

Relazione relativa alle attività svolte nell'ambito del lavoro di supporto alla ricerca consistente in: "Codifica XML e relativa correzione del manoscritto di Ivrea dell'Editto di Rotari ai fini dell'edizione elettronica del testo da far confluire nell'archivio del progetto ALIM"

Periodo di riferimento: 15/10/2018 – 15/01/2019

Responsabile progetto: Prof.ssa Marina Buzzoni

Titolare contratto di collaborazione: dott.ssa Marianna Giardina

Introduzione

Il progetto di codifica XML, oggetto della presente relazione, fa parte di un progetto più ampio frutto della collaborazione tra diversi team di lavoro che hanno finora operato con lo scopo di sviluppare un'edizione digitale dell'intero testo dell'*Editto di Rotari* da far confluire nell'archivio ALIM (Archivio della Latinità Italiana del Medioevo).

ALIM nasce negli anni novanta come progetto di ricerca e si pone come obiettivo quello di raccogliere e pubblicare edizioni critiche di testi in latino composti in Italia fra l'VIII e il XV secolo. Lo scopo di tale raccolta consiste nel rendere i testi latini disponibili in rete e liberamente consultabili per il vasto pubblico di studiosi.

Lo sviluppo del progetto di codifica XML ed edizione digitale dell'*Editto di Rotari* (d'ora in poi ER) rappresenta un'importante iniziativa che porterà alla crescita dell'archivio ALIM e darà la possibilità agli studiosi di analizzare il *corpus* delle leggi longobarde e di condurre studi approfonditi su di esso. In particolare, il punto di partenza del progetto consiste nella creazione dell'edizione diplomatico-interpretativa di due importanti codici manoscritti ("testimoni piemontesi"):

- il primo è il *Codex Vercellensis* CLXXXVIII (Vercelli, Archivio e Biblioteca Capitolare Eusebiana, ms 2 secondo la classificazione introdotta da Bluhme, VIII sec.). Vergato in onciale, costituisce una fonte storica di grande interesse culturale poiché, insieme al Sangallese (ms 1, *Codex Sangallensis* 730, fine VII sec.) è l'unico altro codice antico di epoca pienamente longobarda; trasmette le Leggi fino all'art. 139 del codice di Liutprando;
- il secondo è il *Codex Eporedianus* XXXIV- 5 (Ivrea, Biblioteca Capitolare, ms 3, IX sec.). Vergato in grafia carolina, contiene tutte le *Leges*, l'*Admonitio generalis* e alcuni capitolari (tra cui quelli di Pipino, Carlo Magno, Lotario).

I due testimoni piemontesi, in seguito ad un'accurata codifica strutturale e semantica, saranno pubblicati nello spazio web ALIM come edizioni diplomatiche digitali con il grande vantaggio di poter osservare contemporaneamente sia il testo digitalizzato, sia le immagini dei manoscritti. Questo sarà possibile grazie all'utilizzo e alle potenzialità del software EVT (Edition Visualization Technology) che permetterà di valorizzare l'opera dell'ER e di esplorare in modo interattivo le edizioni digitali.

Il software EVT consentirà l'accesso non solo all'immagine facsimile e alla trascrizione del testo in maniera simultanea, ma anche ad una serie di strumenti utili per la fruizione dei manoscritti. L'obiettivo principale, infatti, è quello di offrire un'edizione che permetta di ampliare la fruizione della fonte storica sfruttando la multidimensionalità dello spazio digitale.

In questo modo, rispetto alla diretta e tradizionale edizione a stampa, sarà possibile:

- visualizzare sia la versione diplomatica, sia quella interpretativa dei manoscritti;
- condurre indagini e analisi più dettagliate per individuare i tratti rilevanti dell'opera;
- ricavare informazioni utili e farne uso per costruire, per esempio, nuove edizioni critiche.

Tale edizione digitale, inoltre, sarà arricchita da strumenti per l'analisi lessicale e testuale che favoriranno l'analisi dei termini longobardi (e delle rispettive glosse latine, se presenti) nel contesto di trasmissione, facilitando in questo modo anche i confronti tra forme e grafie. Entrambi i codici e, in particolare, le loro edizioni diplomatico-interpretative costituiranno la base per la futura creazione di una nuova edizione critica dell'*ER*.

