

Digitalizzare un dizionario oggi dall'acquisizione del testo alla pubblicazione: il caso studio del dizionario latino Lana 1978

Francesca Michelone

Università del Piemonte Orientale
francesca.michelone@gmail.com

Abstract

L'articolo ha l'obiettivo di mostrare attraverso un caso studio una possibile metodologia applicabile alla creazione di un dizionario digitale in accesso aperto a partire da un testo a stampa già esistente. In particolare, è esposta l'operazione di digitalizzazione del dizionario latino pubblicato da Italo Lana nel 1978 con l'editore Paravia. Dal dizionario del 1978 sono stati estratti i vocaboli afferenti all'ambito semantico della natura nel mondo antico. Il contenuto delle voci è stato arricchito tramite lo studio di opere della letteratura e il risultato delle riflessioni ha ricadute sulla codifica stessa del dizionario.

L'approccio al testo nasce dallo studio delle migliori pratiche attualmente diffuse nella lessicografia digitale e si propone di produrre dati in linea con i principi FAIR [20]. Tale metodologia è elaborata nel contesto specifico della lessicografia latina, ma i risultati ottenuti possono essere un esempio replicabile per chiunque affronti questo genere di operazione.

Parole chiave: Dizionari Digitali; TEI Lex-0; Dizionario bilingue latino-italiano; Natura nel mondo antico; Accesso Aperto; #AIUCD2024

The article aims to present, through a case study, a possible methodology for creating an open-access digital dictionary from an existing print edition. Specifically, it describes the digitization process of the Latin dictionary published by Italo Lana in 1978 by the Paravia publishing house. From this dictionary, entries related to the semantic domain of nature in the ancient world have been extracted. The content of the entries has been enriched through the study of literary works, and the reflections derived from this analysis are reflected in the dictionary's encoding.

The approach to the text is based on the study of best practices currently adopted in digital lexicography and seeks to align with the principles of FAIR data production [20]. While developed within the specific context of Latin lexicography, this methodology offers a replicable model for anyone undertaking a similar project.

Keywords: Digital Dictionaries; TEI Lex-0; Latin-Italian Bilingual Dictionary; Nature in the Ancient World; Open Access; #AIUCD2024

Un dizionario latino digitale

Il presente contributo intende analizzare e discutere la metodologia impiegata per la realizzazione di un *Dizionario Digitale Aperto*, concepito non solo come strumento di consultazione ma anche come un luogo di incontro e collaborazione per la comunità scientifica. Il progetto mira alla creazione di un dizionario ad accesso aperto in un ambiente digitale controllato che - tramite annotazione XML - consente ad utenti abilitati di modificare, aggiungere o integrare le informazioni preesistenti, promuovendo così un approccio dinamico e partecipativo alla costruzione del sapere.

L'oggetto dello studio è rappresentato dalla digitalizzazione del dizionario latino-italiano pubblicato da Italo Lana nel 1978, con particolare attenzione ai lemmi riconducibili all'ambito semantico di ciò oggi si indica con la parola "ambiente". L'obiettivo è creare un dizionario digitale ad accesso aperto focalizzato sul lessico utilizzato nella letteratura latina per descrivere l'ambiente naturale¹. Questa scelta tematica ha determinato l'organizzazione dei lemmi non solo in ordine alfabetico, ma anche la loro caratterizzazione secondo categorie concettuali che introducono un livello nuovo di lettura, permettendo di esplorare il testo attraverso connessioni semantiche e tematiche.

Il processo di digitalizzazione ha comportato una serie di fasi metodologiche, ciascuna delle quali è discussa nel dettaglio, evidenziando le criticità incontrate e le soluzioni adottate. Il flusso di lavoro comprende:

- acquisizione del testo tramite OCR: particolare attenzione è stata dedicata alla correzione degli errori di riconoscimento e alla validazione dei dati;
- selezione dei lemmi e codifica: sono stati selezionati i lemmi in base ad un'ontologia preesistente e la divisione dei lemmi in categorie è stata riproposta nella codifica costruita a partire dallo standard TEI LEX-0;
- restituzione digitale: riflessioni e soluzioni adottate per la pubblicazione del dizionario.

Ogni fase è stata progettata con l'intento di costruire una metodologia replicabile e scalabile per la digitalizzazione di dizionari e altri strumenti lessicografici. Il contributo non si limita a descrivere i risultati raggiunti, ma propone una riflessione critica sulle implicazioni metodologiche ed epistemologiche del lavoro, sottolineando l'importanza di un approccio collaborativo e interdisciplinare.

Nel seguente contributo non si affronta il tema della creazione di un dizionario nativamente digitale, tipologia che aprirebbe a riflessioni e strumenti diversi. Per questo scopo sono un riferimento gli strumenti del gruppo Elexis [16], un punto di partenza e confronto per la lessicografia digitale. Tra di essi è necessario menzionare *Elexifier*, lo strumento volto a trasformare dizionari digitali già esistenti secondo gli standard di *Elexis*. Il materiale preesistente può essere in formato .pdf o .xml, ed è possibile apportare modifiche ai dati acquisiti tramite *Lexonomy*, l'applicazione dedicata appunto alla scrittura dei dizionari.

Un esempio virtuoso per la creazione di un dizionario digitale a partire da un dizionario a stampa è dato dal progetto sul dizionario storico portoghese MOR Digital [2]. Per la sua creazione è utilizzata la codifica specifica TEI LEX-0, il cui uso è descritto nel dettaglio da Toma Tasovac sulla piattaforma DARIAH [14] ed è anche il riferimento per *Elexis*. La stessa codifica,

¹ Il progetto nasce all'interno del dottorato in Ecologia dei sistemi culturali e istituzionali, con una borsa PON a tematica GREEN, e si propone di esplorare le potenzialità della lessicografia digitale applicata allo studio del lessico ambientale nel mondo antico.

semplificata e applicata per macroaree [15], è utilizzata anche per il Grande Dizionario della Lingua Italiana² (GDLI).

Nel campo della lessicografia, focalizzando l'attenzione sulle risorse specifiche create per la lingua latina, un importante strumento è costituito dal progetto ERC LiLa-Linking Latin [6][9], conclusosi nell'estate 2023, che raggruppa e permette di ricercare in maniera avanzata più risorse lessicografiche legate al latino. Per il latino classico il dizionario bilingue più diffuso online è quello edito da Lewis-Short nel 1879, consultabile sia nella *Persaeus Digital Library* [3] sia in *Logeion* [7]. All'interno di quest'ultimo sono presenti anche dizionari dal latino al francese (Du Cange del 1887 e Gaffiot del 1934). Quelle citate sono tutte risorse ad accesso aperto, ma occorre ancora ricordare la versione digitale del *Thesaurus Linguae Latinae* (TLL) [18] consultabile *online* in due modalità: ad accesso aperto in formato .pdf scaricandolo dal sito del Thesaurus, oppure a pagamento nel sito dell'editore De Gruyter in una versione basata su annotazione XML-TEI. Il TLL non è solo il dizionario latino più completo attualmente disponibile, perché raccoglie tutti gli usi e le costruzioni di ogni singola parola, ma nei suoi articoli, in modo analogo a quanto accade per l'italiano con il GDLI, ne traccia anche una sorta di "biografia", a partire dalla comparsa nella lingua e dall'etimologia, fino all'uso nel latino tardo; si tratta quindi di una risorsa estremamente preziosa.

