



RiDESIN

Rivista del Dizionario Etimologico
e Storico del Napoletano

IV/1 (2026)



Federico II University Press



fedOA Press



RiDESIN

Rivista del Dizionario Etimologico
e Storico del Napoletano

IV/1 (2026)

Federico II University Press



fedOA Press



RiDESN

Rivista del Dizionario Etimologico e Storico del Napoletano

Direzione

Nicola De Blasi (Università di Napoli "Federico II")

Francesco Montuori (Università di Napoli "Federico II")

Comitato scientifico

Giovanni Abete (Università di Napoli "Federico II"), **Marcello Barbato** (Università di Napoli "L'Orientale"), **Marina Castiglione** (Università di Palermo), **Michele Colombo** (Stockholms universitet), **Paolo D'Achille** (Università di Roma "Roma Tre"), **Chiara De Caprio** (Università di Napoli "Federico II"), **Luca D'Onghia** (Università di Siena), **Rita Fresu** (Università di Cagliari), **Mariafrancesca Giuliani** (Istituto Opera del Vocabolario Italiano (OVI) del CNR), **Pär Larson** (Istituto Opera del Vocabolario Italiano (OVI) del CNR), **Marco Maggiore** (Università di Pisa), **Elda Morlicchio** (Università di Napoli "L'Orientale"), **Alessandro Parenti** (Università di Trento), **Emiliano Picchiorri** (Università di Chieti-Pescara "G. D'Annunzio"), **Rosa Piro** (Università di Napoli "L'Orientale"), **Elton Prifti** (Universität des Saarlandes), **Carolina Stromboli** (Università di Salerno), **Lorenzo Tomasin** (Université de Lausanne), **Giulio Vaccaro** (Università di Perugia), **Zeno Verlato** (Istituto Opera del Vocabolario Italiano (OVI) del CNR), **Raymund Wilhelm** (Universität Klagenfurt).

Comitato scientifico onorario

Patricia Bianchi (Università di Napoli "Federico II"), **Rosario Coluccia** (Università del Salento), **Michele Cortelazzo** (Università di Padova), **Franco Fanciullo** (Università di Pisa), **Claudio Giovanardi** (Università di Roma "Roma Tre"), **Rita Librandi** (Università di Napoli "L'Orientale"), **Carla Marcato** (Università di Udine), **Ivano Paccagnella** (Università di Padova), **Edgar Radtke** (Universität Heidelberg), **Giovanni Ruffino** (Università di Palermo), **Wolfgang Schweickard** (Universität des Saarlandes), **Rosanna Sornicola** (Università di Napoli "Federico II"), **Ugo Vignuzzi** (Università di Roma "La Sapienza").

Comitato editoriale

Lucia Buccheri (Università di Napoli "Federico II"), **Cristiana Di Bonito** (Università di Napoli "Federico II"), **Salvatore Iacolare** (Università di Napoli "Federico II"), **Vincenzina Lepore** (Università di Napoli "L'Orientale"), **Andrea Maggi** (Scuola Superiore Meridionale), **Claudia Tarallo** (Università di Napoli "L'Orientale"), **Lidia Tornatore** (Università di Salerno).

Comitato di gestione

Duilia Giada Guarino

Beatrice Maria Eugenia La Marca

I contributi delle sezioni 1, 2 e 4 sono sottoposti a una revisione a doppio cieco.

In copertina e all'interno della rivista si riproduce un inserto dell'affresco *Fanciulla*, *cd. Saffo*, Napoli, MANN, Affreschi Inv. 9084. La fotografia impressa in copertina, realizzata da Giuseppe Gaeta, è un dettaglio di una vetrata di Palazzo Zevallos (NA).

La «Rivista del Dizionario Etimologico e Storico del Napoletano» è una rivista scientifica semestrale realizzata con Open Journal System ed edita da FedOA - Federico II University Press, Centro di Ateneo per le Biblioteche "Roberto Pettorino", Università degli Studi di Napoli Federico II (Piazza Bellini 59-60 - 80138 Napoli) | ISSN 2975-0806.

Indice

<i>Premessa</i>	6
-----------------	---

Saggi

Paolo Greco, <i>La base di conoscenza testuale del progetto MEDITA e i metadati. Il “Manuale” per la scheda bibliografica</i>	9
Mariafrancesca Giuliani, <i>La lemmatizzazione del corpus MED.ITA: orientamenti e indicazioni operative</i>	35
Fabio Aprea e Jacopo Gesiot, <i>Testi, lessico e ambiti socioculturali: note sull’iperlemmatizzazione</i>	83
Fabio Aprea, <i>Convenzioni grafiche e segni editoriali: dalle edizioni a MED.ITA</i>	95
Gelsomina Razza, <i>Alcuni tratti antinapoletani nei dialetti dell’Agro Nolano: il livello fonetico</i>	105

Autori e testi

Giovanni Urraci, « <i>La crema, cioè la telerina di sopra, ch’è il capo di latte</i> ». <i>Glosse e sinonimi nell’Opera (1570) di Bartolomeo Scappi</i>	145
Claudia Tarallo, Gloria Guida e Claudia Grossi, <i>Documenti degli antichi banchi pubblici napoletani (XVI-XVII sec.): testi e prime considerazioni</i>	229
Giovanni Maddaloni, « <i>Cultore e discernitore delle dovizie del nostro dialetto</i> »: <i>il teatro di Aniello Costagliola in prospettiva lessicografica</i>	305

Discussioni e cronache

<i>Il volgare nel Lazio centrale fra Trecento e Quattrocento</i> . Edizione e commento dei testi a cura di Nicoletta Della Penna, Firenze, Cesati, 2025 [recensione di Francesco Montuori]	335
--	-----

Studi dal laboratorio del DESN

Lucia Buccheri, Marialuce Balsamo, Carmine Caruso, Vincenzo De Rosa, Giorgia Cinzia Di Matteo, Stefano Di Nolfi, Erica Iannaccone, Simone Padalino, Rebecca Patrevita, Gelsomina Razza, <i>La bottega delle parole: l’esperienza del laboratorio di formazione per la redazione di voci del DESN</i>	345
--	-----

Nicola De Blasi, <i>L'Intelligenza Artificiale, citazioni inventate e veri errori: Matilde Serao non ha mai citato la coviglia nelle sue opere. Premessa per una voce del DESN</i>	385
Vincenzo De Rosa, <i>Verso un Repertorio zoonimico napoletano. Prime rilevazioni bibliografiche, con in appendice sette zoonimi per il DESN</i>	401
Cristiana Di Bonito e Francesco Montuori, <i>Il lemmario di base del DESN (P-) e la voce punessa</i>	449
Indice delle voci del DESN	
<i>Le ultime voci del DESN</i>	469
Indice delle forme notevoli	470

Premessa

Siamo dunque giunti al fascicolo n. 7 della nostra rivista, che si consolida nello *status* di “scientifica”.

Il lettore si chiederà, forse, che fine ha fatto l’obiettivo fondativo di pubblicare voci del DESN. Esse, in un certo senso, latitano. Nel fascicolo precedente e in questo sono apparse alcune voci anche interessanti, ma in numero inferiore alle attese.

Ma il lettore attento ci avrà certo già assolto da questo peccato: un po’ perché avrà trovato materiale di interesse storico-etimologico anche tra i saggi, come è il caso del napoletano *coviglia*, nella quarta sezione di questo fascicolo, o del calabrese e siciliano *curina*, nel numero 6 della rivista; un po’ perché avrà capito che i redattori del DESN sono stati assorbiti in lavori collaterali che non sono proprio insignificanti per la compilazione del dizionario: la piattaforma digitale, la cui preparazione ha preso più tempo del previsto, è ormai operativa e, piuttosto che compilare nuove voci, stiamo caricando su di essa le voci già redatte, in una struttura leggermente modificata e con qualche ulteriore aggiornamento redazionale (cfr. L. Buccheri e F. Montuori, *Una piattaforma lessicografica per il Dizionario Etimologico e Storico del Napoletano*, in «Quaderni Veneti», 13, 2024, pp. 175-190); l’approntamento

del *corpus* QM ‘Quattrocento meridionale’ (<http://qmweb.ovi.cnr.it/>) ha una nutrita rappresentanza di testi napoletani e campani, che saranno ancora più numerosi dopo l’aggiornamento previsto per l’ottobre prossimo; è in corso di preparazione il lemmario di base (si veda il contributo di Cristiana Di Bonito e Francesco Montuori in questo fascicolo), la redazione delle cui voci costituirà il primo obiettivo del nostro prossimo lavoro.

Per il resto, la rivista è ormai uscita dalla culla e ha assunto una sua fisionomia.

Anche in questo numero pubblichiamo, tra l’altro, una sezione monografica, ospitando quattro contributi che illustrano alcuni aspetti progettuali ed elementi procedurali di un PRIN recentemente concluso. Si tratta di una banca dati interrogabile e parzialmente lemmatizzata, (MED.ITA) che raccoglie un gran numero di testi medio-latini veneti, toscani e meridionali. L’interesse lessicografico è molto grande, ovviamente, ancor più per la Campania, che manca di una raccolta sistematica e critica di tali materiali. I progressi degli ultimi anni, vivaci presso varie “scuole”, hanno visto un significativo compimento nel lavoro di ideazione e coordinamento svolto da Paolo Greco: tra pochissimi giorni sul portale dell’OVI saranno disponibili due *corpora* con le modalità di ricerca cui la piattaforma informatica Gatto-web ci ha abituato: <http://medita.ovi.cnr.it/> e <http://meditalemm.ovi.cnr.it/>. In autunno sarà pronto anche il corpus MED.ITA-GLOSS, che raccoglie i glossari latino-volgare antichi e moderni già editi.

Si rilevano altre costanti, in questo fascicolo: porgiamo particolare attenzione al lessico e ai testi della gastronomia, il cui rilievo negli studi storico-linguistici siamo felici di contribuire a consolidare; manifestiamo grande considerazione per le dinamiche proprie dell’intera area campana e dei dialetti centro-meridionali, in sincronia e in diacronia; e infine, speranza e vanto di questa rivista, documentiamo le attività di formazione svolte nelle università campane, che forniscono materiali per il dizionario e, contemporaneamente, consentono la crescita di nuovi giovani ricercatori.

Napoli, 30 giugno 2026

Nicola De Blasi – Francesco Montuori



LA LEMMATIZZAZIONE DEL CORPUS MED.ITA: ORIENTAMENTI E INDICAZIONI OPERATIVE*

Mariafrancesca Giuliani

0 Premessa

Presento di seguito il *Manuale per la lemmatizzazione del corpus MED.ITA*.¹

Il corpus è stato creato in ambiente GATTO (*Gestione degli Archivi Testuali del Tesoro delle Origini*, cfr. <http://www.oivi.cnr.it/Il-Software.html>), noto software sviluppato negli anni '90 presso l'Opera del Vocabolario Italiano, e utilizza, dunque, le procedure di lemmatizzazione previste dal software nella versione 3.3, ampiamente adoperate e ben testate con riferimento ai testi volgari contenuti nel Corpus TLIO, base documentaria primaria del *Tesoro della Lingua Italiana delle Origini*.²

* Desidero ringraziare Elisa D'Argenio e Jacopo Gesiot per l'accurata lettura e i suggerimenti che mi hanno aiutato a migliorare il testo.

¹ Per una descrizione del progetto MEDITA, che prevede la costruzione del corpus MED.ITA, a cui si associa il modulo MED.ITA-LEMM, e della piattaforma lessicografica MED.ITA-GLOSS, rimando a Greco–Cotugno–Giuliani 2024 e *idem* i.c.s.; vd. anche l'introduzione nell'articolo di Paolo Greco pubblicato nello stesso numero di questa rivista.

² Per una descrizione dell'evoluzione e delle funzionalità di GATTO, cfr. Boccellari–Iorio Fili 2013.

Tale corpus è, come è noto, una raccolta di testi prodotti nelle più antiche varietà medievali di area italo-romanza, disponibili in edizioni affidabili o comunque utilizzabili ai fini della redazione del vocabolario, e la lemmatizzazione è sorta ed è stata sviluppata con l'idea di ricondurre esiti linguistici e grafici diversi a unità invarianti (lemmi) rappresentative del sistema lessicale dell'italiano antico, grosso modo corrispondenti alle entrate definite per il TLIO.³

La lemmatizzazione risponde, dunque, all'esigenza di ordine teorico e pratico di rendere interrogabile un corpus linguisticamente eterogeneo utilizzando come chiave della ricerca il lessico italiano delle origini, un sistema forse astratto o, in un'ottica prospettica, teleologico, ma valorizzato nella sua consistenza da molteplici studi che confermano la tendenziale convergenza delle antiche varietà italo-romanze, pur in assenza di una norma sovralocale esplicita condivisa.⁴

Potremmo asserire che il latino medievale, con cui le antiche varietà italo-romanze hanno lungamente convissuto, offra una casistica diversa.

La norma grammaticale ricevuta dall'epoca classica persiste anche nelle fasi successive, pur risentendo di modifiche e aggiornamenti continui, ma, nel rispecchiare la varietà degli scenari storico-sociali propri del mondo medievale, il lessico si modifica in maniera significativa dando spazio a prestiti e neologismi che in alcuni casi si diffondono anche a larga distanza rispetto al versante originario di manifestazione.⁵

La lemmatizzazione condotta sul corpus MED.ITA consente di identificare la totalità delle parole rappresentate dalle forme sottoposte alla schedatura lessicale e morfologica. Tale operazione è stata avviata con l'ottica di valorizzare la stratificazione interna al vocabolario testuale, testimone della continuità di un patrimonio lessicale ben noto alla tradizione lessicografica

³ Cfr. Artale 2019. Per una descrizione del lemmario del TLIO, cfr. Mosti 2015.

⁴ Il dibattito sul tema è esaminato da Giuliani 2023, pp. 9-13.

⁵ Per un ampio quadro della complessità e delle contraddizioni proprie del latino medievale, cfr. Stotz 2013.

latina e mediolatina, che accoglie, al contempo, unità nuove, talora mai censite, che rappresentano non di rado testimonianze d'interesse anche per la lessicografia italomanzana.

Il manuale qui pubblicato segue indubbiamente la traccia delle linee guida dell'esperienza di lemmatizzazione sviluppata da diversi decenni all'OVI, adattandosi peraltro alle caratteristiche dell'ambiente di gestione proprio del software GATTO.⁶ Nuovo e diverso è l'oggetto linguistico, rappresentato dal latino medievale: difatti la stesura del testo si è sviluppata per fasi, tenendo conto di riflessioni di natura teorica e operativa al tempo stesso, emerse nel confronto concreto con il nucleo testuale su cui è stata avviata la lemmatizzazione integrale, affiancata da una lemmatizzazione selettiva (di specifico interesse lessicologico) realizzata su un corpus più ampio, sebbene parziale rispetto alla documentazione raccolta nel corpus MED.ITA. Entrambi i gruppi testuali sottoposti a una lemmatizzazione integrale e selettiva (quest'ultima integrata da una iperlemmatizzazione, cfr. § 1.1) sono inclusi nel modulo MED.ITA-LEMM, concepito come base di avvio di una lemmatizzazione più estesa.⁷

La latinità, anche tarda e medievale, è stata nell'ultimo decennio al centro dell'interesse della ricerca computazionale, volta a sviluppare sistemi di annotazione utili per la gestione digitale integrata del repertorio delle numerose banche-dati testuali ad oggi disponibili per questa lingua. Si è tenuto conto anche degli orientamenti emersi in quest'ambito (cfr. § 1), di grande interesse, anche se solo parzialmente attento alla prospettiva dell'utilizzazione lessicografica dei materiali.

