of Iconium;
- volume 6 turns to the Latin writers Hilary of Poitiers, Ambrose of Milan, and Ambrosiaster;
- volume 7 on Didymus the Blind appeared in 2000.

The 270,000 published citation references have now been made freely available online as BIBINDEX. The searchable database and website also give access to around 100,000 further unpublished entries for authors including Athanasius, John Chrysostom, Theodoret of Cyr, Procopius of Gaza, Jerome and works marked as *Spuria* and *dubia* related to the authors of the published volumes. The data has been verified exactly in the same way as for the published volumes, only the last proofreading step is still missing.

In addition to this material, about 550,000 references prepared for the CADP by its usual external contributors – volunteer European researchers, experts in different patristic fields with extensive knowledge of the Bible (among them: J. Fleury, G. Valayer, B. Outtier) –, have been digitized and normalized, i.e. converted in a format consistent with the already available data, and are now ready to be put online. This data is “unverified”, that is to say it didn’t benefit from the thorough scientific and technical verification usually carried out by the permanent CADP staff, but all the references have been established by reliable experts. These citations are taken from 3,000 diverse texts, written in different places and at different times. Among them are *Catena*, hagiographical and liturgical texts, works of Ps.-Chrysostomus, Ephrem the Syrian, Cyril of Alexandria, Gregor the Great, Maximus the Confessor. The

BIBINDEX team is now focusing on the systematic treatment of the remaining works from the fourth and fifth centuries, especially those of Augustine, in order to ensure the exhaustive coverage of this period [Mellerin, 2017-1]. Some statistical studies have already been carried out on the corpus currently available [Bady, 2015]; [Mellerin, 2014; 2016; 2017-2; 2017-3]. Moreover, a prototype geovisualization interface [http3] displaying temporal and spatial data about authors and works has already been developed by the LIG.

### 3. Methodological issues

Work continues to augment the BIBINDEX database with citation references in the format described above (verse number – location in the work). This data has its value: the ability to search by verse reference makes it easy to access synthetic information on the various uses of a biblical passage; and the more references that are available in the database the more illuminating statistical approaches will become. Besides, these references in the form of numbers are, and will always be, freely available online, as they don't depend on any copyright protection issue.

However, the references contained in the database are, of course, the result of careful analysis of the patristic and biblical texts concerned. Unfortunately all trace of this analytical step for the inherited data has typically been lost. The precise delimitation of each text reuse is no longer available: the information consists solely of a line number which only gives a clue to the beginning of the text reuse. The annotations made by the analysts and thus the motivations behind their choices have also been lost. Paradoxically, these reference statements, conceived by the CADP as a research outcome, constitute one of the starting points for our research today: the references listed in the database have to be re-injected in the patristic texts in order to be verified and refined.

When it comes to adding new data to the database, the analysis will start from the text with TEI-compliant markup stage supported by an automatic tool specifically designed for the recognition of text reuse in BIBINDEX<sup>2</sup> [Gesche-Egyed-Zsigmond-Calabretto, 2017]. This time we intend to retain data from all the stages leading to the statement of a text reuse: identification, close delimitation, characterization.

To allow access to the corpus by canonical biblical reference, patristic author's name or work title on the one hand and access to the text of a reuse in its context on the other, the project architecture has two main components: a MySQL<sup>3</sup> database and a collection of texts encoded in TEI-XML. This dichotomy reflects the history of the corpus construction; we are now considering the option of a native XML database.

All the metadata about the authors (dates, biography, additional documents, bibliography, etc.) and about the works (date of composition, topics, summary, etc.) are stored in the MySQL database. In order to make them easily available for a common use as linked Open Data, links to the BnF (Bibliothèque nationale de France) authorities are foreseen where they

---

<sup>2</sup> This is not the place in this paper devoted to the TEI-encoding to discuss in detail the exact role of this tool, especially since it is still under construction: its basic function would be to make easier the human search for biblical quotations by pre-tagging the obvious literal quotations and providing suggestions for the more paraphrastic or allusive ones. At the time being, given that patristic quotations are most of the time integrated in the course of speech, often rely on different transmissions of the biblical texts, and are thus deeply modified compared to the reference Bible, automatic tools can only be adjuvants, but they are far from making the manual work useless. For an illustration of the encountered difficulties on this specific corpus, and some insight into the promising researches in this field, see [Moritz et al., 2016].

<sup>3</sup> MySQL is an open-source relational database management system.

exist<sup>4</sup> – but currently this only concerns about a third of the works in our corpus<sup>5</sup>. Each text reuse corresponds to a record in an association table between a Biblical passage and a passage from a patristic text. The data is entered using a list of abbreviations; the start and end line numbers can be stated.

Moreover, to allow the use of several biblical referencing systems beyond the Bible of Jerusalem – as has been the case until now – biblical texts written in different ancient languages (Hebrew, Greek, Latin, Syriac), as well as modern translations in English and French, were added to the database and closely aligned to one another. This additional biblical data was segmented at the verse level and sometimes at the part of verse level. As a result of this work, a multilingual concordance between Bibles was added to the BIBLINDEX platform and is freely searchable online [http5]. The biblical text itself is now also stored in the MySQL database, which makes it easy to display the verses sought by the user.

The patristic texts are stored as TEI-XML files, which won't be available online in their exhaustive form, because of complex copyright issues, but freely searchable<sup>6</sup>. At present there are only a few test samples, namely three complete works written by Bernard of Clairvaux<sup>7</sup> [Bernard of Clairvaux, 1990, 2009, 2010]. Each work is stored in a separate file. The TEI Header of each file contains the unique identifier of the work, which is also recorded in the database. This link enables all the relevant metadata for a work to be extracted from the database and automatically re-injected into the relevant XML file.

All information regarding the basic structure of the texts are embedded in the XML files:

- the logical structure specific to a work (book, chapter, section, paragraph numbers,...) is marked up with the corresponding TEI tags;
- the physical structure specific to the reference edition (page and line numbers) is expressed with milestones elements (empty tags marking a boundary point).

#### 4. Encoding scheme

Our encoding scheme follows the TEI guidelines. Each patristic work is encoded in a separate file which contains the following key details. The TEI Header title statement <fileDesc/titleStmnt> is used to encode the full and abbreviated title of the work and the name

---

<sup>4</sup> BIBLINDEX also aims at being part of a shared project among the patristic community: building a common and freely available referencing system for authors and works, based on the model provided by the *Clavis Patrum*. As the BnF, in the context of the Bibliissima project [http4], has already begun to classify ancient authors and makes this data available online with persistent IDs, it seems more useful to help improving this existing database than to build an independent one.

<sup>5</sup> A large part of the works we have to deal with are fragments, published in reviews as single articles, and/or which authors are unknown; for the time being, the BnF mainly identifies entire works, published in books, which authorship is known with a good certainty level.

