Umanistica Digitale – ISSN: 2532-8816 – n. 21, 2025
A. Bellandi – *I Dizionari nel Web Semantico: Modelli, Strumenti ed Esempi*
DOI: http://doi.org/10.60923/issn.2532-8816/22199

I Dizionari nel Web Semantico: Modelli, Strumenti ed Esempi

Andrea Bellandi

Istituto di Linguistica Computazionale "A. Zampolli", CNR, Pisa, Italia
andrea.bellandi@ilc.cnr.it

Abstract

I dizionari rappresentano strumenti fondamentali per la documentazione e la trasmissione del patrimonio linguistico e culturale di una società. Con l'avvento delle tecnologie del Web Semantico e del paradigma dei Linked Data, la *lessicografia digitale* ha conosciuto una profonda evoluzione, passando dalla semplice digitalizzazione dei dizionari cartacei a forme più avanzate di rappresentazione e interconnessione dei dati linguistici. In questo contesto si inserisce il Lexicography Module (Lexicog), un modello dati sviluppato dal gruppo OntoLex del W3C, concepito per supportare la modellazione di risorse lessicografiche interoperabili secondo i principi FAIR. L'articolo offre una revisione del modello Lexicog, valutandone l'adeguatezza rispetto alle esigenze del lessicografo nella compilazione di dizionari e illustrandone l'applicazione in tre casi di studio reali riferiti a diverse tipologie lessicografiche. Inoltre, presenta un insieme di servizi informatici open source sviluppati per supportare la costruzione e l'utilizzo di dizionari computazionali. L'obiettivo è contribuire alla diffusione di pratiche lessicografiche digitali aperte, sostenibili e aderenti agli standard del Web Semantico.

Parole chiave: Lessicografia digitale, dizionari, risorse linguistiche, linguistic linked data, lexicog, lexo-server.

Dictionaries are essential tools for documenting and transmitting the linguistic and cultural heritage of a society. With the advent of Semantic Web technologies and the Linked Data paradigm, *digital lexicography* has undergone a significant transformation, evolving from the mere digitization of printed dictionaries to more advanced forms of linguistic data representation and interconnection. Within this context, the Lexicography Module (Lexicog) emerges as a data model developed by the W3C OntoLex community group, designed to support the modeling of interoperable lexicographic resources in accordance with the FAIR principles. This paper provides a review of the Lexicog model, assessing its suitability for the practical needs of lexicographers in dictionary compilation and demonstrating its application through three real-world case studies involving different types of dictionaries. Furthermore, it presents a suite of open-source software tools developed to facilitate the construction and use of computational dictionaries. The goal is to promote open, sustainable digital lexicographic practices aligned with Semantic Web standards.

Keywords: E-Lexicography, Dictionary, Linguistic Resources, Linguistic Linked Data, Lexicog, LexO-server.

1. Introduzione

I dizionari costituiscono una testimonianza della lingua, della mentalità e della cultura di una società nel periodo in cui vengono redatti. Forniscono definizioni, usi, etimologie e relazioni tra le parole, e aiutano nella comprensione e nell'apprendimento linguistico. La loro composizione rientra nell'ambito del lavoro lessicografico che comprende tutte le forme di raccolta e studio del patrimonio lessicale di una lingua su diversi livelli. Ne fanno parte, ad esempio, le analisi approfondite di singoli termini, la ricostruzione della loro evoluzione storica, i repertori lessicali relativi a un'epoca o ad un autore specifico. I dizionari, in questo contesto, rappresentano la sintesi più articolata e completa della pratica lessicografica, configurandosi come la realizzazione più complessa e sistematica a cui un lessicografo possa dedicarsi [30].

Oggi, le tecnologie del web semantico [6] e il paradigma dei Linked Data [7] (LD) consentono la creazione e lo sviluppo di risorse lessicali e lessicografiche in piena conformità con i principi FAIR (Findable, Accessible, Interoperable, and Reusable) [40]. L'adozione di tali tecnologie assicura che le risorse risultanti siano non solo altamente interoperabili, ma anche conformi agli standard contemporanei per la rappresentazione, la diffusione e la fruizione dei dati linguistici.

L'incontro della pratica lessicografica tradizionale con queste nuove tecnologie ha portato a un'evoluzione della *lessicografia digitale*, agli inizi basata principalmente sulla digitalizzazione di dizionari cartacei[1], la consultazione on-line di dizionari[2] e più recentemente sulla definizione di standard XML per la rappresentazione dei dati lessicografici[3]. Oggi, grazie all'integrazione tra lessicografia digitale e web semantico, si sono aperte nuove prospettive che vanno dalla rappresentazione formale del dato linguistico, alla possibilità di connessione tra dizionari, tra dizionari e corpora e tra dizionari e ontologie concettuali.

In questo contesto, il gruppo di lavoro OntoLex[4] del W3C ha sviluppato *Lexicog*[5] (Lexicography Module), un modello dati concepito per la lessicografia digitale e destinato alla modellazione di dizionari computazionali all'interno del web semantico. Come verrà illustrato nel dettaglio nelle sezioni successive, *Lexicog* fornisce una struttura formale per la rappresentazione di dizionari, voci lessicali, definizioni, esempi d'uso e altre informazioni essenziali per la descrizione delle risorse lessicografiche. In questo articolo è nostra intenzione di: i), fornire una revisione del modello *Lexicog*, anche attraverso degli esempi di utilizzo del modello stesso in tre casi di studio che abbracciano differenti tipologie di dizionario; ii), fornire un insieme di servizi informatici a codice aperto che rendono utilizzabile il modello proposto per la costruzione e la fruizione dei dizionari. E' importante sottolineare che, nel contesto del Web Semantico, Lexicog è l'unico

[1] Con l'introduzione dei computer, i lessicografi iniziarono a digitalizzare i dizionari cartacei per facilitare la gestione e l'aggiornamento delle voci lessicali. Nacquero i primi corpora testuali digitalizzati, come il Brown Corpus [21]. Successivamente l'uso dell'elaborazione del linguaggio naturale portò alla creazione di dizionari elettronici e risorse computazionali come WordNet [32].

[2] La diffusione di Internet portò alla creazione di dizionari online (per esempio Wiktionary, Oxford Online Dictionary).

[3] Ci riferiamo a Text Encoding Initiative (TEI), in particolare al modulo per la rappresentazione di dizionari in formato XML. Per approfondimenti si rimanda al seguente link https://www.tei-c.org/release/doc/tei-p5-doc/it/html/DI.html

[4] https://www.w3.org/community/ontolex/ (ultimo accesso: 16/06/2025)

[5] https://www.w3.org/2019/09/lexicog/ (ultimo accesso: 16/06/2025)

modello RDF-nativo, proposto dal W3C per la modellazione di risorse lessicografiche, in particolare di dizionari.

L'articolo è organizzato come segue: la Sezione 2 introduce il concetto di dizionario prendendo in esame l'ampia gamma di tipologie lessicografiche esistenti; la Sezione 3 illustra il modello *Lexicog*, mettendolo a confronto con la prassi lessicografica tradizionale e analizzandone le caratteristiche principali; la Sezione 4 descrive il software sviluppato per la gestione informatica del modello proposto; la Sezione 5 è suddivisa in tre sottosezioni, ciascuna delle quali presenta un caso di studio nel quale viene utilizzato il software sviluppato; infine, la Sezione 6 propone alcune considerazioni conclusive.

2. Dizionario e Tipi di Dizionario

Spesso in letteratura, la parola vocabolario è utilizzata come sinonimo di dizionario. In generale tutti e due i termini indicano l'opera che raccoglie in ordine alfabetico le parole di una determinata lingua, di un particolare sottoinsieme di una lingua, o di più lingue. Nel seguito dell'articolo useremo indifferentemente le parole dizionario e vocabolario.

E' bene evidenziare che all'interno dei dizionari esiste una complessità tipologica che è stata sottoposta a molteplici tentativi di analisi da parte degli studiosi, al fine di individuare i tratti distintivi di ogni vocabolario e di tracciare una griglia entro cui collocare le diverse realizzazioni lessicografiche. Sebbene non sia un obiettivo dell'articolo, vale la pena discutere, seppur molto brevemente, alcune delle contrapposizioni individuate tra alcuni tipi di vocabolario, come riportato in figura 1. In primis, quella tra i dizionari dal contenuto prevalentemente *linguistico* e quelli che danno spazio all'elemento enciclopedico: i primi hanno il compito di informare sulle parole prevalentemente come segni linguistici (con indicazioni sulla loro grafia, sulla loro origine, sui loro significati e i valori d'uso), mentre quelli *enciclopedici* hanno il compito di ragguagliare sulle realtà denotate dalle parole[6]. All'interno di questa dicotomia, ci riferiamo in questo articolo al dizionario linguistico, all'interno del quale troviamo altre tipologie come i dizionari che si pongono nella prospettiva di registrare tutte le parole del lessico, siano esse di un autore, oppure di una comunità linguistica, o di una lingua - in tutto il divenire storico o in un'epoca particolare (*dizionari generali*) oppure di scegliere solo un settore della totalità del lessico di una lingua (dizionari *speciali*). Altre contrapposizioni tipologiche si osservano tra i dizionari *storici*, in particolare quelli *etimologici* che raccolgono una o più ipotesi sulla storia, l'origine o etimologia delle parole di una qualsiasi lingua e quelli *scolastici*, che a loro volta si distinguono tra dizionari *monolingui* e *bilingui*[7]. Comunque, come discusso in [16], molto spesso ogni opera può presentare tratti tipologici diversi che tendono a sovrapporsi e mescolarsi rendendo la classificazione sopra presentata puramente indicativa. Anche alcuni dei casi d'uso presentati rientrano in questa casistica. Mentre infatti il vocabolario discusso in sezione 5.2 relativo al lessico del Boccaccio rientra pienamente nella tipologia di dizionario *generale* → *sincronico* → *di un autore*, gli altri presentano tratti tipologici sovrapposti. La sezione 5.1 presenta un dizionario *speciale*, in

[6] In realtà l'opposizione non è così netta: parlare delle cose implica il ricorrere evidentemente a significati concettualmente ben determinati e viceversa dare informazioni sulle unità linguistiche (a meno che non si prenda in considerazione solo l'aspetto formale delle stesse come nei dizionari di grafia o pronuncia) vuol dire ricorrere talvolta a dati extra linguistici che illuminano sul segno linguistico stesso, di qui l'eterogeneità degli articoli del dizionario.

[7] Per ulteriori approfondimenti si prega di consultare [33].

riferimento alla terminologia medico-botanica dell'antico Occitano, ma *multilingue* e con tratti *enciclopedici*; la sezione 5.3 presenta un dizionario *speciale* delle lingue dell'Italia antica (in particolare il lessico funerario e religioso della documentazione epigrafica in Osco, Falisco e Venetico), ma anche *storico* ed in particolare *etimologico*.

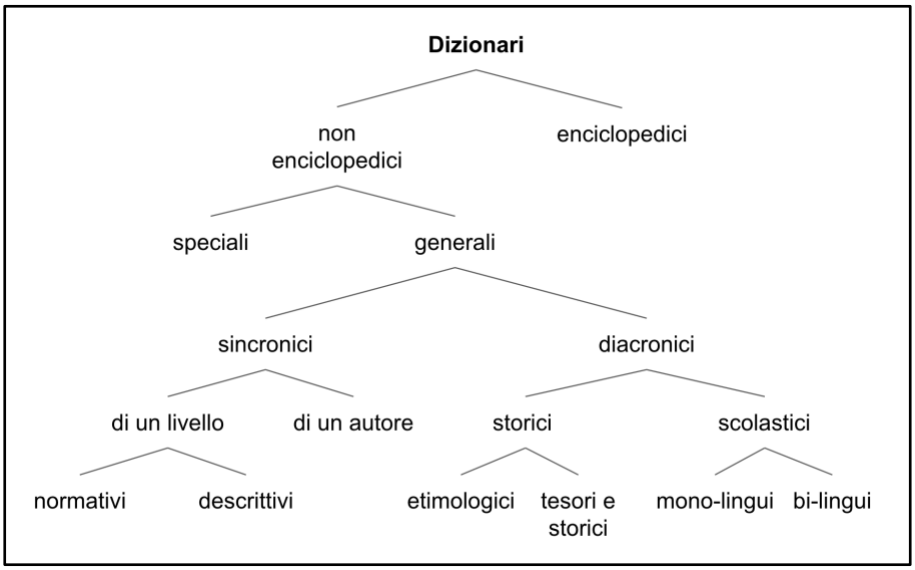


Figura 1. Tipi di dizionario, adattato da [33].

Chiarita l'impossibilità di assegnare una tipologia precisa ad un dizionario, cerchiamo adesso di individuare quelli che sono i suoi principali elementi costitutivi. Come descritto in ([16]; [17]), un dizionario si compone di una *macrostruttura* e una *microstruttura*. La *macrostruttura* è formata principalmente da un insieme di lemmi[8], ognuno dei quali rappresenta la parola collocata in esponente, come entrata di ogni singola voce del vocabolario, che viene poi sviluppata nell'articolo relativo. La *microstruttura* del dizionario invece fa riferimento a tutti gli elementi che compongono una sua voce. Figura 2 presenta un esempio di microstruttura relativa alla voce *"libro"*, all'interno della quale sono evidenziate, mediante lettere, le diverse tipologie di informazioni in essa contenute. In linea generale, la configurazione della microstruttura può variare in funzione della tipologia di dizionario considerata, includendo integralmente oppure solo parzialmente le informazioni riportate in tale voce e descritte qui di seguito (tra parentesi si riportano i casi della figura 2):

A. intestazione della voce (*libro*)9;

B. trascrizione fonetica, divisione in sillabe, indicazione della pronuncia (lì-bro);

C. varianti grafiche (*libri*);

[8] Per completezza, possono essere presenti anche un'eventuale introduzione, un insieme di avvertenze sull'uso dell'opera e delle eventuali appendici.

[9] Spesso nei dizionari cartacei nell'intestazione della voce si comprende anche B-F.

D. indicazioni morfologiche: declinazione di nomi e aggettivi, coniugazione dei verbi, indicazione del genere, ecc. (nome, maschile);

E. categoria grammaticale (sostantivo);

F. etimologia: in genere può influenzare l'ordine dei lemmi e portare alla distinzione di termini omografi sulla base della loro origine storica (dal latino *librum*);

G. definizione: i diversi significati che una parola può assumere in base ai contesti in cui compare vengono definiti e organizzati spesso in maniera gerarchica (5 significati ordinati dal più generale a quello specialistico);

H. fraseologia ed esempi d'uso (*libro in brossura*, *un libro interessante*, …);

I. sottolemmi: elementi che, pur avendo una propria autonomia semantica, non costituiscono vere e proprie unità lessicali e quindi non sono registrati autonomamente, ma vengono relegati in posizione secondaria, in fondo alla trattazione del lemma (*libricino*, *libraccio*, …);

J. sinonimi, contrari e traduzioni (ad esempio *volume* definito come sinonimo);

K. marche d'uso: consistono in abbreviazioni che segnalano l'ambito o il registro d'uso e possono indicare la frequenza d'uso della parola (comune, poco comune, rara, …), il settore disciplinare di appartenenza (geologia, matematica, biologia, …), l'uso figurato o estensivo, l'ambito geografico (toscano, regionale, …), il tono (ironico, spregiativo, …), il registro espressivo (letterario, volgare, …), l'uso grammaticale dell'accezione (sostantivale, aggettivale, ecc.).