⁶ Le basi storiche, teoriche e operative della lemmatizzazione del Corpus TLIO sono descritte nella *Grammatichetta*. Le linee-guida sono state dettagliate, di seguito, in documenti di lavoro ad uso della redazione prodotti da Elena Artale e Diego Dotto: si veda rispettivamente *Note di lemmatizzazione* e *Libro bianco*. Orientamenti specifici definiti in questi testi saranno richiamati in più punti del manuale.

⁷ Una prima descrizione del modulo MED.ITA-LEMM, e del senso e degli orientamenti della lemmatizzazione avviata sui testi, compare in Giuliani–Aprea i.c.s.

Lasciando ogni approfondimento al manuale (§§ 1-3), mi limito qui a mettere a fuoco alcuni aspetti della lingua del corpus che hanno imposto l'individuazione di percorsi di lemmatizzazione capaci di garantire una schedatura coerente ma duttile dei materiali, utilizzabile anche nella prospettiva dello sviluppo di una lemmatizzazione semiautomatica, auspicabilmente supportata dall'utilizzo dell'intelligenza artificiale.

0.1 Una lingua “a norma”, ma innovativa

La lemmatizzazione integrale condotta sul campione testuale selezionato in MED.ITA-LEMM coinvolge le parole lessicali e grammaticali e i nomi propri (antroponimi e toponimi).

Come anticipato, le forme appartenenti ad entrambi gli insiemi sono state associate a parole invarianti (lemmi), adottando come riferimento normalizzante – in presenza di una corrispondenza esplicita – il repertorio lessicale e onomastico censito nei più importanti lessici della latinità e della media latinità (cfr. §§ 1, 1.3 e 3).

Il trattamento di antroponimi e toponimi, legati in misura cospicua a modelli non latini e non di rado costituiti da designazioni polirematiche (caso frequentissimo per i microtoponimi che spesso hanno una *facies* schiettamente volgareggiante) ha richiesto strategie di normalizzazione diverse (cfr. § 2.6.7.1).

Il “principio evolutivo” descritto da Korkiakangas (2020, pp. 74-76) con riferimento alle *Late Latin Charter Treebanks* (LLCT), sintetizza, invece, l'ottica che ha guidato in maniera prevalente la lemmatizzazione delle parole lessicali e grammaticali: «The evolutionary principle reduces the linguistic variants provoked by language evolution to their standard Latin ancestors» (ib., p. 74). Questo tipo di classificazione è pienamente funzionale in una annotazione orientata a favorire ricerche sul livello lessicale, e agevola il trattamento unitario di ogni variante in rapporto genetico con parole latine o mediolatine censite dai lessici:⁸ la forma participiale *contjata*, attestata in

⁸ L'idea del confronto con la norma classica è alla base anche della metodologia adottata da József Herman e successivamente da Béla Adamik per la catalogazione di dati innovativi

ChLA² LIII, 4, 823 (luc.;⁹ p. 36.25) è ricondotta al lemma *conciare* v. al pari dell'infinito *conciare* attestato in CDAmal I, 4, 939 (camp.; p. 7.21), così come la forma *signum*, più volte attestata nel corpus da noi lemmatizzato, è associata al lemma *signum -i* al pari di *sinnus* e *sinnu*, forme attestate in documenti di area marchigiana che portano la traccia di un'evoluzione fonetica divenuta caratteristica del dialetto locale (cfr. Rohlf 1966, § 259). D'altronde le forme note sulla base della registrazione lessicografica orientano non di rado l'interrogazione di un repertorio testuale: così, ad esempio, cercando il lemma *chrysobulium -ii* (< gr. χρυσόβουλλον 'documento, bolla con sigillo d'oro') censito da Du Cange e MLW, l'utente potrà individuare la forma genitivale *grossovoli*, foneticamente evoluta rispetto alla sua base latina.¹⁰

La procedura di lemmatizzazione descritta si basa evidentemente sulla ricerca di corrispondenze tra il latino lessicografico e il latino dei testi, focalizzate sulla componente lessicale, si è dato rilievo, tuttavia, ad alcune evoluzioni morfologiche significative e caratteristiche dell'uso documentario, tralasciando in tal caso la procedura di "livellamento" implicata dal principio evolutivo.

Il lemma *rebus* s.f. affianca il lemma *res rei* s.f., rappresentando i casi in cui il tipo dativo-ablativo plurale è utilizzato visibilmente come forma irrigidita per tutte le funzioni grammaticali. In maniera analoga il lemma *parte -tis* affianca il lemma *pars -tis* rappresentando l'estensione della forma

(nella loro condizione differenziale) riconosciuti nel latino delle iscrizioni di età imperiale. Così Adamik (2021, p. 3) descrive il contenuto del database creato: «its entries consist of deviations of all sorts from "Classical Latin" or what is called the "classical norm", irrespective of whether they are due to linguistic or extra-linguistic (i.e. graphic or technical, etc.) factors».

⁹ Indico di seguito lo scioglimento delle abbreviazioni relative all'area generica di ogni testo citato nell'esemplificazione di questo lavoro: abr.: abruzzese, camp.: campano, luc.: lucano, pugl.: pugliese, tosc.: toscano, umbr.: umbro, ven.: veneto. Per indicazioni sul campo di localizzazione "area generica" (la cui etichetta figura nel titolo abbreviato di ogni testo) rimando all'articolo di Paolo Greco nello stesso numero di questa rivista.

¹⁰ Cortelazzo (1970) cita questa stessa forma s.v. *crusovolo* s.m. 'bolla aurea': con quest'ultima grafia il prestito si attesta in documenti veneziani trecenteschi.

piena anche alla funzione Soggetto, anch'essa contemplata da alcune antiche scritture legali, che forse anticipano esiti romanzi (cfr. § 2.1).

Nuove forme plasmate per la funzione Soggetto,¹¹ così come nuove forme metaplastiche, che attraversano le reti della morfologia nominale e verbale, popolano il lemmario di MED.ITA, rappresentando la tensione innovativa delle fluttuazioni accettate entro alcuni ambienti e livelli di scrittura come microregolarità.¹² Queste ultime distinguono in maniera consistente il latino del corpus MED.ITA dal latino contemplato dalla norma classica.

Un ulteriore aspetto evolutivo notevole coinvolge la semantica di molte parole del corpus, la lemmatizzazione, tuttavia, non classifica questo dato – che è l'oggetto d'interesse della descrizione lessicografica – se non coinvolge l'emergere di nuove associazioni tra forma e senso, che si distanziano da parole di tradizione precedente, come nel caso di *dominus* e *domnus* menzionati in § 2.1.

0.2 Una lingua incline alla ibridazione

L'interferenza domina inevitabilmente la vita del latino medievale nelle sue dinamiche di adattamento agli usi linguistici di una vasta diatopia e di ambienti diversi anche per livello sociale.¹³

¹¹ È doveroso precisare che le opposizioni di caso e funzione proprie del latino classico sono pienamente osservate solo nei testi pratici che evidenziano uno stile alto e pienamente sorvegliato, mentre nei documenti notarili dei secoli IX-XI, che rappresentano una parte fondamentale del modulo MED.ITA-LEMM, e che hanno avuto un ruolo centrale nella costituzione del vocabolario di macchina, la morfologia flessiva appare non di rado scissa dalle antiche celle paradigmatiche (cfr. sul tema Sornicola 2017, pp. 86sgg.). La creazione di un lemma come "*parte -is s.f.*" ha, pertanto, un valore puramente convenzionale e risponde a esigenze di omogeneità rispetto ai criteri di citazione fissati per i sostantivi (cfr. § 2.1).

¹² Si tratta di orientamenti linguistici ricorrenti e dunque presumibilmente accettati entro determinati circuiti di scriventi. Il concetto sociolinguistico di microregolarità è stato utilmente applicato da Rosanna Sornicola allo studio di alcune fenomenologie variazionali emergenti nelle più antiche scritture notarili di area campana, vd. ad es. Sornicola (2017).

¹³ Ha una *facies* ibrida anche il francese medievale d'Italia che è oggetto di raccolta e analisi nel quadro del progetto FrInGE (*The French in/of Italy: Code-MixiNG in Medieval Europe*).

Il principio adottato nella lemmatizzazione dei prestiti dalle varietà romanze in contatto, e dei tecnicismi giuridico-istituzionali e politici importati in maniera consistente dalle varietà germaniche (seppure in gran parte mediati da un tramite italoromanzo), ha seguito le linee guida definite dal *Mittelateinisches Wörterbuch* (MLW) nella strutturazione del proprio lemario. Queste sono chiarite nell'introduzione al primo volume dell'opera (cfr. § 1.3.1), ma il gruppo di ricerca di MEDITA ha potuto far riferimento a un testo descrittivo sul tema, dettagliato da numerosi esempi desunti dal repertorio della più importante fucina lessicografica mediolatina dei secoli XX e XXI, approntato da Monica Niederer, redattrice esperta del MLW recentemente venuta a mancare. Anche tenendo conto dell'infelice congiuntura che ha toccato la sua vita ritengo importante dare valore al ruolo-guida del suo contributo.

Fondamentalmente la lemmatizzazione dei prestiti adotta come principale riferimento la lessicografia storica dell'italiano, lasciando integro, però, l'assetto latinizzato delle forme restituite dai testi.¹⁴

1 Manuale per la lemmatizzazione del corpus MED.ITA

La lemmatizzazione del corpus MED.ITA – attualmente realizzata in maniera integrale solo su un ristretto numero di testi documentari (originali preferibilmente pubblicati in edizione paleografica) raccolti nel modulo MED.ITA-LEMM – è orientata a ricondurre il formario del corpus a un lemario latino appartenente a due stadi cronologici: quello del latino classico e tardo, rappresentato dal *Thesaurus Linguae latinae* (ThLL) e strumenti lessicografici analoghi (Forcellini, Lewis Short, Georges, ecc.), e quello del latino medievale,

Si rimanda a Gambino et al. (2024, in particolare a p. 293sgg.) per un confronto sulle strategie utilizzate nell'annotazione del corpus (DiFrl), con particolare riferimento ai fenomeni d'ibridazione.

¹⁴ Una strategia diversa è seguita da Larson 1996 che, nella sua raccolta di volgarismi registrati in documenti mediolatini di area toscana anteriori al 1200, assume a lemma le forme corrispondenti nell'italiano.

rappresentato dal monumentale Du Cange e da strumenti lessicografici di più recente creazione come il MLW e il Niermeyer.

Il lemmario associato alle forme di MED.ITA-LEMM segue la logica della lemmatizzazione manuale sperimentata all'OVI con riferimento al Corpus TLIO. Si adotta la classificazione delle parti del discorso usata per questo corpus (in funzione della redazione del TLIO) adeguandola però alle specificità di un corpus latino. Si intende operare, tuttavia, anche tenendo conto della prospettiva degli *UPOS tags* (*Universal Parts of Speech*) di *Universal Dependencies* (UD) usati nella comunità scientifica che opera nel quadro dei *Linking Open Data* (LOD):¹⁵ con riferimento alla gestione di database di testi latini rimando in particolare alle linee scientifiche e operative definite per la base di conoscenza testuale *Linking Latin* (LiLa, <https://lila-erc.eu/#page-top>).¹⁶

Lo scopo principale della lemmatizzazione condotta in MED.ITA-LEMM è creare un dizionario di macchina che possa essere implementato ed esteso progressivamente, anche in forma semi-automatica all'intero formario del corpus. Per agevolare tale prospettiva di lavoro la lemmatizzazione opera a maglie larghe e secondo un principio di economia: si cerca di evitare, infatti, ogni tipo di duplicazione lemmatica. Questa potrebbe scaturire, ad esempio, dall'associazione della stessa forma di citazione (rappresentativa della forma del lemma) a categorie grammaticali diverse, o da una schedatura volta a integrare nella scheda di lemmatizzazione informazioni legate alla singolarità dell'occorrenza lemmatizzata.

La lemmatizzazione è adeguata agli standard formali e categoriali dei principali dizionari latini di riferimento per la fase classica e medievale. Si valorizzano, d'altronde, le nuove introduzioni lessicali e le innovazioni formali e categoriali riscontrabili nella lingua del corpus utilizzando le etichette illustrate nei paragrafi seguenti.

Lo scopo del lavoro è rendere conto del modo in cui i materiali testuali raccolti si rapportano con una tradizione secolare ampiamente analizzata,

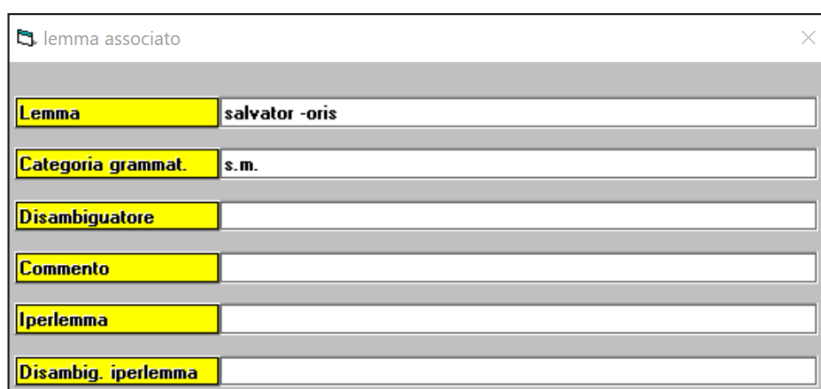
¹⁵ Cfr. <https://universaldependencies.org/u/pos/>.

¹⁶ Per un'introduzione alla collezione lessicale gestita da LiLa e per una descrizione dell'architettura informatica che la sostiene cfr. Passarotti *et al.* 2020.

mettendo in luce, tuttavia, le specificità e le novità che appartengono al corpus, anche nella prospettiva di un utilizzo del corpus MED.ITA nel quadro della lessicografia italoromanza.

