

# Un modello integrato per il *Roman d'Alexandre* del codice Correr 1493: annotazione linguistica e edizione critica digitale

Giacomo Costa<sup>1</sup>, Simone Zenzaro<sup>2</sup>

<sup>1</sup> Università degli Studi di Padova, Italia – giacomo.costa.4@phd.unipd.it

<sup>2</sup> Cnr-Istituto di Linguistica Computazionale "Antonio Zampolli", Italia - simone.zenzaro@cnr.it

## ABSTRACT (ITALIANO)

L'obiettivo di questa proposta è di presentare alcune riflessioni, a livello sia pratico sia teorico, su di un nuovo modello di edizione critica digitale in cui i metadati derivati dall'annotazione linguistica di un testo possano essere assimilati e consultati. Lo scopo è di superare la divisione che spesso intercorre tra il piano ecdotico-filologico e quello dei corpora linguistici, proponendo, pertanto, lo sviluppo di un sistema integrato in cui il metadato linguistico possa essere interrogato in tutte le sue parti direttamente dall'interfaccia dedicata all'edizione critica digitale del testo oggetto di annotazione. L'oggetto di indagine di questo progetto sono testi che presentano particolarità linguistiche, in particolar modo legati a fenomeni di ibridismo tra diverse varietà e code-mixing. La *proof of concept* è in fase di sviluppo su di un testo medioevale appartenente al corpus del cosiddetto francese d'Italia, ovvero l'insieme di testi composti o trascritti in francese da autori della Penisola, tra il XIII e il XV secolo, nei quali i volgari medioevali italiani hanno trovato punti di contatto e di fusione con le varietà del francese antico.

**Parole chiave:** edizioni critiche digitali; annotazione linguistica automatica; corpora linguistici; Francese d'Italia (XIII-XV secolo)

## ABSTRACT (ENGLISH)

*An Integrated Model for the Roman d'Alexandre of the manuscript Correr 1493: Linguistic Annotation and Scholarly Digital Edition*

The aim of this proposal is to present some reflections, both practical and theoretical, on a new model of scholarly digital editing in which metadata derived from the linguistic annotation of a text can be integrated and consulted. The goal is to overcome the division that often exists between the philological-ecdotical level and that of linguistic corpora, proposing the development of an integrated system where linguistic metadata can be queried in all its parts directly from the interface dedicated to the digital edition of the annotated text. The object of investigation for this project is texts that present linguistic peculiarities, related to phenomena of hybridity between different varieties and code-mixing. The proof of concept is under development on a medieval text from the corpus of what is called "The French of Italy", that is texts composed or transcribed by Italian authors between the 13th and 15th centuries, in which the medieval Italian vernaculars encountered points of contact and fusion with varieties of Old French.

**Keywords:** digital scholarly editions; automatic linguistic annotation; linguistic corpora; the French of Italy (13<sup>th</sup> -15<sup>th</sup> century)

## 1. INTRODUZIONE

Lo sviluppo di un'edizione critica digitale integrata alle buone pratiche dell'annotazione linguistica nasce dall'esigenza di rappresentare e strutturare in modo integrale e ordinato una serie di informazioni ecdotiche, linguistiche e letterarie insite nei testi appartenenti al corpus medioevale del cosiddetto francese d'Italia (Gambino & Beretta, 2023). Per questo motivo proponiamo un prototipo di modello di edizione per il *Roman d'Alexandre* del codice Venezia, Biblioteca del Museo Correr, Correr 1493 (per l'edizione si veda La Du, 1937; per il facsimile del codice Benedetti, 1998).