Several authors may be associated to a work, according to the certainty of the attribution; the works have different kind of relationship with each other. The structure of the database makes it possible to create workgroups, to specifically manage the collections of fragments and the works that quote each other. Each work is linked to one or more reference editions, classified according to their date and relevance. The metadata on the 823 authors and 7,585 works [http6] gives some idea of the complexity of the data and their relations.

<sup>6</sup> The internet user will be allowed to see about 5 lines of each text at a time, but texts won't be available for download as full files. As the original print editions of the patristic texts are copyrighted material, it is not possible for us to build from the outset a full open-access architecture comparable to the Leipzig Open Fragmentary Text Series (LOFTS) project's one [http7] that provides open editions of ancient texts linked to born-digital editions of fragmentary works. Nevertheless, our beginning collaboration with the TXM research team [http8] aims to follow such a model on a delimited test corpus.

<sup>7</sup> Bernard of Clairvaux is indeed often referred to as the last Church Father. His work has been chosen as a test case, despite his late period of life compared to most other Fathers, because he "speaks Bible": the about 35,000 biblical reuses found in his work are really embedded in the course of his speech and provide a wealth of information to describe non-literal quotations. Besides, a great part of the biblical analysis on this specific corpus has already been carried out at Sources Chrétiennes.

of the author (including a link to BIBINDEX page for the author). The <body> part consists of a single element <div>. Each <div type="work"> element can be divided in chapters, <div type="chapter"> elements. The chapters contain a mandatory title and one or more paragraphs.

## I. BIBLICAL TARGETS: DELIMITATION AND CHARACTERIZATION

### 1.1 Delimitation of the biblical targets

Biblical targets fall into one of two categories, depending on whether or not they are linked to precise verse numbers.

The first category, described in [Mellerin, 2013:21-24], gathers vague allusions to larger biblical subcorpora such as Johannine writings or healing narratives, which cannot easily be characterized by verse numbers, as well as common biblical formulations, names of characters (e.g. the disciples of Christ, David) and global reminiscences of heterogeneous episodes (e.g. the life of the Hebrew people in the desert). The database contains several records of such subcorpora and named entities in dedicated tables, which will be expanded in the course of our work. In the TEI files, pointers are built from database keys: instead of targeting a canonical reference, they use the identifiers of these records, and/or the PersNames, PlaceNames tags.

---

[Irenaeus, *Adversus Haereses*:II, 24, 4 (244, l. 145)]:

<rs type="bookgroup" key="PentateuchID">In quinque libris legem populo <persName key="MosesID">Moyses</persName> tradidit</rs>

---

It is the second, predominantly major category, of text reuse that we will focus on in this paper. This category comprises of the citations that can be linked to canonical references. At present, any reference to the biblical text is made to a whole verse or set of verses, defined by the name of the Bible edition used as the versification reference, the biblical book, the chapter number(s) and the verse number(s)<sup>8</sup>. This universal referencing system is convenient, insofar as it allows access to the data in BIBINDEX using the verse numbering, but it cannot express the specific details when only a few words of a verse are actually targeted. This system therefore has to be refined when applied to text reuse detected in the XML files: it has to be capable of identifying text reuse at the word level.

We considered targeting only verse numbers and expressing the precise words concerned using the <app> and <rdg> tags, and following a critical apparatus parallel segmentation method. In this system, the quoting Church Father's text would be treated as a witness to the biblical text. This solution has the benefit of exposing the variants clearly in a portion of a verse; but it also requires the insertion of the biblical text to the patristic text file and makes difficult to take word order into account. It therefore seems preferable to create pointers that get down to the word level, expressed in the biblical URI itself.

The canonical expression of biblical references has been designed as such:

- in case of complete verse reuse: Vg:Mc:12:8: *et adprehendentes eum occiderunt et eiecerunt extra uineam;*
- in case of incomplete verse reuse: Vg:Mc:12:8:6-8: *eiecerunt extra uineam.*

---

<sup>8</sup> Here is not the place to address again the issue of the necessarily imperfect choice of biblical referentials taken from modern editions. See [Mellerin, 2013:15-16].

Each URI constructed targets a verse in the Bible parser of the BIBLINDEX website, allowing the visualization of the text in its biblical context. The second example above is represented in the XML as follows<sup>9</sup>:

```
<seg type="bRef" xml:id="T01">
  <bibl type="biblical">
    <ref cRef="Vg:Mc:12:8:6-8">Mc 12, 8</ref>
    <ptr target="http://www.biblindex.info/en/biblical/content/ref/Vg_Mc_12_8:6-8" targetLang="lat"/>
  </bibl>
</seg>
```

Clicking the link expressed in the target opens the relevant page in the Bible parser, represented below in Figure 1.

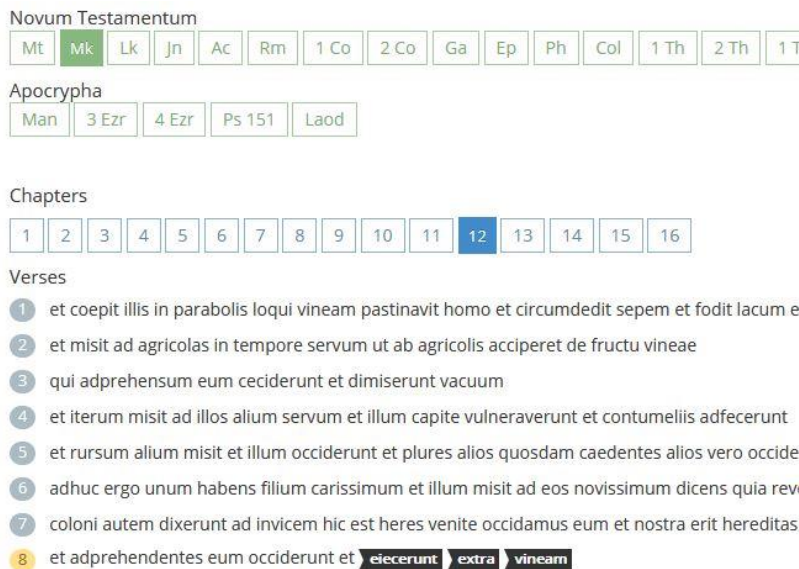


Figure 1: Screenshot of [http://www.biblindex.info/en/biblical/content/ref/Vg\\_Mc\\_12\\_8:6-8](http://www.biblindex.info/en/biblical/content/ref/Vg_Mc_12_8:6-8)

The detailed expression of these canonical references for more complex cases is being developed using semantically meaningful identifiers compliant with the “Canonical Text Service” (CTS)<sup>10</sup>: this allows a future reformulation of our URIs with CTS.

This precise word-level targeting allows to refer a single text reuse to several biblical sources.

Another additional advantage means that a passage of a text written in one language may be related to a Bible written in another, which is not its usual reference Bible.