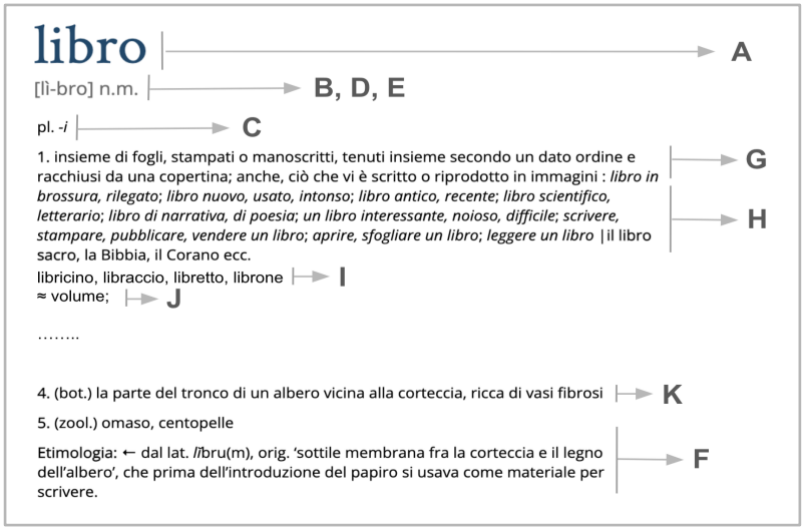


Figura 2. Esempio della voce di dizionario "libro" riadattato da Garzanti Linguistica on-line[10].

[10] https://www.garzantilinguistica.it/ricerca/?q=libro# (ultimo accesso: 16/06/2025)

Non è raro infine, riscontrare in alcune voci di dizionario la giustapposizione di due o più parole che costituiscono un sintagma stabilizzata dall'uso, come nel caso delle collocazioni, oppure la presenza di rimandi ad altre voci del medesimo dizionario. Nella sezione successiva si analizzerà se, e in quale misura, tali aspetti possano essere rappresentati all'interno di un modello dati nativamente concepito per il Web Semantico.

3. La Rappresentazione dei Dizionari nel Web Semantico

3.1. Il Modello Lexicog

Una delle caratteristiche più importanti del Web Semantico e in particolare del paradigma sul quale è basato, è la riusabilità. La riusabilità nei LD si riferisce alla capacità di utilizzare risorse già pubblicate all'interno di nuove risorse, massimizzando così il valore delle informazioni disponibili, evitando duplicazioni e favorendo l'integrazione tra dataset eterogenei. Vedremo, anche in seguito, come diversi modelli (o schemi di dati) già esistenti possano essere utilizzati in combinazione tra loro per rappresentare aspetti diversi di una stessa voce di dizionario. Anche il modello *Lexicog* nasce proprio con questo spirito: è specificamente progettato per rappresentare i dizionari come LD, ed integra – leggi riusa – un insieme di moduli già esistente, chiamato *OntoLex-Lemon* [31], estendendone le funzionalità al fine di supportare in modo efficace le strutture e le annotazioni comunemente impiegate nelle pratiche lessicografiche. Come illustrato in figura 3, i due modelli sono distinti ma strettamente interconnessi.

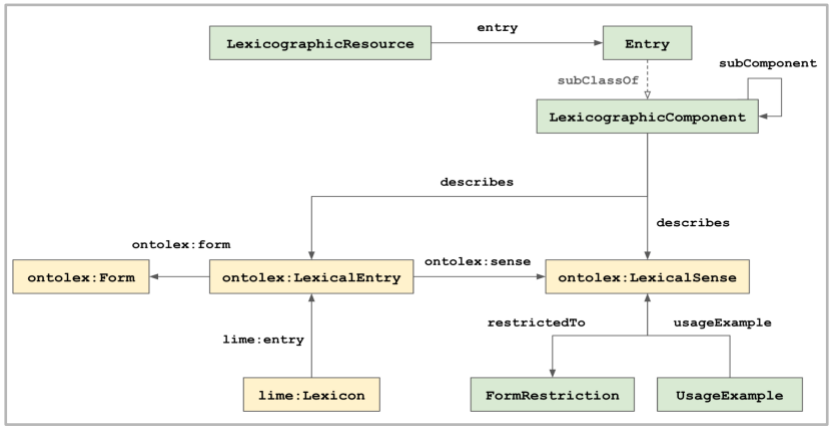


Figura 3. Modello per la rappresentazione dei dizionari – *Lexicog*. In giallo le entità di *OntoLex-Lemon* riusate dal modello e in verde le entità introdotte da *Lexicog*.

OntoLex-Lemon è lo standard di fatto per la rappresentazione di risorse linguistiche — come lessici e terminologie — nel contesto del Web Semantico. Facendo riferimento alle entità in giallo di figura 3, la classe `LexicalEntry` è definita come un'unità lessicale avente un'unica categoriale grammaticale (sostantivo, aggettivo, ecc.), un'unica etimologia e appartenente ad un lessico in una data lingua, rappresentato dalla classe `Lexicon`. Ogni entrata lessicale può essere istanziata da elementi lessicali rappresentati da una singola parola (`Word`), da una parola sintagmatica (`MultiwordExpression`) o da una parte di parola (`Affix`). La classe `Form` è popolata dalle realizzazioni morfologiche di un'entrata lessicale, associate a una o più

rappresentazioni scritte. Tra tutte le forme, è possibile definire una forma canonica, che rappresenta il lemma. La classe `LexicalSense` rappresenta infine l'insieme dei possibili significati dell'entrata lessicale. A questo livello possono essere definite le relazioni semantiche come la sinonima, l'iperonimia, l'antonimia, ecc. Il repertorio di tali relazioni, così come quello delle categorie grammaticali e dei tratti morfologici, è fornito dallo schema linguistico LexInfo[11] [11], un'ontologia di categorie di dati. *Lexicog* è definito in maniera complementare a *OntoLex-Lemon*: organizza, struttura e ordina i lemmi che sono rappresentati come entrate lessicali di uno o più lessici, in voci di dizionario. Tuttavia non vi è sempre una corrispondenza diretta tra le voci di un dizionario e le entrate lessicali (`LexicalEntry`) all'interno di un lessico *OntoLex-Lemon*. Ad esempio, alcune voci di dizionario possono includere traduzioni e sinonimi che, pur non disponendo di una voce dedicata nel vocabolario, dovrebbero essere trattati computazionalmente come entità autonome, ovvero come entrate lessicali appartenenti a uno o più lessici (`Lexicon`). Analogamente possono includere entrate lessicali differenti, comprendendo sia le definizioni come sostantivo che come aggettivo. Ad esempio la voce "*informatica*" di un dizionario italiano potrebbe contenere la sua traduzione inglese "*computer science*". Oppure la voce "*bello*" potrebbe riferirsi sia all'aggettivo che al sostantivo. Per gestire questa complessità, come mostrato in figura 3, il modello introduce la classe `Entry`, che consente di rappresentare la voce di dizionario aggregando in modo strutturato più entrate lessicali, riflettendo così l'organizzazione stabilita dal lessicografo. In riferimento agli esempi di prima, la voce "*informatica*" aggregherà l'entrata lessicale "*informatica*" e l'entrata lessicale "*computer science*" (o, se necessario, il suo particolare significato) appartenente al lessico inglese, legandole con una relazione di traduzione. La voce "*bello*" invece, includerà le entrate lessicali italiane di "*bello*" come aggettivo e "*bello*" come nome. Gli elementi che costituiscono tale strutturazione sono di tipo `LexicographicComponent`. Ognuno di essi può descrivere, tramite la proprietà `describes`, sia un significato lessicale (`LexicalSense`) che una entrata lessicale (`LexicalEntry`). La loro strutturazione può avvenire in diversi modi: possono essere organizzati secondo un ordine specifico tramite il meccanismo delle liste ordinate di RDF[12], oppure possono essere strutturate secondo una gerarchia prestabilita attraverso la proprietà di *Lexicog* chiamata `subComponent`, oppure possono essere semplicemente dichiarati come elementi non ordinati integrati all'interno di una determinata voce di dizionario. Infine, il dizionario viene definito come elemento di tipo `LexicographicResource`, formato dall'insieme delle voci di tipo `Entry` associate attraverso la proprietà `entry`.

I record lessicografici, in genere, includono anche esempi d'uso di una voce per ciascuno dei suoi significati. Essi costituiscono il contesto in cui lo specifico significato della parola viene precisato, illustrano gli usi sintattici della parola nella frase e contribuiscono alla descrizione degli usi linguistici per quanto riguarda modi di dire, locuzioni o espressioni idiomatiche. Sebbene questi esempi possano essere rappresentati con la proprietà stringa `examples` di SKOS[13] (Simple Knowledge Organization System), talvolta le informazioni relative all'esempio vanno oltre il semplice testo, cioè si rende necessario dare informazioni aggiuntive sullo specifico esempio. Se da un lato la prassi dei dizionari, specialmente monolingui, si rivolge sempre più agli esempi plasmati dai lessicografi per illustrare l'uso della lingua, dall'altro non rinuncia all'autorevolezza delle citazioni d'autore. Il modello *Lexicog* introduce quindi la classe `UsageExample` (rappresentata in verde in Figura 3) per rappresentare esempi testuali dell'uso di un significato e

[11] https://lexinfo.net/ (ultimo accesso: 16/06/2025)

[12] https://www.w3.org/TR/rdf-schema/#ch_list (ultimo accesso: 16/06/2025)

[13] SKOS fornisce un modo standard per rappresentare i sistemi di organizzazione della conoscenza utilizzando il RDF, https://www.w3.org/2004/02/skos/ (ultimo accesso: 16/06/2025)

citazioni come entità riferibili nel dizionario. In questo modo, esempi d'uso e citazioni d'autore possono essere collegate a qualsiasi altra informazione rilevante ad essi associata, dalla provenienza — ad esempio, mediante un'apposita proprietà che relaziona l'esempio o la citazione al testo da cui è stato estratto — alle note d'uso. Inoltre, in specifici contesti, può risultare necessario associare una o più forme esclusivamente a un sottoinsieme dei significati possibili di una voce lessicale. Il modello *Lexicog* consente di specificare i tratti morfologici delle forme da collegare ai significati attraverso la classe `FormRestriction`.

3.2. Esempi di Utilizzo del Modello

L'analisi comparativa tra quanto appena descritto e rappresentato in figura 3 e le informazioni presenti nella *microstruttura* del dizionario definite in Sezione 2, mette però in luce l'insufficienza degli elementi forniti dal modello nel rappresentare in maniera esaustiva tutti gli aspetti richiesti da un dizionario. Analizzando la struttura della voce "libro" in figura 2, identifichiamo quali elementi sono adeguatamente rappresentati dal modello e in che modo, e quali invece risultino esclusi o parzialmente trattati. Da qui in poi, gli esempi di codifica dell'informazione saranno illustrati tramite esempi in codice Turle14, un formato di rappresentazione dati interpretabile dal lettore.

L'intestazione della voce è costituita dal lemma, seguito dalle indicazioni sulla pronuncia: tali indicazioni riguardano ad esempio la posizione dell'accento tonico, o il timbro aperto o chiuso delle vocali. Molti dizionari riportano anche la trascrizione in alfabeto fonetico internazionale e la sillabazione, nonché le varianti (o forme) del lemma. Per quanto riguarda la rappresentazione degli aspetti fonetici, di pronuncia, morfologici e grammaticali, il modello *OntoLex-Lemon* e l'ontologia Lexinfo forniscono già le strutture necessarie.

[14] Turtle (Terse RDF Triple Language) è un formato di file ideato per esprimere dati di tipo RDF. Secondo le convenzioni RDF, le informazioni sono rappresentate per mezzo di triple, ciascuna delle quali consiste di un soggetto, un predicato e un oggetto. Per approfondimenti si veda il seguente link: https://www.w3.org/TR/turtle/ (ultimo accesso: 16/06/2025)

```
01  :libro_lex a ontolex:Word ;
02        lexinfo:partOfSpeech lexinfo:noun ;
03        lexinfo:gender lexinfo:masculine ;
04        ontolex:canonicalForm :libro_lemma ;
05        ontolex:otherForm :libro_plur ;
06        ontolex:sense libro_senseGen, libro_senseX,
07                libro_senseBot, libro_senseY,
08                libro_senseZool ;
09        lexinfo:etymology "dal lat. librum, sottile
                membrana fra la corteccia e . . . " .

10  :libro_lemma a ontolex:Form, pdstruct:Word ;
11        lexinfo:number lexinfo:singular ;
12        ontolex:writtenRep "libro"@it ;
13        lexinfo:pronunciation "libro" ;
14        ontolex:phoneticRep "'libro"@it-fonipa ;
15        pdstruct:hasSyllable :libro_syll_1 ;
16        pdstruct:hasSyllable :libro_syll_2 .

17  :libro_syll_1 a pdstruct:Syllable ;
18        pdstruct:positionInWord 1/2 ;
19        pdstruct:nextSyllable :libro_syll_2 ;
20        pdstruct:content "li"@it .

21  :libro_syll_2 a pdstruct:Syllable ;
22        pdstruct:positionInWord 2/2 ;
23        pdstruct:previousSyllable :libro_syll_1 ;
24        pdstruct:content "bro"@it .

25  :libro_plur a ontolex:Form ;
26        lexinfo:number lexinfo:plural ;
27        ontolex:writtenRep "libri"@it ;
28        ontolex:phoneticRep "'libri"@it-fonipa .

29  :libro_senseGen a ontolex:LexicalSense ;
30        skos:definition "Insieme di fogli,
31                stampati o manoscritti . . ." ;
32        lexicog:usageExample :libro_sense_1_ex ;
33        lexinfo:synonym :volume_sensoGen .

34  :libro_senseGen_ex a lexicog:UsageExample;
35        rdf:value "libro in brossura" .

36  :libro_senseZool a ontolex:LexicalSense ;
37        skos:definition "La parte del tronco . . " ;
38        dct:subject <https://dbpedia.org/page/Zoology> .

39  :libro_senseBot a ontolex:LexicalSense ;
40        skos:definition "Omaso, centopelle" ;
41        dct:subject <https://dbpedia.org/page/Botany> .

42  :volume_senseGen a ontolex:LexicalSense ;
43        skos:definition "Complesso di fogli uniti
44                        insieme, sia che . . " ;
45        lexinfo:synonym :libro_senseGen .

46  :volume_lex a ontolex:Word ;
47        lexinfo:partOfSpeech lexinfo:noun ;
48        lexinfo:gender lexinfo:masculine ;
49        ontolex:canonicalForm :volume_lemma ;
50        ontolex:sense volume_senseGen, . . .

51  :volume_lemma a ontolex:Form ;
52        lexinfo:number lexinfo:singular ;
53        ontolex:writtenRep "volume"@it ;
54        ontolex:phoneticRep "voˈlu.me"@it-fonipa .
```

Figura 4. Rappresentazione in Turtle dell'entrata lessicale "libro".

