

Materni, Marta

L'ancien français attraverso la variazione manoscritta : la proposta della nuova base testuale EXOMARE

Études romanes de Brno. 2024, vol. 45, iss. 4, pp. 105-127

ISSN 2336-4416 (online)

Stable URL (DOI): <https://doi.org/10.5817/ERB2024-4-6>

Stable URL (handle): <https://hdl.handle.net/11222.digilib/digilib.81316>

License: [CC BY-SA 4.0 International](https://creativecommons.org/licenses/by-sa/4.0/)

Access Date: 20. 02. 2025

Version: 20250219

Terms of use: Digital Library of the Faculty of Arts, Masaryk University provides access to digitized documents strictly for personal use, unless otherwise specified.

L'ancien français attraverso la variazione manoscritta: la proposta della nuova base testuale EXOMARE

The Old French through Manuscript Variation: the Proposal of the New EXOMARE Textual Base

MARTA MATERNI [marta.materni@unitus.it]

Università della Tuscia, Italia

RIASSUNTO

Il contributo intende illustrare i criteri che hanno presieduto alla costituzione di una nuova base testuale consacrata all'*ancien français*: la base EXOMARE. Il progetto si focalizza su un segmento limitato di uso della lingua: opere appartenenti alla dimensione letteraria dello scritto, in particolar modo al genere *roman*. Come primo caso di studio la base dedicherà una specifica attenzione al sotto-genere dei *romans antiques*, fra il 1150 e il 1250, dal momento che è nei loro testi che si afferma l'espressione *mettre en romanz* nella sua doppia valenza di traduzione linguistica, dal latino al volgare, e formale, dalle fonti all'ottosillabo. L'elemento che dovrebbe caratterizzare la base EXOMARE in confronto alle altre già esistenti è in primo luogo la scelta di costruire il corpus testuale in modo omogeneo rispetto alla tipologia testuale, distinguendo fra testualità letteraria e testualità documentaria (esclusa); e in secondo luogo la proposta di creare dei "corpora d'Opera", di utilizzare cioè, per ciascuna opera, la trascrizione di tutti i manoscritti della sua tradizione, al fine di seguire la variazione linguistica di uno stesso bacino lessicale, quello del testo dell'opera, attraverso i suoi vari testimoni.

PAROLE CHIAVE

Diacronia linguistica; diatopia linguistica; linguistica storica; ancien français; linguistica dei corpora

ABSTRACT

This paper intends to illustrate the criteria that presided over the establishment of a new textual base dedicated to the Old French: the EXOMARE base. The project focuses on a limited segment of language use: the works belonging exclusively to the literary dimension of writing, with a focus on the *roman*. As a first case study the base will dedicate specific attention to the sub-genre of *Romans antiques*, between 1150 and 1250, since it is in their texts that the expression *mettre en romanz* is established in its double value of linguistic translation, from Latin to the vernacular, and formal translation, from the sources to the octosyllable. The element that should characterize the EXOMARE base compared to other existing ones is first of all the choice to build the textual corpus in a homogeneous way with respect to the textual typology, distinguishing between literary textuality and documentary textuality (excluded); and secondly, the proposal to create "Works corpora", that is, to use, for each work the transcription of *all* the manuscripts

of its tradition, in order to follow the linguistic variation of the same lexical set, that of the text of the work, through its various witnesses.

KEYWORDS

Linguistic diachrony; linguistic diatopy; historical linguistics; ancient français; corpus linguistics

RICEVUTO 2024-02-31; ACCETTATO 2024-06-30

Il contributo si propone innanzitutto di giustificare la scelta di mettere in cantiere un nuovo corpus (o meglio base, come si dirà in seguito) linguisticamente annotato dedicato al francese medievale, piuttosto che di contribuire a progetti già esistenti. Si tratta della base EXOMARE *Exploring Old French through Manuscript Readings*, concepita nel contesto di un progetto triennale finanziato dal MUR-Ministero dell'Università e della Ricerca italiano (PNRR Young Researcher Marie Curie n° MSC_0000085, Università della Tuscia, 2022-2025), la quale base rappresenta anche il punto d'arrivo di un percorso di ricerca e creazione di strumenti legati alla rappresentazione e analisi informatizzate della testualità medievale iniziato nel 2017 e realizzati grazie al contributo di due Marie Curie Fellowship (DIGIFLOR n° 745821, *Digital Edition of the 'Roman de Florimont'*, Université Grenoble Alpes 2017-2019¹; PRODIGI n° 886478, *'Prose 2' Digital Lemmatized Edition*, Università di Padova 2020-2022²).

In secondo luogo, il contributo intende illustrare i criteri adottati nella costruzione del corpus e offrire una panoramica dei potenziali sviluppi futuri che, soprattutto, nei *desiderata* di chi ha realizzato questo progetto dovrebbero muoversi nella direzione dell'integrazione possibile fra studi strettamente linguistici e metodo filologico, così che i corpora annotati linguisticamente possano fornire un contributo all'edizione di un testo in un'ottica filologica orientata alla ricostruzione.

Per giustificare l'assenza finale di dati di esempio ricavati dal progetto stesso e il possibile apparente sbilanciamento dell'esposizione nella direzione della puntigliosa contestualizzazione della base EXOMARE nel quadro delle altre iniziative analoghe già in essere, si specifica immediatamente che, alla data di stesura e pubblicazione del contributo, la produzione dei dati stessi è ancora in corso, essendo da poco entrati nella seconda fase, quella cioè della lemmatizzazione/annotazione linguistica, dopo il primo anno dedicato alla trascrizione dei testimoni. Le tappe stesse del processo (trascrizione/codifica TEI, lemmatizzazione/annotazione, allineamento dei testimoni), strettamente sequenziali e consequenziali, implicano forzatamente che le osservazioni sui dati – e in ultima analisi la verifica concreta dell'efficacia del modello proposto – siano realizzabili solo a valle dell'operazione, al termine di tutto il processo. Quel che sembrava da giustificare al momento, anche in assenza di dati, era tuttavia il modello stesso, soprattutto in relazione a una pluralità di esperienze già in essere e ampiamente consolidate rispetto alle quali, senza adeguata giustificazione, esso sarebbe potuto apparire un superficiale duplicato.

1 <http://digiflorimont.huma-num.fr/flsite/florimont.html>.

2 <https://mmaterni.github.io/prodigi/static/prodigi.html>.

A dare lo spunto iniziale per la costituzione di questa base, che come si vedrà pone al centro la scrittura manoscritta nei molteplici manoscritti, è stata una considerazione proposta da Martin Glessgen in un articolo orientato alla valorizzazione del dato linguistico derivante da trascrizione del manoscritto e non da edizione critica: “la variation entre différents manuscrits à l’intérieur d’une même tradition textuelle exige plus de réflexion” (Glessgen 2003: 57). Se il richiamo al dato documentario è abbastanza ovvio provenendo da uno studioso che ha fatto della scriptologia il suo principale ambito di interesse (e non si tratta, a dire il vero, di una sua preoccupazione esclusiva, ma al contrario piuttosto condivisa), il vero elemento di interesse è quel richiamo a prendere in considerazione la variazione fra i Testi di una stessa Opera. Ed è proprio da qui che prende le mosse la proposta EXOMARE.

1. Criteri di costituzione del corpus

1.1 L’acronimo EXOMARE

L’acronimo del progetto, corrispondente allo scioglimento *Exploring Old French through Manuscript REadings*, che rappresenta anche il nome che si intende attribuire alla base testuale-linguistica, focalizza i due elementi principali che vogliono costituire la cifra stilistica della base stessa e che rappresentano anche, ci si augura, la fondamentale giustificazione per la sua esistenza accanto ad altre basi linguistiche con, alle spalle, anni di accumulo di dati e una ricchezza di informazioni ovviamente ineguagliabile per un’iniziativa che muove i suoi primi passi. Per fare un esempio in termini numerici, quantitativi, il NCA (*Nouveau Corpus d’Amsterdam*) può contare su circa 140 elementi testuali distribuiti in circa 300 manoscritti³. La BFM (*Base de Français Médiéval*), con ormai 35 anni di storia (l’inizio di questa impresa testuale si pone nel 1989), annovera oggi 219 testi per 6.400.000 di occorrenze-parole, di cui 612.000 lemmatizzate e un milione annotate. In un certo senso, proprio perché esistono già questi grandi edifici in cui il qualitativo si sposa con un quantitativo su larga scala, rendendo possibile fornire al tempo stesso quadri d’insieme e di dettaglio, il progetto EXOMARE vorrebbe al contrario orientarsi esclusivamente verso il dettaglio, sperimentando un approccio di micro-scala, per certi versi chirurgico. Si tratta di una metodologia rispetto alla quale, ovviamente, è anche intellettualmente onesto dichiarare che non mancano alcune incertezze per quel che concerne i risultati raggiungibili, quelle stesse incertezze ammesse da Glessgen (2003: 65) al termine del già citato contributo: “On peut naturellement se poser la question de savoir si tous ces efforts vont donner des résultats. En d’autres mots: notre vision de l’histoire de la langue changera-t-elle si elle se fonde un jour sur l’interrogation de corpus électroniques constitués suivant les critères exposés?”.

Si rende a questo punto necessario un piccolo passo indietro prima di proseguire con l’analisi dei criteri di composizione. La scelta di optare per la definizione “base” in luogo di “corpus” risponde alle stesse precauzioni che hanno fatto definire come tale, cioè come base, la BFM, e sono state ben illustrate da Serge Heiden in uno dei primi bilanci di sintesi del progetto: “Un corpus de référence est conçu pour fournir une information en profondeur sur une langue. Il

3 Un bilancio della revisione del corpus rispetto allo stato in cui era stato “lasciato in eredità” da Antonij Dees si può leggere in Glessgen – Gouvert 2007.

visé à être suffisamment grand pour représenter toutes les variétés pertinentes de cette langue et son vocabulaire caractéristique, de manière à pouvoir servir de base à des grammaires, des dictionnaires, et d'autres usuels fiables" (Heiden 2004: 103)⁴. È evidente come tale prospettiva non possa, e non voglia, essere quella, ben più limitata, delle iniziative citate. Quando si utilizza il termine corpus in questo contributo lo si fa perciò nel senso anodino di "insieme di testi", e come sinonimo di base.

Tornando all'acronimo e ai criteri costitutivi, in primo luogo la base EXOMARE intende focalizzarsi esclusivamente su una fase dell'evoluzione della lingua, quella dell'*ancien français* inteso in senso strettamente tecnico e non di designazione generica della fase medievale della lingua rispetto all'epoca moderna: lo stadio cioè corrispondente al periodo compreso fra il IX sec. e il 1330 (in opposizione al *moyen français*, 1330-1500)⁵. Con riferimento alle altre iniziative testuali-linguistiche, come orizzonte d'attesa il progetto EXOMARE si vuole quindi innanzitutto collocare in blocco, sul "fronte dizionari", a monte di quello strumento di riferimento imprescindibile che è il DMF (*Dictionnaire du Moyen Français 1330-1500*); ma ovviamente a latere di quella grande impresa che è il DÉAF (*Dictionnaire Étymologique de l'Ancien Français*), il quale "embrasse la période allant de 842 (date des serments d'alliance entre Louis le Germanique et Charles le Chauve contre Lothare I^{er}, transcrits en ancien haut-allemand et en ancien français) jusqu'au milieu du XIV^e siècle (limite vers le moyen français, établie sur la base de faits philologiques et sociologiques)"⁶.

