

Le Risorse Linguistiche nell'era del Web Semantico

Un insieme di servizi informatici per la gestione di lessici e terminologie

ANDREA BELLANDI*

ABSTRACT: The application of digital technologies in lexicography and terminology is becoming increasingly pervasive. The Semantic Web offers a great potential for representing data in order to process it, understand its meaning, and share it across different scientific communities. In this area scientific communities defined computational models for the representation of linguistic data together with best practices for their use. However, applications for managing this data in the context of the Semantic Web, are a few. This paper fills this gap by presenting a set of web services that are general enough to be used as a basis for the development of task-oriented lexicography and terminology applications, such as editing, data visualization, querying, and using for linguistic/conceptual full-text searches. The final part of the contribution will describe some use cases of the presented services.

Keywords: Semantic Web, Lexicography, Terminology, OntoLex-Lemon.

1. Introduzione

Le risorse linguistiche e le tecnologie digitali, hanno recentemente stretto un rapporto di simbiosi che, per molti aspetti, sembra ormai indissolubile. Le potenzialità della rete e delle tecnologie del Web Semantico possono aiutare a studiare la storia delle lingue, aumentando le nostre conoscenze e fornendo nuove modalità di accesso alle informazioni, capaci di interconnettere vocabolari e risorse digitali in funzione di nuovi orizzonti di ricerca (Arcidiacono 2019). La lessicografia diventa così lessicografia elettronica, o computazionale, e prevede la ristrutturazione e lo sfruttamento dei dizionari tradizionali ai fini computazionali; le risorse tradizionali, come il dizionario cartaceo, vengono riorganizzate, ristrutturate e rese trattabili da strumenti informatici. Diventa sempre più importante anche la rappresentazione digitale delle terminologie; dato il rapido sviluppo dell'economia, della cultura, della scienza e della tecnologia in tutto il mondo, il valore della terminologia come risorsa di comunicazione è indiscu-

* Istituto di Linguistica Computazionale "A. Zampolli", Consiglio Nazionale delle Ricerche - Pisa, Italy. andrea.bellandi@ilc.cnr.it.

tibilmente in aumento. Diventa quindi imprescindibile una rappresentazione delle risorse linguistiche in modo tale che le macchine possano: i) elaborarle, ii) comprenderne la semantica, iii) interrogarle, iv) condividerle tra le comunità scientifiche. Questi aspetti pongono delle sfide da affrontare attraverso la collaborazione e la standardizzazione su diversi fronti: linguaggi formali di rappresentazione, interoperabilità e modelli linguistici. Le tendenze attuali nelle risorse linguistiche e nella scienza del dato, consentono oggi di immaginare un ecosistema interconnesso, su larga scala, di un insieme di tecnologie e dati linguistici aperti, interrogabili e standardizzati (Monachini e Khan 2018).

Con l'avvento del Web Semantico, le ontologie sono diventate un metodo sempre più importante per modellare formalmente la semantica del dato e condividere la conoscenza di un dominio sul Web. In questo contesto, è stato sviluppato il modello *lemon* (McCrae et al. 2012; McCrae et al. 2017), (anche chiamato *OntoLex-Lemon*), basato sul linguaggio formale Ontology Web Language (OWL), con lo scopo di fornire una ricca base linguistica per le ontologie concettuali. Esso comprende la rappresentazione delle proprietà morfologiche e sintattiche delle voci lessicali, così come l'interfaccia sintassi-semantica, ovvero la rappresentazione del loro significato rispetto ai concetti di cui esse sono le lessicalizzazioni. Negli ultimi anni, anche i terminologi hanno iniziato ad adottare modelli sviluppati nell'ambito della lessicologia, per descrivere in modo più ricco le relazioni tra i termini, spostandosi dalla metodologia tradizionalmente usata per la descrizione dei termini incentrata più sulle relazioni tassonomiche e meronimiche¹. Molti dei modelli lessicali e terminologici proposti in letteratura, hanno trovato una loro corrispondenza in *OntoLex-Lemon*, dimostrando che il modello è in grado di catturare sia aspetti più lessicografici² (Bosque-Gil, Gracia, e Montiel-Ponsoda 2017) che terminologici³ (Cimiano et al. 2017). Vedremo nella sezione 2, come sia possibile riferire definizioni concettuali sia da un senso, che direttamente da una voce lessicale.

Nel presente articolo si offrono dei servizi informatici, chiamati LexO-server, per gestire risorse lessicografiche e terminologiche secondo il modello *OntoLex-Lemon*. I servizi proposti sono offerti come una Web API, che implementa funzionalità di scrittura, lettura e modifica di repertori lessicali o terminologici, sopra la quale è possibile costruire diverse interfacce utente. Questa separazione netta tra le funzioni sul dato e l'interfaccia utente, è alla

¹ Diversamente, i modelli lessicografici, basati su un approccio orientato alla parola, ed in particolare al senso, tengono conto di un insieme più ricco di relazioni rilevanti, ad esempio quelle lessico-semantiche.

² Vedere il modulo dedicato per la lessicografia al seguente indirizzo <https://www.w3.org/2019/09/lexicog/> (ultima consultazione: 10/03/2022).

³ Strumento per convertire database terminologici nel formato standard per le terminologie Term Base eXchange (TBX) in *OntoLex-Lemon*, <http://tbx2rdf.lider-project.eu/converter/index.html> (ultima consultazione: 10/03/2022).

base delle architetture software REST. I servizi qui presentati si rivolgono agli sviluppatori del dominio linguistico, che possono utilizzare LexO-server come backend delle loro applicazioni lessicali e terminologiche. Il presente lavoro si basa sull'esperienza maturata nello sviluppo dell'applicazione LexO⁴ (Bellandi 2021), un'interfaccia utente web e collaborativa, che rende fruibile il modello *OntoLex-Lemon* ad un vasto target di utenza, astruendo l'utente dalle conoscenze delle tecnologie del Web semantico.

L'articolo è strutturato come segue: nella sezione 2 si introduce il Web Semantico descrivendone caratteristiche e vantaggi; nell'ultima parte è presentato il modello linguistico alla base dei servizi. Nella sezione 3 viene descritto LexO-server, il backend di servizi lessicali, e vengono discussi i lavori simili. Nella sezione 4 si presentano tre casi d'uso mostrando altrettante applicazioni che si basano su LexO-server: un editor di lessici, un sistema di navigazione grafica delle relazioni semantiche di un lessico e un sistema di ricerca full-text basato su informazioni linguistiche.

2. Rappresentazione del dato linguistico nel Web Semantico

La presente sezione descrive i principi base del Web semantico e riassume i vantaggi derivanti dall'utilizzo delle tecnologie su cui esso si basa. In questo contesto viene successivamente presentato il modello linguistico *OntoLex-Lemon*, lo *standard de facto* per la rappresentazione formale di lessici e terminologie, creato nell'ambito del World Wide Web Consortium (W3C).

