

# Francigena

10 (2024)

Il 'francese d'Italia' e il progetto *FrIngE*.  
Panoramica generale e casi di studio

Francesca Gambino *et alii*

(Università degli Studi di Padova, Cnr-Istituto di Linguistica  
Computazionale, Università degli Studi di Genova)



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

*Direzione / Editors-in-chief*

GIOVANNI BORRIERO, Università degli Studi di Bologna  
FRANCESCA GAMBINO, Università degli Studi di Padova

*Comitato scientifico / Advisory Board*

CARLOS ALVAR, Universidad de Alcalá  
ALVISE ANDREOSE, Università degli Studi e-Campus  
FRANCESCO BORGHESI, Università di Modena e Reggio Emilia/University of Sydney  
FURIO BRUGNOLO, Università degli Studi di Padova  
KEITH BUSBY, The University of Wisconsin  
LAURA J. CAMPBELL, Durham University  
DAN OCTAVIAN CEPRAGA, Università degli Studi di Padova  
RACHELE FASSANELLI, Università degli Studi di Padova  
CATHERINE GAULLIER-BOUGASSAS, Université de Lille 3  
JOHN HAJEK, The University of Melbourne  
BERNHARD HUB, Freie Universität Berlin, Germania  
MARCO INFURNA, Università Ca' Foscari di Venezia  
STEPHEN P. MCCORMICK, Washington and Lee University  
ILARIA MOLteni, University of Lausanne  
LUCA MORLINO, Università di Trento  
GIANFELICE PERON, Università degli Studi di Padova  
LORENZO RENZI, Università degli Studi di Padova  
ANDREA RIZZI, The University of Melbourne  
FABIO SANGIOVANNI, Università degli Studi di Padova  
ZENO VERLATO, Opera del Vocabolario Italiano, CNR  
RAYMUND WILHELM, Alpen-Adria-Universität Klagenfurt, Austria  
LESLIE ZARKER MORGAN, Loyola University Maryland

*Redazione / Editorial Staff*

ANDREA BERETTA, Università degli Studi di Padova  
IVO ELIES OLIVERAS, Scuola Superiore Meridionale  
JACOPO FOIS, Università degli Studi di Padova  
MARCO FRANCESCON, Università degli Studi di Padova, chief editor  
FEDERICO GUARIGLIA, Università di Genova  
CLAUDIA LEMME, Università di Chieti-Pescara  
MARTA MATERNI, Università degli Studi della Tuscia  
MARTA MILAZZO, Università degli Studi di Milano Statale  
ELENA MUZZOLON, Università degli Studi di Padova  
ELEONORA POCETTINO, Università degli Studi di Napoli Federico II  
CARLO RETTORE, Università degli Studi di Padova  
BENEDETTA VISCIDI, Université de Fribourg, chief editor

*Francigena is an international peer-reviewed journal with an  
accompanying monograph series entitled "Quaderni di Francigena"*

ISSN 2420-9767

Dipartimento di Studi Linguistici e Letterari  
Via E. Vendramini, 13  
35137 PADOVA

[info@francigena-unipd.com](mailto:info@francigena-unipd.com)

## INDICE

FABIO ZINELLI	
Storia di una grafia ( <i>leuc, fenc, jeuc</i> ): i valori testuali del francese medievale	5
LORENZO TOMASIN	
Gallicismes du vénitien et vénétianismes du français d'Italie	57
ALVISE ANDREOSE	
Esplorazioni lessicali nel <i>Devisement dou monde</i>	83
MASSIMO DAL BIANCO	
Il codice Firenze BML Ash. 123: struttura e <i>scripta</i>	149
LESLIE ZARKER MORGAN	
« <i>Roland, senator romanus</i> »: Origins of the Italian Literary Romance-Epic Trope	181
LEONARDO TERRUSI	
Ancora sui rapporti tra deonomastica italiana e letteratura antico francese	229
ANDREA BIANCO	
Osservazioni sulla presenza di 'friulanismi' nei testi Franco-italiani	261
FRANCESCA GAMBINO, ANDREA BERETTA, SONIA BARILLARI, FLORIANA CERESATO, GIACOMO COSTA, RACHELE FASSANELLI, MANUEL FAVARO, JACOPO FOIS, ELISA GUADAGNINI, FEDERICO GUARIGLIA, MATTEO PARODI, CARLO RETTORE	
Il 'francese d'Italia' e il progetto <i>FrIngE</i> . Panoramica generale e casi di studio	285

**Open Access. ©2024 Francesca Gambino *et alii*. This work is licensed under  
the Creative Commons Attribution 4.0 International License.  
<https://doi.org/10.25430/2420-9767/V10-008>  
DOI: 10.25430/2420-9767/V10-008**

# Il ‘francese d’Italia’ e il progetto *FrIngE*. Panoramica generale e casi di studio\*

Francesca Gambino<sup>^</sup>, Andrea Beretta<sup>^</sup>, Sonia Barillari<sup>°</sup>,  
Floriana Ceresato<sup>^</sup>, Giacomo Costa<sup>^</sup>, Rachele Fassanelli<sup>^</sup>,  
Manuel Favaro<sup>#</sup>, Jacopo Fois<sup>^</sup>, Elisa Guadagnini<sup>#</sup>,  
Federico Guariglia<sup>°</sup>, Matteo Parodi<sup>°</sup>, Carlo Rettore<sup>^</sup>  
francesca.gambino@unipd.it, andrea.beretta@unipd.it, maurasonia.barillari@unige.it,  
floriana.ceresato@unipd.it, giacomo.costa.4@phd.unipd.it, rachele.fassanelli@unipd.it,  
manuel.favaro@ilc.cnr.it, jacopo.fois@unipd.it, elisa.guadagnini@cnr.it,  
federico.guariglia@edu.unige.it, matteoparodi62@gmail.com, carlo.rettore.1@unipd.it

(<sup>^</sup>Università degli Studi di Padova, <sup>#</sup>Cnr-Istituto di Linguistica  
Computazionale, <sup>°</sup>Università degli Studi di Genova)

## ABSTRACT:

Il contributo presenta i primi risultati del progetto *FrIngE - The French in/of Italy: Code-MixiNG in Medieval Europe* (2023-2025, PRIN 2022XRFSXR), che si pone come obiettivo la creazione di un *corpus* annotato dei testi in Francese d’Italia.

The contribution presents the first results of the *project FrIngE - The French in/of Italy: Code-MixiNG in Medieval Europe* (2023-2025, PRIN 2022XRFSXR), which aims to create an annotated *corpus* of French texts written in Italy.

## PAROLE CHIAVE:

Francese d’Italia – franco-italiano – annotazione linguistica – *Pyrrba* – Digital Humanities.

## KEYWORDS:

French of Italy – franco-italian – Linguistic Annotation – *Pyrrba* – Digital Humanities.

## 1. *Introduzione*

Negli ultimi anni è stata elaborata, soprattutto sulla scia degli studi di area anglofona, la nozione di ‘francofonia medievale’ per mettere in evidenza il fatto che, per tutto il Medioevo, varietà oitaniche erano ampiamente diffuse in un vasto territorio che comprendeva l’attuale Regno Unito, tutta l’Europa occidentale e il

\* Questo articolo è il primo frutto del progetto *The French in/of Italy: Code-MixiNG in Medieval Europe (FrIngE Project)* (2023-2025, PRIN 2022XRFSXR). Ci è sembrata buona l’idea di scrivere tutti insieme un intervento che fotografasse il momento in cui ci troviamo, per cominciare a fissare sulle pagine i problemi che abbiamo incontrato e le soluzioni che abbiamo via via trovato durante le nostre riunioni settimanali a distanza, pur nella fluidità del lavoro ancora in corso. Tutte e tutti hanno lavorato in stretta collaborazione, leggendo e modificando ripetutamente ogni singola riga dell’articolo. Dal momento, tuttavia, che le regole accademiche ci impongono di farlo,

Mediterraneo orientale<sup>1</sup>. La presenza del francese si manifesta sia nella circolazione di testi e manoscritti, sia nella produzione testuale autoctona: al di fuori della Francia esisteva un'ampia e spesso originale cultura letteraria in francese e, in particolare, nella Penisola italiana questa cultura è stata talvolta scritta o trasmessa in forme linguistiche ibride, scaturite dalla commistione o dalla commutazione di codice del francese con le varietà volgari locali. La sua natura mescolata ha spesso fatto sì che questa lingua fosse considerata come deteriore e, quindi, indegna di attenzione: tale sorte è toccata, a lungo, a un insieme di testi individuati come 'franco-italiani' (cfr. *infra*, § 2).

Dato il ruolo del francese come lingua 'transnazionale', molto più che come lingua di 'prestigio culturale' (che è stata a lungo la tradizionale chiave interpretativa degli studi), il progetto *FrIngE* si propone di studiare il 'francese in Italia', vale a dire la lingua usata nella copia in Italia di testi francesi composti altrove, e il 'francese d'Italia', testimoniato dai testi composti in francese da autori italiani. A livello diatestuale, dunque, sono oggetto dello studio le trascrizioni italiane di testi francesi realizzati altrove, i rifacimenti italiani di testi francesi e i testi francesi scritti da italiani (anche in Oriente). Le ultime due categorie coincidono con i cosiddetti testi 'franco-italiani' quando l'interferenza tra francese e volgari italiani è particolarmente marcata<sup>2</sup>. Dal punto di vista diatopico, sono inclusi l'intera area italoromanza e anche i territori dell'Oriente latino in cui siano attivi scriventi di origine italiana; l'arco cronologico si estende dal XIII al XV secolo (cfr. *infra*, § 3.1). Lo studio si propone di rispondere a queste domande:

- Ci sono caratteristiche linguistiche comuni all'intero *corpus*? o, più precisamente, ci sono particolari fatti di code-mixing che tendono a caratterizzare questo *corpus*?
- Come sono distribuiti nel *corpus* specifici fenomeni di code-mixing (ad esempio, casi di italianismo morfologico o lessicale)?
- All'interno del *corpus* complessivo, è possibile isolare la lingua della così detta letteratura 'franco-italiana' come varietà a sé stante?

elenchiamo di seguito le principali autrici e autori dei singoli paragrafi: 1 Francesca Gambino; 2 Andrea Beretta; 3 Elisa Guadagnini; 3.1 Francesca Gambino; 3.2 Andrea Beretta (pp. 293-294), Rachele Fassanelli (pp. 295-296); Carlo Rettore (per il paragrafo relativo all'annotazione degli alterati, p. 294); 3.2.1 Rachele Fassanelli (p. 297), Andrea Beretta (p. 298); 3.2.2 Francesca Gambino; 4 Manuel Favaro; 4.1 Jacopo Fois; 4.2 Floriana Ceresato; 5.1 Floriana Ceresato; 5.2 Andrea Beretta; 5.3 Federico Guariglia (pp. 310-311) e Matteo Parodi (p. 312); 5.4 Rachele Fassanelli; 6.1 Giacomo Costa (pp. 314-315, 319-321), Floriana Ceresato (pp. 317-318), Andrea Beretta (p. 316); 6.2 Francesca Gambino; 6.3 Manuel Favaro; 6.4 Elisa Guadagnini; 7 Andrea Beretta.

<sup>1</sup> Cfr. McDonald – Suleiman 2011, Kleinhenz – Busby 2011, Morato – Schoenaers 2018, Zinelli – Lefèvre 2021; i progetti di Simon Gaunt *Medieval Francophone Literary Culture Outside France* (<http://www.medievalfrancophone.ac.uk>) e *The Values of French* (<https://tvof.ac.uk/>).

<sup>2</sup> Queste tipologie testuali, riprese dalla banca dati *RLALFrI*, [www.rialfri.eu](http://www.rialfri.eu), sono da intendersi come categorie ideali: la realtà dei testi rende spesso labile il confine tra le une e le altre.

Siamo consapevoli di una serie di problemi da affrontare che, se considerati isolatamente, sembrano scoraggiare l'iniziativa di studiare il francese in/di Italia (la definizione del *corpus*, la descrizione delle modalità di incontro di due sistemi diversi che non si cristallizzano mai in una vera e propria lingua di contatto; cfr. *infra*, §2); tuttavia, abbiamo ritenuto che analizzare dal punto di vista linguistico, e specificamente lessicale, testi medioevali poco frequentati potesse comunque offrire risultati interessanti alla comunità scientifica. Si tratta, in effetti, di opere di difficile comprensione per il lettore moderno; oltre a questo, i dati già disponibili sono dispersi (nei glossari dei singoli testi, in articoli di non sempre facile accessibilità), talvolta gli studi mancano del tutto.

Il progetto *FrIngE* sta, quindi, avviando lo studio del francese d'Italia, considerando l'intero *corpus* di testi francesi nella Penisola, nel complesso quadro del plurilinguismo dell'Europa medievale<sup>3</sup>, con particolare attenzione alle regioni settentrionali.

La complessità che caratterizza la produzione di testi francesi in Italia richiede la definizione di parametri specifici di ricerca: l'introduzione della tecnologia digitale e di metodi computazionali consente forme di analisi e visualizzazione multiparametrica dei testi – sfaccettata, dinamica, statisticamente analizzabile – che rappresentano da sempre un'aspirazione teorica dei filologi, ma che, in passato, erano tecnicamente impossibili. In questo modo, il progetto *FrIngE* pone le basi per una futura analisi linguistica completa del francese usato in Italia: la disponibilità di dati grafico-fonetici, morfologici, sintattici e lessicali costituisce il fondamento su cui effettuare analisi approfondite di ogni testo incluso nel *corpus*, così come di specifici gruppi di testi e dell'insieme del *corpus* nel suo complesso.

Prima di intraprendere la descrizione dell'insieme e della sua annotazione linguistica, sarà opportuno dare qualche coordinata utile a inquadrare il senso del lavoro che proponiamo nell'orizzonte della tradizione degli studi.

## 2. *L'appropriazione della tradizione antico-francese e la nozione di 'franco-italiano'*<sup>4</sup>

Nel 1856, François Guessard, assieme a Léon Gautier e Henri Michelant, attraversò le Alpi per consultare i testimoni manoscritti di opere francesi conservati a Venezia, e si trovò a constatare una «*affreuse corruption du langage*», dovuta al fatto che «*presque tous les textes d'anciens poèmes français qui se trouvent à Venise ont subi le même genre d'altération: la langue en est plus ou moins italianisée, et par suite la mesure des vers y est horriblement rompue*»<sup>5</sup>.

<sup>3</sup> Cfr. Minervini 1996, Kleinhenz – Busby 2011.

<sup>4</sup> Questa sezione sintetizza quanto argomentato analiticamente da Andrea Beretta in Gambino – Beretta 2023: XXI-XLV.

<sup>5</sup> Guessard 1857: 395.

Qualche anno dopo, Gaston Paris, nella *Histoire poétique de Charlemagne*, avrebbe proposto una tipologizzazione di questa documentazione in lingua ‘più o meno italianizzata’ distinguendo tra le ‘copie’, i ‘rimaneggiamenti’ e le ‘creazioni originali’, che si porrebbero in una successione anche cronologica<sup>6</sup>. Tale ricostruzione – che fu ripresa anche, in Italia, da Giulio Bertoni e Antonio Viscardi<sup>7</sup> – non corrisponde però ai dati documentari: in questa sede, basterà ricordare che l’*Enanchet* e le *Estoires de Venise* di Martin da Canal, opere entrambe duecentesche e quindi tra le più antiche conservate, sono ‘originali’, vale a dire composte da autori italiani (che eventualmente fanno uso di fonti latine preesistenti). Nel tempo, sono state avanzate diverse altre proposte di sistemazione e interpretazione complessiva del fenomeno della scrittura, da parte di scriventi italiani, di testi in francese: in generale, tuttavia, esse sembrano fondarsi sul presupposto dell’esistenza del ‘francese’ e dell’‘italiano’ come componenti distinte, da un lato, e dall’altro sulla possibilità di individuare diverse tipologie (linguistiche e testuali) in base alle modalità di combinazione tra queste due componenti<sup>8</sup>.

Radicalmente diversa da questo approccio è la lettura proposta da Fabio Zinelli, che afferma che «ce n’est pas la France qui a donné le français au monde médiéval mais plutôt celui-ci qui se l’est approprié à son propre compte»<sup>9</sup>. Rispetto alla Penisola italiana, l’inizio del Duecento segna l’intensificarsi di scambi con l’Oriente mediterraneo da parte, in primo luogo, delle Repubbliche affacciate sul mare, Venezia, Pisa e Genova: questa ‘internazionalizzazione’ sembra tradursi anche, rispetto alla cultura scritta, in una appropriazione dei generi letterari antico-francesi, vale a dire composti nella lingua più diffusa dell’epoca. Come ha scritto Gianfranco Folena<sup>10</sup>:

l’espansione del francese in Italia è concomitante all’espansione del francese e dell’italiano in Levante nell’età delle Crociate, degli stati latini di Levante e dell’Impero latino d’Oriente. C’è una relazione tra questi due fatti? Perché la cultura francese in Italia si afferma particolarmente nelle aree, soprattutto in quella veneta e anche in quella lombardo-emiliana, e significativamente in Toscana nell’area pisana, direttamente interessate e coinvolte nelle imprese militari-commerciali d’Oriente? L’esperienza francese sembra legata a una dimensione avventurosa-internazionale e gotico-orientaleggiante che non trova rispondenze nella cultura italiana centro-meridionale con l’eccezione di Pisa (del tutto diversa naturalmente è la situazione del francese nell’Italia meridionale sotto i Normanni o gli Angioini). Il francese s’è affermato in Italia non tanto per contiguità, come adstrato (così più tardi in Piemonte), né per sovrapposizione, come superstrato (come nel caso di Normanni o Angioini), ma piuttosto per scelta culturale, per il suo volto internazionale, come metastrato.

<sup>6</sup> Paris 1865: 163.

<sup>7</sup> Cfr. rispettivamente *La guerra d’Attila*: IX-X e Viscardi 1941: 37-38.

<sup>8</sup> Si ricordano in particolare i lavori di Aurelio Roncaglia, Lorenzo Renzi e Marcello Barbatto: per una descrizione e un commento di queste e altre proposte interpretative si rinvia al lavoro di Beretta già citato (Gambino – Beretta 2023: XXXII-XL).

<sup>9</sup> Zinelli 2020: 199.

<sup>10</sup> Folena 1990: 272-273, n. 5.



Se si assume questo punto di vista, il fenomeno 'franco-italiano' si svincola dalla necessità di far costantemente riferimento ai due poli, distinti, del francese e dell'italiano, per avvicinarsi invece al concetto di *koinè*, come già proposto da Cesare Segre<sup>11</sup>.

Come ha sostenuto Francesca Gambino per il *Bovo udinese*<sup>12</sup> – ma l'osservazione è applicabile in generale ai testi cosiddetti 'franco-italiani' – sembra opportuno adottare per l'analisi il metodo storico-linguistico tradizionale, al fine di (tentare di) ricostruire la trafila lungo la quale un determinato fenomeno osservato è giunto al testo in oggetto. Andrà di volta in volta valutato, infatti, se tale fenomeno si spieghi più probabilmente in virtù del 'contatto' linguistico oppure in altro modo – con riferimento, per es., alle abitudini grafiche e alla cultura dello scrivente: se egli ha una cultura notarile o è un chierico, sarà certamente attivo l'influsso del latino; se, invece, ha una formazione mercantile, è più probabile che la sua scrittura rifletta fatti di lingua o i nascenti usi grafici volgari. La 'linguistica di contatto', certamente utile per comprendere alcuni fatti (si pensi, per es., alla nozione di 'transgrafemizzazione')<sup>13</sup>, deve insomma accompagnarsi a una solida base analitica fondata sulla storia della lingua, che consenta l'analisi minuta delle tradizioni scritte medievali<sup>14</sup>.

### 3. Il corpus DiFrI

Lo studio del francese d'Italia così come sopra definito deve necessariamente fondarsi sul *corpus* dei testi riconducibili alle due grandi fattispecie enunciate, vale a dire i testi francesi composti da italiani e quelli copiati da italiani (ricordando, come detto, che si tratta di due poli ideali più che di due insiemi distinti, separati da un confine netto).

Come è noto, la rivoluzione digitale ha permesso l'inclusione del piano documentario entro la sfera propriamente lessicografica. Per riprendere un celebre passo di Bernard Quemada<sup>15</sup>:

Le temps n'est plus où il fallait se battre pour imposer le recours aux ordinateurs. Au fil des générations informatiques, ses implications (et ses avantages) gagnent en étendue et en diversité et le projet dictionnaire s'en retrouve radicalement modifié. [...] le plus im-

<sup>11</sup> Cfr. Segre 1995: 644.

<sup>12</sup> Cfr. Gambino 2016.

<sup>13</sup> Cfr. Renzi 1970: 63: la 'transgrafemizzazione' è il passaggio di un tratto grafico da una *scripta* all'altra, come avviene per es. nella forma *chant* 'quando', in cui il fonema /k/ iniziale è reso non con *q/qu*, come accade di norma in francese antico, ma con il digramma *cb* tipico delle varietà italo-settentrionali medievali.

<sup>14</sup> Cfr. Zinelli 2018.

<sup>15</sup> Quemada 1987: 234.

portant est sans doute la possibilité de disposer d'une structure à deux niveaux. En aval, le niveau proprement dictionnaire procédant du niveau lexicographique situé en amont. En amont, ces réalisations lexicographiques [...] sont des bases de données lexicographiques, conçues sur le modèle des bases relationnelles ouvertes, dont le rôle est de proposer le plus vaste choix d'éléments ou de données sur les mots et leurs emplois.

La banca dati testuale, insomma, si connota a tutti gli effetti come un elemento dello strumento lessicografico: questa consapevolezza ha guidato le scelte fondamentali riguardanti la composizione del *corpus* e la sua annotazione.

