**DIGITAL LATIN III**

**Digital tools and AI for research on Latin lexicon**

SIENA, 15-16 giugno 2026

Santa Chiara Lab, via di Valmontone 1

Abstracts

**Martina Venuti (Università 'Ca Foscari Venezia), *The glossary* LaLaLexiT. Late Latin Lexicon in Transition**

Presentation of the glossary *LaLaLexiT. Late Latin Lexicon in Transition* (Edizioni Ca' Foscari, 2025–; DOI: 10.30687/LLLXT/2375-1355), an open-access digital scholarly edition featuring lexicographic entries based on a selected corpus of Latin prose and poetry from the 3rd to the 7th centuries AD. The paper presents the principal characteristics of the edition, focusing on its editorial criteria, technological infrastructure, publication model and functionalities, as well as the approaches adopted for the preservation, and assessment of the research results.

**Nadia Rosso, Gianmario Cattaneo (DigilibLT – Univ. Piemonte Orientale), *Il mondo DigilbLT. Prospettive lessicografiche***

Nella prima metà dell'intervento, verrà presentato il nuovo sito del progetto *DiglibLT*, lanciato nel gennaio 2026, a seguito di un vero e proprio redesign al passo con i tempi, mettendone in evidenza upgrade, criticità e nuove sfide; un focus particolare verterà sulle maschere di ricerca e sul passaggio tra il vecchio e il nuovo sistema di ricerca lessicografica. Nella seconda metà, tramite la presentazione di alcuni casi studio, da una parte verranno presentati i modi in cui la ricerca lessicografica può aiutare gli studiosi nella redazione delle schede biobibliografiche che accompagnano i testi caricati su *DigilibLT*, dall'altra verrà mostrato come la ricerca di parole in *DigilibLT* possa fornire all'utente spunti interessanti in ambito

lessicografico e come queste ricerche possano combinarsi con quelle in *corpora* più ampi di testi latini.

**Marina Giani (Università degli Studi di Milano), *Digital Approaches to the Study of Medieval Latin Glossaries***

Medieval Latin glossaries constitute a privileged source for investigating the evolution of Latin lexicon from Late Antiquity through the Middle Ages. Their composite nature, characterized by the accumulation of materials from different periods, the variability of manuscript traditions, and the complex relationships linking lemmata, glosses, and source texts, makes them particularly suitable for digital approaches. This paper presents the proposed *LexLat* project, which builds upon the results of GLAM (*Glossaria Latina Medii Aevi*), a census of early medieval manuscript witnesses transmitting monolingual Latin glossaries. By providing a comprehensive survey of the manuscript tradition, GLAM has laid the groundwork for new editions of glossographical texts and, consequently, for the study of their lexical contents. The project aims to produce, on the one hand, a digital scholarly edition (both reconstructive and comparative) of the early medieval collections *Abba* and *AA*, through the Edition Visualization Technology (EVT) framework. On the other hand, it seeks to create a database-thesaurus of glosses organized around normalized entries linked to the variant forms attested in individual manuscript witnesses. This structure allows lexical data to be explored independently of specific textual witnesses while preserving the evidence of manuscript variation. Particular attention is devoted to interoperability, with the thesaurus ideally designed to interact with other projects, such as Franck Cinato's *Thesaurus Glossariorum Latinorum* (TGL), thereby fostering the integration of lexical resources and facilitating comparative research across traditions.

**Mariafrancesca Giuliani (CNR-OVI, Firenze)**, ***Le risorse del progetto MEDITA (Medieval Latin Documentation and Digital***