Da questa ricognizione emergono due aspetti: il primo è come gli strumenti lessicografici dedicati al latino siano un ambito vivo e in continuo sviluppo;³ il secondo consiste nell'assenza dell'edizione di un dizionario digitale ad accesso aperto latino-italiano relativamente *recente*. Il progetto presentato in questa sede si propone proprio di colmare questo vuoto partendo da un dizionario già esistente a stampa, appunto il Lana 1978 pubblicato dall'editore Paravia che dopo essere stato a catalogo per alcuni anni tornò nelle mani del suo autore e degli eredi che hanno dato il consenso alla digitalizzazione per finalità di studio e ricerca sotto licenza Creative Commons CC BY - NC - SA.

L'articolazione del testo rispecchia le tre fasi di lavoro citate:

- acquisizione e correzione del testo
- selezione dei lemmi e loro annotazione
- restituzione digitale.

Acquisizione e correzione del testo

L'idea alla base del flusso di lavoro è di sfruttare gli elementi formali già presenti nel dizionario come ancoraggi per la marcatura, in particolare i delimitatori di sezione, quali segni grafici e punteggiatura. L'applicazione di questo approccio richiede tuttavia un riconoscimento testuale estremamente preciso.

I programmi di OCR, siano essi a pagamento o ad accesso aperto, sono numerosi. La scelta del programma dipende dal materiale di partenza: nel caso del dizionario Lana 1978 si tratta di un testo bilingue a stampa su due colonne. Ad un primo sguardo, l'acquisizione tramite OCR di un testo moderno non dovrebbe presentare complicazioni perché la stampa è di buona qualità su

² <https://www.gdli.it>.

³ Si pensi anche al lavoro sulla metrica portato avanti da *Musisque Deoque* [10].

carta anch'essa di buona qualità. Nonostante ciò, il dizionario Lana 1978 ha richiesto le lavorazioni delle immagini descritte di seguito e l'utilizzo di un programma di OCR che permettesse un buon livello di personalizzazione: la presenza di caratteri speciali al suo interno e l'uso diffuso dei segni per indicare le quantità vocaliche creano la necessità di uno strumento specifico, come potrebbe accadere anche per dizionari che abbiano ad esempio al loro interno i caratteri dell'alfabeto fonetico.

Una significativa riduzione degli errori di riconoscimento è conseguita all'elaborazione delle immagini acquisite (Figura 1 e Figura 2): esse sono state ingrandite al 200% con filtro bicubico e ritagliate in modo da eliminare il margine. Questi accorgimenti apportati in collaborazione con lo sviluppatore Massimo Ghisalberti hanno facilitato l'acquisizione del testo.



Figura 1

<p>1. A, a, f. e n. indecl.: <i>A, a</i> (prima lettera dell'alfabeto latino) ● (come abbr.), <i>A</i> = <i>Aulus</i>; <i>A. o. a.</i> = <i>annus</i> (comun. anno: <i>a. tertio ab urbe conditā</i>, nell'anno terzo dalla fondazione di Roma); <i>a.</i> (nelle date) = <i>ante: a. diem IV Kalendas Maias</i>, il quarto giorno prima delle calende di maggio (= il 28 aprile).</p> <p>2. a, inter. (di dolore, di gioia, di meraviglia, ecc.): <i>ah!, oh!, ah!</i>.</p> <p>3. a, ab, abs, prep. con l'abl.: può indicare ● origine o provenienza: <i>da; da parte di; a M'</i>. <i>Tullio esse Cic</i> discendere da Manio Tullio; <i>a Marco tabellarius Cic</i> un corriere da parte di Marco; <i>nil (= nihil) tibi est a me periculi (= periculi)</i> TER non c'è alcun pericolo per te da parte mia ● moto da luogo: <i>da, via da;</i> (con nomi di città), <i>dalle vicinanze di; fuga ab Urbe Cic</i> fuga dalla città; <i>Ephesum ab Sardibus est profectus Liv</i> venendo da Sardi, parti per Efeso; <i>ab ostio Tanais Plin</i> a partire dalla foce del Tanai ● distanza: <i>da; a; positis castris a milibus passuum quindecim Ces</i> posto l'accampamento a quindici miglia (di distanza) ● allontanamento o separazione: <i>da; Praeneste ab Latinis ad Romanos descivit Liv</i> Preneste passò dai Latini ai Romani; <i>veri a falso distinctio Cic</i> la distinzione del vero dal falso ● privazione: <i>di; inops ab amicis Cic</i> povero di amici ● tempo: <i>da, a partire da; dopo; ab urbe conditā</i>, dalla fondazione della città; <i>ab adulescentiā</i>, fin dall'adolescenza; <i>secundus a Romulo Liv</i> secondo dopo Romolo ● agente: <i>da; respondit a civis se spoliari malle quam ab hoste venire QUINT</i> rispose che preferiva essere spogliato da un concittadino che venduto schiavo dal nemico ● causa: <i>per, a causa di; perit a morbo NEP</i> morì per una malattia ● relazione: <i>quanto a, di; a morbo valui, ab animo aeger fui Pl</i> sono stato bene fisicamente, malato nell'animo; <i>imparati cum a militibus tum a pecuniā Cic</i> imparati sia quanto a soldati, sia quanto a denaro; <i>philosophi ab ore honestissimi APUL</i> filosofi bellissimi d'aspetto ● paragone: <i>di Plin CRIST MARZ C; scleratus agens ab eis Vulg</i> agendo peggio di loro ● stare ab aliquo, stare dalla parte di qualcuno: <i>multi ab reo RET ER</i> molti difensori dell'imputato; <i>ab re</i>, fuor di luogo; contro l'interesse; <i>a tergo</i>, alle spalle; <i>a fronte</i>, di fronte; <i>servus a pedibus Cic</i> (a manu SVET) corriere (scrivano); <i>(libertus) ab epistulis (a rationibus) SVET</i> segretario (amministratore); <i>laborare ab aliquā re, petere ab aliquo ecc.</i>, — <i>laboro, peto ecc.</i></p> <p>Āaron e Āron, m. indecl.: <i>Aronne</i>.</p> <p>ab: — 3. a.</p> <p>abactio, abactiōnis, f.: <i>abactio pecorum</i> GR furto di</p>	<p>bestiame.</p> <p>ābactor, abactōris, m.: <i>ladro (di bestiame)</i>.</p> <p>1. ābactus, a, um, part. pf. di <i>abigo</i> agg.: <i>passato, trascorso; nox abacta VIRG</i> notte inoltrata.</p> <p>2. ābactus, ūs, m.: <i>sfratto</i> (?) (altri intende: <i>furto di bestiame</i>) PLIN G.</p> <p>ābācus, i, m.: <i>tavola, tavoletta</i> <i>credenza</i> CIC <i>tavoletta da giuoco SVET</i> <i>tavoletta per i calcoli PERS APUL</i> ● <i>pannello decorativo</i> VITR PLIN ● <i>abaco</i> (del capitello) VITR.</p> <p>abaestūmo NIG = <i>autumo</i>.</p> <p>ābaeto = <i>abito</i>.</p> <p>ābālienātio, abalienatiōnis, f.: <i>cessione; vendita</i> CIC ● <i>rinuncia, abbandono</i> CRIST.</p> <p>ābālienātus, a, um, part. pf. di <i>abalieno</i> agg.: <i>avverso, ostile</i> ● (t. med.), <i>paralizzato, privo di vita</i>.</p> <p>ābāliēno, as, āvi, ātum, āre, tr. I: <i>separare, allontanare, distogliere; illa a me abalienatur Pl</i> ella mi viene tolta ● <i>alienare, cedere, vendere; agros populi Romani abalienare Cic</i> alienare i terreni del popolo romano ● <i>inimicare, rendere ostile; totum se a te abalienavit Cic</i> ti è divenuto completamente ostile; <i>totam Africam abalienare NEP</i> spingere tutta l'Africa alla rivolta; <i>Campanos metu abalienavit Liv</i> spaventandoli, rese ostili i Campani ● <i>privare; abalienatus iure civium Liv</i> privato del diritto di cittadinanza ● (t. med.), <i>tramortire, paralizzare</i>. — anche <i>abalienatus</i>.</p> <p>ābāmīta, ae, f.: <i>sorella del trisavolo paterno</i> GIUR.</p> <p>ābantēus, a, um: <i>di Abante</i>.</p> <p>Ābantīādes, ae, m.: <i>Abantiade</i> (figlio o nipote di Abante).</p> <p>Ābas, Abantis, m.: <i>Abante</i>.</p> <p>ābāvīa, ae, f.: <i>trisavola</i> GIUR.</p> <p>ābāvūncūlus, i, m.: <i>fratello della trisavola</i> GIUR.</p> <p>ābāvus, i, m.: <i>trisavolo</i> <i>antenato</i>.</p> <p>abbas, abbātis, m.: <i>abate</i> CRIST.</p> <p>abbibo = <i>adbibo</i>.</p> <p>abbreviātio, abbreviatiōnis, f.: <i>riassunto</i> ● <i>riduzione</i>.</p> <p>abbreviō, as, āre, tr. I: <i>riassumere</i> <i>abbreviare</i> ● <i>indebolire</i>.</p> <p>Abdēra, ae, f. e Abdēra, ōrum, n. pl.: <i>Abdera</i> (nome di più città).</p> <p>Abdēritānus, a, um: <i>abderitano</i> (= stupido) MARZ.</p> <p>Abdērites, ae, m.: <i>abderita, di Abdera</i>.</p> <p>abdīcātio, abdīcātiōnis, f.: <i>abdicazione, rinuncia; abdicatio dictaturae</i> (gen.) Liv rinuncia alla dittatura; <i>abdicatio voluptatum TERT</i> rinuncia ai piaceri ● <i>ripudio, atto di diseredare</i> (un figlio).</p>
---	--

