Umanistica Digitale – ISSN: 2532-8816 – n. 20, 2025: Special Issue
A.A. Nannini - *Sulla rappresentazione delle informazioni semantiche nei dizionari storici digitali*
DOI: http://doi.org/10.6092/issn.2532-8816/21250

Sulla rappresentazione delle informazioni semantiche nei dizionari storici digitali*

Alessandro Alfredo Nannini

Universität des Saarlandes
alessandro.nannini@uni-saarland.de

Abstract

Il presente contributo intende presentare gli sviluppi che si sono susseguiti nel campo della lessicografia storica in merito alla rappresentazione delle informazioni semantiche contenute nelle risorse digitali.

Muovendo da alcune considerazioni sul rapporto tra lessicografia e semantica lessicale nel secolo scorso, cercheremo di indagare se e in che modo la svolta digitale delle discipline umanistiche abbia influenzato l'interazione tra le due discipline. Dopo un'analisi delle modalità di digitalizzazione della lessicografia storica, passiamo in rassegna le possibilità – e le impossibilità – di indagine semantica offerte dalla lessicografia storica digitale.

La seconda parte del contributo prende in esame le possibilità offerte dalle tecnologie del Web alla lessicografia. Nello specifico, mostreremo quali strumenti possono essere utilizzati per rappresentare le informazioni semantiche ed etimologiche all'interno delle risorse Linguistic Linked Data (LLD), proponendo un esempio di modellazione di voci tratte da dizionari storici romanzi.

Parole chiave: #AIUCD2023 -- Lessicografia Digitale – Semantica -- Linguistic Linked Data – Lessicografia Storica – Etimologia -- OntoLex-Lemon

This paper aims to present the developments that have taken place in the field of historical lexicography regarding the representation of semantic information in digital resources.

Starting from some considerations on the relationship between lexicography and lexical semantics in the past century, we will investigate whether and to what extent the digital turn in the humanities has influenced the interaction between these two disciplines. After analyzing the different modes of digitization in historical lexicography, we will review the possibilities – and limitations – of semantic analysis offered by digital historical lexicography.

* Corrected in date 07/08/2025, see the Erratum: https://doi.org/10.6092/issn.2532-8816/22472

Copyright © 2025 The Author(s)
The text in this work is licensed under the Creative Commons BY License.
https://creativecommons.org/licenses/by/4.0/

The second part of the paper examines the opportunities that Semantic Web technologies offer to lexicography. Specifically, we will illustrate which tools can be employed to represent semantic and etymological information within Linguistic Linked Data (LLD) resources, proposing an example of the modeling of entries taken from historical dictionaries of Romance languages.

Keyword: #AIUCD2023 -- Digital Lexicography -- Semantics – Linguistic Linked Data -- Historical Lexicography -- Etimology -- OntoLex-Lemon

1. Lessicografia e semantica

Nel passaggio dal dizionario cartaceo alle risorse lessicografiche digitali, si sono verificate trasformazioni significative che hanno interessato sia la rappresentazione delle informazioni semantiche, sia le modalità di accesso e di consultazione. In questa sezione introduttiva intendiamo illustrare i rapporti tra semantica e lessicografia, operando un distinguo tra lessicografia analogica e lessicografia digitale, e mostrando i vantaggi che la svolta digitale ha apportato alla vocabolaristica.

1.1 Lessicografia analogica

Con la diffusione delle teorie di Ferdinand de Saussure, la linguistica europea comincia ad adottare un approccio strutturalista allo studio del significato lessicale. Nella semantica saussuriana il valore del segno linguistico, unione di significante e significato, è determinabile soltanto in relazione al sistema di cui fa parte, poiché "la lingua è un sistema di cui tutti i termini sono solidali ed in cui il valore dell'uno non risulta che dalla presenza simultanea degli altri" ([64]: 139). Su queste basi si sviluppano teorie strutturaliste che promuovono una visione sistematica del lessico e delle strutture semantiche.