Attività svolte

In base alle esigenze dell'intero progetto, le attività di supporto alla ricerca sono state incentrate principalmente sulla codifica XML delle leggi longobarde contenute nel primo testimone piemontese, il *Codex Vercellensis 188* (ms di Vercelli).

Questa prima parte del lavoro è stata necessaria per garantire l'omogeneità con la codifica XML che verrà poi effettuata sul secondo testimone piemontese, il *Codex Eporedianus 34-5* (ms Ivrea).

Il processo di codifica del manoscritto di Vercelli è stato basato sull'utilizzo del linguaggio di markup XML (*eXtensible Markup Language*) e sullo standard di codifica TEI (Text Encoding Initiative), prendendo come riferimento le linee guida della TEI P5 ed il manuale ad uso interno del progetto contenente le istruzioni di codifica per il testo dell'*ER*.

All'interno del manuale "Istruzioni per la trascrizione e codifica delle *Leges Langobardorum*", infatti, sono riportate tutte le scelte di codifica effettuate in merito agli elementi XML da utilizzare per la struttura generale, gli interventi scribali ed editoriali, la marcatura del contenuto, i termini longobardi e i casi particolari.

Il testo dell'*ER* del *Codex Vercellensis* è stato, quindi, codificato mettendo in luce le peculiarità dell'antico manoscritto che, grazie alla marcatura TEI-XML, è stato predisposto per il livello di edizione diplomatica e interpretativa.

La codifica si è svolta in due fasi:

1. codifica **strutturale**, in cui è stata definita la struttura di base del testo ed è stato prodotto un documento TEI -XML valido; in questa fase è stato anche applicato il foglio di stile *w-wrap.xsl* in modo tale che ogni parola si trovi in un elemento `<w>` separato;
2. codifica **semantica**, in cui è stata verificata la correttezza dei risultati del foglio di stile, per poi proseguire con la marcatura semantica utilizzando degli elementi specifici.

Per eseguire la codifica è stato utilizzato l'editor di testo *oXygen XML Editor v.19.0*, strumento fondamentale per l'elaborazione dei testi xml e per verificare la loro validità rispetto ad uno schema. La codifica strutturale e semantica è stata effettuata solo su alcune leggi del *Codex Vercellensis* (composto da 388 capitoli), continuando in questo modo il lavoro precedentemente iniziato da altri codificatori. Nello specifico, sono stati trattati i capitoli che vanno da 191 a 360 che sono contenuti nei fogli 055v - 093r.

Durante le prime fasi del lavoro di codifica XML, il testo dell'*ER* è stato marcato secondo le norme TEI P5 che prevedono una struttura gerarchica in cui l'elemento `<TEI>` rappresenta la "radice" del documento ed include al suo interno altri elementi come `<teiHeader>` per i metadati relativi al documento e `<text>` per il testo del documento oggetto di codifica; all'interno di `<text>` è poi presente l'elemento `<body>` che contiene l'intero corpo del documento.

In merito ai metadati racchiusi all'interno del <teiHeader>, uno degli elementi importanti da citare è <charDecl> a sua volta racchiuso all'interno dell'elemento <encodingDesc>.

L'elemento <charDecl> assume particolare importanza poiché al suo interno, nel corso delle attività di codifica, sono stati inseriti tutti i caratteri speciali e le legature esistenti ed individuate all'interno del testo latino dell'*ER* di Vercelli.

I caratteri speciali e le legature sono stati identificati tra i caratteri messi a disposizione dal consorzio *Unicode* e dal gruppo di lavoro *MUFI* (Medieval Unicode Font Initiative) nella tabella presente in "MUFI character recommendation".

Dopo aver fornito i metadati necessari e fissato la struttura generale del testo, è stato possibile iniziare con la codifica del contenuto del manoscritto, cioè i capitoli delle leggi longobarde.

I capitoli dell'*ER* presi in esame (leggi 191-360, ff. 055v - 093r) sono stati codificati cercando di porre l'attenzione non solo sulla struttura, ma anche sulle caratteristiche distintive del testo e su casi linguistici particolari: specifica attenzione è stata posta per la codifica XML dei termini longobardi.