***Italo-Romance Lexicography. Integrated Resources for the New Historical and Etymological Lexicography)***

La comunicazione presenterà l'organizzazione, gli intenti, i metodi, gli strumenti tecnologici e i risultati del progetto *MEDITA – Medieval Latin Documentation and Digital Italo-Romance Lexicography. Integrated Resources for the New Historical and Etymological Lexicography*, promosso congiuntamente dall'Università di Napoli Federico II, dall'Università Ca' Foscari di Venezia e dall'Opera del Vocabolario Italiano (OVI-CNR, Firenze) e finanziato dal MUR con il bando PRIN 2022 (codice progetto: 20227CHRR, PI: Paolo Greco). Sorto in stretta relazione con i lavori di tre officine lessicografiche, il *Dizionario Storico-Etimologico Napoletano (DESN)*, il *Vocabolario Etimologico Veneziano (VEV)* e il *Tesoro della Lingua Italiana delle Origini (TLIO),* il progetto ha creato una base di conoscenza testuale di media latinità non letteraria di area italiana articolata in tre moduli: un ampio corpus di testi documentari e statutari disponibili in edizioni affidabili, un sottocorpus annotato – primo nucleo di una lemmatizzazione che mira a diventare estesa all'intero corpus –, e una piattaforma di consultazione integrata di glossari contemporanei che selezionano volgarismi, tecnicismi, prestiti e neoformazioni presenti in fonti di media latinità di area italiana. L'intera risorsa è stata strutturata con un assetto che non avvicina solo gli interessi di lessicografi e linguisti, ma anche di storici del medioevo e paleografi. L'intervento si soffermerà anche sul senso e gli orientamenti della procedura di lemmatizzazione, avviata con l'intento di tracciare lo spazio di tensione tra tradizione e innovazione in cui si colloca il vocabolario testuale.

**Richard Ashdowne and Thomas Wrobel (University of Oxford), *Making the* DMLBS *digital***

The *Dictionary of Medieval Latin from British Sources*, completed in 2013 and reprinted in 2018, covers the use of the Latin language in Britain and by Britons from the mid-sixth to the late sixteenth centuries. Plans for the dictionary first began to form in 1913, and publication

began in 1975 with a fascicule covering A and B. The processes of research, editing and publication were initially entirely manual and based on hundreds of thousands of handwritten slips of paper. During the long evolution of the project, all parts of the preparation came to be partly or wholly digital: this began in the 1980s and 1990s with the use of electronic text databases for research and ended with the implementation of digital editorial processes and the preparation of a dataset for online publication alongside the printed dictionary. There was considerable challenge in moving to digital methods of working while not slowing the publication schedule and, above all, while still processing the massive amounts of essential handwritten material that had been collected over the decades. Ultimately online publication was made possible through licensing the dataset to external providers: this solution was chosen because of the technical and financial impossibility of guaranteeing a sustainable future for an online interface after the conclusion of the funded project.

**Ana Gómez Rabal (CSIC, Barcelona), *From Text to Lemma: Lexicographical Work in the* Glossarium Mediae Latinitatis Cataloniae**

This paper presents the methodological framework underlying the *Glossarium Mediae Latinitatis Cataloniae* (*GMLC*), focusing on the process of constructing lexical knowledge from textual evidence. Within the context of the significant development of editions of Medieval Latin documents in Spain, the project has gradually evolved from a traditional lexicographical enterprise into an integrated system combining philology and digital technologies. At the core of this system lies the *Corpus Documentale Latinum Cataloniae* (*CODOLCAT*), an expanding textual database that collects, describes, and makes searchable the Latin sources of medieval Catalonia. On the basis of this corpus, the different stages of lexicographical work will be examined: from the selection and normalization of textual evidence to excerpting, from the contextual comparison of attestations to the

elaboration of lemmas and the reconstruction of their semantic history. Particular attention will be devoted to the digital tools that support these operations and to the ways in which they transform the relationship between corpus and lexicographical processing, fostering a dynamic interaction between textual data and dictionary-making. Finally, the paper will outline the role of the *GMLC* within the broader landscape of European Medieval Latin lexicography and its contribution to current reflections on the renewal of historical Latin dictionaries, showing how this experience may provide a methodological model transferable to other research contexts.