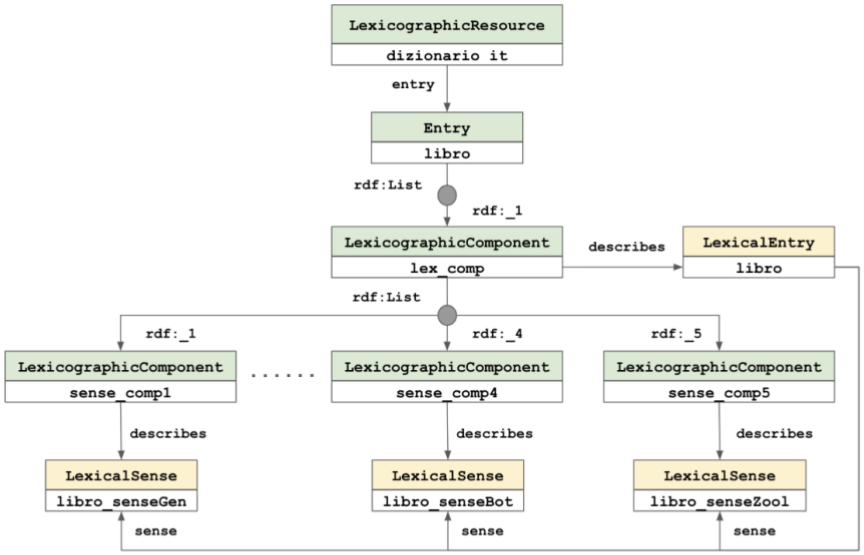
Come mostrato in figura 4, nelle righe 01-09 viene rappresentata l'entrata lessicale di "libro" come una `Word` con categoria grammaticale sostantivo. Si noti che anche il tratto morfologico "genere" col valore "maschile" può essere specificato a questo livello. Ciò implica che quel tratto sarà comune a tutte le forme di quella entrata: nello specifico una canonica – cioè il lemma, righe 10-16 – e una plurale, righe 25-28. Per ognuna di esse vengono specificati i tratti morfologici, la forma grafica (`writtenRep`) e la fonetica (`phoneticRep`) secondo il sistema fonetico IPA[15]. Per il lemma viene anche rappresentata la pronuncia (`pronunciation`). Al contrario, il modello non prevede alcun elemento specifico per la rappresentazione della sillabazione. Pertanto, è necessario ricorrere a ontologie alternative che definiscano strutture adeguate per modellare tale aspetto. Nell'esempio di figura 4, nella parte in grassetto, viene utilizzata l'ontologia POSTDATA[16] (Poetry Standardization and Linked Open Data) [19] che permette di associare alla forma una sequenza ordinata di sillabe. Per fare ciò è necessario assegnare il tipo `Word` di POSTDATA alla forma, come specificato alla riga 10. In questo modo la forma potrà disporre dell'uso della proprietà `hasSyllable`, come specificato alle righe 15-16, per la dichiarazione della sequenza ordinata "li" e "bro" (righe 17-24). Si noti che la soluzione qui proposta potrebbe non essere l'unica possibile ma, a conoscenza dell'autore, l'ontologia POSTDATA sembra essere tra le più autorevoli per questi aspetti.

La definizione costituisce il vero e proprio corpo della voce, il luogo in cui si illustra il significato del lemma, quando è unico, o si sviluppano e definiscono le sue varie accezioni, quando si tratta di lemma polisemico. La distinzione dei vari significati può essere articolata in più accezioni,

[15] https://www.treccani.it/enciclopedia/alfabeto-fonetico_(Enciclopedia-dell'Italiano)/ (ultimo accesso: 16/06/2025)

[16] https://postdata.linhd.uned.es/results/network-of-ontologies/ (ultimo accesso: 16/06/2025)

distinte da numeri progressivi. Le righe 07-09 dichiarano i 5 significati possibili di "libro". La definizione testuale del singolo significato è rappresentata attraverso la proprietà `definition`, appartenente all'ontologia SKOS, come specificato nelle righe 29-31. Comune ai diversi tipi di definizione è l'esigenza primaria di realizzare la sinonimia, ovvero l'equivalenza tra la parola da definire e l'enunciato della definizione stessa; tale equivalenza può attuarsi con una parola sinonimo. La riga 33 rappresenta la corretta accezione del termine *"volume"* (definito alle righe 42-45) come sinonimo della prima accezione di *"libro"*, mediante l'impiego della categoria semantica fornita da Lexinfo. È opportuno osservare che, poiché la relazione di sinonimia è definita come simmetrica, un reasoner sarà in grado di inferire automaticamente anche la relazione inversa, come illustrato alla riga 45. Infine, le righe 32 e 34-35, mostrano un semplice esempio d'uso relativo al primo significato di "libro", utilizzando la classe `UsageExample`. Per casi più complessi sulla modellazione di esempi d'uso e citazioni, riferirsi a ([23];[26]).




```
01  :dizionario_it a lexicog:LexicographicResource ;
02         dct:language "it" ; lexicog:entry :libro .
03
04  :libro a lexicog:Entry ;
05         rdf:_1 :lex_comp ; rdfs:label "libro"@it .
06
07  :lex_comp a lexicog:LexicographicComponent ;
08         rdf:_1 :sense_comp1 ; rdf:_2 :sense_comp2 ;
09         rdf:_3 :sense_comp3 ; rdf:_4 :sense_comp4 ;
10         rdf:_5 :sense_comp5 ;
11         lexicog:describes :libro_lex .
12
13  :sense_comp_1 a lexicog:LexicographicComponent ;
14          lexicog:describes :libro_senseGen ;
15          rds:label "1" .
16
17  . . . . . .
18
19  :sense_comp_4 a lexicog:LexicographicComponent ;
20          lexicog:describes :libro_senseBot ;
21          rds:label "4" .
22
23  :sense_comp_5 a lexicog:LexicographicComponent ;
24          lexicog:describes :libro_senseZool ;
25          rds:label "5" .
26
27  :libro_senseGen lexinfo:synonym :volume_sense .
```

(b)

Figura 5. Ordinamento delle accezioni della voce "libro". (a) rappresentazione grafica del modello - (b) codice *Turtle*.

Le accezioni più tecniche del lemma *libro* (la quarta e la quinta riportate in Figura 2), impiegate in ambiti specialistici, sono contrassegnate rispettivamente con le marche d'uso "botanica" e "zoologia". Sebbene il modello *Lexicog* non preveda un meccanismo esplicito per la codifica di

tali marcature d'uso, il vocabolario *Dublin Core*[17] mette a disposizione la proprietà `subject`, che consente di associare un *tema* (topic) a ciascuna accezione. Il repertorio semantico per i temi può essere selezionato tra diversi schemi esistenti; nel presente lavoro si opta per l'ontologia DBpedia[18], come vocabolario di riferimento. Le righe 38 e 41 mostrano come assegnare la marca "botanica" e "zoologia"[19] alle rispettive accezioni.

Nella redazione di tutti i possibili significati di una voce di dizionario, il lessicografo attribuisce ad essi un ordinamento. I criteri di ordinamento possono essere diversi: l'impiego temporale dell'accezione, la sua frequenza d'uso, o il fatto che un'accezione rappresenti un significato "primario" rispetto ad altre che scaturiscono da essa. L'articolazione delle accezioni di "libro" è identificata da un lista ordinata di numeri progressivi, da 1 a 5. In Figura 5(a) è illustrata una rappresentazione grafica dell'entrata lessicale relativa al lemma "libro", articolata in cinque accezioni distinte. Il modello *Lexicog* fornisce una struttura formale che consente di ordinare tali accezioni, assegnando a ciascuna di esse un identificatore progressivo, come visto in sezione 3.1. Le righe 01-02 di figura 5(b) definiscono la voce di dizionario italiano di "libro", costituita dalla sola entrata lessicale "libro" (righe 04-05) i cui significati sono ordinati da 1 a 5 (righe 07-11). Ogni componente d'ordine del significato si riferisce alla specifica accezione (ad esempio riga 14 per il primo senso) e la registra con un'etichetta numerica (ad esempio riga 15 sempre per il primo senso). È importante osservare che, oltre all'organizzazione lineare delle accezioni illustrata in precedenza, i dizionari prevedono frequentemente una suddivisione ulteriore di tali accezioni in sotto-accezioni, solitamente identificate da lettere alfabetiche (a, b, c, …). Il modello *Lexicog* consente di rappresentare formalmente questa articolazione gerarchica mediante la strutturazione di un `LexicographicComponent`, membro di una lista ordinata, come contenitore di un'ulteriore lista annidata. Tale meccanismo di annidamento può essere applicato ricorsivamente, consentendo una profondità strutturale teoricamente illimitata nella rappresentazione delle voci lessicografiche.

Per quanto concerne l'etimologia, LexInfo propone una soluzione di base attraverso l'impiego della proprietà `etymology`, alla quale può essere associata una stringa contenente una descrizione testuale della natura etimologica della voce (riga 09). Tale modellazione, tuttavia, potrebbe rivelarsi insufficiente qualora si presenti l'esigenza, ad esempio, di riferire esplicitamente un'entità che rappresenti l'etimo, di modellare catene etimologiche articolate o di descrivere relazioni di affinità tra termini appartenenti alla medesima famiglia linguistica. In tali casi, risulta più adeguato l'impiego di un modello specificamente sviluppato a questo scopo, il cui utilizzo e funzionamento verranno illustrati nel caso di studio presentato nella Sezione 5.3.

4. Servizi Informatici per la Gestione dei Modelli

Nel presente contributo è stato sviluppato un insieme di servizi informatici finalizzati alla costruzione, all'accesso e alla fruizione di risorse lessicografiche conformi al modello *Lexicog*.

[17] Dublin Core è uno standard internazionale per la descrizione di risorse digitali, basato su un insieme di metadati semplici e interoperabili. https://www.dublincore.org/specifications/dublin-core/dces/ (ultimo accesso: 16/06/2025)

[18] DBpedia è un progetto che estrae in formato strutturato i contenuti da Wikipedia, rendendoli disponibili come Linked Open Data. https://www.dbpedia.org/ (ultimo accesso: 16/06/2025)

[19] Come valori vengono specificati gli URI dei concetti di "botanica" e "zoologia" all'interno dell'ontologia DBpedia.

Tali servizi sono realizzati sotto forma di un API RESTful[20], il quale espone un insieme organico di funzionalità accessibili da applicazioni esterne (tipicamente interfacce di front-end) mediante la rete. L'infrastruttura è sviluppata come estensione del sistema LexO-server[21] [3], un insieme di servizi che gestisce lessici basati sul modello *OntoLex-Lemon*. Dal punto di vista tecnico, il sistema si fonda sull'utilizzo del protocollo HTTP per la trasmissione delle richieste e delle risposte, adottando il formato JavaScript Object Notation (JSON) come standard per la rappresentazione e lo scambio dei dati. Le interfacce dei servizi sono descritte in conformità alla specifica OpenAPI[22].

Come illustrato in figura 6(a), le unità lessicali che costituiscono il lessico rappresentano il fulcro dell'analisi e sono esaminate sia dal punto di vista semantico, sia in relazione alle loro proprietà grammaticali e stilistiche. Questi aspetti sono rappresentati tramite il modello dati *OntoLex-Lemon*, per il quale LexO-server offre già tutti i servizi necessari. Successivamente, i lemmi selezionati come rappresentanti di tali unità, possono essere *organizzati in* voci di dizionario mediante i nuovi servizi sviluppati, secondo il modello dati *Lexicog* (vedi figura 6(b)). Tali servizi sono concepiti come uno strato che avvolge i servizi già esistenti – un wrapper in termini tecnici – e, in parte, li utilizza per estenderne le funzionalità al fine di supportare in modo efficace tutte le strutture e le annotazioni viste nella sezione precedente e comunemente impiegate nelle pratiche lessicografiche.

L'insieme dei servizi offerti è organizzato in quattro gruppi funzionali, "*data*", "*creation*", "*updating*" e "*deletion*". Il primo gruppo comprende i servizi dedicati all'accesso e alla consultazione delle voci lessicografiche, tramite filtri sull'ortografia, sul tipo e sul grado di polisemia delle voci. Il secondo gruppo include i servizi preposti alla generazione delle voci di dizionario comprendenti la strutturazione gerarchica dei significati all'interno di ciascuna di esse, la creazione di fraseologie ed esempi d'uso, l'assegnazione di marche d'uso e di rimandi ad altre voci. I restanti due gruppi forniscono rispettivamente le funzionalità per la modifica e la rimozione delle informazioni appena descritte.

[20] Un API RESTful è un'interfaccia software che consente la comunicazione tra diversi sistemi informatici, seguendo i principi del modello architetturale REST (Representational State Transfer)

[21] LexO-server è un servizio registrato nell'infrastruttura di ricerca CLARIN-IT (http://hdl.handle.net/20.500.11752/ILC-1004, ultimo accesso: 16/06/2025)

[22] https://www.openapis.org/ (ultimo accesso: 16/06/2025)

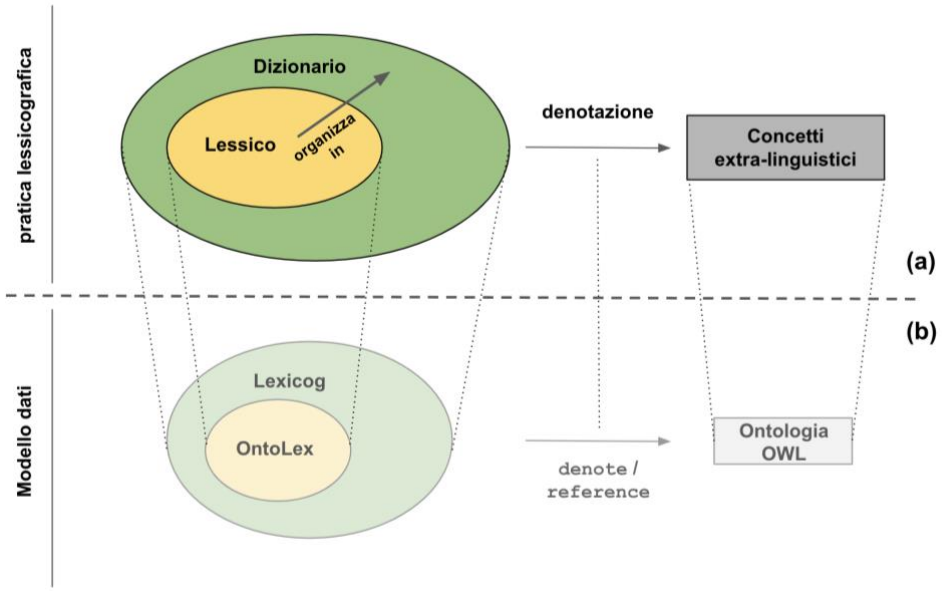


Figura 6. (a) pratica lessicografica nel contesto del Web Semantico - (b) modelli di dati che implementano la pratica lessicografica.