Fissate queste coordinate, sul "fronte basi testuali" quella EXOMARE potrebbe a prima vista apparire come un doppione di una di più ampio respiro e in parte coincidente a livello cronologico, oltre che ovvio punto di riferimento per qualsiasi progetto di questo genere, vale a dire la BFM, i cui testi alimentano anche il DMF. La coincidenza con la BFM è tuttavia più a livello di concezione originale di quella base, cioè la BFM, che di stato attuale della sua evoluzione, così sintetizzato nella sua stessa presentazione: "[*proponendosi come complemento del DMF, la BFM*] a pendant longtemps contenu majoritairement des textes antérieurs à 1320 et des textes littéraires. Elle s'est beaucoup diversifiée ensuite, étendue vers le très ancien français, et enrichie de textes non littéraires (religieux, historiques, scientifiques ou didactiques)"⁷. Quando al contrario, ancora nel 2004, si poteva definire "la langue de la BFM [...] comme étant essentiellement celle des textes littéraires narratifs écrits du français médiéval (IX^e-XVI^e siècle)" (Heiden – Lavrentiev 2004: 103).

Ma soprattutto la base EXOMARE potrebbe sembrare sovrapporsi al NCA, che riduce la forbice cronologica dei testi e dei manoscritti presi in conto a un periodo compreso fra inizio XII e fine XIV sec., benché poi di fatto non includa se non in percentuale piccolissima i manoscritti posteriori al XIII sec.⁸, facendo grosso modo coincidere (come tendenza generale seppur con eccezioni) il periodo ammesso per le opere con il periodo ammesso per i manoscritti; si vedrà

4 Riflessione che si appoggiava a sua volta su Sinclair 1996.

5 Fra le sintesi più recenti di introduzione a questo stadio evolutivo della lingua si ricordano Joly 2018, Hasenohr – Raynaud de Lage 2019.

6 Dalla pagina di presentazione *Le dictionnaire* della versione online del DÉAF (<https://www.hadw-bw.de/fr/recherche/centre-de-recherche/dictionnaire-etymologique-de-lancien-francais-deaf/le-dictionnaire>).

7 Dalla pagina di presentazione della BFM (<http://bfm.ens-lyon.fr/spip.php?article339>).

8 "L'on a donc essentiellement affaire à des manuscrits du XIII^e siècle reproduisant toutefois des œuvres de la seconde moitié du XII^e et de la première moitié du XIII^e" (Glessgen – Gouvert 2007: 67).

di seguito che, al contrario, la continuità e varietà cronologica (fino a tutto il XIV sec. e con un ultimo testimone del 1465) della tradizione del *Roman de Florimont*, l'opera individuata per inaugurare la base e testare la metodologia EXOMARE, è stata proprio una delle ragioni della sua scelta in quanto funzionale all'idea che si vuole proporre di una base linguistica che permetta di seguire l'evoluzione delle forme nei “testi del Testo” di un'Opera.

1.2 Parola chiave: delimitare (gli usi della lingua)⁹

Rispetto alle basi testuali-linguistiche menzionate, dal punto di vista cronologico quella EXOMARE si focalizza su un secolo circa – fra il 1150 ca. e il 1250 ca. – e si circoscrive a un ambito esclusivo della testualità, quella letteraria, la quale in questo arco cronologico così delimitato inventa un “genere” e una “forma testuale” (il *roman in octosyllabe*) che segnano lo smarcamento dalla tematica epica e religiosa. “Letterarietà” e “genere”, e sotteso l'arco cronologico, sono quindi i primi criteri di costituzione della base (e quindi di selezione dei testi da includere) da giustificare.

Per poter rendere chiaro il discorso, pur con tutte le difficoltà insite nel cercare di proporre in una forma lineare sequenziale una serie di riflessioni che in realtà si intersecano e si richiamano l'un l'altra, si deve però prima terminare l'analisi dell'acronimo del progetto, appuntando l'attenzione sull'ultimo elemento: *through MANuscript REadings*. Si tratta anche dell'elemento che può distinguere in modo più interessante la base EXOMARE dalle altre citate. Innanzitutto, fra i “deux approches aparamment opposées de la source des données sur lesquelles il convient de baser le corpus” (Heiden – Lavrentiev 2004: 100), cioè edizioni critiche ricostruttive o edizioni del testimone, EXOMARE opta esclusivamente per le seconde: è il primo senso dell'espressione *through manuscript readings*. Si tratta di una scelta minoritaria, anche per le difficoltà – leggi tempo necessario – di realizzazione, ma della quale, con il suo ricondurre il materiale linguistico alla forma documentaria, reale e non filtrata dall'elaborazione editoriale (qual che sia l'impostazione editoriale), è anche intuitivo percepire tutti i benefici. Se le conseguenze di un utilizzo delle edizioni critiche possono essere più arginabili nel caso di edizioni realizzate in decenni vicini a noi, secondo un metodo scientifico rigorosamente stabilito e soprattutto esplicito e verificabile, le edizioni ottocentesche e dei primi decenni del secolo seguono talvolta principi non più accettabili, come, fra i vari esempi possibili, l'edizione ri-picardizzata del *Jeu de la feuille* di Adam de la Halle a opera di Ernest Langlois (1923), o lo sforzo di Léopold Constans, nel 1890, di ricostruire la “lingua originale” del *Roman de Thèbes*, creando una sorta di dialetto dell'Ovest razionalizzato a tavolino¹⁰. La conseguenza principale, comunque, della prevalenza dell'approccio orientato all'edizione critica – scelta d'altro canto obbligata nel momento in cui si rendeva necessario costruire la panoramica generale dell'*ancien français* e si aveva bisogno di una grande mole di dati immediatamente fruibile – è il fatto che “Toute la lexicologie aussi bien que la lexicographie de l'ancien français repose sur les formes plus ou moins normalisées des éditions critiques” (Glessgen 2003: 58).

9 Sul grado di rappresentatività dei corpora cfr. le riflessioni di Habert 2000 e Prévost 2008.

10 Su questo problema si veda il paragrafo *Fiabilité des éditions* in Buridant (2019: 201-204), e le considerazioni di Glessgen – Gouvert (2007) che, proprio sulla base di valutazioni di questo tipo, hanno finito per escludere alcuni testi, utilizzati da Dees, dal NCA.

Proprio l'esistenza di grandi basi già in essere, che hanno sostanzialmente permesso di delineare il quadro generale, e non solo, dell'*ancien* (e del *moyen*) *français*, consente di inoltrarsi ora nella dimensione del manoscritto individuale, in una prospettiva di ampliamento, forse in qualche caso (sicuramente minoritario) correzione, delle conoscenze sulla lingua acquisite attraverso le edizioni critiche. La generalizzazione e standardizzazione dell'uso dello strumento informatico facilita questo tipo di approccio, anzi, è forse il presupposto principale che lo rende sostanzialmente possibile. Se i grandi corpora con storie ormai trentennali (oltre ai già citati NCA e BFM si ricordano anche l'ANH [*Anglo-Norman On-line Hub*]¹¹, il TFA [*Textes de Français Ancien*]¹² dell'Università di Ottawa, o la piattaforma PALM [*Plateforme d'Analyse Linguistique Médiévale*]¹³) sono tutti fondati su edizioni critiche, i progetti successivi, complice anche la loro nascita in un'epoca informatica non più agli albori e in fase sperimentale ma ormai decisamente adulta, tendono sempre più spesso verso la dimensione documentaria: è il caso, fra gli ultimi, dell'OGR (*Old Gallo-Romance Corpus*), su cui si ritornerà; e, a mo' di ulteriore esempio, si citano il progetto *Khartès*, dedicato alle carte della Wallonia in lingua vallone anteriori al 1300¹⁴, e tutto quanto ruota intorno all'iniziativa dei *Plus anciens documents linguistiques de la France*. Ma, sul piano più strettamente letterario, è il caso soprattutto dell'esperienza BFM-Mss, inaugurata a partire dal prototipo dell'edizione interattiva della *Queste de saint Graal* (v. oltre): essa comprende attualmente, oltre all'edizione integrale della *Queste*, 24 estratti da testi manoscritti e 4 da incunaboli.

Si tratta di una tendenza che va di pari passo con la valorizzazione, per le singole opere, di edizioni a doppia visualizzazione, diplomatica e/o semidiplomatica vs. interpretativa¹⁵, rese possibili grazie al ricorso allo strumento digitale: si consideri, fra gli ultimi (in mero ordine cronologico) esempi, l'edizione dell'*Histoire ancienne* nell'ambito del progetto TVOF (*The Values of French*)¹⁶, concretizzatasi nell'edizione, visualizzabile in modalità semidiplomatica e interpretativa, dei manoscritti London, BL, Royal 20 D I e Paris, BnF, fr. 20125 (scelti come rappresentanti rispettivi delle due versioni dell'opera). Alle spalle di iniziative di questo tipo c'è ovviamente un padre nobile (sentito come tale da un duplice punto di vista nel caso della base EXOMARE, come si espliciterà), vale a dire il pionieristico *Charrette Project* di Karl Uitti¹⁷ che introduceva almeno

11 <https://anglo-norman.net>.

12 <https://francaisancien.net/base-tfa.html>.

13 <http://palm.huma-num.fr/PALM/>, si veda inoltre Mairey – Aouin (2020).

14 Per un'introduzione a questo progetto, si veda Mazziotta (2004).

15 Per chiarire la terminologia e di conseguenza il prodotto editoriale a cui si fa riferimento, sempre partendo dal presupposto che ci si sta muovendo nell'ambito di "edizioni di testimoni" e non di edizioni critiche ricostruttive: l'edizione diplomatica *stricto sensu* prevede la riproduzione fedele del documento incluse le varianti delle lettere e i segni abbreviativi (non sciolti), oltre ovviamente alla segmentazione originale e alla disposizione del testo nello spazio di scrittura; l'edizione semidiplomatica (la nozione più fluida fra le tre e con alcuni margini di discrezionalità) aumenta il grado di leggibilità almeno normalizzando la separazione delle parole, sciogliendo le abbreviazioni, con indicazione delle lettere restituite, e riducendo la registrazione degli interventi scribali; l'edizione interpretativa offre il massimo della leggibilità all'occhio moderno dal momento che, oltre a superare la *scriptio* scribale e sciogliere (in modo silente) le abbreviazioni, introduce la punteggiatura e i segni diacritici e normalizza l'uso delle maiuscole, svincolando del tutto il testo dalla riproduzione mimetica della *mise en page* manoscritta (nel caso della semidiplomatica è una scelta eventuale).