Nel *corpus* delle *Leges Langobardorum* e, nel nostro caso, nell'*ER* del *Codex Vercellensis* e del *Codex Eporedianus*, infatti, sono molte le tracce di termini longobardi, spesso latinizzati con il ricorso a desinenze della lingua di Roma (es. *mundius...*) o di origine longobarda (es: *angargathungi*, *faida*, *aldio*, *gastaldo*); sono, inoltre, presenti parole germaniche spesso affiancate da una traduzione in latino oppure formule latine spiegate in longobardo (es.: *faida*, *id est inimicitia* Roth. 326, ff. 84v).

Proprio per questo motivo, nell'edizione elettronica dell'*ER*, si è deciso di marcare i termini longobardi e di utilizzare una precisa codifica volta ad evidenziare questi particolari elementi linguistici. La marcatura accurata di tutti i termini longobardi consentirà non solo di facilitare la loro individuazione all'interno del testo, ma anche di collegare ciascun termine ad un glossario. In questo modo, il potenziale fruitore del testo avrà la possibilità di esplorare il materiale lessicale nel contesto immediato del testo che sta esaminando, di visualizzare una traduzione o un commento senza interrompere lo studio del brano e di operare confronti con le varianti ortografiche proprie di altri testimoni.

Al fine di consentire le operazioni sopra citate e le analisi lessicografiche automatizzate, la codifica per i termini longobardi è stata predisposta utilizzando gli elementi <term> ("termine tecnico") e <gloss> ("glossa o definizione"), come si può vedere dall'esempio che segue tratto dal *Codex Vercellensis* (foglio 078r, cap. 277):

```
<term xml:lang="lng">
  <w>
    <choice>
      <orig>aistant</orig>
      <reg resp="alim-pi beyerle">haistan</reg>
    </choice>
  </w>
</term>
<gloss>
  <w>
    <choice>
      <sic>it</sic>
      <corr resp="alim-pi">id</corr>
    </choice>
  </w>
  <w>est</w>
  <w>
    <choice>
      <sic>furonem</sic>
      <corr resp="alim-pi beyerle">furorem</corr>
    </choice>
  </w>
</gloss>
```

Nel caso dei termini longobardi, il tag <term> è stato usato per marcare la singola parola germanica e, ad esso, è associato l'attributo @xml:lang con valore "lng" che identifica la lingua longobarda.

Il tag <gloss>, invece, è stato usato per codificare il termine o l'espressione che fornisce una glossa o definizione per altre parole o espressioni.

Nell'esempio riportato sopra, la coppia di elementi <term> e <gloss> è stata utilizzata per marcare il termine longobardo *haistan* al quale è associata una glossa latina che ne spiega il significato.

Grazie a questo tipo di codifica, sarà possibile associare ad ogni termine longobardo la glossa corrispondente e generare un glossario separato.

Inoltre, questo tipo di marcatura renderà più semplice la ricerca dei termini longobardi in quanto risulteranno evidenziati nel testo con l'obiettivo di facilitare il lavoro degli studiosi per condurre studi linguistici sui termini germanici dell'*ER*.

In merito a questo meccanismo di codifica, non appena sarà verificato e concluso nel contesto del *Codex Vercellensis* sarà poi possibile estenderlo nel contesto del *Codex Eporedianus* ed anche alle altre edizioni diplomatiche che costituiranno la base dell'edizione critica finale.

In concomitanza con l'attività di codifica dei termini longobardi, un'altra attività svolta all'interno del progetto è stata quella relativa al controllo delle varianti ortografiche del latino medievale, grazie anche all'utilizzo del dizionario *DuCange* consultabile in rete.

Una questione importante, infatti, riguarda le correzioni e le regolarizzazioni nel caso di parole che presentino un tipo di ortografia non standard o nel caso di termini evidentemente errati.