Figura 2

Quando il progetto ha avuto inizio nel 2022 i due software più adatti per l'acquisizione del testo di questo dizionario potevano essere *Transkribus* e *Tesseract*. Da un lato *Transkribus* [19] è una piattaforma nata per la trascrizione dei manoscritti di semplice utilizzo e negli ultimi anni con l'implementazione di processi di Machine Learning si prospetta sempre più avanzata e predisposta alle personalizzazioni dei modelli; la piattaforma prevede un sistema di crediti mensili gratuiti ai quali se ne possono aggiungere altri a pagamento per l'acquisizione del testo dalle immagini. Di fatto, per digitalizzare un dizionario sarebbe necessario versare un contributo per il numero di pagine e plausibilmente per una buona caratterizzazione del modello. Dall'altro lato *Tesseract* consente l'accesso al codice sorgente e, se si hanno le competenze, permette un livello di personalizzazione estremamente dettagliato. Per *Tesseract* sono disponibili dei file .traineddata per le diverse lingue ed è possibile modificarli secondo le esigenze. Nel caso del dizionario Lana si è optato per l'uso di *Tesseract* 5.3.4 abbinato ai .traineddata *Latin* (4.1.0) e *Latincustom* creati su misura da *Rescribe* Ltd. [13] in una prima fase del lavoro di digitalizzazione. Questa scelta è stata guidata dalla volontà di privilegiare, in tutte le fasi del processo di digitalizzazione, strumenti disponibili in accesso aperto e, ove possibile, open source. L'OCR prodotto con *Tesseract* nel modo indicato genera un output in formato .txt, che costituisce l'input della fase successiva che

avviene tramite il software *Dicnorm*⁴ prodotto dallo sviluppatore Ghisalberti. Questo programma agisce in varie fasi:

- 1) analisi del testo sorgente grezzo derivato dall'OCR;
- 2) generazione di un file normalizzato, applicando regole di sostituzione di caratteri o parole, individuazione degli autori in base a una tabella di sostituzioni⁵, riunione di linee spezzate;
- 3) analisi del file normalizzato e produzione dell'output, ancora in .txt.

Pare importante sottolineare come la seconda fase, tramite un file di configurazione .yml, consenta di procedere ad una correzione per la maggior parte automatica. Se da una parte il file .traineddata addestrato da *Rescribe* ha ridotto il numero di errori di riconoscimento, il punto di svolta è stato l'applicazione del software per la normalizzazione e l'eliminazione degli errori ricorrenti perché, benché la fase di revisione manuale non sia stata eliminata, è stata di molto alleggerita, dando così la possibilità di agire con più concentrazione sull'uniformità del dizionario ai fini della codifica.

La revisione manuale del testo si articola in due fasi principali. La prima fase consiste nella correzione degli errori formali residui e nell'uniformazione del testo, al fine di ottenere una corretta codifica del contenuto nel passaggio successivo. In questo contesto, i separatori di sezione del dizionario rappresentano un elemento cruciale: ogni macrosezione all'interno della voce deve essere delimitata in modo chiaro in tutto il dizionario per mezzo di tali simboli.⁶

Un aspetto problematico riscontrato durante la revisione riguarda la gestione delle varianti dei lemmi. Talvolta, le varianti compaiono all'interno della stessa voce, altre volte sono trattate come rimandi separati, oppure inserite nelle particolarità grammaticali alla fine della voce. Questa varietà di approcci ha complicato l'applicazione della codifica automatica, rendendo necessaria un'operazione di uniformazione il più possibile estesa. Per facilitare sia la ricercabilità del testo sia l'automazione della codifica, si è deciso di trattare le varianti al nominativo come lemmi separati, corredandole di un rimando alla voce principale. Interventi manuali sono stati necessari anche nei casi in cui, accanto al nominativo, venivano riportate particolarità grammaticali. In tali situazioni, le particolarità sono state spostate nella sezione dedicata alla fine del lemma, per facilitare la codifica e mantenere una coerenza interna.