16 *The Values of French Language and Literature in the European Middle Ages*. ERC Advanced Grant (n. 620726) at King's College London (<https://tvof.ac.uk>).

17 <https://www.princeton.edu/~lancelot/ss/index.shtml>, si vedano inoltre Uitti 1997 e Pignatelli – Robinson 2002.

tre elementi di interesse in campo filologico: 1. la trascrizione diplomatica integrale dei testimoni manoscritti 2. codificata informaticamente attraverso il linguaggio SGML, con tutto quello che ne conseguiva in termini di markup semantico, visualizzazione digitale, analisi testuale computazionale e loro ricadute sulla ricostruzione filologica; 3. la trascrizione integrale dell'intera tradizione manoscritta di una singola opera, nel caso specifico il *Chevalier de la charrette* di Chrétien de Troyes. A questo progetto va affiancato l'altro prodotto esemplare in questo ambito, vale a dire la già citata edizione interattiva della *Queste de saint Graal*, secondo il ms. Lyon, BM, p. a. 77, a opera di Christiane Marchello-Nizia e Alexei Lavrentiev¹⁸. Per un estratto di circa 10.000 parole la *Queste* è stata anche annotata sintatticamente all'interno del progetto SCRMF (*Syntactic Reference Corpus of Medieval French*)¹⁹.

Sempre sotto il segno di Chrétien de Troyes va ovviamente citato l'ultimo progetto la cui impostazione è confluita come suggestione in quella della base EXOMARE: il DÉCT (*Dictionnaire Électronique de Chrétien de Troyes*)²⁰. Si ricorda che il DÉCT – che si configura quindi, e lo si sottolinea in relazione all'impostazione EXOMARE, al tempo stesso come corpus e dizionario di un uso specifico della lingua, quello cioè di un singolo autore nel complesso della sua opera – è costruito sulle trascrizioni semidiplomatiche dei cinque romanzi di Chrétien nella forma offerta dalla copia Guiot, Paris, BnF, fr. 794, considerata come “le meilleur manuscrit et le seul à conserver le texte intégral des cinq romans”. Allo stato attuale il DÉCT non è ancora evoluto (almeno pubblicamente) verso la seconda tappa annunciata: “La seconde étape du DÉCT (DÉCT2) se caractérisera par un enrichissement considérable par rapport à la première version; il s'agit essentiellement de l'ajout: des mots grammaticaux [...]; des variantes des autres manuscrits (variations lexicales, variations également dans la collocation des termes [...]); des axes de synonymie et d'antonymie”²¹. I dati non sono ancora confluiti nel DÉCT ma il sito del LFA (*Laboratoire de Français Ancien*)²², da cui muove il DÉCT, nella collezione TEL (*Textes en Liberté*)²³ ospita la trascrizione semidiplomatica integrale del *Chevalier au lyon*, realizzata da Kajsa Meyer, preceduta da quella di 7 mss. dello stesso testo a opera di Pierre Kunstmann.

Sulla scorta di tutte queste suggestioni, ritornando all'espressione *through manuscript readings*, la base EXOMARE intende fondarsi su questi due pilastri: 1. i dati linguistici deriveranno solo da trascrizioni dei manoscritti, ponendosi quindi nel solco della BFM-Mss e delle iniziative che ruotano intorno al nome di Chrétien, 2. per ciascuna opera scelta verranno trascritti tutti i testimoni sul modello *Charrette* ma, a differenza di quest'ultimo, orientato solo verso la componente editoriale, tali trascrizioni saranno lemmatizzate (come il DÉCT) e annotate linguisticamente (a differenza del DÉCT).

Soffermandosi su questo secondo punto, cioè la pluritestimonialità, nella prospettiva che si vuole proporre ciascun testo dell'Opera è concepito innanzitutto come un testo a sé, come un documento linguistico connotato cronologicamente e spazialmente, fotografia di uno stato della

18 https://txm-bfm.huma-num.fr/txm/catalog/qgraal_cm.

19 <http://scrmf.org>.

20 <http://zeus.atilf.fr/dect/>.

21 Dalla pagina del sito del DÉCT *Accueil/Un dictionnaire en deux étapes*.

22 <https://www.francaisancien.net>.

23 <https://www.francaisancien.net/tel.html>.

lingua. Al tempo stesso però i singoli documenti indipendenti sono strettamente connessi fra loro dal fatto di condividere in percentuale altissima lo stesso insieme di lemmi, coincidenti in percentuale altissima dal punto di vista della categoria grammaticale e della declinazione morfologica, e articolati in percentuale altissima nella medesima strutturazione sintattica: in una percentuale tale da far riconoscere cioè una molteplicità di testi come espressione della stessa opera, al di là della forma grafico-fonetica e morfologica assunta dai lemmi (o variazione linguistica) e dalla percentuale di discordanza testuale (o variante filologica).

Il sub-corpus così creato (“sub-corpus d’opera” rispetto alla macrostruttura della base EXOMARE) permetterà quindi di far convergere, o meglio ci si auspica che possa far convergere, in un unico contesto l’analisi della evoluzione diacronica e quella della variazione diatopica: un dato insieme di forme di lemmi – limitato e quindi analizzabile nell’estremo dettaglio, compensando in tal modo sul piano qualitativo quanto si potrebbe perdere sul piano quantitativo – viene seguito nello spazio e nel tempo attraverso le istantanee rappresentate da ciascun documento manoscritto. Nel caso di partenza che si illustrerà, il *Roman de Florimont*, la presenza di un manoscritto del 1465, ultima fotografia a oggi nota della diffusione di un’opera concepita indicativamente intorno al 1188, si rivela in questo senso particolarmente interessante perché permette di verificare cosa resta e cosa è diventato incomprensibile del sistema lessicale-morfologico-sintattico di un testo redatto circa tre secoli prima dell’epoca della copia.

“Dato documentario” e “pluralità testimoniale” dell’opera sono quindi in conclusione i due significati sottesi all’espressione *through manuscript readings*.

Tornando alla questione di letterarietà e genere come criteri costitutivi della base, per quel che riguarda il primo punto, cioè la rigida letterarietà, la base EXOMARE si distingue da questo punto di vista dalla BFM nella sua versione attuale, cioè ampliata: nata infatti con un orientamento prettamente letterario, quest’ultima si è, come visto, non solo aperta cronologicamente tanto verso il *très ancien français* quanto verso il *moyen français*, ma ha anche inglobato, accanto ai testi non-documentari²⁴, delle scritture archivistiche, come, a titolo di esempio, il *Cérimonial d’une épreuve judiciaire*, (inizio XII sec., ms. BnF lat. 2403) o le *Chartes de l’Aube* e le *Chartes du Hainaut* (1230-1271) judiciaire (inizio XII sec., ms. BnF lat. 2403) o le *Chartes de l’Aube* e le *Chartes du Hainaut* (1230-1271).

L’impostazione EXOMARE viene a coincidere invece, anche se casualmente, con la fisionomia attuale del NCA: la perdita del database di Dees contenente i dati linguistici delle *chartes* ha fatto sì che il NCA si identificasse esclusivamente con una fisionomia letteraria. Le riflessioni portate avanti successivamente da Martin Glessgen a proposito dei *lieux d’écritures* (che sovrappongono al dato linguistico scripturale una dimensione socio-culturale), delle *scriptae* e della metodologia di definizione dei profili scriptologici di un testimone letterario²⁵, hanno però sistematizzato un metodo, facendo dei dati ricavati dalle *chartes* un parametro di confronto per i dati

24 Nel caso della BFM utilizzo l’espressione “non-documentario” in quanto, nell’ecosistema BFM, la definizione *littéraire* ha un significato specifico e limitato rispetto all’utilizzo che se ne fa nel contesto EXOMARE nel senso di categoria testuale opposta a quella documentaria.

25 Per un’introduzione alla tematica della scriptologia si veda Glessgen 2012. Videsott 2019: 181, appoggiandosi su Glessgen, ben sintetizza la definizione di *scripta* (“la forme langagière des manuscrits”) e scriptologia (“l’étude de la variation linguistique diasystématique de textes produits dans des conditions de normalisation et de standardisation linguistiques encore peu affermées”). Sul problema della standardizzazione si veda, per una breve introduzione al problema, Lodge 2008 e soprattutto Pfister 1993.

letterari, e conducendo anche alla rettifica di alcune conclusioni di Dees circa l'assegnazione di alcuni testimoni alle *scriptae* dell'Est. Si tratta dell'impostazione che viene pienamente condivisa dal progetto EXOMARE.

Tale scelta parte dalla considerazione di una problematica che affligge in modo endemico i testi letterari dell'*ancien français*: quella cioè del *décalage* sistematico tra composizione e attestazione, sia sul piano cronologico che sul piano linguistico, problematica ovviamente ben nota e alla radice della predilezione degli studi scriptologici per il dato documentario²⁶.

1.3 Quale *ancien français*? Composizione e attestazione

La scrittura letteraria della fase dell'*ancien français* che si prende qui in esame presenta una sistematica discrepanza (che non è esclusiva in assoluto di questo periodo ma è pur sempre in questo periodo sistematica): quella tra la data di composizione dell'opera e la data della prima attestazione manoscritta²⁷ giunta fino a noi, discrepanza che si aggira intorno a un cinquantennio nella maggioranza dei casi, e che in alcuni può elevarsi fino a un secolo pieno e oltre²⁸. Ad esempio:

Roman de Thèbes, redazione (red.) 1150 ca., attestazione (att.) secondo terzo XIII sec.;

Roman de Brut di Wace, red. 1155, att. inizio XIII sec.;

Roman d'Enéas, red. 1150 ca., att. 1200 ca.;

Floire et Blancheflor, red. 1150-1160 ca., att. 1289;

Pyramus et Thisbé, red. 1155-1160, att. 1325 ca.;

Tristan di Thomas, red. 1172-1176, att. fine XII sec.;

Tristan di Beroul, red. 1165-1200, att. seconda metà XIII sec.;

Eracle di Gautier d'Arras, red. ca. 1176-1184, att. fine XIII sec²⁹.

26 Videsott (2019: 182): "Il est vrai que les études scriptologique sont plus sûres si elles sont menées sur la base de manuscrits datés et localisés et dont le contexte de genèse est connu, ce qui explique d'ailleurs la préférence habituellement accordée aux textes documentaires et originaux".