<sup>9</sup> The examples that follow do not use the full encoding system finally chosen, but present a simplified version, highlighting only the precise aspects the example is illustrating. The complete encoding model is presented in the third part of this paper.

<sup>10</sup> CTS [http9] is a component of the CITE citation architecture designed by the Homer Multitext Project. The CTS URNs (Uniform Resource Names) follow a naming pattern that has the capability to represent with precision the cited entity. Hence the CTS scheme allows to cite at the text level or at the level of a passage of text. The version (or instance) of the text or the level of the generic notional work that the text represents may also be expressed with this citation scheme. As it follows a hierarchical structure that is aligned to the BIBLINDEX database structure, it will be easy to retrieve metadata information of text passages from CTS canonical reference in the BIBLINDEX platform.

---

[Jerome, *Epistula 36 ad Damasum* 12 (59, l. 17)], whose text is mostly related to the Vulgata, quotes sometimes the Septuagint.

```
<seg type="scripturalQ" xml:id="B01">
  <seg type="insertion">ubi Septuaginta posuerunt</seg>
  quinta autem generatione ascenderunt filii Israhel de terra Aegypti</seg>
<note type="scripturalNote" subtype="translation_latin"11>
  <seg type="bRef">
    <bibl type="biblical">
      <ref cRef="LXX:Ex:13:18:15-26"/>
      <ptr target="http://www.bibliindex.info/fr/biblical/content/ref/LXX_Ex_13_18:15-26"/>
    </bibl>
  </seg>
</note>
```

---

In addition to the reference Bibles stored in the BIBLINDEX database, “Bible” entities under construction can be targeted. In Greek for example, the translation made from the Hebrew by Aquila (2<sup>nd</sup> c. AD) is assumed as a work on the whole Bible, but has to be reconstructed from fragments. If reuse of Aquila’s text is identified, it will be related to a virtual target, prefixed AQ instead of LXX, with a verse number equivalent to the LXX verse number. The entity “Aquila’s version” will be a posteriori gradually reconstructed by compiling the discovered instances of text reuse. Similarly, in the Latin tradition, each time an author quotes the *Vetus Latina* and not a text form similar to the Vulgate’s text form, a record in the entity “Vetus Latina” will be created.

Table 1 lists further examples of (partial) biblical versions we have already identified in our corpora which could be (partially) reconstructed using our text reuse discoveries.

---

<b>Aq.</b>	Aquila
<b>Symm.</b>	Symmaque
<b>Theod.</b>	Theodotion
<b>Quinta</b>	<i>Quinta</i>
<b>Sexta</b>	<i>Sexta</i>
<b>V2</b>	“The Others”
<b>Sin.</b>	text of Codex <i>Sinaiticus</i>
<b>V3</b>	text B in Rahlfs’ edition
<b>Heb.</b>	<i>Psalmi iuxta Hebraeos</i>
<b>Diat.</b>	<i>Diatessaron</i>
<b>VL</b>	<i>Vetus Latina</i> text of a specific Father

---

Table 1. Other biblical versions.

## 1.2 Description of the textual reality expressed by the canonical reference

This information regarding biblical verses and words described in the previous section will be supplemented by specifying the nature of the relationship between the text reuse and the target.

Abbreviations, such as the sign  $\neq$  used in the printed volumes of the Sources Chrétiennes series, may be used to indicate morpho-syntactic and/or lexical modifications which have been made in the textual reality expressed by a canonical reference, with respect to the reference edition chosen. These modifications are usually caused by the introduction of the reuse in the thread of the discourse, examples of this include changes in the word order, recasting in direct or indirect form, addition of particles or conjunctions, changes in tense,

---

<sup>11</sup> See infra, 1.2 (p.7), the definition of this subtype.



mood or person, and so on. A fine-grained TEI markup could be used to refine the description of these “morpho-syntactic modifications”, ad locum. But such an approach would be a considerable burden in the encoding work and would not be useful for the further data processing. Instead these differences will be directly visualized by comparing texts in the user interfaces of BIBINDEX. This solution is favored by the project as, rather than data mining, its primary aim is to provide human users with an easy way of seeing text reuses both in their original and their reuse contexts. To indicate a text reuse that exhibits changes from the original, the attribute value “inexactQuotation” is used.

---

[Bernard of Clairvaux, *Liber ad milites Templi de novae militiae*, Prologus:48, l. 1]

```
... Hugoni, militi Christi et magistro militiae Christi, Bernardus Claraevallis solo nomine abbas:
<seg type="scripturalQ" xml:id="lat.sQ.Tpl.Pr.a">bonum certamen certare</seg>
<note type="scripturalNote" type="bRef">
  <seg xml:id="lat.b.Tpl.Pr.a" >
    <bibl type="biblical">
      <ref cRef="Vg:2_Tm:4:7:1-3"/>
      <ptr target="http://www.bibindex.info/en/biblical/content/ref/Vg_2%20Tm_4_7:1-3"/>
    </bibl>
  </seg>
  <link target="#lat.sOcc.Tpl.Pr.a12 #lat.b.Tpl.Pr.a" type="inexactQuotation"/>
</note>
```

---

An author may also quote a translation or a transliteration of the Bible taken as a reference, which led to the creation of further subtypes:

---

[Jerome, *Epistula XXXIV ad Marcellam* 2:45, l. 3]

```
<seg type="scripturalQ" subtype="normal" xml:id="lat.sQ.Ep.34.2.a">
  <seg type="insertion" ana="#explicitScripturalAttribution">Vnde ad Hebraeum recurrens inueni pro</seg>
pane doloris</seg>
<note type="scripturalNote" xml:id="lat.sNote.Ep.34.2.a">
  <seg type="bRef" xml:id="lat.b.Ep.34.2.1">
    <bibl type="biblical">
      <ref cRef="Vg:Ps:126:2:12-13"/>
      <ptr target="http://www.bibindex.info/en/biblical/content/ref/Vg_Ps_126_2:12-13"/>
    </bibl>
  </seg>
  <link target="#lat.sOcc.Ep.34.2.a #lat.b.Ep.34.2.113" type="inexactQuotation" subtype="explicit"/>
</note>

<seg type="scripturalQ" subtype="transliteration_hebrew" xml:id="lat.sQ.Ep.34.2.b">
  <seg type="insertion" ana="#genericScripturalAttribution">scriptum esse</seg>
leem aasabim</seg>
<note type="scripturalNote" xml:id="lat.sNote.Ep.34.2.b">
  <seg type="bRef" xml:id="lat.b.Ep.34.2.2">
    <bibl type="biblical">
      <ref cRef="BHS:Ps:127:2:2-3"/>
      <ptr target="http://www.bibindex.info/en/biblical/content/ref/BHS_Ps_127_2:2-3"/>
    </bibl>
  </seg>
  <link target="#lat.sOcc.Ep.34.2.b #lat.b.Ep.34.2.2" type="inexactQuotation" subtype="explicit"/>
</note>

<seg type="scripturalQ" subtype="normal" xml:id="lat.sQ.Ep.34.2.c">
  <seg type="insertion" ana="#explicitScripturalAttribution">quod Aquila interpretatus est</seg>
ἄρτον τῶν διαπονημάτων</seg>
<note type="scripturalNote" xml:id="lat.sNote.Ep.34.2.c">
  <seg type="bRef" xml:id="lat.b.Ep.34.2.3">
    <bibl type="biblical">
      <ref cRef="AQ:Ps:126:2:13-15"/>
      <ptr target="http://www.bibindex.info/en/biblical/content/ref/...14n/>
    </bibl>
  </seg>
```

---

<sup>12</sup> ID of the quoting passage in Bernard’s work.