Nella seconda fase della revisione manuale sono invece state inserite le integrazioni al dizionario in modo che potessero essere successivamente codificate in TEI contestualmente al resto. In particolare, è stato utilizzato un sistema di parentesi quadre per indicare dove poi sono stati inseriti tag annidati più o meno complessi, come l'inserimento di link per la referenziazione esterna, o alcune informazioni sul testo inserito.⁷

⁴ <https://gitlab.com/minimalprocedure/dicnorm>.

⁵ Per esempio, la sostituzione del gruppo “bp” con “öp” oppure di “Ď” con “ö”.

⁶ Le sezioni all'interno della voce sono divise da un pallino (●) o dalla doppia stanghetta (| |): nel primo caso si prosegue con un significato indipendente da quello precedente, nel secondo, invece, si tratta di un senso derivato. Gli usi proverbiali sono preceduti da un quadratino (■), quelli cristiani da una croce (+), le particolarità grammaticali da una losanga (◇). I simboli uguale (=) e freccia (→) introducono invece la referenziazione ad un'altra voce del dizionario.

⁷ Si può vedere un esempio pratico sotto *abellana* nel paragrafo “Integrazioni al dizionario”.

Da *Dicnorm* vengono prodotti diversi file intermedi per un'eventuale ulteriore revisione manuale e i file annotati in .xml la cui codifica è descritta nel paragrafo successivo.

Poiché il primo esito della digitalizzazione del *Lana* 1978 consisterà in un dizionario specialistico dei vocaboli latini impiegati per descrivere la natura e l'ambiente - denominato *Lexicon Orbis Terrarum* -, al lavoro di correzione si è affiancata l'attività di selezione dei lemmi. Trattandosi di un'operazione specifica di questo progetto, in questa sede ci si limiterà a dire che la selezione ha riguardato, in linea generale, le voci del *Lana* 1978 riconducibili alle quattro categorie (cielo e atmosfera, superficie terrestre, piante, animali) della sezione *Universo* del sistema concettuale elaborato da Hallig e von Wartburg (1963) [5], con alcune limitazioni e adattamenti pensati in funzione del contesto del mondo classico. In questo senso, l'idea di un dizionario digitale risulta particolarmente efficace perché rimane potenzialmente aperto non solo all'aggiornamento, ma anche alla discussione teorica sulle categorie che regolano la nostra lettura del tema ambientale nell'antichità. I vocaboli selezionati non hanno quindi lo scopo di esaurire il discorso sull'ambiente nel mondo antico, ma di facilitare la ricerca in questo settore.

Codifica

Perché il dizionario sia effettivamente digitale è necessaria una codifica del testo dopo che esso sia stato acquisito e corretto. La codifica condivisa per i dizionari digitali è, come anticipato, quella nota con il nome di TEI Lex-0, una selezione pensata di XML-TEI che stabilisce uno *standard* per l'uso coerente dei marcatori nei dizionari. L'elemento critico che emerge, come però già nella fase precedente, è la mole di materiale da annotare. Si tratta di un'operazione difficilmente gestibile manualmente, e si è dunque cercato un sistema che permettesse una codifica automatica del testo. A questo scopo, è stata valutata la possibilità di utilizzare GROBID [8], "a machine learning library for extracting, parsing and re-structuring raw documents such as PDF into structured XML/TEI encoded documents". Tuttavia, l'utilizzo di questo software non è stato implementato, in quanto richiede un processo di addestramento su un campione di pagine annotate secondo le modalità desiderate, affinché l'annotazione possa poi essere applicata all'intero documento. Sebbene si tratti di uno strumento open source corredato da una documentazione ampia, il suo utilizzo efficace richiede tempo sia per acquisire competenze specifiche sia per addestrare il modello. Dalla collaborazione con Ghisalberti, l'informatico già citato, è nata una strategia alternativa: l'inserimento di tag e marcatori in modo semiautomatico tramite l'utilizzo di script che operano in presenza di specifici elementi del testo. L'impiego di *Dicnorm* anche per la codifica si è rivelato una scelta coerente con la fase precedente e maggiormente in linea con gli obiettivi e le tempistiche del progetto. La codifica realizzata in questo modo, pur non consentendo di restituire tutta la complessità dello standard TEI Lex-0, si è rivelata utile per produrre una prima annotazione di base, che potrà essere arricchita in un secondo momento, qualora necessario. Qui di seguito si propone un lemma campione annotato con le indicazioni base di TEI Lex-0 applicate in modo automatico con *Dicnorm*.

Ai fini espositivi si riporta di seguito la voce del dizionario nella sua forma testuale e la codifica divisa in sezioni.

aether, aethēris, m.: ètere (regione superiore del cielo) || cielo; sublatus ad aethera clamor [Virg] le grida salgono fino al cielo • (poet.), aria; liquidum trans aethera vectae [Virg] volando attraverso l'aria limpida • (poet.), regione (del cielo o della terra) ◇ (acc. aethēra); gen. anche aethēros; acc. anche aethērem.

Lemma e informazioni grammaticali

```
<entry xml:lang="la" xml:id="L.cm5o6a8jy001so3lauckuljzr">
  <lbl ana="head">aether</lbl>
  <form type="taxonomy">
    <lbl type="category">cielo_atmosfera</lbl>
  </form>
  <form type="lemma">
    <orth xml:lang="la">aether</orth>
    <orth xml:lang="la">aethëris</orth>
  </form>
  <gramGrp>
    <gram>m.</gram>
  </gramGrp>
```

entry è il tag per indicare l'unità base del dizionario e contiene l'intera voce. Al suo interno presenta due attributi la cui presenza è obbligatoria:

- **@xml:id** che permette il riconoscimento univoco del lemma;
- **@xml:lang** che indica invece la lingua di appartenenza, in questo caso il latino.

Il tag **form** contiene a sua volta **orth** e l'attributo **@type=lemma** che servono a dare la forma ortograficamente corretta del lemma. TEI Lex-0 raccomanda l'uso del tag **form** con **@type=inflected** o **@type=paradigm** per sostantivi, aggettivi e verbi che subito dopo la forma ortograficamente corretta presentino una o più forme flesse. In una codifica automatica però a causa della varietà delle forme e delle numerose irregolarità presenti non è stato possibile avere questo dettaglio di codifica. Si è dunque optato per una versione semplificata, comunque compatibile e che distinguesse le varie forme, senza però specificarle. Un ragionamento simile è stato applicato alle varianti che si trovano alla fine della voce.