1.1 La lemmatizzazione delle parti del discorso: lemmi e categorie grammaticali

La scheda di lemmatizzazione operativa in ambiente GATTO, prevede cinque o sei campi a seconda della versione del software, si veda l'esempio del lemma *salvator -oris* s.m., secondo la scheda di GATTO 3.3:



Lemma	salvator -oris
Categoria grammat.	s.m.
Disambiguatore	
Commento	
Iperlemma	
Disambig. iperlemma	

fig. 1

Il campo “lemma” accoglie la forma di citazione (lemma) selezionata per rappresentare una o più forme del corpus e corrisponde, nella maggior parte dei casi, all’entrata lessicale riconosciuta dai dizionari di riferimento. Il lemma incorpora anche informazioni morfologiche minime, secondo un formato che sarà dettagliato per le singole parti del discorso. In assenza di riferimenti nelle risorse lessicografiche a disposizione, la forma del lemma sarà opportunamente determinata dal responsabile della lemmatizzazione, tenendo conto delle regole di normalizzazione stabilite per la grafia (vd. § 1.4) e ricostruendo la forma di citazione opportuna per la funzione grammaticale o onomastica svolta dalla forma-parola analizzata (vd. § 2).

Segue il campo “categoria grammaticale” che accoglie l’etichetta che definisce la parte del discorso associata al lemma. La lista delle categorie grammaticali ammesse, coincidente in grossa misura con quelle utilizzate per la lemmatizzazione del Corpus TLIO è riportata nella tabella seguente, che

esplicita le corrispondenze con i tag corrispondenti a *P(arts) O(f) S(peech)* elaborati nel sistema di *Universal Dependencies* e utilizzati da LiLa:

Categoria grammaticale TLIO	UD / LiLa POSTags
a.g. = ambiguità grammaticale	
a.t. = ambiguità testuale	
g.a. = grafia alternativa	
volg. = volgare, volgarismo	
agg. = aggettivo	adjective
avv. = avverbio	adverb
cong. = congiunzione	coordinating conjunction / subordinating conjunction
dim. = aggettivo / pronome dimostrativo	determiner / pronoun
escl. = esclamazione	interjection
indef. = aggettivo / pronome indefinito	determiner / pronoun
interr. = aggettivo / pronome interrogativo	determiner / pronoun
n.p. = nome proprio	proper name
num. = numerale	numeral
pers. = pronome personale	pronoun
poss. = aggettivo / pronome possessivo	determiner / pronoun
prep. = preposizione	adposition
rel. = pronome relativo	pronoun
s.f. = sostantivo femminile	common noun
s.i. = sostantivo di genere incerto	common noun
s.m. = sostantivo maschile	common noun
s.n. = sostantivo neutro	common noun
v. = verbo	verb

Se per un lemma i lessici di riferimento contemplano più di una funzione grammaticale, alla categoria grammaticale primaria (quella descritta come principale) potrà essere associato il simbolo “+”, si vedano gli esempi seguenti relativi a *intro*, avverbio e preposizione, e a *quicumque*, pronome indefinito usato anche con funzione di indefinito relativo:¹⁷

1	ChLA ² LIII, 1, 809 (... RO.. 18.8	intro avv.+ X
loquo pro merce anima meam, simul et tertija parte de casa quem abere bisus sum intro civitatem Copersano in s(uprascripto) monasterio offeruit; excepto tertija partem de familia mea, masculi et femine,		
2	ChLA ² LIII, 1, 809 (... RO.. 18.11	intro avv.+ X
portjonis de omnem substantija mea et de familia, masculi et femine, et case quantecumque abeo intro hanc Tarantinam civitatem, quem demisi ad Frisso f(i)l(i)o meo quam etiam et ad alios f(i)l(i)is		
3	SMTrastevere 2, 1026... RO.. 293.2.17	intro avv.+ X
beati Prankatii martiris in locum qui appellatur Pantano Maio(re), v(e)l quo alio vocabulo nuncupatur, et intr' affines a primo la(tere) via karraria p(u)bl(i)ca, et a secundo la(tere) nvumqu(e) Galeria, et a		
4	PergCap 1, 1091 (cam... RO.. 70.12	intro avv.+ X
predictis omnibus terris de s(uprascript)is locis, que et qualiter superius dixi, una cum omnia ibi intro abentibus subter et super et cum uie sue ibidem intrandi et exiendi. Ego, q(ui) s(upra),		

fig. 2

1	ChLA ² LIII, 4, 823 (1... RO.. 34.4	quicumque indef.+
Chr(ist)i e(t) salbatori mei misericordia, cepit iudicares de res meas ut, si iustu(m) Dei iudiciu(m) quacunq ue mo(r)te occupatus fuero, non inpareato ante eternu iudiciu(m) ductus fiat. Primis omnium offero adq(ue		
2	ChVer 255, 955 (ven...) RO.. 394.16	quicumque indef.+
proprietatis mee que ego habere et possidere visus sum hic in regno Italico per ceteris quibuscunq ue locis et notabiliter ab ipsis iam prenomatis castris et rebus. Id est in primis castris		
3	AntiCaMod 9, 1144 (c... RO.. 27.7	quicumque indef.+
et possideant securiter inde faciendo omnia quecumque voluerint, sine nostra nostrorumque heredum contradictione et sine cuiuscunq ue requisitione et per nostram defensionem omni tempore ab omnibus		

fig. 3

Nel caso di aggettivi usati anche come sostantivi o di participi presenti o passati usati anche come aggettivi si ricorrerà alla marca grammaticale dedicata.

Gli ulteriori campi presenti nella scheda sono “disambiguatore”, “iperlemma” e “disambiguatore iperlemma”: tali campi, non compilati nell’esempio precedentemente illustrato (“*salvator*, -oris s.m.”), dovranno essere riempiti solo in casi specifici, secondo le indicazioni fornite di seguito nel manuale (vd. §§ 1.2, 1.2.1, 1.2.2, 2.6.5, 2.6.6, 2.6.7 e 2.6.7.1).

¹⁷ Si mutua, in tal senso, la strategia di lemmatizzazione (funzionale all’automatizzazione del processo) indicata nelle *Note di lemmatizzazione*, soprattutto con riferimento agli avverbi utilizzati anche come connettivi (ad es. *anzi*) e agli avverbi relativi come *dove* e *dovunque*.

Nel quadro del progetto MEDITA il campo “iperlemma” è stato compilato per schedare in maniera solo esemplificativa, su un campione di 28 testi, gli ambiti socioculturali più significativi richiamati dal lessico dei testi raccolti, con un riferimento prevalente a neologismi, volgarismi e tecnicismi. In questo sviluppo del progetto l’iperlemma è rappresentato, dunque, da alcune etichette standard concordate dai ricercatori (ad esempio *diritto, economia, cultura materiale, agricoltura e marineria*) per delineare una prima tassonomia descrittiva.¹⁸

La scheda di lemmatizzazione di GATTO 3.3 include, infine, anche il campo “commento”. Tale campo è stato utilizzato per segnalazioni bibliografiche utili per identificare specifici antroponimi e toponimi (vd. § 2.6.7).

1.2 L’uso del disambiguatore: indicazioni generali con riferimento a tutte le parti del discorso

Il campo definito “disambiguatore” deve essere utilizzato, come già anticipato, quando è necessario introdurre un’informazione che si aggiunge a quanto noto sulla base della schedatura della forma di citazione del lemma e della categoria grammaticale o che commenta aspetti notevoli propri dell’associazione forma-lemma.

1.2.1 L’uso del disambiguatore nella lemmatizzazione di antroponimi e toponimi

Una specifica etichetta informativa è inserita di default da GATTO nel campo “disambiguatore” quando si associa una forma a un “nuovo antroponimo” o a un “nuovo toponimo” utilizzando i tasti appositi disponibili nel menu di lemmatizzazione: in entrambi i casi la categoria grammaticale è “n.p.” (= “nome proprio”).

L’etichetta inserita dal programma è “antr.::” nel caso della lemmatizzazione di un *antroponimo*, “n.g.::” (= *nome geografico*) nel caso della lemmatizzazione di un toponimo. Si vedano le due schede seguenti:

¹⁸ Per dettagli sull’iperlemmatizzazione avviata nel quadro del progetto MEDITA rimando al contributo di Fabio Aprea e Jacopo Gesiot in questo stesso numero della rivista.

Lemma	andreas
Categoria grammat.	n.p.
Disambiguatore	antr.::

fig. 4

Lemma	beneventum
Categoria grammat.	n.p.
Disambiguatore	n.g.::

fig. 5

1.2.2 L'uso del disambiguatore nella lemmatizzazione degli elementi lessicali

Con riferimento al lessico, l'uso del disambiguatore dovrà seguire la casistica e le indicazioni indicate di seguito.

Come già nella pratica di lemmatizzazione avviata sul Corpus TLIO, il campo "disambiguatore" deve essere utilizzato per disambiguare due o più lemmi omografi, si rimanda in proposito al paragrafo *Individuazione e distribuzione dei lemmi* del *Libro bianco*:

«in disambiguatore viene inserito il significato o l'etimo di entrambi i lemmi o di uno dei due se esiste una forte disparità nell'attestazione dei due lemmi (*disambiguatore semantico*). In genere è preferibile la prima soluzione alla seconda.»

Si vedano le coppie di lemmi del Corpus TLIO "*mondo s.m. (universo)*" e "*mondo s.m. (puro)*" o ancora "*macchia s.f. (boscaglia)*" distinta da "*macchia s.f.*". Un ottimo esempio per il latino è rappresentato dalla coppia "*venire v.*" (prima persona: *věňō*), noto verbo di movimento, e *venire v.* 'essere venduto' (prima persona: *věňō*), seppure l'omografia riguarda solo alcune forme del paradigma verbale.

Il campo può essere utilizzato, inoltre, per raccogliere altre segnalazioni di rilievo correlate alle associazioni forma-lemma: 1) usi grammaticali specifici; 2) la combinazione con componenti enclitici, 3) l'appartenenza della forma a una stringa lessicale più estesa, ovvero la scrizione analitica (e in questo senso polirematica) di una specifica parola (rappresentata da un lemma) o,

all'inverso, la probabile scriizione unverbata, talora d'interpretazione dubbia (per un errore dello *scriptor*), di una sequenza di parole.

Tutti questi casi sono già contemplati dalla lemmatizzazione applicata al Corpus TLIO,¹⁹ ne seguiremo, dunque, gli orientamenti pur semplificando la forma e la sintassi delle annotazioni.

Il caso 1) contempla l'uso di *habere* ed *esse* come ausiliari (nei tempi composti e nel perfetto passivo): il disambiguatore accoglierà l'annotazione "ausiliare":

Lemma	esse
Categoria grammat.	v.
Disambiguatore	ausiliare

fig. 6

Il caso 2) contempla l'integrazione grafica, nella stringa che realizza l'occorrenza, di una particella enclitica: vd. ad es. la forma *ideoque* nell'esempio riportato di seguito:

«*Ideo*que ego Alip(er)tu f(i)l(ius) q(uon)d(am) Alip(er)ti q[ui] fuit sc[u]l[d]ais...»
(ChLA² LIII, 1, 809 (pugl.), p. 18.2)

La forma è lemmatizzata "*ideo avv.*" e nel disambiguatore è segnalata la presenza della congiunzione enclitica ("^encl.: *que* cong."). L'etichetta standardizzata in questi casi sarà introdotta da "^encl.", di seguito, dopo i due punti (:), sarà indicata la particella enclitica, nell'esempio "*que* cong."

¹⁹ Cfr. *Note di lemmatizzazione* nei paragrafi *Forme con clitici*, *Come trattare le polirematiche e Casi particolari (1) Verbi* e *Libro bianco* nei paragrafi *Trattamento delle parole composte: il disambiguatore] polir.* e *Il disambiguatore] ausiliare per avere e essere.*

La standardizzazione della segnalazione nel disambiguatore agevolerà la ricerca delle forme enclitiche indicate in questo campo:

“[^]encl.: xxx c.g.” (xxx: forma di citazione dell’enclitica, c.g.: categoria grammaticale registrata nei lessici per l’enclitica).

Si lemmatizzano in questa maniera anche le forme pronominali *mecum*, *tecum* e *secum*: ad es., forma *secum* → lemma: *se*, c.g.: *pers.*, disambiguatore: “[^]encl.: *cum* prep.”.

Il caso 3) contempla, in primo luogo, la presenza di scritture non univerbate per forme lessicalizzate, come gli avverbi *inantea* e *desuper* (quest’ultimo anche preposizione) o la congiunzione *idest*, attestati – anche nella documentazione raccolta dai lessici storici – nelle forme graficamente scisse (per scelte specifiche degli editori) *in antea*, *de super*, *id est*. Per questa casistica si avrà cura di lemmatizzare separatamente ogni componente secondo la specifica funzione grammaticale (ad es. “*in* prep.” e “*antea* avv.”), ma il disambiguatore dei singoli componenti conterrà l’informazione standardizzata “polir.”, utile a chiarire che i singoli morfemi schedati compongono anche unità lessicali realizzate secondo una scrittura “polirematica”. Qualora la scrittura scissa isoli componenti non significativi e non funzionali, la categoria grammaticale per ciascun componente sarà g.a. = “grafia alternativa” e il disambiguatore ospiterà l’etichetta “polir.” seguita, dopo i due punti, dall’indicazione del lemma trascritto in forma scissa. Così ad es. la grafia *at que* per *atque* sarà classificata nei due lemmi “*at* g.a. (disambiguatore: “polir.: *atque*)” e “*que* g.a. (disambiguatore: “polir.: *atque*)”: cfr. ulteriori dettagli in § 2.6.6.

La probabile scrittura univerbata e dubbia di una sequenza di parole è descritta nel paragrafo dedicato alle “ambiguità testuali” (c.g.: a.t.), rimando, pertanto, a § 2.6.5.

Per le necessità proposte dal corpus MED.ITA, il campo “disambiguatore” potrà essere utilizzato anche come contenitore di etichette standardizzate volte a segnalare innovazioni lessicali, formali e categoriali proprie della lingua ospitata dal corpus, censita o meno dai lessici mediolatini.

Si presenta di seguito la lista delle annotazioni utilizzabili:

- **“mlat.”** (= mediolatino): l’annotazione dovrà essere utilizzata nei casi in cui il lemma sia attestato esclusivamente in lessici mediolatini.
- **“innov.”** (= innovazione): l’annotazione, integrata da un’opportuna aggettivazione (vd. i dettagli di seguito), sarà inserita di seguito all’etichetta “mlat.” (“mlat., innov. ...”) segnalando innovazioni non registrate o non sistematizzate dai lessici.