27 Sul rapporto fra opera e testi nella letteratura medievale è d'obbligo far riferimento a Zumthor 1972.

28 Si veda Buridant 2019: 197-198.

29 Non mi sembra ozioso riportare il quadro tracciato da Glessgen (2007: 54-55) per il NCA e che ben sintetizza la casistica e le percentuali relative: "Le décalage en question existe dans l'immense majorité des cas; très rarement, nous sommes en face d'un manuscrit pour lequel nous ne pouvons pas supposer ou identifier un antécédant défini. Dans le cas idéal, il s'agit de manuscrits autographes, ce qui reste néanmoins une exception. [...] Plus fréquents sont les manuscrits qui, sans être autographes, sont contemporains de l'écriture de l'oeuvre [...]. Dans des nombreux cas (66 au moins), Dees a intégré des manuscrits très proches de la genèse de l'oeuvre [...]. Dans tous ces cas, textes et manuscrits peuvent être considérés comme isochrones: un quart du Corpus [...]. Les trois quarts restants témoignent au contraire des décalages habituels pour les genres en question; la moitié des textes comporte une distance chronologique d'environ trente à quatre-vingt ans. [...] Il reste un dernier quart (70 textes au moins) pour lequel l'écart dépasse légèrement cent ans. [...] Les décalages plus importants restent exceptionnels, puisque Dees n'a intégré qu'assez peu de manuscrits postérieurs au XIII^e siècle: la trentaine de codex du XIV^e siècle figurant dans le Corpus s'inscrivent pour la plupart dans un écart d'un siècle".

Ma anche nei casi in cui questa forbice cronologica si attesta intorno al valore più basso, cioè un cinquantennio circa, non mancano altre forme di complessità nel rapporto tra l'opera e il testo che la veicola: per fare un esempio, il più antico manoscritto del *Roman d'Enéas*, databile come opera intorno al 1155-1160, è sì un manoscritto della fine del XII-inizio del XIII sec. ma, a compensazione di questa precocità cronologica (rispetto ad altri casi), si tratta di un manoscritto copiato in Italia (parliamo del ms. Firenze, Biblioteca Medicea Laurenziana, Plut. XLI 44). Siamo cioè di fronte alla complessa problematica sintetizzabile nell'opposizione lingua dell'autore vs. lingua dello scriba, della "question séculaire du rapport entre l'auteur et ses copistes" (Leonardi 2019: 173).

Nel momento in cui si costruisce un corpus, per queste opere/testi non è quindi possibile adottare il criterio che presiede ad esempio alla costituzione del recente OGR (*Old Gallo-Romance Corpus*) (Rainsford 2022), che raccoglie tutti i testi gallo-romanzi copiatii prima del 1130. Il "primat du manuscrit" sostenuto da questa iniziativa non si traduce solo nella scelta di un sistema di trascrizione dei testi fedele al manoscritto – opzione condivisa dalla base EXOMARE e in linea (per entrambi) con quello che è il punto di riferimento per le trascrizioni di questo tipo, cioè la BFM-Mss³⁰. Ma, nel caso dell'OGR, questo primato si concretizza anche "dans le choix d'inclure seulement les textes préservés dans un manuscrit copié avant 1130 sans consideration de la date de composition du texte lui-même". Il che conduce all'individuazione di 22 testi contenuti in 16 manoscritti che rispondono a questi criteri. Il numero di testi e manoscritti calerebbe drasticamente se applicassimo lo stesso criterio, fissare cioè un limite cronologico interno alla cronologia dello stadio stesso della lingua preso in esame, alla produzione *romanesque* della seconda metà del XII sec. Né i pochi testi e manoscritti che eventualmente rispondessero positivamente a un criterio analogo proiettato sui nostri riferimenti cronologici avrebbero la stessa significatività che hanno i relativamente pochi documenti dell'OGR, testimoni di una fase sperimentale che registra "des traits linguistiques qui disparaissent avant l'essor des traditions textuelles" (Rainsford 2022: 1). Se consideriamo poi le sole attestazioni letterarie ante 1130, come la *Sequense de Sainte Eulalie*, la *Passion de Clermont*, le *Boëci* ecc., constatiamo che la forbice tra data di composizione e data dell'attestazione (tendenzialmente un testimone unico) è molto ridotta rispetto alla prassi delle opere qui in esame. Questa problematica, cioè lo scarto cronologico, è quella che ha spinto a fissare, tra i criteri di composizione della base EXOMARE, l'esclusione dei testi documentari per focalizzarsi solo sull'ambito della testualità letteraria (letteraria in senso lato, da intendersi come opposizione a documentaria). Potremmo per sommi capi definire il documento come un atto testuale che deve in primo luogo rispondere a una richiesta di autenticità e che deve svolgere una funzione pragmatica rispetto al reale. Atto scritturale e documento scrittorio collasano per definizione l'uno sull'altro fondendosi nel concetto di "originale": in questo caso non può quindi esserci distanza cronologica fra composizione e attestazione ed è quindi evidente come il livello di rappresentatività della testualità documentaria rispetto allo stato della lingua (lo stato della lingua ovviamente del diasistema di riferimento, superata ormai l'illusione, coltivata inizialmente dalla scriptologia, che la lingua documentaria sia riflesso diretto della lingua della comunicazione) sia differente da quello della testualità letteraria. In altre parole, almeno idealmente (possono esistere è vero copie dei documenti, ma la copia del documento risponde

30 Lavrentiev 2008.

a una logica profondamente diversa dalla copia diffusoria dell'opera letteraria, che della copia vive, mentre il documento vive dell'originale³¹), Opera e Testo nel caso del documento archivistico coincidono e tendono verso il legame unico e irripetibile³².

Come accennato, le nostre opere trascinano poi con sé un altro elemento di complessità che costituisce un argomento a favore della valorizzazione della molteplicità dei manoscritti e della proposta di introdurre in modo sistematico come sottofondo dell'analisi le nozioni di “coordinate d'attestazione” e “coordinate d'origine (dell'opera/del testo)”. Il XII secolo, soprattutto la seconda metà, assiste allo sviluppo delle *scriptae* regionali, cosicché, come evocato, il divario fra origine e attestazione a livello cronologico si incrocia e si interseca con il divario a livello di *scripta* in una catena di variazione in crescendo: *scripta* del copista differente dalla *scripta* dell'*exemplar* da cui copia, differente a sua volta dalla *scripta* dell'autore dell'opera. Torniamo come esempio alla più antica testimonianza dell'*Enéas*, che risulta copiata in Italia, entrando in tal modo nell'orbita dei testi francesi copiati nella Penisola, caso estremo di *scripta* francofona extra-territoriale. Di fronte a queste tradizioni manoscritte, ciò con cui abbiamo a che fare è dunque una sequenza di attestazioni, ovvero testi manoscritti, cronologicamente e geograficamente definite (le coordinate di attestazione), di un'opera il cui primo “testo di espressione” ha/può avere una collocazione crono-spaziale, ovvero luogo e data di composizione (le coordinate di origine dell'opera), differente/molto differente rispetto al primo “testo di attestazione”. Il sistema della copia precipuo del mondo pre-stampa fa sì che si arrivi alla situazione che vede la prima attestazione di un'opera linguisticamente legata all'*ancien français* in piena epoca di evoluzione della lingua allo stadio del *moyen français*: mi sembra una situazione che marca in modo efficace e chiaro la differenza fra la testualità letteraria e quella documentaria, e che ricontestualizza la valenza delle forme che si inseriscono, pur di diritto, in una base dedicata all'*ancien français* in quanto relative a un'opera composta a una determinata altezza cronologica. In quest'ottica, i corpora linguistici legati alla testualità documentaria dovrebbero, a mio avviso, essere separati da quelli legati alla testualità non documentaria; con la loro tendenza alla coincidenza fra dato linguistico prodotto e dato linguistico attestato, essi rappresentano tuttavia un fondamentale strumento di controllo e verifica per i corpora non documentari, instaurando tra le due realtà del *paysage de l'écrit* un rapporto di fruttuosa complementarietà.

Per tornare alle coordinate di attestazione, esse possono essere, nel momento in cui si intraprende il lavoro, note, ipotizzate o non ancora ipotizzate: l'analisi completa del testo, auspicabilmente, fornirà conferme nel secondo caso e permetterà di formulare ipotesi nel terzo; la potenzialità *in progress* implicita nell'uso dello strumento informatico garantisce ovviamente, in qualsiasi momento, la possibilità di integrare e modificare i dati. Mentre in questo caso, quello delle coordinate di attestazione, siamo di fronte a un'unica informazione che deve essere associata automaticamente a tutte le forme del singolo testo preso in esame, più complesso – e per il momento un *desideratum* che sarà soddisfatto solo in minima parte nello spazio di vita del progetto

31 Sulle differenti ripercussioni del processo di copia nell'ambito documentario in opposizione a quello letterario si vedano, per un'introduzione, Coluccia 2000 e Greuss 2007.

32 “[*Nel caso delle carte*] leurs caractéristiques externes (mise en page, calligraphie, forme et qualité du parchemin, mode de scellement) proviennent du même contexte de genèse que leurs caractéristiques internes (l'état linguistique et le contenu textuel). Le processus de copie, habituel pour les textes littéraires et les textes d'un savoir spécialisé, dissout ce lien immédiat et anéantit toutes les caractéristiques externes. Celles-ci répondent alors au moment de la copie et non plus à celui de la première rédaction du texte” (Glessgen 2008: 414-415).

(la cui fine è prevista per dicembre 2025)³³ – è il discorso rispetto alle coordinate di origine. Per riprendere la questione, la nozione di coordinate di origine è come si è detto duplice. Avremo innanzitutto le “coordinate di origine dell’opera”, corrispondenti al luogo e alla data di composizione dell’opera e di sua espressione in un testo; anche in questo caso, come per le coordinate di attestazione, si tratta di un dato unico e omogeneo per tutte le forme di qualsivoglia testo che veicoli quell’opera, ed è quindi ovviamente assegnabile automaticamente a ciascun elemento.

Più complessa, più interessante, e di più difficile realizzazione è invece l’esplicitazione delle “coordinate di origine delle forme”, le singole forme, di un determinato testo manoscritto, cioè uno dei vari testi che attestano il Testo dell’Opera: si tratta infatti dello spazio di annotazione in cui emergeranno i dati relativi alle *scriptae* letterarie e ai tratti dialettali. Se in una certa percentuale questo dato coincide con quello delle coordinate di attestazione³⁴, sussiste poi lo spazio per forme in contrasto con il tessuto linguistico preponderante, sia in termini di tratti scripturali altri sia in termini di arcaismi³⁵. Il grado di dettaglio di un’analisi di questo tipo è evidente, così come la conseguente complessità di realizzazione: credo però che sia anche altrettanto evidente il potenziale interesse di una classificazione così minuziosa delle forme. Ci si orienta cioè verso le raccomandazioni di Glessgen (Glessgen – Vachon 2013: 193): “Une telle réflexion porte alors sur les éléments de graphématique et de morphologie qui montrent la plus forte intervention des copistes; ces domaines permettent en même temps de tirer le plus grand nombre de conclusions linguistiques sur les caractéristiques des manuscrits individuels”. Risalendo dall’individuale al sopra-individuale, la definizione dei profili dei manoscritti individuali può evolvere verso l’individuazione di un *lieu d’écriture*, ricollocando l’individuo in un tessuto socio-scritturale.