Nel gruppo racchiuso dal marcatore **GramGroup** sono contenute le informazioni grammaticali. In questo punto dovrebbe essere inserita la parte del discorso (POS), ma nel dizionario latino digitalizzato essa è spesso sottintesa perché evidente, ad esempio, dal **paradigma** e l'indicazione esplicita dovrebbe essere inserita manualmente. Sono state quindi inserite qui in modo automatico le informazioni relative al gruppo grammaticale di appartenenza sulla base del fatto che dopo la forma base del lemma, nel dizionario vengono date le informazioni grammaticali e quindi esse sono annotate automaticamente.

Si è poi ipotizzato di affinare la codifica in un secondo passaggio rendendola conforme *in toto* a TEI Lex-0 mediante l'utilizzo di strumenti basati sull'intelligenza artificiale applicata all'analisi linguistica. In questa direzione sono solo stati fatti però alcuni esperimenti su forme verbali, sostantivi e aggettivi. L'esito è positivo e sembra promettente, ma il campione non è ancora significativo per dare risultati certi.

A scopo esemplificativo si propone la codifica così rielaborata come descritto:

```
<entry xml:lang="la" xml:id="L.cm5o6a8jy001so3lauckuljzr">
  <lbl ana="head">aether</lbl>
  <form type="lemma">
    <orth>aether</orth>
    <orth>aethëris</orth>
  </form>
  <gramGrp>
    <gram type="pos">noun</gram>
    <gram type="gender">masculine</gram>
    <gram type="declension">3rd</gram>
  </gramGrp>
  <form type="inflected">
    <gramGrp>
      <gram type="case">nominative</gram>
      <gram type="number">singular</gram>
    </gramGrp>
    <orth>aether</orth>
  </form>
  <form type="inflected">
    <gramGrp>
      <gram type="case">genitive</gram>
      <gram type="number">singular</gram>
    </gramGrp>
  </form>
```

Traducenti ed esempi

```
<sense xml:id="L.cm5o6a8jy001so3lauckuljzr.nor">
  <cit type="translation" xml:lang="it">
    <form>
      <orth>ètere (regione superiore del cielo)</orth>
    </form>
```

```
</cit>
<cit type="translation" xml:lang="it">
  <form>
    <orth>cielo</orth>
  </form>
</cit>
<cit resp="Lana">
  <quote xml:lang="la">sublatus ad aethera clamor</quote>
  <quote xml:lang="it">le grida salgono fino al cielo</quote>
  <bibl>
    <author ref="phi0690">
      <persName type="usualname">Publius Vergilius
Maro</persName>
      <persName type="abbrname">Virg</persName>
      <date from="-70" to="-19"/>
    </author>
    <title xml:lang="la"></title>
    <publisher></publisher>
    <date></date>
    <biblScope></biblScope>
  </bibl>
</cit>
<note motivation="editing">
  <span>ètere (regione superiore del cielo) || cielo</span>
  <span>sublatus ad aethera clamor [Virg] le grida salgono fino
al cielo</span>
</note>
</sense>
<sense xml:id="L.cm5o6a8jy001so3lauckuljzr.acc">
  <cit type="translation" xml:lang="it">
    <form>
      <orth>(poet.), aria</orth>
```

```

    </form>
  </cit>

  <cit resp="Lana">
    <quote xml:lang="la">liquidum trans aethera vectae</quote>
    <quote xml:lang="it">volando attraverso l'aria limpida</quote>
    <bibl>
      <author ref="phi0690">
        <persName type="usualname">Publius Vergilius
Maro</persName>
        <persName type="abbrname">Virg</persName>
        <date from="-70" to="-19"/>
      </author>
      <title xml:lang="la"></title>
      <publisher></publisher>
      <date></date>
      <biblScope></biblScope>
    </bibl>
  </cit>
  <note motivation="editing">
    <span>(poet.), aria</span>
    <span>liquidum trans aethera vectae [Virg] volando attraverso
l'aria limpida</span>
  </note>
</sense>

```

Un altro tag obbligatorio in TEI Lex-0 è **sense**, che viene utilizzato per definire il significato o dare una traduzione. Ogni **sense** ha un suo **@xml:id** univoco e generato automaticamente.

Sense prevede l'utilizzo del tag **cit** che a sua volta può contenere la traduzione. In questo caso, come nelle forme esemplificate, è presente l'attributo **@type="translation"** e il riferimento alla lingua della traduzione, in questo caso l'italiano. Tale tag include poi i già citati **form** e **orth** in cui è indicata la forma ortograficamente corretta della traduzione. Come nell'esempio possono esserci più **sense** all'interno di una voce. Nel caso si inserisca un esempio d'uso, accanto al tag **cit** è indicato l'attributo **@resp** che nelle citazioni identifica chi ha inserito e tradotto una determinata frase (automaticamente è inserita per tutti i lemmi la responsabilità all'autore originario, Italo Lana). Questo è pensato nella prospettiva di un dizionario aperto collaborativo

in cui i contributi dei singoli non si dissolvono nell'insieme ma rimangono attribuiti, visibili, selezionabili e potenzialmente pubblicabili in Linked Data.

Ogni autore antico citato è stato associato a un codice di identificazione univoco, uniformato, ove possibile, ai codici assegnati dal Packard Humanities Institute (PHI) nel loro database, per garantire una maggiore interoperabilità. Tali codici sono adottati anche da piattaforme di riferimento come Logeion e Perseus. L'estensione diacronica degli autori citati nel dizionario Lana 1978, però, si protende anche nel periodo tardoantico e include gli autori cristiani; due categorie non presenti all'interno del canone del PHI. Gli autori latini tardi in prosa sono stati identificati tramite i codici attribuiti dal progetto DigilibLT [4], mentre per gli autori cristiani e per i poeti tardoantichi citati, per il momento, è stata generata una stringa alfanumerica univoca in modo casuale affinché gli sia abbinato un codice identificativo interno. Per ciascun autore sono state riportate, ove disponibili, indicazioni cronologiche utili a una più precisa identificazione. Inoltre, il nome completo degli autori è stato mantenuto in lingua latina per preservare la coerenza con le convenzioni bibliografiche della tradizione classica. Gli autori citati sono inseriti all'interno del tag **author** come descritto, ma quest'ultimo è a sua volta contenuto nel tag **bibl**, nel quale possono essere annotate informazioni sulle opere specifiche. In questo momento tali informazioni non sono presenti, ma la codifica è predisposta in modo da poterle eventualmente aggiungere in un secondo momento.

Dal momento che questa parte dell'annotazione avviene in modo automatico si è deciso di mantenere la forma originaria del testo alla fine di ogni **sense**, collocata in un tag **note** con attributo **@editing**. Questo rimane per utilizzo interno e di correzione per ricostruire i passaggi più agevolmente nel caso si rilevi qualche errore.