**Dominique Longrée (LASLA, Univ. de Liège), Laurent Vanni (BCL-UMR 7320 CNRS – Univ. Côte d'Azur), *Deep Learning vs statistical textual data analysis: how do they complement each other in the study of Latin texts? The example of Hyperbase and Hyperdeep***

For a few years, the software Hyperdeep has been used to detect "deep motifs" in the LASLA latin files. The main aim of this research was double: on the one side, we were looking for recurring "textual motifs" structuring and characterizing texts; on the other side, we were trying to detect "motifs" being traces of intertextuality. Based on an attribution task and a deconvolution process of both Convolutional Neural Network and Transformers, we searched inside the passages appearing as the most significative for the attribution task and mined "Deep motifs", various patterns on which the attribution task was based. In this paper, we will show how a statistical textual data analysis with Hyperbase, combined with a philological study, allowed us to sort, amongst these "Deep Motifs", various more complex patterns of already known "textual motifs", new "Textual Motifs" characteristic of the various authors and traces of intertextuality. We will also assess the individual performances of the CNN and of the Transformers for this task.

**Philipp Roelli (University of Zurich), *Future lexicographic applications of* Corpus Corporum**

*Corpus Corporum* (CC) is a free and open digital corpus and research platform dedicated primarily to Latin texts from Antiquity, the Middle Ages, and the Early Modern period. We can feed it with TEI XML encoded Latin text editions, which we take from many open sources. CC offers a number of search options as well as tools to identify text-reuse and metrical analysis of verses. The bibliographic data is linked to outside resources providing information about authors and texts. Currently we have approximately 226m Latin words (tokens) and 1.8m distinct ones (types) across thirty corpora. The texts are linked to important Latin dictionaries; currently twelve such dictionaries are integrated into the site. There is a total of 288k distinct lemma entries in these dictionaries. Besides we use a morphological word-form list we extracted from Wiktionary. The most recent version contains 1.9m Latin word-forms. Now we link CC vocabulary through the morphology list to the dictionary entries so that users can click words and see the dictionary entries. In this talk I discuss how well this currently works with concrete numbers and I highlight problem cases. Future tasks are to enrich our dictionary display with information extracted from the texts: direct links to cited passages, information about the earliest use, word statistics. As a long-term goal, it may become possible to enrich the dictionaries by automatically extracting lemmata unknown to the dictionaries from the text material. We also plan to introduce AI models to enable automated translation and semantic analysis.

**Andrea Balbo (Università di Torino), *Intelligenza artificiale e studi latini***

My contribution aims to provide an overview of the prospects for the use of artificial intelligence in Latin studies, focusing primarily on research perspectives, while also offering some remarks on its applications in teaching and third-mission activities.

**Barbara McGillivray (King's College London),** ***AI-Driven Approaches to Latin Lexical Semantics: the COALA Project***

This paper presents COALA (COmputational Corpus Annotation for Quantitative Analysis of LAtin Lexical Semantics), a five-year long ERC-selected project which brings together computational methods, corpus linguistics, and lexicography to study semantic variation and change across the full diachronic span of Latin, from the 7th century BCE to the 20th century CE. Latin's long lexicographic tradition provides rich qualitative evidence for semantic description but lacks the quantitative dimension needed for large-scale diachronic analysis. COALA addresses this gap by developing a large sense-annotated corpus of Latin texts designed to support both semasiological and onomasiological research questions: how individual word senses emerge and shift over time, and how conceptual domains are lexicalised across periods. COALA will contribute a replicable annotation framework, new empirical findings on Latin lexical change, and a foundation for future large-scale quantitative analyses. The project operates on two parallel streams, a curated manual annotation track for detailed analysis and a computational track for large-scale analyses. Sense inventories are drawn from multiple dictionary resources (Lewis & Short, Oxford Latin Dictionary, Thesaurus Linguae Latinae, Latin WordNet). I will present initial experiments employing large language models for word sense disambiguation. Results show strong performance with sense inventories that have clear, well-delineated sense boundaries, and highlight challenges posed by fine-grained or overlapping senses, particularly in WordNet-based inventories.