Talvolta comunque, la compilazione di un dizionario non implica solo l'analisi della struttura linguistica, ma può richiedere anche una comprensione della cultura della comunità che utilizza quella lingua, specialmente in base al tipo di dizionario che si intende creare. Il lessicografo è talvolta coinvolto in risposte a domande sulle cose denotate, alle tecniche e alle scienze cui la parola rinvia. Occorre cioè disambiguare e formalizzare i concetti sottesi ai diversi significati lessicali, includendo riferimenti a domini tecnici e scientifici. Tale attività non si limita più quindi alla descrizione linguistica, ma si estende alla rappresentazione semantica strutturata di entità, processi e conoscenze specialistiche cui i termini rimandano. Il software sviluppato fornisce un meccanismo tale per cui una voce di dizionario possa denotare un concetto, cioè un'entità extra-linguistica formalizzata in una risorsa esterna al dizionario (vedi figura 6(a)). Nel contesto del Web Semantico, la modellazione della conoscenza concettuale avviene tramite lo sviluppo di ontologie nel linguaggio OWL[23] (Ontology Web Language).

Componente del dizionario	***Lexicog/OntoLex-Lemon***	**Modello alternativo / integrativo**	**Gestione in LexO-server**
lemma	si	-	✅
sottolemma	no	-	❌

[23] https://www.w3.org/OWL/ (ultimo accesso: 16/06/2025)

fonetica	si	-	✓
pronuncia e traslitterazione	no	LexInfo	✓
sillabazione	no	POSTDATA	✗
categoria grammaticale	no	LexInfo	✓
morfologia	no	LexInfo	✓
forme	si	-	✓
etimologia	no	LemonEty	✓
significati (accezioni)	si	-	✓
fraseologia ed esempi d'uso	si		✓
collocazioni	no	OntoLex-FrAC	✓
marche d'uso	si	Dublin Core, LexInfo, SKOS	✓
sinonimi e contrari	no	LexInfo	✓
traduzioni	si	LexInfo	✓
rimandi	no	RDF(s)	✓
concetto (componente extra-linguistica)	si	-	✓

Tabella 1. Stato di sviluppo dei servizi per la gestione dei dizionari.

Esistono ontologie già sviluppate come DBpedia[24], una risorsa di concetti enciclopedici estratti da Wikipedia e resi disponibili nel Web Semantico, Wikidata[25] anch'essa una base di dati per il

[24] https://www.dbpedia.org/ (ultimo accesso: 16/06/2025)

[25] https://www.wikidata.org/wiki/Wikidata:Main_Page (ultimo accesso: 16/06/2025)

Web Semantico che raccoglie e archivia dati strutturati provenienti da Wikipedia e da altre fonti, oppure è ovviamente possibile creare ontologie di dominio proprie, attraverso strumenti software di libero utilizzo come Protégé[26]. I servizi consentono di associare a un concetto sia una voce lessicale sia ciascuno dei suoi singoli significati, rispettivamente tramite le proprietà `denote` e `reference` del modello *OntoLex-Lemon* (vedi figura 6(b)), assicurando un collegamento formale tra la dimensione linguistica e quella concettuale ([9];[18]). Ciò permette l'interoperabilità con risorse esterne del Web Semantico, come ontologie di dominio, thesauri o sistemi di classificazione specialistica, che rappresentano strumenti cruciali per garantire la coerenza, l'interoperabilità e il riutilizzo dei dati linguistici in ambienti digitali e computazionali. Infine, l'integrazione di riferimenti concettuali espliciti consente l'accesso onomasiologico al dizionario, ovvero, in fase di ricerca, dà la possibilità di partire da un concetto per risalire alle voci che lo denotano – eventualmente nelle diverse lingue o varianti.

Per offrire un quadro dello stato di sviluppo dei servizi per la gestione di un dizionario, si rimanda alla Tabella 1. Nella prima colonna sono riassunti gli aspetti contenuti nella microstruttura di un dizionario presentati nella Sezione 2. Mentre alcuni di questi sono previsti nel modello *Lexicog*, o nel modello *OntoLex-Lemon* su cui esso si fonda, altri non trovano una rappresentazione. In accordo con i principi dei Linked Data si rende opportuno utilizzare schemi già esistenti per modellare tali aspetti – ove possibile ovviamente – collegandoli e integrandoli tra loro. Nella terza colonna della tabella vengono pertanto proposti modelli che offrono strutture e categorie linguistiche atte a completare la rappresentazione della voce lessicografica in tutti i suoi aspetti. In particolare, LexInfo fornisce un repertorio di proprietà di rappresentazione (fonetica e di traslitterazione), di morfologia e di semantica (sinonimia e traduzione). Come descritto in Sezione 3, il modello POSTDATA permette la rappresentazione della sillabazione, mentre gli aspetti etimologici e le collocazioni possono essere codificati tramite specifici modelli basati su *OntoLex-Lemon* rispettivamente chiamati *LemonEty* [24] e *OntoLex-FrAC*[27] [10] di cui vedremo degli esempi di utilizzo nelle sezione successive. Le marche d'uso invece presentano un elevato grado di variabilità nella modellazione, in funzione della tipologia della marca stessa. In particolare, per le marche che indicano l'ambito o il settore disciplinare (ad esempio, botanica, zoologia, ecc.), come illustrato nella Sezione 3, è possibile ricorrere all'impiego di metadati Dublin Core e di ontologie esterne, come DBpedia, per rappresentare il valore semantico della marca. Per quanto riguarda, invece, le marche relative alla frequenza d'uso e alla datazione, LexInfo mette a disposizione rispettivamente la proprietà `frequency`, con valori quali `commonly used`, `infrequently used`, `rarely used`, e la proprietà `dating`, con valori come `old` e `modern`. Per le restanti tipologie di marche, per quanto noto all'autore, non sono attualmente disponibili categorie ontologiche predefinite in grado di supportarne una rappresentazione formale. La soluzione proposta per ovviare a tale mancanza, è stata quella di fornire il software sviluppato di una serie di servizi per la creazione di tassonomie personalizzate mediante SKOS [15], attraverso le quali è possibile modellare, tra l'altro, anche marche d'uso specifiche. Anche di quest'ultimo caso daremo un esempio nelle sezioni successive. I rimandi tra voci vengono modellati tramite lo schema di dati RDF(S), in particolare attraverso la proprietà `seeAlso`. L'unico aspetto che richiede una totale personalizzazione di rappresentazione è quello di "sotto lemma". Non esiste infatti, per quanto noto all'autore, una categoria di dati che definisca tale tipo di voce lessicografica. Come mostrato nella quarta colonna di Tabella 1, tale aspetto, insieme alla gestione della sillabazione, non è gestito dal software sviluppato. Come

[26] https://protege.stanford.edu/ (ultimo accesso: 16/06/2025)

[27] https://github.com/acoli-repo/ontolex-frac (ultimo accesso: 16/06/2025)

vedremo in Sezione 5.1 in questo caso è necessaria una soluzione ad-hoc che comporta la scrittura di codice personalizzato in base al progetto.

Un'ultima considerazione riguarda la rappresentazione della fraseologia e degli esempi d'uso. Come discusso nella Sezione 3 e implementato nei servizi presentati, il modello Lexicog offre una modellazione di base per tali aspetti. Tale struttura costituisce un punto di partenza solido, predisposto per essere esteso mediante l'integrazione con ulteriori categorie di dati o ontologie specifiche. Queste ultime possono fornire una descrizione più articolata, anche sotto il profilo bibliografico, delle fonti testuali e dei relativi metadati, con l'obiettivo di garantire l'interoperabilità dei dati bibliografici e di facilitare la gestione, l'analisi e il collegamento delle informazioni nel contesto del Web Semantico. Anche se non è il focus dell'articolo vale la pena citarne alcune: l'ontologia FRBRoo[28] [5], un'ontologia formale che ha lo scopo di definire e rappresentare la semantica alla base dell'informazione bibliografica per modellare le informazioni bibliografiche e culturali in contesti digitali e archivistici complessi; l'ontologia CIDOC CRM[29] (Doerr, 2003) (CIDOC Conceptual Reference Model), un modello di riferimento concettuale sviluppato dall'International Council of Museums (ICOM) che fornisce una struttura formale per descrivere le informazioni sul patrimonio culturale (quindi anche manoscritti e testi a stampa); le ontologie SPAR[30] (Peroni and Shotton, 2018) (*Semantic Publishing and Referencing Ontologies), un insieme di ontologie OWL sviluppate per rappresentare in modo semantico e strutturato le pubblicazioni scientifiche, le citazioni, i riferimenti bibliografici, i ruoli degli autori, i processi editoriali, e altri aspetti del publishing accademico. Tutte queste ontologie permettono di rappresentare, a vari livelli di dettaglio, la fonte originaria dell'esempio d'uso, sia essa un testo a stampa, un manoscritto o un manufatto, specificandone tutti i metadati.*

5. Casi d'uso

In questa sezione vengono illustrati tre casi d'uso relativi a dei progetti di costruzione di dizionari nell'ambito del Web Semantico. In ciascuno di essi sono stati sviluppati dei front-end dedicati all'editing e alla visualizzazione dei dati basati sull'utilizzo dei servizi software precedentemente descritti. Le sottosezioni seguenti forniscono una descrizione sintetica di ciascun progetto mostrandone le peculiarità e analizzando le scelte di modellazione adottate per ciascun tipo di dizionario dimostrando così l'applicabilità e la generalità dei servizi sviluppati, in differenti contesti.

5.1. DiTMAO: Il Dizionario Medico-Botanico dell'Antico Occitano

Il progetto "Dictionnaire de Termes Médico-botaniques de l'Ancien Occitan" (DiTMAO)[31] mira alla creazione di un dizionario della terminologia medico-botanica dell'antico occitano, lingua romanza parlata nel sud della Francia in epoca medievale ([12]; [13]; [14]). Tale lingua ebbe un

[28] https://cidoc-crm.org/frbroo (ultimo accesso: 16/06/2025)

[29] https://cidoc-crm.org/ (ultimo accesso: 16/06/2025)

[30] http://www.sparontologies.net/ (ultimo accesso: 16/06/2025)

[31] DiTMAO è stato un progetto congiunto dei responsabili scientifici Gerrit Bos (Università di Colonia), Andrea Bozzi (Istituto di Linguistica Computazionale "Antonio Zampolli" del CNR), Maria Sofia Corradini (Università di Pisa) e Guido Mensching (Università Georg-August di Gottinga). Il progetto è finanziato dalla Deutsche Forschungsgemeinschaft (DFG).

ruolo rilevante nella scienza medica grazie a centri come Montpellier e Tolosa. Il dizionario si propone come strumento utile non solo ai linguisti, ma anche a studiosi di testi medico-botanici medievali in ebraico contenenti termini occitani. Questi testi risultano infatti parzialmente inaccessibili senza una conoscenza specifica del lessico antico. Dal punto di vista tipologico, il dizionario si colloca nell'ambito dei "*dizionari speciali*", pur presentando caratteristiche che lo qualificano anche come multi-alfabetico — per l'impiego di caratteri ebraici e arabi — e multilingue, grazie alla presenza di traduzioni in francese e in inglese. Inoltre, si distingue per un marcato orientamento enciclopedico, che ne amplia la portata oltre la semplice funzione lessicografica. Lo strumento informatico con il quale viene redatto il dizionario è LexO (Bellandi, 2021), una interfaccia utente che utilizza i servizi software qui presentati.

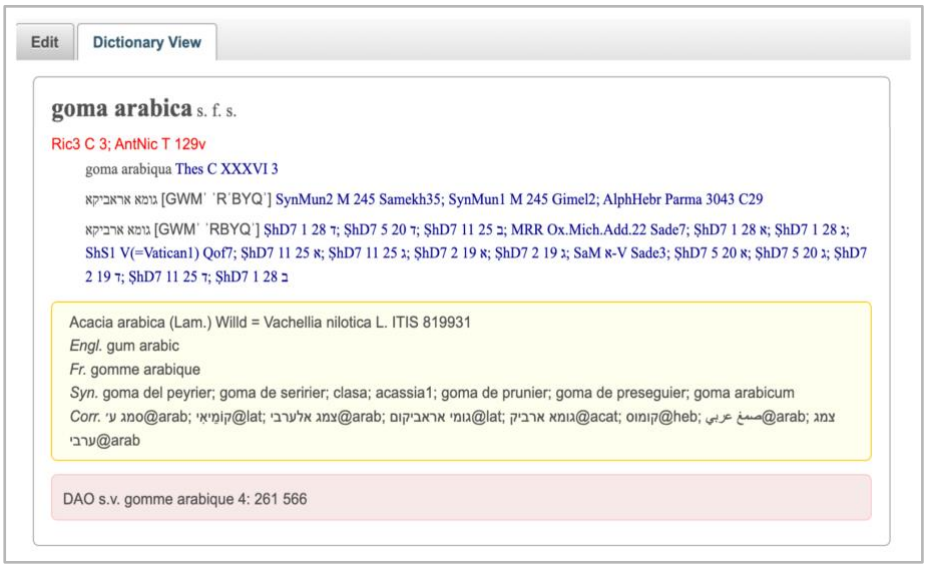


Figura 7. Entrata "goma arabica" – rendering del tool LexO per la fruizione del dizionario.

Aspetto lessicografico.

Come esempio di codifica mostriamo la voce di dizionario "goma" in antico occitano, che denota una resina vegetale utilizzata in ambito medico. In particolare ci concentriamo su una delle sue sottovoci – o sottolemma, – che indica un particolare tipo di resina, ovvero la "goma arabica". La Figura 7 presenta una visualizzazione complessiva degli aspetti informativi relativi alla sottovoce, generata automaticamente attraverso la piattaforma LexO. La rappresentazione include l'intestazione della voce corredata dei tratti morfologici, riferimenti di attestazione, un elenco delle varianti con le relative occorrenze all'interno del corpus di riferimento, il nome scientifico della specie designata, una serie di relazioni semantiche con altre accezioni appartenenti ad altre voci, nonché un riferimento, ove disponibile, alla medesima voce codificata nel *Dictionnaire de l'Ancien Occitan* (DAO)[32].