Particolarità grammaticali

```
<sense xml:id="L.cm5o6a8jy001so3lauckuljzr.pgr">
  <metamark function="pgr"></metamark>
  <form type="variant">
    <orth xml:lang="la">(acc. aethëra)</orth>
  </form>
  <form type="variant">
    <orth xml:lang="la">gen. anche aethëros</orth>
  </form>
  <form type="variant">
    <orth xml:lang="la">acc. anche aethërem</orth>
  </form>
  <note motivation="editing">
    <span>(acc. aethëra)</span>
    <span>gen. anche aethëros</span>
    <span>acc. anche aethërem</span>
  </note>
</sense>
```

```

</note>
</sense>
<sense xml:id="L.cm5o6a8jy001so3lauckuljzr.see">
  <metamark function="see"></metamark>
</sense>
<sense xml:id="L.cm5o6a8jy001so3lauckuljzr.equ">
  <metamark function="equ"></metamark>
</sense>

```

Alla fine della voce, prima della chiusura del tag **entry** – che contiene la voce di dizionario - è stato poi predisposto uno spazio all'interno del quale sono rilevati eventuali richiami ad altre voci del dizionario o particolarità grammaticali, come in questo caso. L'inserimento automatico si basa sulla presenza di segni caratterizzanti, preesistenti nel dizionario, dai quali viene riconosciuta una sezione specifica: la losanga (◊) precede delle notazioni grammaticali, i simboli uguale (=) e freccia (→) corrispondono a rimandi a voci equivalenti e derivate e, infine, il quadrato nero (■) precede utilizzi proverbiali. Tutte queste sezioni avrebbero una codifica specifica e costruita in TEI Lex-0, ma per favorire l'automatizzazione del processo sono stati inseriti dei tag **sense** che delimitano le sezioni che poi si sviluppano ognuna nella sua specificità. Nell'esempio presentato, la codifica è simile a quella della parte grammaticale, ma in questi casi è utilizzato il tag **form** con attributo **@type** che ha però come valore al suo interno "**variant**", appunto per indicare una variante.

Integrazioni al dizionario

La digitalizzazione del dizionario Lana 1978 si è accompagnata allo studio del lessico utilizzato nei testi latini per descrivere la natura. Il risultato di questa ricerca esemplifica - con il caso studio delle piante - come un dizionario digitale può essere ampliato, assolvendo anche a nuove funzioni rispetto al riferimento cartaceo. Si presenta di seguito un altro esempio, contenente l'annotazione della voce "abellana" al fine di mettere in evidenza le integrazioni rispetto al dizionario originale e la loro codifica.

äbellāna, ac, f.: nocciola [ref|target:
<https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:295446-1>]
 [metamark:{identificazione di André (2010,1)}]; nuces abellanae capitis dolorem faciunt et
 inflationem stomachi, corpori et pinguetudinis conferunt plus quam sitveri simile [Plin-M] Le
 nocciole causano mal di testa e gonfiore dello stomaco, e contribuiscono ad aumentare il peso
 corporeo più di quanto sembri verosimile • nocciolo (albero) [Plin].

```

<entry xml:lang="la" xml:id="L.cm68587ha00010dlaop6nh1yu">
  <lbl ana="head">äbellāna</lbl>
  <form type="taxonomy">
    <lbl type="category">vegetali</lbl>
  </form>
  <form type="lemma">
    <orth xml:lang="la">äbellāna</orth>

```

```
<orth xml:lang="la">ae</orth>
</form>
<gramGrp>
  <gram>f.</gram>
</gramGrp>

<sense xml:id="L.cm68587ha00010dlaop6nh1yu.nor">
  <metamark function="acc"> identificazione di André
(2010,1)</metamark>
  <ref
target="https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:
295446-1"/>
  <cit type="translation" xml:lang="it">
    <form>
      <orth>nocciola</orth>
    </form>
  </cit>
  <cit resp="Michelone">
    <quote xml:lang="la">nuces abellanae capitis dolorem faciunt
et inflationem stomachi, corpori et pinguetudinis conferunt plus
quam sitveri simile</quote>
    <quote xml:lang="it">Le nocciole causano mal di testa e
gonfiore dello stomaco, e contribuiscono ad aumentare il peso
corporeo più di quanto sembri verosimile</quote>
  <bibl>
    <author ref="phi0978">
      <persName type="usualname">Gaius Plinius
Secundus</persName>
      <persName type="abbrname">Plin</persName>
      <date from="23" to="79"/>
    </author>
    <title xml:lang="la"></title>
    <publisher></publisher>
    <date></date>
```

```

    <biblScope></biblScope>
  </bibl>
</cit>
<note motivation="editing">
  <span>nocciola [ref|target:
https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:295446-
1]</span>
  <span>nuces abellanae capitis dolorem faciunt et inflationem
stomachi, corpori et pinguetudinis conferunt plus quam sitveri simile
[Plin] Le nocciole causano mal di testa e gonfiore dello stomaco, e
contribuiscono ad aumentare il peso corporeo più di quanto sembri
verosimile</span>
</note>
</sense>

<sense xml:id="L.cm68587ha00010dlaop6nh1yu.acc">
  <cit resp="Lana">
    <cit type="translation" xml:lang="it">
      <quote xml:lang="it">nocciòlo (albero)</quote>
    </cit>
  <bibl>
    <author ref="phi0978">
      <persName type="usualname">Gaius Plinius
Secundus</persName>
      <persName type="abbrname">Plin</persName>
      <date from="23" to="79"/>
    </author>
    <title xml:lang="la"></title>
    <publisher></publisher>
    <date></date>
    <biblScope></biblScope>
  </bibl>
</cit>
<note motivation="editing">
  <span>nocciòlo (albero) [Plin]</span>

```

```
</note>
</sense>

<sense xml:id="L.cm68587ha00010dlaop6nh1yu.loc">
  <metamark function="loc"></metamark>
</sense>

<sense xml:id="L.cm68587ha00010dlaop6nh1yu.pgr">
  <metamark function="pgr"></metamark>
</sense>

<sense xml:id="L.cm68587ha00010dlaop6nh1yu.see">
  <metamark function="see"></metamark>
</sense>

<sense xml:id="L.cm68587ha00010dlaop6nh1yu.equ">
  <metamark function="equ"></metamark>
</sense>
</entry>
```

Due attributi previsti dallo standard, sono utilizzati in modo originale. Il primo è **@type=category** che colloca il lemma all'interno di una o più categorie semantiche, nel caso dell'esempio "piante". Il presente dizionario segue l'ontologia di Hallig-Wartburg (1963) [5] che fornisce uno schema gerarchico di concetti e un vocabolario controllato; essa è stata trasportata in OWL e funge come punto di riferimento per l'analisi di testi storici [17]. Per questo dizionario si è fatto riferimento alla sua parte dedicata all'universo. Le voci del dizionario Lana 1978 sono state suddivise in quattro categorie assegnate durante la fase di selezione e annotate automaticamente: cielo e atmosfera, superficie terrestre, piante, animali.⁸ L'inserimento di queste o altre categorie in un dizionario permette di implementare in una fase successiva una ricerca mirata e un livello di lettura ulteriore. Questo elemento, dopo aver portato nella dimensione digitale i lemmi tramite la codifica, arricchisce il dizionario aprendolo all'interoperabilità con altri strumenti del web semantico.