**Lucia Pinelli e Matteo Salvestrini (SISMEL, Firenze),** ***Intelligenza artificiale e Mirabile Web***

La prima parte dell'intervento riflette sul ruolo che le tecnologie di intelligenza artificiale possono assumere nella necessaria evoluzione

del portale Mirabile. Mirabile rappresenta un insieme complesso di archivi integrati, connotati ciascuno da caratteristiche diverse sia sul piano scientifico sia su quello tecnologico. Si presenta quindi, come primo approccio alle possibilità offerte dall'IA, una soluzione relativa al dialogo tra due archivi onomastici presenti in Mirabile e alla loro proiezione finale sul *Mirabile Atlas*. La seconda parte illustra nel dettaglio l'applicazione, articolata in cinque stadi: normalizzazione delle grafie secondo regole esplicite; calcolo della distanza di Levenshtein; validazione contestuale tramite modello linguistico, cui vengono fornite anche le datazioni come contromisura agli omonimi; verifica umana; ed export in GeoJSON dei soli abbinamenti approvati. Se ne mostrano i limiti: la similarità percentuale non è un fatto neutro — formule diverse indirizzano lo stesso caso a corsie decisionali diverse — e un modello linguistico può produrre errori *plausibili ma falsi*, concordi con un'alta somiglianza di stringa. Per questo ogni decisione è conservata e tracciata in un registro *append-only*: confidenza graduata, provenienza reversibile, responsabilità umana. La stessa architettura è riusabile per luoghi, opere e manoscritti, e verso le autorità esterne (VIAF, CERL Thesaurus, Wikidata, GeoNames).

**Elisa D'Argenio, Alessandro Papini (ERC DILADI, Università 'Ca Foscari Venezia), *Innovazioni lessicali nel database LLDB: il caso dei documenti toscani***

The ERC project DiLaDi (Digital Latin Dialectology; https://pric.unive.it/projects/diladi/home) investigates the variation of the Latin Language during the first millennium AD from both a diatopic and a diachronic perspective, on the basis of deviations from the norm of Classical Latin attested in non-literary sources, such as inscriptions and charters of private law. These deviations are recorded and classified in the LLDB database (https://lldb.elte.hu/en/database/), which makes it possible to query the data according to different linguistic and extralinguistic

parameters and to carry out both quantitative and qualitative analyses of the attested phenomena. The present paper focuses on the lexical innovations recorded in the LLDB database and found in eighth-century notarial documents from Tuscany. The analysis considers the elements classified as innovative with respect to Classical Latin, with particular attention to those labelled as *voces medii aevi*. This category includes both internal developments of Latin and lexemes of Celtic or Germanic origin. The study aims to assess whether the distribution of lexical innovations is influenced by factors such as chronology, place of issue, document type, or the scribe, or whether it depends above all on factors internal to the individual texts, including subject matter, the presence of specific formulae, or particular terminological needs. The analysis will therefore shed light on the extent to which the innovative lexicon of this corpus reflects broader processes of linguistic change or choices linked to the function of the texts and to writing

**Timo Korkiakangas (University of Helsinki), *Quanto possiamo fidarci dei corpora parserizzati? Misurare la classicità lessicale in testi mediolatini***

L'affidabilità dei corpora automaticamente annotati è una delle questioni più importanti per il linguista del latino che vorrebbe fare linguistica computazionale. Anche se i parser hanno compiuto notevoli progressi negli ultimi anni, rimane aperta la questione fino a che punto i dati così ottenuti possano essere utilizzati per lo studio del lessico latino (medievale).

Questo contributo esamina questo problema attraverso un case study dedicato alla misurazione della classicità lessicale. Il materiale analizzato è costituito dal corpus HistHag, che comprende 71 testi storiografici e agiografici prodotti in Italia e in Francia tra il VII e il XIII secolo (ca. 869.000 token). Come termine di confronto viene utilizzato il corpus LASLA, assunto come rappresentazione del latino classico.