<sup>13</sup> Simplified encoding.

<sup>14</sup> Not available yet.

```

</seg>
<link target="#lat.sOcc.Ep.34.2.c #lat.b.Ep.34.2.3" type="inexactQuotation" subtype="explicit"/>
</note>
<seg type="scripturalQ" subtype="translation_latin" xml:id="lat.sQ.Ep.34.2.d">
  <seg type="insertion">id est</seg>
  panem elaborationum</seg>
<note type="scripturalNote" xml:id="lat.sNote.Ep.34.2.d">>
  <seg type="bRef" xml:id="lat.b.Ep.34.2.4">
    <bibl type="biblical">
      <ref cRef="AQ:Ps:126:2:13-15"/>
      <ptr target="http://www.bibindex.info/en/biblical/content/ref/..."/>
    </bibl>
  </seg>
  <link target="#lat.sOcc.Ep.34.2.d #lat.b.Ep.34.2.4" type="inexactQuotation" subtype="explicit"/>
</note>

```

---

Finally, each time several canonical references are related to the same text reuse, either a probability coefficient must be specified<sup>15</sup>, or the canonical references must be ordered. This question especially arises in cases of internal biblical reuses (quotations from the Old Testament in the New, synoptic versions of the Gospels, etc. [Mellerin, 2013:28-30]); but it also applies to phrases common to several verses.

---

[Bernard of Clairvaux, *Homilia I in Laudibus Matris* 3:112, l. 7]

```

<seg type="scripturalQ" xml:id="lat.sQ.Miss.H1.3.c">
  <seg type="insertion" xml:id="lat.sI.Miss.H1.3.c.1">scriptum est : </seg>
  Nisi Dominus Sabaoth reliquisset nobis semen, sicut Sodoma fuisset et sicut Gomorraha similes essemus</seg>
<note type="scripturalNote" xml:id="lat.sNote.Miss.H1.3.c">
  <seg type="bRef" xml:id="b.Miss.H1.3.3">
    <bibl type="biblical">
      <ref cRef="Vg:Rm:9:29:5-19"/>
      <ptr target="http://www.bibindex.info/fr/biblical/content/ref/Vg_Rm_9_29:5-19"/>
    </bibl>
  </seg>
  <link target="#lat.sOcc.Miss.H1.3.c #b.Miss.H1.3.3" type="inexactQuotation" subtype="explicit" ana="#priority_rank1"/>
  <seg type="bRef" xml:id="b.Miss.H1.3.4">
    <bibl type="biblical">
      <ref cRef="Vg:Is:1:9">Is 1, 9</ref>
      <ptr target="http://www.bibindex.info/fr/biblical/content/ref/Vg_Is_1_9"/>
    </bibl>
  </seg>
  <link target="#lat.sOcc.Miss.H1.3.c #b.Miss.H1.3.4" type="inexactQuotation" subtype="explicit" ana="#priority_rank2"/>
</note>

```

---

## II. TEXT REUSE IN THE PATRISTIC TEXT: DELIMITATION AND CHARACTERIZATION

### 2.1 Delimitation of the patristic text reuse

“The text reuse object considered is a **text unit defined as a part of a patristic text in which an author refers, explicitly or implicitly, through quotation, mention or reminiscence, to one or more parts of the biblical text.** Multiple biblical references may be found in a single unit. In each text unit, the author’s thought and the biblical contents are nested together in such a way that they are often impossible to separate. This is why a unit is not constituted by the exact words of the Bible which also appear in the Fathers’ text, but is defined as the coincidence between a patristic sense unit and a biblical one. There may be no common words between the text unit and the biblical referential, for example in the case of a complete paraphrase of the biblical text. A single word, if relevant, can be a biblical reference and a unit.” [Allenbach, 1975:20-21].

---

<sup>15</sup> A three-level coefficient has been planned: high, medium or low probability that this specific text has been selected in the Father’s memory rather than this other. As the criteria are often very tenuous, we can’t think of something more precise.

### 2.1.1 Delimitation of the quotation

We first had to decide whether empty elements as boundary delimiters (milestones) or container elements (bloc or segment elements) should be used to indicate the beginning and end of each text reuse. The use of milestones and generic anchors avoids problems of overlapping hierarchies in the encoding; but non-empty elements are more suited to describing precisely a segment of text with help of attributes and types.

The following examples, taken from [Bernard of Clairvaux, *De Diligendo Deo*], show what the encoding would look like using each of these two methods.

The first example highlights a case of overlapping reuses.

---

[Bernard of Clairvaux, *De diligendo Deo* 27:130, l. 24-26]

*Corpus mortis* is a literal quotation of Rm 7:24:8-9; *corpus mortis aggravat* is a modified quotation of Sp 9:15:1.3-5 (*mortis* is considered a paraphrase of *quod corrumpitur* read in Sp 9:15).

#### ANCHORS

perturbat diei malitia

```
<anchor xml:id="anchor1"/>corpus mortis<anchor xml:id="anchor2"/>aggravat<anchor xml:id="anchor3"/>sollicitas carnis necessitas defectus corruptionis non sustinet.
```

```
<span from="#anchor1" to="#anchor2" source16="#Vg:Rm7:24:8-9">biblical_occurrence</span>
```

```
<span from="#anchor1" to="#anchor3" source="#Vg:Sp:9:15:1.3-5">biblical_occurrence </span>
```

#### SEG TAGS

perturbat diei malitia

```
<seg type="scripturalQ" xml:id="seg1" source="#Vg:Rm7:24:8-9">corpus mortis</seg>
```

```
<seg type="scripturalQ" xml:id="seg1a" source="#Vg:Sp:9:15:5">aggravat</seg>
```

sollicitas carnis necessitas defectus corruptionis non sustinet

```
<join result="seg" target="#seg1 #seg1a" xml:id="seg2" source="#Vg:Sp:9:15:1.3-5">
```

---

The encoding with anchors is simple and intuitive, it sticks to the actual delimitation of the text reuse; but in order to characterize this reuse it is still necessary to create segments, using the anchors' identifiers, in the form of `<span>` tags which are independent of the text structure.