2.1. Caratteristiche e vantaggi del Web Semantico

Riprendendo quanto affermato in (Berners-Lee, Hendler, e Lassila 2001), «il Web semantico è un tentativo di *rappresentare e collegare* i dati (risorse linguistiche, nel nostro caso) in un modo che sia comprensibile per le macchine». Il compito di *rappresentazione* viene svolto mediante l'utilizzo di linguaggi formali basati su sottoinsiemi della logica del primo ordine, in particolare la Logica Descrittiva (Baader et al. 2017). I dati vengono resi *collegabili* attraverso il paradigma dei Linked Data (LD), consentendo a ciascuna entità di un dataset (concetti, relazioni, attributi e così via) di essere identificata in modo univoco da un Internationalized Resource Identifier (IRI) e di essere disponibile sul Web tramite il Protocollo di trasferimento ipertestuale (HTTP).

L'uso di un linguaggio di rappresentazione formale, come l'OWL, fornisce un modo comune per rappresentare e codificare la semantica dei dati garantendo al contempo l'interoperabilità; consente inoltre di eseguire ragionamenti automa-

⁴ Il codice di LexO è open source e disponibile al seguente indirizzo <https://github.com/andreabellandi/LexO-lite> (ultima consultazione: 10/03/2022).

tizzati sui dati, ad esempio per garantire coerenza logica, calcolare chiusure di relazioni (ad esempio la transitività dell'iperonimia) e dedurre nuove conoscenze sulla base di tassonomie di classi e gerarchie di proprietà. Il collegamento dei dati consente di federarsi con contenuti provenienti da risorse esterne, consentendo di dedurre nuovi fatti attraverso il web, per la scoperta di nuova conoscenza. Inoltre facilita il riutilizzo dei dati stessi, in virtù della natura intrinseca dei LD.

I vantaggi della rappresentazione del contenuto lessicografico e terminologico secondo gli standard del Web Semantico sono quindi la reperibilità, la riutilizzabilità, l'accessibilità e la visibilità su scala web, secondo i principi FAIR (Wilkinson et al. 2016). Negli ultimi anni, ciò ha portato a una serie di attività guidate da comunità che hanno promosso l'adozione del paradigma dei LD per la pubblicazione di dati linguistici, in particolare l'Open Linguistics Working Group (Chiarcos, Hellmann, e Nordhoff 2012) e l'Open Knowledge Foundation (Chiarcos, Hellmann, e Nordhoff 2011). L'approccio del paradigma dei LD facilita sia lo sviluppo distribuito della rete di risorse, sia la collaborazione tra i ricercatori che forniscono e utilizzano questi dati e che impiegano un insieme condiviso di tecnologie⁵.

Il paradigma dei LD richiede l'utilizzo del modello di dati Resource Description Framework (RDF). I dati devono essere rappresentati in una struttura a tripla, costituita da un soggetto, un predicato e un oggetto, dove il soggetto è una risorsa, il predicato è una proprietà e l'oggetto è una risorsa, ad esempio `<lemon, meronym, lemon_tree>` o un *datatype* (una stringa, un intero, un decimale, ecc.), ad esempio `<lexicon, hasPhonetic, "lɛksikən">`. Insieme a RDF, RDF Schema (RDFS) aiuta a descrivere istanze, gruppi di risorse correlate e proprietà che collegano queste risorse. Ad esempio, `<synonym, hasDomain, LexicalSense>` e `<synonym, hasRange, LexicalSense>`, significa che la relazione `synonym` collega i sensi lessicali tra loro; `<MultiWordExpression, subclassOf, LexicalEntry>`, significa che ogni espressione `multiword` è un tipo di voce lessicale; `<hypernym, inverseOf, hyponym>`, significa che la relazione di iperonimia è l'inversa della relazione di iponimia.

Le ontologie aggiungono una rappresentazione semantica degli elementi definiti con RDF, introducendo, per esempio, operatori logici come *and* e *or*, operatori insiemistici come *unione*, *intersezione*, *disgiunzione*, il quantificatore *esistenziale* e quello *universale*⁶. Una voce lessicale, ad esempio, può essere formalmente definita in OWL come una parola, un composto o un affisso che può essere caratterizzata da una specifica parte del discorso, un modello morfologico, un'etimologia, un insieme di sensi, avente almeno una forma ortografica e al massimo una

⁵ Vedi il Linguistic Linked Open Data Cloud, <https://linguistic-lod.org/> (ultima consultazione: 10/03/2022).

⁶ Per una completa definizione di OWL si rimanda al seguente link <https://www.w3.org/TR/owl2-overview/> (ultima consultazione: 10/03/2022).

forma canonica. Il modello linguistico formale, scritto in OWL, che andremo a presentare nella sezione successiva, contiene appunto questo tipo di definizioni.

2.2. *Il Modello linguistico*

Il modello LEXicon Model for ONtologies (*lemon*) (McCrae et al. 2017), sviluppato dal gruppo della comunità W3C OntoLex⁷, è diventato uno standard di fatto per rappresentare e pubblicare risorse lessicali nel Web Semantico. Il modello, spesso chiamato anche *OntoLex-Lemon*, nasce accogliendo un'ampia serie di casi d'uso rilevanti, e astruendo i requisiti generali da modellare. Tutti i problemi e le decisioni del modello sono pubblicamente disponibili e accessibili sul web⁸. L'architettura di *OntoLex-Lemon* è suddivisa in sei moduli, come illustrato nella Figura 1. Ciascun modulo tiene conto di aspetti specifici nella modellazione delle informazioni lessicali. Di seguito daremo una breve spiegazione di ogni modulo fornendo, laddove opportuno, dei piccoli esempi⁹.

Il modulo principale indicato come *CORE*, Fig. 1(a), si basa sul principio noto come “semantic by reference” (Buitelaar 2010), ovvero il modello permette di considerare la descrizione del comportamento morfologico e sintattico di una parola, come separata dalla descrizione ontologica dei concetti a cui la parola si riferisce. Gli elementi principali del *CORE* sono quattro: la classe `LexicalEntry` che rappresenta un insieme di voci lessicali classificate in base al loro tipo, cioè, parola singola (`Word`), parola composta (`MultiWordExpression`) o parte di una parola (`Affix`); la classe `Form` che istanzia tutte le forme flesse di una voce lessicale, incluso il lemma; la classe `LexicalSense` che rappresenta una reificazione tra la voce lessicale e il concetto; la classe `LexicalConcept` che permette di esprimere il fatto che una certa voce lessicale evochi un certo concetto mentale o un'unità di pensiero (ad esempio il verbo “*morire*” può evocare il concetto di “*morte*”). I concetti lessicali possono anche raggruppare insiemi distinti di sensi, come ad esempio i *synsets* in WordNet (Miller 1995). Per quanto riguarda la dimensione concettuale invece, in un'ottica semasiologica, in cui il senso è al centro della rappresentazione, il concetto può essere riferito dal singolo senso lessicale, tramite la proprietà *reference*. Nell'ambito di un approccio onomasiologico, in cui il

⁷ <https://www.w3.org/community/ontolex> (ultima consultazione: 10/03/2022).