Questo prototipo di edizione critica digitale integrato all'insieme dei dati derivati dal lavoro di annotazione linguistica ha l'obiettivo di configurarsi come modello unitario e generale per la banca data *RIALFrI* (*Repertorio Informatizzato Antica Letteratura Franco-Italiana*)<sup>1</sup> che raccoglie in linea il corpus dei testi del francese d'Italia. Al momento, difatti, le edizioni dei testi presenti nel repertorio si potrebbero configurare

---

<sup>1</sup> <https://www.rialfri.eu/>.

come edizioni «digitalizzate» poiché costituiscono digitalizzazioni di edizioni a stampa già esistenti. Tuttavia, con l'avviarsi di un progetto PRIN (*FrIngE - The French in/of Italy: Code-MixiNG in Medieval Europe* [2023-2025, PRIN 2022XRFSXR]), che ha portato all'annotazione linguistica completa di un certo numero di testi franco-italiani, è sorta anche la necessità di ripensare al modello di edizioni, per le quali il digitale non fosse più solo un medium di rappresentazione, ma anche e soprattutto la base di partenza per il lavoro, con un valore trasformativo su quest'ultimo (Pierazzo & Mancinelli, 2020: 9).

## 2. IL PROCESSO DI EDIZIONE

Il prototipo del modello di edizione è stato sviluppato a partire dal *Roman d'Alexandre* del codice Venezia, Biblioteca del Museo Correr, Correr 1493. La versione del *Roman* contenuta in questo manoscritto, databile al secondo quarto del XIII secolo, presenta una serie di caratteristiche linguistiche (tratti di ibridismo) ed ecdotico-materiali tali per cui risulta essere un campo di prova ideale per lo sviluppo di un modello di edizione digitale integrato.

Dopo una prima fase di lavoro filologico, durante la quale il testo è stato accuratamente trascritto e editato, quest'ultimo è stato sottoposto al lavoro di annotazione linguistica. Il testo è stato annotato linguisticamente tramite l'applicazione web *Pyrrha*, sviluppata dall'*École nationale des chartes* per l'analisi del francese antico (Clérice et al., 2024). L'utilizzo di uno strumento allenato sulle varianti del francese antico su di un testo franco-italiano ha posto, fin dal principio, alcuni problemi legati alla scelta del *référentielle* per tutte quelle forme non appartenenti allo standard oitanico e alla necessità di allenare ulteriormente il modello affinché potesse essere efficiente anche per i testi del francese d'Italia (su questo aspetto si noti il lavoro già citato del progetto PRIN: *FrIngE - The French in/of Italy: Code-MixiNG in Medieval Europe* [2023-2025, PRIN 2022XRFSXR]; per i primi risultati cfr. Gambino et al., 2024). Tramite questo strumento di annotazione e post-correzione, ad ogni forma è stato associato il lemma, la *part of speech* e il *morph tag*<sup>2</sup>. Inoltre, sono state applicate delle specifiche etichette all'analisi morfologica che segnalano la presenza di una forma ibrida o italianismo (SPEC=it), di un hapax causato dalla posizione in rima (SPEC=rim), di latinismi (SPEC=lat), la presenza di occitanismi (SPEC=occ) e, infine, di eventuali forme filologicamente complesse o di difficile attribuzione lemmatica (SPEC=probl). In particolar modo, l'utilizzo dell'etichetta «SPEC=it» permette di individuare tratti italianizzanti per certi aspetti devianti rispetto allo "standard" oitanico, consentendo, dunque, di svolgere delle indagini sulla distribuzione e sulla tipologia degli italianismi. Si riporta in Figura 1 un esempio di lavoro di annotazione. Le stringhe bianche sono state analizzate correttamente dal software, mentre il colore blu indica un intervento di correzione del risultato da parte dell'annotatore umano:

Id	Form	Lemma	POS	Morph	Context	Similar	Save	+
801	27rb	croire	OUT	MORPH=empty	PONtbl c. 27rb De Eumenedus que	0	Save	+
802	De	de	PRE	MORPH=empty	c. 27rb De Eumenedus que pasma	63	Save	+
803	Eumenedus	Eumenedus	NOMpro	NOMB.=s GENRE=m CAS=r	. 27rb De Eumenedus que pasma por	8	Save	+
804	que	que2	PROrel	SPEC=it NOMB.=s GENRE=m CAS=n	27rb De Eumenedus que pasma por lo	144	Save	+
805	pasma	pasmer	VERcjg	MODE=ind TEMPS=psp PERS.=3 NOMB.=s	De Eumenedus que pasma por lo sanc	0	Save	+
806	por	por2	PRE	MORPH=empty	Eumenedus que pasma por lo sanc que	20	Save	+
807	lo	le	DETdef	NOMB.=s GENRE=m CAS=r	que pasma por lo sanc que il	4	Save	+
808	sanc	sanc	NOMcom	NOMB.=s GENRE=m CAS=r	pasma por lo sanc que il avoit	1	Save	+

Figura 1. Annotazione del testo

L'integrazione di tutte queste informazioni linguistiche all'interno di un'edizione critica trova nel mezzo digitale, e in particolar modo nella codifica in XML/TEI, una sua congenita collocazione. Il supporto digitale, difatti, risulta l'unico efficace per questo obiettivo, sia per quanto riguarda la possibilità di conservazione a lungo termine dei dati sia per rendere disponibili a tutti una serie di informazioni altrimenti difficilmente rappresentabili tramite il supporto cartaceo.

Il software di annotazione *Pyrrha* permette di esportare i dati dell'annotazione linguistica direttamente in

<sup>2</sup> Il sistema di etichette per l'annotazione linguistica segue lo schema di Cattex (<http://bfm.ens-lyon.fr/spip.php?article176>).

formato XML/TEI grezzo, di cui la Figura 2 ne è un esempio, che costituisce la base di partenza per lo sviluppo dell'edizione digitale.

```
<w xml:id="t1" n="1" lemma="conte2" type="POS=NOMcom|NOMB.=s|GENRE=m|CAS=r" >Conte</w>
<w xml:id="t2" n="2" lemma="vouloir" type="POS=VERc|MODE=ind|TEMPS=psp|PERS.=1|NOMB.=s" >voil</w>
<w xml:id="t3" n="3" lemma="dire" type="POS=VERinf|MORPH=empty" >dire</w>
<w xml:id="t4" n="4" lemma="par" type="POS=PRE|SPEC=it|MORPH=empty" >per</w>
<w xml:id="t5" n="5" lemma="rime4" type="POS=NOMcom|NOMB.=s|GENRE=f|CAS=r" >rime</w>
<w xml:id="t6" n="6" lemma="et" type="POS=CONcoo|MORPH=empty" >e</w>
<w xml:id="t7" n="7" lemma="par" type="POS=PRE|MORPH=empty" >per</w>
<w xml:id="t8" n="8" lemma="léonime" type="POS=NOMcom|NOMB.=s|GENRE=f|CAS=r" >léonine</w>
<w xml:id="t9" n="9" lemma="de" type="POS=PRE|MORPH=empty" >De</w>
<w xml:id="t398" n="10" lemma="le" type="POS=DETdef|NOMB.=s|GENRE=m|CAS=r" >#l</w>
<w xml:id="t10" n="11" lemma="fil2" type="POS=NOMcom|NOMB.=s|GENRE=m|CAS=r" >fil</w>
<w xml:id="t11" n="12" lemma="Félice" type="POS=NOMpro|NOMB.=s|GENRE=m|CAS=r" >Félice</w>
<w xml:id="t12" n="13" lemma="le" type="POS=DETdef|NOMB.=s|GENRE=m|CAS=r" >le</w>
<w xml:id="t13" n="14" lemma="roi2" type="POS=NOMcom|NOMB.=s|GENRE=m|CAS=r" >roi</w>
<w xml:id="t14" n="15" lemma="de" type="POS=PRE|MORPH=empty" >de</w>
<w xml:id="t15" n="16" lemma="Macedoine" type="POS=NOMpro|NOMB.=s|GENRE=f|CAS=r" >Macedoine</w>
```