### 3.1. *Consistenza del corpus*

Il nucleo centrale del *corpus DiFrI* è costituito dalle opere originali scritte in francese da autori italiani tra il secondo quarto del XIII e la fine del XV secolo. All'epoca la Penisola era uno dei più importanti *partners* commerciali della corona francese, che controllava direttamente parte dell'Italia meridionale: le relazioni economiche e politiche favorirono l'afflusso di materiale artistico, letterario e linguistico francese e occitano. Dal punto di vista diatopico, è stata presa in considerazione l'intera area italo-romanza, con l'aggiunta dell'Oriente latino (per quanto riguarda gli scrittori italiani). Le opere selezionate secondo questi criteri sono abbastanza eterogenee da un punto di vista linguistico, dato che si va dal francese estremamente corretto di Brunetto Latini alla lingua pesantemente influenzata dai volgari italo-settentrionali della *Geste Francor*, nella quale le varietà francesi e italiane coesistono in un equilibrio dinamico. Questa diversificazione tra gli esiti linguistici possibili ci è parsa, tuttavia, un aspetto positivo, in quanto risultano rappresentate e indagabili in questo modo molte differenti varietà di 'francese d'Italia'.

Ai testi selezionati con questo primo criterio, vale a dire le opere originali di redazione italiana, si sono aggiunti alcuni rifacimenti di testi francesi realizzati in Italia che risultavano particolarmente interessanti per motivi linguistici. Rientrano in questo insieme, tra gli altri, la copia della *Chanson de Roland* conservata alla Biblioteca Nazionale Marciana dal manoscritto francese 4 (*Roland V4*), la versione del *Foucon de Candie* trådita da un altro manoscritto marciano, il francese 20, e poi ancora il *Bovo d'Antona udinese* (Udine AC Fondo Nuovi manoscritti 736.28). La *Geste Francor* (Venezia BNM fr. Z 13) è stata naturalmente considerata per intero e non solo per le parti presumibilmente originali.

Del *corpus DiFrI* fanno, dunque, parte innanzitutto le opere della letteratura 'franco-italiana' o 'franco-veneta' propriamente detta, che comprende soprattutto *chansons de gestes* e altri testi narrativi in versi (romanzi e racconti agiografici) caratterizzati da un elevato grado di ibridismo linguistico<sup>16</sup>. Si tratta di testi localiz-

<sup>16</sup> Cfr. Holtus – Wunderli 2005 e Zinelli 2016. Per i riferimenti bibliografici dei testi citati si rinvia alle schede e alla bibliografia del *RLALFrI*; per un commento critico di queste etichette tradizionali, cfr. *supra*, § 2.

zabili nel Nord-Est della Penisola italiana (Lombardia, Emilia, Veneto) che potevano talvolta essere fruiti anche attraverso la recitazione orale e sono stati scritti da italiani in una lingua letteraria artificiale plasmata nel tempo allo scopo di rendere più comprensibili le opere francesi al pubblico autoctono, senza rinunciare al prestigio del modello d'Oltralpe.

Complementare a quest'area è il Nord-Ovest della Penisola, con il *Livre du Chevalier Errant* di Tommaso di Saluzzo, la *Bataille de Gamenario* e altre testimonianze di francese d'Italia quattrocentesche, che ci giungono soprattutto dal Piemonte.

Una diversa tipologia testuale è rappresentata dalle opere in prosa di scrittori italiani che hanno usato il francese per rivolgersi a un pubblico più vasto, dal momento che – come spiega nelle *Estoires de Venise* Martin da Canal – «Lengue françoise cort parmi le monde»<sup>17</sup>. Il racconto di viaggio di Marco Polo, nella cui lingua si riconoscono almeno tre diversi strati linguistici (l'idioletto dell'autore, gli usi scrittori di Rustichello e l'influsso dei copisti), e i trattati di falconeria *Moamin* e *Ghatrif* tradotti dall'arabo attraverso un intermediario latino da Daniele Deloc di Cremona presentano fenomeni di mescolanza in parte simili a quelli che si rinvengono nella letteratura 'franco-italiana'. Gli stessi fenomeni di interferenza, pur se meno invasivi, compaiono nella *Compilazione* di Rustichello da Pisa<sup>18</sup>, la cui lingua subisce in più l'interferenza di alcune *scriptae* toscane. In altre opere, invece, l'influenza dei volgari italiani si riduce a ben poco. La lingua del trattato di igiene di Aldobrandino da Siena e dell'enciclopedia di Brunetto Latini, che scrivevano in francese mentre si trovano in Francia, è, infatti, paragonabile a quella utilizzata dagli autori madrelingua dell'epoca. Anche il veneziano Martin da Canal scrive intorno al 1275 in un buon francese, venato però da qualche tratto d'Oltremare.

Al contesto mediterraneo ci conduce pure uno scrittore come Filippo da Novara, attivo a Cipro: gli italianismi che si rinvengono nella sua opera potrebbero non essere dovuti tanto alla sua origine, quanto al fatto che erano ormai vocaboli diffusi nella *koinè* d'Oltremare creatasi in seguito ai viaggi e alle migrazioni dei popoli romanzi nel Mediterraneo, durante il periodo delle crociate. Nel francese dei regni di Levante tratti provenienti da diversi dialetti si coniugano con alcuni occitanismi e con un lessico caratterizzato da prestiti italiani e catalani, nonché arabi e bizantini per il contatto con le popolazioni locali. Marco Polo, Martin da Canal e persino Brunetto Latini accolgono nelle loro opere alcuni di questi vocaboli. Le intersezioni tra il lessico degli autori italiani che usano il francese per raggiungere un pubblico internazionale e il vocabolario d'Oltremare<sup>19</sup> rendono opportuno includere nel nostro *corpus* anche Filippo da Novara, che, oltre a corrispondere perfettamente al criterio di 'autore italiano che scrive in francese', ci consente di tenere sullo sfondo questo contesto mediterraneo più ampio.

<sup>17</sup> *Estoires de Venise*, 1, I.

<sup>18</sup> Cfr. *Romanzo arturiano di Rustichello da Pisa*.

<sup>19</sup> Cfr. Zinelli 2016.

Un'ulteriore area geografica italiana che deve essere considerata nella storia del francese d'Italia è costituita da Napoli e dal Regno di Sicilia. I dialetti del Sud Italia conservano tracce del passaggio dei Normanni e degli Angioini<sup>20</sup>: a Napoli il francese è stato per un certo periodo la lingua della corte e dell'amministrazione<sup>21</sup>, con una produzione testuale di tipo sia documentario che letterario. Per il momento abbiamo pubblicato in rete l'*Ystoire de li Normant* di Amato di Montecassino<sup>22</sup>, traduzione di una perduta cronaca latina, opera che presenta fenomeni di ibridismo simili a quelli che si riscontrano nei testi scritti al Nord, ma che si distingue per il fatto che il volgare italiano che interferisce con il francese non è di tipo settentrionale o toscano ma meridionale e può, dunque, fornire proficue occasioni di comparazione e integrazione.

Il *corpus DiFri* è integrato con le opere iconografiche ed epigrafiche italiane legate alla Francia, i motti nobiliari, gli epitaffi e altre testimonianze estemporanee mai considerate prima. Sono, infine, inclusi carotaggi di manoscritti di opere francesi copiate in Italia: pur se con un grado di ibridazione inferiore, essi possono presentare fenomeni di interferenza grafici, fonetici e morfologici dovuti alla prassi più o meno consapevole dei copisti medievali di attualizzare la lingua del testo che trascrivevano<sup>23</sup>. Per quest'ultima via si riesce a valorizzare l'apporto di un'altra area importante secondo la prospettiva panitaliana che ci siamo posti, quella tosco-ligure<sup>24</sup>. I circa cinquanta codici trascritti nelle prigioni genovesi da copisti per lo più pisani presentano una *scripta* abbastanza uniforme, ma hanno relativamente poche novità lessicali, per cui essi saranno utilizzati solo in modo sporadico per eventuali raffronti. Dal punto di vista lessicale, in generale, risultano interessanti soprattutto due delle tre tipologie testuali di cui è composta la letteratura 'franco-italiana' secondo lo schema di Bertoni e Viscardi, adottato poi anche dal *RLALFri*<sup>25</sup>: nelle semplici trascrizioni, infatti, l'interferenza linguistica si limita per lo più al livello grafico-fonetico e morfologico<sup>26</sup>.

Dal punto di vista cronologico il *corpus DiFri* si estende dal più antico testo letterario scritto in francese nell'Italia medievale giunto fino a noi, l'*Enanchet* (databile al secondo quarto del Duecento, 1226-1252), fino al passo in francese incluso ne *Lo balzino* di Rogeri da Piacenza (1497-1498); le opere letterarie interamente in francese più recenti sono l'*Huon d'Auvergne* padovano (fine XIV – inizio XV sec.) e l'*Aquilon de Bavière* di Raffaele da Verona (1379-1407). Sebbene

<sup>20</sup> Cfr. Varvaro 1981 e Valenti 2014.

<sup>21</sup> Cfr. Minervini 2015.

<sup>22</sup> Cfr. <https://www.rialfri.eu/texts/amatodiMontecassino|001>.

<sup>23</sup> Cfr. Varvaro 1996: 533. Sulle diverse tipologie di commutazione di codici nei testi antichi cfr. anche la recente messa a punto di Baglioni 2016: 9-35.

<sup>24</sup> Cfr. Cigni 2011.

<sup>25</sup> Cfr. *supra*, § 2.

<sup>26</sup> Cfr. Morlino 2015: 32.

gli studi sul franco-italiano facciano convenzionalmente terminare il fenomeno con Raffaele da Verona, si osserva che la 'moda' del francese rimane abbastanza vivace anche nel corso del Quattrocento: ne sono testimoni i *Tituli del castello della Manta* (1416-1426), alcuni motti nobiliari (comuni per tutto il XV sec.; il motto del marchese di Vigevano Gian Giacomo Trivulzio *Ne t'esmai* 'non perderti d'animo, non temere' è datato addirittura al 1503 e 1509), iscrizioni, epitaffi e le Scritte avventizie del manoscritto Ashburnham 1076 (XIV-XV sec.).

### 3.2. Teoria e criteri dell'annotazione<sup>27</sup>

Per l'annotazione del *corpus DiFrI* si è deciso di avvalersi del *software Pyrrha*, sviluppato dall'École Nationale des Chartes per l'analisi del francese antico (cfr. *infra*, § 4): il lemmario segue quello dell'*Altfranzösisches Wörterbuch* di Adolf Tobler e Erhard Lommatzsch (T-L, eventualmente integrato), l'attribuzione della *Part-of-Speech* (POS) si riferisce allo schema sviluppato nel contesto del progetto *Cat-tex*<sup>28</sup>. L'adozione di *Pyrrha*, tuttavia, si è accompagnata da subito alla necessità di mettere a punto una serie di modifiche e aggiustamenti.

La descrizione di una varietà linguistica ibrida pone, come si sa, problemi peculiari: la logica contrastiva (propria di qualunque analisi linguistica) deve necessariamente evolvere in una logica *fuzzy*, per cui tra i due poli opposti sono ammissibili gradi intermedi di verità (cfr. *infra*). L'analisi lessicografica, in particolare, sia che la si consideri 'a monte' (banca dati testuale) sia che la si consideri 'a valle' (dizionario), si trova confrontata a diverse questioni, problematiche tanto sul piano teorico quanto su quello pratico.

Quando si annota e descrive una varietà linguistica ibrida, deve essere tenuta ben presente la natura convenzionale del lemma, che è appunto la forma del paradigma (verbale o nominale) con cui convenzionalmente si registra il lessema (infinito per i verbi, singolare per i sostantivi, singolare maschile per gli aggettivi). Per il *DiFrI*, dato un certo etimo, si adotta il lemma nella forma in cui compare nel T-L (ma cfr. *infra*); si adotta un lemma 'italiano' (secondo la forma di *TLIO*) soltanto se nessuno strumento lessicografico del francese registra un esito dalla base etimologica di interesse. Questo significa che l'attribuzione del lemma prescinde dalla forma grafica (o grafico-fonetica) e dalla semantica dell'occorrenza: la forma *che* del pronome relativo è lemmatizzata *que*<sup>2</sup> (e la forma *che* della congiunzione è lemmatizzata *que*<sup>4</sup>); il fatto che un'occorrenza in forma *che* del relativo (o della congiunzione) costituisca un italianismo è annotato soltanto dal tag «SPEC=it» che estende la descrizione morfologica della parola (cfr. *infra*, § 3.2.1.).

<sup>27</sup> Qui e in seguito per annotazione si intende l'aggiunta di informazioni specificamente linguistiche. Per tipi diversi di annotazione cfr. *infra*, § 6.1.

<sup>28</sup> Cfr. Guillot – Prévost – Lavrentiev 2013a e Guillot – Prévost – Lavrentiev 2013b.

Tale scelta permette da un lato la piena interoperabilità del *corpus* e del dizionario con gli altri strumenti del francese, dall'altro consente di isolare in modo netto i casi di 'italianismo irriducibile' (per così dire), vale a dire i casi in cui una certa forma risulta assolutamente aliena rispetto al sistema linguistico oitanico (considerato nel suo complesso, rispetto allo stato attuale delle conoscenze).

Punto centrale adottato in ogni aspetto della nostra lemmatizzazione è il T–L, che viene considerato come primo riferimento per l'apposizione del lemma, come già detto *supra*. Qualora la forma non ricorra nel T–L, il lemmatizzatore farà ricorso al *DÉAF*; se, ancora, il *DÉAF* non contribuisce a risolvere il problema, ulteriori strumenti sono costituiti dal Godefroy e dal *DMF* (il dizionario di Matsumura, a sua volta debitore di T–L e *FEW*, può essere un ulteriore utile punto di confronto). Se, infine, si è di fronte a italianismi per i quali non sono attestati lemmi corrispondenti negli strumenti per l'antico francese, il riferimento principe è il *TLIO* (affiancato dal *GDLI*)<sup>29</sup>: l'adozione di un lemma 'italiano' serve a evitare la creazione *ad hoc* di una etichetta di forma francesizzante che risulterebbe non attestata altrove. Come sempre, tuttavia, il lemma ha esclusivamente la funzione di individuare univocamente il tipo lessicale in oggetto; l'informazione linguistica che la parola è un italianismo è ancora affidata alla presenza della proprietà «SPEC=it» (cfr. *infra*, § 3.2.1). Altro strumento di riferimento per la lemmatizzazione è il *Thesaurus Linguae Latinae*, cui fare ricorso nel caso di eventuali inserti latini, non così rari nel *corpus*: anche in questo caso, l'adozione di un lemma latino solleva l'annotatore dall'onere di dover selezionare o costruire *ad hoc* un lemma (di forma) francese e ha l'unico scopo di facilitare l'individuazione unica del tipo lessicale. Come nel caso degli 'italianismi', il dato linguistico che la parola è un xenismo latino è veicolato dall'attribuzione dell'etichetta «SPEC=lat»<sup>30</sup>.

Nel caso specifico degli alterati, qualora una parola formata per alterazione non sia registrata nei dizionari di riferimento si ricostruisce un lemma composto dalla base e dal suffisso alterante. È questo il caso, per es., di *Canzoni francesi*, XVI – 35, v. 2, in cui è attestata la forma diminutivo-vezzeggiativa – evidentemente italiana – *albeta*: essa è lemmatizzata *aubete*, da *aube* 'alba' e il suffisso *-ete* (nella forma in cui esso compare a lemma nel T–L)<sup>31</sup>; il fatto che la parola sia un italianismo è dichiarato dall'etichetta «SPEC=it».

Analogamente a quanto avviene per gli alterati, si ricostruisce il lemma per gli avverbi in *-ment* che non hanno una voce nel T–L. I numeri romani presenti a

<sup>29</sup> Per i venezianismi lo strumento di riferimento è adesso il *VEV*.

<sup>30</sup> Cfr. *infra*, § 3.2.1.

<sup>31</sup> Cfr. T–L s.v. *aube* e, per il suffisso, cfr. a titolo d'esempio *amorete*, *bergerete*, *bochete*, *espinete*, *erbeta*, *pucelete*, *rosete*, *violete*<sup>1</sup> e *passim*. Nel caso particolare, *aubete* in effetti non è un hapax ed è attestato in lessicografia: Gdf s.v. *aubete*<sup>1</sup> riporta un paio di esempi piuttosto tardi (non anteriori al Quattro-Cinquecento). Notevole, comunque, l'assenza di una voce *albeta* anche nel lemmario del *TLIO*: l'esempio in questione potrebbe in qualche modo fornire una prima attestazione dell'alterato in ambito sia italiano che francese.

testo sono lemmatizzati. Le forme pronominali *moi*, *me* e *mi* sono ricondotte al lemma «moi» (e non «je», secondo l'uso del T-L); per *ni* «ADVneg» o «CONcoo», il lemma sarà «ni1»; il possessivo di terza persona è annotato come lemma «suen» (seguendo l'entrata del T-L) e non «son4» (come proposto da *Pyrrha*)<sup>32</sup>; *lequel* è ricondotto al lemma «quel1», *messire* e *mesire* a «monseigneur». Infine, la preposizione italiana *con* è lemmatizzata «con3» (ed etichettata «SPEC=it»), per distinguerla dagli omografi francesi.

Un breve approfondimento, infine, merita la questione dei nomi propri. Toponimi e antroponimi occupano spesso una sezione a parte di vocabolari e strumenti lessicografici (non di rado l'appendice), quando non siano da essi totalmente esclusi: fanno talvolta eccezione i casi di nomi comuni usati all'occorrenza (o esclusivamente) come nomi personali e, anche in questa eventualità, si richiede di solito un'attestazione particolarmente antica (magari 'prima') o comunque degna di rilievo per qualche motivo<sup>33</sup>. Tuttavia, come noto, il nome proprio si rivela particolarmente utile ai fini della storia della lingua per ragioni differenti<sup>34</sup>: può, ad esempio, conservare inalterati, e, dunque, attestare tratti linguistici antichi o, al contrario, riflettere fenomeni propri della lingua del territorio di appropriazione (questo soprattutto il caso che ci interessa: si pensi per es. a *Zarles*, *Belenzer*, *Zanevre*, a testimoniare l'avanzamento articolatorio dell'affricata palatale ad affricata dentale). In attesa dell'allestimento di uno studio *ad hoc* dedicato ai nomi propri all'interno del progetto, si è deciso di promuovere a lemma la forma occorrente nei singoli luoghi, senza ricondurla a un lemma convenzionale unico (che oscillerebbe a seconda del repertorio onomastico di riferimento): così, a titolo esemplificativo, si troveranno i lemmi *Carles*, *Charles*, *Çarlemagne*, *Çarlemaine*, *Çarles*, *Çarlons*, ecc., non ricondotti a una forma onomastica unica. Tra i toponimi, un caso particolare è rappresentato dalle forme sintagmatiche che *Pyrrha* divide automaticamente nei singoli membri, ciascuno dei quali è poi ricondotto, nella fase di post-correzione, alla rispettiva categoria di appartenenza: *Asie-la-Capelle* (Aix-la-Chapelle) sarà dunque annotato come una successione di «NOMpro», «DET-def» e «NOMcom».

L'annotazione del *corpus* comprende non soltanto l'attribuzione del lemma, ma anche la specificazione del POS [*Part-of-Speech*] e la descrizione morfologica.

Per quanto concerne il POS, per cui *Pyrrha* segue il sistema *Cattex*, è utile segnalare che per il *corpus DiftI* si è deciso di seguire la prassi di annotazione del *corpus TLIO* e, dunque, si etichetta come «NOMcom» il secondo elemento delle locuzioni avverbiali del tipo *par certain*, *en bas* e simili; «NOMcom» sono anche gli infiniti sostantivati e «ADJqua» i participi passati con funzione aggettivale. Ri-

<sup>32</sup> Sebbene *Pyrrha* segua, in linea di principio, il lemmario del T-L, capita che singoli lemmi si presentino in forme diverse.

<sup>33</sup> Cfr. Norme per la redazione del *TLIO* e Beltrami 1999.

<sup>34</sup> Cfr., tra gli altri, Pfister 1999, Marcato 2016 e Sanfilippo 2016: 52.

spetto a toponimi e antroponimi, *Pyrrha* contrassegna come ‘nome proprio’ i nomi di popolo e divinità (quando usato, quest’ultimo, al singolare), altrove considerati nomi comuni, mentre mantiene la categoria «NOMcom» per i nomi comuni personificati e assunti come antroponimici. Abbiamo ritenuto opportuno etichettare come «NOMcom» e non come «NOMpro» gli etnonimi (per es. *Saracins, Français, Paiens, Cristiens, ...*); allo stesso modo, *Dieu, Damnediex* e simili sono «NOMcom», mentre *Jesus* è «NOMpro». Ancora, rispetto a locuzioni come *en estant* (e simili), l’elemento *estant* è da annotare come «VERppa», salvo quando sia accompagnato da avverbio o aggettivo (*en mon estant*, ad esempio), nel qual caso riceve il POS «NOMcom».

Sempre in relazione alla questione dei POS, ma rispetto al piano della tokenizzazione, si è deciso di separare ogni forma composta presente nel *corpus* (a partire dalle preposizioni articolate): per es., *au* è analizzato come «a# #u»: *a#* è descritto come lemma «a3», POS «PRE», «MORPH=empty»; *#u* è ricondotto al lemma «le», POS «DETdef», e MORPH «NOMB.=s | GENRE=m | CAS=r»<sup>35</sup>.

Come si nota negli esempi appena proposti, l’annotazione prevede anche un campo dedicato alla morfologia. Da questo punto di vista, ci siamo distaccati dalle prassi di *Pyrrha* rispetto segnatamente a due aspetti. In primo luogo, abbiamo considerato per i vocativi la funzione sintattica e non quella morfologica, etichettandoli quindi come *cas sujet* anche se non rispondono alle norme grammaticali *standard*. In secondo luogo, rispetto ai pronomi riflessivi, abbiamo deciso di etichettare come *cas régime* il pronome riflessivo che assume la funzione di oggetto diretto nei verbi pronominali.