[32] Il DAO dell'occitano è un progetto scientifico che mira a documentare e descrivere in modo sistematico il lessico dell'antico occitano.

In figura 8(a) viene riportata la rappresentazione Turtle della voce di dizionario "goma": le righe 01-02 la dichiarano[33], mentre la riga 03 specifica l'entrata lessicale, descritta alle righe 07-08, che ne andrà a definire gli aspetti morfologici e semantici, per brevità non specificati in figura. Le righe 04-06 rappresentano un insieme non ordinato di sotto componenti della voce[34], che in questo caso descriveranno i sotto lemmi della voce stessa. Come si nota quindi "goma arabica" non viene elevata a voce di dizionario, ma è contenuta all'interno della voce "goma" e qualificata come espressione multi parola di tipo sotto lemma (righe 09-10). Come criterio generale (lessicografico), le espressioni polirematiche sono considerate sublemmi solo se il loro significato è non composizionale e, nel caso specifico del progetto, se possono essere considerate termini tecnici. Ad esempio, il significato di "goma arabica" non può essere dedotto dal significato delle sue parti, "goma" e "arabica"[35]. Come già evidenziato in sezione 4, per quanto noto all'autore non esiste una categoria di dati che definisca il tipo di voce lessicografica "sottolemma" e quindi è stata creata appositamente e definita come sottoclasse di `LexicalEntry` del modello OntoLex-Lemon. Le righe 14-19 definiscono gli aspetti grammaticali, le forme e le accezioni in maniera standard tramite LexInfo e OntoLex-Lemon. Le righe 12-13 completano la descrizione della sottovoce specificando un riferimento bibliografico alla stessa entrata definita nel DAO. Vista l'indisponibilità del DAO come risorsa nel Web Semantico e la conseguente impossibilità di riferirla secondo le buone pratiche dei Linked Data, è stato scelto di utilizzare la proprietà `seeAlso`[36] valorizzata con una stringa di testo che rappresenta l'identificativo della voce nel DAO.

[33] Per l'occitano antico, lo standard ISO propone il tag linguistico *pro* (@pro), derivato dal termine *provenzale*. Tuttavia, il provenzale, così come il guascone, il limosino, il linguadociano e l'Alverniate, deve essere considerato un dialetto dell'occitano antico. Pertanto, è stata adotta la denominazione *occitano antico* (in francese *ancien occitan*) come iperonimo corretto, e viene definito un nuovo tag linguistico *aoc* (@aoc) per il progetto DiTMAO.

[34] Si usa la proprietà `subComponent` di *Lexicog*, invece del meccanismo delle liste ordinate di RDF.

[35] Al contrario, il significato di un termine polirematico come "goma de gingibre" è desumibile dalle sue parti, "goma", "de" e "gingibre" ("resina", "di", e "zenzero") e quindi il suo significato non è lessicalizzato. Da un punto di vista morfologico tale espressione costituisce un composto sintagmatico o una collocazione.

[36] I possibili valori per la proprietà `seeAlso` sono elementi di tipo rdfs:Resource. Siccome rdfs:Literal è una sottoclasse di rdfs:Resource, è possibile associare alla proprietà una valore di tipo stringa.

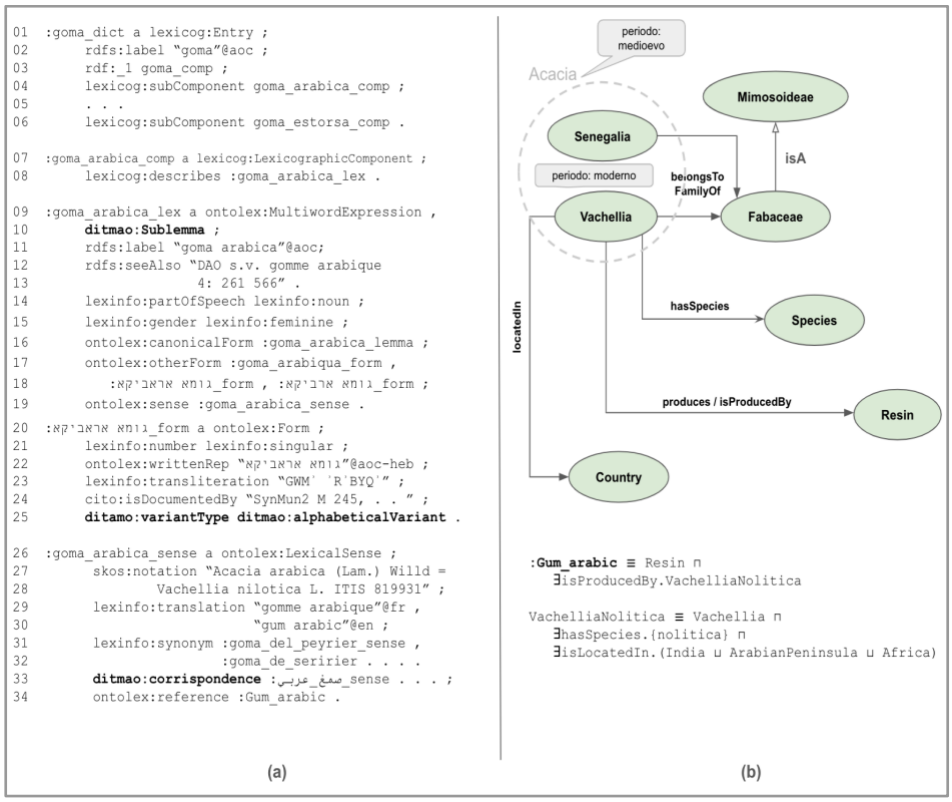


Figura 8. (a) *Aspetto lessicografico*: rappresentazione in Turtle del sottolemma "goma arabica" - (b) *Aspetto enciclopedico*: definizione in logica descrittiva del concetto di "goma arabica".

Il progetto ha posto delle problematiche anche dal punto di vista morfologico. La base testuale del lessico infatti è composta da testi medico-botanici scritti anche in latino ed ebraico. Tra le fonti ebraiche, spiccano liste di sinonimi che includono numerosi termini in antico occitano trascritti in caratteri ebraici; esistono anche varianti dei termini che differiscono nella grafia e nella pronuncia. In questo caso si è reso necessario introdurre una piccola categoria di dati personalizzata per modellare questi tipi di forme dei termini. In particolare sono stati introdotti tre tipi, `morphologicalVariant`, `alphabeticalVariant`, `graphoPhoneticVariant` e una proprietà `variantType` che associa alla forma – o variante, – uno dei tre tipi. Ad esempio, alla riga 20 è definita una forma in caratteri ebraici di "goma arabica" e la riga 25 definisce la variante come di tipo alfabetico, tramite la categoria di dati appena descritta. La variante alfabetica può essere formalizzata aggiungendo un tag di script al tag della lingua, ad esempio *aoc-heb*, come mostrato alla riga 22. Per fornire la traslitterazione della forma ebraica, si è adottata la proprietà `transliteration` di Lexinfo (riga 23). Infine, le attestazioni delle forme negli scritti del corpus documentale, sono modellate con l'ontologia CiTO[37] [36] definita come modulo delle ontologie SPAR, come mostra la riga 24.

[37] https://github.com/SPAROntologies/cito (ultimo accesso: 16/06/2025)

Dal punto di vista semantico, alla riga 26 è dichiarato il significato dell'espressione "goma arabica". La riga 27 specifica la nomenclatura della voce mediante l'impiego della proprietà `notation` del modello SKOS, a cui è associato un valore testuale rappresentante il nome scientifico del termine. Come menzionato precedentemente, il corpus del DiTMAO contiene termini corrispondenti in altre lingue antiche, che gli autori dei manoscritti consideravano sinonimi. Tuttavia, anche qualora tali termini avessero esattamente lo stesso significato, non dovrebbero essere considerati sinonimi secondo l'accezione moderna del termine, poiché non appartengono alla stessa lingua. Per modellare questo tipo di relazione si è reso necessario introdurre una nuova proprietà chiamata `correspondence`. Essa collega i sensi di due voci lessicali appartenenti a lessici distinti di lingue antiche. Per indicare invece un termine corrispondente in francese e inglese moderno, si utilizza la relazione di traduzione. È necessario mantenere distinte queste relazioni per due motivi principali: i termini corrispondenti e le traduzioni appartengono infatti a fasi storiche differenti e a registri linguistici distinti. I primi sono termini tecnici medievali, mentre i secondi rappresentano denominazioni comuni moderne. Inoltre, il corpus include termini in antico occitano che risultano sinonimi secondo l'accezione moderna del termine per i quali viene utilizzata la relazione di sinonimia di LexInfo. Le righe 29-33 mostrano l'utilizzo di tali relazioni. Si noti che le traduzioni in lingua inglese e francese alle righe 29 e 30 indicate come semplici valori testuali corrispondenti alle rispettive forme ortografiche nelle due lingue, riflettono un approccio semplificato adottato nel progetto; tuttavia, in un'ottica conforme ai principi dei Linked Data, sarebbe stato più appropriato riferirsi ai traducenti come entità già definite all'interno di risorse lessicali strutturate per l'inglese e il francese. Infine, la riga 34 stabilisce un collegamento formale tra il significato del termine e la sua definizione concettuale espressa in un'ontologia esterna al dizionario.

Aspetto enciclopedico.

Il dominio concettuale del DiTMAO mira a descrivere il significato di ciascun termine attraverso ontologie relative ai campi della botanica, zoologia, mineralogia, anatomia umana, malattie e terapie (farmaci, strumenti medici). L'obiettivo è integrare la descrizione onomasiologica, ove possibile, con una classificazione scientifica moderna per almeno la maggior parte dei nomi di piante e con una classificazione medievale delle piante e di altri farmaci semplici[38]. Figura 8(b) presenta un frammento ontologico relativo alla modellazione delle piante appartenenti alla famiglia delle *Fabaceae*. Sulla base di tale frammento, viene fornita, a titolo esemplificativo, una definizione del concetto di "goma arabica" espressa in logica descrittiva (Baader, 2003; Baader et al., 2008). La "goma arabica" è una resina naturale prodotta esclusivamente da una pianta della famiglia delle Fabaceae, appartenente al genere Vachellia e alla specie nilotica, presente in diverse aree geografiche tra cui l'Africa, l'India e la penisola araba. Una descrizione del genere può fornire la possibilità di un accesso onomasiologico al dizionario immaginando interrogazioni tipo "Quali sono i termini che designano resine prodotte da piante che si trovano in India ?", "Quali sono i termini che denotano sostanze prodotte da piante della famiglia delle Fabaceae ?", e viceversa "Quali sottolemmi di goma derivano da piante di genere Senegalia ?". Inoltre, esaminando figura 8(b) si nota che nel medioevo la "goma arabica" era considerata un genere di Acacia. Solo nel corso del tempo c'è stata una riconcettualizzazione di tale elemento. In seguito a studi filogenetici moderni infatti, tale genere è stato suddiviso in più generi (tra cui Vachellia e Senegalia). La modellazione di questi mutamenti concettuali nel corso del tempo può rendere ancor più ricco e interessante l'insieme delle possibili interrogazioni per un accesso

[38] Al momento della scrittura dell'articolo sono state solo delineate delle prospettive su questi aspetti. E' in corso la richiesta di rifinanziamento del progetto DiTAMO per lo sviluppo di queste tematiche.

onomasiologico in diacronia rispetto alle differenti classificazioni scientifiche introdotte nella storia dei termini medico-botanici dell'antico Occitano.

Nel presente caso d'uso è stato mostrato come la struttura offerta da *Lexicog* (e da *OntoLex-Lemon*) consenta una compilazione efficiente del dizionario, mentre LexInfo supporti la costruzione di una rete di relazioni semantiche tra le diverse voci lessicografiche, a parte il nuovo concetto di "corrispondenza" introdotto nel DiTMAO. I servizi sviluppati si sono dimostrati idonei nella gestione di tali aspetti. Le peculiarità del progetto hanno portato tuttavia all'esigenza di introdurre ulteriori elementi, quali il sottolemma e la categorizzazione sui tipi di forme non previsti né nei modelli citati nell'articolo, né in schemi di categorie di dati noti. A tale scopo, sono state presentate le soluzioni personalizzate adottate. Ovviamente l'integrazione di tutti gli elementi introdotti appositamente per il caso d'uso con il modello *Lexicog*, ha richiesto lo sviluppo di specifiche componenti di codice personalizzato per garantirne una gestione corretta all'interno dei servizi.

5.2. VocaBo: Il Dizionario di Boccaccio

Il progetto VocaBO[39] – "Vocabolario di Boccaccio Online", avviato alla fine del 2022, ha come obiettivo la creazione del vocabolario digitale della lingua volgare di Giovanni Boccaccio. Il *Decameron*[40], costituisce il punto di avvio del progetto, in quanto riconosciuto come la prima grande opera in prosa della letteratura italiana nonché modello di riferimento per la narrativa europea. Sebbene l'importanza della produzione di Giovanni Boccaccio sia ampiamente riconosciuta, l'analisi sistematica del suo lessico ha ricevuto finora un'attenzione limitata, con alcuni studi recenti focalizzati su aree circoscritte, quali gli hapax [39] e il lessico artistico (Murru, 2019). In confronto al lessico dantesco, quello boccacciano appare di particolare rilievo poiché si colloca nella fase conclusiva del periodo indagato dal *Tesoro della Lingua Italiana delle Origini*[41] (TLIO), offrendo spunti significativi per la comprensione degli sviluppi linguistici successivi. Un ulteriore elemento distintivo risiede nel carattere di "lessico d'autore", supportato dalla presenza di testimonianze autografe. Tra gli obiettivi principali del progetto si annoverano l'elaborazione di un lemmario di riferimento e la redazione delle corrispondenti voci lessicografiche in un dizionario d'autore attraverso uno strumento informatico chiamato MAIA[42] [22], sviluppato nel contesto del progetto. Lo strumento integra anche la gestione del corpus di indagine e la possibilità di annotare i testi con gli elementi lessicografici creati nel dizionario. Le sue caratteristiche generali permettono una sua applicazione anche in altri progetti simili. Come mostrato in figura 9, l'interfaccia permette la creazione delle voci del dizionario a partire dal lessico del Boccaccio estratto dal Decameron. Utilizza i servizi qui presentati sia per l'editing di tutti gli aspetti del dizionario, sia per la produzione di un'anteprima a stampa delle voci stesse.