⁸ I casi d'uso sono disponibili al seguente link https://www.w3.org/community/ontolex/wiki/Specification_of_Use_Cases (ultima consultazione: 10/03/2022); i requisiti generali sono disponibili al seguente link https://www.w3.org/community/ontolex/wiki/Specification_of_Requirements (ultima consultazione: 10/03/2022); problemi e decisioni del modello sono disponibili al seguente link <https://lists.w3.org/Archives/Public/public-ontolex/> (ultima consultazione: 10/03/2022).

⁹ Per una completa guida al modello riferirsi al seguente link <https://www.w3.org/2016/05/ontolex/> (ultima consultazione: 10/03/2022).

concetto è al centro della rappresentazione, quest'ultimo può essere denotato direttamente dall'entrata lessicale, tramite la proprietà *denote*.

Il modulo in Fig. 1(b), denominato *DECOMP*, è dedicato a rappresentare quali sono i componenti di un composto. È possibile modellare sia l'ordine che i tratti morfologici di ogni componente.

Tutte le relazioni tra le parole, sono modellate dal modulo Variazione e Traduzione nella Fig. 1(c), denominato *VARTRANS*. Vengono presi in considerazione due tipi di relazioni: le relazioni lessicali, che collegano direttamente tra loro voci lessicali, ad esempio `<LexO, acronymOf, Lexicon_and_Ontology>`¹⁰, e le relazioni tra sensi, che collegano direttamente tra loro i sensi lessicali, ad esempio la sinonimia. Se è necessario rappresentare informazioni aggiuntive sulla relazione, è possibile reificare la relazione stessa e modellarla come classe. In questo modo è possibile specificare alcune proprietà della relazione, ad esempio un grado di confidenza. Consideriamo il caso della traduzione (Gracia et al. 2014). Supponiamo che un algoritmo di allineamento automatico proponga che “casseroola” sia la traduzione italiana di “pot” in inglese con un grado di confidenza pari a 0,8. La classe *Translation* avrà un elemento *t* con un *source* `<t source casseroola>`, un *target* `<t target pot>` e un grado di confidenza `<t confidence 0,8>`.

¹⁰ *OntoLex-Lemon* è agnostico ad uno specifico vocabolario linguistico. Le categorie grammaticali, le relazioni semantiche, e tutti gli oggetti linguistici in genere, sono definiti in vocabolari esterni. Nel nostro caso, come da raccomandazione della comunità, LexO-server utilizza il vocabolario LexInfo <https://lexinfo.net/> (ultima consultazione: 10/03/2022).

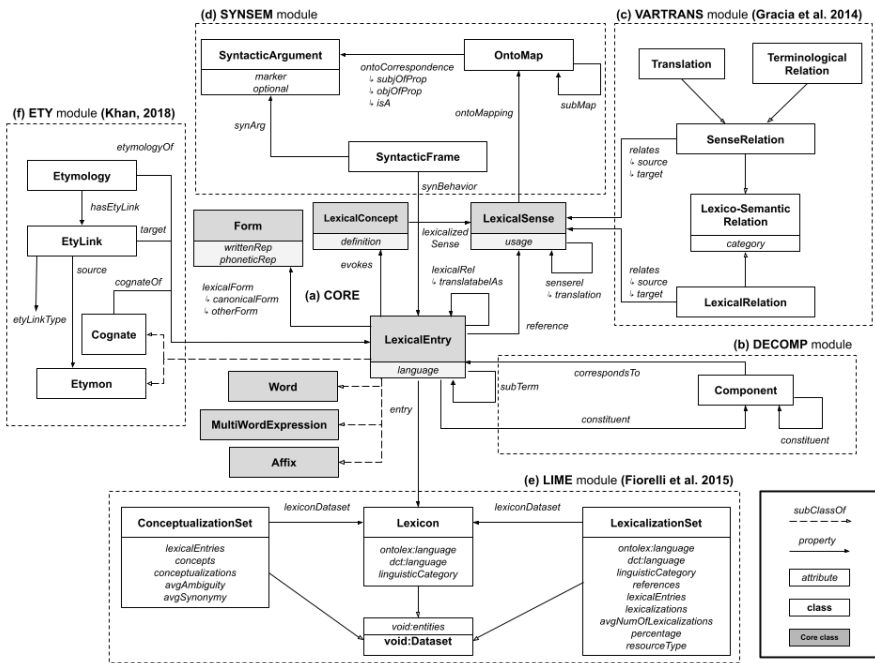


Figura 1: Architettura del modello *OntoLex-Lemon*.

Il modulo Sintassi e Semantica illustrato nella Fig. 1(d), denominato *SYNSEM*, consente di descrivere i frames sintattici di un elemento lessicale, specificando il tipo (frame verbale, frame transitivo, frame sostantivo preposizionale e così via), gli argomenti sintattici, le loro funzioni (soggetto, oggetto, ecc.), la loro posizione e l'eventuale obbligatorietà di presenza nel frame. Il modulo permette infine di far corrispondere ogni argomento del frame a un'entità di un'ontologia concettuale. Consideriamo ad esempio il frame transitivo “X possiede Y”. Il predicato può essere corrisposto su una proprietà ontologica “*possiede*”, dove il soggetto grammaticale X e il complemento oggetto Y si riferiscono, rispettivamente, al dominio e al codominio della proprietà, che saranno classi ontologiche. Il modulo è in grado di rappresentare anche strutture multi-argomentali più complesse.

Il modulo Metadata (Fig. 1(e)), denominato *LIME* (Fiorelli, Paziienza, e Stellato 2013), permette di rappresentare informazioni quantitative e qualitative a grana grossa sul dataset linguistico, ad esempio il numero di entrate, di sensi lessicali, di concettualizzazioni, i metadati descrittivi su un sottoinsieme di un insieme di lessicalizzazioni (per il sottoinsieme che rappresenta tutte le lessicalizzazioni per un certo tipo di entità concettuale riferita), o l'endpoint per accedere alla risorsa. Lo scopo principale è quello di consentire all'uomo e

alle macchine di conoscere quale materiale lessicale è disponibile e quindi di capire meglio come utilizzarlo per scopi specifici.

L'ultimo modulo che presentiamo (Fig. 1(f)), è denominato *ETY* e non fa ancora parte dell'architettura ufficiale di *OntoLex-Lemon* e gestisce l'etimologia. Il modulo prevede una prima semplice gestione dell'aspetto etimologico delle parole (Khan 2018). Le etimologie (classe *Etymology*) rappresentano la storia di una parola tracciando una sorta di albero genealogico linguistico o genealogia. Per questo la classe *LexicalEntry* del modulo *core* è estesa con due sottoclassi: la classe *Etymon* e la classe *Cognate*. Una parola può essere quindi associata ai suoi etimi e affini direttamente o indirettamente prendendo in esplicita considerazione altri etimi/affini attraverso relazioni che rappresentano dei processi storico-linguistici, rappresentati dalla classe *EtymLink*. Questi processi sono: il prestito, che si riferisce al processo mediante il quale gli elementi linguistici sono trasferiti da una lingua all'altra tramite il contatto linguistico (valore *borrowing* della proprietà *etymLinkType* in Fig. 1(f)), e l'eredità che si riferisce all'eredità di parole da una lingua madre, o a fasi precedenti della stessa lingua (valore *inheritance* della proprietà *etymLinkType* in Fig. 1(f)).