Per evitare ogni ambiguità, nei possessivi le persone sono numerate da 1 a 6 (e non 1-3 sing./pl.): per es., in *vostre terre*, il possessivo sarà etichettato in questo modo: lemma «vostre», POS «DETpos», MORPH «PERS.=5 | NOMB.=s | GENRE=f | CAS=r». Per i sostantivi come *arme* e *honor* si adotta l’etichetta «GENRE=f/m», a meno che non ci siano elementi che esplicitino il genere (come avviene, per es., in *ma arme, la vestre honor*). Si associa «PROind» a *on* e «PROimp» alle occorrenze del *si* passivante italiano (lemma «si2»).

Va segnalato, infine, che *Pyrrha* non consente l’analisi delle forme verbali composte: non prevede la funzione di ausiliare né la distinzione tra diatesi attiva e passiva. Per ragioni eminentemente pratiche, per il momento non siamo intervenuti sulle analisi restituite dal sistema e abbiamo, dunque, conservato la scomposizione nei singoli elementi delle forme verbali composte.

### 3.2.1. I tag «SPEC»

Vista la natura della documentazione, si è ritenuto opportuno introdurre nel-

<sup>35</sup> Il processo di separazione delle forme composte attualmente avviene in fase di revisione manuale; in un secondo momento questi dati potrebbero idealmente contribuire all’addestramento di nuovi modelli di tokenizzazione automatica (cfr. § 4).



l'annotazione alcune etichette *ad hoc*<sup>36</sup>, utili a individuare fenomeni peculiari e rilevanti che costituiscono specificità linguistiche e testuali: queste etichette rispondono al tipo «SPEC=» e sono inserite all'inizio del campo MORPH dell'annotazione.

È stato adottato il tag «SPEC=it» per individuare le occorrenze che presentano almeno un tratto riconoscibile come 'italianismo'.

Il fatto che si proceda ad annotare *a priori* come 'italianismi' determinati fenomeni, entro una banca dati che ha la finalità – tra l'altro – di permettere l'individuazione degli italianismi, e quindi di descrivere e valutare il tasso di ibridismo, di un testo o insieme di testi o (sotto)varietà linguistica, comporta una innegabile circolarità. Tuttavia, riteniamo che tale circolarità possa ritenersi virtuosa, alla luce di diverse considerazioni. In primo luogo, alcuni tratti (soprattutto grafici, eventualmente passibili di valore fonetico) sono da tempo riconosciuti come italo-romanzi. Il grado di certezza dell'etichettatura in questi casi è altissimo<sup>37</sup>. Inoltre, va sottolineato il fatto che l'etichettatura individua oggetti linguistici che risultano problematici per l'annotatore automatico, vale a dire oggetti linguistici assenti (o assai rari) nel *corpus* di addestramento del *software*: la loro marcatura segnala in ogni caso fatti che, in ottica contrastiva, risultano devianti rispetto alla norma (intesa naturalmente in senso puramente quantitativo). In prospettiva, immaginando in particolare una collaborazione a lungo termine con l'École des Chartes, che ha in progetto di lavorare alle diverse varietà oitaniche di epoca storica, sarà interessante studiare gli oggetti taggati da *DiFrI* e valutare se effettivamente rispondano a una logica diatopica (se risultino tratti caratteristici del 'francese d'Italia') oppure se si spieghino meglio rispetto ad altri parametri, osservabili soltanto entro l'orizzonte vasto delle varietà oitaniche coeve. Per portare un solo caso, sarà così possibile porsi la domanda del valore di forme dell'infinito come *aver*, ricorrenti nel *corpus DiFrI* ma attestate anche, per es., in diversi testi d'Oltremare. Avendo a disposizione dati sistematici, sarà possibile comprendere (o per lo meno porre correttamente la domanda) se questa terminazione sia un tratto tipico di determinate varietà diatopiche, se sia effettivamente un italianismo, se sia un tratto verosimilmente poligenetico (un italianismo nei testi francesi d'Italia e un latinismo grafico nei testi di traduzione, per dire), ecc.

In generale, come sempre, va considerata la differenza tra l'annotazione di un singolo fatto (testimoniato da una specifica occorrenza) e la descrizione di una varietà linguistica: a caratterizzare una varietà è la presenza simultanea di un fascio descrivibile di tratti; un singolo tratto di per sé può occorrere più volte, entro varietà complessivamente distinte. Il senso della taggatura come 'italianismo', per il momento (vale a dire in un momento in cui si sta lavorando all'annotazione del *corpus DiFrI* e non è ancora disponibile un *corpus* complessivo delle varietà oitani-

<sup>36</sup> Per il *tagset standard* cfr. Prévost (*et alii*) 2013.

<sup>37</sup> Cfr. Gambino – Beretta 2023: XLV-XLVIII.

che medievali), è questo: si tratta di rilevare tratti occorrenti nel *corpus* dei testi francesi d'Italia e che risultano devianti rispetto a quanto è attestato e noto altrove (cioè in tutti gli altri testi oitani). Tra questi tratti, certamente si troveranno fatti di 'sicura italianità' (cioè ascrivibili senza dubbio all'interferenza con l'italiano) e altri fatti, in cui il contatto con l'italiano costituisce una spiegazione parziale o probabile, a gradi diversi (in questo senso abbiamo parlato di 'logica fuzzy'). È noto, infatti, che alcuni tratti possono ad esempio essere comuni ad alcuni dialetti della Penisola italiana o della Francia (ad. es. il piccardo). Si tratta, quindi, di valutare caso per caso quale sia l'ipotesi più economica.

Dal punto di vista computazionale, sarà interessante verificare se (ed entro quali limiti) la macchina può essere addestrata a riconoscere questo tipo di tratto. Dal punto di vista linguistico, l'annotazione sistematica ed esaustiva di questi tratti permetterà una descrizione più accurata e profonda delle caratteristiche del francese d'Italia. Oltre ai fatti grafici, eventualmente passibili di valore fonetico, che rappresentano il piano linguistico attualmente meglio noto di questa varietà (o fascio di varietà), l'annotazione permetterà di cogliere meglio tratti morfologici e lessicali eventualmente caratteristici.

Casi di «SPEC=it» sono, per es., gli avanzamenti delle affricate (*civaler*); gli usi irrazionali/parassiti di /j/ comportanti dittonghi non attestati altrove in antico francese (a quanto risulta dagli strumenti e banche dati di riferimento; per es.: *oit*); uso di occlusive velari sorde in luogo di affricate palatali sorde non attestati in anglo-normanno o in piccardo (*scanper*); la prostesi di *a-* (*averbés*); l'utilizzo di forme verbali alla terza persona singolare in presenza di soggetto plurale (*li baron s'en rist*). Sono, poi, «SPEC=it» quelle parole riconducibili a lemmi solo italiani, vale a dire per le quali non risulta attestato un lemma corrispondente in area oitana: ne è un esempio *erese* 'erede', annotato con lemma – italiano – «erede» (cfr. *supra*, § 3.2) e, appunto, etichetta «SPEC=it».

Sono marcate come «SPEC=rin» le forme che sembrano adattamenti occasionali dovuti esclusivamente all'occorrenza in posizione di rima: il fenomeno appare molto diffuso in tutta l'epica, come per es. nella *Guerra d'Attila* di Nicolò da Càsola.

Come si è anticipato (cfr. *supra*, § 3.2), i latinismi sono etichettati come «SPEC=lat».

Abbiamo, infine, introdotto l'etichetta «SPEC=probl» per isolare singole occorrenze problematiche dal punto di vista linguistico o filologico: sarà descritto *infra* (§ 5.4) il caso di *scerp*.

### 3.2.2. La granularità dell'annotazione

Per descrivere alcuni tratti del franco-italiano sarebbe auspicabile raggiungere, già a livello di annotazione, una granularità più raffinata rispetto a quella attuale, marcando i possibili livelli di interferenza tra francese e volgari italiani settentrionali per le singole forme secondo il possibile schema:

	Grafia e Fonetica	Morfologia	Semantica
Forma francese	GF F	M F	S F
Forma ibrida (franco-italiana)	GF FI	M FI	S FI
Forma italiana	GF I	M I	S I

In una futura fase del lavoro cercheremo di suddividere le diverse componenti di una parola, distinguendo tra grafia e fonetica, morfologia e, infine, semantica, applicando questa griglia di analisi nel processo di annotazione allo scopo di ottenere nuovi risultati. L'utilità di tale operazione emerge dagli esempi che seguono.

Prendiamo il verbo *andare* alla terza persona singolare del *passé simple* / passato remoto. Le forme *standard* sono *al-a* in francese e *and-ò* in italiano, ma nel nostro corpus troviamo *al-ò* («El alò a'l çival, si le voloit furer», *Geste Francor, Chevalerie Bovo*, 4712)<sup>38</sup>, che ha la radice lessicale francese e la desinenza morfologica italiana, mentre in *and-ée* («Sia nostra gente, ke veçu l'a, andée», *Foucon de Candie*, 5128) l'origine dei costituenti è opposta, italiana e francese.

Gli aggettivi numerali ordinali latineggianti *decimonone*, *quatordecime*, *quintodecime*, *decimottave*, che compaiono nell'*Aquilon de Bavière* di Raffaele da Verona (cfr. 2, XXXVIII, 9, ecc.), rinviano alle forme italiane letterarie *decimonono*, *decimottavo*<sup>39</sup>, ma presentano nell'ultima sillaba la morfologia del francese.

Il sostantivo femminile *escherne*<sup>40</sup> corrisponde al maschile antico francese *eschar*. Il femminile in francese non è attestato e si tratta di un italianismo: nell'esempio del *Devisement dou monde* di Marco Polo<sup>41</sup> non sappiamo se la parola possa essere fatta risalire a Marco Polo oppure al pisano Rustichello<sup>42</sup>; altri esempi di femminile si rinvencono in alcuni manoscritti del *Tresor* di Brunetto Latini (*eschiernes*, 2, 67, 2)<sup>43</sup>, in un passo in cui l'alta percentuale di varianti può indicare la difficoltà dei copisti di fronte a una lezione avvertita come problematica ma che, proprio per questo, ha possibilità di essere originale. *Scherna* è forma attestata in Toscana e in diversi dialetti italiani; nel corpus *OVI* se ne rinvencono alcune occorrenze toscane, tra cui una in rima nel *Tesoretto* dello stesso Brunetto Latini (*scherna* 1465 : *taverna*); *esquerna* compare anche in occitano.

<sup>38</sup> I testi citati in questa sezione rinviano implicitamente alle edizioni e alla bibliografia del *RIAL-FrI*.

<sup>39</sup> Cfr. *TLIO* s.v.

<sup>40</sup> Cfr. *FEW* s.v. \**skirnjan*; *GDLI* s.v. *scherna*.

<sup>41</sup> «Il en fasoient si grant gas et si grant *escherne* qu'ele vindrent devant le Grant Chan. Et quant le Grant Chan oï ce, il dist maus a celz que gas en fasoient devant elz», *Devisement dou monde*, LXXIX, 9.

<sup>42</sup> Cfr. Zinelli 2016: 229.

<sup>43</sup> Cfr. Perugi 2008: 501.

Talvolta la forma di un vocabolo è francese, ma è la semantica ad essere italiana (o viceversa). Il verbo *avancer*, ad esempio, è ben attestato in francese antico<sup>44</sup>, ma mai con il significato di ‘essere rimanente di qualcosa, di un insieme, di una quantità’, che troviamo in «De quella carne qe li parse *avançer*, | El s’*à pris en seno a covoter*» (*Geste Francor, Orlandino*, prima metà XIV sec., Nord-Est, ed. Morgan 2009, 10970), e va, quindi, considerato un italianismo.

Il sostantivo femminile *bilotte* ‘biglia’ secondo il *FEW* deriva dall’antico franccone *bikkil* ‘dado’<sup>45</sup>. Un’altra etimologia possibile, tuttavia, potrebbe essere quella dal longobardo *palla* / franco *balla*, per cui è attestata anche la variante *balote*: cfr. *FEW* 1, 216b germ. BALLA, che per il fr. *balot*, *balete* attesta solo il significato di «gros paquet de marchandises, enveloppé de toile, pour être expédié» (*DÉAF* s.v. *bale*<sup>1</sup>). Il punto per noi è, comunque, che nelle *Estoires de Venise* di Martin da Canal la parola indica la ‘pallina di cera usata a Venezia nelle procedure di voto’ ed è, quindi, un venezianismo<sup>46</sup>. *Ballotta* è anche termine toscano usato nel corso del Medioevo a indicare egualmente la ‘pallina usata per le votazioni’ (cfr. *TLIO* s.v. *pallotta*<sup>2</sup>).

Il sostantivo *bise* indica in a.fr. un ‘uccello rapace’, *bisse* la ‘femmina del cervo’. Nel *DMF* *bisse*<sup>1</sup> è un termine tecnico dell’araldica<sup>47</sup>. Nel *Moamin* di Daniele Deloc, invece, il femminile *bisse* indica genericamente un ‘serpente non velenoso’ e, in particolare, nel sintagma *bixe escuiere* / *escuere*, la ‘biscia, serpente d’acqua’<sup>48</sup>. Si tratta, quindi, di un italianismo, presente soprattutto nel Nord Italia; il *TLIO* registra *biscia scudaia* in Umbria a Gubbio<sup>49</sup>.

Un altro vocabolo interessante è il sostantivo maschile *duc*<sup>50</sup>. Con la parola *dus* Martin da Canal indica, indipendentemente dal ruolo del sostantivo nella frase, il titolo che dall’VIII secolo ebbe il capo dello stato nella Repubblica di Venezia, il ‘doge’<sup>51</sup>, distinguendolo da *duc*, che nelle *Estoires de Venise* designa invece ‘colui

<sup>44</sup> Cfr. *FEW* s.v. *\*abantiare*, der. di *abante* ‘avanti’ (XXIV, 11b); T–L s.v. *avancier*.

<sup>45</sup> Cfr. *FEW* s.v. *\*bikkil*; Gdf s.v. *billote*; *DMF* s.v. *billotte*; cfr. inoltre T–L s.v. *bille*<sup>1</sup>; *DMF* s.v. *bille*; *TLFi* s.v. *bille*<sup>1</sup>.

<sup>46</sup> «Les nobles consillers firent celui jor meesme après none assenbler li Grant Consoil, et firent por chascun que au Consoil estoit une petite belote de cire; et dedens trente de ces belotes de cire avoit un petit de parchemin escrit en chascune, que disoit “lector”; et en fu prise por chascun une», *Estoires de Venise*, 2, CVIII. Cfr. *DiFrI* s.v. *belote*.

<sup>47</sup> Cfr. *DMF* s.v. *bisse*<sup>1</sup>: «HÉRALD. Serpent (couleuvre, et plus partic. encore couleuvre de Milan)»; cfr. anche T–L s.v. *bisse*; *TLFi* s.v. *bisse*<sup>1</sup>.

<sup>48</sup> «Et puis li donez a maingier de char d’une *bixe escuiere* – des celes qi vivent en terre es bois, et non pas des celes qe vivent es fossez, la qelle l’en appelle entre nos tortuche – ointe de butre et baignee [de vin], selonc ce q’il vos sera avis q’il puisse souffrir, qar molt est ce profitable», *Moamin*, II, 84, 3.

<sup>49</sup> Cfr. *TLIO* s.v. *biscia*. Cfr. anche *FEW* s.v. *bestia*; *LEI* s.v. latino *bestia*, latino volgare *bistia* (5, 1280-1282); *GDLI* s.v. *biscia*; *AIS* III 450; Glessgen 1996: 2, 499: «bissa scutaia cioè testuyna»; Zinelli 2016: 234: «littérament ‘couleuvre à écuelle’».

<sup>50</sup> Cfr. *FEW* s.v. *dux*; T–L s.v. *duc*<sup>2</sup>; *DMF* s.v. *duc*<sup>1</sup>; *TLIO* s.v. *duca*; *GDLI* s.v. *duca*.

<sup>51</sup> Cfr. «et por l’enor de mesire Renier li noble *dus* de Venise et por henor de cele noble cité que

che aveva la signoria di un ducato’, il grado di nobiltà più elevato dopo quello del principe (*duc de Carantan* 1, XXX; *duc de Crit* 2, LIII, ecc.). *Dus* sarebbe la forma del caso soggetto francese, *duc* quella del caso obliquo, e Martin introduce, dunque, nella declinazione dello stesso sostantivo una differenziazione semantica arbitraria<sup>52</sup>.

Lessemi che rinviano alle istituzioni di Venezia sono anche il verbo *\*ducar* ‘Esercitare il dogado, essere doge’ (*Estoires de Venise*, 1, VI, ecc.), per cui il *GDLI* attesta solo il significato di ‘investire della dignità ducale’<sup>53</sup>, e, inoltre, i sostantivi *\*ducat<sup>1</sup>*, nelle due accezioni ‘territorio retto da un doge, dogado’ e ‘titolo e potere di doge’, e *\*ducat<sup>2</sup>* ‘moneta d’argento fatta coniare dal doge Enrico Dandolo (secondo Martin da Canal)’ e ‘moneta d’oro coniata a Venezia a partire dal 1284’<sup>54</sup>.

#### 4. *Il versante Digital: l’annotazione del corpus in prospettiva FAIR, il corpus gold e la sperimentazione nel campo del NLP*

Il *corpus*, come si è detto, è uno strumento lessicografico propedeutico alla redazione del vocabolario ma è passibile anche di un impiego a sé. Posto che il *DiFri* punta a conformarsi ai principi *FAIR* (*Findability, Accessibility, Interoperability, and Reusability*), al fine di rendere la banca dati interoperabile e riutilizzabile, è necessario porsi l’obiettivo di adottare e implementare gli strumenti allo stato dell’arte nel campo del *NLP* (*Natural Language Processing*) conformandosi allo *standard* (cfr. *infra*).

Tra gli strumenti oggi disponibili per il francese antico ricordiamo il *Classical Language Toolkit* (*CLTK*, <http://cltk.org/>), originariamente implementato per il greco antico e il latino, nonché il parser di dipendenze per il francese antico studiato da Achim Stein<sup>55</sup>. La piattaforma di analisi testuale *TXM* (<https://txm.gitpages.huma-num.fr/textometrie/index.html>) fornisce accesso online a metodologie semiautomatiche di analisi del *corpus* che combinano strumenti qualitativi e quantitativi applicabili a testi di varia natura. *LGeRM* (*Lemmes Graphies*

l’en apelle Venise et por honor de la gentillesse et dou peuple venesiens, je Martin da Canal sui entremis de translater de latin en francis les henorees victoires que ont eües les Veneciens au servise de sainte Yglise et au servise de sa noble cité», *Estoires de Venise*, 1, I, 1 e *passim*.

<sup>52</sup> Cfr. Zinelli 2016: 183.

<sup>53</sup> Cfr. *GDLI* s.v. *ducare*; cfr. *FEW* s.v. *dux*.

<sup>54</sup> Cfr. *FEW* s.v. *dux* (III, 196b); *DMF* s.v. *ducat*; *TLIO* s.v. *ducato*; *GDLI* s.v. *ducato*. Per *ducat* ‘territorio retto da un doge, dogado’, cfr. «Et en leu de joie et de leece fist monsignor li dus convoier li patriarce de Grade et trestuit li evesque de son *ducat* et les autres perlas et li Frere menors et li Frere preschors et tos autres relegios et li prestres et tote la clergie de Venise» *Estoires de Venise*, 2, LIX; per ‘titolo e potere di doge’, cfr. «Or dit li contes, et la veraie estoire des Venisiens le tesmoi | gne, que mesure Marin Moresin fu esleüs dus de Venise après ce que mesure Jaques Teuples refusa le *ducat*», *ivi*, 1, CXXVIII.

<sup>55</sup> Cfr. Stein 2016, Johnson (*et alii*) 2021.

*et Règles MORPHologiques*) è un lemmatizzatore progettato per gestire la variazione grafica storica del francese: è stato inizialmente sviluppato per il *moyen français* (1330-1500) e poi adattato al francese del XVI e XVII secolo. Il *software Pyrrha* dell'École Nationale des Chartes è una applicazione Web sviluppata seguendo i dettami del framework Python Flask per annotare e velocizzare la postcorrezione di corpora lemmatizzati e taggati morfo-sintatticamente (<https://dh.chartes.psl.eu/Pyrrha/>): è un'applicazione web open source che integra metodi di *Interactive Machine Learning* per correggere i risultati ottenuti e riaddestrare automaticamente il sistema per renderlo più efficace<sup>56</sup>.

Per lemmatizzare e taggare POS/MSD (cioè aggiungere informazioni relative alla Part-of-Speech e al piano morfo-sintattico) il *corpus* descritto, implementeremo uno strumento dedicato estendendo *Pyrrha*: poiché non si tratta di creare un *software* da zero, ma di specializzarlo in modo che sia in grado di gestire varietà linguistiche miste, è necessario costruire un *corpus gold*, annotato a mano, che serva a (ri)addestrare e affinare il *software*.

Per il francese antico, come si è detto, il lemmario segue quello del T-L (eventualmente integrato) mentre i POS si conformano allo schema di *Cattex* (<http://bfm.ens-lyon.fr/spip.php?article176>). Da segnalare, tuttavia, che è già stato messo a punto un sistema di conversione dei POS secondo il sistema *Universal Dependencies (UD)*, che attualmente si configura come lo *standard* di riferimento per l'annotazione linguistica<sup>57</sup>.

Le banche dati facenti parte delle *Universal Dependencies* raccolgono primariamente testi *standard* di lingue contemporanee, ma sono presenti anche raccolte riferite ad alcune varietà non *standard* e alle lingue antiche (incluso l'antico francese)<sup>58</sup>. Un ulteriore vantaggio nell'utilizzo di *UD* riguarda la possibilità di impiegare *tool* per l'annotazione 'UD-compliant' allo stato dell'arte, come *UDPipe*<sup>59</sup> e il più recente *Stanza*<sup>60</sup>, grazie ai quali è possibile ottenere modelli per l'annotazione con un elevato livello di accuratezza, mediante l'implementazione di avanzate tecniche di *machine learning*<sup>61</sup>.