La figura 10 illustra la rappresentazione della voce "dolce" in alcuni dei suoi aspetti grammaticali e semantici. La voce di dizionario è costituita da: l'entrata lessicale "dolce" (riga 25), nella quale

[39] Il progetto è promosso dall'Ente Nazionale Giovanni Boccaccio in collaborazione con l'Università per Stranieri di Siena, il CNR-Istituto di Linguistica Computazionale "Antonio Zampolli" e l'Accademia della Crusca. La direzione è affidata a Giovanna Frosini, docente di Storia della lingua italiana presso l'Università per Stranieri di Siena e presidente dell'Ente Nazionale Giovanni Boccaccio.

[40] https://www.treccani.it/magazine/lingua_italiana/speciali/VocaBO/ (ultimo accesso: 16/06/2025)

[41] http://tlio.ovi.cnr.it/TLIO/ (ultimo accesso: 16/06/2025)

[42] https://github.com/klab-ilc-cnr/Maia (ultimo accesso: 16/06/2025)

vengono strutturate tutte le sue accezioni tramite l'utilizzo degli elementi `LexicographicComponent`; un sottocomponente che raggruppa in maniera non ordinata le collocazioni della voce (riga 26); un sottocomponente che raccoglie l'insieme delle locuzioni della voce (riga 27); un rimando alla voce di dizionario "dolciato" rappresentata dalla proprietà `seeAlso`. Gli aspetti lessicografici su cui ci concentreremo in questo caso d'uso, sono i), la fitta strutturazione gerarchica dei significati delle voci in accezioni e sotto-accezioni con le proprie marche d'uso e ii) la rappresentazione dei sintagmi delle voci.

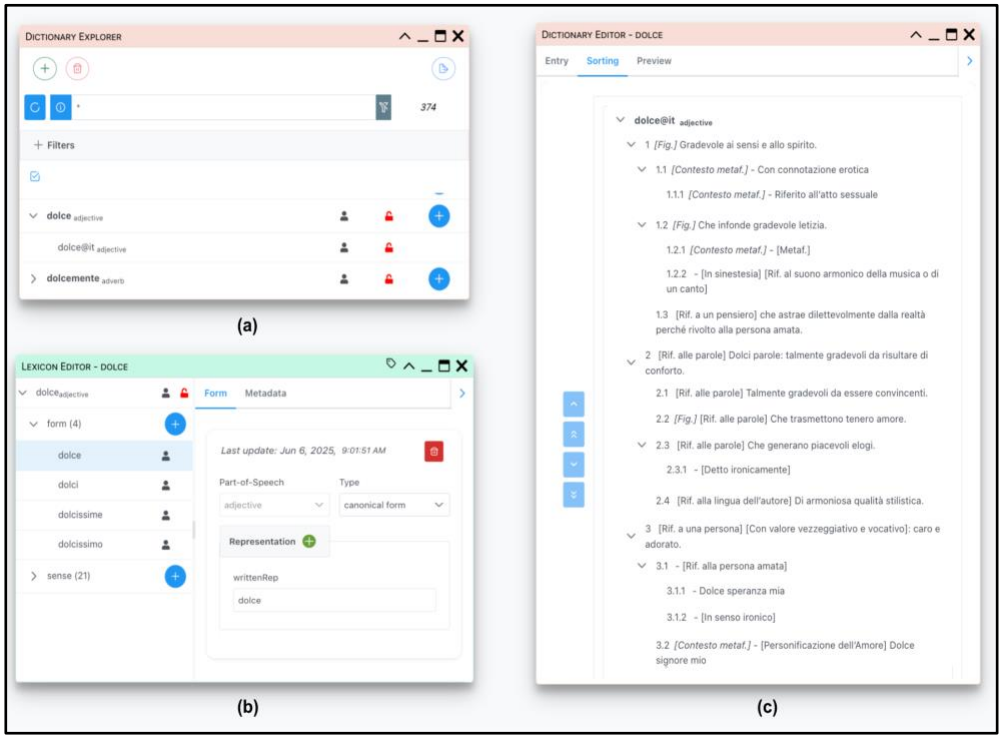


Figura 9. Pannelli per la gestione degli aspetti lessicografici dell'interfaccia di MAIA per la voce "dolce" - (a) lista delle voci di dizionario - (b) editing delle entrate lessicali - (c) pannello di ordinamento gerarchico delle accezioni di una voce.

Marche d'uso e struttura delle accezioni.

La codifica lessicale delle accezioni delle parole richiede una rappresentazione gerarchica formalmente strutturata, in cui ciascuna accezione può essere accompagnata da marche di vario tipo. Queste ultime risultano fondamentali tanto per indicare il dominio lessicale di appartenenza (quali, ad esempio, arte, scienza, agricoltura, ecc.), quanto per esplicitare i tratti semntici dell'accezione, come l'uso figurato, traslato o ulteriori impieghi non letterali del significato. Dal punto di vista grammaticale, si rende altresì necessaria l'annotazione di informazioni relative a particolari usi grammaticali, quali l'impiego sostantivale di aggettivi, l'uso impersonale dei verbi, e simili. Come anticipato nella Sezione 4, attualmente non risultano disponibili, almeno secondo le conoscenze dell'autore, categorie di dati standard che consentano di rappresentare in maniera sistematica tali aspetti. Tuttavia, il modello *OntoLex-Lemon* prevede un meccanismo generale attraverso il quale i significati lessicali possono essere associati a uno o più elementi di una

tassonomia (o semplicemente una lista) definita in SKOS (`LexicalConcept`), i quali possono anche fungere da rappresentazione astratta delle caratteristiche possedute da una determinata accezione[43]. Nel progetto VocaBO si è deciso di utilizzare questo meccanismo definendo tre insiemi di marche rappresentate come concetti SKOS, in accordo con quanto detto sopra: quelle grammaticali (ad esempio usi intransitivi, usi impersonali), quelle semantiche (ad esempio metafora, sineddoche, metonimia) e quelle d'uso (ad esempio musica, teologia, veterinaria). I servizi sviluppati permettono sia di creare le tassonomie SKOS, sia di collegare ogni elemento di esse alle accezioni delle voci.

```
:dolce_lex a ontolex:Word ;
      rdfs:label "dolce"@it ;
      lexinfo:partOfSpeech lexinfo:adjective ;
      ontolex:canonicalForm :dolce_lemma ;
      ontolex:otherForm :dolce_forma1 ,
          :dolce_forma2 , . . . , :dolce_forman ;
      ontolex:sense :dolce_sense1 ,
          :dolce_sense2 , . . . , :dolce_sensen .

:dolce_lemma a ontolex:Word ;
      lexinfo:gender lexinfo:masculine ;
      lexinfo:number lexinfo:singular ;
      ontolex:writtenRep "dolce"@it .

:dolce_sense1 a ontolex:LexicalSense ;
     dc:subject <https://dbpedia.org/page/Emotion> ;
     ontolex:isLexicalizedSenseOf :marche_sem_fig ;
     skos:definition "Gradevole ai sensi e
                      allo spirito."@it .

:dolce_sense2 a ontolex:LexicalSense ;
     dc:subject <https://dbpedia.org/page/Emotion> ;
     ontolex:isLexicalizedSenseOf :marche_sem_metaf ;
     skos:definition "Con connotazione erotica"@it .

:dolce_sense3 a ontolex:LexicalSense ;
     dc:subject <https://dbpedia.org/page/Emotion> ;
     ontolex:isLexicalizedSenseOf :marche_sem_metaf ;
     skos:definition "Riferito all'atto sessuale"@it .

:dolce_dict a lexicog:Entry ;
      rdfs:label "dolce"@it ;
      rdf:_1 :dolce_lex_comp ;
      lexicog:subComponent :docle_coll ;
      lexicog:subComponent :docle_loc ;
      rdfs:seeAlso :dolciato_dict .

 :dolce_lex_comp a lexicog:LexicographicComponent ;
      lexicog:describes :dolce_lex ;
      rdf:_1 :dolce_signore_acc1 .

:dolce_signore_acc1 a lexicog:LexicographicComponent ;
      lexicog:describes :dolce_sense1 ;
      rdf:_1 :dolce_signore_acc1_1 .

:dolce_signore_acc1_1 a lexicog:LexicographicComponent ;
      lexicog:describes :dolce_sense2 ;
      rdf:_1 :dolce_signore_acc1_1_1 ;

:dolce_signore_acc1_1_1 a lexicog:LexicographicComponent ;
      lexicog:describes :dolce_sense3 .

:dolce_coll a lexicog:LexicographicComponent ;
      rdf:_1 :dolce_signore_coll ;
      . . . .

:dolce_loc a lexicog:LexicographicComponent ;
      rdf:_1 :dolce_sale_loc ;
      . . . .

:dolce_signore_coll a frac:Collocation ;
      lexinfo:example "dolce signore mio" ;
      frac:head :dolce_senseX ;
      rdf:_1 :dolce_senseX ;
      rdf:_2 :signore_senseY ;
      frac:frequency [ rdf:value 6 ] .

:dolce_sale_loc a frac:ContextualRelation ;
      dc:description "di tipo locuzione" ;
      lexinfo:example "dolce di sale" ;
      frac:head :dolce_lex ;
      rdf:_1 :dolce_lex ;
      rdf:_2 :sale_lex ;
      frac:frequency [ rdf:value 4 ] ;
      skos:definition "sciocco, privo di sale in zucca" .
```

Figura 10. Rappresentazione in Turtle della voce "dolce" del Decameron.

In Figura 10, le righe 01-08 presentano la struttura dell'entrata lessicale relativa all'aggettivo "dolce", articolata in un insieme di forme flessive, tra cui il lemma in forma maschile singolare, esplicitato nelle righe 09-12, e un insieme di sensi associati. Le righe 13-22 illustrano tre accezioni selezionate per il caso d'uso, ciascuna riferita, a titolo esemplificativo, a un campo semantico

[43] Per approfondimenti consultare il seguente link https://www.w3.org/2016/05/ontolex/#lexical-concept (ultimo accesso: 16/06/2025)

emozionale che comprende esperienze sia di ordine percettivo sia psicologico. Nello specifico, le righe 14, 18 e 21 stabiliscono un collegamento tra le accezioni e tale campo semantico, modellato tramite un concetto estratto dall'ontologia DBpedia, mediante la proprietà `subject` del Dublin Core. Le righe 15, 20 e 24 invece associano le marche semantiche alle accezioni, (uso figurato o metaforico), create come classi SKOS e associate tramite la proprietà `isLexicalizedSenseOf`.

Come rappresentato in figura 9(c), le accezioni sono organizzate secondo una gerarchia concettuale, che va da quella più generale (accezione 1) a quella più specifica (accezione 1.1.1). Le righe 32-39 mostrano, attraverso il meccanismo dei `LexicographicComponents` e delle liste ordinate, come l'accezione di "gradevolezza ai sensi e allo spirito" venga ulteriormente specializzato in una connotazione erotica, la quale può a sua volta essere riferita a un atto sessuale specifico.

Collocazioni e locuzioni.

Nel progetto è stato ritenuto opportuno isolare i sintagmi associati alle singole voci lessicali. Per quanto mostrato fino ad ora, né il modello Lexicog né il modello *OntoLex-Lemon* prevedono elementi che ne permettano una rappresentazione. Il modello *OntoLex-FrAC* arrichisce *OntoLex-Lemon* con un insieme di classi e proprietà che permettono di rappresentare co-occorrenze tra parole. In particolare, la classe `ContextualRelation` fornisce una relazione tra due o più elementi lessicali (siano essi voci o accezioni), caratterizzata da una proprietà `description` che descrive la particolare natura della relazione, dalla possibilità di specificare un corpus dal quale la relazione è stata inferita e da una valutazione del peso o della probabilità della relazione. Le collocazioni sono definite come specializzazioni di tale classe, rappresentate dalla classe `Collocation`. I redattori del dizionario, hanno scelto di rappresentare, ove possibile, sia le combinazioni di parole che co-occorrono con alta frequenza e in maniera preferenziale — in cui ciascun componente mantiene un'autonomia semantica — ossia le collocazioni, sia le espressioni fisse o semifisse costituite da più parole che funzionano come un'unica unità strutturale e/o semantica, ovvero le locuzioni. Figura 10 presenta di seguito un esempio dell'applicazione di *OntoLex-FrAC* relativo all'entrata lessicografica della voce dolce in collocazione con signore. A partire dalla riga 49, la collocazione viene definita come istanza della classe `Collocation`. La proprietà `head` specifica il lessema che rappresenta la testa della collocazione, mentre la proprietà `example` fornisce l'esempio selezionato dal lessicografo (in questo caso, "dolce signore mio"). L'uso delle proprietà `rdf:_1` e `rdf:_2` consente di rappresentare l'ordine degli elementi all'interno della collocazione. Si osservi che tali elementi corrispondono alle accezioni specifiche di "dolce" e "signore", in quanto il significato complessivo della collocazione risulta fondamentalmente composizionale. Infine, la proprietà `frequency` permette di annotare nel dizionario il numero di occorrenze della collocazione all'interno del testo del Decameron.

Per quanto riguarda le locuzioni, il modello *OntoLex-FrAC* non implementa una classe distinta da quella delle collocazioni. Nel progetto è stato deciso di marcare le locuzioni con la classe generica `ContextualRelation` ed utilizzare la proprietà `description` per codificare testualmente la particolare natura della relazione, come mostrato alle righe 55-56 per la locuzione "dolce di sale". Analogamente a quanto descritto per la collocazione riportata sopra, le righe 57-61 specificano un esempio, la testa, le componenti e la frequenza della locuzione. Si noti che nel caso delle locuzioni, il significato non è mai dato dai singoli significati delle parole che compongono la locuzione e per questo le righe 59-60, che indicano tali componenti, si riferiscono alle voci lessicali piuttosto che a specifiche accezioni delle stesse. Questo implica però l'esigenza di creare un nuovo significato specifico per la locuzione. La soluzione adottata nel progetto è stata quella di associare una definizione come stringa di testo direttamente alla

locuzione stessa. La riga 62 mostra la definizione del sintagma riferito a una persona poco intelligente, ottusa. Si noti che questa soluzione dal punto di vista computazionale potrebbe non essere soddisfacente, in quanto il significato della locuzione non viene rappresentato come entità riferibile da/verso altre entità, ma bensì come mera stringa. Rappresentare il significato come elemento della classe `LexicalSense` comporterebbe tuttavia la creazione di un'entrata lessicale alla quale associare il senso stesso. Registrare la locuzione quindi come entrata lessicale di tipo `MultiWordExpression` potrebbe non essere lessicograficamente rilevante[44].