Tra queste, la teoria del campo lessicale, nella formulazione di Jost Trier [70], considera il lessico come sistema in cui i significati hanno dei confini ben delineati, mentre la semantica relazionale di John Lyons [43] pone l'accento sulla natura relazionale dei significati. Sul versante della semantica storica, è di grande importanza la trattazione strutturalista del cambio semantico proposta da Stephen Ullmann [71].

Le teorie strutturaliste hanno contribuito a generare e alimentare la discussione sulla forma dei dizionari scientifici che ha luogo tra gli anni Venti e gli anni Sessanta del Novecento. In quel periodo si fa strada una visione alternativa del dizionario, ordinato non più per significanti (metodo *semasiologico*, diretto dai significanti ai significati) ma per significati (o concetti; metodo *onomasiologico*, diretto dai significati ai significanti).

> "Semasiology takes its starting point in the word as a form, and charts the meanings that the word can occur with; onomasiology takes its starting point in a concept, and investigates by which different expressions the concept can be designated, or named. Between the two, there is a difference of perspective: semasiology starts from the expression and looks at its meanings, onomasiology starts from the meaning and looks at the different expressions." ([26]: 23)

Il dizionario onomasiologico viene considerato più adatto dell'omologo semasiologico per la rappresentazione della natura sistematica e strutturata del vocabolario.

La commistione di semantica e lessicografia è ben presente nel *discurso* sul *Nuevo concepto del diccionario de la lengua* di Julio Casares [17], negli scritti di Walther von Wartburg sulla compilazione dei dizionari scientifici ([75], [76]), nella compilazione del *Begriffssystem als Grundlage für die Lexikographie* da parte di Ruldolf Hallig e dello stesso von Wartburg [29], così come nei lavori sia teorici che lessicografici di Kurt Baldinger, raro esempio di lessicografo semanticista del secolo scorso ([2], [3], [4], [8]). Questa discussione, condotta quasi esclusivamente nell'Europa continentale, ha portato all'ideazione e alla compilazione di alcuni dizionari onomasiologici di tipo sincronico ([17], [6], [7]).

Per quanto riguarda la lessicografia storica, in quei decenni i grandi progetti europei continuano a essere concepiti e redatti nel solco della tradizione delle accademie di lingua. Si tratta spesso di dizionari semasiologici in cui ogni voce registra le accezioni assunte da un lessema nel tempo. Talvolta le accezioni vengono accompagnate da marche metalessicografiche atte a evidenziare – spesso senza sistematicità – l'azione di un qualche cambiamento semantico.

Nelle Figure 1, 2 e 3 vengono mostrate tre accezioni presenti sotto la voce *cane* del *Grande Dizionario della Lingua Italiana* (GDLI) [9]. Per l'accezione 6. (Figura 2) il dizionario presenta una marca d'uso riferita a un uso figurato, mentre non viene registrata nessuna marca per il significato 8., altrettanto *figurato*.

Cane[1], sm. (femm. *cagna* [v.]). Zool. Genere di Mammiferi carnivori appartenente alla famiglia Canidi, dal pelo folto variamente colorato, con pupilla rotonda, voce ringhiante, latrante, abbaiante; comprende molte specie selvatiche (sciacallo, coyote, lupo, dingo), e il cane domestico, che di tutti gli animali è il più fedele e affezionato all'uomo; e ne esiste un grande numero di razze, classificate, secondo l'uso a cui sono destinate, in quelle dei *cani da guardia e d'utilità* (lupo, molosso, alano, poliziotto); *da pastore*; *di lusso* (volpino, maltese, pechinese, giapponese); *da salvataggio* (Terranova, San Bernardo); *da caccia* (bracco, spinone); *da ferma*: avvezzato a puntare l'animale, per dar tempo al cacciatore di colpirlo; *da giungere*: che raggiunge la preda; *da presa*: che afferra coi denti l'animale.

Figura 1 Il primo significato della voce *cane* nel GDLI.

6. Figur. Cantante, attore, scrittore, maestro estremamente inabile.

Figura 2 Sezione 3 della voce *cane*.