<sup>56</sup> Cfr. Camps (*et alii*) 2021.

<sup>57</sup> Cfr. De Marneffe (*et alii*) 2021.

<sup>58</sup> Per la lista delle lingue trattate e delle relative banche dati si rimanda al sito di *UD*, <https://universaldependencies.org/>.

<sup>59</sup> Cfr. Straka – Straková 2017.

<sup>60</sup> Cfr. Qi (*et alii*) 2020.

<sup>61</sup> Si vedano, a tal proposito, gli esperimenti di addestramento su varietà storiche di italiano effettuati con il *software Stanza* (Favaro [*et alii*] 2022 e 2023).

4.1. *Il corpus gold: consistenza e caratteristiche*

Tra gli obiettivi già conseguiti dal progetto *FrIngE* c'è quello della costituzione di un 'test corpus', vale a dire un campione testuale interamente corretto a mano contenente *specimina* testuali rappresentativi dei diversi piani di variazione (diacronica, diatopica, diatestuale) documentati nel *corpus* complessivo. Il *test corpus* costituito si configura dunque come *gold standard* del francese d'Italia, un punto di riferimento sia per la valutazione dei modelli di annotazione preesistenti (*Pyrrha*), sia per l'addestramento di nuovi modelli, con particolare riferimento a quelli derivati dalle banche dati della galassia *UD*, *standard* di riferimento per l'annotazione linguistica (cfr. *supra*, § 4).

Il bilanciamento del *test corpus* è avvenuto secondo l'ottica di rappresentare in primo luogo qualitativamente le caratteristiche principali dei testi del francese d'Italia, al fine di consentire valutazioni che non si limitassero a misurare il livello di accuratezza dell'annotazione. Si è deciso, dunque, di non effettuare una divisione rigida dei *tokens* degli *specimina*, in virtù sia dei parametri scelti per il bilanciamento, sia per l'inserimento di tipologie testuali brevi e frammentarie, ossia tutte le iscrizioni e i motti presenti nel *corpus* complessivo, che rappresentano un ulteriore elemento di complessità e di novità rispetto alle tipologie di testi generalmente impiegati per il trattamento automatico della lingua.

L'estensione attuale del *test corpus* è di circa 55.000 *tokens*, ma tale ampiezza è destinata a essere incrementata nel prosieguo del progetto. Le parole grafiche differenti (*types*) sono circa 9000, generando un rapporto *type-token* (la cosiddetta *type-token ratio*, in breve *TTR*) complessivamente attorno al 16%; i lemmi distinti sono circa 3700. La tabella sottostante mostra le caratteristiche principali dei testi scelti, ossia il titolo, l'autore, la forma (prosa o versi), il luogo e il periodo di composizione; in aggiunta, la sezione testuale presa in esame (inizio, fine, tutto il testo ecc.) e la tipologia testuale di appartenenza, secondo la classificazione adottata per il *corpus RLALFrI* (nel caso del *test corpus* si sono considerati testi appartenenti alle tre sole categorie delle copie di testi francesi eseguite in Italia, delle opere originali scritte da autori italiani in francese e dei testi in francese d'Oltremare).

Per ragioni pragmatiche, la selezione delle opere da includere nel *test corpus* a partire da un repertorio ricco e sostanzialmente aperto come quello del *RLALFrI* ha comportato l'esclusione di testi che pure avrebbero potuto entrarvi con profitto. La scelta del campione presentato è stata infatti operata tenendo in considerazione la necessità di bilanciare innanzitutto la presenza di testi in versi (otto) e in prosa (sette, comprese anche le testimonianze di carattere particolare quali i motti, le iscrizioni e le note per il miniatore del *Roman d'Alexandre* del museo Correr), quindi premurandosi di considerare tempi e luoghi di produzione diversificati. Per quanto dichiaratamente parziale, il campione che si è così costituito risulta in definitiva sufficientemente rappresentativo delle varietà testuali oggetto di studio.

La scelta delle porzioni di testo all'interno delle singole opere, di lunghezza variabile tra alcune centinaia e 5.000 *tokens* circa<sup>62</sup>, è stata quindi affidata ai rispettivi annotatori, che hanno individuato in base alle proprie competenze specifiche estratti paradigmatici delle caratteristiche linguistiche e formali dei testi di provenienza.

Titolo	Autore	Forma	Luogo	Datazione	Sez.	Tip. testuale
<i>Attila</i> in prosa	anonimo	prosa	Venezia	XIII	Inizio	Originale
<i>Bataille de Gamenario</i>	anonimo	versi	Piemonte	XIV (seconda metà)	Intero	Originale
<i>Dittamondo</i> (vv. fr.)	Fazio degli Uberti	versi	Toscana	XIV (metà)	brani sparsi	Originale
<i>Entrée d'Espagne</i>	anonimo padovano	versi	Padova (?)	XIV (inizio)	Metà	Originale
<i>Estoires de Venise</i>	Martin da Canal	prosa	Venezia	XIII (metà)	Fine	Originale
<i>Geste Francor</i> (ed. Morgan 2009)	anonimo	versi	It. sett.	XIV (inizio)	inizio della <i>Berta da li pé grant</i>	Originale
<i>Iscrizioni</i>	multipli	prosa			Intero	Originale
<i>La guerra d'Attila</i>	Nicolò da Casola	versi	Ferrara/ Emilia	XIV (seconda metà)	Inizio	Originale
<i>Canzoni Francesi</i>	multipli anonimi (?)	versi	Toscana	XIV/XV	Intero	Originale
<i>Ghatrif</i>	Daniele da Cremona	prosa	It. sett.	XIII (metà)	Inizio	trad. dal latino
<i>Moamin</i>	Daniele da Cremona	prosa	It. sett.	XIII (metà)	Metà	Originale
<i>Motti</i>	multipli	prosa			Intero	Originale
<i>Rubriche e note al miniatore</i> ( <i>Roman d'Alexandre</i> Correr 1493)	anonimo	prosa	Bologna (?)	XIII (fine)	Intero	Originale
<i>Passion</i> di Venezia	anonimo	versi	ignoto	XIV (metà)	intero	Originale
<i>Quatre âges de l'homme</i>	Filippo da Novara	prosa (e inserti in v.)	Cipro	XIII sec. (metà)	inizio	francese d'Oltremare
<i>Roland</i> (V4, ed. Beretta 1995)	anonimo	versi	Treviso	XIV (metà)	seconda metà	Copia

Grazie al *test corpus*, siamo riusciti a condurre una prima valutazione dei risultati, ossia l'analisi delle prestazioni attuali del *software Pyrrha*, mettendo a confronto i

<sup>62</sup> Fanno eccezione iscrizioni, motti e note per il miniatore che, data la relativa esiguità dei *tokens*, sono stati lemmatizzati integralmente.



singoli *files gold* con il *file* così detto *predicted*, ossia il testo annotato automaticamente in prima istanza, senza la revisione manuale<sup>63</sup>.

La seconda tabella che qui riportiamo rappresenta i valori di accuratezza di *Pyrrha* applicato al *corpus gold* rispetto alla lemmatizzazione, all'annotazione grammaticale del POS e delle proprietà morfologiche (colonna «MORPH»). I dati, come si osserva, sono piuttosto elevati, soprattutto in rapporto al lemma e al POS, con una media rispettivamente dell'86 e 87%, mentre l'annotazione delle etichette presenti in MORPH presenta maggiori difficoltà, con una media di accuratezza del 74%.

Titolo	Autore	N. <i>tokens</i>	N. <i>types</i>	TTR	N. <i>lemmi</i> distinti	Acc. Lemma	Acc. POS	Acc. MORPH
<i>Attila</i> in prosa	anonimo	4030	868	22%	537	93%	95%	87%
<i>Bataille de</i> <i>Gamenario</i>	anonimo	5334	1221	23%	770	87%	92%	78%
<i>Dittamondo</i> (vv. fr.)	Fazio degli Uberti	869	320	28%	239	97%	96%	86%
<i>Entrée</i> <i>d'Espagne</i>	anonimo padovano	4640	1399	31%	846	86%	90%	74%
<i>Estoires de</i> <i>Venise</i>	Martin da Canal	4990	902	18%	596	85%	81%	77%
<i>Geste Francor</i> (ed. Morgan 2009)	anonimo	2857	791	28%	479	79%	83%	65%
<i>Iscrizioni</i>	multipli	272	161	59%	130	73%	86%	66%
<i>La guerra</i> <i>d'Attila</i>	Nicolò da Casola	996	450	45%	350	81%	88%	68%
<i>Canzoni</i> <i>Francesi</i>	multipli (?)	4495	820	18%	454	84%	65%	64%
<i>Gbatrif</i>	Daniele da Cremona	4703	935	20%	627	90%	92%	74%
<i>Moamin</i>	Daniele da Cremona	5188	1001	19%	655	87%	93%	78%
<i>Motti</i>	multipli	266	110	29%	77	87%	86%	70%
<i>Rubriche</i> <i>e note al</i> <i>miniature</i> ( <i>Roman</i> <i>d'Alexandre</i> Correr 1493)	anonimo	3063	708	23%	375	87%	87%	71%
<i>Passion di</i> <i>Venezia</i>	anonimo	6080	1681	28%	1027	80%	84%	67%
<i>Quatre âges</i> <i>de l'homme</i>	Filippo da No- vara	2438	596	24%	387	99%	99%	99%
<i>Roland</i> (V4, ed. Beretta 1995)	anonimo	4943	1294	26%	669	75%	79%	58%

<sup>63</sup> Per garantire l'allineamento dei *files*, condizione essenziale per la comparazione dei risultati a

I dati qui rappresentati sono soltanto un primo passo per un'analisi più approfondita, e saranno un punto di riferimento per comparare l'efficacia dei nuovi modelli che verranno implementati in formato *UD*, una volta che sarà completata la conversione. I prossimi passi, infatti, prevederanno, come accennato all'inizio, la costituzione di un *test corpus* parallelo, che verrà usato per la creazione di modelli di annotazione per il francese d'Italia in formato *standard UD*. Una volta costruita tale risorsa, verrà svolta una valutazione più approfondita delle prestazioni, non soltanto in relazione alle specifiche dei singoli testi, ma anche sistematicamente ai singoli errori, seguendo le strategie di analisi adottata in Favaro (*et alii*) 2023 sul *test corpus* del *VoDIM* (*Vocabolario Dinamico dell'Italiano Moderno*)<sup>64</sup>.

#### 4.2. *Gli errori di Pyrrha*

In riferimento a quanto accennato in chiusura del precedente paragrafo, si tenterà ora di mostrare i primi risultati dell'analisi, sia qualitativa sia quantitativa, degli errori compiuti da *Pyrrha* in fase di annotazione.

Il primo problema riguarda la tokenizzazione, ovvero la suddivisione del testo in singoli *tokens*, un processo che, dal punto di vista computazionale, riguarda anche la suddivisione delle frasi (*sentence splitting*) e la corretta assegnazione di unità complesse, come per esempio forme articolate, polirematiche, ecc. (cfr. *supra*, § 3.2). In effetti, il principale problema riscontrato nel processo di tokenizzazione effettuato dal *software* di *Pyrrha* riguarda la separazione dell'apostrofo dalla rispettiva forma apostrofata e la conseguente errata attribuzione delle etichette in fase di annotazione. In tali casi, *Pyrrha* conta l'apostrofo come un *token* a parte e lo etichetta «PONfbb», analizzando poi la forma in origine portatrice di apostrofo alla stregua della forma omografa priva di apostrofo. Tra i casi più significativi, segnaliamo la sovrapposizione del pronome personale *e'* < EGO e della congiunzione *e* < ET. Poiché le due forme vengono a coincidere, *Pyrrha* procede ad annotarle tutte allo stesso modo: lemma «et», POS «CONCOO», «MORPH=empty». Lo stesso accade per le voci verbali *a'* < HABEO, HABES e *de'* < DEBET, considerate erroneamente occorrenze delle preposizioni *a* e *de*. I casi, che interessano le serie di monosillabi omografi, sono numerosi e variano in rapporto al testo lemmatizzato. Problemi di segmentazione dei *tokens* come questi rappresentano un grosso ostacolo nell'annotazione, in quanto comportano, come si è visto dagli esempi precedenti, errori di riconoscimento di tutte le etichette, dal lemma al POS alle proprietà morfologiche. Essi non sono, tuttavia, un *unicum* attribuibile a *Pyrrha*; si

tutti i livelli dell'annotazione (lemmatizzazione, POS-tagging e descrizione morfologica), l'unico intervento sui *files* annotati automaticamente ha riguardato le ipo-segmentazioni dei *tokens*, per es. nelle preposizioni articolate, spesso non riconosciute dal modello di *Pyrrha* e non separate in due o più *tokens*; ciò ha comportato che questi venissero esclusi dagli errori compiuti da *Pyrrha*: la stima qui riportata è da considerarsi, dunque, lievemente in difetto.

<sup>64</sup> Sul *VoDIM* cfr. Marazzini – Maconi 2018.

tratta, invece, di un problema piuttosto comune nel trattamento automatico delle varietà storiche: numerosi casi sono stati osservati, per esempio, in alcune banche dati rappresentative di varietà storiche di italiano<sup>65</sup>.

Errori di questo genere si inseriscono nel più vasto ambito della corretta distinzione delle forme omografe in *Pyrrha*. Sia nella lemmatizzazione di testi antico-francesi che di testi franco-italiani, il *software* ha dimostrato i propri limiti nel riconoscere, ad esempio, *a* preposizione e *a* forma verbale, *oit* voce del verbo *oir* e voce del verbo *avoir*, *son* possessivo e *son* voce del verbo *estre*. Si registrano, inoltre, casi specifici di confusione relativi a forme italiane o italianizzate, quali la preposizione *cum* riconosciuta come la congiunzione subordinante *come*<sup>1</sup>, il numerale *doç* riportato all'aggettivo *douç* e non al lemma *doçe*, la voce verbale è associata al verbo *avoir* e non al verbo *estre*.

Infine, ulteriore fattore di difficoltà riguarda l'attribuzione automatica dei nomi propri (cfr. *infra*, § 3.2): sia nella sovraestensione del POS «NOMpro» relativamente ai *tokens* che seguono un punto fermo e che quindi iniziano con una maiuscola; sia nel riconoscimento del corretto nome proprio in fase di lemmatizzazione. Se la capacità di attribuzione del POS «NOMpro» è oscillante e varia di testo in testo, l'impatto invece sulla lemmatizzazione è impressionante: sui circa 7.400 errori totali, oltre 1.300 coinvolgono nomi propri (17,6%)<sup>66</sup>.

Per quanto riguarda invece le etichette morfologiche, bisogna escludere dal computo degli errori quelli riguardanti il tag «SPEC=ib»: trattandosi di un'etichetta innovativa che segnala la presenza di un italianismo (cfr. *supra*, § 3.2.1), il modello attuale di *Pyrrha* non è addestrato al riconoscimento di tali valori.

## 5. Studio di casi

### 5.1. Il Roland di V4

Riportiamo di seguito alcuni casi di particolare interesse emersi durante l'annotazione di brani sparsi del *Roland* (V4, ed. Beretta 1995)<sup>67</sup>.

*aer*<sup>68</sup>: il sostantivo maschile conta due sole occorrenze, entrambe associate a *solibione* (cfr. *infra*) all'interno di una dittologia che si riferisce a fenomeni atmosferici straordinari<sup>69</sup>. Nei

<sup>65</sup> Si vedano per es. gli esempi relativi al *corpus Voci della Grande Guerra*, De Felice (*et alii*) 2018, Lenci (*et alii*) 2020; oppure quelli riscontrati nel *corpus VoDIM*, Favaro (*et alii*) 2023.

<sup>66</sup> Visto l'interesse primariamente linguistico del *corpus*, per il momento si è rinviata l'elaborazione automatica dei nomi propri, processo che solitamente è demandato a tecniche di NER o *Entity Linking* di cui non si esclude un impiego in fasi successive del progetto.

<sup>67</sup> Per alcune considerazioni su questo testo (e la sua edizione) cfr. *infra*, § 6.1.

<sup>68</sup> Cfr. *Roland* (V4, ed. Beretta 1995), Glossario, sv.

<sup>69</sup> Cfr. *Roland* (V4, ed. Beretta 1995): 249: «Si me mandai dal cel et aer et sulipion», v. 3899; «Quel corno li manda aher et solibione», v. 3905.

passaggi testuali coinvolti, tale forma designa molto probabilmente una precisa condizione meteorologica, che può rientrare nello spettro di significati di *air*<sup>1</sup> del T-L, designato quale lemma di riferimento<sup>70</sup>. L'aggiunta, inoltre, dell'etichetta «SPEC=ib» nel campo dell'annotazione morfologica permette di segnalare la peculiarità delle forme *aer/aber*.

*adrelī*: voce non registrata nel glossario dell'edizione Beretta, ma riconducibile al verbo franco-italiano *adreler* (scelto quale lemma di riferimento), nel significato di 'mettre derrière', come riporta il *DÉAF*<sup>71</sup>. Si tratterebbe di una forma del congiuntivo<sup>72</sup>, che è stata quindi annotata dal punto di vista morfologico «SPEC=it|MODE=sub|TEMPS=pst|PERS.=3|NOMB.=s».

*imprenderò*<sup>73</sup>: indicativo futuro alla prima persona singolare, che nell'annotazione ha acquisito l'etichetta «SPEC=ib» non solo per l'evidente morfologia italiana, ma anche per lo specifico significato di 'imparare, venire a conoscere' che la voce verbale assume nel contesto del verso<sup>74</sup>. Tale accezione, documentata in altri testi franco-italiani e italiani<sup>75</sup>, pare invece non essere attestata nell'antico francese *empredre*, scelto come lemma di riferimento.

*noé*<sup>76</sup>: il sostantivo femminile compare un'unica volta nel testo, in sede di rima, e non risulta avere altre attestazioni. Rappresenta il secondo elemento di una dittologia oppositiva<sup>77</sup> ed è stato annotato come segue: lemma «nuit»; POS «NOMcom»; MORPH «SPEC=rīm|NOMB.=s|GENRE=f|CAS=r».

*recontra*<sup>78</sup>: il sostantivo femminile, che incontriamo nel testo una sola volta in chiusura di verso<sup>79</sup>, è registrato come voce franco-italiana dal *DÉAF* col significato di 'risposta' e ricondotto al verbo *rencontrer*<sup>80</sup>. In assenza di un'ipotesi etimologica convincente e dato che la forma sembra essere un conio occasionale, ci si è discostati dalla scelta del *DÉAF* e si è deciso di creare un nuovo lemma – *recontra*, appunto – aggiungendo poi nel campo morfologico l'etichetta «SPEC=rīm».

*se pise*: voce del verbo *peser*, usata in sede di rima alla forma riflessiva nell'accezione di 'caricarsi'<sup>81</sup>; tale uso riflessivo sembra non essere attestato altrove. È stata annotata come segue: lemma «peser», POS «VERcǵ», MORPH «SPEC=rīm|MODE=ind|TEMPS=pst|PERS.=3|NOMB.=s».

<sup>70</sup> Si segnala che il sostantivo ricorre in *Ystoire de li Normant*, 5, [11], 2 con il diffuso significato di 'aria'.

<sup>71</sup> Cfr. *DÉAF* s.v. *adreler*, <https://deaf.hadw-bw.de/lemme/adreler>.

<sup>72</sup> Cfr. *Roland* (V4, ed. Beretta 1995): 60: «Eo sum li terço, or s'en adrelī el quart!», v. 843.

<sup>73</sup> *Ivi*, Glossario, s.v.

<sup>74</sup> *Ivi*: 16: «Si impre<n>derò ses costumi et son talan». v. 201.

<sup>75</sup> *Ivi*, Glossario, s.v., e *GDLI* s.v. *imprenderere*.

<sup>76</sup> *Ivi*, Glossario, s.v. Beretta, per ragioni contestuali e fonetiche, esclude l'interpretazione di *noé* come 'nottata' e propone una diversa lettura della lezione del manoscritto: *l'anoé*, col significato di 'anno'.

<sup>77</sup> *Ivi*: 261: «El me remembra li çorn et la noé», v. 4134.

<sup>78</sup> *Ivi*, Glossario, s.v.

<sup>79</sup> *Ivi*: 22: «Al roi Marsilio portarò la recontra», v. 292.

<sup>80</sup> Cfr. *DÉAF* s.v. *contre*, <https://deaf.hadw-bw.de/lemme/contre#rencontrer>.

<sup>81</sup> Cfr. *Roland* (V4, ed. Beretta 1995): 65: «Che non fa set mul, quand d'aver se pise», v. 919.

*solibione*<sup>82</sup>: il sostantivo maschile conta due sole occorrenze, entrambe associate ad *aer* (cfr. *supra*) all'interno di una dittologia che si riferisce a fenomeni atmosferici straordinari<sup>83</sup>. Beretta afferma che l'eventuale passaggio etimologico SULPHUR > *solibione/sulipion* non è chiaro, benché si possa ipotizzare l'impiego metonimico di *zolfo* per 'fulmine'; Luca Morlino ipotizza una derivazione etimologica da \*SUBLIVIONE(M) < SUBLUVIUM e propone di intendere *solibione* come 'pioggia di fango'<sup>84</sup>. Ai fini della lemmatizzazione si è deciso di creare un nuovo lemma, *solibione*, e di etichettare le due occorrenze nel modo seguente: POS «NOMcom»; MORPH «SPEC=it|NOMB.=s|GENRE=m|CAS=r».

## 5.2. *La Geste Francor*

Nel corso dell'annotazione della *Berta da li pé grant* (tuttora in fase di svolgimento) sono emersi punti critici che potrebbero destare interesse per il progetto nel suo complesso<sup>85</sup>. Di seguito, diamo nota di tali passaggi.

*cason* («Questo non po avenire por nesune cason», v. 1360): si tiene a lemma il parallelismo con l'antico francese *ochaison* e si indica «SPEC=it» in MORPH – lemma «ochaison», POS «NOMcom», MORPH «SPEC=it|NOMB.=s|GENRE=f|CAS=r».