Preciso di seguito i principali tipi di *innovazione* che dovranno essere segnalati:

- **“innov. lessicale”**: l’annotazione dovrà essere utilizzata per i lemmi totalmente privi di riscontro nella tradizione lessicografica latina e mediolatina (vd. ad es. *impariatus* agg. [“impunito” < lat. *pariare*], lemma associato alla forma *impareato*). Formula da inserire nel disambiguatore: “mlat., innov. lessicale”.
- **“innov. categoriale”**: l’annotazione dovrà essere utilizzata nei casi in cui il lemma, pur attestato nel latino classico, sia attestato nel corpus con un assetto morfo-sintattico non atteso e, talora, non esplicitamente registrato (con un’entrata distinta) dai lessici mediolatini, per un cambio di genere, di funzione grammaticale (ad es. un sostantivo derivante da un aggettivo) o di paradigma flessivo; cfr. il caso di *merces -is* (in luogo di *merces -edis*): il cambio paradigmatico è suggerito dalla duplice occorrenza, nello stesso testo (ChLA² LIII, 1, 809 (pugl.), p. 18.7 e 9), di una forma *merce* (di caso ablativo) dipendente dalla preposizione *pro*.²⁰ Formula da inserire nel disambiguatore: “mlat., innov. categoriale”.
- **“innov. morfologica”**: l’annotazione dovrà essere utilizzata nei casi in cui si renda opportuno mettere a lemma la forma tipica dei casi obliqui, perché attestata anche per la funzione Soggetto

²⁰ L’accusativo *mercem* è attestato già nel latino classico, cfr. Lewis Short s.v. *merces*, *-edis*.

o perché divenuta base di un adattamento che ne lessicalizza l'assetto per tutte le funzioni: vd. i casi di *potestatem* in luogo di *potestas* e di *rebus* in luogo di *res* illustrato in § 2.1, vd., inoltre, il sost. *argilis -is* "terrapieno" («et est *argele* ... de nostro fundamento» SGiovEvTorc, 1065-66 (ven.), p. 8.21), evoluzione del lat. *agger / arger, -eris* (con dissimilazione *-r-r- > -r-l-*), che ha particolare continuità in area veneta (cfr. LEI s.v. *arger*). Formula da inserire nel disambiguatore: "mlat., innov. morfologica".

- **"prev. xxx"** (= prevalentemente c.g.): l'annotazione potrà essere utilizzata per quei componenti lessicali registrati dai lessici con più di una funzione grammaticale, ma attestati nel corpus soprattutto con una funzione: è il caso, ad esempio, di *post avv.+*, *contra avv.+* e *ante avv.+*, frequentissimi nei testi del corpus MED.ITA soprattutto con funzione preposizionale (per cui il disambiguatore reca l'annotazione "prev. prep.").
- **"sic"** (= forma di citazione del lemma adeguata alla forma attestata nel corpus): l'annotazione dovrà essere utilizzata quando la forma lemmatizzata è un *hapax* lessicale (ma non un'ambiguità testuale, per cui cfr. § 2.6.5) privo di riscontro nei lessici e non facilmente riconducibile a una parola nota, una forma per cui non si dispone, dunque, di un modello di normalizzazione valido. L'annotazione dovrà essere utilizzata per gli antroponomi e i toponimi per i quali non si riesce a ricostruire la forma di citazione del nominativo e che rappresentano probabili varianti di un modello non restituito in maniera univoca dai testi e dai repertori di riferimento (cfr. § 2.6.7).
- **"forma invariabile"**: l'annotazione dovrà essere utilizzata per le forme del relativo non accordate nel genere e nel numero con l'antecedente e quindi probabilmente cristallizzate e sovra-estese ben oltre l'uso originario (cfr. § 2.6.2), vd. gli esempi per *quod* nell'immagine seguente:

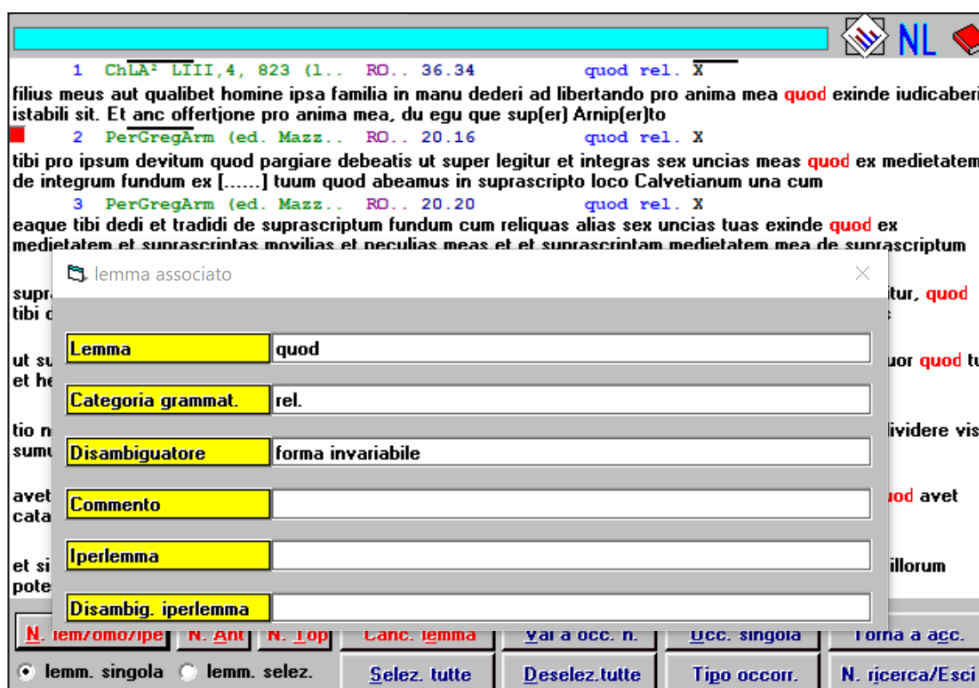


fig. 7

1.3 La lingua della lemmatizzazione: quale latino?

Come anticipato nella premessa di questo manuale (§ 1), lo scopo della lemmatizzazione è rendere conto del modo in cui i materiali testuali raccolti si rapportano alla tradizione secolare del latino, nella sua veste lessicografica classica e medievale.

Tra i due stadi della lingua c'è una continuità dinamica, capace di contenere una variazione che tocca spesso la semantica, ma anche il livello formale.

Ma quale lemma si deve scegliere nei casi in cui la lessicografia dia conto di sviluppi che distanziano il latino medievale dal latino classico?

La forma *pignorandum* di AntiCaMod (9, 1144, (camp.), p. 27.29) trova corrispondenza nel lemma *pignoro* (*pignorare*) del Du Cange e *pignorare* del Niermeyer. Si tratta di una variante della forma classica *pignerare*, già attestata a partire dalla tarda latinità: il ThLL contempla, difatti, il rinvio «**pignor-** v. **pigner-**» e il Gaffiot affianca al lemma *pignērō* la variante *pignōro*.

Per il principio di “livellamento” alla base della strategia di lemmatizzazione avviata in MEDITA (cfr. § 0.1), gli sviluppi prodotti dall’evoluzione linguistica saranno ricondotti in ogni caso agli antecedenti latini e, in tal senso, le forme del latino classico saranno da privilegiare rispetto alle forme del latino tardo e medievale. In questo caso, dunque, il lemma di riferimento sarà “*pignerare v.*” a cui potranno essere associate la forma *pignorandum*, ed anche le forme in *pigner-* documentate in altri testi del corpus MED.ITA. Si osserva in questa maniera il principio di economia già enunciato: un solo lemma potrà essere associato a più varianti e il livellamento raggiunto a livello lemmatico sarà comunque sciolto dalla variazione messa in luce dal formario.

In altri casi si dovrà scegliere tra lemmi diversi proposti dai lessici medio-latini: un buon esempio è discusso di seguito in § 2.1, muovendo dal quesito posto dalla lemmatizzazione della forma *offersionis* attestata in uno dei più antichi documenti presenti nel corpus MED.ITA.

Come si evidenzierà in § 3, in caso di disparità, le soluzioni lessicografiche del MLW e del Niermeyer dovranno essere preferite a quelle del Du Cange, che offre un campionario di forme spesso imprescindibile, ma non evidenzia una strategia ponderata di distinzione tra invarianti e varianti lessicali.

1.3.1 La lemmatizzazione delle forme assenti nei lessici mediolatini

In mancanza di riferimenti lessicografici, i prestiti mai censiti riconosciuti nel tessuto latino dei testi di MED.ITA saranno lemmatizzati secondo le regole definite dal MLW e ulteriormente precisate ed esemplificate da Monica Niederer – che ha avuto un ruolo rilevante nella creazione del lemmario dell’opera – in un documento di lavoro prodotto a vantaggio dei ricercatori di MEDITA (cfr. *Lemmatisierung MLW*).²¹ Si opererà, nel dettaglio, seguendo una strategia di bilanciamento tra l’assetto della parola nella lingua fonte o

²¹ Le norme redazionali del MLW sono dichiarate nell’introduzione al I volume (A-B), cfr. *Vorwort*, pp. V-IX; le indicazioni relative ai criteri d’inclusione e normalizzazione dei prestiti dalle “lingue nazionali” (*Nationalsprachen entlehnten Wörter*) compaiono a p. VII.

tramite, e le attestazioni effettivamente riconosciute nella documentazione analizzata, avendo cura di normalizzare le grafie che rappresentano fonemi non latini secondo l'uso dell'italiano antico rappresentato principalmente dal lemmario del TLIO.

La forma *moglaççum* di *Stat. San Gimignano*, 1255 (tosc., p. 135.25), ad esempio, è stata associata al lemma *mogliazzum -i*, creato sul modello dell'it. ant. *mogliazzo* 'matrimonio, stipulazione di matrimonio' (cfr. TLIO s.v.), così come la forma *boççam* dello stesso testo (p. 98.10) è lemmatizzata *bozza -ae*, con rispondenza nell'it. ant. *bozza* 'vittima di adulterio' (cfr. TLIO s.v. *bozza* (2)).

Si esclude, dunque, la possibilità di ricostruire forme latine non attestate come **muliaceum*. Con la stessa logica, la forma *granaccie* di CDAmal I, 4, 939 (camp., p. 6.21) è associata al lemma *granaccia -ae*, con riscontro nell'it. *granaccia* 'miscuglio di grano e altri cereali' (cfr. TB s.v.), tralasciando l'ipotesi di costruire un lemma **granacea*, mai registrato dalla lessicografia latina e mediolatina.

Come riferimento per questa prassi è utile tenere conto di esempi provenienti dal lemmario del MLW: è emblematico il caso delle neoformazioni in *-aticum/-agium*, spesso documentate nelle fonti considerate dal succitato vocabolario nella forma in *-agium* di evidente conio o tramite romanzo. Il MLW registra sia neoformazioni spesso riferite a tasse e tributi come *ancoraticum*, *biveraticum*, *caespitaticum*, *clusaticum* ed *escaticum*, sia prestiti latinizzati come *ancoragium*, *aquagium*, *beginagium*, *bevrarium*, *confinagium*, *eremitagium*, *fenagium*, ecc. Già solo in questa selezione si evidenziano coppie di lemmi etimologicamente corrispondenti, distanziati solo dalla diversa trafila; talora i due esiti sono allineati nella stessa entrata (come nel caso di *companaticum*, *-i* et *companagium*, *-i*), ma, in generale, le parole attestate solo con il suffisso *-agium* sono lemmatizzate secondo l'esito romanzo senza essere rilatinizzate.²²

²² «Ein wichtiges Kriterium für den Lemma-Ansatz ist, dass wir keine "Geisterwörter" schaffen wollen. Wir wollen keine künstlich latinisierten Lemma-Ansätze erfinden,

I prestiti caratterizzati da una *facies* dialettale saranno associati a lemmi dalla veste grafico-fonetica di tipo toско-italiano se i lessici storici offrono riferimenti per questo modello: per la forma *cavadicia*, ad esempio, attestata in un documento di area veneziana (SGiorMag II, 37, 1078 (ven.), p. 2.109.13), è stato coniato il lemma *cavaticia* -ae s.f., coerente con l'assetto fonetico dei tipi *cavaticcio* e *cavaticcia*, documentati da LEI s.v. *cavare* (vol. 13, col. 497). Questa prassi consentirà di associare al lemma *cavaticia* -ae anche eventuali forme in -t- che potrebbero emergere in fonti di provenienza centro-meridionale.

Diversamente, per la forma *breganega* («fundamentum *Canna Breganega*»), attestata in SGiovEvTorc 47 (1165 (ven.), p. 74.1) il lemma è *breganegus* agg. (disambiguatore: “mlat., innov. lessicale, polir. in n. g.”), coerente con *bregànego* agg. del TLIO (sempre riferito alla canna), perché si tratta di una denominazione botanica tipica dell'area veneta: cfr. LEI s.v. *canna* (vol. 10, col. 1012).

Per la lemmatizzazione dei prestiti italo-romanzi, dunque, i lessici storico-etimologici italo-romanzi costituiscono lo strumento interpretativo e regolativo prioritario. A seconda dell'informazione fornita da questi repertori, il lemma sarà adattato a un modello toско-italiano o confermerà la diatopicità delle forme attestate. Il lessico italo-romanzo ospitato dalla documentazione mediolatina sarà così lemmatizzato assecondando la logica applicata alle forme toscane e non toscane del Corpus TLIO.²³

1.4 La grafia dei lemmi

Diversamente dall'uso dei vocabolari latini, i lemmi coniat per il corpus MED.ITA non riportano indicazioni sulla quantità vocalica.

wenn alle unsere Belege eine romanische Lautung zeigen: wenn zum Beispiel alle Belege eines Worts mit dem Suffixkonglomerat -aticum/-agium die Form -agium zeigen, dann setzen wird das Wort nur mit -agium an.» (Monica Niederer in *Lemmatisierung MLW*).

²³ Cfr. Artale 2019.

Il grafema <i> sarà utilizzato per rappresentare sia la vocale palatale alta /i/ che l'approssimante palatale /j/, quindi non *jejunium* -ii s.n., come in Lewis Short, ma *ieiunium* -ii s.n., come in Georges. Si distinguerà inoltre tra <u> per la vocale velare alta e <v> per l'approssimante labiovelare sonora, secondo l'uso riprodotto dalla maggior parte dei lessici latini. Per le forme germaniche, che talora nei testi sono rappresentate con attacco in <uu->, si userà <w->. Per le parole che si attestano con grafie alternanti tra <w-> e <gu-> si opterà per la forma del lemma che appare maggioritaria nella tradizione lessicografica mediolatina e che talora ha continuità nel lessico italo-romanzo: vd., ad esempio, *quantus* -i s.m. associato alla forma *wantos*, ma *wasonus* -i per la forma *wassonum*, che trova riscontro nelle entrate *waso*, *gua-*, *va-* -sio del Niermeyer e *wasonnus* del Du Cange.

2 Indicazioni per la lemmatizzazione di forme lessicali appartenenti alle diverse parti del discorso

2.1 I sostantivi

I sostantivi si lemmatizzano secondo la forma di citazione con cui compaiono nei lessici latini o mediolatini di riferimento: così, ad esempio, la forma *dei*, genitivo del sostantivo *deus*, già presente nel repertorio latino classico, va associata al lemma "*deus* -i" menzionato dal ThLL; la forma *offersionis* (attestata in ChLA² LIII, 1, 809 (pugl.), p. 20.21), genitivo di *offersio*, variante del sostantivo *offertio* assente nei repertori di latino classico, va lemmatizzata secondo la forma di citazione presente nei più affidabili lessici mediolatini di riferimento: Du Cange registra *offersio* come entrata distinta, ma Niermeyer la accosta come variante al lemma *offertio*. Tenendo conto del fatto che un metodo razionale ed esplicito di lemmatizzazione è stato affinato solo dai vocabolari mediolatini avviati nel sec. XX, si accoglie il lemma proposto dal Niermeyer, ovvero "*offertio* -onis s.f." (disambiguatore: "mlat.").