3. LexO-server

In questa sezione viene proposta una breve analisi dello stato dell'arte relativo alle applicazioni e ai servizi web disponibili per la gestione di risorse linguistiche, in particolare facendo riferimento a soluzioni che operano nel contesto del Web Semantico e col modello *OntoLex-Lemon*. Successivamente verrà fornita una descrizione dettagliata dell'architettura e del funzionamento dei servizi proposti.

3.1. Relazione con lavori simili

In generale, esistono diversi sistemi o servizi per il trattamento di risorse linguistiche, ognuno basato su modelli differenti e con scopi diversi, ad esempio per la costruzione di lessici di lingua, per la retro digitalizzazione di dizionari, o per la gestione di terminologie. Nel contesto del Web Semantico, a conoscenza dell'autore, il lavoro qui proposto è unico nel suo genere. Per quanto riguarda il modello *OntoLex-Lemon* infatti, esistono pochi strumenti dedicati alla sua fruizione, ma sono tutti applicazioni *full stack* che si rivolgono direttamente all'utente finale, al lessicografo o al terminologo e non servizi generici come quelli qui proposti, rivolti a chi costruisce applicazioni finali. Daremo di seguito una breve descrizione di tali strumenti. *Lemon source*¹¹ è uno strumento in stile Wiki per la manipolazione e la pubblicazione di risorse

¹¹ <https://lemon-model.net/download/source.php> (ultima consultazione: 10/03/2022).

lemon. Consente di caricare un lessico e condividerlo con altri utenti. È un progetto open source, ed è disponibile gratuitamente online per l'uso. Tuttavia, gestisce versioni vecchie del modello *OntoLex-Lemon* e sembra essere uno strumento non più mantenuto. Lemonade (Rico e Unger 2015), è un editor di *lemon* basato su un insieme di *lemon patterns* (McCrae e Unger 2014), una grammatica che consente di costruire entrate lessicali esprimendole in un semplice linguaggio formale orientato all'utente, senza occuparsi della loro codifica in OWL. Il linguaggio però non permette di specificare le relazioni tra voci lessicali o tra sensi lessicali. VocBench (Stellato et al. 2017) è uno strumento web e collaborativo per costruire ontologie OWL, thesauri Simple Knowledge Organisation System (SKOS) e dataset RDF generici multilingue. Tra le altre caratteristiche offre anche la funzionalità di editing per risorse *OntoLex-Lemon*. Infine LexO (Bellandi 2021) è un sistema dedicato all'editing di risorse *OntoLex-Lemon* con un'interfaccia utente che astrae le complessità del modello allo studioso. Una comparazione tra LexO e VocBench è discussa dagli autori in (Fiorelli et al. 2020).

Esistono invece lavori che propongono servizi Representational State Transfer (REST) simili a quelli qui descritti¹², ma che non si basano sul modello *OntoLex-Lemon*. Ad esempio, un set di API¹³ è stato sviluppato da K Dictionaries¹⁴ (ex Kernerman Dictionaries), per la fruizione di molti dizionari per lo più bilingui in oltre 50 lingue, inclusa una serie di dizionari per studenti apprendenti di varie lingue. L'Università di Oxford ha invece sviluppato una API¹⁵ che contiene diversi dizionari di inglese monolingue (destinati a madrelingua inglese) e dizionari bilingue in oltre 30 lingue abbinate all'inglese. Infine l'Università di Cambridge ha sviluppato una API¹⁶ per l'accesso al loro dizionario monolingua Inglese per studenti. Tutti questi servizi sono comunque orientati esclusivamente all'accesso e alla fruizione di dizionari mono e pluri lingue già esistenti, mentre la proposta dell'autore è focalizzata anche sull'offerta di servizi per la costruzione, in modo da servire un ampio set di possibili tasks per lo sviluppo di applicazioni dedicate, per la lessicografia e la terminologia, da parte di terzi. Un altro aspetto che caratterizza e distingue LexO-server dai servizi sopra citati, è la possibilità di riferire ontologie extra-linguistiche per assegnare una descrizione formale al senso o direttamente all'entrata, permettendo quindi anche un accesso concettuale alla risorsa terminologica e lessicale.

¹² Un'interessante rassegna può essere consultata al link <https://www.lexiconista.com/dictionary-apis/> (ultima consultazione: 10/03/2022).

¹³ <https://api.lexicala.com/> (ultima consultazione: 10/03/2022).

¹⁴ <https://lexicala.com/> (ultima consultazione: 10/03/2022).

¹⁵ <https://developer.oxforddictionaries.com> (ultima consultazione: 10/03/2022).

¹⁶ <https://dictionary-api.cambridge.org> (ultima consultazione: 10/03/2022).

3.2. Architettura e struttura dei servizi

LexO-server è una API REST che espone un insieme di funzionalità (o servizi) che permettono di essere invocate da programmi esterni via rete e, a seguito di un'elaborazione, restituiscono un risultato. Nello specifico LexO-server si basa sul protocollo HTTP e scambia i dati in formato JavaScript Object Notation (JSON)¹⁷. I servizi sono conformi a OpenAPI¹⁸, una specifica per file di interfaccia leggibili dalle macchine per descrivere, produrre, consumare e visualizzare servizi REST¹⁹.

I servizi sono raggruppati nelle seguenti categorie, in base alla tipologia di funzionalità che realizzano:

- *Linguistic Vocabulary*. Questo gruppo di servizi è volto a fornire la lista dei valori del vocabolario linguistico utilizzato da LexO-server (*LexInfo* come descritto in sezione 2.2), ad esempio le categorie grammaticali, le relazioni semantiche, i tratti morfologici, i tipi di frames sintattici, ecc. Lo scopo di questi servizi è quello di permettere alle interfacce delle applicazioni che li useranno di mostrare tali valori come possibili scelte, in opportuni menù in fase di editing della risorsa.
- *Lexicon Creation*. I servizi di questo gruppo permettono di creare tutti gli elementi del modello presentato in sezione 2.2. Per ognuno di essi, i servizi registrano l'utente creatore e la data di creazione (si veda in seguito la gestione degli utenti). Le entrate lessicali vengono gestite con uno stato a tre valori: i) *working*, per indicare che l'utente ha creato un'entrata e che questa è in fase di lavorazione (inserimento forme, sensi, ecc.); ii) *completed*, per indicare che l'utente ha terminato la compilazione dell'entrata; iii) *revised*, per indicare che l'entrata è definitiva e non è più modificabile. Questa gestione può essere utile nel contesto dello sviluppo di editors collaborativi in cui, ad esempio, è prevista una figura di utente revisore che ha il compito di validare le entrate redatte da utenti lessicografi o terminologi.
- *Lexicon Data*. Questo gruppo colleziona tutti i servizi per accedere ai dati della risorsa. È previsto un recupero dei dati di ogni elemento sia di sintesi, cosicché ad esempio un'interfaccia possa presentarli come liste o come alberi, sia di dettaglio, ad esempio per la loro visualizzazione in tabelle o forms per l'editing. I servizi di ricerca del dato permettono la specifica di parametri di input, che implementano ricerche avanzate.
- *Lexicon Update*. I servizi di questo gruppo permettono di modificare gli elementi del modello, aggiungendo o modificando proprietà e valori degli stessi.