L'obiettivo è valutare tre diverse misure di classicità lessicale: una basata sulla sovrapposizione lessicale con il LASLA, una fondata sui lemmi più caratteristici del corpus classico (keyness) e una basata su weighted log-ratio scores. I risultati vengono successivamente analizzati usando modelli di regressione che tengono conto della cronologia, del genere testuale, del luogo geografico e della lunghezza dei testi. L'analisi mostra che le tre misure convergono in modo significativo nell'identificazione dei testi meno classici. Inoltre, il genere testuale risulta il fattore più costantemente associato alla classicità lessicale. Nonostante numerosi errori di lemmatizzazione, le tendenze osservate a livello del corpus risultano relativamente stabili. I risultati suggeriscono quindi che i corpora medievali latini parserizzati possano costituire una base sufficientemente affidabile almeno per questo tipo di studio quantitativo delle distribuzioni lessicali.

**Michele Ciletti (Università di Foggia), *LLM-Driven, Open-Vocabulary Word Sense Disambiguation for Latin Digital Lexicography***

Word-Sense Disambiguation remains a central challenge in computational linguistics, and the task becomes particularly relevant for historical and low-resource languages, where comprehensive lexical resources are often incomplete or unavailable. This project introduces Inspicio, a language-agnostic, LLM-based framework for open-vocabulary disambiguation that bypasses predefined candidate lists by retrieving synsets from the Open English Wordnet, and evaluates it through a case study on Latin verbs. The approach investigates whether target words in a historical language can be mapped effectively to a large English-based semantic network, and how hybrid retrieval can manage the extreme granularity of Wordnet while preserving genuinely distinct meanings.

The pipeline begins by generating two English translations of each Latin target sentence, one literal and one natural. Using these

translations together with the original sentence, the model proposes candidate English substitute lemmas and dictionary-style glosses for the target word, ordered by likelihood and able to represent multiple literal or metaphorical interpretations. All verb synsets in the Open English Wordnet are embedded by combining glosses, examples, lemmas, hypernyms, and lexnames. Retrieval combines dense cosine similarity with sparse lemma matching, boosts synsets supported by multiple retrieval signals, and applies a soft Maximal Marginal Relevance filter to avoid burying alternative meanings beneath minor variants of the dominant sense. The system outputs a top-k candidate pool. Preliminary experiments on Latin perception verbs and selected LatinISE data show promising results, with Recall@20 reaching 90% under the most effective combination of LLM and embedding model. Future work will aim to add an LLM reranker, broaden evaluation beyond verbs, test smaller open-weight models, and adapt the framework to other historical languages.

**Giuseppe Ferrara (Università degli Studi di Siena), *(Quasi) sinonimi latini. Le distinzioni della tradizione lessicografica nello spazio semantico di LaBerta e Latin BERT***

La tradizione lessicografica ha classificato i (quasi) sinonimi latini attraverso l'analisi qualitativa dei loro contesti d'uso, senza disporre di una misura quantitativa dell'effettiva distanza semantico-distribuzionale tra di essi. In questo contributo, si indaga se la geometria degli *embedding* contestuali prodotti da modelli *Transformer* colga, ed eventualmente arricchisca, le distinzioni effettuate dagli studi lessicografici. L'analisi, condotta su un *corpus* di circa 7.900 testi, utilizza gli *embedding* estratti da due modelli, LaBerta (RoBERTa) e Latin BERT (BERT), relativi a 21 lemmi afferenti a cinque campi semantici (paura, ira, bellezza, grandezza e parlare) e a 13 lemmi non correlati impiegati come *baseline*.

I risultati mostrano che i raggruppamenti effettuati dalla tradizione lessicografica trovano un corrispettivo nello spazio degli *embedding*. La

similarità coseno media si rivela una metrica efficace per separare le coppie di lemmi dello stesso campo (sinonimi) da quelle di campi diversi (AUC-ROC = 0,976; F1 = 0,873). L'algoritmo di Louvain applicato a un grafo di prossimità lessicale recupera, autonomamente, quattro dei cinque campi semantici (NMI = 0,749). Il *Rocchio classifier*, utilizzato per verificare se i singoli (quasi) sinonimi siano distinguibili all'interno del loro campo semantico e abbiano sfumature effettivamente diverse, raggiunge un'accuratezza del 93,4-97,6%. L'analisi qualitativa dei profili distribuzionali dei lemmi mostra che le sfumature semantiche della maggior parte di essi corrispondono alle definizioni del Forcellini, di cui restano però opache le distinzioni di grado e di registro. La *Word Sense Induction*, infine, individua per alcuni lemmi una bipartizione tra usi classici e cristiano-teologici che arricchisce il quadro lessicografico tradizionale.