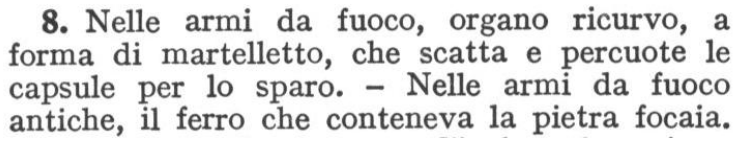
8. Nelle armi da fuoco, organo ricurvo, a forma di martelletto, che scatta e percuote le capsule per lo sparo. – Nelle armi da fuoco antiche, il ferro che conteneva la pietra focaia.

Figura 3 Sezione 8 della voce *cane*.

Decisamente diversa appare la situazione della vocabolaristica anglossassone. Il primo grande dizionario storico della lingua inglese, l'*Oxford English Dictionary* (OED) [52], viene compilato tra il 1884 e il 1928. L'OED è un dizionario semasiologico in cui le informazioni semantiche vengono presentate secondo un criterio genetico-cronologico. L'opera, così dettagliata semanticamente, viene *rovesciata*[1] onomasiologicamente all'interno di uno dei più importanti progetti onomasiologici del mondo. Infatti, a partire dagli anni Sessanta, le entrate lessicali dell'*Oxford English Dictionary* vengono ordinate per concetti nell'*Historical Thesaurus of Oxford English Dictionary* (HTOED) [33], dizionario storico-onomasiologico che registra il lessico anglosassone – considerato in tutta la sua estensione diacronica – all'interno di un sistema concettuale.

Tuttavia, nonostante le eccezioni appena viste, la lessicografia continuava a contraddistinguersi per un approccio prettamente pratico, spesso distante dalle riflessioni teoriche del suo tempo. In tale contesto si inseriscono alcune critiche che la disciplina ha attirato nella seconda metà del Novecento. Ad esempio, Uriel Weinreich rimproverava alla lessicografia l'utilizzo di molti approcci pratici a discapito di una metodologia motivata teoricamente [77]. Sulla stessa linea, in un contributo del 1982, il romanista svizzero Max Pfister, fondatore del *Lessico Etimologico Italiano* (LEI) [57], riconosceva un abisso tra le ricerche di semantica teorica e la loro applicazione pratica in lessicologia e lessicografia [53]. Pfister, pur segnalando alcune importanti eccezioni (soprattutto per l'area linguistica galloromanza), individuava l'origine di questa lacuna nella mancata identità tra lessicografi e semanticisti. In buona sostanza, chi fa lessicografia non si occupa di semantica, mentre chi si occupa di semantica non si interessa di lessicografia. Una considerazione simile si trova qualche anno più tardi nelle parole della linguista polacca Anna Wierzbicka, la quale ribadisce l'esistenza dei quell'abisso tra il lavoro dei semanticisti e quello dei lessicografi: "Semantic theories flourished, often propped up on a single 'example' […], while lexicographers described thousands of words regardless of any semantic theories whatsoever" ([78]: 5).

Come abbiamo visto, non mancano linguisti novecenteschi interessati alla cooperazione tra le due discipline in funzione di concezioni alternative del dizionario, ma sarà soltanto in ambito informatico che tali concezioni conosceranno uno sviluppo concreto nei nuovi modelli lessicografici digitali.

1.2 La svolta digitale

Mentre alcuni studiosi orientano il proprio lavoro verso rappresentazioni alternative del dizionario, una spinta decisiva in questo senso giunge dal connubio tra informatica, linguistica e scienze cognitive di ambito nordamericano.

[1] Prendiamo in prestito l'immagine del rovesciamento da Mariafrancesca Giuliani (http://www.ovi.cnr.it/files/Giuliani_Contatti_e_non_solo_parole.pdf).