Sia nel caso di *deus* che nel caso di *offertio*, alla forma di citazione al nominativo si accosta l'uscita del genitivo corrispondente, preceduta dal trattino (*deus* -i; *offertio* -onis).

L'assetto formale delle entrate proposto dalla maggior parte dei lessici (ad es. quelli raccolti nella base di dati del *Corpus Corporum*) sarà riproposto,

scrivendo, ad esempio, per esteso il genitivo dei sostantivi con nominativo monosillabico: *ius iuris* (e non *ius -ris*), *res rei* (e non *res -ei*), *crux crucis* (e non *crux -cis*) e *lex legis* (e non *lex -gis*). Per i lemmi il cui tema termini in *-i*, il genitivo sarà *-ii* (vd. ad es. *auxilium -ii*, *bysantius -ii*, *colludium -ii*, *monasterium -ii*, *pretium -ii*, ecc.).

La stringa indicata per i lemmi relativi a sostantivi potrà richiamare in questa maniera l'unità lessicale e la relativa caratterizzazione morfologica (ovvero la classe di declinazione). La soluzione scelta è volta a dare rilievo ai casi in cui la documentazione raccolta nel corpus metta in luce cambi di declinazione significativi rispetto agli schemi paradigmatici del latino. Naturalmente la ricerca di un lemma di tipo sostantivale dovrà tenere conto del fatto che la forma di citazione prescelta include anche l'indicazione sintetica della forma attestata al genitivo: in fase di ricerca si consiglia, pertanto, di utilizzare preferibilmente una stringa seguita da un asterisco: "*civit**" o "*civitas**" serviranno, ad esempio, per individuare il lemma "*civitas -atis* s.f."

Qualora un testo presenti il genitivo di tipo arcaico *familias* il lemma presenterà il morfema di genitivo in *-as* per tutte le occorrenze della parola presenti nello stesso testo, ma solo per quelle (negli altri casi il lemma sarà *familia -ae*).

I prestiti e i neologismi mediolatini potranno essere lemmatizzati senza l'indicazione della forma del genitivo se compaiono nei testi e nei lessici in forma invariabile per qualsiasi funzione: è il caso, ad es., del longobardismo *sculdais* s.m. 'capo di una circoscrizione' e di *habere* s.n. 'possesso', forma sostantivata del verbo corrispondente.

La scelta della forma del lemma dovrà essere orientata dal principio di economia enunciato nell'introduzione di questo manuale: è opportuno sollevare una variante a lemma solo quando l'uso linguistico (evidenziato dai testi e dall'esame lessicografico) consenta di accertare la lessicalizzazione di significati distinti rispetto alla forma da cui deriva. Si devono certamente lemmatizzare distintamente le occorrenze per *dominus*, *domina* e *domnus*, *domna*, queste ultime varianti sincopate dei lemmi precedenti, già proprie dell'uso latino (cfr. ThLL s.v. *dom(i)nus* e *dom(i)na*): in numerosi contesti del corpus i tipi *domnus* e *domna* sono attestati come titoli onorifici in contiguità

con nomi propri, mentre *dominus* è un appellativo riferito prevalentemente a Gesù Cristo.

Come si è visto nel caso della forma *offersionis*, le scelte operate dalla lessicografia mediolatina più recente fanno da guida in numerosi casi controversi. Nella maggior parte dei casi la forma più vicina alla base etimologica deve essere ragionevolmente accolta come lemma.

La categoria grammaticale da assegnare al lemma sarà naturalmente del tipo “s.f.”, “s.m.”, “s.n.” e inoltre “s.i.” nel caso in cui non sia possibile assegnare al sostantivo una categoria di genere univoca. Alle forme attestate nei contesti seguenti è associato, ad esempio, il lemma “*arbor -is* s.i.”:

«tertjia partem de omnem substantjia meam sibe territorio culto bel inculto, *arboribus* fructiferi bel infructiferis, aquis, silbis...» (ChLA² LIII, 1, 809 (pugl.), p. 18.5);

«omnia quae de superscripta corte nostra pertinent in ipsa villa in integrum transfundimus [...] vineis, campis, cultis et incultis, *arboribus* pomiferis et impomiferis...» (CDVer 18, 890 (ven.), p. 22.31).

Il lat. *arbor* è un sostantivo di genere femminile che, tuttavia, si attesta già in epoca tarda e nell’*Itala* anche come maschile. Negli esempi illustrati il sistema degli accordi non consente di assegnare *arboribus* a un genere specifico. Scelte diverse saranno fatte quando gli accordi segnalano in maniera inequivocabile il genere femminile o maschile del sostantivo.

Un caso particolare è costituito dagli imparisillabi latini della terza declinazione che presentano forme accusativi per la funzione Soggetto: si veda l’esempio di *potestatem* nei contesti proposti di seguito:²⁴

«in s(upra)s(crip)to monasterio remaneat *potestatem*» (ChLA² LIII, 1, 809 (pugl.), p. 18.15)

«de superscriptum fundum [...] in tui tuisque heredibus sint *potestatem* quodque exinde facere volueritis...» (PerGregArm (ed. Mazzoleni), 1, 911 (camp.), p. 20.24).

²⁴ Un’ampia casistica per questa fattispecie è presentata da Sornicola (2012, pp. 59-60).

In questo caso la forma attestata è accolta come lemma (*potestatem -is* s.f.) e il disambiguatore registra la segnalazione “innov. morfologica” volta ad evidenziare la “novità” rappresentata dalla morfologia del lemma. Una soluzione siffatta sarà riproposta solo per ulteriori forme caratteristiche dei casi obliqui reiterate nella funzione Soggetto, tanto da configurare una microregolarità.

Appartiene a questa casistica anche la forma *rebus*, quando “irrigidita” per tutte le funzioni sintattiche e, talora, anche per l’indicazione del numero (cfr. Sornicola 2012, p. 58), tanto da rendere preferibile la creazione di un lemma dedicato, con categoria grammaticale “s.f.”:

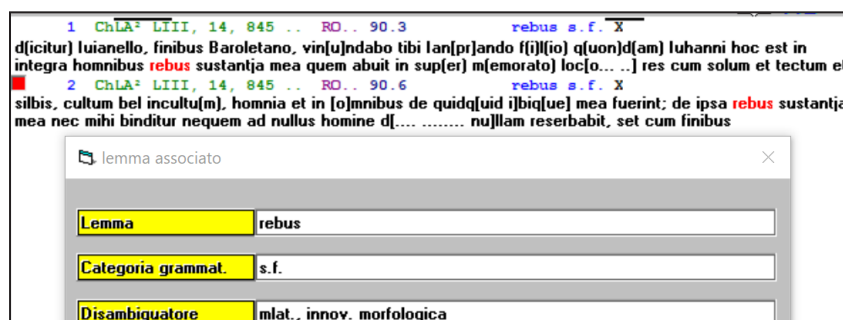


fig. 8

2.1.1 Le denominazioni riferite a popoli

Se un sostantivo definisce al plurale una popolazione, la forma del lemma sarà al plurale e la categoria sarà del tipo “s.m.pl.” o “s.f.pl.”. Tale opzione è coerente con l’entrata scelta dai lessici per molti etnonimi. Si veda l’esempio seguente:



fig. 9

2.2 Gli aggettivi

Gli aggettivi si lemmatizzano secondo la forma di citazione con cui compaiono nei lessici latini o mediolatini di riferimento, ma è sufficiente indicare la sola forma del maschile associando ad essa la categoria grammaticale “agg.”.

Potranno essere marcati come aggettivi anche i participi presenti o passati usati come attributi e apposizioni. I testi volgari offrono una ricca casistica per questo uso, che propone non poche ambiguità nella lemmatizzazione del Corpus TLIO, così come nella redazione lessicografica propria del TLIO, ambiguità generalmente risolte caso per caso.²⁵ I testi mediolatini di carattere pratico sembrano proporre per questa fattispecie una casistica meno ambigua. Saranno classificati certamente come aggettivi i tipi participiali *supradictus* e *suprascriptus* che compaiono di frequente nelle sezioni formulari dei testi documentari e in particolare nelle soprascrizioni o nelle sottoscrizioni:

Quam bero cartul(a)m offersionis te Precopi not(arium) scribere rogabi in *s(uprascripta)* ratjonem. (ChLA² LIII, 1, 809 (pugl.), p. 20.22);

Sig(num) m(anus) *s(upra)s(crip)ti* Roppuli qui an cartula bend(itjonis) ut supra(m) me scrib(ere) rogabet (ChLA² L, 4, 816 (camp.), p. 30.17).

Le forme del comparativo e del superlativo formate aggiungendo al morfema radicale dell’aggettivo i morfemi dedicati dovranno essere associate al lemma dell’aggettivo di riferimento, per cui, ad esempio, la forma *sanctissimus* sarà associata al lemma *sanctus* agg. Diversamente, le forme di comparativo e superlativo che differiscono dall’aggettivo di grado positivo (vd. ad es. *maior* e *maximus* vs *magnus*) saranno lemmatizzate distintamente secondo la forma di citazione al maschile (lemma: “*maior* agg.” e “*maximus* agg.”).

²⁵ Ne dà conto *Libro Bianco* nel paragrafo *Criteri per la distinzione tra aggettivo e verbo nei participi*, che indica «dei criteri diagnostici da adottare o quanto meno una serie di problemi da considerare nella lemmatizzazione dei participi». La *Grammatichetta* aveva fissato già degli orientamenti interpretativi di massima, suggerendo per i casi problematici una prassi sintetizzata nella formula «in dubio pro verbo» (p. 157).

Saranno classificati come aggettivi anche gli etnici, in ogni tipo di occorrenza. Questa classificazione sarà utilizzata anche nei casi in cui gli etnici identificano l'appartenenza di terre, confini, corsi d'acqua e alture a regioni o comuni specifici, cfr. gli esempi seguenti:

«cum ipsa libra de dinarios *veronenses*» (CDVer 255, 955 (ven.), p. 395.9)

«castro meo qui positus est in finibus *Veronense*» (ib., p. 395.5)

forme: *veronenses*, *veronense*, lemma: *veronensis*, c.g.: “agg.”.

Saranno classificati in questa maniera anche gli etnici non censiti dai lessici, ma identificabili sulla base della morfologia derivativa (ad es. *barolitanus* e *prominianensis*): in tali casi il disambiguatore recherà l'annotazione: “mlat., innov. lessicale”.

Le forme sostantivate di un aggettivo, generalmente segnalate anche dai lessici in entrate autonome o in sotto-partizioni di una voce aggettivale, saranno da trattare secondo la prassi definita per i sostantivi (vd. § 2.1), cfr. ad es. *incultum* -i s.n., *integrum* -i s.n., *praepositus* -i s.m. e *praesens* -entis s.n. Si segnalano in particolare i “pluralia tantum” *inferiora* -um e *superiora* -um (già cit. in ThLL s.v. *infer* 7,1, 1395, r. 26 e in Georges s.v. *superus*), ricorrenti nelle formule che descrivono l'assetto di un cespite sottoposto a transazioni: entrambi sono da classificare come sostantivi plurali di genere neutro (c.g. “s.n.pl.”).

2.3 Gli avverbi

Gli avverbi si lemmatizzano secondo la forma di citazione dei lessici latini o mediolatini di riferimento, il campo “categoria grammaticale” recherà l'etichetta “avv.”.

Qualora un avverbio che deriva dalla composizione di due elementi (ad es. *deintus*, *desuper* e *inantea*) sia registrato con una scrizione analitica, le forme relative ai due componenti saranno lemmatizzate secondo la procedura descritta in § 1.2.2 e recheranno nel disambiguatore l'annotazione “polir.” (= parte di una parola polirematica).

Se l'avverbio si attesta nel corpus e nella lessicografia di riferimento anche con funzioni preposizionali e congiunzionali si potrà adottare la categoria

grammaticale polifunzionale “avv.+”, certamente opportuna per gli avverbi relativi *ubi* e *ubicumque*, così come per gli avverbi *contra*, *intro*, *deintro*, *foris*, *foras*, *post*, *adversum* e forme analoghe usate anche con funzione preposizionale (cfr. § 1.1). Si potrà segnalare l’uso prevalente di una specifica funzione utilizzando l’etichetta “prev. xxx” (cfr. § 1.2.2).

2.4 I verbi

I verbi si lemmatizzano all’infinito secondo il modello consueto nella lessicografia delle lingue romanze e già proposto, nell’ambito della lessicografia mediolatina, dal Niermeyer. Qualora l’infinito di un verbo non sia documentato nell’intero corpus MED.ITA, dovrà essere ricostruito sulla base degli schemi morfologici evidenziati dalle altre forme del paradigma. La lessicografia latina e mediolatina offrirà naturalmente i riferimenti necessari, ma si dovrà porre attenzione agli eventuali metaplasmi di coniugazione non registrati da questi strumenti.

Questa soluzione è volta a dare immediato rilievo alla classe di coniugazione rappresentativa per ogni lemma verbale attestato nel corpus.

Qualora nella diatesi passiva un verbo sviluppi un significato nettamente distinto da quello sviluppato all’attivo, le forme al passivo saranno associate all’infinito corrispondente: è rappresentativo il caso di *videri* ‘apparire, sembrare’, trattato come lemma distinto dall’attivo *videre* anche da alcuni lessici latini e mediolatini (cfr. Gaffiot s.v. *videor*; Du Cange e Niermeyer s.v. *videor* e *videri*).

2.5 Le preposizioni e le congiunzioni

Le preposizioni, uniformate nella forma di citazione all’entrata lessicale presente nei lessici latini e mediolatini, adotteranno la categoria grammaticale “prep.” (in LiLa *adposition*).²⁶ Due preposizioni in sequenza come *de ab* saranno lemmatizzate singolarmente, ma acquisiranno l’etichetta “polir.”

²⁶ Per l’etichetta “adposition” è utile fare riferimento al paragrafo a ciò dedicato in *Universal Dependencies* (<https://universaldependencies.org/u/pos/ADP.html>): si applica alle preposizioni e alle posposizioni, anche espresse in forma polirematica, equiparabili per la funzione connettiva svolta nella frase.

(componente di una parola polirematica) nel disambiguatore, se compongo, in scrizione analitica, una parola schedata dalla lessicografia di riferimento, cfr. *deab* in MLW.

Le congiunzioni, uniformate nella forma di citazione all'entrata lessicale presente nei lessici latini e mediolatini, adotteranno la categoria grammaticale "cong."