The encoding with non-empty elements artificially creates a segment (seg1a) which is only a part of the text reuse. The second text reuse should still be expressed as formed by the addition of segments. The code is not as "clean" as when milestones are used, but the amount of work to mark up the text is similar.

The second example presents a case of a quotation interrupted by an insertion.

---

[Bernard of Clairvaux, *De diligendo Deo* 31:138, l. 6-7]

#### ANCHORS

```
<anchor xml:id="anchor1"/>Comedite<anchor xml:id="anchor2"/>inquit<anchor xml:id="anchor3"/>amici et bibite et inebriamini carissimi<anchor xml:id="anchor4"/>
```

```
<span from="#anchor1" to="#anchor2" xml:id="seg1a"/>
```

```
<span from="#anchor3" to="#anchor4" xml:id="seg1b"/>
```

```
<join target="#seg1a #seg1b" result="seg" type="biblical_occurrence" xml:id="seg1" source="#Vg:Ct:5:36-41"/>
```

#### SEG TAGS

```
<seg type="scripturalQ" source="#Vg:Ct:5:36-41">Comedite
```

```
<seg type="insertion">inquit</seg>
```

```
amici et bibite et inebriamini carissimi</seg>
```

---

This time, the encoding using anchors is cumbersome; the encoding using `<seg>` elements is faster and more intuitive. As it is clearly better for simple cases and interrupted text reuses,

---

<sup>16</sup> The 'source' attribute is not used in the real encoding, but as it is not the point of the following examples, a simplified markup is shown.

and as complex as the encoding with milestones in cases of overlaps, the encoding with <seg> tags has been chosen and applied to the whole corpus.

The semantically neutral <seg> tag was retained rather than the quote-specific tags proposed by the TEI, <q>, <cit> and <quote>: <q>, as a graphical marker, is not relevant to the ancient texts in general; <cit> assumes that the source is explicitly mentioned, which is rarely the case in our corpus; <quote> assumes the explicit intention to quote, which is also irrelevant in most cases. Moreover, we have left aside the tags specifying the status of the enunciation, <said>, <mentioned>, <soCalled>, because these topics find their place in the commentary notes.

In its simplest case, the reuse is composed exclusively of words quoted exactly, from a single biblical passage and without any discursive interruption. However, such text reuse occurs rarely in our corpus, rather they are frequently interrupted by inserted segments.

### 2.1.2 Delimitation of the inserted segments

A text reuse may include introductory texts, binding texts and paraphrased sections, which, while part of the text reuse from the author's point of view, do not refer to the biblical target. The content that directly refers to the biblical target must be isolated, and the other parts may be annotated with specific attributes.

The next example (which is the same text reuse instance as the previous example) presents an instance of text reuse that includes an explicit insertion formula<sup>17</sup> mixed with the biblical text quoted.

---

Audi et  
<seg type="scripturalQ" source="#Vg\_Ct\_5:36-41">  
<seg type="insertion" ana="#explicitScripturalAttribution">sponsum in Canticis ad profectum hunc trimodum  
invitantem</seg>  
Comedite  
<seg type="insertion" ana="#genericScripturalAttribution">, inquit,</seg>  
amici, et bibite, et inebriamini carissimi</seg>.

---

The text parts where a biblical text reuse is observed are marked with the element <seg>. This element includes the words borrowed from the Bible, but also introductory formulas and possible insertions. However, this system of sub-segmentation is not sufficient in the most complex cases, in particular when text reuses containing references to several biblical texts and with various insertions must be handled. Let's take as an example [Jerome, *Prologue to Samuel-Kings*]. The introduction formula is common to all texts, the terms taken from the book of Exodus are nested and present each in several verses.

---

[Jerome, *Prologus in libro Regum*]

[In Tabernaculum Dei [offert] (25:3; 35:5) unusquisque quod potest : alii [aurum et argentum et lapides pretiosos] (25:3; 35:5; 35:9), alii [byssum et purpuram, coccum offerunt et hyacinthum] (25:7; 35:6; 35:23) ; nobiscum bene agetur, si [obtulerimus pelles et caprarum pilos] (25:4-5; 35:7; 35:23)].

---

The encoding of such a text by segmentation and sub-segmentation is an insoluble puzzle. Due to the frequency of such complex cases in our corpus, we have decided to generalize the encoding by tokenization of words; it makes the management of simple cases more complicated but is essential for complex ones. A preliminary step of processing the text files has been planned. The files will be imported into TXM [http8] associated with TreeTagger [http10] to automate the preliminary marking of each word. This phase will also include lemmatization.

---

<sup>17</sup> About the typing of insertions, see below, 2.2.

The following example illustrates the encoding eventually chosen to solve complex cases of overlaps, using tokenisation and lemmatization. The `<span>` element, used with the ‘target’ attribute, allows a very precise segmentation.

---

[Jerome, *Prologus in libro Regum*]

```
<w xml:id="w_2">Tabernaculum</w> <w xml:id="w_3">Dei</w>
  <seg type="scripturalQ"><w xml:id="w_4">offert</w>
    <seg type="insertion" xml:id="sI_1"><w xml:id="w_5">unusquisque</w> <w xml:id="w_6">quod</w> <w
      xml:id="w_7">potest</w> <pc xml:id="lat.pc_1">:</pc> <w xml:id="w_8">alii</w> </seg>
    <w xml:id="w_9">aurum</w> <w xml:id="w_10">et</w> <w xml:id="w_11">argentum</w> <w xml:id="w_12">et</w> <w
      xml:id="w_13">lapides</w> <w xml:id="w_14">pretiosos</w>
    <seg type="insertion" xml:id="sI_2"><pc xml:id="lat.pc_2">,</pc> <w xml:id="w_15">alii</w> </seg>
    <w xml:id="w_16">byssum</w> <w xml:id="w_17">et</w> <w xml:id="w_18">purpuram</w> <pc xml:id="pc_3">,</pc>
    <w xml:id="w_19">coccum</w> <w xml:id="w_20">offerunt</w> <w xml:id="w_21">et</w> <w xml:id="w_22">hyacinthum</w>
    <seg type="insertion" xml:id="sI_3"><pc xml:id="lat.pc_4">,</pc> <w xml:id="w_23">nobiscum</w> <w
      xml:id="w_24">bene</w> <w xml:id="w_25">agetur</w> <pc xml:id="lat.pc_5">,</pc> <w xml:id="w_26">si</w> </seg>
    <w xml:id="w_27">obtulerimus</w> <w xml:id="w_28">pelles</w> <w xml:id="w_29">et</w> <w xml:id="w_30">caprarum</w>
    <w xml:id="w_31">pilos</w>

    <span target="#w_4 #w_9 #w_10 #w_11 #w_12 #w_13 #w_14 #w_16 #w_17 #w_18 #w_19 #w_20 #w_21 #w_22 #w_27 #w_28
      #w_29 #w_30 #w_31" type="scripturalOcc" xml:id="lat.sOcc.a.1"/>
    <span target="#w_16 #w_17 #w_18 #w_19 #w_22 #w_28 #w_29 #w_30 #w_31" type="scripturalOcc" xml:id="lat.sOcc.a.2"/>

  </seg>
<note n="a" place="foot" style="txt_Note" type="scripturalNote" xml:id="sNote.a">
  <seg type="bRef" xml:id="b.1">
    <bibl type="biblical">
      <ref cRef="Vg:Ex:25:2:12_25:3:7-10_25:4:1-4.8-10_25:5:2_25:7:1.4">Ex 25, 2-5.7</ref>
    </bibl>
  </seg>
  <link subtype="implicit" target="#lat.sOcc.a.1 #b.1" type="inexactQuotation"/>
  <seg type="bRef" xml:id="b.2">
    <bibl type="biblical">
      <ref cRef="Vg:Ex:35:5:11.14-17_35:6:1-3.7-9_35:7:2_35:9:1.4">Ex 35, 5-7.9</ref>
    </bibl>
  </seg>
  <link subtype="implicit" target="#lat.sOcc.a.1 #b.2" type="inexactQuotation"/>
  <seg type="bRef" xml:id="b.3">
    <bibl type="biblical">
      <ref cRef="Vg:Ex:35:23:4-6.9-13">Ex 35, 23</ref>
    </bibl>
  </seg>
  <link subtype="implicit" target="#lat.sOcc.a.2 #b.3" type="inexactQuotation"/>
</note>
```