Nel campo dell'intelligenza artificiale, a partire dagli anni Sessanta vengono sviluppati modelli informatici per la rappresentazione delle informazioni semantiche che prendono il nome di *semantic network*. Il concetto di rete semantica viene ideato da M. Ross Quillian ([60], [61]), il quale sviluppa un modello computerizzato (*memoria*) composto da *unità* e *proprietà*; ogni unità corrisponde al concetto "of some object, event, idea, assertion, etc." ([61]: 462). Le unità presenti nella memoria (indicate in MAIUSCOLO) vengono distinte dalle unità lessicali memorizzate esternamente in un dizionario; le unità lessicali puntano alle unità della memoria che ne rappresentano il significato. Le unità semantiche sono collegate tra di loro da rapporti gerarchici (ad esempio, MAN risulta gerarchicamente inferiore a PERSON che, a sua volta è gerarchicamente inferiore ad ANIMAL). In base al significato rappresentato, le queste unità vengono specificate da una serie di proprietà, proprio come accade nelle definizioni lessicografiche del tipo *genus proximum et differentia specifica* (Figura 4).

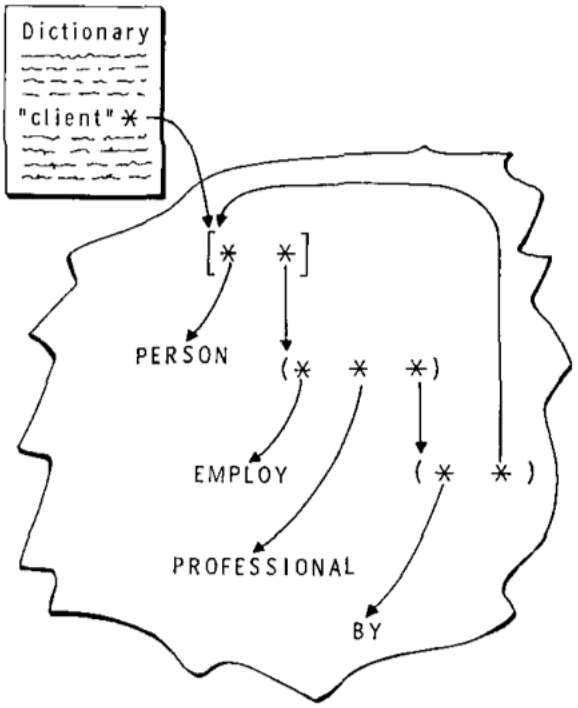

Figura 4 Una porzione di informazione nella memoria di Quillian ([61]: 462)

Come si vede, seppur in un contesto disciplinare diverso, il concetto di *semantic network* può essere avvicinato alla concezione sistematica del vocabolario sviluppata pochi decenni prima in ambito lessicografico.

Novità e vantaggi di questo modello sono rappresentati dalla possibilità di raggiungere ogni elemento da percorsi differenti.

> "Ciascuno dei termini definienti può diventare a propria volta type di un'altra serie di tokens. Per esempio animale sarebbe esemplificato da cane, ma anche da gatto e comprenderebbe quadrupede e bipede; oppure,

> individuato un nodo gatto, questo sarebbe definito da alcuni nodi comuni alla definizione di cane, quali animale o quadrupede, ma rinvierebbe anche a nodi come felino, comune anche a tigre, e così via." ([24]: 62)

La rete semantica si sposa bene con l'ideazione di nuovi strumenti che prende piede nella neonata lessicografia digitale. Tra i benefici apportati dall'informatica alla vocabolaristica viene presto riconosciuta la possibilità di descrizioni semantiche più precise e il superamento del metodo semasiologico. Il tema trova discreto riscontro in un convegno pisano [80], dove i relatori non sembrano dubitare di questi vantaggi. Infatti, un database lessicografico, a differenza di un dizionario cartaceo, risulta consultabile da vari punti di accesso (formali, semantici, cronologici, ecc.). A tal proposito Harald H. Zimmermann parla di *dizionario multifunzione* [81], mentre Martin Kay propone esplicitamente la soluzione della rete semantica per la compilazione lessicografica: "a large number of elementary items of information, linked together in a very richly interconnected structure" ([34]: 172). Antonio Zampolli, presentando il database lessicale *Dizionario Macchina Italiano* (DMI), illustra le possibilità di analisi del sistema definitorio del database al fine di ordinare il materiale lessicografico per campi semantici e, all'interno di essi, per relazioni semantiche [79]. In questo clima di interesse per concezioni alternative del dizionario nel nuovo contesto digitale, non mancano studi mirati a illustrare le possibilità offerte dall'informatica all'approccio onomasiologico (cf. [42], [15]).