2.6 Determinanti e pronomi

2.6.1 Aggettivi e pronomi attestati con funzione dimostrativa, indefinita, interrogativa e possessiva

Descrivo in questo paragrafo la strategia da utilizzare per lemmatizzare quei componenti che si associano a un sostantivo definendone le condizioni di relazione con il referente contestuale e testuale. Ci si riferirà dunque alla condizione definita o indefinita, alla quantificazione, alle relazioni di distanza e vicinanza e alla relazione di possesso espresse da componenti che la grammatica classica definisce dimostrativi, indefiniti, interrogativi e possessivi, rappresentati nella frase da aggettivi e pronomi.

Il sistema categoriale accolto da LiLa, coerente con UD, contempla l'etichetta *determiner* per gli articoli e gli aggettivi (usati in alcuni contesti anche con funzione pronominale) che svolgono le funzioni sopra descritte: cfr. ad es. il POS associato ad *alius*, *ille*, *quidam*, *multus*, *meus*. Un indefinito come *nemo*, invece, sostituisce costantemente un nome e quindi in LiLa, secondo gli schemi delle UD, è classificato come *pronoun*.

La lemmatizzazione condotta sul corpus MED.ITA conserverà la descrizione funzionale già adottata per il Corpus TLIO utilizzando le etichette grammaticali "dim.", "indef.", "interr." e "poss." e rinunciando a precisare la funzione aggettivale e pronominale assunta da questi componenti nelle singole occorrenze. Preciso che quantificatori come *multus*, *totus* e *paucus* potranno essere classificati come "indef."

Per la lemmatizzazione dei possessivi, degli indefiniti e degli aggettivi interrogativi, come per l'intera classe degli aggettivi (vd. § 2.2), sarà sufficiente fare riferimento alla forma di citazione del nominativo maschile, ad es. *suus* poss., *aliqui* indef., *aliquis* indef., *quis* indef., *paucus* indef., *quilibet* indef.,

quivis indef., *quispiam* indef., *omnis* indef., *nullus* indef., *quidam* indef., *quisquis* indef., *quicumque* indef., *qui* interr. Lo stesso varrà per i dimostrativi *hic*, *haec*, *hoc* (il lemma sarà “*hic* dim.”) e *idem*, *eadem*, *idem* (il lemma sarà “*idem* dim.”). L’aggettivo e pronome interrogativo e relativo *quantus*, -a, -um sarà lemmatizzato “*quantus* interr.+” (disambiguatore: “prev. rel.”, cfr. § 1.2.2).

Il pronome interrogativo *quis*, *quid* (come i suoi composti *quisnam*, *ecquis* ecc.) sarà lemmatizzato distinguendo il genere (dunque *quis* o *quid* a seconda che la forma del testo indichi il maschile / femminile o il neutro); la categoria grammaticale sarà sempre “interr.”.

Per i dimostrativi *is*, *ea*, *id*; *ille*, *illa*, *illud*; *ipse*, *ipsa*, *ipsum*; *iste*, *ista*, *istud* la gran parte dei dizionari latini indica come semanticamente distintivi gli usi aggettivali e pronominali: nella lemmatizzazione del corpus MED.ITA la forma pronominale sarà classificata come “pers.” e quella aggettivale come “dim.” avendo cura di lemmatizzare distintamente, in entrambi i casi, la forma del maschile, del femminile e del neutro.

2.6.2 I pronomi relativi

Per il relativo *qui*, *quae*, *quod*, che LiLa classifica come “pronoun”, si adotta la marca funzionale “rel.” avendo cura di distinguere la forma di citazione sulla base del genere, quindi *qui* rel., *quae* rel., *quod* rel. a seconda delle caratteristiche dell’occorrenza e dell’accordo stabilito dal relativo con l’antecedente nominale o pronominale.

Qualora ogni tipo di accordo ipotizzabile sulla base della norma del latino classico risulti disatteso, il lemma coinciderà con la forma attestata, avendo cura di inserire nel disambiguatore l’annotazione “forma invariabile” (cfr. § 1.2.2).²⁷

Le varianti grafiche non consuete per il relativo saranno ricondotte alle forme consuete del latino classico: pertanto la forma *que* attestata in alcune

²⁷ I relativi invariabili censiti in MED.ITA-LEMM sono quattro: *qui*, *quae*, *quem* (attestato solo per la funzione Oggetto, cfr. l’analisi semantica e sintattica in Greco 2025) e *quod*.

fonti sarà associata al lemma *quae* rel. (quando sia proforma per un femminile singolare o plurale) o *quod* rel. (quando sia grafia per *quae*, forma del relativo neutro plurale).

2.6.3 I pronomi personali

I pronomi personali (in LiLa “pronoun”) saranno classificati con la categoria “pers.”. La forma di citazione sarà quella che esprime i casi del latino classico: ad esempio, per la prima persona, “*ego* pers.”, “*mei* pers.”, “*mihi* pers.”, “*me* pers.”. Se la forma che in latino esprime l’ accusativo e l’ ablativo risulterà sovra-estesa e utilizzata, ad esempio, in contesti in cui ci si aspetterebbe il dativo, la forma di citazione del lemma sarà comunque “*me* pers.”.

La procedura di lemmatizzazione consigliata per questa serie pronominale è quella ad oggi adoperata per il repertorio formale del Corpus TLIO:²⁸ il principale vantaggio di questa procedura consiste nella possibilità di offrire all’utente un quadro della presenza e della distribuzione di tutte le forme del paradigma dei personali, indipendentemente dalla funzione sintattica effettiva svolta nei singoli contesti di occorrenza (aspetto che richiede un’analisi raffinata).

Per la lemmatizzazione delle forme *mecum*, *tecum* e *secum* rimando al § 1.2.2, nella sezione dedicata all’uso del disambiguatore per la segnalazione dell’uso delle enclitiche.

In mancanza di un pronome personale che rappresenti la terza persona singolare e plurale nella funzione Soggetto, per la forma pronominale dei dimostrativi *is*, *ea*, *id*; *ille*, *illa*, *illud*; *ipse*, *ipsa*, *ipsum*; *iste* *ista*, *istud* – che in più casi esprimono tale concetto – si adotterà la categoria grammaticale “pers.” (cfr. anche § 2.6.1).

²⁸ Cfr. *Note di lemmatizzazione*, nel paragrafo *Articoli e pronomi*. Nel piano di lavoro enunciato dalla *Grammatichetta* (§ 28, p. 155) tutti i personali risultavano esclusi dalla lemmatizzazione ad eccezione di *si*, *se*, *ne*, *costui*, *costei*, *costoro*, *colui*, *colei*, *coloro*.

2.6.4 I numerali

La categoria grammaticale “num.” (= numerale) sarà associata ai lemmi che rappresentano numeri cardinali e ordinali.

La forma di citazione del lemma sarà corrispondente alla variante al nominativo riportata dai lessici di riferimento per il latino. Si considera in questa serie anche il lemma *ambo*.

2.6.5 Le ambiguità grammaticali e testuali

La categoria grammaticale “a.g.” (= ambiguità grammaticale) potrà essere accostata alle forme d’incerta classificazione utilizzate, ad esempio, come espressioni scherzose o parole di scongiuro (cfr. *Note di lemmatizzazione* nel paragrafo *Scongiuri*).

Per l’uso della categoria grammaticale “a.t.” (= ambiguità testuale) riporto la descrizione presente in *Libro bianco*:

«Con questa categoria si segnalano le occorrenze delle forme riconducibili a corruzione testuale. Di regola la forma da mettere a lemma è la forma stessa senza alcuna normalizzazione. In disambiguatore può comparire un’annotazione sull’origine della corruzione: un rinvio al lemma a cui si riferisce la forma corretta secondo la formula “v. lemma” se il lemma sia documentato nel Corpus TLIO o “err. per lemma” se il lemma non è documentato nel Corpus TLIO, o la probabile ragione della corruzione (lacuna, crux, cassatura)».

Nel corpus MED.ITA la marca “a.t.” potrà essere usata per segnalare forme graficamente o morfologicamente aberranti, pur riconducibili, almeno ipoteticamente, a lemmi specifici; si consiglia tuttavia di limitare la creazione di lemmi-*hapax* che rischiano di restare al di fuori delle ricerche automatiche e che, senza dubbio, non potranno supportare le procedure della lemmatizzazione semi-automatica. Riporto di seguito l’esempio correlato alla lemmatizzazione della forma *pascuit* inserita in una sequenza nominale: forma di per sé priva di senso (probabile errore per un plurale *pascuis*), è stata associata al lemma *pascuum* -i s.n. Il confronto con contesti analoghi per struttura formulare avvalora questa ipotesi:

«tertjia partem de omnem substantjia meam [...] arboribus fructiferi bel in-fructiferis, aquis, silbis, *pascuit*, bineas, clausoria...» (ChLA² LIII, 1, 809 (pugl.), p. 18.5);

«transfundimus iure proprietario ac perdonamus, casis videlicet terris, hortis, pratis, *pascuis*, vineis, campis...» (CDVer 18, 890 (ven.), p. 22.30);

«terra sementaricia pedica una in in(tegro) [...] una cum terris sementariciis, kamps hac *pascuis*, cum terminis limitibusq(ue) suis, cultum v(e)l incultu(m)...» (SMTrastevere 2, 1026 (laz.), 2, p. 293, col. 2.9).

Quanto più la forma ambigua risulterà distante dal suo probabile antecedente, tanto più sarà preferibile lemmatizzarla come “ambiguità testuale”: si veda il caso di *quabit* a.t. (*quodlibet?*), a confronto con un contesto documentario utile per ricostruire la probabile forma corrotta:

«quid si vobis menime inantistare potuerimus aud si nos ipse bind(itor) p(er) *quabit* ingenio retornare q(ue)sierimus» (ChLA² LIII, 14, 845 (pugl.), p. 90.12);

«si nos ipsis ex eadem nostra venditio(ne) vobis cumcausaverimus et eam vobis per *quodlibet* ingenio tol(lere) vel contrariare seu remove quesierimus» (CDTr II, 32, 1043 (pugl.), p. 104.9).

Naturalmente per alcune ambiguità testuali sarà il contenuto del testo a sostenere un’ipotesi ricostruttiva plausibile.

La forma *ano* dell’esempio riportato di seguito è un errore dello scrivente o comunque una forma corrotta riconducibile probabilmente a *grano* (lemma: *ano* a.t. [*granum* -i?]), parola presente nel co-testo:

«et betura a *ano* una et alia ad fenum et hopera una ad arare...» (SMTrastevere 2, 1026 (laz.), p. 293, col. 2.42)

«modia de *grano* quinque et de ordeum quatuor ad modium iuxtu(m)» (ib., p. 293, col. 2.39).

L’etichetta “a.t.” sarà utilizzata anche per le scrizioni erronee e di interpretazione dubbia che presuppongono la probabile univervazione di due o più componenti; anche in questo caso il disambiguatore presenterà l’interpretazione probabile, seguita dal punto interrogativo. Così, ad esempio, il

lemma per la forma *assilvis* (CDVer 255, 955 (ven.)) sarà “*assilvis* a.t.” (disambiguatore: “ac silvis?”).

2.6.6 Grafie alternative

Sono da classificare come *grafie alternative* (“g.a.”) le forme unverbate di sintagmi ricorrenti come *imperpetuum*, da ricondurre a *in perpetuum*, o, al contrario, le varianti graficamente scisse di forme lessicali erroneamente interpretate dagli scriventi come somma di due componenti, ad es. *at que* = *atque*.

Nei due casi la lemmatizzazione procederà come indicato di seguito:

forma *imperpetuum* → lemma: *in perpetuum*, c.g.: *g.a.*;

forma *at* → lemma: *at*, c.g.: *g.a.* (disambiguatore: “polir.: atque”); forma *que* → lemma: *que*, c.g.: *g.a.* (disambiguatore: “polir.: atque”).

Per le grafie scisse di lemmi le cui “parti” siano parole funzionali si utilizzerà la classificazione indicata al § 1.2.2. Esempifico la procedura proponendo la lemmatizzazione di *in simul*, g(rafia) a(lternativa) per *insimul*:

forma *in* → lemma: *in*, c.g.: *prep.* (disambiguatore: “polir.”);

forma *simul* → lemma: *simul*, c.g.: *avv.* (disambiguatore: “polir.”)

Per le scritture erronee e d’interpretazione dubbia che presuppongono la probabile unverbazione di due o più parole, cfr. il § 2.6.5.

2.6.7 Antroponimi e toponimi

Come già anticipato nel paragrafo che descrive l’uso del disambiguatore (§ 1.2.1), gli antroponimi e i toponimi adottano come categoria grammaticale l’etichetta “n.p.” (= nome proprio).

Come nel caso degli elementi lessicali, per gli antroponimi e i toponimi appartenenti al repertorio latino o confluiti in esso, la forma di citazione sarà il lemma al nominativo schedato dai repertori lessicografici ed enciclopedici di riferimento (*Onomasticon*, De Vit; DizTop; Rossebastiano–Papa 2005).

Per i nomi di epoca medievale di diversa origine si dovrà ricorrere, spesso, ad altre fonti e, in molti casi, a soluzioni differenti.

Rientrano tra gli antroponimi tutti gli elementi utilizzati nei testi per denominare individui specifici: pertanto, devono essere classificati come antroponimi anche i soprannomi e gli eponimi, per lo più trascritti con la prima lettera maiuscola, che seguano un nome proprio: è il caso, ad esempio, di *Christus* (n.p. (antr.:)), che cooccorre quasi sempre con *Iesus* (n.p. (antr.:)), ma anche di eponimi deaggettivali come *Maior* o *Minor*, talora accostati a un nome proprio. Non è, invece, un componente onomastico l'aggettivo *sanctus* che precede il nome di un santo; tuttavia, nel caso in cui la stringa “*sanctus + nome*”, ad es., *Sanctus Iohannes*, designi un luogo specifico – e sia, dunque, un toponimo – ciascun componente della stringa sarà lemmatizzato come parte del *nome geografico* in questa maniera:

- (1) lemma: *sanctus*, c.g.: agg., disambiguatore: “polir. in n.g.” (= [componente di una formazione] polirematica [attestata] in [un] nome geografico);
- (2) lemma: *iohannes*, c.g.: n.p., disambiguatore: “antr.: polir. in n.g.”

Saranno da lemmatizzare in questa maniera gli agionimi riferiti a centri di piccola o grossa estensione, ma non a chiese e monasteri e più in generale a istituzioni religiose e a costruzioni *dedicate a specifici santi*.