---

## 2.2 Characterization of the patristic text reuse

A text reuse is partially characterized by the segment containing the quoted biblical text. Thus, we can specify if it is a lemma, that is to say a biblical section specifically commented on in a commentary, an homily or a biblical chain (*catena*), or a reuse of this lemma in the text.

---

[Bernard of Clairvaux, *Homilia in laudibus Virginis matris* 1, 1:106]

```
<seg type="scripturalQ" subtype="lemma" xml:id="B01" source="#Vg:Lc:1:26">Missus est angelus Gabriel a Deo in civitatem Galilae, cui
nomen Nazareth, ad Virginem desponsatam viro, cui nomen erat Joseph de domo David, et nomen Virginis Maria, et reliqua.</seg>
```

And a little further in the same homily:

```
<seg type="scripturalQ" subtype="lemmaReuse" xml:id="B02" source="#Vg:Lc:1:26:1-6" sameAs="#B01">
  <seg type="insertion" ana="#explicitScripturalAttribution"> Ait itaque: </seg>
Missus est angelus Gabriel a Deo</seg>
```

---

Further information will be given in the structure of the text itself: e.g., a biblical text reuse included in a quotation from another text will be identified as such by the nesting of tags.

---

Quotation of Ap:1:9 in the *Letter of the Gallic Martyrs to pope Eleutherios*, quoted itself by [Eusebius of Caesarea, *Historia Ecclesiastica* V, 4, 2:28, l. 5]:

```
<cit>
  <quote> (...)
    <seg type="scripturalQ" source="#LXX:Ap:1:9:3-7" >τὸν ἀδελφὸν ἡμῶν καὶ κοινώνον</seg> (...)
  </quote>
  <ref cRef="ID Bx.work">Gallic Martyrs, Letter to Eleutherios</ref>
</cit>
```

---

Most of the characterization elements, however, will be provided by the analysis of the segments inserted in the text reuse, mentioned in the previous section (2.1.2). As BIBLINDEX aims to remain as neutral as possible, and rather provide the user with materials for his/her own hermeneutic path, we decided not to include degrees of intentionality, even though they are at the heart of the text reuse ontologies commonly used (CITO, LAWDI, etc.): each analyst can indicate his/her own interpretation of the author's act of quoting in the commentary notes.

Among the formulas used to insert the biblical text in the reuse context, we have distinguished between neutral insertions (e.g. one reads, he says), and explicit ones, which show in one way or another that the author is aware that his text is borrowed from Scripture. The latter may be general mentions (e.g. according to the Scriptures) or more explicit attributions such as a more or less precise author's name (the Prophet, the Apostle, Job, etc.) or other clues, sometimes even canonical references. The quotations that are explicitly introduced can be categorised as correct, false (in case of a mistake due to the author for instance), agraphon (words attributed to Jesus without being in any gospel), or unknown which includes words not attested in any biblical book, but quoted as belonging to the Scripture by the patristic author.

### III. SUMMARY OF THE SELECTED ENCODING

The following is a complete encoding scheme for all these descriptive elements.

#### 3.1 Example: [Bernard of Clairvaux, *Homiliae in laudibus Virginis matris* I, 1]

---

```
<w xml:id="w_Miss.H1_355">Plena</w> <w xml:id="w_Miss.H1_356">quippe</w> <w xml:id="w_Miss.H1_357">sunt</w> <w
xml:id="w_Miss.H1_358">omnia</w> <w xml:id="w_Miss.H1_359">supernis</w> <w xml:id="w_Miss.H1_360">mysteriis</w><pc
xml:id="pc_Miss.H1_361">,</pc> <w xml:id="w_Miss.H1_362">ac</w> <w xml:id="w_Miss.H1_363">caelesti</w>
<w xml:id="w_Miss.H1_364">singula</w> <w xml:id="w_Miss.H1_365">dulcedine</w> <w xml:id="w_Miss.H1_366">redundantia
</w><pc xml:id="pc_Miss.H1_367">,</pc> <w xml:id="w_Miss.H1_368">si</w> <w xml:id="w_Miss.H1_369">tamen</w>
<w xml:id="w_Miss.H1_370">diligentem</w> <w xml:id="w_Miss.H1_371">habeant</w> <w xml:id="w_Miss.H1_372">inspectorem
</w><pc xml:id="pc_Miss.H1_373">,</pc> <w xml:id="w_Miss.H1_374">qui</w> <w xml:id="w_Miss.H1_375">noverit</w>
<seg type="scripturalQ" xml:id="sQ.Miss.H1.1.b">
  <w xml:id="w_Miss.H1_376">sugere</w> <w xml:id="w_Miss.H1_377">mel</w> <w xml:id="w_Miss.H1_378">de</w>
  <w xml:id="w_Miss.H1_379">petra</w> <w xml:id="w_Miss.H1_380">oleumque</w> <w xml:id="w_Miss.H1_381">de</w>
  <w xml:id="w_Miss.H1_382">saxo</w> <w xml:id="w_Miss.H1_383">durissimo</w>
  <span type="scripturalOcc" from="#w_Miss.H1_376" to="#w_Miss.H1_383" xml:id="sOcc.Miss.H1.1.b"/>
</seg>
<note type="scripturalNote" xml:id="sNote.Miss.H1.1.b">
  <seg type="bRef" xml:id="b.Miss.H1.1.2">
    <bibl type="biblical">
      <ref cRef="Vg:Dt:32.13:11-18">Dt 32, 13</ref>
      <ptr target="http://www.biblindex.info/fr/biblical/content/ref/Vg_Dt_32_13:11-18"
targetLang="lat"></ptr>
    </bibl>
  </seg>
  <link type="inexactQuotation" subtype="implicit" target="#sOcc.Miss.H1.1.b #b.Miss.H1.1.2"/>
</note>
<pc xml:id="pc_Miss.H1_389">.</pc>
```

---

### 3.2 Text reuse

	Content	Attributes		Value
		Name	Presence	
<seg>	Place of a text reuse w   pc   span   seg   note   text()	type	required	scripturalQ
		subtype	optional	lemma   lemmaReuse   lemmaAnticipatedUse   quotationReuse
		xml:id	required	

Table 2. Element <seg>

This element <seg type="scripturalQ"> contains an element <span> establishing words related to the biblical text.