Nel caso d'uso analizzato, i modelli *Lexicog* e *OntoLex-Lemon* si sono rivelati adeguati nel supportare le scelte redazionali adottate dai lessicografi del progetto, in particolare per quanto concerne la complessa strutturazione delle accezioni delle voci lessicali. Per la descrizione degli aspetti specifici delle singole accezioni, si è tuttavia reso necessario definire categorie di dati ad hoc, al fine di rappresentare le marche d'uso, le marche grammaticali e quelle semantiche. Il modello *OntoLex-FrAC* ha offerto una soluzione efficace per la rappresentazione delle collocazioni e, in forma semplificata, anche delle locuzioni. In particolare, la scelta di descrivere la natura della relazione contestuale (ad esempio, una locuzione) mediante una stringa testuale assegnata alla proprietà `description` ha consentito l'impiego diretto dei servizi software sviluppati, senza richiedere modifiche al codice esistente. Rimangono comunque valide, sebbene non supportate dalle implementazioni attuali dei servizi software, soluzioni alternative più strutturate. Tra queste, si può citare la possibilità di specializzare la classe `ContextualRelation` attraverso una classe definita ad hoc, ad esempio `Phrase`, che, analogamente a `Collocation`, erediti le proprietà illustrate negli esempi. In alternativa, si potrebbe ricorrere a LexInfo, attribuendo alla locuzione un valore specifico della proprietà `termType`, come `idiom`, `phraseologicalUnit` o `setPhrase`.

5.3. ItAnt: Lingue e Culture dell'Italia Antica

Il progetto ItAnt[45] – "Lingue e culture dell'Italia antica", si propone di indagare le culture dell'Italia antica a partire dalla loro documentazione linguistica epigrafica[46]. Molte delle lingue parlate nell'antichità infatti, sono pervenute attraverso testimonianze scritte che, in alcuni casi, risultano estremamente limitate sia dal punto di vista quantitativo che qualitativo. Per queste lingue si utilizza la denominazione *Restsprachen*, ovvero "lingue di attestazione frammentaria" come l'osco, il falisco, il venetico e il celtico cisalpino, poiché i loro corpora possono essere costituiti da un numero molto ridotto di testi, talvolta poche decine, per lo più tipologicamente limitati alla forma epigrafica. La formalizzazione digitale e la rappresentazione semantica delle *Restsprachen* costituiscono di per sé un valore, in quanto strumenti per la condivisione e la conservazione del sapere esistente. Tra gli obiettivi del progetto vi è quello della creazione di un repertorio di termini presenti nelle epigrafi strutturato come un dizionario specialistico con una forte connotazione storica. Lo strumento informatico con il quale viene redatto il dizionario è EpiLexO[47] (Quochi et al., 2022; Mallia et al., 2024), un'interfaccia utente per la creazione di

[44] Una soluzione del genere richiederebbe un adattamento dei servizi software.

[45] https://www.ilc.cnr.it/progetti/itant/ (ultimo accesso: 16/06/2025)

[46] Progetto P.R.I.N. 2017 frutto della collaborazione scientifica tra Università degli Studi di Firenze (unità coordinata dalla prof.ssa Francesca Murano), Università Ca' Foscari Venezia (prof.ssa Anna Marinetti, coordinatrice nazionale) e Istituto di Linguistica Computazionale "A. Zampolli" del CNR di Pisa (unità coordinata dalla dott.ssa Valeria Quochi).

[47] Il codice è disponibile al seguente link https://github.com/DigItAnt/Epilexo (ultimo accesso: 16/06/2025)

dizionari sviluppata appositamente per il progetto, di cui si riporta uno screenshot in figura 11, che si avvale dei servizi software presentati nel presente contributo. Brevemente, sul lato sinistro vengono riportate le voci contenenti le forme (eventualmente ricostruite) attestate nelle epigrafi, il significato e l'etimologia. Nella parte centrale è possibile editare tutti gli aspetti della voce, mentre nella parte di destra è possibile gestire la bibliografia a supporto sia dell'ipotesi etimologica che della ricostruzione delle forme.

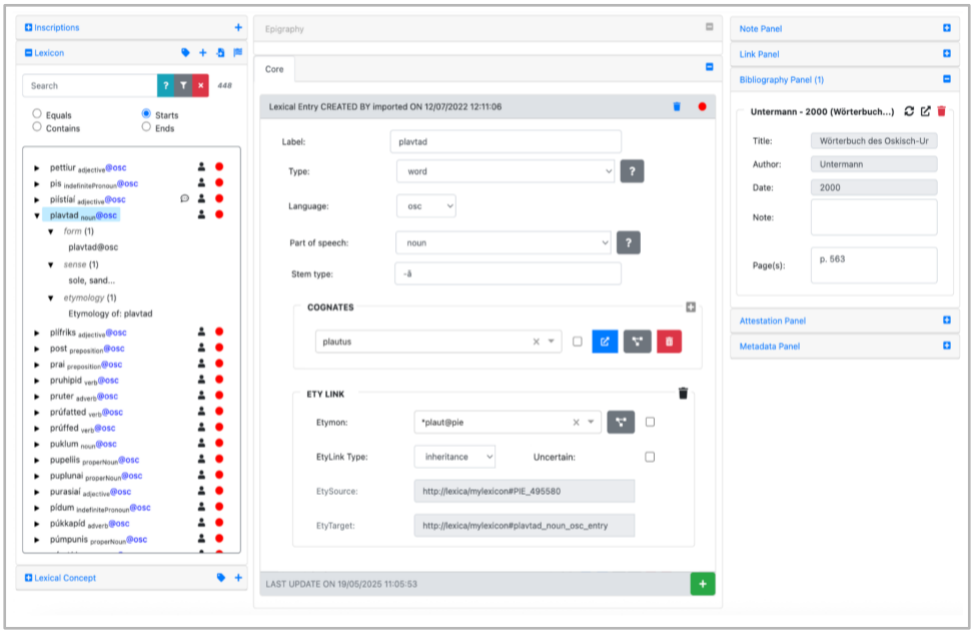


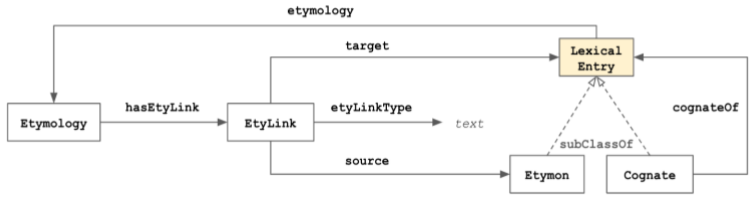
Figura 11. Interfaccia dell'editor EpiLexO. Da sinistra a destra: elenco delle voci, dettaglio della voce "plavtad", bibliografia a supporto dell'ipotesi etimologica.

In figura 12(a) è rappresentato il codice Turtle per la voce di dizionario in Osco "plavtad" (il codice ISO della lingua "osc"). Le righe 01-03 definiscono un componente relativo all'entrata lessicale di "plavtad" (righe 04-05) e uno relativo a quella di "plautus" che rappresenta la voce latina che condivide la stessa radice etimologica di "plavtad" (righe 06-07). Le righe 08-14 definiscono nel dettaglio il sostantivo "plavtad". Nel caso delle *Restsprachen*, l'assenza di un paradigma completo rende difficile l'individuazione di un lemma nella forma tradizionale. Per questo motivo, la voce lessicale è associata a realizzazioni linguistiche non normalizzate e non viene formalizzata alcuna forma canonica. La riga 10 definisce le generiche forme tramite la proprietà `lexicalForm` che sottospecifica se trattasi di lemma o forma flessa. Le forme in ItAnt corrispondono alle attestazioni reali, risultanti dalla lettura del curatore ed eventualmente comprensive di interventi editoriali come, ad esempio, il ripristino di lettere danneggiate o mancanti. Pertanto le attestazioni devono essere registrate e codificate per ciascuna forma, come avviene di norma nei dizionari storici tradizionali. Per il progetto è stata poi introdotta una proprietà non presente nei modelli esistenti denominata `stemType`, che indica approssimativamente le classi nominali e aggettivali, ad esempio i temi in *-ā*, come mostrato nella

riga 11, ovvero i temi che terminano in *-ā* < *PIE -eh₂*[48] che appartengono a un tipo specifico di declinazione.

```
:plavtad_dict a lexicog:Entry ;
   rdfs:label "plavtad"@osc ;
   rdfs:member :plavtad_comp , :plautus_comp .

:plavtad_comp a lexicog:LexicographicComponent ;
   lexicog:describes :plavtad_lex .

:ara_comp a lexicog:LexicographicComponent ;
   lexicog:describes :plautus_lex .

:plavtad_lex a ontolex:Word ;
   lexinfo:partOfspeech lexinfo:noun ;
   ontolex:lexicalForm :plavtad_form1, :plavtad_form2 ;
   itant:stemType "-ā < PIE -eh₂" ;
   lemonEty:etymology :plavtad_ety ;
   lemonEty:cognate :plautus_lex ;
   ontolex:sense :plavtad_sense .

:plautus_lex a ontolex:Word , lemonEty:Cognate ;
   ontolex:canonicalForm
         [ ontolex:writtenRep "plautus"@lat ;
           lexinfo:gender lexinfo:masculine ;
           lexinfo:number lexinfo:singular ] ;
   ontolex:sense :plautus_sense .
   ### in alternativa LiLA link
   rdfs:seeAlso http://lila-erc.eu/data/id/lemma/118101 .

:plavtad_sense a ontolex:LexicalSense ;
   skos:definition "sole"@en ;
   ontolex:isLexicalizedSenseOf [
       a ontolex:LexicalConcept ;
       skos:prefLabel "Clothing & Adornment"@en ;
       owl:sameAs https://lrc.la.utexas.edu/lex/semantic/category/CA .
   ] .

:plavtad_ety a lemonEty:Etymology ;
   lexinfo:confidence 1.0 ;
   dct:references :plavtad_ety_documentation ;
   lemonEty:hasEtyLink [
       a lemonEty:EtyLink ;
       lemonEty:etySource :plavtad_lex ;
       lemonEty:etyTarget [ a lemonEty:Etymon ;
                            rdfs:label "*plau"@pie ] ;
       lemonEty:etyType "inheritance" .
   ] .

:plavtad_ety_documentation a rdf:Description ;
   dct:contributor "Untermann" ;
   dct:title "Wörterbuch des Oskisch-Umbrischen" ;
   dct:publisher "Z4XB2N9S" ;
   skos:note "pag. 563" .
```

(a)



(b)

Figura 12. (a) Rappresentazione in Turtle della voce "plavtad" in Osco - (b) *LemonEty*, il modello etimologico adottato.

Dal punto di vista della linguistica storica che caratterizza il progetto ItAnt, le informazioni etimologiche e il relativo grado di certezza assumono un ruolo centrale rispetto a descrizioni di relazioni lessicali o sintattiche e semantiche. Il modello lessicale rappresenta i dati etimologici sfruttando il modello *LemonEty* [24], già adottato in alcuni progetti di rilievo, tra cui Linking Latin (*LiLa*)[49] [29]. Una rappresentazione grafica del modulo etimologico è data in figura 12(b). Le informazioni etimologiche sono associate a una `Lexical Entry` attraverso la classe `Etymology` e si applicano a tutte le forme ad essa collegate. Per ciascuna voce lessicale, le radici ricostruite del proto-italico e/o del proto-indoeuropeo sono rappresentate e codificate come istanze della classe `Etymon`. Le relazioni etimologiche che collegano una parola ai suoi etimi — intese come espressione dei processi storico-linguistici che ne hanno determinato

[48] "-ā < PIE -eh₂" significa che la desinenza -ā in una lingua storica (come il latino o altre lingue italiche) deriva dal suffisso -eh₂ del proto-indoeuropeo (codice lingua "PIE"), spesso associato ai nomi femminili. Questo è un classico esempio di ricostruzione etimologica, utile per comprendere l'origine morfologica delle parole.

[49] https://lila-erc.eu/#page-top (ultimo accesso: 16/06/2025)

l'evoluzione — si articolano, nel modello adottato, in due categorie principali, corrispondenti ai possibili valori della proprietà `etyLinkType`: prestito e eredità. Il prestito fa riferimento al trasferimento di elementi linguistici da una lingua all'altra attraverso il contatto linguistico; l'eredità, invece, riguarda la trasmissione di parole (o altri elementi linguistici) da una lingua madre o da una fase precedente della stessa lingua [25]. Nel caso esemplificato in figura 12(a), alla riga 12 è dichiarata l'etimologia della voce "plavtad", definita nelle righe 29–38 come istanza della classe Etymology collegata all'etimo rappresentato dalla radice ricostruita "*plau*" del proto-indoeuropeo (contrassegnata con il tag @PIE). La relazione etimologica specificata è di tipo ereditario (riga 37), indicando che il termine osco "plavtad" deriva direttamente dal proto-indoeuropeo, e non da un'altra lingua tramite prestito. Il grado di certezza dell'etimologia è stato codificato con la proprietà `confidence` di LexInfo, alla quale può essere associato un numero reale compreso tra zero (etimologia incerta) e uno (etimologia certa) come nel caso dell'esempio (riga 30). L'etimologia è stata successivamente collegata a una fonte documentaria (riga 31), fornendo così un riferimento utile ai fini della verifica e dell'approfondimento. Il sistema EpiLexO consente l'accesso diretto alla base bibliografica del progetto, gestita tramite Zotero[50], e, grazie agli appositi servizi di LexO-server — i quali permettono la creazione di descrizioni RDF generiche di entità — è stato possibile generare una risorsa bibliografica (riga 39) arricchita con proprietà Dublin Core e SKOS, valorizzate automaticamente con i dati estratti dai record di Zotero. Le righe 40–43 definiscono rispettivamente l'autore dell'opera citata, il titolo, la chiave identificativa del record all'interno di Zotero e l'intervallo di pagine pertinenti al riferimento etimologico.

I termini cognati attestati in lingue sorelle sono codificati come istanze di un altro sottotipo di `Lexical Entry` la classe `Cognate`. La riga 13 fa riferimento a un cognato di "plavtad" definito alle righe 15-21, come la parola latina "plautus". Vale la pena notare che, in conformità col principio dei LD, e per evitare la creazione di risorse di dati isolate, i cognati latini, così come gli etimi e, quando rilevanti, le etimologie, sono collegati alla knowledge base LiLa, rispettivamente alla *LiLa Lemma Bank*[51] [35] e all'*EDLIL*[52] [28]. Un esempio è mostrato alla riga 21, in cui il collegamento con la *LiLa Lemma Bank* è implementato utilizzando la relazione di *seeAlso*. Ovviamente l'inserimento di un link esterno è, in genere, alternativo al fornire una descrizione del cognato, a meno che non siano informazioni che vanno ad integrare quanto già descritto dall'entità riferita.