**Silvana Maiello (Università per Stranieri di Siena), *Reconstructing the Implicit: Grammar, Lexicon and Digital Analysis in Basilio Brollo's* Dictionarium sinico-latinum *(1694)***

This paper explores how the digitalisation of Basilio Brollo's Dictionarium sinico-latinum (1694) has transformed a large manuscript dictionary from a consultable philological source into a structured and searchable corpus, creating new opportunities for linguistic and lexicographical research. Developed within the PRIN PNRR 2022 project CHIN-DICTIONARY, the digital corpus enabled the systematic investigation of metalinguistic information embedded in the Latin definitions, revealing a largely implicit grammatical layer. The research adopts an inductive methodology based on data extraction and corpus interrogation. Through searches of recurring lexical roots, metalinguistic markers, and Chinese romanisations, it becomes possible to identify patterns that, while accessible through traditional manuscript consultation, would require considerably more time and effort to detect systematically. Rather than applying predefined categories, the analysis reconstructs thirteen

metalinguistic macro-categories emerging directly from the data, including classifiers, grammatical particles, and pronouns. Particular attention is devoted to the role of digital pattern detection. A case study centred on the recurrent romanisation jen^ demonstrates how corpus-wide searches made it possible to identify the corresponding character 然 (rán) and verify Brollo's description of its adverb-forming function across multiple entries. This example illustrates how digital methods facilitate the formulation and testing of philological and linguistic hypotheses, enabling large-scale analyses that would be feasible only with significant additional time through manuscript consultation alone. The paper argues that digitalisation does not replace philological interpretation; rather, it enhances it by accelerating data retrieval, exposing hidden connections, and enabling systematic analyses that substantially expand research possibilities on early modern Sino-European linguistic works.

**Elisabetta Bartoli-Filippo Costanti- Carmen Esteban Martínez-Giulio Quaresima (Università degli Studi di Siena),** ***Per una implementazione di ALIM con strumenti lessicografici digitali: ricerche su antroponimi e toponimi ed esperimenti sul lessico retorico***

This paper presents an ongoing collaborative project aimed at enhancing ALIM as a digital environment for targeted lexical, onomastic, and semantic research on medieval Latin texts. The first line of work stems from the UMAS project on the medieval universities of Arezzo and Siena, and from the need to map and index personal and place names in large documentary collections, especially the Aretine collection edited by Robert Black and the Chartularium Studi Senensis. A second line of work addresses the rhetorical lexicon of ALIM, aiming to isolate the rhetorical sense of terms also used in other semantic or disciplinary fields. The first case study concerns NER (Named Entity Recognition) in medieval Latin charters, with a focus on person and place names. Since NER in this field is hindered

by limited annotated data, formulaic variation, rich morphology, and free word order, we experiment with a graph-enhanced architecture combining contextual representations from Latin-BERT with syntactic dependency structures. A two-stage transfer-learning pipeline, based on pre-training on a larger corpus and fine-tuning on annotated charters, is used to evaluate whether the integration of syntactic information can reduce entity-boundary errors. The second case study addresses the contextual classification of rhetorical vocabulary in ALIM. Since word-sense induction remains an open problem in NLP, the experiment focuses on a restricted set of lexical items with both rhetorical and non-rhetorical senses.
The experiment starts with an unsupervised classification of those lexical items in context, generated using word-embeddings and clustering techniques, and then verifies the results against human-classified word-context pairs.