Negli stessi anni (1986) presso l'Università di Princeton viene progettata la rete semantica *WordNet*[2] in cui lessemi e i gruppi di lessemi sono interrelati per mezzo di relazioni lessicali e concettuali.

Senza voler continuare un resoconto storico che risulterebbe inevitabilmente incompleto, ci limitiamo a riconoscere che, almeno a partire dagli anni Ottanta, il dizionario digitale viene concepito come uno strumento completamente diverso dal suo corrispettivo cartaceo, tradizionalmente costretto in una camicia di forza semasiologica e alfabetica.[3]

Spesso il nuovo oggetto lessicografico non è soltanto uno strumento multifunzionale e multidimensionale, ma arriva a trascendere i confini stessi del dizionario tradizionale, partecipando a quel processo di ibridazione con altre risorse per cui Reinhard Hartmann [30] individua una dozzina di esempi. Tra questi, è di particolare interesse per la semantica il *dictionary cum thesaurus*, opera ibrida nella quale un layer concettuale interagisce con il materiale lessicografico.

In conclusione, la svolta digitale ha favorito riflessioni di tipo metalessicografico, aprendo a nuove forme di ibridazione e promuovendo una maggiore consapevolezza rispetto alla strutturazione delle informazioni. Queste, progressivamente svincolate dalla loro configurazione testuale lineare, vengono oggi organizzate in sistemi relazionali che permettono una descrizione articolata dei fatti linguistici, facilitano il recupero delle informazioni e abilitano il collegamento con altre risorse.

[2] https://wordnet.princeton.edu.

[3] Di questo si compiace Beryl Atkins al convegno di Euralex del 1996: "and now at last we are liberated from the straitjacket of the printed page and alphabetical order" ([1]: 516).

2. Lessicografia storica digitale e rappresentazione delle informazioni semantiche

Nel quadro brevemente abbozzato rientra sicuramente l'interesse informatico per dizionari monolingui o bilingui relativi agli stadi contemporanei delle lingue, risorse fondamentali in determinati campi quali l'elaborazione del linguaggio naturale (o, in inglese, *Natural Language Processing*, NLP) e la traduzione automatica [49]. Un discorso a parte deve essere fatto per la lessicografia storica, focus del presente contributo. Molti dizionari storici in corso d'opera sul finire del Novecento erano stati concepiti come strumenti analogici prima dello sviluppo della lessicografia digitale o in contemporanea con esso,[4] il che motiva spesso una dichiarata rinuncia al cambiamento, come risulta evidente dalle parole di Max Pfister: «*j*e suis d'avis que l'emploi de l'ordinateur devrait être prévu tout au commencement d'un projet; au cours d'une telle enterprise, un changement présente des difficultés sensibles» ([54]: 189). Tuttavia, anche la lessicografia storica ha dovuto fare i conti con l'inevitabilità della svolta digitale, la quale si è articolata nelle tre fasi individuate dal linguista francese Bernard Cerquiglini (citato in [59]).

1) La prima fase riguarda l'ausilio dell'informatica nel processo di compilazione dei dizionari in formato analogico;

2) la seconda riguarda la retrodigitalizzazione delle opere cartacee;

3) la terza è orientata alla creazione di dizionari nativi digitali.