Per quanto riguarda la parte semantica, poiché per le *Restsprachen* spesso non è possibile ricostruire con precisione il contenuto semantico delle parole, i significati forniti sono per lo più generici e le voci presentano generalmente un solo significato. La sua codifica, come mostrato dalle righe 22-28, è quindi ridotta al minimo: viene specificata una definizione, e viene associato un campo semantico al significato. Per quest'ultimo scopo è stata creata una tassonomia SKOS dei campi semantici basata sull'elenco di campi semantici di Buck [8][53]. Tra i lavori relativi alla

[50] Zotero è uno strumento gratuito che aiuta a raccogliere, organizzare, annotare, citare e condividere la ricerca. https://www.zotero.org/ (ultimo accesso: 16/06/2025)

[51] https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/OPEN-532 (ultimo accesso: 16/06/2025)

[52] https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/OPEN-533 (ultimo accesso: 16/06/2025)

[53] Le classi della tassonomia SKOS vengono rappresentate in LexO-server come oggetti di tipo `LexicalConcept`, una classe del modello OntoLex e vengono poi collegate ai sensi con le

semantica indoeuropea, quella di Buck è una delle poche opere ad aver organizzato il lessico indoeuropeo secondo categorie, seguendo una struttura tassonomica.

Le righe 24-26 definiscono la classe "Abbigliamento e ornamento" come campo semantico del significato "suola della scarpa". La riga 27 stabilisce infine che la classe creata è esattamente quella definita nella tassonomia originaria di Buck.

Nel caso d'uso esaminato, i modelli *Lexicog*, *OntoLex-Lemon* e LexInfo si sono rivelati adeguati alla rappresentazione strutturata delle diverse componenti del dizionario. Le peculiarità di natura storica e bibliografica che caratterizzano il progetto hanno richiesto l'integrazione di ulteriori elementi descrittivi, in particolare quelli relativi all'etimologia e alla documentazione delle fonti, gestiti nativamente attraverso i servizi informatici sviluppati appositamente. Per quanto riguarda invece la categorizzazione dei temi delle voci, si è reso necessario introdurre una proprietà personalizzata, la cui implementazione ha comportato un lieve adattamento del codice dei servizi. Il software presentato nel presente contributo è stato infine impiegato anche per lo sviluppo dell'interfaccia utente finale[54], dedicata alla fruizione del dizionario costruito mediante la piattaforma EpiLexO.

6. Conclusioni

Con l'evoluzione del Web Semantico e l'adozione del paradigma dei Linked Data, la lessicografia digitale ha progressivamente integrato nuovi modelli per la rappresentazione e l'interconnessione dei dati linguistici. In questo contesto, il modello *Lexicog*, proposto dal gruppo W3C OntoLex, si è affermato come una soluzione efficace per la descrizione strutturata e interoperabile delle risorse lessicografiche, in conformità con i principi FAIR. Il presente contributo ha inteso valutare l'efficacia e la flessibilità di tale modello attraverso l'analisi di tre casi d'uso, affiancando alla riflessione teorica lo sviluppo di un insieme di servizi software open-source finalizzati alla costruzione di dizionari digitali aperti e sostenibili. Come emerso dagli esempi pratici analizzati, la realizzazione di un dizionario digitale non si esaurisce nella semplice codifica di lemmi e definizioni, ma può richiedere l'integrazione di molteplici tipologie di informazioni: dati morfologici e sintattici, varianti ortografiche, etimologie, attestazioni d'uso in corpora testuali, equivalenti in altre lingue, nonché metadati relativi alla provenienza e alla qualità delle fonti. La natura e la granularità di tali informazioni possono variare in funzione del tipo di risorsa da compilare, che si tratti di un dizionario storico, bilingue, terminologico o descrittivo. Lexicog si è dimostrato in grado di gestire la complessità dei vari tipi di dizionari. Da un lato il modello fornisce una struttura base che riesce ad essere generale per i vari tipi di dizionario presentati, dall'altro, grazie al principio del riuso del paradigma dei Linked Data, il contenuto di tale struttura può essere specializzato attraverso l'utilizzo di vocabolari e ontologie che modellano aspetti specifici, quali l'etimologia, la morfologia, il multilinguismo, gli esempi d'uso e le attestazioni.

Dal punto di vista informatico è stato sviluppato un insieme di servizi REST, chiamato LexO-server, che implementa tutte le componenti del modello *Lexicog* e si integra con altri modelli di supporto per la completa gestione dei principali aspetti di un dizionario. La natura modulare e orientata ai servizi del software presentato, permette sia di offrire un solido back-end per lo sviluppo di interfacce front-end rivolte ai lessicografi – siano esse di visualizzazione di dati che

opportune proprietà. Per approfondimenti fare riferimento a https://www.w3.org/2016/05/ontolex/#lexical-concept (ultimo accesso: 16/06/2025)

[54] https://www.ilc.cnr.it/digitant/ (ultimo accesso: 16/06/2025)

di editing degli stessi, sia di potersi facilmente integrare all'interno di pipeline di elaborazione linguistico-computazionale. Tuttavia, come spesso accade nell'ambito applicativo, ciascun caso d'uso può presentare specificità tali da rendere necessario l'adattamento o l'estensione dei modelli esistenti. Alcuni aspetti lessicografici, infatti, possono risultare non immediatamente rappresentabili con le sole risorse ontologiche standard. Come evidenziato nei casi discussi nella sezione 5, il paradigma dei Linked Data consente comunque di sviluppare soluzioni personalizzate, integrabili armonicamente con i modelli di riferimento. La disponibilità open source del codice dei servizi proposti, consente la conseguente modifica e personalizzazione dei servizi in funzione delle esigenze specifiche del progetto lessicografico.

7. Ringraziamenti

progetto VocaBO - "Vocabolario di Boccaccio Online"; progetto DiTAMO "Dictionary of Old Occitan medico-botanical terminology"; progetto PRIN 2017XJLE8J "Languages and Cultures of Ancient Italy - Historical Linguistics and Digital Models"; progetto "Rut - modelli, risorse, metodologie e strumenti per la rappresentazione di risorse terminologiche e ontologiche".

References

[1] Baader, F., ed. 2003. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge: Cambridge University Press.

[2] Baader, F., I. Horrocks, and U. Sattler. 2008. "Description Logics." In *Foundations of Artificial Intelligence*, vol. 3, edited by Frank van Harmelen, Vladimir Lifschitz, and Bruce Porter, 135–179. Elsevier.

[3] Bellandi, Andrea. 2023. "Building Linked Lexicography Applications with LexO-Server." *Digital Scholarship in the Humanities* 38 (3): 937–52. https://doi.org/10.1093/llc/fqac095

[4] Bellandi, Andrea. 2021. "LexO: An Open-Source System for Managing OntoLex-Lemon Resources." *Language Resources and Evaluation* 55 (4): 1093–1126. https://doi.org/10.1007/s10579-021-09546-4

[5] Bekiari, C., M. Doerr, P. Le Boeuf, and P. Riva. 2017. *Definition of FRBRoo: A Conceptual Model for Bibliographic Information in Object-Oriented Formalism (Version 2.4)*. International Federation of Library Associations and Institutions (IFLA). https://repository.ifla.org/handle/20.500.14598/659.

[6] Berners-Lee, T., Hendler, J., Lassila, O. 2001. "The semantic web." *Scientific American* 284 (5): 34-43.

[7] Bizer, Christian, Tom Heath, and Tim Berners-Lee. 2009. "Linked Data: The Story so Far." *International Journal on Semantic Web and Information Systems* 5 (July):1–22. https://doi.org/10.4018/jswis.2009081901

[8] Buck, C. D. 1949. *A Dictionary of Selected Synonyms in the Principal Indo-European Languages: A Contribution to the History of Ideas*. Chicago: University of Chicago Press.

[9] Cabré, M. T. 1999. *Terminology. Theory, Methods and Applications*. Amsterdam/Philadelphia: John Benjamins.

[10] Chiarcos, C., K. Gkirtzou, M. Ionov, B. Kabashi, A. F. Khan, and C. O. Truică. 2022. "Modelling Collocations in OntoLex-FrAC." In *Proceedings of GlobaLex-2022*, Marseille, France.

[11] Cimiano, P., P. Buitelaar, J. McCrae, and M. Sintek. 2011. "LexInfo: A Declarative Model for the Lexicon-Ontology Interface." *Journal of Web Semantics* 9 (1): 29–51. https://doi.org/10.1016/j.websem.2010.11.001

[12] Corradini, M. Sofia and Mensching, Guido (2017): "Le DiTMAO (Dictionnaire des Termes Médico-botaniques de l'Ancien Occitan): caractères et organisation des données lexicales", in: Carrera, Aitor & Grifoll, Isabel (eds.) : *Occitània en Catalonha. De tempses novèls, de novèlas perspectivas*. Actes de l'XIen Congrès de l'Associacion Internacionala d'Estudis Occitans, Lhèida, 16-21 june 2014. Lhèida : Generalitat de Catalunya / Institut d'Estudis Ilerdencs, 125-138.

[13] Corradini, M. Sofia, Mensching, Guido, and Zwink, Julia (2021): "Le DiTMAO (Dictionnaire des termes médico-botaniques de l'ancien occitan) – innovations et évolution récente", in: Courouau, Jean-François (éd.) : *Fidélités et dissidences*. Actes du XIIe Congrès de l'Association Internationale d'Etudes Occitanes, Albi, 10-15 juillet 2017. Toulouse: SFAIEO, II, 907-920.

[14] Corradini, M. Sofia, and Mensching, Guido (sous presse): "L'apport original du DiTMAO à la lexicographie scientifique de l'ancien occitan", in : *L'occitan à la rencontre des études romanes*, Actes du XIVe Congrès de l'Association Internationale d'Etudes Occitanes, Munich 11-16 septembre 2023. Open Publishing LMU.

[15] Costa, Rute, Ana Salgado, and Bruno Almeida. 2021. "SKOS as a Key Element for Linking Lexicography to Digital Humanities." In *Information and Knowledge Organisation in Digital Humanities*, by Koraljka Golub and Ying-Hsang Liu, 1st ed., 178–204. London: Routledge. https://doi.org/10.4324/9781003131816-9

[16] Della Valle, Valeria. 2024. *Dizionari italiani: storia, tipi, struttura*. Roma: Carocci, collana Bussole.

[17] De Mauro, Tullio. 2005. *La fabbrica delle parole: il lessico e i problemi di lessicologia*. Torino: UTET Università.

[18] Depecker, L. 2002. *Entre signe et concept: Éléments de terminologie générale*. Paris: Presses Sorbonne Nouvelle.

[19] Diez Platas, M. L., H. Bermúdez, S. Ros, E. González-Blanco, O. Corcho, O. K. Gómez, L. Hernández-Lorenzo, M. De Sisto, J. de la Rosa, Á. Pérez, A. Diez, e J. L. Rodriguez. 2022. "Description of Postdata Poetry Ontology V1.0." In *Tackling the Toolkit: Plotting Poetry through Computational Literary Studies*, a cura di P. Plecháč, R. Kolár, A.-S. Bories, e J. Říha, 19–34. Prague: Institute of Czech Literature of the Czech Academy of Sciences.

[20] Doerr, M. 2003. "The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata." *AI Magazine* 24(3): 75.

[21] Francis, W. N., and H. Kučera. 1964. *A Standard Corpus of Present-Day Edited American English, for Use with Digital Computers (Brown)*. Providence, RI: Brown University. Revised 1971, 1979.

[22] Giovannetti, E., D. Albanesi, A. Bellandi, S. Marchi, M. Papini, and F. Sciolette. 2024. "Maia: An Open Collaborative Platform for Text Annotation, E-Lexicography, and Lexical Linking." *Umanistica Digitale* (18): 27–52.

[23] Khan, A. F., and F. Boschetti. 2018. "Towards a Representation of Citations in Linked Data Lexical Resources." In *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*, 137–147.

[24] Khan, A. F. 2018a. "Towards the Representation of Etymological Data on the Semantic Web." *Information* 9(12): 304.

[25] Khan, A. F. 2018b. "Towards the Representation of Etymological and Diachronic Lexical Data on the Semantic Web." In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan*.

[26] Khan, A. F., and A. Salgado. 2021. "Modelling Lexicographic Resources Using CIDOC-CRM, FRBRoo and Ontolex-Lemon." In *SWODCH*, 1–12.

[27] Mallia, M., M. Bandini, and V. Quochi. 2024. "An Interface for Linking Ancient Languages." *Cybernetics and Information Technologies* 24(4).

[28] Mambrini, F., and M. Passarotti. 2020. "Representing Etymology in the LiLa Knowledge Base of Linguistic Resources for Latin." In *Proceedings of the 2020 Globalex Workshop on Linked Lexicography*, 20–28.

[29] Mambrini, F., F. M. Cecchini, G. Franzini, E. Litta, M. C. Passarotti, and P. Ruffolo. 2020. "LiLa: Linking Latin Risorse linguistiche per il latino nel Semantic Web (AIUCD 2019)." *Umanistica Digitale* (8).

[30] Massariello Merzagora, Giovanna. *La Lessicografia*. 1st ed. Bologna: Nicola Zanichelli, 1983.

[31] McCrae, J. P., J. Bosque-Gil, J. Gracia, P. Buitelaar, e P. Cimiano. 2017. "The OntoLex-Lemon Model: Development and Applications." In *Proceedings of eLex 2017 Conference*, 19–21.

[32] Miller, George A. 1994. "WordNet: A Lexical Database for English." In *Human Language Technology: Proceedings of a Workshop Held at Plainsboro, New Jersey, March 8–11, 1994*.

[33] Muljačić, Žarko. 1971. *Introduzione allo studio della lingua italiana*. Torino: Einaudi.

[34] Murru, Chiara. 2019. "«Sanza che alla mia penna non dee essere meno d'autorità conceduta che sia al pennello del dipintore». Considerazioni sulla pittura di Giotto da Giovanni Boccaccio a Roberto Longhi." *Studi sul Boccaccio*, 12. http://digital.casalini.it/4606737.

[35] Passarotti, M., F. Mambrini, G. Franzini, F. M. Cecchini, E. Litta, G. Moretti, P. Ruffolo, and R. Sprugnoli. 2020. "Interlinking through Lemmas. The Lexical Collection

of the LiLa Knowledge Base of Linguistic Resources for Latin." *Studi e Saggi Linguistici* LVIII(1): 177–212.

[36] Peroni, S., and D. Shotton. 2012. "FaBiO and CiTO: Ontologies for Describing Bibliographic Resources and Citations." *Journal of Web Semantics* 17: 33–43.

[37] Peroni, S., and D. Shotton. 2018. "The SPAR Ontologies." In *Proceedings of the 17th International Semantic Web Conference (ISWC 2018)*, 119–136.

[38] Quochi, V., A. Bellandi, M. Mallia, A. Tommasi, and C. Zavattari. 2022. "Supporting Ancient Historical Linguistics and Cultural Studies with EpiLexO." In *CLARIN Annual Conference Proceedings*, vol. 39.

[39] Ricotta, V. 2019. "Con animi e con vocaboli onestissimi si convien dire. Prime attestazioni di hapax in Boccaccio." *Studi di lessicografia italiana* 36: 67–102.

[40] Wilkinson, M., et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (2016). https://doi.org/10.1038/sdata.2016.18