¹⁷ <https://www.json.org/json-en.html> (ultima consultazione: 10/03/2022).

¹⁸ <https://www.openapis.org/> (ultima consultazione: 10/03/2022).

¹⁹ <https://licodemo.ilc.cnr.it/LexO-backend-beta/> (ultima consultazione: 10/03/2022).

- *Lexicon Deletion*. I servizi di questo gruppo permettono di eliminare elementi o singole proprietà. Al momento i servizi adottano una politica di cancellazione *lazy*, nel senso che se un elemento da eliminare è collegato ad altri elementi della risorsa, il servizio non procede con la cancellazione ma ritorna un codice di errore.
- *Lexicon Statistics*. Questo gruppo di servizi può essere di supporto a differenti tasks, fornendo alcuni elementi quantitativi della risorsa (numero di lingue, numero di voci per lingua, numero di parole per tipo, ecc.)
- *Graph Visualization Support*. I servizi implementano funzionalità tipiche della navigazione di un grafo. Al momento viene ipotizzato che i nodi siano i sensi e gli archi tutte le relazioni possibili tra sensi²⁰. Alcune relazioni possono avere delle caratteristiche specifiche, ad esempio la sinonimia può essere simmetrica, l'iperonimia transitiva, l'olonimia l'inversa della meronimia, ecc. Di conseguenza, i servizi specificano anche se l'informazione relativa ad una relazione è inferita o esplicita. Questo può essere utile in un task di visualizzazione perché, ad esempio, un arco potrebbe essere colorato in maniera diversa a seconda della natura della relazione che rappresenta.
- *Query Expansion Support*. I servizi di questo gruppo sono funzionali alla ricerca full-text e quindi prevedono l'espansione di: un'entrata con le sue forme flesse e i relativi tratti morfologici; un concetto con le forme dei sensi che lo riferiscono; un senso con le sue relative forme.

Per quanto riguarda la parte concettuale, LexO-server offre la possibilità di costruire un'ontologia SKOS²¹. SKOS consiste in una serie di classi RDFS e proprietà di RDF che sono usate per rappresentare il contenuto e la struttura di base dei cosiddetti concept-schema cioè thesaurus, schemi di classificazione, elenchi di titoli di argomento, tassonomie, glossari e altri tipi di vocabolario. La classe `LexicalConcept` del modulo core di *OntoLex-Lemon* è definita come una classe di SKOS e quindi i concetti lessicali possono essere strutturati tra di loro, ad esempio, tramite relazioni di gerarchia. LexO-server comunque gestisce il vocabolario SKOS anche indipendentemente dalla formalizzazione del concetto lessicale e permette quindi di costruire ontologie concettuali in SKOS, che possono rappresentare i concetti le cui lessicalizzazioni sono descritte tramite il modello linguistico.

Come riportato in Fig. 2, i componenti principali dell'architettura sono, i) la parte di logica dei servizi, ii) il repository semantico e iii) un insieme di

²⁰ Al momento i servizi ritornano oggetti secondo il formato della libreria Cytoscape <https://cytoscape.org/>. È una delle librerie free e open source più utilizzate per la costruzione di sistemi di visualizzazione (ultima consultazione: 10/03/2022).

²¹ Simple Knowledge Organisation System (SKOS) <https://www.w3.org/2004/02/skos/> (ultima consultazione: 10/03/2022).

servizi REST esterni e integrabili in LexO-server, tramite dei dati di configurazione.

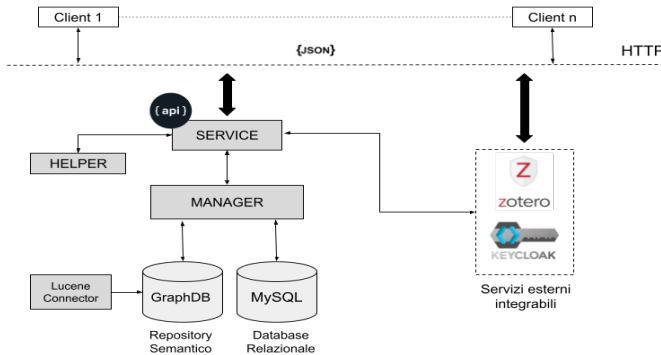


Figura 2: Architettura di LexO-server.

Riguardo la parte di logica, i servizi sono scritti in Java 15²². Ogni servizio effettua la verifica sintattica della propria invocazione, validando gli eventuali parametri in input con l'aiuto del modulo di Helper. Il servizio poi demanda l'elaborazione della funzionalità che esso implementa, al modulo del Manager che si interfaccia con il repository. Quest'ultimo risponde al Manager e, in base al tipo di risposta, il manager chiede al modulo di Helper di preparare il dato di risposta nella struttura dati opportuna. L'Helper trasmette il JSON al servizio, che poi lo invia al chiamante. In caso di problemi o errori, il servizio risponde con il relativo codice HTTP di stato.

GraphDB è la tecnologia scelta per la realizzazione del repository semantico. È un'implementazione dell'interfaccia di storage e reasoning del framework RDF4J. Gestisce basi di conoscenza in OWL 2, e ha delle ottime performance di scalabilità su dataset contenenti bilioni di triple. Si integra con Apache Lucene²³, un sistema di indicizzazione del testo, grazie al quale si ottengono alte prestazioni in fase di interrogazione e recupero di informazioni. I servizi non fanno riferimento ad una particolare istanza del repository. Chi utilizza i servizi dovrà installare un'istanza di GraphDB, anche su una macchina fisica o virtuale diversa da quella dove sono installati i servizi, e configurare il proprio repository, vuoto o importando una risorsa *OntoLex-Lemon* esistente.

LexO-server è attualmente integrabile con Zotero²⁴, un gestore di record bibliografici e con KeyCloack²⁵, un manager per la gestione degli utenti. Per quanto riguarda la bibliografia, ogni entità lessicale (entrata, forma, senso ed etimologia) può essere associata ad elementi bibliografici presenti in una li-

²² <https://openjdk.java.net/projects/jdk/15/> (ultima consultazione: 10/03/2022)

²³ <https://lucene.apache.org> (ultima consultazione: 10/03/2022).

²⁴ <https://www.zotero.org> (ultima consultazione: 10/03/2022).

²⁵ <https://www.keycloak.org/> (ultima consultazione: 10/03/2022).

breria di Zotero, tramite la sua API REST²⁶. Per quanto riguarda gli utenti, la cui anagrafica, i diritti e i permessi sono gestiti da KeyCloack, i servizi di LexO-server possono essere configurati in modo tale che richiedano un'autenticazione basata sulla base dati di KeyCloack, interfacciandosi con la sua API REST²⁷. La parte di autorizzazione invece (diritti e permessi), richiede un po' di lavoro in più rispetto alla semplice configurazione, ovvero la definizione dei ruoli utente e dei diritti sulle risorse, che dipendono dall'applicazione finale che andrà ad essere sviluppata. Il codice di ogni servizio dovrà essere modificato di conseguenza.