*çubler* (v. 1218 e *passim*): il sostantivo appare tipico del francese d'Italia<sup>86</sup>, ma ha ovviamente un parallelo nell'antico francese *jogleor*; si è deciso, dunque, di mantenere il lemma francese e di inserire «SPEC=it» – lemma «jogleor», POS «NOMcom», MORPH «SPEC=it|NOMB.=s|GENRE=m|CAS=r».

*çugolarie* (in rima, v. 1327): qui si mantiene il lemma parallelo antico francese *joglerie*, e si inserisce «SPEC=it» in MORPH – lemma «joglerie», POS «NOMcom», MORPH «SPEC=it|NOMB.=s|GENRE=f|CAS=r».

*ma* (v. 1238 e *passim*): la congiunzione coordinante è sicuramente italiana, ma a lemma abbiamo usato il parallelo antico francese *mais*<sup>1</sup>, con «SPEC=it» – lemma «mais1», POS «CONcoo», MORPH «SPEC=it|MORPH=empty».

*machon* («Cun grant avoir e cun grande machon», v. 1367): si lemmatizza *maconne*, che secondo il *TLIO* sarebbe un tipo di lana, qui metonimico per 'corredo vestiario' trattandosi della descrizione di una dama (cfr. «Prendés una dame de qualche region...», vv. 1363 e ss.). Nel glossario della *Geste Francor* (ed. Morgan 2009), l'editore pensa invece a un altopro di *mangon* 'moneta d'oro', probabilmente in dittologia sinonimica con *avoir*. Abbiamo però un altro luogo in cui *machon* varrebbe 'tessuto', nel poema *La guerra d'Attila* di Nicolò da Càsola: «L'autre part regarde Forest dou pavilon / Mout richement ovré et de grant machon / A pieres precieuses, a or et a charbon» (I, III, vv. 106-108) – lemma «maconne», POS «NOMcom», MORPH «SPEC=it|NOMB.=s|GENRE=f|CAS=r».

<sup>82</sup> Cfr. *Roland* (V4, ed. Beretta 1995), Glossario, s.v.

<sup>83</sup> Cfr. *supra*, n. 69.

<sup>84</sup> Morlino 2010: 79-82.

<sup>85</sup> Per alcune considerazioni su questo testo (e la sua edizione) cfr. *infra*, § 6.1.

<sup>86</sup> Per un approfondimento sull'etimologia della forma si rimanda a Beretta 2021: 265, n. 33.

*quille* («cun quille de Clermon», v. 1357): esempio di una forma con radice italiana *quill-* e morfologia latamente antico-francese *-e* (manca la *-s* finale): il fatto che il tipo lessicale sia esclusivamente italiano implica l'adozione del lemma italiano, *quello*; conformemente alle norme (cfr. *supra*, § 3.2.1), si indica «SPEC=it» in MORPH – lemma «quello», POS «PROdem», MORPH «SPEC=it|NOMB.=p|GENRE=m|CAS=n».

*soit* («Lengue el soit de plesore mainer», v. 1228; «D'Ongarie soit e l'insir e l'intrer», v. 1232): voce verbale dall'antico francese *savoir*, si può intendere come perfetto con *i* irrazionale, e si annota dunque con «SPEC=it» in MORPH – lemma «savoit», POS «VERc|g», MORPH «SPEC=it|MODE=ind|TEMPS=psp|PERS.=3|NOMB.=s».

*vonter* (v. 1277): si tratta di un avverbio tipico in antico dell'area milanese e lombarda orientale, e lo si lemmatizza con il corrispondente antico francese *volentiers*, con «SPEC=it» in MORPH – lemma «volentiers», POS «ADVgen», MORPH «SPEC=it|MORPH=empty».

### 5.3. Esempi dal Moamin e dal Ghatrif: tratti del francese d'Italia e lessico scientifico

All'interno del panorama dei testi 'franco-italiani'<sup>87</sup>, si conservano due trattati di falconeria e cinegetica, comunemente conosciuti con il nome di *Moamin* e *Ghatrif*, conservati nei codici *a* (*Moamin* e *Ghatrif*) e *k* (solo *Moamin*, fino a IV, 7, 3)<sup>88</sup>.

I due trattati di falconeria sono stati oggetto di annotazione, in modo da valutare la risposta di *Pyrrha* a fronte di una lingua complessa come quella che Daniele utilizza per classificare le varie tipologie di falconi e cani da caccia, per classificare una serie di elementi naturali, animati (ad es. il lessico fitonimico) o inanimati (pietre, polveri, ...) impiegati nelle ricette veterinarie.

La *scripta* di *a*<sup>89</sup> è costituita da un «droit français», che lascia, però, intravedere elementi riconducibili alle *koine* veneto-centrali: si pensi, ad esempio, ad alcuni fenomeni linguistici che possono essere ricondotti a Verona, come l'evoluzione AU > *ons* (es. *chonses*)<sup>90</sup>, ben attestata in Nicolò da Verona; il passaggio da *-al-* ad *-ao-* (es. *faoconers*, I, 23, 10), oppure alcuni lessemi particolari che rimandano alla città scaligera, come *bisse escuere* che ricalca *bissa scuara*<sup>91</sup>. Per quanto riguarda i tratti veronesi, si è utilizzato il marcatore «SPEC=it», poiché non esistono, per il momento, categorie iponime nel quadro dell'annotazione linguistica. Altri *markers* rimandano più genericamente a testi del Francese d'Italia, come l'assenza di *e-prostetica* o la poca consuetudine con il comparto dei dittonghi<sup>92</sup>.

<sup>87</sup> Cfr. Gambino – Beretta 2023: VII-XXX.

<sup>88</sup> Per un quadro bibliografico cfr. *Moamin* e *Ghatrif*; Beretta Spampinato 1990; Marruncheddu 2008: 29; Guariglia 2022: 136-139.

<sup>89</sup> L'annotazione dei due testi è stata condotta solo su *a*, in modo tale da poter avere, per i due trattati, una base comune. Sarà interessante, in futuro, condurre la stessa operazione sul codice *k*, per verificare la risposta di *Pyrrha* alla *scripta* del codice belga.

<sup>90</sup> Cfr. Beretta – Palumbo 2015: 75-78.

<sup>91</sup> Per i tratti veronesi cfr. *Testi veronesi*: 61-64; Zinelli 2016: 245. Per le *scriptae* dei due codici cfr. Guariglia 2022: 148-156.

<sup>92</sup> Si rimanda sempre all'introduzione linguistica di Gambino – Beretta 2023, in particolare XLV-XLIX.

In sostanza, la *scripta* di *a* è un compromesso tra almeno tre 'sistemi' linguistici: 1) il 'francese' – con significato di 'le *koinè* dialettali che compongono ciò che comunemente chiamiamo lingua d'*oïl*' – che reagisce con xenismi quali arabismi, latinismi, occitanismi, italianismi lessicali; 2) la lingua di Daniele da Cremona, presumibilmente il lombardo meridionale; 3) le lingue dei copisti (per le quali possiamo quantomeno individuare in Verona il centro di copia di *a*). Per discutere il lessico di Daniele è, dunque, opportuno tenere a mente la complessa stratigrafia linguistica che ha dato luogo alle *scriptae* dei due codici superstiti.

L'annotazione operata attraverso lo strumento *Pyrrha* ci ha permesso di isolare alcuni dei *markers* che costituiscono l'insieme linguistico di Daniele e valutare la risposta della macchina. Numerosi sono stati gli errori in lemmi che presentano elementi italiani, come nel caso dell'inserzione di *-i-* irrazionale nella desinenza verbale *-er* (es. *translatier*, *Gh Pr.*, II *et passim*, lemmatizzata con *translacion*), delle forme aferetiche (es. *spice* II, 2, 3 *et passim*, lemmatizzata con *sipos*, o *scorçe*), dell'oscillazione nel dittongamento o dell'avanzamento delle affricate. Meno difficoltà si sono registrate in lemmi in cui si rilevano tratti di convergenza con le *koinè* nord-orientali (per cui *Phyrra* è risalito alla corretta entrata), come ad esempio, la mancata palatalizzazione in contesti *k+a*, es. *capitres*, o le forme del dimostrativo in *tiele* (II, 4, 14 *et passim*)<sup>93</sup>, forme già attestate in testi francesi nord-orientali.

Le osservazioni più interessanti riguardano, però, le scelte lessicali di Daniele che concernono la resa dei termini tecnico-scientifici<sup>94</sup>. Il lessico dell'autore è un complesso intrecciarsi di parole di provenienza differente, dal latino all'arabo, dal francese alle *koinè* padane: quando Daniele deve indicare un referente reale, opta per soluzioni differenti a seconda della sua conoscenza dell'oggetto o dell'organismo da classificare.

Nell'analisi degli xenismi è bene ricordare, però, due elementi di non secondaria importanza:

- 1) Alcuni lemmi non possono essere riportati a un solo dominio linguistico; è il caso della convergenza tra forme latine e italiane, le quali possono aver concorso mutuamente alla costruzione della forma. Il *Moamin* e il *Ghatrif* franco-italiani sono stati «*translatiez de latin en françois*» (*Moamin Prol.*), la quale deve aver giocato un importante ruolo nella costruzione lessicale di Daniele. Non stupisce, dunque, che opti spesso per l'utilizzo dei termini scientifici latini.
- 2) Alcuni lemmi sono, invero, attestati nei vocabolari del francese antico, ma tali attestazioni possono essere a) tarde; oppure b) essere degli *unica* attestati solamente nei trattati di falconeria e cinegetica. È il caso, ad esempio, di *uve* (II, 47, 28) che è attestata nei vocabolari francesi o come forma tarda,

<sup>93</sup> Cfr., per es., Zinelli 2021: 84.

<sup>94</sup> Sul lessico scientifico dei trattati di falconeria, cfr. Glessgen 1998: 437-441.

in coppia con *passee*, oppure proprio con riferimento al *Moamin*, spiegata con il significato ‘raisin’. Il *modus operandi* da seguire in queste circostanze prevede la segnalazione della specificità del lessema mediante l’attribuzione dell’etichetta «SPEC=it» o «SPEC=lat» (cfr. *supra*, § 3.2.1).

Nella sezione analizzata, si segnalano di seguito alcune forme che, a causa della loro assenza nei vocabolari francesi, hanno messo in difficoltà *Pyrrha*. Un primo gruppo racchiude le forme di origine orientale. Tra queste si registrano prestiti non adattati, che mantengono cioè una *facies* orientale (per es. *saadeg*, II, 14, 4; *chyatrech* II, 32, 2; *suuch* II, 82, 3), e prestiti ‘francesizzati’, come nel caso di *cacule* (II, 3, 1; 44, 2; 44, 7) o *sambace* (II, 46, 7; 91, 20). Nei casi succitati, si è proceduto all’introduzione del lemma in *Pyrrha*.

Numerose sono le forme desunte dal latino: spesso si tratta di prestiti adattati, poiché Daniele sceglie, come già visto per gli xenismi orientali, di adattare al sistema morfologico volgare le forme latine. Per le forme certamente latine è stato utilizzato il marcatore «SPEC=lat», a fronte di un’iniziale classificazione erronea di *Pyrrha*. Come anticipato *supra*, però, non sempre riusciamo a discernere se le forme di Daniele siano latinismi *tout court* o vi sia l’influenza delle *koinè* padane. Tra i latinismi osserviamo, ad esempio, *acetose* (IV, 18, 14, *Rumex acetosa*), *amoron* (II, 91, 16, *Radicem amoroni*), *atriple* (II, 48, 10, *Atriplex hortensis*), *cassie* (II, 2, 3, *Cassia lignea*), *challains artmatic* (Gh. 56, 2, *Calamus aromaticus*), *chochine* (II, 91, 20, *Cochina*), *choest* (II, 9, 2, *Choestus*), *coloquintide* (II, 55, 13, *Cumis colocynthis*). Per le forme di convergenza, si è scelto di valutare altri elementi, come l’attestazione delle forme nei vocabolari italiani di riferimento (*in primis* il *TLIO*) e nei testi coevi. È il caso di *aqacie* (III, 13, 4), già presente nel *Régime du Corps*, o di *liquirice* (I, 34, 9; *et passim*, *Glycyrrhiza glabra* L.; *k* presenta numerosi allografi: *liqirize*, *liquiricie*, *liquirice*, *leqirice*), *majorane* (II, 62, 2; III, 1, 7, *Origanum majorana*; *k* *maiorane*), *malvavische* (II, 7, 2; 27, 4; III, 8, 34, it. ‘malvavischio’). Nei casi dove l’eziologia italiana appariva preferibile – e la forma non era attestata nei vocabolari francesi – si è scelto di utilizzare il marcatore «SPEC=it». Lo stesso marcatore è stato utilizzato per forme certamente riconducibili alle *koinè* italiane, come *çename* (II, 3, 2; 4, 10), *çevole* (II, 46, 22; 46, 23, *Allium cepa*), *zimin* (II, 68, 6, *Cuminum cyminum*), *pevre* (II, 4, 10 *et passim*, *Piper* ‘pepe’; *k* *pever*, *puere*, *pevre*), che *Pyrrha* in prima istanza aveva registrato erroneamente. Il marcatore «SPEC=it» è stato quindi utilizzato sia per segnalare fenomeni grafico-fonetici imputabili all’*usus* franco-italiano, come l’avanzamento delle affricate, sia per lemmi italiani con morfologia francese<sup>95</sup>.

<sup>95</sup> Si rimanda alla trattazione in Gambino 2016: 85-86, § 3.



#### 5.4. Altri casi

Di seguito si segnalano altri punti degni di nota e riflessione.

*averbés* (*Entrée d'Espagne*, v. 154, in rima: «Mielz valt sovant taisir q'estre trop averbés»): l'aggettivo ha il significato di 'verbosi, loquaci' e presenta prostesi di *a-* tipicamente franco-italiana. Si mantiene a lemma la forma *verbos*, registrata in T-L (XI, 249) come hapax con rinvio a Gdf (X, 844a s.v. *verbeux*), e si inserisce l'etichetta «SPEC=it» nell'analisi morfologica.

*bauzaine* (*Bataille de Gamenario*, v. 411, in rima: «blanche et vermeille elle est bauzaine»; v. 462: «la bauzaine, que Dieu maintieigne»): nel lavoro di annotazione si è adottato il lemma *baucenc* (< BALTEUS), aggettivo antico francese riferito al manto del cavallo (cfr. T-L I, 884, 49), ma con l'aggiunta «SPEC=it» per segnalare l'italianismo semantico. Nel testo in esame, infatti, *bauzaine*, utilizzato in forma prima aggettivale e poi sostantivale, si riferisce sempre all'insegna bicolore a bande orizzontali, secondo il significato documentato nel *TLIO* (s.v. *balzana* s.f. e *balzano* agg./s.m.).

*corin* (*Entrée d'Espagne*, v. 11532, in rima: «Se vient vos vant ni levant ni corin»): la forma antico francese *chorus* (< CAURUS 'vento di nord-ovest, maestràle') non figura nel T-L e nel Gdf, ma è documentata nel *DEAF* con tre esempi, ai quali si affianca, in un'entrata specifica, il caso in esame. Il vocabolo attestato nel poema del *Patavian* è risultato dell'evoluzione dal diminutivo latino CAURINUS, al pari del sostantivo femminile *corina*<sup>2</sup> registrato nel *TLIO* (sebbene lì sia glossato, seguendo Contini, come 'vento di sud'), per cui, in base ai criteri adottati, è posta a lemma la forma *corin* e nell'analisi morfologica viene indicato il tag «SPEC=it».

*estremie* (*Bataille de Gamenario*, v. 558, in rima): il lessema, riferendosi ad un contesto di battaglia, sembra combinare le forme *estormie* ed *escremie*, ma non è attestato nei dizionari di antico francese. Dato che il sostantivo è presente anche nella *Continuazione dell'Entrée d'Espagne* (v. 1879) sempre col significato di 'fracasso, strepito'<sup>96</sup>, si è deciso di creare il lemma apposito *estremie*.

*pensonciaus* (*Entrée d'Espagne*, v. 10118, 'fringuellini'): questa forma diminutiva, collocata all'interno di verso, non si rintraccia nei dizionari di antico francese e pare creazione dell'Anonimo Padovano. Seguendo il criterio invalso per gli alterati, si è posto a lemma *pinçoncel*, con base nell'antico francese *pinçon* (T-L VII, 957, 28) seguito da suffisso alterante.

*scerp* (*Entrée d'Espagne*, v. 10129): il vocabolo, glossato 'cervi' da Thomas (*Entrée d'Espagne*: II, 403), si riferisce piuttosto a 'botoli', come chiarito da Marco Infurna che lo collega a *schirp*, attestato in esempi mantovani e veronesi, e alle occorrenze di *grepo* in Bonvesin<sup>97</sup>. Visti i dubbi nell'individuazione dell'etimo, si è preferito al momento lemmatizzare *scerpo* e segnalare la criticità con l'etichetta «SPEC=probl» (POS «NOMcom»; MORPH «SPEC=probl|SPEC=it|NOMB.=p|GENRE=m|CAS=n»).

<sup>96</sup> Si rinvia, a tal proposito, al glossario curato da Franca Di Ninni in Niccolò da Verona, s.v. *Cerato* traduce invece 'assalto': cfr. *Bataille de Gamenario*: 418.

<sup>97</sup> Si vedano, a riguardo, le opportune osservazioni di Infurna 2011: 133-136.

6. *Prospettive future*6.1. *Nuove edizioni*<sup>98</sup>

Il *Repertorio Informatizzato Antica Letteratura Franco Italiana (RLALFrI)*, la banca dati all'interno della quale si colloca il *corpus DiFrI*, ha come primo obiettivo quello di restituire una «biblioteca digitale dinamica»<sup>99</sup> nella quale vengono raccolte le edizioni dei testi del così detto 'francese d'Italia'<sup>100</sup>. Quest'ultime possono essere vecchie edizioni riviste per la messa in linea, oppure edizioni più recenti; in entrambe le situazioni, in caso di modifiche da parte dell'editore-filologo, il testo viene accompagnato da una scheda di apparato che riporta le variazioni rispetto all'edizione di riferimento. Si veda, ad esempio, il *Bovo d'Antona* udinese, a cura di Francesca Gambino, che nella sua versione digitalizzata nel *Repertorio* integra, oltre al testo, anche l'apparato critico e le varianti rispetto alla precedente edizione Rajna<sup>101</sup>; oppure il testo della *Battaille de Gamenario* secondo l'edizione Cerrato per la quale, in apparato, vengono segnalate le proposte di correzione di Formisano<sup>102</sup>. Secondo questo modello, le edizioni *RLALFrI* potrebbero essere definite come 'edizioni digitalizzate', dal momento che si configurano come trasposizioni o revisioni critiche dell'edizione a stampa e sono state sviluppate, dunque, tramite una metodologia di lavoro 'tradizionale', nella quale si sono innestate solo successivamente delle pratiche digitali<sup>103</sup>. Tuttavia, di recente, soprattutto in concomitanza con le prime annotazioni, per alcuni testi digitalizzati nel *Repertorio* ci si è avvicinati alla metodologia e ai modelli delle più recenti edizioni digitali. È il caso, per esempio, de *l'Enfances Bovo* della *Geste Francor*, lemmatizzato da Sira Rodighiero<sup>104</sup>: al testo, oltre alla scheda delle modifiche all'edizione, sono stati integrati anche i metadati linguistici derivati dall'annotazione. Questi ultimi sono consultabili tramite dei pop-up direttamente dal testo edito, e riportano le informazioni seguenti: lemma, analisi morfologica della forma e la specifica di 'italia-

<sup>98</sup> Il lavoro per lo sviluppo di un nuovo modello di edizioni *RLALFrI* è possibile anche grazie al prezioso aiuto di Angelo Mario Del Grosso e Simone Zenzaro (Cnr-Istituto di Linguistica Computazionale «A. Zampolli»).

<sup>99</sup> Vale a dire una biblioteca di testi suscettibili di evoluzione; cfr. <https://www.rialfri.eu/il-progetto>.

<sup>100</sup> Cfr. Gambino – Beretta 2023.

<sup>101</sup> Cfr. Rajna 1887, <https://www.rialfri.eu/rialfriPHP/public/testo/testo/codice/rialfri%7CbovoUdinese%7C001>

<sup>102</sup> Cfr. Formisano 1978, <https://www.rialfri.eu/rialfriPHP/public/testo/testo/codice/rialfri%7CbbattagliaGamenario%7C001>.

<sup>103</sup> Per la differenza tra edizioni 'digitali' e edizioni 'digitalizzate' cfr. Sahle 2016: 19-40; cfr. inoltre Pierazzo – Mancinelli 2020: 11, Robinson 2013 e Shillisburg 2014.

<sup>104</sup> L'edizione digitalizzata e annotata è consultabile al seguente *link*: [https://www.rialfri.eu/rialfriPHP/public/testo/testo/codice/rialfri%7CgesteFrancor\\_1%7C001](https://www.rialfri.eu/rialfriPHP/public/testo/testo/codice/rialfri%7CgesteFrancor_1%7C001).

nismo' («SPEC=ib»), dove necessaria. Questo tipo di sistema restituisce un ambiente che non si limita più solo a presentare un'edizione in formato digitale, ma permette anche un'interazione diretta da parte del lettore.

Con l'avvio del progetto *FrInGE* e l'annotazione massiccia dei testi, ci si è interrogati sui modelli e sulla qualità delle edizioni raccolte nel *Repertorio*. La banca dati *RLALFrI* vuole raccogliere testi e edizioni scientificamente validi, andando così a evitare il così detto 'effetto internet'<sup>105</sup>. Infatti, con l'esponenziale crescita di testi *open access* disponibili in rete di dubbia validità scientifica, garantire la massima scientificità del testo e rendere limpide e chiare le scelte dell'editore tramite apparati risulta, ancor di più, un obiettivo di fondamentale importanza<sup>106</sup>. Per questo motivo, in alcuni casi, si possono consultare più edizioni dello stesso testo: si veda, per esempio, la *Berta* della *Geste Francor*, consultabile sia nell'edizione Morgan 2009, sia nell'edizione Scattolini 2009<sup>107</sup>.