Esemplifico questa casistica nelle due immagini presentate di seguito, che illustrano il trattamento diverso dell'antroponimo *Maria* – parte della sequenza *Santa Maria* – con riferimento a una chiesa (fig. 10) e a uno specifico territorio (fig. 11). L'agionimo è considerato a tutti gli effetti un toponimo solo nel secondo caso:

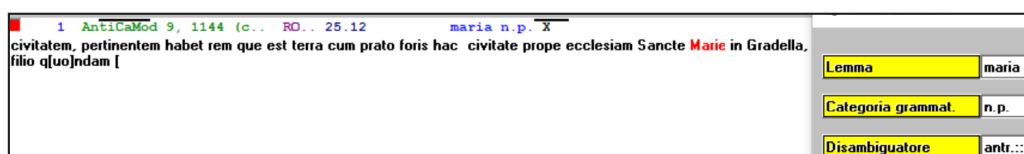


fig. 10

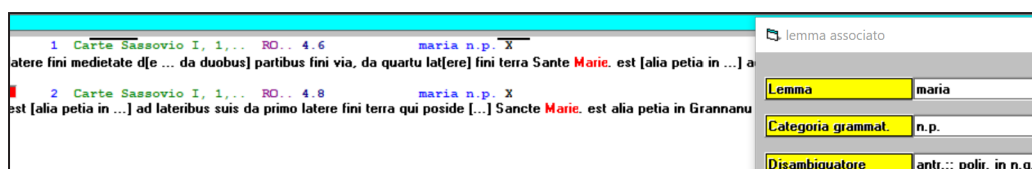


fig. 11

Per gli antroponimi germanici o di altra provenienza, quando la forma del nominativo non sia facilmente ricostruibile, si dovranno consultare i principali repertori di riferimento: nel lemmatizzare i testi di MED.ITA-LEMM abbiamo utilizzato fondamentalmente le monografie di Morlicchio (1985) e Francovich Onesti (1999), che esaminano l'onomastica longobarda documentata in fonti archivistiche ed epigrafiche di area italiana, e quella di Förstermann (1900-1916), silloge di onomastica germanica basata sullo spoglio di numerose fonti medievali, in parte anche italiane.

In mancanza di dati di contrasto, l'antroponimo potrà essere normalizzato almeno provvisoriamente secondo la forma attestata: in tal caso il disambiguatore recherà, dopo l'etichetta "antr.:" seguita da uno spazio, l'annotazione "sic" (= lemma coincidente con la forma attestata nel corpus). La stringa standardizzata che comparirà nel disambiguatore sarà dunque "antr.: sic".

Gli antroponimi non latini compaiono nella documentazione di aree diverse (e di fasi diverse) con un'enorme variabilità grafica: il modello ideale, in questo caso, è la forma latinizzata registrata dai repertori onomastici mediolatini oppure la variante registrata più anticamente e più a nord nel quadro delle attestazioni di area italiana. Si veda ad es. *Alahis*, nome già registrato nei sec. VII-VIII in documenti provenienti da diverse aree italiane (cfr. Francovich Onesti 1999, p. 200): a questa forma di citazione è stato ricondotto l'antroponimo *Alasi* documentato in un atto redatto a Taranto nell'anno 809 (ChLA² LIII, 1), con riscontro nel tipo *Alaisi*, nominativo individuato da Morlicchio (1985: 87) in carte salernitane del IX secolo; si veda ancora *Vvarnifrida*, individuato nelle *Epistole* di Gregorio Magno da Francovich Onesti (1999: 220), lemma a cui è associato l'antroponimo *Vuarnifreda* documentato nel già menzionato atto tarantino: la variante *Uuarnifrit*, documentata nel 715 a Siena e a Pavia, è cronologicamente successiva e indubbiamente meno affine alla forma del corpus.

È utile segnalare che per molti antroponimi germanici schedati nel modulo MED.ITA-LEMM il campo "commento" interno alla scheda di lemmatizzazione è stato utilizzato per registrare il dato di riscontro fornito dagli studi prima citati; non sono pochi, d'altronde, gli antroponimi assenti nelle raccolte di onomastica menzionate. Essi costituiscono un campione di antroponimia germanica potenzialmente meritevole di attenzione.

2.6.7.1 I microtoponimi dalla struttura polirematica

Una prassi di lemmatizzazione specifica dovrà essere adottata per le denominazioni volgari o volgareggianti applicate a luoghi di estensione areale ristretta, spesso noti solo all'interno di specifiche comunità locali. I cosiddetti microtoponimi, utilizzati per denominare appezzamenti di terra, fiumiciattoli, canali, fossi, piccole alture, ecc., sono spesso introdotti nei testi da formule come (*locus*) *qui dicitur / vocatur*, formule che segnano fenomeni di commutazione di codice o di adattamento allo scritto di un dato connesso a una tradizione prettamente orale: un chiaro segno del passaggio al volgare è spesso dato dal ricorso a denominazioni strutturate nella forma di indicazioni di appartenenza, specificazione o adiacenza che contengono l'articolo determinativo, cfr. ad es. *fons de la Mortella*, *Rigo de li Porcili*, *locus qui dicitur a li Capitilli*, *in locum qui dicitur a le Cese*. L'articolazione del microtoponimo in questi casi è di tipo polirematico per cui si ricorrerà alla strategia di lemmatizzazione indicata per i lemmi attestati anche in forma analitica (cfr. § 1.1): ogni "componente" della denominazione dovrà essere annotato come "polir. in n.g." (= [componente di una formazione] polirematica [attestata] in [un] nome geografico. Per il lemma corrispondente all'articolo, componente estraneo alla struttura grammaticale del latino, si utilizzerà la categoria "volg.". Ogni forma rappresentativa dell'articolo sarà posta a lemma: i lemmi saranno, dunque, "*la* volg.", "*le* volg.", "*lu* volg.", "*lo* volg.", "*li* volg.", e tutti recheranno nel disambiguatore l'annotazione "polir. in n.g.". ²⁹

Propongo un esempio di lemmatizzazione per i componenti delle sequenze *Rigo de li Porcili* (1), attestata in CDTTr II, 29, 1041 (p. 94.13;

²⁹ Si utilizzerà la categoria "volg." anche per gli articoli disseminati nelle formule documentarie che precisano i confini di una proprietà, cfr. ad es. «de uno latu fine *la* terra Sancti Iohannis» (CDTr II, 40, 1049 (abr.), p. 129.26), «a quartus lato fini *la* terra de Cese» (Carte Sassovivo I, 13, 1079 (umbr.), p. 22.27). In questi casi il lemma dedicato non avrà annotazioni nel disambiguatore. La stessa categoria deve essere utilizzata per le forme dell'articolo indeterminativo, se questa funzione risulta esplicita: cfr. «si cu *nu* omine coludiu factu vel facere voluerimus» (Carte Sassovivo I, 84, 1091(?) (umbr.), 84, p. 131.1).

forme *rigo, de, li, porcili*) e (*loco qui dicitur Montis Rothari* (2), attestata in AntiCaMod 9, 1144 (camp.) (p. 25.9; forme *montis, rothari*). Preciso che, con riferimento a (1), si considera parte della sequenza onomastica solo il sintagma introdotto dall'articolo:

(1)

lemma: <i>rigus -i</i>	c.g.: s.m.	disambiguatore: "mlat."
lemma: <i>de</i>	c.g.: prep.	
lemma: <i>li</i>	c.g.: volg.	disambiguatore: "polir. in n.g."
lemma: <i>porcile -is</i>	c.g.: s.n.	disambiguatore: "mlat., polir. in n.g."

(2)

Lemma: <i>mons montis</i>	c.g.: s.m.	disambiguatore: "polir. in n.g."
lemma: <i>rothari</i>	c.g.: n.p.	disambiguatore: "antr.: polir. in n.g."

Questa strategia di lemmatizzazione consente di escludere la citazione a lemma di denominazioni analitiche, che potrebbero essere registrate anche in modo variabile da operatori diversi, a seconda della stringa concretamente menzionata dai testi (solitamente non normalizzata); consente, inoltre, di registrare come componenti lessicali anche le occorrenze di parole utilizzate nell'onomastica (ad es. *porcile -is*).

Nomi comuni come *vallis, mons, vallo, campus, fundus, terra*, ecc. seguiti da un nome proprio o da un determinante usato per identificare il luogo, saranno considerati parte della sequenza onomastica (di tipo polirematico) solo quando l'intera sequenza si rinverrà introdotta da formule come *locus qui / ubi dicitur / appellatur*, o risulterà allineata ad altri toponimi di tipo monorematico. Cfr. (2) e vd., inoltre, gli esempi seguenti:

(3) «in locum qui appellatur *Pantano Maio(re)*» (SMTrastevere 2, 1026 (laz.), p. 293, col. 2.15):

lemma: <i>pantanum -i</i>	c.g.: s.n.	disambiguatore: "mlat., polir. in n.g."
lemma: <i>maior</i>	c.g.: agg.	disambiguatore: "polir. in n.g."

(4) «infra fines Pusterna et *Fosa Rabeda*» (CDVer 18, 890 (ven.), p. 22.26):

lemma: *fossa -ae* c.g.: s.f. disambiguatore: "polir. in n.g."

lemma: *rapidus* c.g.: agg. disambiguatore: "polir. in n.g."

ma

(5) *ipsa res mea de campo Calia*» (ChLA² LIII,4, 823 (luc.), p. 34.13):

lemma: *campus -i* c.g.: s.m.

lemma: *calia* c.g.: n.p. disambiguatore: *n.g.::*

Questa procedura consente di limitare la duplicazione dei lemmi e arricchisce il serbatoio delle occorrenze disponibili all'associazione con lemmi di tipo esclusivamente lessicale.

Per il trattamento degli etnici si rimanda al § 2.2.

3 Fonti lessicografiche di riferimento

Indico di seguito in forma sintetica le risorse lessicografiche cartacee e digitali da utilizzare per il controllo e la selezione delle forme di citazione rappresentative dei lemmi.

Il ThLL rappresenta la risorsa principale per verificare l'appartenenza di una forma lessicale o onomastica al repertorio latino. Per un controllo rapido si può ricorrere naturalmente anche a vocabolari compatti bilingui di uso anche scolastico come Calonghi; si consiglia peraltro di consultare le risorse lessicografiche accessibili alla ricerca simultanea per parole chiave disponibili nel portale di *Corpus Corporum*: la piattaforma lessicografica allestita nel quadro del progetto (<https://mlat.uzh.ch/dictionaries>) include repertori celebri come il Lewis-Short, il Georges e il Gaffiot per il latino classico e il Du Cange, il MLW (solo per il segmento A-C), il *Novum Glossarium* e altri lessici meno noti, comunque di riferimento. La consultazione dei contenuti lessicografici del portale consente di identificare rapidamente l'attestazione classica o solo medievale delle forme riconosciute nei testi del corpus. È di grande utilità anche il *Database of Latin Dictionaries* (DLD), una risorsa ad accesso limitato, consultabile, tuttavia, attraverso i servizi di accesso digitale di numerose istituzioni scientifiche.

Per riscontri sulla classificazione delle parole grammaticali di attestazione mediolatina si dovrà fare riferimento principalmente alle indicazioni esplicite del MLW e alla forma del lemma presentata da Niermeyer; il Du Cange è privo di annotazioni di questo genere e veicola informazioni morfosintattiche prevalentemente attraverso le definizioni.

In generale le strategie di lemmatizzazione del MLW e del Niermeyer saranno preferite a quelle del Du Cange che non evidenzia un metodo lessicografico ragionato ed esplicito che consenta di distinguere le forme che, nella ricca variantistica del latino medievale, possono adeguatamente rappresentare un sistema lessicale testimoniato da molteplici fonti di diversa provenienza, e dunque aperto e fortemente incline alla variazione.³⁰

È uno strumento prezioso anche l'Arnaldi-Smiraglia, che ha focalizzato lo spoglio lessicale su fonti documentarie e letterarie mediolatine di area italiana ed è ad oggi accessibile digitalmente tra le risorse lessicografiche di DLD.

Riferimenti bibliografici citati

Adamik 2021 = Béla Adamik, *Romanization and Latinization of the Roman Empire in the light of data in the Computerized Historical Linguistic Database of Latin Inscriptions of the Imperial Age*, in «Journal of Latin Linguistics» 20.1 (2021), pp. 1-19.

Arnaldi-Smiraglia = *Latinitatis Italicae Medii Aevi Lexicon (saec. V ex. – saec. XI in.)*. Editio altera, a cura di Francesco Arnaldi e Pasquale Smiraglia, Firenze, SISMEL – Edizioni del Galluzzo, 2001.

Artale 2019 = Elena Artale, *I lemmi non toscani nel Corpus TLIO*, in *Italiano antico, italiano plurale. Testi e lessico del Medioevo nel mondo digitale*. Atti del convegno tenutosi a Firenze, il 13 e 14 settembre 2018, presso la sede dell'OVI, a cura di Lino Leonardi e Paolo Squillacioti, «Bollettino dell'Opera del Vocabolario Italiano», Supplementi, 7 (2019), pp. 159-172.

Boccellari-Iorio Fili 2013 = Andrea Boccellari e Domenico Iorio Fili, *Il supporto dell'informatica al vocabolario*, in *Diverse voci fanno dolci note. L'Opera del*

³⁰ Questi aspetti, legati alla struttura e al diverso approccio dei più importanti lessici mediolatini ad oggi disponibili, sono approfonditi da Giuliani-Aprea i.c.s.

- Vocabolario Italiano per Pietro Beltrami*, a cura di Pär Larson, Paolo Squillacioti e Giulio Vaccaro, Alessandria, Edizioni dell'Orso, 2013, pp. 15-30.
- Calonghi = Ferruccio Calonghi, *Dizionario latino-italiano*, 3ª edizione interamente rivista ed aggiornata del dizionario Georges-Calonghi, 11ª tiratura, Torino, Rosenberg & Sellier, 1972.
- Corpus Corporum = *Corpus Corporum Project*, repository of Latin texts: <https://mlat.uzh.ch/home>.
- Corpus TLIO = *Corpus TLIO per il vocabolario*, diretto da Pär Larson, Elena Artale, Diego Dotto, Firenze, Istituto Opera del Vocabolario Italiano, <<http://tlioweb.ovi.cnr.it>> (ultimo aggiornamento: 8 settembre 2025).
- Cortelazzo 1970 = Manlio Cortelazzo, *L'influsso linguistico greco a Venezia*, Bologna, Pàtron, 1970.
- DLD = *Database of Latin Dictionaries (2026)*, creato da Tim Denecker, Brepols online: <https://www.brepols.net/series/DLD-O>.
- De Vit = *Totius latinitatis onomasticon, opera et studio doct. Vincentii De-Vit lucubratum*, Prati, Typis Aldinianis, 1859-1887.
- DiFri = *Dizionario del Franco-Italiano*, diretto da Francesca Gambino: <https://www.rialfri.eu/il-progetto>.
- DizTop = Giuliano Gasca Queirazza et al., *Dizionario di toponomastica: storia e significato dei nomi geografici italiani*, Torino, UTET, 1990.
- Du Cange = *Glossarium Mediae et Infimae Latinitatis*, conditum a Carolo du Fresne Domino Du Cange, [...] Editio nova aucta pluribus verbis aliorum scriptorum a L. Favre, Niort, Favre 1883-1887; online: <http://ducange.enc.sorbonne.fr/>.
- Forcellini = *Lexicon totius Latinitatis* ab Aegidio Forcellini... lucubratum, deinde a Iosepho Furlanetto... emendatum... nunc vero curantibus Francisco Corradini et Iosepho Perin..., Patavii, Typis Seminarii, 1940.
- Förstermann 1900-1916 = Ernst Wilhelm Förstermann, *Altdeutsches Namenbuch*, vol. I: *Personennamen*, II: *Ortsnamen*, 2 voll., Bonn, P. Hanstein, 1900 (zweite, völlig umgearbeitete Auflage) – 1916 (dritte, völlig neu bearbeitete erweiterte Auflage, herausgegeben von Hermann Jellinghaus).
- Francovich Onesti 1999 = Nicoletta Francovich Onesti, *Vestigia longobarde in Italia, 568-774: lessico e antroponimia*, Roma, Artemide, 1999.