Alla luce di quanto detto, in un'apposita tabella di configurazione gestita in MySQL di LexO-server, è possibile specificare, tra gli altri:

- l'endpoint del repository semantico (GraphDB);
- l'endpoint della libreria *Zotero* al quale ci si vuole connettere;
- l'endpoint di *Keycloak* per la gestione degli utenti;
- un parametro per stabilire se i servizi sono aperti o il loro utilizzo deve richiedere un'autenticazione;
- il tipo di lessicalizzazione dei concetti SKOS (utilizzo della proprietà *label* del vocabolario RDFS o utilizzo delle proprietà *prefLabel*, *altLabel* e *hiddenLabel* del vocabolario SKOS);
- l'insieme delle lingue ammissibili per le lessicalizzazioni degli elementi di SKOS e la lingua di default per la loro visualizzazione.

I servizi funzionano con un web server come Apache Tomcat o Jetty. Nei casi d'uso che saranno presentati nella sezione successiva, i servizi utilizzano Tomcat 9.03.

4. Casi d'uso

In questa sezione vengono presentati tre strumenti informatici che utilizzano LexO-server. Ognuno di essi è dedicato a un caso d'uso diverso: il primo, sviluppato nel contesto del P.R.I.N. "Lingue e culture dell'Italia antica: linguistica storica e modelli digitali"²⁸, è dedicato alla creazione di lessici multilingua; il secondo permette una visualizzazione sofisticata delle relazioni tra sensi di un lessico; il terzo, sviluppato all'interno del "Progetto di Traduzione del Talmud Babilonese"²⁹, permette la ricerca su base linguistica all'interno di un

²⁶ https://www.zotero.org/support/dev/web_api/v3/basics (ultima consultazione: 10/03/2022).

²⁷ <https://www.keycloak.org/docs-api/15.0/rest-api/index.html> (ultima consultazione: 10/03/2022).

²⁸ <https://www.prin-italia-antica.unifi.it/> (ultima consultazione: 10/03/2022).

²⁹ <https://www.talmud.it/> (ultima consultazione: 10/03/2022).

testo. Le tre sottosezioni seguenti sono dedicate ad una breve descrizione dei sopra elencati sistemi.

4.1. Editing

Lo strumento chiamato EpiLexO (Quochi et al. 2022), mostrato in Fig. 3, è un editor web collaborativo di lessici storici multilingua. Nello specifico rappresenta un esempio di utilizzo di LexO-server per la costruzione di un lessico a partire da epigrafi. Come mostrato in Fig. 3, la colonna alla sinistra dell'interfaccia mostra la struttura del lessico con alcuni dati di sintesi come lo stato di editing della voce lessicale e l'utente che l'ha creata. In alto è possibile effettuare delle ricerche avanzate in base ad alcuni parametri. Questa è la parte di interfaccia che utilizza i servizi del gruppo Lexicon Data. Al centro ci sono i dati di dettaglio delle entità selezionate nella colonna di sinistra. Questi possono essere creati, modificati ed eliminati tramite i servizi dei gruppi *Lexicon Create*, *Lexicon Update*, e *Lexicon Deletion*. Nella colonna più a destra vengono editate le informazioni di contesto di ciò che è stato selezionato nella colonna di sinistra come la bibliografia, i links esterni, le note e alcuni metadati. Anche questa parte dell'interfaccia utilizza gli stessi gruppi di servizi usati dalle altre.

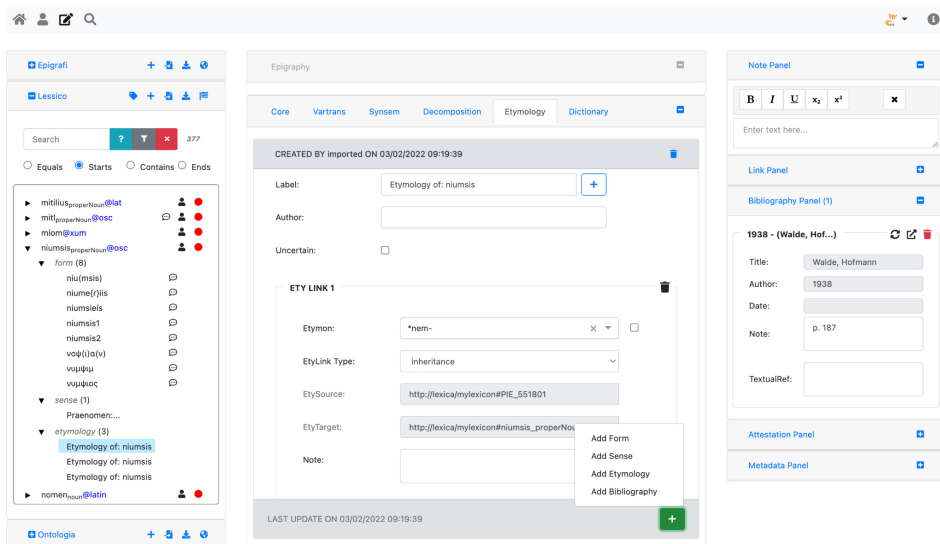


Figura 3: EpiLexO, un editor di lessici.

In questo caso i servizi sono configurati anche col gestore degli utenti (*Keycloak*); l'utilizzatore del servizio dovrà quindi autenticarsi per poter invocare LexO-server.

4.2. Visualizzazione

Il lavoro presentato in (Colombo e Giovannetti 2022), è un primo esperimento di strumento per permettere la navigazione di “PAROLE-SIMPLE-CLIPS” (PSC), un lessico computazionale dell’italiano, sviluppato dal 1996 al 2003 dell’Istituto di Linguistica Computazionale “A. Zampolli” (Rumy et al. 2002). Lo strumento mostra i sensi lessicali e le loro relazioni tramite un grafo. Come riportato in Fig. 4, l’interfaccia si divide in due parti. Nella colonna a sinistra viene mostrata la lista dei sensi disponibili nel lessico con la possibilità di filtrarli in base a vari parametri. Questa parte di interfaccia utilizza i servizi del gruppo Lexicon Data. Al centro è possibile navigare graficamente i sensi in base alle loro relazioni. Nell’esempio di Fig. 4 è mostrato un esempio di relazioni di meronimia e olonimia del senso di “nocca” e “unghia”. La relazione che i due sensi sono meronimi di “dito” è esplicita. In questo caso viene chiesto al sistema di ricostruire, tramite le regole di inferenza, la relazione di meronimia fino al senso di braccio (visualizzata in un colore diverso). I servizi relativi a questo tipo di navigazione sono quelli del gruppo *Graph Visualization Support*.