Anche in questo caso non si tratta di una novità concettuale in assoluto: questa doppia parametrizzazione, opera/manoscritto, è presa in conto anche dal NCA. Quel che cambia, nel caso di EXOMARE, è: 1. il grado di dettaglio dell’annotazione, dal momento che si punta alle forme in modo sistematico, 2. la tipologia di edizione dei testi su cui si basa il corpus, e 3. l’impostazione pluritestimoniale.

33 Tale analisi verrà limitata a due manoscritti, quello di Vienna e quello di Monza, fra loro legati a livello stemmatico. Nel primo caso, l’analisi linguistica e scriptologica completa, a fronte di una datazione esplicita (1465), dovrebbe permettere di formulare un’ipotesi circa la sua origine: si tratta infatti del manoscritto più negletto, del tutto sconosciuto a Hilka. Nel secondo caso, l’analisi dovrebbe invece permettere, si spera, di sciogliere dubbi: a lungo considerato di produzione italiana, gli ultimi spogli hanno avanzato invece l’ipotesi di un’origine francese meridionale, che è peraltro la prima ipotesi accennata anche per il ms. di Vienna. Confermare questo quadro, in considerazione anche della vicinanza stemmatica, significherebbe rafforzare l’immagine di un ben affermato ramo meridionale nella tradizione del *Florimont*.

34 O in senso inverso, in un circolo virtuoso, l’individuazione di questi tratti permetterà di avanzare un’ipotesi per la collocazione crono-spaziale del manoscritto.

35 Nel NCA le indicazioni diatopiche di Dees introducevano l’interessante e utile concetto di “coefficient régional” cioè “le degré de cohérence du texte avec les caractéristiques graphématiques du lieu” (Glessgen 2007: 57), interesse del concetto che mi sembra sussistere al netto delle conclusioni talvolta contestate in materia di localizzazioni effettive realizzate sulla base del suo metodo geolinguistico.

1.4 Le intersezioni fra storia degli usi linguistici e storia delle forme letterarie

Per concludere sui criteri definitori della base EXOMARE dal punto di vista della selezione dei materiali, la terza scelta operata dopo quella dell'arco cronologico e della testualità esclusivamente letteraria è quella dell'individuazione dei testi sulla scorta della nozione di genere letterario, che si riconnette, con un legame che si scinde solo per esigenze espositive, alla questione cronologica. Si tratta ovviamente di una classificazione, quella dei testi medievali per generi, che si scontra con una serie di difficoltà: la coscienza dei generi interna alla scrittura medievale stessa, il rischio dell'anacronismo nella sovrapposizione di strutture moderne su realtà altre³⁶, la costruzione di una teoria e di una classificazione sulla base di materiali sostanzialmente lacunosi (quanti testi e, potenzialmente, quanti generi abbiamo perduto?). Non è intenzione del progetto EXOMARE inoltrarsi in una discussione teorica rispetto a questa problematica così spinosa, e ci si limiterà quindi ad alcune considerazioni, forse semplificate, ma che appaiono funzionali alla costituzione della base stessa.

Intorno alla BFM è stata condotta in questo senso una riflessione molto approfondita sfociata nella messa a punto di un sistema di classificazione dei testi a un triplice livello³⁷, con l'identificazione: della forma (verso, prosa, mista, glossa); del *domaine*, legato alla "funzione" dell'opera (religioso, letterario, giuridico, storico, didattico); del genere, legato alla "struttura" (agiografia, epica, romanzo, lirica, trattato, giuramento, cronaca, memoria, enciclopedia ecc.). In questo contesto, la parola *littéraire* corrisponde solo e unicamente alla funzione testuale del 'divertire'. Il sistema regge molto bene sui 219 testi che compongono la BFM, ma alcuni casi possono essere ridiscussi: se il *Roman d'Enéas* viene descritto attraverso la funzionalità letteraria e il genere romanzo, la classificazione del *Brut* di Wace come appartenente all'ambito storico e al genere cronaca apre lo spazio alla riflessione. Come detto, il progetto EXOMARE attribuisce un significato differente al termine 'letterario', adottando, come Sophie Prévost (2008: 9 n. 9), "une terminologie très simplifiée, en opposant genres littéraires (roman, nouvelle, chanson de geste, poésie, théâtre...) et genres non-littéraires (texte juridique, charte, coutumier...)". Che è anche sostanzialmente l'opposizione sulla quale si fondava la classificazione dei testi nella prima fase della BFM, quando si parlava di *textes littéraires* in opposizione a *textes utilitaires*³⁸. Quanto alla nomenclatura dei generi e all'assegnazione di ciascun testo a uno di essi, ci si rifà sostanzialmente alle classificazioni cristallizzate dalla tradizione filologica. Per lo meno in partenza.

Il legame fra genere testuale e lingua emerge con sempre maggiore forza dagli studi nell'ambito della fraseologia, da cui la scelta qui operata di "ragionare per generi". In un movimento di andata e ritorno in cui è difficile separare i due momenti, il fatto linguistico determina il genere e il genere determina il fatto linguistico. Nelle parole di Julie Sorba, che ha dedicato vari studi al tema della fraseologia in connessione con la letteratura francese medievale, in collaborazione con Corinne Denoyelle³⁹, si tratta da una parte di "montrer comment les unités phraséologiques

36 La problematica è ben sintetizzata in Guiot – Lavrentiev (2009: 8-9): "Trois taxonomies au moins sont possibles: – les genres tels qu'on pense que les médiévaux les voyaient; – les genres médiévaux tels qu'on les voit aujourd'hui (du point de vue de la tradition philologique); – les genres tels qu'on les définit dans la société actuelle".

37 Guiot – Lavrentiev 2009: 8-11.

38 Heiden – Lavrentiev 2004: 102.

39 Sorba – Denoyelle 2020, 2021, 2023; Denoyelle 2014. Cfr. anche Malrieux – Rastier 2001 e, sul tema del motivo

caractérisent les genres textuels et permettent de les distinguer les uns des autres. [...] Partant de l'hypothèse que la langue littéraire se caractérise par la surreprésentation statistiquement significative de phraséologismes⁴⁰. E dall'altra, mutando prospettiva, di verificare l'ipotesi che "l'appartenance d'un texte à un genre conditionne les variations lexicales, morphosyntaxiques et discursives qui s'y trouvent en comparaison des autres genres"⁴¹. Come ben esposto anche in precedenza da Sophie Prévost (2008: 51): "On peut envisager d'élaborer une typologie "inductive" des textes (c'est à dire fondée sur des traits linguistiques), avec la perspective d'établir une corrélation ultérieure entre critères "externes" (situationnels) et "internes" (linguistiques)". Parole a cui fa eco la "question de recherche" posta da Julie Sorba: "l'entrée phraséologique pourrait-elle nous permettre de contribuer efficacement à caractériser les genres médiévaux?". Il fatto di caricarsi dell'onere di procedere alla lemmatizzazione e annotazione linguistica di un testo, vale a dire i dati a partire dai quali si costruisce un'analisi fraseologica, va in prospettiva anche in questa direzione di studi, facendo così di un corpus o base che dir si voglia il punto di incontro fra la linguistica e la filologia intesa in quanto storia delle forme letterarie.

Se in un'ottica futura il corpus EXOMARE vorrebbe articolarsi in sub-corpora di genere, aprendosi in particolare alla *chanson de geste* e alla scrittura storiografica – sempre rimanendo tuttavia nell'ambito della letterarietà –, per porre una prima pietra dell'edificio il progetto si interessa al momento al solo genere *roman*, considerando appunto l'invenzione del romanzo cortese come un cambio di passo rispetto all'oggetto della letterarietà di tipo finzionale (con l'irrompere dell'avventura e del meraviglioso) e al pubblico interessato (laici, nobili e borghesi). A sua volta, tuttavia, il mondo del *roman* è tutt'altro che un'entità monolitica, né è ormai più accettabile in blocco, senza sfumature, la classificazione tematica, per certi versi semplificata e a compartimenti stagni, cristallizzata da Jean Bodel nella celebre tripartizione delle *matières* – *de France, de Rome, de Bretagne* –, con riferimento nel primo caso alla *chanson de geste*, nel secondo ai *romans d'Antiquité* e nel terzo al romanzo arturiano. Solo per limitarsi agli ultimi due elementi della lista, di nostro più diretto interesse in questo momento, è ormai d'uso corrente parlare di una serie di *romans* definiti da Marie-Madeleine Castellani come *romans antique* della seconda generazione (*Athis et Procelias, Hypomedon, Partonopeus, Florimont* ecc.), i quali proiettano l'elemento *féérique* e dell'*aventure*, proprio della materia bretone, su più o meno accentuati scenari anticheggianti e integrano nell'ambientazione l'elemento orientale. È il caso esattamente del testo da cui si è deciso di partire per sperimentare la costruzione del corpus EXOMARE, il *Roman de Florimont* di Aymon de Varennes, databile intorno al 1188 secondo la data esplicitata dai colophon. Il corpus-*Florimont* rappresenta quindi il primo tassello della base EXOMARE, che ci si augura di poter incrementare nel corso degli anni puntando immediatamente a un secondo corpus d'opera dedicato al *Roman d'Enéas*, il cui lessico e sintassi amorosa, in particolare, si propagano attraverso la scrittura di contemporanei ed epigoni.

Quello che definirei il sotto-genere, rispetto al *roman*, dei *romans d'Antiquités* si rivela di notevole interesse sotto il duplice punto di vista letterario e linguistico.

Letterariamente, questa sequenza (*Roman d'Alexandre*, 1150 ca.; *Roman de Thèbes*, 1150 ca.; *Roman de Brut*, 1155 ca.; *Roman d'Enéas*, 1160 ca.; *Roman de Troie*, 1165 ca.; *Roman de Jules*

letterario, Longrée – Mellet 2013.

40 Sorba – Denoyelle 2020: 1.

41 *Ibid.*: 2.

César, 1250 ca.), che segna la nascita del *roman* come genere, si apre e si chiude simbolicamente con due opere – il *Roman d’Alexandre* da una parte e il *Roman de Jules César* dall’altra – che tendono la mano verso i generi dominanti rispettivamente nel periodo precedente e in quello successivo: se il *Roman d’Alexandre*, anomalo a partire dalla sua forma – l’alessandrino appunto, in luogo di quell’ottosillabo a rima piatta che diventa la forma di espressione di un genere –, si muove sul crinale della *chanson de geste*, il *Roman de Jules César* si impone sulla scena della scrittura, chiudendo l’epoca dell’*Antiquité en roman*, cioè della storia versificata, nel momento in cui si stanno imponendo le grandi compilazioni storiografiche in prosa. Fra queste, si collocano per primi i *Faits des Romains*, datati al 1213-1214, cioè la vita dello stesso Cesare; *Faits des Romains* che, chiudendo il cerchio e operando un passaggio di testimone a doppio giro, costituiscono, insieme alla *Farsalia* e ai *Commentaria*, una delle fonti del successivo *roman*. A un secolo di distanza, ancora per chiudere un cerchio, l’ultimo dei *romans d’Antiquité*, ibrido e problematico dal punto di vista del genere, condivide col primo la forma: 9.500 alessandrini articolati in lasse monorima sul modello del *Roman d’Alexandre* e della sua ambiguità epico-romanzesca. A seguire, il panorama della scrittura storiografica conoscerà la monumentale *Histoire ancienne jusqu’à César* di Wauchier de Danain.