	Content	Attributes		Value
		Name	Presence	
<span>	Words related to the biblical text	from	optional	#[xml:id of the first biblical word in the text reuse]
		to	optional	#[xml:id of the last biblical word in the text reuse]
		target	optional	#[xml:id of the first biblical word in the text reuse] #[xml:id of the second biblical word in the text reuse] {etc.}
		type	required	scripturalOcc
		xml:id	required	

Table 3. Element <span>

If the words borrowed from the Bible are contiguous, the ‘from’ and ‘to’ attributes are used. If they do not follow each other immediately the ‘target’ attribute is used.

The <seg type="scripturalQ"> element may contain insertions which are tagged with a <seg type="insertion"> element. Encoding focuses on the classification of segments that assign the text reuse to a speaker or source.

	Content	Attributes		Value
		Name	Presence	
<seg>	Words in the text reuse not referred to the biblical text w   pc   note   text()	ana	required	#explicitScripturalAttribution   #genericScripturalAttribution   #noAttribution
		type	required	insertion
		xml:id	required	

Table 4. Element <seg> applied to insertions

<seg ana="#explicitScripturalAttribution"> elements may contain a <note> element if some precision needs to be added.

	Content	Attributes		Value
		Name	Presence	
<note>	Precision	type	required	attributionNote

Table 5. Element <seg> applied to insertions (precisions)

### Special case of overlapping quotations

In some cases of overlapping text reuse the <join> element may be used to link two discontinuous <seg> elements that form a single text reuse.

	Content	Attributes		Value
		Name	Presence	
<join>		type	required	scripturalQ
		result	required	seg
		target	required	#[xml:id of the first <seg>] #[xml:id of the second <seg>]
		xml:id	required	

Table 6. Element <join>

### Encoding of the first text reuse segment

	Content	Attributes		
		Name	Presence	Value
<seg>	Place of a text reuse w   pc   span   seg   note   text()	type	required	scripturalQ
		next	required	#[xml:id of the next joined <seg>]
		xml:id	required	[xml:id of the <join> el.]_1

Table 7. First element <seg> in case of disjunction

### Encoding of the second text reuse segment

	Content	Attributes		
		Name	Presence	Value
<seg>	Place of a text reuse w   pc   span   seg   note   text()	type	required	scripturalQ
		prev	required	#[xml:id of the previous joined <seg>]
		xml:id	required	[xml:id of <join> el.]_2

Table 8. Second element <seg> in case of disjunction

	Content	Attributes		
		Name	Presence	Value
<span>	Words related to the biblical text	from	optional	#[xml:id of the first biblical word in the text reuse]
		to	optional	#[xml:id of the last biblical word in the text reuse]
		target	optional	#[xml:id of the first biblical word in the text reuse] #[xml:id of the second biblical word in the text reuse] {etc.}
		type	required	scripturalOcc
		xml:id	required	

Table 9. Element <span> in case of disjunction

### Special case of quotation reuse

When text reuse is repeated, it will be specified in the <span> element using the values of the ‘ana’ attribute and specifying the ‘sameAs’ attribute.



Element name	Content	Attributes		
		Name	Presence	Value
<span>	Words related to the biblical text	from	optional	#[xml:id of the first biblical word in the text reuse]
		to	optional	#[xml:id of the last biblical word in the text reuse]
		target	optional	#[xml:id of the first biblical word in the text reuse] #[xml:id of the second biblical word in the text reuse] {etc.}
		sameAs	required	#[xml:id of the <span> el. defining the reused text]
		type	required	scripturalOcc
		xml:id	required	

Table 10. Lemma reuse

## Scriptural notes

Each text reuse is marked with a scriptural note (each <seg type="scripturalQ"> is followed by an element <note type="scripturalNote">).

	Content	Attributes		
		Name	Presence	Value
<note>	seg   link	n	required	[n <sup>o</sup> ]
		type	required	scripturalNote
		xml:id	required	

Table 11. Scriptural note

The <note type="scripturalNote"> element consists of a <seg type="bRef"> element containing the canonical reference and a <link> element connecting the quoting patristic text and the quoted biblical text and specifying the nature of that relationship. There are as many <seg> elements associated to a <link> element as canonical Biblical references associated to the words encoded.

	Content	Attributes		
		Name	Presence	Value
<seg>	bibl   note	type	required	bRef
		xml:id	required	

Table 12. Canonical reference (<seg>)

	Content	Attributes		
		Name	Presence	Value
<bibl>	ref   ptr	type	required	biblical

Table 13. Canonical reference (<bibl>)

The <ref> element contained in the <bibl> element gives the canonical reference in written form.

	Content	Attributes		
		Name	Presence	Value
<ref>	ref	cRef	required	[abbreviated title of the Bible]:[abbreviated title of the book]:[chapter number].[verse number]: [ordering number of words in the verse]

Table 14. Canonical reference (<ref>)

The <ptr> element contains the URN of the biblical target in BIBLINDEX.

	Content	Attributes	Presence	Value
		Name		
<ptr>		target	required	[URI of the passage in BIBLINDEX]
		targetLang	required	

Table 15. Canonical reference (<ptr>)

The <note> element indicates the source of the biblical text quoted.

	Content	Attributes	Presence	Value
		Name		
<note>		type	required	originNote

Table 16. Canonical reference (<note>)

The <link> element establishes the relationship between a quoting text segment and one or more quoted biblical passages. It also specifies the nature of this relationship: exact or inaccurate, explicit or implicit quote.

	Content	Attributes	Presence	Value
		Name		
<link>		type	required	exactQuotation   inexactQuotation
		subtype	required	
		ana	optional	#allusion
		target	required	#[xml:id of the <span> element giving the words of the quoting text] #[xml:id of the <seg> element defining the biblical text quoted]

Table 17. Canonical reference (<link>)

## Comment notes

	Content	Attributes	Presence	Value
		Name		
<note>		n	required	[n <sup>o</sup> ]
		type	required	commentaryNote
		xml:id	required	

Table 18. Comment notes

## Conclusion

Trying to take into account every possible scenario, we have defined the markup scheme given above. It is currently implemented on our sample corpus of Bernard of Clairvaux's works. The encoding of other texts in real production conditions will allow us to clarify and refine this scheme.