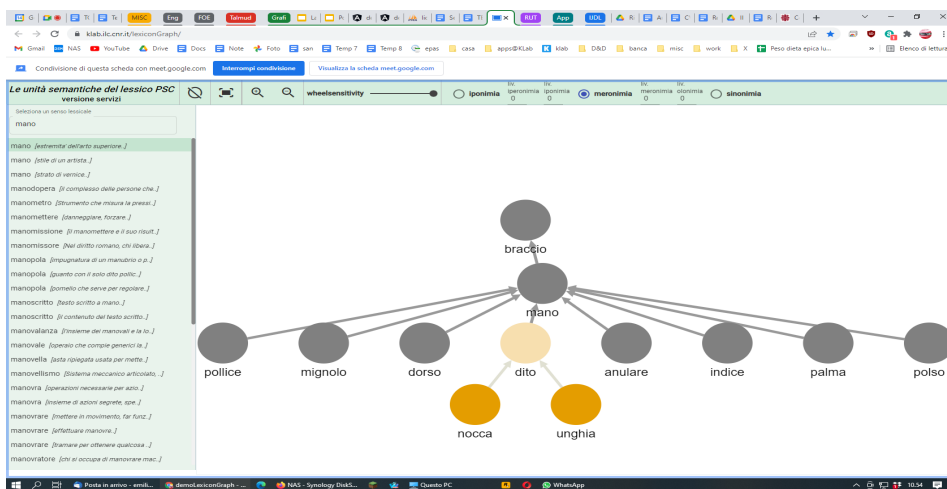


Figura 4: Visualizzatore di relazione semantica.

4.3. Supporto alla ricerca full-text

In (Giovannetti et al. 2021), gli autori hanno sviluppato uno strumento per accedere ad un testo su base linguistica e concettuale. In questo caso vengono utilizzati i servizi del gruppo Lexicon Data e quelli del gruppo Query Expansion Support. In Fig. 5 viene mostrato un esempio di ricerca linguistica sul testo. Nell’ambito del “Progetto di Traduzione del Talmud Babilonese”

i testi utilizzati sono le traduzioni italiane di alcuni trattati del testo sacro della cultura ebraica³⁰. In questo caso, l'utente chiede i contesti della parola "recipiente". Come mostrato in Fig. 5, l'interfaccia propone le voci lessicali ambigue ritornate da LexO-server. L'utente seleziona i sensi desiderati della categoria grammaticale "nome" e specifica che le forme desiderate devono essere al plurale. LexO-server espande l'interrogazione con gli iponimi diretti delle forme scelte (Fig. 5 in basso) e invia l'elenco delle forme risultanti al modulo di ricerca nel testo. Nella parte destra dell'interfaccia è possibile vedere alcuni contesti restituiti al sistema. I servizi del gruppo *Query Expansion Support*, partendo dal lemma "recipiente" hanno permesso di ottenere risultati come "le anfore ...", "... proprietari di barili ...", "... bicchieri colmi di vino ...". La risorsa linguistica usata è PSC. Siccome, ogni senso riferisce uno dei tipi rappresentati nell'ontologia SIMPLE (Lenci et. al. 2001), è possibile anche accedere ai contesti desiderati del testo partendo da caratteristiche concettuali (quindi extra-linguistiche). Componendo un'interrogazione relativa ai sensi riferiti al campo semantico di "Animale del cielo", si otterranno ad esempio i contesti relativi alle forme (solo lemma o anche forme flesse) dei sensi di "uccello", "mosca", "cavalletta" e così via.

The screenshot displays the LexO-server interface. On the left, there's a sidebar for 'Limit the research context' with a list of 19 resources, all checked. The main search area has 'recipiente' entered in the 'Word' field. Below it, the 'Computational Lexicon' section shows selected senses: 'recipiente - noun - [number: plural:]', 'Un recipiente d'acqua [43-Amount]', and 'involucro che possa contenere liquidi o materiali incoerenti! [333-Container]'. The 'Expand with' section is set to 'hyponymy'. On the right, a table titled 'Computational Lexicon' lists forms corresponding to the parameters entered. Below the table, there's a search results section with an index and occurrences of the word in context.

Form	Lemma	Sense	Relation	Range	Target Lemma	Target Sense	Forms
recipiente - noun							recipienti -
recipiente	Un recipie...	iponimia	1	anfora	vaso a due anse	anfore -	
recipiente	Un recipie...	iponimia	1	casseruola	recipiente da cucina...	casseruole -	
recipiente	Un recipie...	iponimia	1	tegame	recipiente da cucina	tegami -	

Items per page: 25 | 1 - 25 of 209

Index Occurrence

1.1.9 E Ghilón e cento uomini che erano con lui vennero al bordo del campo all'inizio della veglia di mezzo, quando avevano appena stabilito le guardie, e suonarono lo shofar e ruppero le anfore che erano nelle loro mani (Ouid. 7:19);

1.9.26 Le anfore intatte vanno al fiume,

8.2.3 Quelli sono proprietari di barili (pittanti) pieni di vino pregiato.

1.7.42 rav Chisda lo incoronava circondandolo di bicchieri colmi di vino.

1.7.42 E anche i bicchieri che circondano il bicchiere con cui si recita la benedizione si riempiono con vino "chià".

1.8.2 In effetti, la mishna si riferisce al caso in cui non c'è abbastanza vino per due bicchieri, ma ce n'è più di quanto basta per riempirne uno.

Figura 5: Interfaccia di ricerca full-text basata su conoscenza linguistica.

5. Conclusioni

In questo articolo è stato presentato LexO-server, un insieme di servizi informativi, sotto forma di API REST, per la gestione di lessici e terminologie, nel contesto del Web Semantico. I servizi sono di libero utilizzo e a codice

³⁰ Il testo si intende linguisticamente annotato.

aperto. Sono rivolti a chi vuole sviluppare applicazioni lessicali e/o terminologiche che si basano sugli standard del Web Semantico. Sono stati presentati tre casi d'uso che mostrano la generalità e la flessibilità di LexO-server, mostrando il loro utilizzo in altrettanti strumenti.

I lavori futuri proseguiranno su due fronti, ovvero il miglioramento dei servizi attuali e il concepimento di nuovi servizi per la gestione di ontologie concettuali in OWL.

Per quanto riguarda il primo punto, è prevista un'attività di revisione, risoluzione di bugs, eventuale aggiornamento e ampliamento dei servizi, sulla base anche dei feedback ricevuti da chi sviluppa applicazioni che utilizzano i servizi stessi.

Riguardo la seconda linea, si intende fornire dei servizi di supporto sia alla costruzione e fruizione di ontologie OWL, sia alle interrogazioni avanzate che sfruttano la potenza espressiva di OWL per accedere in maniera più sofisticata dal concetto al dato linguistico.

Riferimenti bibliografici

- Arcidiacono, Salvatore. 2019. *Per una Lessicografia Elettronica degli Antichi Volgari Italiani*. Quaderni di Artesia. Centro Studi Filologici e Linguistici Siciliani.
- Baader, Franz, Ian Horrocks, Carsten Lutz, and Uli Sattler. 2017. *An Introduction to Description Logic*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781139025355>.
- Bellandi, Andrea. 2021. "LexO: an Open-source System for Managing OntoLex-Lemon Resources." *Language Resources and Evaluation* 55, no. 4: 1093-126.
- Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. "The Semantic Web." *Scientific American* 284, no. 5: 34-43.
- Bosque-Gil, Julia, Jorge Gracia, and Elena Montiel-Ponsoda. 2017. "Towards a Module for Lexicography in OntoLex." In *Proceedings of the 1st Workshop on the OntoLex Model 18 June 2017 Galway, Ireland*, 74-84.
- Buitelaar, Paul. 2010. "Ontology-based Semantic Lexicons: Mapping between Terms and Object Descriptions." In *Ontology and the Lexicon. A Natural Language Processing Perspective*, edited by Chu-ren Huang, Nicoletta Calzolari, Aldo Gangemi, Alessandro Lenci, Alessandro Oltramari, Laurent Prevot, 212-23. Cambridge: Cambridge University Press.
- Chiarcos, Christian, Sebastian Hellmann, and Sebastian Nordhoff. 2011. "Towards a Linguistic Linked Open Data cloud: The Open Linguistics Working Group." *TAL* 52, no. 3: 245-75.