Linguisticamente, il lessico dei *romans d’Antiquité*, e quindi questo bacino lessicale specifico dell’*ancien français*, si forgia in un atto di traduzione, traduzione di fonti latine, cosicché l’espressione icastica “*metre en romanz*” assume il duplice valore di atto fondativo di un genere letterario, il *roman* appunto, e di atto di mediazione linguistica tra il latino e il *romanz* in quanto lingua, che significa in ultima analisi atto di mediazione tra *auctoritas* e pubblico laico ad opera di quella figura fondamentale di *passeur* culturale che è il *clerc* del XII sec. Col risultato che il termine di designazione di una lingua e di un genere letterario si sovrappongono e fondono indicando implicitamente tutta la pregnanza culturale⁴² di quel genere, per il quale è opportuno ricordare le parole di Michel Zink (1987: 6-7):

En vain chercherait-on en lui les traces de l’oralité. Il est la première forme littéraire en langue vulgaire destinée à être lue, et non chantée (comme la lyrique ou l’épique). Le roman ne dérive donc pas de la chanson de geste, dont tout le sépare : la forme, le mode supposé de diffusion, le cadre de l’action, l’époque où elle située, et très vite la sensibilité. Il se rapproche davantage des vies des saints, très tôt écrites en octosyllabes et rimes plates et traduites comme lui du latin. Il en est une sorte de laïcisation, reportant sur les philosophes, c’est-à-dire sur les auteurs de l’antiquité païenne, l’autorité des textes scripturaires et hagiographiques.

Il primo corpus con cui si inaugura il progetto EXOMARE è tuttavia, come si è detto, quello relativo al *Roman de Florimont*, non un romanzo d’antichità della prima generazione. La scelta dettata da ragioni pragmatiche – vale a dire la conoscenza approfondita del testo e della tradizione manoscritta da parte di chi scrive, che se ne sta occupando dal 2017 e sta terminando un’edizione critica del ramo testuale diffusosi in Italia, accompagnata da traduzione in italiano – si è rivelata però anche giustificata dal punto di vista della funzionalità scientifica, come si illustrerà nel paragrafo successivo. Non è certo l’ultimo degli argomenti a favore della scelta di

42 Per una sintesi complessiva sul significato e su tutte le implicazioni (linguistiche e letterarie, e quindi culturali) di quest’atto si rimanda alla monografia di Mora-Le Brun 2008.

questo testo, anche se forse è il più banale, il fatto che, a differenza di *Thèbes* e dell'*Enéas* nella loro interezza e di *Troie* e dell'*Alexandre* in ampi passaggi (solo per citare i testi più importanti e rappresentativi), il *Florimont* è un grande assente in ogni iniziativa testuale-linguistica: il suo testo, nella forma del ms. Paris, BnF, fr. 792, è utilizzato esclusivamente dal *Dictionnaire Godefroy*. Considerando le caratteristiche della sua tradizione, che si illustreranno a seguire, l'immissione nei contesti di analisi linguistiche di questo bacino di lemmi e forme appare un contributo potenzialmente degno di interesse.

2. Il corpus-*Florimont*

Con la sua narrazione che si colloca in un'epoca imprecisata ma "nell'Antichità" (successivamente alla guerra di Troia, di cui si citano personaggi, e anteriormente alla nascita di Alessandro Magno, di cui *Florimont*, padre di Filippo di Macedonia, è il nonno paterno) e in uno scenario mediterraneo (tra Babilonia e Cartagine, Durazzo e la Puglia) che risale verso l'interno macedone-ungherese, e che vede la presenza di elementi che afferiscono al *féerique* e al meraviglioso, il *Florimont* si inserisce in maniera esemplare nella categoria dei *romans antiques* di seconda generazione. Esso si propone quindi come utile complemento al corpus DÉCT, cioè all'opera omnia di Chrétien de Troyes che, con l'eccezione parziale del *Cligés*, possiamo collocare in blocco nell'alveo della *matière de Bretagne*. Rispetto all'opera di Chrétien, ben nota e analizzata, il *Florimont* rappresenta quindi un'apertura su un altro potenziale filone lessicale, altro in un duplice senso: altro rispetto alla *matière* e altro rispetto alla figura autoriale.

Il testo del *Florimont* presenta una lunghezza che lo pone in linea con i *romans antiques* propriamente detti: 12.532 ottsillabi nella versione più breve, circa 13.200 in quella più lunga; *Roman de Thèbes*, 10.541 vv. nella versione breve, 14.626 in quella lunga; *Roman d'Enéas*, 10.156 vv. Si consideri di contro la lunghezza dei *romans* di Chrétien: *Yvain* 6.818 vv., *Cligés* 6.664 vv. ecc.

La corposità del testo, che significa per conseguenza possibilità di verificare su una buona scala usi lessicali e sintattici, si accompagna alla corposità della tradizione manoscritta da un punto di vista quantitativo e alla sua diversificazione da un punto di vista cronologico e geografico⁴³. Ad oggi – e sottolineo ad oggi perché io stessa ho avuto la possibilità di individuare un nuovo frammento, quello del ms. Paris, BnF, nouv. acq. lat. 1034⁴⁴, correttamente inventariato (è doveroso dirlo) dai redattori delle schede bibliografiche della BnF ma finora mai segnalato e del tutto sconosciuto agli studiosi di quest'opera – il corpus manoscritto del *Florimont* si compone di: 13 mss. completi, 1 ms. mutilo dell'inizio e della fine, 1 ms. incompleto della fine, 1 ms. con una lacuna materiale corrispondente alla perdita di un fascicolo, 2 frammenti (rispettivamente di 261 e 105 vv.). ARLIMA segnala l'esistenza di un manoscritto del XIV sec., di 152 cc., con *mise en page* su due colonne di 45 linee, messo in vendita nel 1767 presso Vallat-la-Chapelle e la cui localizzazione è attualmente sconosciuta. A questa lista si deve aggiungere una copia moderna del XIX sec., il ms. Albi, BM, Rohegude 005, di interesse in quanto non corrispondente a nessuno dei testimoni conosciuti. Possiamo quindi in conclusione e sintesi parlare di: 14 mss. completi (virtualmente 15 considerando la copia la cui localizzazione è sconosciuta), 3 mss. incompleti

43 L'ultimo bilancio codicologico rispetto alla tradizione manoscritta del *Florimont* è quello di Busby 2020.

44 Materni 2024.

e 2 frammenti. Si consideri, come termini di paragone, che: *Thébes* è testimoniato da 5 mss. completi e 1 frammento di 221 vv.; *Enéas* da 8 mss. completi e 1 ms. con lacuna iniziale e finale; *Erec* da 7 mss. completi e 3 frammenti rispettivamente di 273, 306 e 901 vv.; *Cligés* da 7 mss. completi, 1 ms. lacunoso e 3 frammenti di 25, 550 e 244 vv. Con l'eccezione del *Roman de Troie*, il *Florimont* eguaglia e anzi supera la tradizione di un testo come il *Perceval*, con i suoi 15 mss. e 1 frammento. Quanto alla vitalità del testo in termini di rielaborazione, si aggiunga: una traduzione in medio-neerlandese, due prosificazioni francesi, una traduzione-prosificazione in veneto⁴⁵, una tradizione a stampa.

A tutto questo fa paradossalmente da *pendant* una sostanziale lacuna negli studi colmata solo minimamente da una serie di contributi sparsi e da un'unica edizione risalente al 1932, a opera di Alfons Hilka, immediatamente criticata e in più, attualmente, difficilmente accessibile: attraverso quell'edizione, l'opera-*Florimont* è stata identificata col testo-*Florimont* di un ramo minoritario della tradizione, rappresentato da due soli manoscritti e copiato nella Lorena della fine del XIII sec. Se si aggiunge all'entità della tradizione manoscritta (che potrebbe ascrivere anche a mera fortuna di sopravvivenza) la ricchezza della fortuna testuale (prosificazioni, traduzioni, stampa), la valorizzazione di questo testo e la sua scelta per sperimentare l'approccio proposto da EXOMARE mi sembrano sufficientemente giustificati.

Ad essere interessante però non è solo il numero dei manoscritti in sé per sé ma anche la loro distribuzione cronologica e geografica. Per quel che riguarda datazioni e localizzazioni, alla data attuale si fa riferimento a quelle "ufficiali", emerse in modo non organico da vari contributi e da *expertises* personali realizzate soprattutto da Christine Ruby e Keith Busby, di cui si può leggere un bilancio in un contributo dello stesso Busby del 2020 e nei dati presenti nella base *Jonas*, curati da Ruby. Fra le attività collaterali del progetto EXOMARE, in questa prima fase sotto il segno del *Florimont*, è prevista una revisione completa *ex novo* di tutta la tradizione dal punto di vista paleografico-codicologico.

Per le datazioni: tre datazioni esplicite, 1295 (London, BL, Harley 4487), 1323 (London, BL, Harley 1323), 1465 (Wien, ÖN, 3434); metà XIII sec. (Paris, BnF, fr. 1374); 2 mss. ultimo quarto XIII sec. (Paris, BnF, fr. 792 e fr. 24376); 2 mss. XIII *ex.* (Monza, BC, b-21/137; Tours, BM, 941); 2 mss. XIII *ex.*-XIV *in.* sec. (Paris, BnF, fr. 1376 e fr. 1491); 2 mss. inizio XIV sec. (Paris, BnF, fr. 15101 e nouv. acq. fr. 5094); 1 ms. primo quarto XIV sec. (Paris, BnF, fr. 353); 3 mss. secondo quarto XIV sec. (Venezia, BnM, fr. Z.XXII; Paris, BnF, nouv. acq. lat. 1034; Torino, BU, L.II.16); 1 ms. metà XIV sec. (Montpellier, BU Médecine, H 252).

Per le localizzazioni a oggi appurate o rispetto alle quali si sono formulate delle ipotesi anche sulla scorta dell'analisi dell'apparato decorativo: Borgogna (fr. 1376), Metz (fr. 15101), Lorena (Harley 4487 e 1323), Italia (fr. 24376, Torino, Venezia), Sud-Est della Francia (Tours), Midi? (fr. 792, fr. 1491, nouv. acq. fr. 5094, nouv. acq. lat. 1034), Vienne? (fr. 1374), Nord della Francia? (fr. 353).