Il secondo è **@resp**, che nell'esempio indica il nome di chi ha inserito quella determinata citazione. Tramite questo attributo è possibile identificare chi arricchisce il dizionario con nuovi esempi, come in questa sede accade per i nomi delle piante o degli animali, che, per questioni di spazio e obiettivi nel dizionario a stampa, non erano particolarmente approfonditi. Si è ritenuto che in un dizionario digitale si potesse dare più spazio agli esempi per chiarire l'uso e il significato

⁸ Inizialmente le categorie erano cinque e includevano anche "uomo" inteso come essere umano. La presenza di quest'ultimo però comportava problemi nell'attribuzione delle categorie stesse e in particolare interferiva con quella "animale". La concezione dell'ontologia in un'epoca diversa ne ha determinato una visione antropocentrica, forse vicina a quella dei testi antichi, ma sempre più distante da quella moderna.

delle parole legate alla natura e proprio per mezzo di @resp è possibile arricchire il dizionario attribuendo le responsabilità degli interventi. In questo caso specifico le nuove citazioni sono state integrate a partire da ricerche tra i testi latini del PHI tramite Diogenes e, fin dove possibile, dagli articoli pubblicati dal Thesaurus Linguae Latinae. Come però è stato mostrato nel lemma campione, lo stesso ragionamento può essere ampliato per analogia ad infinite altre casistiche e interessi di studio. Il dizionario diventa così un luogo di confronto e un punto di incontro e va oltre alla sua funzione originaria di consultazione.

In linea con il focus ambientale del progetto di dottorato, è stato approfondito lo studio di specifiche piante nei testi latini. Laddove possibile, le piante sono state identificate mediante il confronto con il lessico edito da André nel 2010 [1] e all'interno delle relative voci sono stati inseriti i collegamenti al database Plants of the World Online (POWO) [11] mediante il tag di referenziazione esterna <ref> con attributo @target. Questo collegamento veicola ulteriori informazioni sul vocabolo, facilita la comprensione dei lemmi relativi alle piante per i lettori non esperti in botanica e può aprire nuove prospettive per l'analisi e l'interpretazione dei testi. In altre parole, la ricerca su questi vocaboli mira a restituire e valorizzare la specificità dei nomi, delle immagini e delle funzioni attribuite alla natura nei testi antichi.⁹

Dove però sono inserite nuove informazioni si è reso necessario prevedere la presenza di note alle identificazioni o riferimenti bibliografici. A questo scopo è stato utilizzato il tag **metamark**, già previsto nella codifica automatica per esprimere informazioni aggiuntive su simboli specifici come quelli legati alle sezioni particolari. Il tag **note**, che sarebbe più indicato a questo scopo, è già impiegato per contenere il testo ai fini di revisione editoriale e quindi si creerebbe confusione. L'uso di **metamark** per annotare meta-informazioni sul testo non è da manuale, ma ha permesso di inserire agevolmente le informazioni bibliografiche che potranno eventualmente esser rielaborate in un secondo momento.

Restituzione digitale

Il progetto mira a rendere disponibile il lavoro svolto in una forma versatile, in grado di rispondere alle esigenze di una vasta gamma di utenti. La visualizzazione dei contenuti prevede la trasformazione dei file XML in formato Markdown, da cui è possibile generare output in .pdf, .txt ed .epub. Questa varietà di formati è stata progettata per soddisfare le necessità di diverse categorie di utilizzatori del dizionario: la versione stampata è rivolta principalmente a scopi divulgativi o didattici, mentre il dizionario digitale è concepito come un ambiente dinamico, non solo per la consultazione, ma anche per il lavoro collaborativo tra studiosi.

In collaborazione con la casa editrice *InRiga*, è in fase di sviluppo un ambiente cloud per l'archiviazione delle versioni aggiornate del dizionario, associato a un sistema di stampa on demand. Questo approccio non solo garantisce che ogni copia stampata corrisponda sempre all'ultima versione disponibile, ma consente anche una gestione sostenibile della stampa. Inoltre, eventuali errori, inevitabili nella digitalizzazione di un progetto complesso come un dizionario,

⁹ Incrociare questi dati, ad esempio tra piante e animali, aprirebbe nuove possibilità per l'analisi di testi medico-farmacologici, in cui frequentemente si trovano lunghe liste di ingredienti che, senza strumenti adeguati, rischiano di restare nomi opachi e isolati. In riferimento a questo si riporta il pensiero di Rackham (1996, 17) "Plants and animals are not a generalized nature, not the passive recipients of whatever mankind chooses to inflict on them: they are thousands of individual species, each with its own behaviour which has to be understood" [12].

potranno essere segnalati attraverso la piattaforma DigilibLT e corretti rapidamente, con aggiornamenti tempestivi sia nel formato digitale che in quello a stampa.

Per quanto riguarda la pubblicazione digitale, il dizionario sarà ospitato sul sito di DigilibLT, in una forma interoperabile e integrata con la biblioteca digitale. Parallelamente, una versione consultabile sarà disponibile su Logeion, una piattaforma di grande visibilità che contribuirà a migliorare l'accessibilità e la diffusione del dizionario. La presenza del dizionario su queste piattaforme, entrambe sostenute da istituzioni universitarie, rappresenta un elemento chiave per garantire la sua conservazione e fruibilità nel tempo.

Conclusioni

Questo contributo illustra un metodo di lavoro replicabile per la digitalizzazione di dizionari, fondato sui principi FAIR (Findable, Accessible, Interoperable, Reusable). Il metodo proposto mira a coniugare la specificità del testo di partenza con l'aderenza agli standard condivisi, garantendo al contempo flessibilità e sostenibilità per progetti futuri.

Il dizionario digitale rappresenta una risposta concreta alla lacuna presente nelle risorse lessicografiche digitali per il latino. Sebbene la tradizione dei dizionari bilingue latino-italiano sia ricca, essa non trova ancora un'adeguata rappresentazione nei progetti digitali esistenti. Questo progetto, nato nell'ambito di una ricerca dottorale, offre un primo nucleo di una risorsa accessibile e versatile, capace di rispondere a esigenze diverse: dalla consultazione per scopi didattici alla ricerca accademica. Le diverse modalità di fruizione – dalla versione digitale a quella stampata on demand – sono pensate per arrivare a diversi tipi di utenza, con un'attenzione particolare agli studenti, per i quali i dizionari in inglese disponibili ad accesso aperto possono rappresentare una barriera linguistica.