- Chiarcos, Christian, Sebastian Hellmann, and Sebastian Nordhoff. 2012. "Linking linguistic resources: Examples from the open linguistics working group." In *Linked Data in Linguistics*, 201-16. Berlin Heidelberg: Springer.
- Cimiano, Philipp, John Philipp McCrae, Víctor Rodríguez-Doncel, Tatiana Gornostay, Asunción Gómez-Pérez, Benjamin Siemoneit, and Andis Lagzdins. 2017. "Linked Terminologies: Applying Linked Data Principles to Terminological Resources." In *Proceedings of the Electronic Lexicography in the 21st Century 19-21 September 2017 Leiden, the Netherlands*, 504-17.
- Colombo, Marianna, e Emiliano Giovannetti. 2022. "La Visualizzazione Grafica di Sensi e Relazioni Semantiche di un Lessico Computazionale della Lingua Italiana." In *Proceedings of AIUCD 2022 Digital Cultures. Intersections: Philosophy, Arts, Media 01-03 June 2022 Lecce*, 155-60.
- Fiorelli, Manuel, Armando Stellato, Tiziano Lorenzetti, Andrea Turbati, Peter Schmitz, Enrico Francesconi, Najeh Hajlaoui, and Brahim Batouche. 2020. "Editing OntoLex-Lemon in VocBench 3." In *Proceedings of the 12th Conference on Language Resources and Evaluation, Marseille, 11-16 May 2020*, 7196-205.
- Fiorelli, Manuel, Maria Teresa Pazienza, and Armando Stellato. 2013. "LIME: Towards a Metadata Module for Ontolex." In *Proceedings of the 2nd Workshop on Linked Data in Linguistics: Representing and Linking Lexicons, Terminologies and Other Language Data September 2013 Pisa, Italy*.
- Giovannetti, Emiliano, Davide Albanesi, Andrea Bellandi, Simone Marchi, Mafalda Papini, and Flavia Sciolette. 2021. "The Role of a Computational Lexicon for Query Expansion in FullText Search." In *Proceedings of the 8th Italian Conference on Computational Linguistics 26-28 June 2022 Milan*.
- Gracia, Jorge, Elena Montiel-Ponsoda, Daniel Vila-Suero, and Guadalupe Aguado-de-Cea. 2014. "Enabling Language Resources to Expose Translations as Linked Data on the Web." In *Proceedings of 9th Language Resources and Evaluation Conference May 2014 Reykjavik, Iceland*, 409-13.
- Khan, Anas Fahad. 2018. "Towards the Representation of Etymological Data on the Semantic Web." *Information* 9, no. 12: 304.
- Lenci, Alessandro, Nuria Bel, Federica Busa, Nicoletta Calzolari, Elisabetta Gola, Monica Monachini, Antoine Ogonowski, Ivonne Peters, Wim Peters, Nilda Ruimy, et al. 2000. "SIMPLE: A General Framework for the Development of Multilingual Lexicons." *International Journal of Lexicography* 13, no. 4: 249-63.

- McCrae, John Philipp, Julia Bosque-Gil, Jorge Gracia, Paul Buitelaar, and Philipp Cimiano. 2017. "The Ontolex-Lemon model: development and applications." In *Proceedings of the Electronic Lexicography of the 21st Century September 2017 Leiden, the Netherlands*, 19-21.
- McCrae, John Philipp, and Christina Unger. 2014. "Design Patterns for Engineering the Ontology-Lexicon Interface." In *Proceedings of the Towards the Multilingual Semantic Web*, edited by Paul Buitelaar, Philipp Cimiano, Berlin Heidelberg: Springer, 15–30.
- McCrae, John Philipp, Guadalupe Aguado de Cea, Paul Buitelaar, Philipp Cimiano, Thierry Declerck, Asunción Gómez-Pérez, Jorge Gracia, Laura Hollink, Elena Montiel-Ponsoda, Dennis Spohr, et al. 2012. "Interchanging Lexical Resources on the Semantic Web." *Language Resources and Evaluation* 46, 701-19.
- Miller, George A. 1995. "WordNet: a lexical database for English." *Communications of the ACM* 38, no. 11:39-41. <https://doi.org/10.1145/219717.219748>.
- Monachini, Monica, and Anas Fahad Khan. 2018. "Towards the Construction of a Lexical Data and Technology Ecosystem: The Experience of ILC-CNR." In *Proceedings of the Language Resources and Evaluation Conference, Workshop Globalex 7-12 May 2018 Miyazaki, Japan*, 52-4.
- Quochi, Valeria, Andrea Bellandi, Fahad Khan, Michele Mallia, Francesca Murano, Silvia Piccini, Luca Rigobianco, Alessandro Tommasi, and Cesare Zavattari. 2022. "From Inscriptions to Lexica and back: A Platform for Editing and Linking the Languages of Ancient Italy." In *Proceedings of the 13th Language Resources and Evaluation Conference (LREC 2022) 20-25 June 2022 Marseille*.
- Rico, Mariano, and Christina Unger. 2015. "Lemonade: A Web Assistant for Creating and Debugging Ontology Lexica." *Natural Language Processing and Information Systems NLDB 2015 Lecture Notes in Computer Science* 9103, edited by Chris Biemann, Siegfried Handschuh, Andre Freitas, Farid Mezziane, Elisabeth Méitas, Cham: Springer, 448-52. https://doi.org/10.1007/978-3-319-19581-0_45.
- Ruimy, Nilda, Monica Monachini, Raffaella Distanti, Elisabetta Guazzini, Stefano Molino, Marisa Ulivieri, Nicoletta Calzolari, and Antonio Zampolli. 2002. "Clips, a Multi-level Italian Computational Lexicon: A Glimpse to Data." In *Proceedings of the 3rd International Conference on Language Resources and Evaluation May 2002 Las Palmas, Canary Islands, Spain*.

- Stellato, Armando, Andrea Turbati, Manuel Fiorelli, Tiziano Lorenzetti, Eugeniu Costetchi, Christine Laaboudi, Willem Van Gemert, and Johannes Keizer. 2017. "Towards VocBench 3: Pushing Collaborative Development of Thesauri and Ontologies Further Beyond." In *Proceedings of the 17th European Networked Knowledge Organization Systems Workshop 21 September 2017 Thessaloniki, Greece*, CEUR-WS, 39-52.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, et al. 2016. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3.