Nel suo complesso quindi la distribuzione geografica dei manoscritti del *Florimont* corre dal Sud al Nord-Est della Francia, proiettandosi oltre la regione francese, a nord verso la regione neerlandese e a sud verso l'Italia settentrionale, dall'arco alpino alla regione padovana; mentre la copertura cronologica si estende dalla metà/ultimo quarto del XIII sec. a coprire tutto il XIV

45 Per una panoramica su quest'opera si veda Materni 2023.

sec., e con l'interessantissimo testimone *recentior* del 1465. Tutto questo al netto di alcune localizzazioni ancora fortemente incerte e che potrebbero confermare o ampliare questo asse di diffusione a tendenza verticale.

La casistica manoscritta, attraverso l'approccio pluritestimoniale della base, permette quindi di presentare testi la cui *scripta* è parzialmente complementare rispetto a quelle presenti alla data attuale nella BFM: lorrain, franco-picard, picard, anglo-normand, normand, amiénois, cambrésien, arrageois, parisien, vendéen. E nel NCA (l'ordine di citazione non è casuale ma corrisponde all'ordine decrescente di percentuale di attestazione): picard, anglo-normand, francien, lorrain, normand, chamenois, bourguignon, wallon, poitevin, orléanais, tourangeon, berrichon; a queste aree di origine dei manoscritti si contrappone in parte l'ordine della distribuzione geolinguistica delle opere: picard, anglo-normand, champenois, francien, norman, bourguignon, angevin, wallon, lorrain, orléanais, poitevin, berichon, affinités occitanes e francoprovençal. Vediamo in particolare che la tradizione manoscritta del *Florimont* permette, permetterà, di aprirsi verso le *scriptae* del Sud-Est.

3. Trattamento tecnico dei dati testuali

La base EXOMARE si compone di due elementi, virtualmente connessi attraverso una chiave ma, al momento, ancora distinti. La connessione fra le due strutture nella risposta alle *query* rappresenterà il prossimo step nell'implementazione della suite EXOMARE. Allo stato attuale dello sviluppo abbiamo a disposizione: da una parte le trascrizioni dei testi, visualizzabili in modalità diplomatica-semidiplomatica⁴⁶ vs. interpretativa (o modernizzata), abbinata alle immagini dei manoscritti, e informaticamente costruite a partire da una codifica XML-TEI; dall'altra parte il corpus dei testi lemmatizzati e annotati morfo-sintatticamente, secondo un modello di annotazione UD (*Universal Dependencies*)⁴⁷ e stoccati in un database da cui vengono estratti dei file .csv. Il legame fra le due strutturazioni dei dati, vale a dire fra le due differenti strutturazioni degli stessi dati secondo differenti prospettive di analisi, è garantita dalla presenza di un ID per ciascuna unità lessicale. Il limite attuale della risposta fornita alle *query* è nel fatto che per il momento la visualizzazione contestuale della forma, ovvero la concordanza, si realizza rispetto alla forma del testo utilizzata per realizzare la lemmatizzazione-annotazione, cioè una forma modernizzata o interpretativa che dir si voglia, alla quale è stato cioè applicato lo scioglimento silente delle abbreviazioni, la regolarizzazione della distribuzione delle lettere ramiste, e l'introduzione di alcuni segni diacritici con funzione disambiguante. Il che significa che, per il momento,

46 Nel caso della modalità di visualizzazione EXOMARE, che sfrutta quella messa a punto durante il progetto DIGIFLOR (versi) e poi ulteriormente perfezionata con il progetto PRODIGI (prosa), uso l'espressione diplomatica-semidiplomatica perché l'uso dello strumento informatico ha permesso praticamente di far collassare l'una sull'altra le due modalità di rappresentazione del dato documentario: in questa modalità di visualizzazione infatti si rispetta la segmentazione scribale del testo e la disposizione del testo sulle colonne, si sciolgono le abbreviazioni segnalando le lettere restituite ma si permette, passando con il cursore sopra l'abbreviazione sciolta, di visualizzare l'immagine dell'abbreviazione originale, si rispetta lo stato originale delle lettere ramiste, si introducono tutti gli interventi di correzione del copista e le note a margine, arrestandosi sostanzialmente solo di fronte alla variante delle forme delle lettere, che non è presa in considerazione.

47 Si tratta del formato di annotazione linguistica verso cui si sta orientando anche l'équipe di Lione, dove la BFM è nata invece sotto il segno del *jeux d'étiquettes* Cattex: <https://universaldependencies.org>.

vengono meno nella risposta le informazioni aggiuntive legate allo stato grafico pre-normalizzato, alla presenza di abbreviazioni e alla segmentazione scribale della scrittura (agglutinazione e deglutinazione). Per queste informazioni è necessaria una ricerca distinta in un altro ambiente, quello appunto dell'edizione. L'obiettivo finale è invece quello di restituire una concordanza che punti automaticamente all'edizione del testo, la quale a sua volta è (e lo è già) visualizzabile in modalità diplomatica-semidiplomatica e in modalità interpretativa. Questa certamente non auspicabile scissione deriva da una scelta ponderata, considerata necessaria, fatta a livello di formattazione dei dati: la scelta cioè di non sovraccaricare inutilmente e farraginosamente il codice XML, dal momento che la ripetitività delle informazioni linguistiche (i *token* della stessa forma presentano tutti la stessa annotazione che, nella struttura a database, viene associata con la forma ripercuotendosi poi automaticamente, ma senza repliche dell'informazione, sui *token*, mentre nella codifica XML andrebbe ripetuta per ogni singolo *token*) mal si concilia con la verbosità, già di difficile gestione, che è il tallone d'Achille di questo linguaggio.

Il modello di codifica XML-TEI applicato è stato messo a punto in occasione del progetto Marie Curie DIGIFLOR, quando è stato anche realizzato un editor, TEI-MEDit (*TEI MEDiator EDitor*)⁴⁸, che permettesse di realizzare agevolmente la codifica molto dettagliata a cui si aspirava, puntando da una parte alla semi-automatizzazione del processo, in modo da minimizzare l'errore, e dall'altra a facilitare il compito del codificatore, in modo da massimizzare il tempo-lavoro. La codifica DIGIFLOR, orientata verso la struttura testuale in versi, è stata ulteriormente affinata successivamente in occasione del progetto Marie Curie PRODIGI, orientato invece verso la scrittura in prosa. Tale codifica si fonda su una granularità a livello delle parole (non grafiche ma lessicali: in questo modo, ad esempio, le particelle pronominali enclitiche sono codificate in modo distinto dalla parola a cui si appoggiano), ciascuna individuata da un tag <w> (*word*) associato a un ID. Sono inoltre codificate tutte le abbreviazioni e parimenti le sequenze di scrittura originali, vale a dire agglutinazioni e deglutinazioni.

L'esperienza PRODIGI ha invece permesso di proporre un *tagset* per l'annotazione morfo-sintattica che si appoggia allo standard UD, introducendo/proponendo però alcune varianti che sono sembrate necessarie nel momento in cui l'aspirazione universalistica di questo standard si scontrava in modo troppo violento con le caratteristiche e con le abitudini descrittive delle lingue romanze: è sembrato ad esempio imprescindibile introdurre le etichette distintive per l'articolo definito e quello indefinito, così come la nozione di preposizione semplice ed articolata; è stato inoltre aggiunto un campo *Function* che permette di descrivere in modo si spera più soddisfacente situazioni come quelle dei verbi ausiliari (avremo un verbo con *funzione* ausiliare e non una categoria a parte come nel *tagset* UD) o della discrepanza fra forma e uso, come nel caso degli infiniti sostantivati o dei participi con funzione aggettivale. Analogamente a quanto avvenuto per il progetto DIGIFLOR con la messa a punto dell'editor TEIMEDit, la riflessione teorica è stata accompagnata da una ricerca volta all'ergonomia degli strumenti a disposizione per realizzare le operazioni sui testi: è stata così creata la struttura per la lemmatizzazione-annotazione ULA (*Universal Language Annotation*) che, coniugando i vantaggi della lemmatizzazione-annotazione per liste con quelli della verifica per contesti, e sfruttando la parziale automatizzazione

48 Il manuale di codifica DIGIFLOR è contenuto in Materni 2020a; la descrizione del funzionamento dell'editor TEI-MEDit si legge in Materni 2020b. Per un'ulteriore discussione delle maggiori problematiche che nascono dal tentativo di rappresentare informaticamente la complessità grafica manoscritta, si veda Materni 2021.

delle assegnazioni delle etichette, ha consentito a un singolo ricercatore l'annotazione completa di un testo come il *Roman de Troie en prose* nello spazio di circa un anno e mezzo⁴⁹.

Il progetto EXOMARE può quindi oggi tirare le fila e raccogliere i frutti, in termini di strumenti e procedure, dell'esperienza maturata negli ultimi sei anni, proponendosi così, anche, come un progetto ragionevolmente realizzabile.

Conclusioni

Si è aperto il contributo partendo da una citazione di Martin Glessgen, e lo si chiude allo stesso modo (Glessgen 2007: 51): “Une forme linguistique qui n'est pas précisément ancrée dans le diasystème linguistique en termes de temps, espace, genre textuel et état de transmission reste en dernière instance un fantôme. Elle est inutilisable pour des quantifications et pour des interprétations sur le changement et la variation linguistique”.

Su queste premesse, nell'affrontare il *mare magnum* dell'*ancien français*, il progetto EXOMARE ha proceduto delimitando spazi – la testualità esclusivamente letteraria, una particolare declinazione della testualità letteraria, uno specifico segmento cronologico – nella convinzione che in presenza di una “langue sans locuteur” (Prévost 2008: 3) qual è una lingua antica, ogni pretesa di esaustività sia minata alla radice dalla condanna all'accesso alla lingua esclusivamente attraverso i suoi testi, peraltro testi sopravvissuti. In secondo luogo, il fulcro della proposta EXOMARE è l'idea che la fonte dei dati non sia rappresentata da un'edizione critica o dalla trascrizione di un singolo manoscritto, bensì dalla trascrizione di tutti i testi attraverso cui si è perpetuata un'Opera: è in questo senso che il macrocorpus EXOMARE dedicato all'*ancien français* letterario del 1150-1250 si comporrà progressivamente di subcorpora d'opera/d'autore, corpora costituiti cioè dalle trascrizioni di tutte le testimonianze manoscritte di una data opera, al fine, in primo luogo, di seguire la variazione linguistica delle forme del testo di un'opera attraverso lo spazio e il tempo.

Il progetto EXOMARE però non vuole presentare solo dati testuali annotati. L'onere della lemmatizzazione e dell'etichettatura vogliono condurre anche a un'altra tipologia di prodotto: far confluire cioè i dati dell'analisi di una singola opera verso l'organizzazione delle informazioni in un dizionario d'opera/d'autore (da replicare per ciascuna opera affrontata) sul modello del DÉCT, così da sfruttare la lingua come strumento per la creazione di profili autoriali nell'impossibilità, che corrisponde alla maggioranza dei casi per l'epoca medievale, di tracciare dei profili biografico-artistici.