Tra le peculiarità del dizionario digitale spicca il suo carattere aperto e collaborativo. La sua pubblicazione su due piattaforme istituzionali ad accesso aperto – DigilibLT e Logeion – ne garantirà al tempo stesso una grande visibilità e una lunga conservazione. Logeion, con la sua ampia diffusione internazionale, contribuirà a valorizzare il progetto presso una vasta comunità di utenti, mentre l'ambiente digitale su DigilibLT consentirà un arricchimento progressivo del contenuto, favorendo la connessione con la biblioteca digitale e incoraggiando il riutilizzo costante delle risorse. Inoltre, l'adozione di un linguaggio di codifica standardizzato, unitamente all'uso dei codici di autore PHI, facilita una maggiore interoperabilità con altre risorse digitali.

La possibilità di arricchimenti progressivi esplicita anche un'altra caratteristica di un dizionario digitale: la sua natura dinamica che permette di ampliarlo in tutte le sue parti (lemmi, traduzioni, esempi, traduzioni di esempi, etc.) ne fa uno strumento vivo, capace di evolversi nel tempo e di rispondere a nuove esigenze.

Per il futuro, si può prevedere una codifica estesa dell'intero dizionario, resa possibile anche dal supporto di strumenti di intelligenza artificiale, che potranno migliorare l'annotazione delle parti grammaticali e offrire ulteriori livelli di analisi. Questo approccio non solo potrà rendere il dizionario una risorsa più ricca, ma permetterà di integrare nuove funzionalità e di portare avanti nuove prospettive interdisciplinari.

Creare dizionari è tradizionalmente un lavoro collettivo, e il digitale esplicita e amplifica questa natura. L'idea di un ambiente digitale di lavoro, in cui sia possibile operare in modo supervisionato e con un'attribuzione chiara della responsabilità sui lemmi, trasporta quella collaborazione anche nella lessicografia digitale e la esplicita.

Bibliografia

- [1] André, Jacques. *Les noms des plantes dans la Rome antique* Paris Les Belles Lettres, 2010.
- [2] Costa Rute, Salgado Ana, Fahad Khan Anas, Carvalho Sara, Romary Laurent, et al. «MOR Digital: The Advent of a New Lexicographical Portuguese Project.» *Electronic lexicography in the 21st century. Proceedings of the eLex 2021 conference*, edited by Kosem, I., Cukr, M., Jakubiček, M., Kallas, J., Krek, S. & Tiberius, C., pp. 312-324.
- [3] Crane, Gregory R. «Perseus Digital Library». Tufts University. Accessed January 18, 2024. <https://www.perseus.tufts.edu/hopper/>.
- [4] «DigilibLT». “Biblioteca digitale di testi latini tardoantichi “. Università del Piemonte Orientale. Accessed January 29, 2025.
- [5] Hallig Rudolf, Wartburg Walther von, *Begriffssystem als Grundlage für die Lexikographie / Système raisonné des concepts pour servir de base à la lexicographie*, Berlin (Akademie-Verlag), [1952] 1963.
- [6] «LiLa: Linking Latin». Università Cattolica del Sacro Cuore. Accessed May 14, 2024. <https://lila-erc.eu>
- [7] «Logeion». University of Chicago. Accessed January 18, 2024. <https://logeion.uchicago.edu>.
- [8] Lopez, Patrice. «Kermitt2/Grobid». January 11, 2024. <https://github.com/kermitt2/grobid>.
- [9] Mambrini Francesco, Cecchini Flavio Massimiliano, Franzini Greta, Litta Eleonora, Passarotti Marco Carlo, Ruffolo Paolo. «LiLa: Linking Latin Risorse Linguistiche per Il Latino Nel Semantic Web (AIUCD 2019)». *Umanistica Digitale*, no. 8 (2020): 63-78.
- [10] «Musisque Deoque». Università Ca' Foscari. January 6, 2024. <https://mizar.unive.it/mqdq/public/index>.
- [11] «Plants of the World Online». “Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet; <https://powo.science.kew.org/> Accessed January 28, 2025.
- [12] «Rackham, Oliver. 1996. «Ecology and pseudo-ecology: the example of ancient Greece». In *Human landscapes in classical antiquity: environment and culture*, a cura di D. Graham J. Shipley e John Salmon, 6:16–43. *Leicester-Nottingham studies in ancient society*. London: Routledge.
- [13] «Rescribe OCR». Durham University. Accessed January 13, 2024. <https://rescribe.xyz/>.
- [14] Romary, Laurent, and Toma Tasovac. «TEI Lex-0: A Target Format for TEI-Encoded Dictionaries and Lexical Resources». *TEI Conference and Members' Meeting, Sep 2018*. <https://inria.hal.science/hal-02265312>

- [15] Sassolini, Eva, Sebastiana Cucurullo, e Marco Biffi. “L’informatizzazione del GDLI [Grande Dizionario della Lingua Italiana di G. Battaglia]: risultati, prospettive, sfide future.” *Me.Te. Digitali. Mediterraneo in rete tra testi e contesti, Proceedings del XIII Convegno Annuale AIUCD2024*, Di Silvestro, A. & Spampinato, D. (Eds.). Catania: AIUCD, pp. 281–286. ISBN 978-88-942535-8-0. DOI: 10.6092/unibo/amsacta/7927. Quaderni di Umanistica Digitale, p. 595.
- [16] Tiberius, Carole, Simon Krek, Katrien Depuydt, Polona Gantar, Jelena Kallas, Iztok Kosem, and Michael Rundell. «Towards the Elexis Data Model: Defining a Common Vocabulary for Lexicographic Resources». *Electronic lexicography in the 21st century. Proceedings of the eLex 2021 conference*, edited by Kosem, I., Cukr, M., Jakubiček, M., Kallas, J., Krek, S. & Tiberius, C. pp. 56-77.
- [17] Tittel, Sabine, Frances Gillis-Webber, and Alessandro A. Nannini. «Towards an Ontology Based on Hallig-Wartburg’s Begriffssystem for Historical Linguistic Linked Data». *Proceedings of the 7th Workshop on Linked Data in Linguistics (LDL-2020)*. Edited by European Language Resources Association. pp. 1–10. <https://aclanthology.org/2020.ldl-1.1>.
- [18] «TLL Open Access: Thesaurus Linguae Latinae». Bayerische Akademien der Wissenschaften. Accessed January 11, 2024. <https://thesaurus.badw.de/tll-digital/tll-open-access.html>.
- [19] «Transkribus». READ-COOP. Accessed January 12, 2024. <https://readcoop.eu/it/transkribus/>.
- [20] Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, e Philip E. Bourne. 2016. «The FAIR Guiding Principles for scientific data management and stewardship». *Scientific data* 3 (1): 1–9.