Per la lemmatizzazione ci si è avvalsi, dunque, di edizioni già presenti nella banca dati, le quali, tuttavia, grazie a uno studio minuzioso legato proprio all'annotazione linguistica, sono state revisionate dai ricercatori e dalle ricercatrici che collaborano con il progetto. È il caso, per esempio, della *Berta* nella *Geste Francor* (ed. Morgan 2009), che prima di essere lemmatizzata è stata rivista da Andrea Beretta direttamente sul manoscritto. Si è deciso di prendere a punto di riferimento per l'annotazione della *Geste Francor* l'ed. Morgan 2009, l'ultima di una serie cospicua di attenzioni critico-filologiche al testo in questione (si ricorderà qui almeno l'ed. Rosellini 1986 dell'intero ciclo epico; il database *RLALFrI* presenta per l'intera *Geste* l'ed. Morgan, mentre per la *Berta* e per il *Macario* ha rispettivamente anche le edizioni Scattolini 2009 e Bergo 2016). Alla luce, però, di un approfondimento filologico sulla *Berta* che ha ripreso in esame direttamente il manoscritto, si è deciso di aggiornare il testo Morgan 2009 della specifica sezione del ciclo, con inevitabili ricadute anche sul resto dell'opera, che pertanto andrà rivista alla luce di una riconsiderazione globale e del codice e della resa del testo stesso. Di seguito, presenteremo le modifiche apportate al testo Morgan 2009 sulla base della revisione globale compiuta sul segmento della *Berta* (circa 1750 vv.).

Dapprima vediamo gli errori di lettura del manoscritto (le forme del codice si presentano già trattate diacriticamente):

<sup>105</sup> Cfr. Zaccarello 2017: 148-162.

<sup>106</sup> Cfr. Italia 2016: 13-26.

<sup>107</sup> Cfr. rispettivamente il *link* riportato alla n. 94 e <https://www.rialfri.eu/rialfriPHP/public/testo/testo/codice/rialfri%7Cberta%7C001>.

	<i>manoscritto</i>	<i>ed. Morgan 2009</i>
R. 17	avron	Avrò
R. 17	buem	Buens
1167	là	Li
R. 18	baron	Baroz
R. 18	du	d·i
1249	Saragoca	Saragoça
1285	vestra	Vestre
1285	corte	Cort
1306	Et	E
1310	parola	Parole
1312	et	E
1314	parola	Parole
1329	no	Non
1339	dame	Dama
R. 21	Aqilun	Aq(u)iluz
R. 21	Pepim	Pepin
1442	tinent	Tinent
1446	vos	Ves
1483	tenpo	Tempo
1484	fist	Sunt
1489	Pipin	Pepin
1582	unqes	Anqes
1608	vos	(v)os <sup>108</sup>
1638	vos	v(o)s
R. 30	la	Sa
1694	Françe	França
R. 32	Baivere	Baviere
1773	sentil	Çentil
1829	delivrea	Deliverea
1855	lusor	Busor
1891	acorler	a acorler
1957	v'averò	n'averò
R. 39	leto	Leito
2010	Baiver	Bavier
2118	e	O
2357	euriter	(eu ri)ter
2473	non	Nen
2531	par	Por
2535	ella	Elle
R. 60	ba[r]un <sup>109</sup>	ba(r)on
2724	noiamant	Inojamant
2761	fogo	Foço
2823	le çucemant	li çu(ç)emant
2864	casuna	Çascuna

<sup>108</sup> Le parentesi tonde nella *Geste Francor* (ed. Morgan 2009) indicano un'aggiunta editoriale.

<sup>109</sup> Segnaliamo con le parentesi quadre le integrazioni editoriali.

Per quanto concerne, invece, la resa del testo, si è deciso di adottare i *Conseils* dell'École des Chartes<sup>110</sup> per l'accentazione e si è anche regolarizzato l'uso della *j*, che in antico francese indica l'affricata palatale sonora /dʒ/, mentre in *Geste Francor* (ed. Morgan 2009) è impiegata anche per la semiconsonante /j/ (un esempio in *inojamant*, qui *supra*).

Diverso, invece, il caso del *Roland* (V4). Nella banca dati *RLALFrI* sono state digitalizzate e caricate le edizioni Robertson-Mellor e Beretta<sup>111</sup>, quest'ultima scelta quale testo di riferimento per l'annotazione<sup>112</sup>. In questa prima fase di lavoro, diversamente da quanto avvenuto per la *Berta*, testo rivisto direttamente sul manoscritto, per il *Roland* (V4) si è seguita un'altra impostazione metodologica, che ha previsto la completa adozione dell'interpretazione critica di Beretta, poi 'tradotta' per quanto possibile nella rispettiva etichettatura.

Nell'introduzione al proprio lavoro, Carlo Beretta specifica che «l'edizione che qui si presenta non vuole e non può essere un'edizione critica»<sup>113</sup>. Lo scopo dell'editore è, infatti, rendere il più fedelmente possibile il diasistema del manoscritto attraverso un'edizione interpretativa, i cui criteri siano volti a rispettare il sovrapporsi dei sistemi che hanno portato al testo di V4<sup>114</sup>. Tale approccio metodologico, come è logico, si è ripercosso anche sull'attività di annotazione, nel momento in cui ci si è dovuti confrontare con un gruppo non irrilevante di lezioni corrotte del manoscritto lasciate a testo dall'editore perché appartenenti al «sistema stilistico-linguistico-metrico dello scriba»<sup>115</sup>. Di seguito si riportano alcuni punti critici incontrati e le rispettive proposte di annotazione.

Nel caso di lezioni palesemente erranee, non interpretabili e non correggibili, si è optato per l'etichetta «SPEC=probl|MORPH=empty» nel campo della morfologia, mentre per POS e lemma si è deciso di mantenere provvisoriamente i valori assegnati in automatico da *Pyrrha*, benché non esatti. Una tale scelta, di certo discutibile e non risolutiva, è dipesa dalla volontà di non ricorrere all'etichetta «OUT» prevista da *Cattex* per «des mots ou segments graphiques qui ne doivent pas être pris en compte dans l'analyse linguistique du texte»<sup>116</sup>. Uno degli scopi della nostra annotazione lessicale, infatti, è quello di tenere traccia anche di forme al momento di difficile comprensione e valutazione, in attesa di poter affinare maggiormente l'analisi e, eventualmente, di poter effettuare una revisione del testo. In questo primo gruppo di casi rientrano ad esempio le forme *uentignira*,

<sup>110</sup> Viellard – Guyotjeannin 2014.

<sup>111</sup> Cfr. rispettivamente *Roland* (V4, ed. Robertson-Mellor 1980) e *Roland* (V4, ed. Beretta 1995).

<sup>112</sup> Al momento di licenziare questo lavoro erano state analizzate e annotate le lasse 1-70 e 281-306.

<sup>113</sup> *Roland* (V4, ed. Beretta 1995): XXI.

<sup>114</sup> Per la nozione di diasistema cfr. Segre 1979: 53-70.

<sup>115</sup> *Roland* (V4, ed. Beretta 1995): XXXIII.

<sup>116</sup> Cfr. Guillot – Prévost – Lavrentiev 2013b: 31.

*innertent* ed *escre* (nei passi «Mai quel culvert non uentignira mia», «Quatro baxans i è inuertent ad òr fi», «Pluvia non cait, rosee n'i ait escre»)¹¹⁷.

Ci sono poi le cosiddette «correzioni mentali»¹¹⁸, nozione che Beretta mutua da Segre, discutendole di volta in volta nelle note al testo. In questi casi, l'annotazione lessicale che ne è seguita ha tenuto conto delle ipotesi formulate dall'editore nel commento, mantenendo a testo la forma grafica dell'edizione che riproduce quella del testimone manoscritto. Ad esempio, la forma *fin* in «Là sede Marsilio, ke Spagna ten in fin» (*Roland* (V4 ed. Beretta 1995): 24, v. 320) è interpretabile come una deformazione di *fin* 'feudo', causata probabilmente dalla ripetizione dell'aggettivo omografo in chiusura del verso precedente. Tale occorrenza è stata quindi lemmatizzata come segue: lemma «fief», POS «NOMcom», MORPH «SPEC=r|NOMB.=s|GENRE=m|CAS=r». Allo stesso modo è stato possibile lemmatizzare la forma *enfrançi* in «Les enfrançi responderà: "El est Gaino li licer"» (ed. Beretta 1995: 252, v. 3964), considerata errore per *enfant*: lemma «enfant», POS «NOMcom», MORPH «SPEC=it|NOMB.=p|GENRE=m|CAS=n». La forma *çe* di «Atant çe Malprimos de Borgal» (ed. Beretta 1995: 60, v. 844), probabile fraintendimento dell'avverbio *et*, è stata annotata: lemma «ez», POS «ADVgen», MORPH «SPEC=probl|MORPH=empty».

Da segnalare in questa sede, inoltre, un altro tipo di 'correzione mentale', che ha comportato un'ulteriore ricaduta nel processo di annotazione, e che è rappresentata ad esempio dalla forma *aprer* in «Au roi Marxillio, a Roncival aprer» (*Roland* (V4, ed. Beretta 1995): 263, v. 4174). Ritenuta storpiatura dell'espressione *as prez* o *al pré*, per essere etichettata ha in primo luogo richiesto l'intervento diretto sul testo caricato in *Pyrrha* e la suddivisione in due *tokens* differenti, *a#* e *#prer*. In secondo luogo, le due forme sono state etichettate rispettivamente: lemma «a3», POS «PRE», MORPH «empty»; lemma «pré», POS «NOMcom», MORPH «SPEC=r|NOMB.=s|GENRE=m|CAS=r».

Oltre a questi aspetti più strettamente ecdotici, l'aumentare dei preziosi dati linguistici risultanti dall'annotazione ha portato a riflettere sulle modalità e sui metodi tramite i quali rappresentarli. Lo sviluppo di un nuovo modello virtuoso di edizione dei testi nella quale possano essere integrate e, successivamente, anche interrogate tutte queste informazioni linguistiche e lessicografiche diventa, dunque, un'esigenza sempre più viva. Non si tratterebbe di edizioni cartacee successivamente digitalizzate, ma di vere e proprie edizioni per le quali il digitale non è più solo il *medium* tramite il quale questi testi vengono fruiti, ma anche, e soprattutto, la base di partenza per il lavoro, con un ruolo trasformativo per la metodologia utilizzata¹¹⁹. Certo, il processo richiede tempo e verrà in un primo momento applicato solo alle nuove edizioni in via di sviluppo: non si andrà, dun-

¹¹⁷ *Roland* (V4, ed. Beretta 1995): 12, v. 143; 39, v. 538; 65, v. 922.

¹¹⁸ Cfr. Segre 1974.

¹¹⁹ Cfr. Pierazzo – Mancinelli 2020: 9.

que, a sostituire in toto tutte le edizioni presenti nel *Repertorio*, ma ci si affiancherà a quest'ultime, nella speranza che questa pratica possa diventare un modello virtuoso da seguire negli anni a venire. Il nuovo modello si fonda sulle buone pratiche delle *Digital Scholarly Editions*<sup>120</sup> e prevede di integrare:

- la riproduzione digitale del manoscritto (laddove il *copyright* lo permetta), secondo il protocollo di scambio dati *IIIF* (International Image Interoperability Framework)<sup>121</sup>;
- la trascrizione diplomatica e la trascrizione interpretativa, che possono dunque essere direttamente confrontate con il codice;
- il testo e l'apparato critico;
- i metadati linguistici dell'annotazione.

Proprio su quest'ultimo punto è in previsione lo sviluppo di uno strumento sofisticato di interrogazione del testo e di tutti i metadati linguistici (Lemma, *Part of speech* e *MORPH tagging*) che permetterà anche di affinare le ricerche con i *tags* utilizzati durante l'annotazione, come l'etichetta «SPEC=it», e tramite il quale sarà possibile porre delle *query* e svolgere così delle analisi sia dal punto di vista quantitativo, sia da quello qualitativo. Inoltre, l'accostamento del *facsimile* del manoscritto all'edizione permette una lettura e un'analisi da vicino per ogni carta del codice, stimolando così il lettore a non fermarsi solo alla lettura del testo, ma a concentrare altresì l'attenzione sugli aspetti codicologici dei manoscritti medievali.

La struttura gerarchica del documento, così come previsto dal modello di *mark-up* utilizzato, ovvero quello delle linee guida *TEI*<sup>122</sup>, ha ispirato anche la modalità di lettura del testo, che ha come obiettivo quello di guidare il lettore nella complessa granularità dei testi franco-italiani:

Carta del manoscritto  
 Episodio del testo  
     Lassa  
         Verso  
             *Token*  
                 Lemma, POS, MORPH

<sup>120</sup> Cfr. per es. Young 2015.

<sup>121</sup> Cfr. <https://iiif.io/>.

<sup>122</sup> Cfr. TEI P5: *Guidelines for Electronic Text Encoding and Interchange*.

<sup>123</sup> Una delle caratteristiche che più contraddistinguono le edizioni digitali è proprio quella di permettere di rappresentare un insieme di dati altrimenti non raggruppabili e presentabili tramite un'edizione cartacea: cfr. per es. Buzzoni 2016.

A questa struttura di tipo gerarchico viene, inoltre, affiancato l'apparato critico, l'analisi di tutti gli elementi paratestuali, dove presenti, e, infine, la marcatura semantica delle regioni di testo sul facsimile del codice (lassa, verso, miniature, rubriche, *maniculae*, ecc.) e la rispettiva corrispondenza nella trascrizione. L'obiettivo è quello di presentare una tale quantità di dati, facilmente esplorabile tramite un'interfaccia *user friendly*, che non sarebbe altrimenti possibile rappresentare tramite il supporto cartaceo tradizionale<sup>123</sup>.

È chiaro, dunque, che un modello di questo tipo, fondato su una modellizzazione e su una marcatura minuziosa del testo, si distanzia in modo netto dalle edizioni digitalizzate finora presenti nel *RLALFrI*. Un cambio di direzione che porterebbe alla costruzione di un'edizione nella quale il testo viene arricchito con una serie di metadati selezionati in base agli scopi dell'editore, andando perciò a modificare il metodo di lavoro e comportando così che «the act of transcribing will therefore now be considered together with the act of encoding the transcribed information»<sup>124</sup>.

Il testo che funge da prototipo per lo sviluppo delle nuove edizioni scientifiche digitali *RLALFrI* è il *Roman d'Alexandre* del manoscritto Venezia, Biblioteca del Museo Correr, Correr 1493, dal momento che, per la varietà di fenomeni che presenta, risulta essere un campo di prova ottimale per la costruzione di un modello per l'intero *corpus*. Sulla base del lavoro svolto sul codice, si propone di seguito un'ipotesi di rappresentazione digitale in formato XML/TEI di alcuni fenomeni<sup>125</sup>.

Lo scioglimento delle abbreviazioni dalla trascrizione diplomatica a quella interpretativa verrà rappresentato tramite l'elemento **<choice>**, all'interno del quale vengono annidati gli elementi **<abbr>** e **<expand>**:

```
<choice> <abbr type= "NotaTironiana" > 7 </abbr>
<expand cert="high" resp="#ed1" > et </expand>
</choice>.
```

Per l'ammodernamento del testo (come l'utilizzo di segni diacritici, la separazione delle parole secondo l'uso moderno o l'inserimento delle maiuscole) vengono utilizzati gli elementi **<orig>** e **<reg>** sempre racchiusi in **<choice>**:

```
<choice> <orig> dalexandre </orig> <reg
type="InserimentoDiacriticoeMaiuscola" > d'Alexandre
</reg> </choice>.
```

<sup>124</sup> Pierazzo 2011: 468.

<sup>125</sup> La codifica che si propone in questa sede deve essere considerata come un'ipotesi. Pertanto, in futuro, potrà essere soggetta a raffinamenti, modifiche e migliorie, poiché lo stato del lavoro è ancora *in fieri*.



Le correzioni sul manoscritto saranno definite tramite gli elementi `<sic></sic>` all'interno del quale viene posta la forma errata e `<corr resp="#ed1"> </corr>` dove invece viene proposta la correzione, sempre racchiusi dall'elemento `<choice>`.

Note e commenti discorsivi al testo verranno marcati tramite il tag `<note>`, distinguendo anche la nota editoriale da eventuali note a margine tramite l'attributo `type`, mentre l'attributo `target` viene utilizzato per indicare l'elemento cui la nota si riferisce. Nonostante la maggior parte dei testi del francese d'Italia siano tramandati da un unico testimone, l'apparato critico avrà comunque il suo spazio fondamentale all'interno dell'edizione e verrà costruito con l'utilizzo del tag `<app>`<sup>126</sup>. Tramite questo elemento sarà possibile, infatti, riportare varianti di lettura (`<app> <lem> ...</lem> <rdg> ...</rdg> </app>`) oppure lacune o impossibilità di lettura tramite il tag `</gap>`. L'elemento `<app>`, in ogni caso, permette una certa libertà di movimento, cosicché ogni apparato possa essere adattato alle caratteristiche filologiche e testimoniali di ogni singolo testo.

In presenza di marginalia, come rubriche o note al miniatore, il fenomeno è stato rappresentato tramite l'elemento `<additions>` nella descrizione fisica del codice. Questo tag permette di segnalare la presenza di miniature:

```
<decoDesc> <decoNote type="miniatures"> Presente al
lato sinistro della colonna di scrittura una
miniatura rappresentante sette maestri che
impartiscono lezioni al giovane Alessandro <list>
<item> <locus> fol. 1r</locus> <mentioned> De li
.vii. maistre que aprendoit Alexandre </mentioned>
</item> </list> </decoNote> </decoDesc>;
```

o la presenza di annotazioni aggiuntive al testo (note al miniatore, appunti del copista e così via), con la possibilità anche di collegare la scrittura di altre mani con l'attributo `<hand>` e richiamando la definizione data in `<handNote>`.

Infine, per quanto riguarda l'annotazione linguistica e l'interrogazione delle forme la codifica è ancora in fase di sperimentazione. Gli elementi che verranno rappresentati sono il lemma, *POS* e *MORPH* e le eventuali etichette di specifica («SPEC=it», «SPEC=rim» eccetera) racchiusi nell'elemento `<w>` (ad esempio: `<w lemma="sustance" pos="NOMcom" msd="S =it|N=s|G=f|C=r">sustançe</w>`). Il lavoro, in questo momento, si sta concentrando sulla definizione di un metodo efficace per l'integrazione dei metadati derivati dall'annotazione linguistica direttamente nel documento di codifica XML/TEI

<sup>126</sup> Al momento si sta utilizzando un approccio di codifica *inline*.

nel modo più automatico possibile, cercando dunque di associare direttamente ogni parola del testo alla rispettiva annotazione, evitando un lavoro manuale del ricercatore che risulterebbe altamente dispendioso.

## 6.2. *La redazione del DiFrI*

Il *DiFrI* dovrebbe consentire di indagare l'elaborazione di una *scripta* letteraria, articolata in più varietà, che al francese amalgama in varia misura e con dinamiche da stabilire volta per volta forme e voci locali, regionali e sovraregionali della Penisola italiana.

Come abbiamo anticipato, il nostro dizionario di riferimento è il T–L: le entrate del *DiFrI* ne riprendono i lemmi, come accade anche nel *software* per l'annotazione del francese antico implementato dall'École des Chartes di Parigi, *Pyrrha* (cfr. *supra*, § 3.2).

Rispetto a questa scelta un problema non indifferente è rappresentato dalla necessità di inserire nuovi lemmi. Ci si può, ad esempio, chiedere se l'aggettivo dimostrativo *questo* debba essere ricondotto al francese *cist* (< ISTE) oppure se sia preferibile assegnargli un'entrata autonoma. In questo caso la decisione è stata facile da prendere, perché l'etimo dell'aggettivo *questo* (< ECCU(M) ĪSTU) è diverso da quello del corrispondente francese (*i*)*cest* (< ECCE ISTU) e appare del tutto legittimo creare due voci con all'interno dei rinvii reciproci.

Volendo portare un secondo esempio, l'etimologia è naturalmente una guida importante anche per l'avverbio *ja* 'già' dal latino IAM<sup>127</sup>, che per alcune delle sue forme (*c'*, *ça*, *gia*, *ja*, *jasoit*, *ya*, *z'a*) è omofono a *ça* 'qua' dal latino ECCU(M) HAC, secondo la tipica omofonia franco-italiana non sempre riconosciuta nei glossari delle edizioni.

Se il lemma creato non è presente nel T–L, esso è preceduto da un asterisco (cfr., ad. es., *\*quilleroil*, 'strumento a percussione usato come richiamo per attirare le quaglie o altri uccelli, quagliere'). Quando ciò accade, si pone il problema di quale forma scegliere per l'entrata. In questo caso, si è optato per la forma caratteristica effettivamente attestata. Le forme flesse sono lemmatizzate tra parentesi quadre secondo l'entrata tradizionale dei dizionari (la forma del singolare per i sostantivi, del caso obliquo singolare maschile per gli aggettivi e dell'infinito per i verbi) e può capitare che questi simboli siano compresenti, perché talvolta è stato necessario ricostruire l'ipotetica entrata del singolare di un lemma non presente nel T–L, come accade per [*\*juber*] 'giubettaio', di cui è attestata solo la forma plurale *jubers*, o [*\*naiscencē*] 'escrescenza, tumore' per l'attestato *naiscences*, e così via.