**Figura 2. XML/TEI di partenza del testo annotato**

I metadati linguistici in formato XML/TEI vengono successivamente rimodellati secondo la struttura del testo in versi e integrati con gli elementi di carattere filologico e letterario, come l'apparato critico, le varianti, note di commento e, inoltre, la riproduzione facsimilare del codice. Ad arricchire questo già complesso sistema di informazioni linguistiche e testuali, si riscontra nel *Roman d'Alexandre* del codice Correr 1493 un abbondante drappello di elementi paratestuali, come rubriche e note al miniatore, che sono stati altresì oggetto di annotazione linguistica, e che richiedono, dunque, di essere integrati nell'edizione critica digitale. Essendo gli elementi paratestuali portatori di preziose informazioni linguistiche, la sfida è stata quella di cercare di integrare quest'ultimi all'interno del testo in versi, rispettando la loro posizione, in questo caso - almeno per i nostri scopi - non più liminale, all'interno del codice. Sono state vagliate diverse alternative per riuscire a rappresentare la presenza di marginalia all'interno di una carta del codice, come la possibilità di inserirli tra due lasse o di raggrupparli tutti all'interno della descrizione del codice nel `<teiHeader>`, per arrivare, alla fine, alla soluzione mostrata in Figura 3.

```
<pb type="carta" n="1r"/>
<noteGrp type="paratesto">
  <note xml:id="min_rub1ba" target="#lassa3" n="notal" type="miniatura_rubricata">
    <figure type="miniatura" n="clba" place="margine lato dx">
      <head type="attribuito"> La nascita di Alessandro </head>
      <figDesc>
        <mentioned rendition="rubrica">
          <w xml:id="t146" n="146" lemma="le" pos="DETdef" msd="NOMB.=s|GENRE=m|CAS=r" >Li</w>
          <w xml:id="t147" n="147" lemma="jor" pos="NOMcom" msd="NOMB.=s|GENRE=m|CAS=r" >jor</w>
          <w xml:id="t148" n="148" lemma="que2" pos="PROrel" msd="NOMB.=s|GENRE=m|CAS=r" >qu&#39;</w>
          <w xml:id="t149" n="149" lemma="Alexandre" pos="NOMpro" msd="NOMB.=s|GENRE=m|CAS=n" >Alexandre</w>
          <w xml:id="t150" n="150" lemma="naistre" pos="VERc|MODE=ind|TEMPS=psp|PERS.=3|NOMB.=s" >
          <w xml:id="t151" n="151" lemma="estre1" pos="VERc|MODE=ind|TEMPS=psp|PERS.=3|NOMB.=p" >furent</w>
          <w xml:id="t152" n="152" lemma="le" pos="DETdef" msd="NOMB.=p|GENRE=m|CAS=n" >li</w>
          <w xml:id="t153" n="153" lemma="signel" pos="NOMcom" msd="NOMB.=p|GENRE=m|CAS=n" >signes</w>
          <w xml:id="t154" n="154" lemma="en1" pos="PRE" msd="MORPH=empty" >en</w>
          <w xml:id="t155" n="155" lemma="ciel" pos="NOMcom" msd="NOMB.=s|GENRE=m|CAS=r" >ciel</w>
          <w xml:id="t156" n="156" lemma="." pos="PONfrt" msd="MORPH=empty" >.</w>
        </mentioned>
      </figDesc>
    </figure>
  </note>
```

**Figura 3. Integrazione degli elementi paratestuali nel testo in versi**

Tutti gli elementi paratestuali sono stati inseriti all'inizio di ogni carta, all'interno del tag `<noteGrp>` personalizzato dal `<type>` "paratesto". All'interno è stata segnalata l'eventuale presenza di rubriche o di miniature rubricate, con la rispettiva analisi linguistica, oppure di istruzioni per il miniatore, anch'esse annotate linguisticamente. Ogni elemento è stato associato a un "id" a sua volta collegato alla lassa a cui questi elementi si riferiscono.