Thanks to this markup, all the queries we have thought of are made possible. For instance, it can be searched for all reuses of a specific verse or set of verses in a defined patristic corpus; for all the quotations characterized by an explicit introduction; for the verses quoted near a specific verse. It's also possible to find a specific lemma in different verses, even when text reuses are embedded within each other. Different ways used by a specific author to insert a biblical passage in the course of his speech can be studied, as well as comparisons between several authors, etc. The next step is to design the search interfaces, in order to express all of these possible queries in the most user-friendly way possible.

In the meantime, any feedback from you, the readers, are welcome: please send a message to [bibindex.sc@mom.fr](mailto:bibindex.sc@mom.fr) if you have any comments or ideas for improvement.

## References

- Allenbach, J. *Étapes, moyens et méthode d'analyse pour la constitution du fichier microphotographique des citations de l'Écriture chez les Pères*. Faculté de théologie protestante (Strasbourg), 1967.
- Bady, G. Le miel des Écritures : la fréquence des références bibliques chez Jean Chrysostome et les trois Cappadociens ; Bibles et canons des Trois hiérarques : Basile de Césarée, Grégoire de Nazianze et Jean Chrysostome. [Marculescu-Mellerin, 2015]:121-179.
- Bernard of Clairvaux, *À la louange de la Vierge Mère (Homiliae in laudibus Virginis matris)*. Sources Chrétiennes, Éditions du Cerf (Paris), 2009<sup>2</sup>:390.
- Bernard of Clairvaux. *Éloge de la Nouvelle Chevalerie (Liber ad milites Templi de novae militiae)*. Sources Chrétiennes, Éditions du Cerf (Paris), 1990:367.
- Bernard of Clairvaux. *L'Amour de Dieu (De Diligendo Deo)*. Sources Chrétiennes, Éditions du Cerf (Paris), 2010<sup>2</sup>:393.
- Biblia Patristica, Index des citations et allusions bibliques dans la littérature patristique*. Éditions du CNRS (Paris): 1. *Des origines à Clément d'Alexandrie et à Tertullien*, 1975 – 2. *Le troisième siècle (Origène excepté)*, 1977 – 3. *Origène*, 1980 – 4. *Eusèbe de Césarée, Cyrille de Jérusalem, Épiphanie de Salamine*, 1987 – 5. *Basile de Césarée, Grégoire de Nazianze, Grégoire de Nysse, Amphiloque d'Iconium*, 1991 – 6. *Hilaire de Poitiers, Ambroise de Milan, Ambrosiaster*, 1995 – 7. *Didyme d'Alexandrie*, 2000 – *Supplément. Philon d'Alexandrie*, 1982.
- Eusebius of Caesarea, *Histoire Ecclésiastique V-VIII (Historia Ecclesiastica)*. Sources Chrétiennes, Éditions du Cerf (Paris), 1955:41.
- Gesche, S., Egyed-Zsigmond, E., Calabretto, S. References and Citations in Ancient Greek Documents, 2017: <hal-01497850>.
- http1 <http://www.bibindex.info>
- http2 <http://www.sourceschretiennes.mom.fr/>
- http3 <http://cartodialect.imag.fr/bibindex/ouvrages/>
- http4 <http://www.bibliissima-condorcet.fr/>
- http5 <http://www.bibindex.info/en/biblical/list/correspondence>
- http6 [http://www.sources-chretiennes.mom.fr/index.php?pageid=auteurs\\_anciens](http://www.sources-chretiennes.mom.fr/index.php?pageid=auteurs_anciens)
- http7 <http://www.dh.uni-leipzig.de/wo/lofts/>
- http8 <http://textometrie.ens-lyon.fr/>
- http9 <http://cite-architecture.github.io/cite/specs/cite/ctsUrn/Contents.html>
- http10 <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>
- Irenaeus (Irénee de Lyon). *Contre les Hérésies*. Sources Chrétiennes, Éditions du Cerf (Paris), 1982:294.
- Jerome. *Lettres*. CUF, Les Belles Lettres (Paris), 1951:III.
- Jerome. *Préfaces aux livres de la Bible*. Sources Chrétiennes, Éditions du Cerf (Paris), 2017:592 (forthcoming).
- Junod E. La Bible et les Pères, conférence donnée pour les 30 ans du CADP (Strasbourg), 1995: <[https://bibindex.hypotheses.org/a-propos/historique\\_projet/la-bible-et-les-peres-par-eric-junod](https://bibindex.hypotheses.org/a-propos/historique_projet/la-bible-et-les-peres-par-eric-junod)>.
- Marculescu, S., Mellerin, L. *Le miel des Écritures, Cahiers de BIBINDEX 1*. Cahiers de Biblia Patristica 15 (Strasbourg), 2015.
- Mellerin L. BIBINDEX, Index en ligne des références scripturaires chez les Pères de l'Église : état des lieux et perspectives. *Cahiers de Biblia Patristica*. 2017;17:145-170. (= Mellerin, 2017-1).
- Mellerin, L. Approches statistiques de la réception du livre de l'Éclésiaste chez les Pères. Un exemple d'utilisation de BIBINDEX. *La réception du livre de Qohélet dans l'Antiquité et au Moyen Âge*. Éditions du Cerf (Paris), 2016:79-112.
- Mellerin, L. Étude des usages bibliques d'Irénee de Lyon à l'aide de BIBINDEX. *Irénee de Lyon et l'émergence de la Bible chrétienne*. Brepols, *Instrumenta Patristica et Mediaevalia* (Turnhout), 2017 (= Mellerin, 2017-2).
- Mellerin, L. Methodological Issues in BIBINDEX, An Online Index of Biblical Quotations in Early Christian Literature. *Biblical Quotations in Patristic Texts*. Peeters (Leuven-Paris-Walpole), 2013:11-32.
- Mellerin, L. New Ways of Searching with BIBINDEX, the online Index of Biblical Quotations in Early Christian Literature. *Digital Humanities in Biblical, Early Jewish and Early Christian Studies*, Brill (Leiden), 2014;11:175-192.
- Mellerin, L. Prolégomènes à une approche statistique des citations scripturaires dans l'œuvre de Jérôme, à l'aide de BIBINDEX. *L'exégèse de saint Jérôme*. PUSE (Saint-Étienne), 2017. (= Mellerin, 2017-3).
- Moritz, M., Wiederhold, B., Pavlek, B., Bizzoni, Y., Büchler, M., Non-Literal Text Reuse in Historical Texts: An Approach to Identify Reuse Transformations and its Application to Bible Reuse, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, Texas, November 1-5, 2016:1849-1859. [<https://aclweb.org/anthology/D16-1190>].