Al T–L si affiancheranno i glossari delle singole edizioni e gli studi sul franco-italiano e sul francese d'Italia. Ritourneremo, inoltre, più volte sui testi, che saranno

<sup>127</sup> Cfr. *FEW* s.v. *jam*; *DEAF* s.v. *ja*.

letti tutti integralmente. Man mano che l'annotazione andrà avanti, potremo affinare la redazione delle voci, individuando altre forme grafiche o lemmi inizialmente sfuggiti, e si entrerà progressivamente sempre di più nel dettaglio. Uno dei vantaggi dei mezzi informatici è proprio quello di poter modificare e migliorare i risultati della ricerca integrando le novità in modo non troppo oneroso. È per questo che, accanto alla data della messa in linea di ogni voce, è stata posta la data che registra l'ultimo aggiornamento. Per la lettera Q questo processo è stato ben testato e si trova in una fase più avanzata, perché le entrate non sono troppo numerose e, dunque, questa lettera è stata scelta come una sorta di campo di prova per verificare la bontà del percorso di affinamento ideato.

I modelli che abbiamo seguito per la redazione delle voci sono il *Tesoro della Lingua Italiana delle Origini* (TLIO), il *Dictionnaire étymologique de l'ancien français* (DÉAF) e il *Dictionnaire du Moyen Français* (DMF), quest'ultimo soprattutto per come sono organizzati i diversi campi semantici secondo una progressiva ramificazione a partire dal significato di base, fino a comprendere le eventuali accezioni metaforiche e figurate.

Nelle voci del *DIFrI* gli esempi sono citati all'interno di ogni rubrica in ordine cronologico. Nell'ordine delle accezioni viene data la priorità al senso proprio, cui seguono i sensi figurati oppure gli usi per estensione, restrizione e metonimia.

### 6.3. Interoperabilità del corpus gold con i corpora UD

La creazione del *corpus gold* del francese d'Italia ha permesso sia di presentare alcuni dati preliminari sulle prestazioni dell'annotazione di *Pyrrha*, sia permetterà nel prossimo futuro una conversione nel formato CoNLL-U (<https://universaldependencies.org/format.html>) previsto dalle *Universal Dependencies* (cfr. *supra*, § 4); si tratta di un processo tutt'altro che semplice, sia per le caratteristiche del formato CoNLL-U, che prevede ben dieci colonne (molte delle quali dedicate, tuttavia, alle dipendenze sintattiche)<sup>128</sup>, sia per l'impiego di un *tagset* 'universale', cioè pensato per essere adattabile a tutte le lingue trattate, e di conseguenza più scarso rispetto ai sistemi di annotazione non *standard* e maggiormente adattivi rispetto alla varietà di riferimento. Ciononostante, l'esperimento di conversione sarà essenziale alla creazione di una risorsa che segue uno *standard de facto* riconosciuto come punto di riferimento internazionale per l'annotazione linguistica, e che potrà essere quindi interoperabile con altre risorse UD riferite a lingue antiche, in particolare il *corpus PROFITEROLE* ([https://universaldependencies.org/treebanks/fro\\_profiterole/index.html](https://universaldependencies.org/treebanks/fro_profiterole/index.html)), una raccolta allo stato dell'arte e in continuo aggiornamento, contenente testi annotati di francese antico dal IX al XV secolo,

<sup>128</sup> Cfr. <https://universaldependencies.org/format.html>.

ottenuto anch'esso dalla conversione in CoNLL-U a partire da *Cattex*<sup>129</sup>. L'interoperabilità delle due risorse sarà un aspetto fondamentale, poiché permetterebbe di impiegare i dati messi a disposizione, secondo la logica di *UD*, per l'addestramento di nuovi modelli di annotazione linguistica.

#### 6.4. *Per la conservazione a medio e a lungo termine: CLARIN*

Se, per molti versi, appare oggi quasi scontata la scelta di sfruttare, per il proprio progetto di studio, le potenzialità delle *Digital Humanities*, va detto anche che per il momento sono disponibili relativamente pochi correttivi per la grande fragilità dei prodotti digitali: il problema della rapidissima obsolescenza è ormai ben noto, così come le difficoltà – economiche, informatiche, relative alla questione della sicurezza, ecc. – che rendono spesso assai complicata (e talvolta impossibile) la sopravvivenza di questo tipo di oggetti.

Per assicurare ai prodotti scaturiti dal nostro progetto di studio del francese d'Italia una continuità nel medio e lungo termine, e d'altro canto per renderli facilmente accessibili alla comunità degli studi, abbiamo deciso di depositare il *corpus* in *CLARIN*.

*CLARIN* (Common Language Resources and Technology Infrastructure; <https://www.clarin.eu/>) è un'infrastruttura digitale distribuita, con centri partecipanti – università, centri di ricerca, biblioteche e archivi pubblici – situati in più stati europei, che fornisce un accesso facile e sostenibile a un'ampia gamma di dati multimodali (testo, audio, video) e strumenti linguistici. Esso rende inoltre disponibili strumenti avanzati che permettono di esplorare, interrogare, analizzare o combinare i repertori e collezioni di dati che conserva.

Il deposito del *corpus DiFrI* in *CLARIN* risponde pienamente, come si vede, alla volontà di conformarsi ai principi *FAIR* e della scienza aperta.

#### 7. *Prime conclusioni*

Nei paragrafi precedenti abbiamo descritto le diverse linee lungo le quali si va sviluppando lo studio del 'francese d'Italia': la costruzione e l'annotazione del *corpus DiFrI*, così come li abbiamo brevemente presentati nei paragrafi precedenti, ne rappresentano un momento essenziale.

Il progetto *FrIngE*, però, si giova dell'esperienza già maturata dal gruppo di lavoro: come si accennava (cfr. *supra*, § 3.2), l'annotazione linguistica fa propri anche risultati già raggiunti grazie allo studio dei testi francesi d'Italia. Si deve, in particolare, ad Andrea Beretta una prima analisi linguistica complessiva dei testi contenuti nell'*Antologia del francese d'Italia*<sup>130</sup>: essa evidenzia la presenza di tratti che

<sup>129</sup> Cfr. Prévost (*et alii*) 2024.

<sup>130</sup> Cfr. Gambino – Beretta 2023: XLV-XLVIII.

appaiono tipici del ‘francese internazionale’, frutto di un processo di koinizzazione che porta alla diffusione della lingua d’*oïl* in Europa e nel Mediterraneo. Accanto a questi tratti, se ne osservano altri che paiono invece riconducibili direttamente all’appropriazione in area italo-romanza della tradizione linguistica e letteraria antico-francese<sup>131</sup>.

In questo senso, come si è detto, parliamo di ‘francese d’Italia’ per denominare quella varietà oitanica (o quell’insieme di varietà a base oitanica) che caratterizza i testi francesi copiati o composti da scriventi di lingua madre italiana (cfr. *supra*, §§ 1-3). Rispetto alla tradizione degli studi che identifica questa varietà linguistica – o suoi sottoinsiemi – come ‘franco-italiano’, riteniamo che sia più corretto parlare di ‘francese d’Italia’: adottare questa denominazione significa, infatti, rinunciare programmaticamente a una visione che analizzi i dati linguistici in un’ottica rigidamente dicotomica che oppone il ‘francese’ e l’‘italiano’, in favore di una valutazione qualitativa che descriva i singoli fatti osservabili nel quadro ampio della Romània medievale (cfr. *supra*, § 2). L’ipotesi è che l’appropriazione della *koinè* francese internazionale avvenuta in area italo-romanza secondo la modalità del ‘metastrato’<sup>132</sup> abbia portato nella Penisola (con singole specificità diatopiche) alla diffusione di una nuova *koinè*. Il ‘francese d’Italia’ si configura, dunque, come una particolare branca di quell’espansione che, tra Due- e Trecento, vide affermarsi la lingua e la tradizione francesi come veicolo di espressione e di cultura in tutto il bacino del Mediterraneo.

## Bibliografia

### I. Manoscritti

<i>a</i>	Venezia	Biblioteca Nazionale Marciana	Str. App. 14 = 279
<i>BovoU</i>	Udine	Archivio Capitolare	Fondo Nuovi manoscritti 736.28
<i>k</i>	Bruxelles	Bibliothèque Royale	IV 1208
V4	Venezia	Biblioteca Nazionale Marciana	francese 4
V13	Venezia	Biblioteca Nazionale Marciana	francese 13
V20	Venezia	Biblioteca Nazionale Marciana	francese 20
Z	Zagreb	Zagrebačke Metropolitane	MR 92

<sup>131</sup> Sono stati rilevati fatti per lo più grafici e fonologici: cfr. Gambino – Beretta 2023: XLV-XLVIII.

<sup>132</sup> Si riprende qui il passo di Folena citato *supra*, § 2.

## II. Opere

*Attila in prosa*

*The Story of Attila in Prose*, A Critical Edition and Translation of the *Estoire d'Atile en prose*, edited and translated by Roberto Pesce and Logan E. Whalen, London – New York, Routledge, 2022 («Routledge Medieval translations»).

*Bataille de Gamenario*

Giuseppe Cerrato, *La Battaglia di Gamenario (MCCCXLV)*. Testo antico francese da un codice ms. della Cronica del Monferrato di Benvenuto San Giorgio nell'Archivio Generale di Stato di Torino, con illustrazioni e schiarimenti, in «Atti della Società Ligure di Storia Patria», XVII/1 (1885), pp. 381-542.

*Berta da li pé grant*

*Berta da li pé grandi*, a cura di Michela Scattolini, Alessandria, Edizioni dell'Orso, 2009 («Gli orsatti», 31).

*Canzoni francesi*

Francesca Gambino, *Trentatré liriche franco-italiane trascritte nel codice Strozzi-Magliabechiano Cl. VII (1040) della Biblioteca Nazionale Centrale di Firenze*, nuova edizione per il RLALFrI, 2015.

*Devisement dou monde*

Marco Polo, *Le 'Devisement dou monde'*, t. 1, *Testo*, a cura di Mario Eusebi, t. 2, *Glossario*, a cura di Eugenio Burgio, Edizioni Ca' Foscari, Venezia, 2018 («Filologie medievali e moderne», 16; «Filologie medievali e moderne. Serie occidentale», 13).

*Dittamondo*

Fazio degli Uberti, *Il Dittamondo e le rime*, a cura di Giuseppe Corsi, vol. I, *Il Dittamondo*, Bari, Laterza, 1952 («Scrittori d'Italia»).

*Enanchet*

*Enanchet. Dottrinale franco-italiano del XIII secolo sugli stati del mondo, le loro origini e l'amore*, edizione, traduzione e commento a cura di Luca Morlino, Padova, ESEdra, 2017 («Filologia veneta. Testi e studi», 9).

*Entrée d'Espagne*

*L'Entrée d'Espagne, chanson de geste franco-italienne*, publiée d'après le manuscrit unique de Venise par Antoine Thomas, 2 voll., Paris, Didot, 1913 («Société des Anciens Textes Français», 61-62); ristampa anastatica con una premessa di Marco Infurna, Firenze, Olschki, 2007 («Biblioteca Mantovana», 7).

*Estoires de Venise*

Martin da Canal, *Les estoires de Venise*, cronaca veneziana in lingua francese dalle origini al 1275, a cura di Alberto Limentani, Firenze, Olschki, 1972 («Civiltà veneziana. Fonti e testi», 12).

*Foucon de Candie*

*Foucon de Candie*. Edizione della versione franco-italiana conservata dai manoscritti Venezia, Biblioteca Nazionale Marciana, fr. Z 20 e fr. Z 19 (già Zanetti 233 e 232). Versi 1-8809, con apparato e note, a cura di Francesca Gambino, Padova, RLALFrI, 2019 (terza versione).

*Geste Francor*

*La Geste francor di Venezia*, edizione integrale del codice XIII del fondo francese della Marciana, con introduzione, note, glossario, indice dei nomi a cura di Aldo Rosellini, Brescia, La Scuola, 1986 («Pubblicazioni del Centro di linguistica dell'Università cattolica. Saggi e monografie», 6).

*La Geste Francor*, edition of the *Chansons de geste* of MS. Marc. Fr. XIII (=256), with glossary, introduction and notes by Leslie Zarker Morgan, Tempe, ACMRS, 2009 («Medieval & Renaissance texts & studies», 348).

*Ghatrif*

*Moamin et Ghatrif, traités de fauconnerie et des chiens de chasse*. Édition princeps de la version franco-italienne avec 3 planches hors texte, par Håkan Tjerneld, Stockholm, Fritze; Paris, Thiébaud, 1945 («Studia Romanica Holmiensia», 1); edizione alle pp. 255-299.

*Iscrizioni*

Corpus composto da: *Iscrizioni della porta della pescheria di Modena* (1110-1120), *Iscrizione del Duomo di Verona* (1139-1145), *Iscrizione in francese del Krak dei Cavalieri in Siria* (1240-1250), *Iscrizione in francese della Cattedrale di San Nicola a Famagosta* (1311), *Graffito in versi di Bonifacio I di Challant nel castello di Fénis in Valle d'Aosta* (1402), *Iscrizioni nel bassorilievo Aldemoresco della Chiesa di San Lorenzo Maggiore a Napoli* (1421), *Iscrizione muraria del castello di Issogne in Valle d'Aosta* (1489), edizioni digitalizzate per il RLALFrI, 2015-2019.

*La guerra d'Attila*

Giulio Bertoni, *Attila. Poema franco-italiano di Nicola da Casola*, Fribourg, Gschwend, 1907.

*Macario*

Matteo Bergho, *Il 'Macario' della 'Geste Francor' (ms. Marciano Fr. Z 13, 256)*, in-

troduzione, revisione del testo, traduzione e glossario, tesi di laurea magistrale in filologia moderna, relattrice Francesca Gambino, Padova, Università degli Studi di Padova, 2015.

#### *Moamin*

*Moamin et Ghatrif, traités de fauconnerie et des chiens de chasse*. Édition princeps de la version franco-italienne avec 3 planches hors texte, par Håkan Tjerneld, Stockholm, Fritze; Paris, Thiébaud, 1945 («*Studia Romanica Holmiensia*», 1); edizione alle pp. 83-254.

#### *Motti*

*Corpus* composto da: *Motto del re di Cipro Pietro I di Lusignano nel Palazzo Corner-Loredan Piscopia di Venezia* (1346), *Motto di Bernabò Visconti* (1370-1385), *Motto nel soffitto dipinto dello Steri di Palermo* (1377-1380), *Motti in francese di Valentina Visconti* (1389-), *Motto in un cofanetto di area lombarda o piemontese* (1390-1410), *Motto in francese di Giorgio Valperga, conte di Mazze* (1418), *Motto in francese nel tabernacolo della piazza San Pancrazio di Bergamo* (1419-), *Motto dell'impresa di Gian Galeazzo Visconti (ante 1430)*, *Motto del cane alano di Gianfrancesco Gonzaga (ante 1432)*, *Motti in francese nei tarocchi dei Visconti-Sforza* (1442-1450), *Motto della rocca di Vignola* (1444), *Motti delle figlie di Niccolò III d'Este, marchese di Ferrara* (1444-1448), *Motto dell'impresa di Ludovico II Gonzaga* (1448), *Motto in francese di Filippo Maria Visconti nei castelli di Fagnano Olona e di Fontaneto d'Agogna* (1451-1456), *Motto in francese di Giovanni de' Medici* (1463), *Motto della Camera degli Sposi del Castello di San Giorgio a Mantova* (1465-1474), *Motto del broncone di Lorenzo il Magnifico* (1469), *Motto del vessillo di Giuliano de' Medici* (1475), *Motto di Palazzo Pigafetta a Vicenza* (1481), *Motto dei Cavassa di Carmagnola e di Saluzzo* (1483/1518), *Motto del pappagallo di Lorenzo il Magnifico* (1485-1490), *Motto dei Della Chiesa nella "Casa di Davide" a Saluzzo* (1491), *Motto in francese degli Challant, signori di Issogne* (1494/1506-1510), *Motto degli arazzi dei "Mesi" Trivulzio* (1503-1509), edizioni digitalizzate per il RLALFrI, 2014-2020.

#### Niccolò da Verona

Niccolò da Verona, *Opere. Pharsale', 'Continuazione dell'Entrée d'Espagne', Passion'*, a cura di Franca Di Ninni, Venezia, Marsilio, 1992 («Medioevo veneto»).

#### *Passion* di Venezia

*La Passion de Venise*, cod. marc. franc. VI (=226), a cura di Virginio Bertolini, Verona, Bi & Gi, 1986.

#### *Quatre âges de l'homme*

Filippo da Novara, *Les .iiij. tenz d'aage d'ome*. Testo, traduzione e note a cura di Silvio Melani, Padova, Dipartimento di studi linguistici e letterari, Università degli studi di Padova, 2020 (Quaderni di «Francigena», 1).



*Roland* (V4)

*The Franco-Italian Roland (V4)*, edited by Geoffrey Robertson-Mellor, Salford, University of Salford Reprographic Unit, 1980.

*Il testo assonanzato franco-italiano della Chanson de Roland*, cod. marciano fr. IV (= 225), edizione interpretativa e glossario a cura di Carlo Beretta, Pavia, Università degli Studi, 1995.

*Romanzo arturiano di Rustichello da Pisa*

*Il romanzo arturiano di Rustichello da Pisa*, edizione critica, traduzione e commento a cura di Fabrizio Cigni, premessa di Valeria Bertolucci Pizzorusso, Pisa, Pacini, 1994.

*Rubriche e note al miniatore (Roman d'Alexandre Correr 1493)*

“*Le roman d'Alexandre*”. Riproduzione del manoscritto Venezia, Biblioteca Museo Correr, Correr 1493, a cura di Roberto Benedetti, Tricesimo, Vattori, 1998, pp. 35-45.

*Testi veronesi*

*Testi veronesi dell'età scaligera*. Edizione, commento linguistico e glossario, a cura di Nello Bertolotti, Padova, Esedra, 2005 («Vocabolario storico dei dialetti veneti», 6).

*Viaggio di Carlomagno in Oriente*

*Il viaggio di Carlomagno in Oriente*, a cura di Massimo Bonafin, Parma, Pratiche, 1993<sup>3</sup> («Biblioteca medievale», 3).

*Ystoire de li Normant*

*Ystoire de li Normant*. Édition du manuscrit BnF fr. 688 par Michèle Guéret-Laferté, Paris, Champion, 2011 («Les classiques français du Moyen Âge», 166).

III. Studi e strumenti

Baglioni 2016

Daniele Baglioni, *Per una fenomenologia della commutazione di codice nei testi antichi*, in «La lingua italiana», 12 (2016), pp. 9-35.

Beltrami 1999

Pietro Beltrami, *Il Tesoro della lingua italiana delle origini (TLIO) e l'onomastica*, in «Rivista Italiana di Onomastica», V/2 (1999), pp. 349-362.

Beretta 2021

Andrea Beretta, *Nuove ricerche sull'Attila Flagellum Dei di Nicolò da Càsola*, in «Zeitschrift für Romanische Philologie», 137/1 (2021), pp. 252-286, <https://doi.org/10.1515/zrp-2021-0008>.

Beretta – Palumbo 2015

Carlo Beretta e Giovanni Palumbo, *Il franco-italiano in area padana: questioni, problemi e appunti di metodo*, in «Medioevo romanzo», 39 (2015), pp. 52-81.

Beretta Spampinato 1990

Margherita Beretta Spampinato, Voce «Deloc, Daniele (Daniele da Cremona)», in *Dizionario Biografico degli Italiani*, Roma, Istituto della Enciclopedia Italiana, vol. 38, 1990, pp. 190-192.

Buzzoni 2016

Marina Buzzoni, *A protocol for Scholarly Digital Editions? The Italian Point of View*, in *Digital Scholarly Editing: Theories and Practices*, ed. by Matthew J. Driscoll and Elena Pierazzo, Cambridge, Open Book Publisher, 2016 («Digital Humanities Series», 4), pp. 59-82.

Camps (et alii) 2021

Jean-Baptiste Camps, Thibault Clérice, Frédéric Duval, Lucence Ing, Naomi Kanaoka, Ariane Pinche, *Corpus and Models for Lemmatisation and POS-tagging of Old French*, in «Journal of Data Mining and Digital Humanities», 2021, <https://arxiv.org/pdf/2109.11442> [cons. 7. VII. 2024].

Cigni 2011

Fabrizio Cigni, *Manuscrits en français, italien, et latin entre la Toscane et la Ligurie à la fin du XIII<sup>e</sup> siècle*, in *Medieval Multilingualism. The Francophone World and its Neighbours*, edited by Christopher Kleinhenz and Keith Busby, Turnhout, Brepols, 2011 («Medieval texts and cultures of northern Europe», 20), pp. 187-217.

De Felice (et alii) 2018

Irene De Felice, Felice Dell'Orletta, Giulia Venturi, Alessandro Lenci, Simionetta Montemagni, *Italian in the Trenches: Linguistic Annotation and Analysis of Text of the Great War*, in *Proceedings of 5th Italian Conference on Computational Linguistics (CLiC-It)*, a cura di Elena Cabrio, Alessandro Mazzei, Fabio Tamburini, Torino, Accademia University Press, 2018, pp. 160-164.

De Marneffe (et alii) 2021

Marie-Catherine De Marneffe, Christopher D. Manning, Joakim Nivre, Daniel Zeman, *Universal Dependencies*, in «Computational Linguistics», 47/2 (2021), pp. 255-308.

*DÉAF*

*Dictionnaire étymologique de l'ancien français*, fondé par Kurt Baldinger; avec la collaboration de Jean-Denis Gendron et Georges Straka; [puis] publié sous la direction philologique de Frankwalt Möhrenéd, Québec – Tübingen – Paris, PU Laval – Niemeyer – Klincksieck, 1974-2016, <https://deaf.hadw-bw.de/> [cons. 26. VI. 2024].

*DMF*

*Dictionnaire du Moyen Français (1330-1500)*, version 2023 (DMF 2023), ATILF – CNRS – Université de Lorraine, 2009-, <http://www.atilf.fr/dmf> [cons. 27. VI. 2024].

Favaro (*et alii*) 2022

Manuel Favaro, Elisa Guadagnini, Eva Sassolini, Marco Biffi, Simonetta Montemagni, *Towards the Creation of a Diachronic Corpus for Italian: A Case Study on the GDLI Quotations*, in *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA 2022)*, a cura di Rachele Sprugnoli, Marco Passarotti, Marseille, European Language Resources Association (ELRA), 2022, pp. 94-100, <https://aclanthology.org/2022.lt4hala-1.0> [cons. 7. VII. 2024].