Allo stato attuale la lessicografia storica è interessata principalmente dalle fasi 2) e 3): le nuove imprese vengono concepite totalmente in digitale (3), mentre i progetti terminati in formato analogico vengono retrodigitalizzati (2). Queste fasi possono essere viste anche come approcci,[5] ai quali aggiungiamo un'ulteriore categoria (4)) dedicata all'approccio misto di retrodigitalizzazione e compilazione nativa digitale che interessa specialmente quei dizionari che, pubblicati inizialmente su carta, sono transitati verso il digitale in corso d'opera.[6]

Nella Tabella 1 mostriamo lo stato di alcune risorse storico-lessicografiche, distinguendo tra quelle retrodigitalizzate, quelle digitalizzate in corso d'opera e progetti nativi digitali.

Approcio alla digitalizzazione	Esempi
Retrodigitalizzazione di opere pubblicate in formato cartaceo	• *Altfranzösisches Wörterbuch* (TL);[7] • *Digitales Wörterbuch der deutschen Sprache* (DWDS);[8]

[4] Le ragioni sono anche economiche e di mercato. Nella rassegna di Carla Marello [46] relativa alle lessicografia digitale (lì chiamata *lessicografia computazionale*), l'OED è l'unico dizionario storico in versione elettronica che viene menzionato. A quell'altezza cronologica, la studiosa riconosce che la compilazione di dizionari basati su grandi corpora (come avviene nella lessicografia storica) è appannaggio di "gruppi di ricerca agguerriti e con finanziamenti adeguati" ([46]: 156), mentre l'editoria elettronica concentra i propri sforzi sulla creazione di versioni digitali di dizionari generici.

[5] Vale a dire *fasi* nella storia particolare dei singoli dizionari.

[6] La digitalizzazione delle risorse che hanno intrapreso la fase 1) deve passare inevitabilmente per la fase 3) o la fase 4).

[7] https://www.steiner-verlag.de/Altfranzoesisches-Woerterbuch/9783515081672.

[8] https://www.dwds.de. La risorsa consiste in un corpus lessicografico che raccoglie le retrodigitalizzazioni di dizionari sincronici e diacronici della lingua tedesca.

	• *Dicţionarul Limbii Române* (DLR);[9] • *Französisches Etymologisches Wörterbuch* (FEW);[10] • *Grande Dizionario della Lingua Italiana* (GDLI);[11] • *Middelnederlandsch Woordenboek* (MNW);[12] • *Romanisches Etymologisches Wörterbuch* (REW);[13] [14] [15] • *Trésor de la langue française informatisé* (TLFi)[16]
Digitalizzazione in corso d'opera	• *A Dictionary of Old Norse Prose* (ONP);[17] • *A Thesaurus of Old English* (TOE);[18] • *Dictionnaire Étymologique de l'Ancien Français* (DEAF);[19] • *Historical Thesaurus of Oxford English Dictionary* (HTOED);[20] • *Lessico Etimologico Italiano* (LEI);[21] • *Mittellateinisches Wörterbuch* (MLW);[22] • *Novum Glossarium Mediae Latinitatis* (NMGL);[23] • *Oxford English Dictionary* (OED);[24]

[9] https://dlr1.solirom.ro.

[10] https://lecteur-few.atilf.fr.

[11] https://www.gdli.it.

[12] I dizionari olandesi presenti in tabella sono disponibili sulla piattaforma *Geïntegreerde Taalbank* (GTB) (https://gtb.ivdnt.org).

[13] https://archive.org/details/romanischesetymo00meyeuoft.

[14] https://nbn-resolving.org/urn:nbn:de:bvb:355-ubr07799-0.

[15] https://www.rew-online.gwi.uni-muenchen.de.

[16] http://atilf.atilf.fr/tlf.htm.

[17] https://onp.ku.dk.

[18] https://oldenglishthesaurus.arts.gla.ac.uk.

[19] https://deaf.hadw-bw.de.

[20] https://www.ht.ac.uk.

[21] https://lei-digitale.it.

[22] https://woerterbuchnetz.de/?sigle=MLW&lemid=A00088.

[23] https://glossaria.eu/ngml.

[24] https://www.oed.com.