Nel momento in cui si sono presentati questi casi, si è deciso di ricorrere all'elemento <choice> e agli elementi sic/corr e orig/reg in modo tale da determinare i due livelli di edizione. Tuttavia, non è facile stabilire con certezza se si tratti realmente di un errore del copista o se sia frutto della scarsa conoscenza della sintassi latina da parte dello scriba o, ancora, se si tratti di una forma particolare di latino medievale.

Per questo motivo, è stato opportuno eseguire un controllo linguistico per poi stabilire quali parole latine dovevano essere regolarizzate (usando le coppie orig/reg) oppure corrette (usando sic/corr).

Per valutare al meglio gli errori reali, si è deciso di tenere in considerazione tre edizioni critiche del testo dell'*ER*: la prima è Azzara, Gasparri, 2005, *Le leggi dei Longobardi. Storia, memoria e diritto di un popolo germanico*, la seconda è F. Beyerle, 1962, *Leges Langobardorum: 643-866* e la terza è Fr. Bluhme, 1868, *Leges Langobardorum* (MGH, LL IV).

Tenendo conto di questi riferimenti è stato, dunque, possibile eseguire un controllo linguistico attraverso un'operazione di confronto tra il testo delle leggi del *Codex Vercellensis* e il testo delle leggi delle edizioni critiche.

Da questo confronto è emerso che sono presenti una o più varianti per lo stesso termine, quindi, è utile riportare le lezioni registrate dagli editori precedenti.

In generale, si è deciso di:

- adottare una soluzione tesa a favorire la marcatura con <orig>/<reg> piuttosto che con <sic>/<corr> in tutti i casi possibili;
- indicare in <reg> e <corr> i responsabili usando l'attributo "@resp".

Inoltre, per indicare i responsabili di un'edizione è stato necessario inserire, all'interno del *teiHeader* e sotto l'elemento <titleStmt>, il seguente codice:

```
<editionStmt>
  <edition>Edizioni precedenti</edition>
  <respStmt xml:id="azzara">
    <resp>Editor</resp>
    <name ref="#Azzara2005">Azzara-Gasparri</name>
  </respStmt>
  <respStmt xml:id="beyerle">
    <resp>Editor</resp>
    <name ref="#Beyerle1962">Beyerle</name>
  </respStmt>
  <respStmt xml:id="bluhme">
    <resp>Editor</resp>
    <name ref="#Bluhme1868">Bluhme</name>
  </respStmt>
</editionStmt>
```

In questo modo, si avrà una lista ben precisa dei responsabili delle edizioni precedenti ai quali potrà essere attribuita la variante di un certo termine.

Un esempio di codifica che presenta l'uso di *orig/ reg* e registra le lezioni degli editori è il seguente:



```
<w>si</w>
<w>eum</w>
<w>
  <choice>
    <orig>occisissit</orig>
    <reg resp="alim-pi azzara bluhme beyerle">occidisset</reg>
  </choice>
</w>
```

Come è possibile osservare nell'esempio riportato sopra, l'elemento <choice> permette di raggruppare un numero di codifiche alternative per la stessa porzione di testo e di determinare i diversi livelli di edizione. L'elemento<choice> contiene gli elementi <orig> e <reg>: il primo contiene la forma originale di una lezione ed il termine viene marcato per come si presenta nel manoscritto, il secondo contiene una lezione che è stata regolarizzata o normalizzata e il termine viene indicato nella sua forma ortografica considerata regolare.

In conclusione, sulla base delle attività sopra descritte e svolte nel periodo di riferimento, gli obiettivi intermedi prefissati sono stati raggiunti.

In particolare, l'attività di controllo generale e di codifica per l'ER di Vercelli - al fine di poter poi garantire l'omogeneità tra i manoscritti - è quasi terminata.

Nei prossimi mesi, si prevede il raggiungimento dei successivi obiettivi dell'intero progetto. Nello specifico saranno portate avanti le seguenti attività: trascrizione e codifica XML dell'ER del manoscritto di Ivrea e creazione di un glossario XML da collegare a tutte le edizioni dei vari manoscritti. Verrà, dunque, proposta una struttura per la messa a punto di un glossario longobardo-latino che sarà allestito grazie agli elementi del modulo TEI per la creazione di dizionari, descritto nel capitolo 9. *Dictionary* di TEI P5-Guidelines.