**Paola Mocella, Martina Paccara (Università degli Studi di Siena), *Building a Word List of Lexical Innovations in Asian Latin. AI-Assisted Experiments on Martino Martini's* De bello Tartarico Historia**

This paper presents an AI-assisted experiment aimed at building a word list of lexical innovations in Asian Latin, using Martino Martini's *De bello Tartarico historia* as a case study. Starting from the concept of the Lexical Innovation Rate, introduced by Francesco Stella and Martina Paccara in 2023 and defined as the ratio of software-unlemmatized forms to the total number of word types in a corpus, the study tests whether LLMs can refine the identification, classification, and interpretation of lexical phenomena arising from Latin texts composed in contact with Asian cultures. The experiment used ChatGPT to identify lexical items absent from major Classical, Medieval, and Neo-Latin dictionaries and to classify them as transliterations or Latinized forms, new formations, or semantic shifts. It was carried out on both TXT and TEI-XML versions of

Martini's text, and the outputs were compared with those produced by Collatinus, a traditional Latin lemmatizer. The final dataset contains 223 lemma-level entries, corresponding to a lexical innovation rate of 4.5%. Most items are Asian anthroponyms, toponyms, and geonyms, but the list also includes new formations and cases of resemanticization in administrative, political, religious, and cultural vocabulary. The study shows that AI can accelerate lexicographical research on Asian Latin, while confirming the need for philological verification.

## Tavola rotonda su *Il lessico botanico latino: da Columella all'*Hortulus *alla* Flora Sinensis

**Francesca Boldrer (Università di Macerata)**
Il lessico botanico latino rientra nel *sermo rusticus*, componente fondamentale della lingua latina. Un testimone autorevole è Columella, di cui si evidenziano i principali aspetti tecnico-linguistici con attenzione al libro X, *sui generis* nel suo trattato *De re rustica* per la forma poetica, sulle orme di Virgilio georgico, ma nondimeno specialistico e denso di terminologia botanica relativa a ortaggi, fiori, alberi e frutti, oltre a definizioni di attrezzi, concimi e attività agricole svolte dall'uomo. Ne risulta un vocabolario preciso e organizzato, che si presta ad approfondimenti semantici, anche nel confronto con Catone, Varrone, Virgilio e Palladio, e a trattamenti informatici.

**Elisa Petri**
Il presente contributo analizza brevemente il lessico botanico nel *De cultura hortorum* di Valafrido Strabone (IX secolo), evidenziando la duplice natura di un'opera tesa tra erudizione letteraria e conoscenza botanica. Attraverso l'esame di nodi significativi, quali l'eredità dei modelli classici (Virgilio e Columella), declinata sia sul piano formale sia nell'antropomorfismo espressivo; l'adozione di fitonimi greci legati

alla tradizione medica e l'uso di termini che denotano un realismo descrittivo, viene mostrata la sapiente versatilità dell'autore.

**Chiara Ombretta Tommasi (Università di Pisa)**
This paper examines the botanical vocabulary employed by Michał Boym in the *Flora Sinensis*, recently published in a critical edition by C. Tommasi and L. Businarolo (2025). Written in Latin, the work draws on the botanical lexicon of both Classical and Medieval traditions whenever suitable terms were available. More frequently, however, Boym resorts to lexical innovation, coining neologisms to describe previously unknown plants and products. These formations are often mediated through Portuguese, the principal linguistic channel of transmission. In several cases, the terms ultimately derive from Indian or Southeast Asian languages, though they invariably reached Boym's Latin via Portuguese intermediaries.

**Ilaria Bonini (Università degli Studi di Siena)**
The evolution of botanical vocabulary reflects our understanding of plants. Latin, which originated among Roman farmers, evolved to meet the challenges of the times: with Columella, it codified agricultural techniques; in the works of Strabo, it took on spiritual symbolism; finally, with the Flora Sinensis, it became a bridge to the exotic, inventing a new vocabulary. This journey demonstrates how a rural language transformed itself and, with Linnaeus, became the universal language of science, capable of giving a unique name to global biodiversity. Today, the International Code of Nomenclature governs the system and is updated every six years.

.