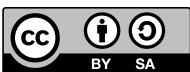
49 Il tagset PRODIGI e le funzionalità di ULA sono descritte in Materni 2022.

Riferimenti bibliografici

- Buridant, C. (2019). La morphologie au carrefour de la description linguistique. In F. Duval, C. Guillot-Barbance, & F. Zinelli (Eds.), *Les introductions linguistiques aux éditions de textes* (pp. 197-222). Paris: Classiques Garnier.
- Busby, K. (2020). Le *Florimont* dans ses contextes codicologiques. In M. Materni (Ed.), *Autour du 'Roman de Florimont'. Approches multidisciplinaires à la complexité textuelle médiévale* (pp. 9-24). Padova: Università degli Studi di Padova. <DOI 10.25430/2724-0975/2>
- Coluccia, R. (2000). L'edizione dei documenti e i problemi linguistici della copia (con tre appendici un po' stravaganti intorno a Guglielmo Maramauro). *Medioevo Romanzo*, 24, 231-255.
- Denoyelle, C. (2014). La formulation des requêtes dans quelques textes littéraires médiévaux. In D. Lagorgette, & P. Larrivée (Ed.), *Représentations du sens linguistique 5* (pp. 91-114). Chambéry: Presses Universitaires de Savoie.
- Duval, F.; Guillot-Barbance, C.; & Zinelli, F. (Eds.) (2019). *Les introductions linguistiques aux éditions de textes*. Paris: Classiques Garnier.
- Glessgen, M.-D. (2003a). L'élaboration philologique et l'étude lexicologique des plus anciens documents linguistiques de la France à l'aide de l'informatique. In F. Duval (Ed.), *Frédéric Godefroy. Actes du X^e Colloque international sur le moyen français (Metz, 12-14 juin 2002)* (pp. 371-386). Paris: Éditions de l'École des chartes.
- . (2003b). La lemmatisation de textes d'ancien français: méthodes et recherches. In P. Kunstmann, F. Martineau, & D. Forget, *Ancien et moyen français sur le web: enjeux méthodologique et analyse du discours* (pp. 55-75). Ottawa: Éditions David.
- . (2008). Les lieux d'écriture dans les chartes lorraines du XIII^e siècle. *Revue de Linguistique Romane*, 72, 413-540.
- . (2012). Trajectoires et perspectives en scriptologie romane. *Medioevo romanzo*, 36, 5-23.
- Glessgen, M.-D.; & Gouvert, X. (2007). La base textuelle du *Nouveau corpus d'Amsterdam*: ancrage diasystématique et évaluation philologique. In P. Kunstmann, & A. Stein (Eds.), *Le Nouveau Corpus d'Amsterdam (Actes de l'atelier de Mauterbad, 23-26 février 2006)* (pp. 51-84). Stuttgart: Steiner.
- Glessgen M.-D.; Kihai, D.; & Videsott, P. (2010). L'élaboration philologique et linguistique des *Plus anciens documents linguistiques de la France*. Édition électronique. Trois études. *Bibliothèque de l'École des Chartes*, 168, 5-94.
- Glessgen, M.-D.; & Vachon, C. (2013). L'étude philologique et scriptologique du *Nouveau Corpus d'Amsterdam*. In E. Casanova Herrero, & C. Calvo Rigual, *Actes du XXVI^e Congrès International de Linguistique et Philologie Romanes (6-11 sept. 2010, València)* (pp. 191-204). Berlin: De Gruyter. <<https://doi.org/10.1515/9783110300017.191>>
- Greub, Y. (2007). Sur un mécanisme de la préstandardisation de la langue d'oïl. *Bulletin de la Société Linguistique de Paris*, 102, 429-434.
- Guiot, C.; & Lavrentiev, A. (2009). *Manuel de description des textes pour la Base de Français Médiéval, version 2.3*. <http://ccfm.ens-lyon.fr/IMG/pdf/Manuel_Descripteurs_BFM.pdf>
- Habert, B. (2000). Des corpus représentatifs: de quoi, pour quoi, comment?. In M. Bilger (Ed.), *Linguistique sur corpus. Études et réflexions* (pp. 11-58). Perpignan: Presses Universitaires de Perpignan.

- Hasenohr, G.; & Raynaud de Lage G. (2019). *Introduction à l'ancien français*. Paris: Armand Colin.
- Heiden, S.; & Lavrentiev, A. (2004). Ressources électroniques pour l'étude des textes médiévaux: approches et outils. *Revue française de linguistique appliquée*, 1, vol. IX, 99-118. <<https://doi.org/10.3917/rfla.091.0099>>
- Joly, G. (2018). *Précis d'ancien français. Morphologie et syntaxe*. Paris: Armand Colin.
- Lavrentiev, A. (2008). *Manuel d'encodage XML-TEI étendu des transcriptions de manuscrits dans le projet BFM-Manuscrits, Version 2.1*. <http://ccfm.ens-lyon.fr/IMG/pdf/BFM-Mss_Encodage-XML.pdf>
- Leonardi, L. (2019). Langue du copiste et langue de l'auteur. In F. Duval, C. Guillot-Barbance, & F. Zinelli (Eds.), *Les introductions linguistiques aux éditions de textes* (pp. 167-180). Paris: Classiques Garnier.
- Lodge, A. (2008). Les débuts de la standardisation du français. In J. Durand, B. Habert, & B. Laks B. (Eds.), *Congrès Mondial de Linguistique Française – CMLF'08, Paris, Institut de Linguistique Française, Section Diachronie, histoire de la langue* (pp. 367-376). <<https://doi.org/10.1051/cmef08320>>
- Longrée, D.; & Mellet, S. (2013). Le motif: une unité phraséologique englobante? Étendre le champ de la phraséologie de la langue au discours. *Langages*, 189, 65-79.
- Mairey, A.; & Aouini, M. (2020). PALM: Un modèle neuronal pour l'étiquetage morphosyntaxique des textes médiévaux. In *JADT 2020, 15^{es} Journées Internationales d'Analyse statistiques des Données Textuelles*
- Malrieux, D.; & Rastier, F. (2001). Genres et variations morphosyntaxiques. *Traitement automatique des langues*, 42, 2, 547-577.
- Materni, M. (2020a). DigiFlorimont: une occasion de réflexion philologique et numérique autour de la représentation de la complexité textuelle médiévale. In M. Materni (Ed.), *Autour du Roman de Florimont. Approches multidisciplinaires à la complexité textuelle médiévale* (pp. 165-184). Padova: Quaderni di Francigena. <<http://doi.org/10.25430/2724-0975/2>>
- . (2020b). Complessità della codifica ed ergonomia strumentale nel contesto XML-TEI: dove siamo? Bilancio a partire da un nuovo progetto di edizione digitale medievale). *Umanistica Digitale*, 8, 123-143.
- . (2021). Les Guidelines TEI à l'épreuve de la complexité textuelle et graphique médiévale. *Magnificat. Cultura i Literatura Medievals*, 8, 1-32.
- . (2022). Note di lavoro intorno alla creazione di una struttura di analisi lessicale (*Roman de Troie Prose 2*, ms. Grenoble BM 861). *Francigena*, 8, 231-281.
- . (2023). Il *Florimont*. In F. Gambino, & A. Beretta (Eds.), *Antologia del francese d'Italia (XIII-XV secolo)* (pp. 173-194). Bologna: Pàtron Editore.
- . (2024). Il *Florimont* e i foyers di copia orientali: edizione e analisi linguistica del frammento Paris, BnF, nouv. acq. lat. 1034. *Translat Library*, 6, 7, 1-25. <https://doi.org/10.7275/tl.2792>
- Mazziotta, N. (2004). Le texte dans tous ses états. Philosophie d'encodage du projet Khartès. In *JADT 2004* (pp. 793-803).
- Mora-Le Brun, F. (2008). *“Metre en romanz”: les romans d'antiquité du XI^e siècle et leur postérité (XIII^e-XIX^e siècle)*, Paris: Champion.
- Pignatelli, C.; & Robinson, M. C. (2002). Chrétien de Troyes, *Le Chevalier de la Charrette* (Lancelot), Le 'Projet Charrette' et le renouvellement de la critique philologique des textes. Tübingen: Gunter Narr.
- Pfister, M. (1993). *Scripta et koinè* en ancien français aux XII^e et XIII^e siècles? In P. Knecht, & Z. Marzys (Ed.), *Écriture, langues communes et normes* (pp. 17-41). Genève: Droz.

- Prévost S. (2008), Corpus informatisés de français médiéval: contraintes sur leur constitution et spécificités de leurs apports, *Corpus*, 7 (= Constitution et exploitation des corpus d'ancien et de moyen français). <https://doi.org/10.4000/corpus.1500>
- Rainsford, Th. (2022). Old Gallo-Romance (OGR) Corpus: annotation phonologique et métrique des plus anciens textes gallos-romans. In *Congrès Mondial de Linguistique Française, SHS Web of Conferences*, 138, 02007. <https://doi.org/10.1051/shsconf/202213802007>
- Sinclair, J. (1996). Preliminary recommendations on Corpus Typology. <https://www.ilc.cnr.it/EAGLES96/corpus/typ/corpus.html>
- Sorba, J.; & Denoyelle, C. (2020). L'approche phraséologique du roman médiéval: une voie de caractérisation générique. In *7^e Congrès Mondial de Linguistique Française, SHS Web of Conferences* 78 <<https://doi.org/10.1051/shsconf/20207805005>>
- . (2022). Analyse phraséologique du marquer discursif *ce m'est avis* et de ses variantes dans un corpus d'ancien français. *Linguisticae Investigationes*, 45.2, 251-275.
- . (2023). Phraséologie et genres textuels. Une étude pilote dans le roman médiéval autour des verbes *donner* et de *mettre*. *Studia Linguistica Romanica*, 9, 24-46.
- Uitti, K. (1997). A Brief History of the 'Charrette Project' and Its Basic Rationale. <<https://www.princeton.edu/~lancelot/projhist.html>>
- Videsott, P. (2019). Diatopie et *scripta*. In F. Duval, C. Guillot-Barbance, & F. Zinelli (Eds.), *Les introductions linguistiques aux éditions de textes* (pp. 181-196). Paris: Classiques Garnier.
- Zink, M. (1987). Chrétien et ses contemporaines. In N. J. Lacy, D. Kelly, & K. Busby, *The Legacy of Chrétien de Troyes: Chrétien et ses contemporaines* (pp. 5-32). Amsterdam: Rodopi.
- Zumthor, P. (1972). *Essai de poésie médiévale*. Paris: Seuil.



This work can be used in accordance with the Creative Commons BY-SA 4.0 International license terms and conditions (<https://creativecommons.org/licenses/by-sa/4.0/legalcode>). This does not apply to works or elements (such as images or photographs) that are used in the work under a contractual license or exception or limitation to relevant rights.