Favaro (*et alii*) 2023

Manuel Favaro, Marco Biffi, Simonetta Montemagni, *POS Tagging and Lemmatization of Historical Varieties of Languages. The Challenge of Old Italian*, in «*Italian Journal of Computational Linguistics*», 9/2 (2023), <http://journals.openedition.org/ijcol/1325> [cons. 7. VII. 2024].

*FEW*

*Französisches etymologisches Wörterbuch. Eine Darstellung des galloromanischen Sprachschatzes*, Walter von Wartburg (ed.), Bonn [...], Schroeder [...], 1922-1989, <https://apps.atilf.fr/lecteurFEW/> [cons. 7. VII. 2024].

Folena 1990

Gianfranco Folena, *Culture e lingue nel Veneto medievale*, Padova, 1990 («*Filologia veneta. Testi e studi*», 1).

Formisano 1978

Luciano Formisano, *Per il testo della «Battaglia di Gamenario»*, in «*Studi Piemontesi*», VII (1978), pp. 341-351.

Gambino 2016

Francesca Gambino, *Code-mixing nel Bovo d'Antona udinese*, in «*Francigena*», 2 (2016), pp. 35-130.

Gambino – Beretta 2023

*Antologia del francese d'Italia. XIII-XV secolo*, a cura di Francesca Gambino e Andrea Beretta, Bologna, Pàtron, 2023 («Storia e testi», 4).

Gdf

Frédéric Godefroy, *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX<sup>e</sup> au XV<sup>e</sup> siècle* [...], 10 voll., Paris, Vieweg, 1881-1902 [New York, Kraus Reprint, 1961], <http://micmap.org/dicfro/search/dictionnaire-godefroy> [cons. 27. VI. 2024].

GDLI

*Grande dizionario della lingua italiana*, di Salvatore Battaglia (poi diretto da Giorgio Bàrberi Squarotti), 21 voll., Torino, UTET, 1961-2002; con *Supplemento 2004* e *Supplemento 2009*, diretti da Edoardo Sanguineti, Torino, UTET, 2004 e 2008, e *Indice degli autori citati nei volumi I-XXI e nel Supplemento 2004*, a cura di Giovanni Ronco, Torino, UTET, 2004, <https://www.gdli.it/> [cons. 7. VII. 2024].

Glessgen 1996

Martin D. Glessgen, *Die Falkenbeilkunde des 'Moamin' im Spiegel ihrer volgarizzamenti. Studien zur Romania Arabica*, 2 voll., Band I: *Edition der neapolitanischen und der toskanischen Version mit philologischen Kommentar*, Band II: *Der medizinisch-biologische Wortschatz und seine Übersetzung*, Tübingen, Niemeyer, 1996 («Beihefte zur Zeitschrift für romanische Philologie», 269-270).

Glessgen 1998

Martin D. Glessgen, *Contatti di cultura e contatti di lingua nelle traduzioni dall'arabo in Italia*, in *Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza* (Centro di studi filologici e linguistici siciliani Università di Palermo 18-24 settembre 1995), a cura di Giovanni Ruffino, sezione 6, *Storia Linguistica e culturale del Mediterraneo*, Tübingen, Niemeyer, 1998, pp. 431-444.

Guariglia 2022

Federico Guariglia, *Moamin et Ghatrif. Prolégomènes à une nouvelle édition*, in «Francigena», 8 (2022), pp. 131-170.

Guessard 1857

François Guessard, *Notes sur un manuscrit français de la bibliothèque de S. Marc*, in «Bibliothèque de l'École des Chartes», 18 (1857), pp. 393-414.

Guillot – Prévost – Lavrentiev 2013a

Céline Guillot, Sophie Prévost, Alexei Lavrentiev, *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) [cons. 7. VII. 2024].

Guillot – Prévost – Lavrentiev 2013b

Céline Guillot, Sophie Prévost, Alexei Lavrentiev, *Manuel de référence du jeu Cattetex09, (Version 2.0)*, Lyon, École normale supérieure de Lyon, 8 avril 2013, [http://bfm.ens-lyon.fr/IMG/pdf/Cattetex2009\\_manuel\\_2.0.pdf](http://bfm.ens-lyon.fr/IMG/pdf/Cattetex2009_manuel_2.0.pdf) [cons. 7. VII. 2024].

Holtus – Wunderli 2005

*Franco-italien et épopée franco-italienne*, éd. par Gunter Holtus et Peter Wunderli, in *Grundriss der romanischen Literaturen des Mittelalters*, vol. 3, t. 1/2, fasc. 10 *Les épopées romanes*, Heidelberg, Winter, 2005.

IIIF

*International Image Interoperability Framework*. IIIF is a set of open standards for delivering high-quality, attributed digital objects online at scale. It's also an international community developing and implementing the IIIF APIs. IIIF is backed by a consortium of leading cultural institutions, cfr. <https://iiif.io/> [cons. 7. VII. 2024].

Infurna 2011

Marco Infurna, *Per il testo dell'Entrée d'Espagne*, in *Metafora medievale. Il 'libro degli amici' di Mario Mancini*, a cura di Carlo Donà, Marco Infurna e Francesco Zambon, Roma, Carocci, 2011 («Biblioteca Medievale. Saggi», 29), pp. 121-137.

Italia 2016

Paola Italia, *Il lettore Google*, in «Prassi Ecdotiche della Modernità Letteraria», 1 (2016), pp. 13-26, <https://doi.org/10.13130/2499-6637/6971> [cons. 7. VII. 2024].

Johnson (*et alii*) 2021

Kyle P. Johnson, Patrick J. Burns, John Stewart, Todd Cook, Clément Besnier, William J. B. Mattingly, *The Classical Language Toolkit: An NLP Framework for Pre-Modern Languages*, in *Proceedings of the 59<sup>th</sup> Annual Meeting of the Association for Computational Linguistics and the 11<sup>th</sup> International Joint Conference on Natural Language Processing: System Demonstrations*, a cura di Heng Ji, Jong C. Park, Rui Xia, Online, Association for Computational Linguistics (ACL), 2021, <https://aclanthology.org/2021.acl-demo.3.pdf> [cons. 7. VII. 2024].

Kleinhenz – Busby 2011

*Medieval Multilingualism. The Francophone World and its Neighbours*, edited by Christopher Kleinhenz and Keith Busby, Turnhout, Brepols, 2011 («Medieval texts and cultures of northern Europe», 20).

*LEI*

*Lessico etimologico italiano*, edito per incarico della Commissione per la Filologia Romanza da Max Pfister; poi edito per incarico della Commissione per la Filologia Romanza da Max Pfister e Wolfgang Schweickard, Wiesbaden, Ludwig Reichert, 1984-, <https://lei-digitale.org/> [cons. 27. VI. 2024].

Lenci (*et alii*) 2020

Alessandro Lenci, Simonetta Montemagni, Federico Boschetti, Irene De Felice, Stefano Dei Rossi, Felice Dell’Orletta, Michele Di Giorgio, *Voices of the Great War: A Richly Annotated Corpus of Italian Texts on the First World War*, in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, Marsiglia, European Language Resources Association, 2020, pp. 911-918.

Marcato 2016

Carla Marcato, *Introduzione. Nomi di persona, nomi di luogo e storia della lingua*, in *Nomina sunt...? L’onomastica tra ermeneutica, storia della lingua e comparatistica. Atti delle giornate di studio (Venezia 3-4 marzo 2016)*, a cura di Maria Pia Arpioni, Arianna Ceschin, Gaia Tomazzoli, Venezia, Edizioni Ca’ Foscari, 2016 («Studi e Ricerche», 3), pp. 17-29.

Marazzini – Maconi 2018

Claudio Marazzini, Ludovica Maconi, *Il ‘Vocabolario dinamico dell’italiano moderno’ rispetto ai linguaggi settoriali. Proposta di voce lessicografica per il redigendo VoDIM*, in «Italiano digitale», 7 (2018), pp. 101-120.

Marruncheddu 2008

Sara Marruncheddu, *La versione franco-italiana di un trattato di falconeria: il Moamin di Daniele di Lodi nel confronto con la tradizione latina e i volgarizzamenti italiani*, Tesi di dottorato, Università degli studi di Torino, 2008.

Matsumura

Takeshi Matsumura, *Dictionnaire du français médiéval*, Paris, Les Belles Lettres, 2015.

McDonald – Suleiman 2011

*French Global. A New Approach to Literary History*, edited by Christie McDonald, Susan Rubin Suleiman, New York, Columbia University Press, 2011.

*MFLCOF*

*Medieval Francophone Literary Culture Outside France*, dir. Simon Gaunt, London, King’s College, 2014, <http://www.medievalfrancophone.ac.uk> [cons. 7. VII. 2024].

Minervini 1996

Laura Minervini, *La lingua franca mediterranea*, in «Medioevo romanzo», 20 (1996), pp. 231-301.

Minervini 2010

Laura Minervini, *Le français dans l'Orient latin (XIII<sup>e</sup>-XV<sup>e</sup> siècles). Éléments pour la caractérisation d'une scripta du Levant*, in «Revue de Linguistique Romane», 74 (2010), pp. 119-198.

Minervini 2015

Laura Minervini, *Il francese a Napoli (1266-1442). Elementi per una storia linguistica*, in *Boccaccio e Napoli. Nuovi materiali per la storia culturale di Napoli nel Trecento*. Atti del Convegno «Boccaccio angioino. Per il VII centenario della nascita di Giovanni Boccaccio» (Napoli – Salerno, 23-25 ottobre 2013), a cura di Giancarlo Alfano, Emma Grimaldi, Sebastiano Martelli, Andrea Mazzucchi, Matteo Palumbo, Alessandra Perriccioli Saggese, Carlo Vecce, Firenze, Cesati, 2015, pp. 151-174.

Morato – Schoenaers 2018

*Medieval Francophone Literary Culture Outside France*, edited by Nicola Morato and Dirk Schoenaers, Turnhout, Brepols, 2018 («Medieval texts and cultures of northern Europe», 28).

Morlino 2010

Luca Morlino, *Contributi al lessico franco-italiano*, in «Medioevo letterario d'Italia» 7 (2010), pp. 65-85.

Morlino 2015

Luca Morlino, *Spunti per un riesame della costellazione letteraria franco-italiana*, in «Francigena», 1 (2015), pp. 5-81.

Norme per la redazione del TLIO

Norme per la redazione del *Tesoro della Lingua Italiana delle Origini*, a cura di Pietro G. Beltrami, con la collaborazione dei redattori e revisori del TLIO, versione del 28/03/2024, <http://tlio.ovi.cnr.it/TLIO/NormeTLIO.pdf> [cons. 7. VII. 2024].

Paris 1865

Gaston Paris, *Histoire poétique de Charlemagne*, Paris, Franck, 1865.

Perugi 2008

Maurizio Perugi, «*La parleüre plus delitable*». Osservazioni sulla lingua del Tresor,

in *A scuola con ser Brunetto. La ricezione di Brunetto Latini dal Medioevo al Rinascimento*, Atti del Convegno internazionale di studi (Università di Basilea, 8-10 giugno 2006), a cura di Irene Maffia Scariati, Firenze, Sismel-Edizioni del Galluzzo, 2008, pp. 493-513.

Pfister 1999

Max Pfister, *L'importanza della toponomastica per la storia della lingua nella Galloromania e nell'Italoromania*, in «Rivista Italiana di Onomastica», V/2 (1999), pp. 449-464.

Pierazzo 2011

Elena Pierazzo, *A Rationale of Digital Documentary Editions*, in «Literary and Linguistic Computing» 26, 4, pp. 463-477, <https://doi.org/10.1093/llc/fqr033> [cons. 7. VII. 2024].

Pierazzo – Mancinelli 2020

Elena Pierazzo e Tiziana Mancinelli, *Che cos'è un'edizione scientifica digitale*, Roma, Carocci, 2020 («Le bussole», 599).

Prévost (*et alii*) 2013

Sophie Prévost, Céline Guillot, Alexei Lavrentiev, Serge Heide, *Jeu d'étiquettes morphosyntaxiques CATTEX2009, (Version 2.0)*, Lyon, École normale supérieure de Lyon, 8 avril 2013, [http://bfm.ens-lyon.fr/IMG/pdf/Cattex\\_2009\\_2.0.pdf](http://bfm.ens-lyon.fr/IMG/pdf/Cattex_2009_2.0.pdf) [cons. 16. IX. 2024].

Prévost (*et alii*) 2024

Sophie Prévost, Loïc Grobol, Mathieu Dehouck, Alexei Lavrentiev et Serge Heiden, *Profiterole: un corpus morpho-syntaxique et syntaxique de français médiéval*, in «Corpus», 25 (2024), <https://doi.org/10.4000/corpus.8538> [cons. 7. VII. 2024].

*Pyrrha*

*Pyrrha*, A language independent post correction app for POS and lemmatization, <https://dh.chartes.psl.eu/pyrrha> [cons. 7. VII. 2024].

Qi (*et alii*) 2020

Peng Qi, Zhang Yuhao, Zhang Yuhui, Jason Bolton, and Christopher D. Manning, *Stanza: A Python Natural Language Processing Toolkit for Many Human Languages*, in *ACL2020 System Demonstration*, Online, 2020, <https://aclanthology.org/2020.acl-demos.14/> [cons. 7. VII. 2024].

Quemada 1987

Bernard Quemada, *Notes sur lexicographie et dictionnaire*, in «Cahiers de lexicologie», 51 (1987), pp. 229-242.



Rajna 1887

Pio Rajna, *Frammenti di redazioni italiane del Buovo d'Antona, I. Nuovi frammenti franco-italiani*, in «Zeitschrift für romanische Philologie», 11 (1887), pp. 153-184.

Renzi 1970

Lorenzo Renzi, *Per la lingua dell'Entrée d'Espagne*, in «Cultura neolatina», 30 (1970), pp. 59-87.

RLALFrI

*Repertorio Informatizzato Antica Letteratura Franco-Italiana (RLALFrI)*, diretto da Francesca Gambino, Università degli Studi di Padova, Dipartimento di Studi Linguistici e Letterari, versione 2.0, 2022, [www.rialfri.eu](http://www.rialfri.eu) [cons. 7. VII. 2024].

Robinson 2013

Peter Robinson, *Towards a Theory of Digital Editions*, in «The Journal of the European Society for Textual Scholarship», «Variants», 10 (2013), pp. 110-131, [https://doi.org/10.1163/9789401209021\\_009](https://doi.org/10.1163/9789401209021_009).

Sahle 2016

Patrick Sahle, *What is a Scholarly Digital Edition?*, in *Digital Scholarly Editing: Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, Cambridge, Open Book Publisher, 2016 («Digital Humanities Series», 4), pp. 19-24, <https://www.openbookpublishers.com/books/10.11647/obp.0095> [cons. 7. VII. 2024].

Sanfilippo 2016

Carla Maria Sanfilippo, *L'onomastica ferrarese del primo Trecento e gli Instrumenta fidelitatis*, Padova, libreriauniversitaria.it edizioni, 2016 («Storie e linguaggi», 18).

Shillingsburg 2014

Peter Shillingsburg, *Development Principles for Virtual Archives and Editions*, in «The Journal of the European Society for Textual Scholarship», «Variants», 11 (2014), pp. 9-28, [https://doi.org/10.1163/9789401212113\\_002](https://doi.org/10.1163/9789401212113_002).

Segre 1974

Cesare Segre, *Correzioni mentali per la Chanson de Roland*, in Id., *La tradizione della 'Chanson de Roland'*, Milano-Napoli, Ricciardi, 1974, pp. 184-193.

Segre 1979

Cesare Segre, *Semiotica filologica. Testo e modelli culturali*, Torino, Einaudi, 1979 («Einaudi paperbacks», 100).

## Segre 1995

Cesare Segre, *La Letteratura franco-veneta*, in *Storia della letteratura italiana*, diretta da Enrico Malato, vol. 1. *Dalle Origini a Dante*, Roma, Salerno, 1995, pp. 631-647.

## Stein 2016

Achim Stein, *Old French Dependency Parsing: Results of Two Parsers Analysed from a Linguistic Point of View*, in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, a cura di Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, Paris, European Language Resources Association (ELRA), 2016, pp. 355-375.

## Straka – Straková 2017

Milan Straka, Jana Straková, *Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe*, in *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, a cura di Jan Hajič, Dan Zeman, Vancouver, Association for Computational Linguistics, 2017, pp. 88-99.

TEI P5: *Guidelines for Electronic Text Encoding and Interchange*

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> [cons. 7. VII. 2024].

## T–L

*Alfranzösisches Wörterbuch*, Adolf Toblers nachgelassene Materialien, bearbeitet und herausgegeben von Erhard Lommatzsch, weitergeführt von Hans Helmut Christmann, vollendet von Richard Baum und Willy Hirdt unter Mitwirkung von Brigitte Frey, 12 voll., Berlin – Wiesbaden – Stuttgart, Steiner, 1925-2002.

## TLFi

*Trésor de la langue Française informatisé*, ATILF - CNRS & Université de Lorraine, <http://www.atilf.fr/tlfi> [cons. 7. VII. 2024].

## TLIO

*Tesoro della lingua italiana delle Origini*, dir. Paolo Squillaciotti, CNR-OVI, <http://tlio.ovi.cnr.it/TLIO/> [cons. 27. VI. 2024].

## UD

*Universal Dependencies*, versione 2.14 (283 treebanks, 161 lingue), rilasciata su

LINDAT/CLARIAH-CZ il 15.V.2024 (handle: <http://hdl.handle.net/11234/1-5502>), <https://universaldependencies.org/> [cons. 9. VII. 2024].

Valenti 2014

Iride Valenti, *Gallicismi e francesismi nel siciliano. Vocabolario storico-etimologico*, Leonforte, Euno, 2014 («Collana di studi e ricerche. Sezione linguistica e letteratura»).

Varvaro 1981

Alberto Varvaro, *Lingua e storia in Sicilia*, Palermo, Sellerio, 1981.

Varvaro 1996

Alberto Varvaro, 116. *Gemeinromanische Tendenzen XII. Literatursprachenbildung / Tendenze comuni alle lingue romanze XII. La formazione delle lingue letterarie*, in *Lexikon der Romanistischen Linguistik (LRL). Band II/1. Latein und Romanisch: Historisch-vergleichende Grammatik der romanischen Sprachen*, eds. Günter Holtus, Michael Metzeltin, Christian Schmitt, Tübingen, Niemeyer, 1996, pp. 528-537.

VEV

*Vocabolario storico etimologico del veneziano*, diretto da Lorenzo Tomasin e Luca D'Onghia, sviluppato dall'Univeristà di Losanna, dalla Scuola Normale Superiore di Pisa e dall'Istituto OVI - Opera del Vocabolario italiano del Cnr di Firenze, 2024, <http://vev.ovi.cnr.it/>.

VF

*The Values of French*, dir. Simon Gaunt, London, King's College, 2024, <https://tvof.ac.uk/> [cons. 7. VII. 2024].

Vielliard – Guyotjeannin 2014

*Conseils pour l'éditions des textes médiévaux*, vol. 1, *Conseils généraux*, a cura di Françoise Vielliard e Olivier Guyotjeannin, Paris, Ecole nationale des Chartes, 2014.

Viscardi 1941

Antonio Viscardi, *Letteratura franco-italiana*, Modena, Società Tipografica Modenese, 1941.

Young 2015

John Young, *Considering the Scholarly Edition in the Digital Age: A White Paper of the Modern Language Association's Committee on Scholarly Editions*, Modern Language Association, Committee on Scholarly Editions, 2015, <https://scholarlyeditions.mla.hcommons.org/cse-white-paper/> [cons. 7. VII. 024].

Zaccarello 2017

Michelangelo Zaccarello, *La letteratura italiana nel contesto della svolta digitale: serve più 'teoria dell'edizione'?*, in «Ecdotica», 14 (2017), pp. 148-162.

Zinelli 2016

Fabio Zinelli, *Espaces franco-italiens: les italianismes du français médiéval*, in *La régionalité lexicale du français au Moyen Âge*. Volume thématique issu du colloque de Zurich (7-8 sept. 2015) organisé sous le patronage de la Société de Linguistique Romane, édition par Martin D. Glessgen et David A. Trotter, Strasbourg, ELiPhi, 2016 («Travaux de linguistique romane. Lexicologie onomastique et lexicographie»), pp. 207-269.

Zinelli 2018

Fabio Zinelli, *Inside/Outside Grammar: The French of Italy between Structuralism and Trends of Exoticism*, in *Medieval Francophone Literary Culture Outside France*, edited by Nicola Morato and Dirk Schoenaers, Turnhout, Brepols, 2018 («Medieval texts and cultures of northern Europe», 28), pp. 31-72.

Zinelli 2020

Fabio Zinelli, *De la France-Italie à l'Italo-France (ou de l'histoire littéraire comme délocalisation)*, in *Transferts culturels franco-italiens au Moyen Âge / Trasferimenti culturali italo francesi*, études réunie par/a cura di Roberto Antonelli, Joëlle Ducos, Claudio Galderisi, Arianna Punzi, Turnhout, Brepols, 2020 («Bibliothèque de Transmédié», 8), pp. 169-199.

Zinelli 2021

Fabio Zinelli, *Traditions manuscrites d'Outremer ('Trésor', 'Sidrac', 'Histoire Ancienne jusqu'à César')*, in *En français hors de France. Textes, livres, collections du Moyen Âge*, études recueillies par Fabio Zinelli et Sylvie Lefèvre, avec la collaboration de Sarah Delale, Strasbourg, ELiPhi, 2021 («Travaux de littératures romanes»), pp. 59-107.

Zinelli – Lefèvre 2021

*En français hors de France. Textes, livres, collections du Moyen Âge*, études recueillies par Fabio Zinelli et Sylvie Lefèvre, avec la collaboration de Sarah Delale, Strasbourg, ELiPhi, 2021 («Travaux de littératures romanes»).