- Gaffiot = Felix Gaffiot, *Dictionnaire illustré latin-français*, Paris, Impr. de Vaugirard, 1934.
- Gambino 2024 = Francesca Gambino et alii, *Il francese d'Italia e il progetto FrIngE*. Panoramica generale e casi di studio, in «Francigena» 10 (2024), pp. 285-340.
- Georges = Karl Ernst Georges, *Ausführliches lateinisch-deutsches Handwörterbuch. Aus den Quellen zusammengetragen und mit besonderer Bezugnahme auf Synonymik und Antiquitäten unter Berücksichtigung der besten Hilfsmittel ausgearbeitet*; 11ª edizione: ristampa dell'ottava edizione migliorata e ampliata da Heinrich Georges, Leipzig, Hahn, 1913.
- Giuliani 2023 = Mariafrancesca Giuliani, *Variazione e omogeneità nel più antico repertorio lessicale italiano*, in «Bollettino dell'Atlante Lessicale degli Antichi Volgari Italiani» 11-2018 (2023), pp. 9-44.
- Giuliani–Aprea i.c.s. = Mariafrancesca Giuliani e Fabio Aprea, *Il latino medievale allo specchio: la prospettiva della lessicografia italo-romanza*, in corso di stampa in «Testi e Linguaggi», 20, numero monografico dedicato ai “Cantieri di lessicografia storica dell'italiano e dei dialetti d'Italia”, a cura di Sergio Lubello, Carolina Stromboli e Francesco Crifò, 2026.
- Grammatichetta = Piero Esperti, *Grammatichetta della lingua italiana ad uso del calcolatore*, in *Al servizio del vocabolario della lingua italiana*, a cura di d'Arco Silvio Avalle, Firenze, Accademia della Crusca, 1979, pp. 123-87.
- Greco 2025 = Paolo Greco, *Sintassi e semantica del pronome quem nelle carte notarili della Langobardia Minor (IX secolo)*, in „Parla, e sie breve e arguto“. *Festschrift für Maria Selig / Studies in honor of Maria Selig*, a cura di Laura Linzmeier, Alexander M. Teixeira Kalkhoff ed Evelyn Wiesinger, Tübingen, Narr, pp. 65-69.
- Greco–Cotugno–Giuliani 2024 = Paolo Greco, Alessio Cotugno e Mariafrancesca Giuliani, *Il progetto MEDITA. La documentazione non letteraria mediolatina e la lessicografia storico-etimologica italo-romanza*, “Quaderni Veneti” 13 (2024), pp. 209-228.
- Greco–Cotugno–Giuliani i.c.s. = Paolo Greco, Alessio Cotugno e Mariafrancesca Giuliani, *Nuovi strumenti per la lessicografia latina e romanza: la documentazione latina medievale di area italiana e il progetto MEDITA*, articolo presentato per *Proceedings of the 15th International Colloquium on Late and Vulgar Latin*, Bavarian Academy of Sciences and Humanities (Munich, Germany, 2-6 September, 2024), a cura di M. Selig, in corso di stampa.

- Korkiakangas 2020 = Timo Korkiakangas, *Theoretical and pragmatic considerations on the lemmatization of non-standard Early Medieval Latin charters*, in «Studi e Saggi Linguistici», 58,1 (2020), pp. 67-94.
- Larson 1996, = Pär Larson, *Glossario diplomatico toscano avanti il 1200: studi sull'italiano preletterario*, Stockholms universitet, Edsbruck, Akademitryck (ristampa con integrazioni e una nuova introduzione dell'edizione 1995, Firenze, Accademia della Crusca).
- LEI = *Lessico Etimologico Italiano*, fondato da Max Pfister, diretto da Elton Prifti e Wolfgang Schweickard, Wiesbaden, Reichert, 1979- (<<https://lei-digitale.it/>>).
- Lemmatisierung MLW*, indicazioni sugli orientamenti seguiti dal MLW nella costituzione del lemmario, sintesi redatta nel 2025 da Monica Niederer per i ricercatori del progetto MEDITA.
- Lewis Short = Charlton T. Lewis e Charles Short, *A Latin Dictionary. Founded on Andrews' edition of Freund's Latin dictionary. Revised, enlarged, and in great part rewritten*, Oxford, Clarendon Press, 1879.
- Libro bianco* = Dotto, Diego, *Proposta per un Libro bianco della lemmatizzazione del Corpus TLIO*, documento di lavoro ad uso della lemmatizzazione del Corpus TLIO (ultimo aggiornamento: 10.09.2021).
- Morlicchio 1985 = Elda Morlicchio, *Antroponimia longobarda a Salerno nel IX secolo: i nomi del Codex diplomaticus Cavensis*, Napoli, Liguori, 1985.
- MLW = *Mittellateinisches Wörterbuch bis zum ausgehenden 13. Jahrhundert*, hrsg. von der Bayerischen Akademie der Wissenschaften und der Deutschen Akademie zu Berlin, redigiert von Otto Prinz, München, Beck, 1959 e sgg; online: <https://mlw.badw.de/das-projekt.html>; in versione database lessicografico: <https://www.woerterbuchnetz.de/?sigle=MLW&lemid=A00001>
- Mosti 2015 = Rossella Mosti, *Il lemmario del Tesoro della Lingua Italiana delle Origini*, in «Bollettino dell'Opera del Vocabolario Italiano», 19-20 (2015), pp. 405-425.
- Niermeyer = Jan Frederik Niermeyer, *Mediae Latinitatis lexicon minus. Lexique latin médiéval-français-anglais / A medieval Latin-french-English dictionary*, 2 voll., Leiden, Brill, 1954-1976.
- Note di lemmatizzazione*, documento di lavoro ad uso della lemmatizzazione del Corpus TLIO, a cura di Elena Artale (ultimo aggiornamento 25.05.2020).

- Novum Glossarium* = Franz Blatt *et al.*, *Novum glossarium mediae Latinitatis ab anno DCCC usque ad annum MCC*, Union académique internationale, Hafniae; Bruxelles; Genève, Munksgaard; Droz, 1959-.
- Onomasticon* = *Thesaurus Linguae Latinae. Onomasticon*, vol. I: A – B, vol. II: C – Cyzistra, vol. III: D – Dzoni, 1900-.
- Passarotti *et al.* 2020 = Marco Passarotti, Francesco Mambrini, Greta Franzini, Flavio Massimiliano Cecchini, Eleonora Litta, Giovanni Moretti, Paolo Ruffolo, Rachele Sprugnoli, *Interlinking through lemmas. The lexical collection of the LiLa Knowledge Base of linguistic resources for Latin*, in “Studi e Saggi Linguistici” 58 (1), 2020, pp. 177-212.
- Rohlf 1966 = Gerhard Rohlf, *Grammatica storica della lingua italiana e dei suoi dialetti*, traduzione di Temistocle Franceschi e Maria Caciagli Fancelli, 3 voll., I (*Fonetica*), Torino, Einaudi, 1966 (nuova edizione con presentazione di Claudio Marazzini e materiali inediti, Bologna, Il Mulino – Firenze, Accademia della Crusca, 2021).
- Rossebastiano–Papa 2005 = Alda Rossebastiano ed Elena Papa, *I nomi di persona in Italia: dizionario storico ed etimologico*, Torino, UTET, 2005.
- Sornicola 2012 = Rosanna Sornicola, *Bilinguismo e diglossia dei territori bizantini e longobardi del Mezzogiorno. Le testimonianze dei documenti del IX e X secolo*, Napoli, Giannini, 2012.
- Sornicola 2017 = Rosanna Sornicola, *La morfologia nominale: polimorfismo e polifunzionalismo nei sistemi di flessione*, in *Sistemi, norme scritture. La lingua delle più antiche carte cavensi*, a cura di Rosanna Sornicola, Elisa D’Argenio e Paolo Greco, Napoli, Giannini, 2017, pp. 85-134.
- Stotz 2013 = Peter Stotz, *Il latino nel Medioevo. Guida allo studio di un’identità linguistica europea*, ed. italiana a cura di Luigi G. G. Ricci, Firenze, SISMEL – Edizioni del Galluzzo, 2013.
- TB = Niccolò Tommaseo e Bernardo Bellini, *Dizionario della lingua italiana*, Torino, UTET, 1861-1879; online: <https://www.tommaseobellini.it/#/>.
- ThLL = *Thesaurus Linguae Latinae*, Lipsia, Teubner, 1900-; online: <https://thesaurus.badw.de/tll-digital/tll-open-access.html>.
- TLIO = *Tesoro della lingua italiana delle Origini*, diretto da Paolo Squillaciotti [fondato da Pietro G. Beltrami, poi diretto da Lino Leonardi], Firenze, Istituto Opera del Vocabolario italiano, <<http://tlio.ovi.cnr.it/>>.

Fonti citate

AntiCaMod = *Le più antiche carte dell'Abbazia di San Modesto in Benevento (secoli VIII–XIII)*, a cura di Franco Bartoloni, *Regesta chartarum Italiae* 33, Roma, Istituto storico italiano per il Medio Evo, 1950.

Carte Sassovivo I = *Le carte dell'abbazia di S. Croce di Sassovivo*, pubblicate dalla Scuola speciale per Archivisti e Bibliotecari dell'Università di Roma, a cura di Giorgio Cencetti e Alessandro Pratesi, 7 voll., vol. I (1023-1115), Firenze, Olschki, 1973-1974.

CDAmal I = *Codice diplomatico amalfitano. Le pergamene di Amalfi esistenti nel Real Archivio di stato di Napoli*, a cura di Riccardo Filangieri di Candida, 2 voll., vol. I (907-1200), Napoli, Tipografia Silvio Morano, 1917.

CDTr = *Codice diplomatico del monastero benedettino di S. Maria di Tremiti (1005–1237)*, a cura di Armando Petrucci, *Fonti per la storia d'Italia* 98, Roma, ISIME, 3 voll., 1960.

CDVer = *Codice diplomatico veronese del periodo dei re d'Italia*, a cura di Vittorio Fainelli, Venezia, a spese della Deputazione, 1963.

ChLA² L = *Chartae Latinae Antiquiores: facsimile-edition of the Latin charters prior to the ninth century*, 2nd series, Ninth Century, 50: Italy 22: Cava dei Tirreni, a cura di Maria Galante, Dietikon-Zurich, Urs Graf Verlag, 1997.

ChLA² LIII = *Chartae Latinae antiquiores: facsimile-edition of the Latin charters prior to the ninth century*, 2nd series, Ninth Century, 53: Italy 25: Montecassino, Trani, Barletta, Benevento, a cura di Francesco Magistrale, Pasquale Cordasco e Clelia Gattagrisi, Dietikon-Zurich, Urs Graf Verlag, 1999.

PerGregArm (ed. Mazzoleni) = *Le pergamene del monastero di S. Gregorio Armeno di Napoli. I. La scrittura curialesca napoletana con n. XVI tavole*, a cura di Jole Mazzoleni, Napoli, Libreria Scientifica Editrice, 1973.

SGiorMag II = *San Giorgio Maggiore. Vol. II: Documenti (982-1159)*, a cura di Luigi Lanfranchi, Venezia, Comitato Pubblicazione delle Fonti relative alla Storia di Venezia, 1968.

SGiovEvTorc = *S. Giovanni Evangelista di Torcello (1024-1199)*, a cura di Luigi Lanfranchi, Venezia, Alfieri, 1948.

SMTrastevere = Paolo Radiciotti. *Le pergamene di Santa Maria in Trastevere. Storia del fondo ed edizione delle pergamene anteriori al 1200*, in «Mélanges de l'École française de Rome – Moyen Âge», 122 (II), pp. 279-317.

Stat. San Gimignano, 1255 = Statuto di San Gimignano del 1255, a cura di Silvia Diacciati e Lorenzo Tanzini, Firenze, Olschki, 2016.

RIASSUNTO Il contributo è incentrato sugli orientamenti e le procedure che sono state seguite per la lemmatizzazione integrale di una parte dei testi mediolatini documentari contenuti nel modulo MED.ITA-LEMM, e che si dovranno applicare per esteso all'intero corpus MED.ITA. Focus dell'articolo è un manuale operativo (§§ 1-3) che definisce le regole e gli standard della lemmatizzazione, al momento realizzata nell'ambiente di gestione digitale di GATTO 3.3 e strutturata per molti aspetti secondo le norme di lemmatizzazione del Corpus TLIO. Nella premessa (§ 0) si indicano il senso e i principi della lemmatizzazione, basata su un vocabolario di macchina esemplato fondamentalmente sul repertorio lessicale dei lessici latini e mediolatini di riferimento, che tiene conto, tuttavia, del modello lessicale italoromanzo da cui provengono i numerosi prestiti immessi nel latino dei testi. Il vocabolario di macchina, in questi termini, non deriva dall'importazione del lemmario di vocabolari specifici ma è il prodotto di un confronto ragionato tra le parole del corpus e il repertorio dell'insieme dei lessici assunti come riferimento.

Parole chiave: Latino medievale, Lessicografia italoromanza, Lessicografia mediolatina, Testi pratici medievali, Lemmatizzazione, Linguistica dei corpora, Digital humanities, Norma e variazione.

ABSTRACT - The paper focuses on the guidelines and procedures that have been followed for the complete lemmatisation of part of the Medieval Latin documentary texts contained in the MED.ITA-LEMM module, and which will have to be fully applied to the entire MED.ITA corpus. The work focuses on an operating manual (§§ 1-3) that defines the rules and standards of lemmatisation, currently implemented in the GATTO 3.3 digital management environment and structured in many respects according to the lemmatisation rules of the TLIO Corpus. The introduction (§ 0) explains the meaning and principles of lemmatisation, based on a machine vocabulary modelled essentially on the lexical repertoire of the Latin and Medieval Latin

reference lexicons, which nevertheless takes into account the Italo-Romance lexical model from which the numerous loanwords introduced into the Latin of the texts originate. In these terms, the machine vocabulary doesn't derive from the importation of the lexical list coming from specific dictionaries, but is the product of a reasoned comparison between the words in the corpus and the lexical repertoire of the set of dictionaries taken as a reference.

Keywords: Medieval Latin, Italo-Romance lexicography, Middle Latin lexicography, Medieval practical texts, Lemmatization, Corpus linguistics, Digital humanities, Norm and variation.

Contatto dell'autrice: giuliani@ovi.cnr.it; mariafrancesca.giuliani@cnr.it