Per garantire la massima fruizione dell'edizione digitale in XML/TEI che riesca a rappresentare in modo ordinato tutti gli elementi, è stato necessario accompagnarla con una piattaforma Web personalizzata, nella quale è possibile consultare le immagini facsimilari del manoscritto (tipicamente giustapposte al testo), la trascrizione interpretativa, l'apparato, la presenza di marginalia e, infine, alcune note di accompagnamento alla lettura. A fianco a questa struttura di partenza, verrà introdotto un sistema di interrogazione lessicale sia di tipo qualitativo - tramite una visualizzazione a colori degli schemi sintattici delle parole nel testo ispirata a *Iliadoscope* (Bouvier, Jambé, Zenzaro, 2020) - sia quantitativo - tramite

ricerche di corrispondenza lessicale e morfologica che rendano conto dei diversi gradi di ibridismo linguistico dei testi del francese d'Italia.

## BIBLIOGRAFIA

- Benedetti, R. (1998). *Le Roman d'Alexandre*, riproduzione del ms. Venezia, Biblioteca Museo Correr, Correr 1493. Udine: Roberto Vattori Editore.
- Bouvier, D., Jambé, A., & Zenzaro, S. (2020). *Iliadoscope*, Le devenir numérique d'un texte fondateur: l'Iliade et le Genavensis Graecus 44 (1.0.0). <https://www2.unil.ch/iliade/#/proj>.
- Clérice, T., Janès, J., Pilla, J., Camps, J-B., Pinche, A., Gille-Levenson, M., & Jolivet, V. *Pyrrha*, A language independent post correction app for POS and lemmatization (4.0.0) [Computer software], Zenodo. <https://doi.org/10.5281/zenodo.2325427>.
- Gambino, F., & Beretta, A. (2023). *Antologia del francese d'Italia: XIII-XV secolo*. Bologna: Pàtron.
- Gambino, F., Beretta, A., Barillari, S., Ceresato, F., Costa, G., Fassanelli, R., Favaro, M., Fois, J., Guadagnini, E., Guariglia, F., Parodi, M., & Rettore, C. *Il francese d'Italia e il progetto "FrIngE". Panoramica generale e casi di studio*. *Francigena V*. 10 (2024), 285-340. <https://www.francigena-unipd.com/index.php/francigena/article/view/100>.
- Guillot, C., Prévost S., & Lavrentiev, A. (2013). *Principes d'annotation Cattetex09 (Version 2.0)*, Lyon, École normale supérieure de Lyon, 8 avril 2013, [http://bfm.ens-lyon.fr/IMG/pdf/Cattetex2009\\_principes\\_2.0.pdf](http://bfm.ens-lyon.fr/IMG/pdf/Cattetex2009_principes_2.0.pdf).
- La Du, M. S. (1937). *The Medieval French Roman d'Alexandre*, vol. I, Text of the Arsenal and Venice Versions, prepared with an introduction and a commentary by Milan S. La Du. Princeton: Princeton University Press; rist. (1965) New York: Kraus Reprint Corporation ("Elliott Monographs", 36).
- Pierazzo, E., & Mancinelli, T. (2020). *Che cos'è un'edizione scientifica digitale*. Roma: Carocci («Le bussole» 599).
- RIALFrI - Repertorio Informatizzato Antica Letteratura Franco – Italiana. Diretto da Francesca Gambino, Università degli Studi di Padova, Dipartimento di Studi Linguistici e Letterari, versione 2.0, 2021. [www.rialfri.eu](http://www.rialfri.eu).
- TEI (Text Encoding Initiative), TEI P5: Guidelines for Electronic Text Encoding and Interchange, by the TEI Consortium; Originally edited by C.M. Sperberg-McQueen and Lou Burnard for the ACH-ALLC-ACL; Text Encoding Initiative Now entirely revised and expanded under the supervision of the Technical Council of the TEI Consortium, 2023. <https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>.