	• *Thesaurus Linguae Latinae* (TLL);[25] [26] • *Woordenboek der Nederlandsche Taal* (WNT)
Progetti nativi digitali	• *ALMA. Wissensnetzen in der mittelalterlichen Romania* [58]; • *Diachroon seMAntisch lexicon van de Nederlandse Taal* (DiaMANT);[27] • *Dictionnaire Étymologique Roman* (DÉRom);[28] • *Dictionnaire de Termes Médico-botaniques de l'Ancien Occitan* (DiTMAO);[29] • *Diccionario histórico de la lengua española* (DHLE);[30] • *LiLa: Linking Latin*;[31] • *Languages and Cultures of Ancient Italy. Historical Linguistics and Digital Models* (ItANt) [62]; • *Oudnederlands Woordenboek* (ONW); • *Mittelhochdeutsche Begriffsdatenbank* (MHDBDB);[32] • *Tesoro della Lingua Italiana delle Origini* (TLIO);[33] • *Vroegmiddelnederlands Woordenboek* (VMNW)

Tabella 1 Classificazione di alcune risorse lessicografiche digitali in base al tipo di digitalizzazione.

Le risorse in formato digitale offrono all'utente nuove opzioni di consultazione, così come l'informatizzazione dei processi redazionali permette rappresentazioni lessicografiche più coerenti e dettagliate.

Se i dizionari retrodigitalizzati in formato fotografico con OCR non consentono interrogazioni che vadano oltre alla semplice ricerca full-text, la multifunzionalità è senz'altro una caratteristica di quelle risorse che strutturano i dati all'interno di database o mediante la codifica testuale.

Per quanto riguarda lo studio dei significati, le interfacce di queste opere consentono ricerche per stringhe nel campo definizione, per marche d'uso o altri elementi che possono avere un

[25] https://thesaurus.badw.de/tll-digital/tll-open-access.html.

[26] https://tll.degruyter.com.

[27] https://diamant.ivdnt.org/diamant-ui/.

[28] http://stella.atilf.fr/DERom/.

[29] https://www.uni-goettingen.de/en/487498.html.

[30] https://www.rae.es/dhle/.

[31] https://lila-erc.eu.

[32] https://mhdbdb.sbg.ac.at.

[33] http://tlio.ovi.cnr.it.

qualche valore semantico. Alcuni dizionari hanno mappato semanticamente i loro materiali, creando così dei *dictionaries cum thesauri* e fornendo strumenti per effettuare query onomasiologiche. È il caso del DÉRom che permette ricerche a partire da una lista di concetti collegati agli etimi degli articoli [44]. In altri casi, come nel DHLE, sono presenti marche relative agli sviluppi semantici (*metafora*, *metonimia*, ecc.), utilizzabili in prospettiva diacronica [16] (Figura 5).

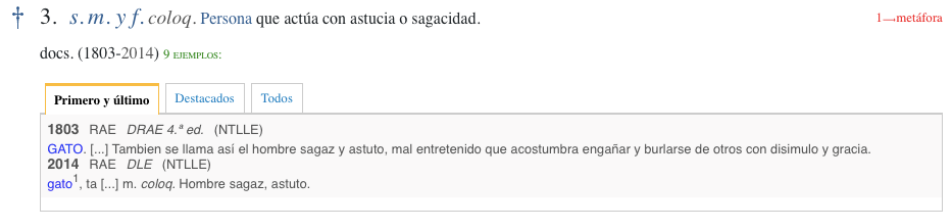

Figura 5 Sezione 3 della voce gato nel DHLE.

Nell'OED la totalità del materiale lessicale risulta ricercabile a partire dai concetti presenti in sistema concettuale gerarchico (Figura 6). La possibilità di utilizzare il dizionario come un *thesaurus* è frutto degli sforzi del progetto parallelo HTOED (poi *Historical Thesaurus of English*, HTE),[34] anch'esso nato analogico e poi transitato verso il digitale, così come il TOE, suo equivalente per l'inglese antico.

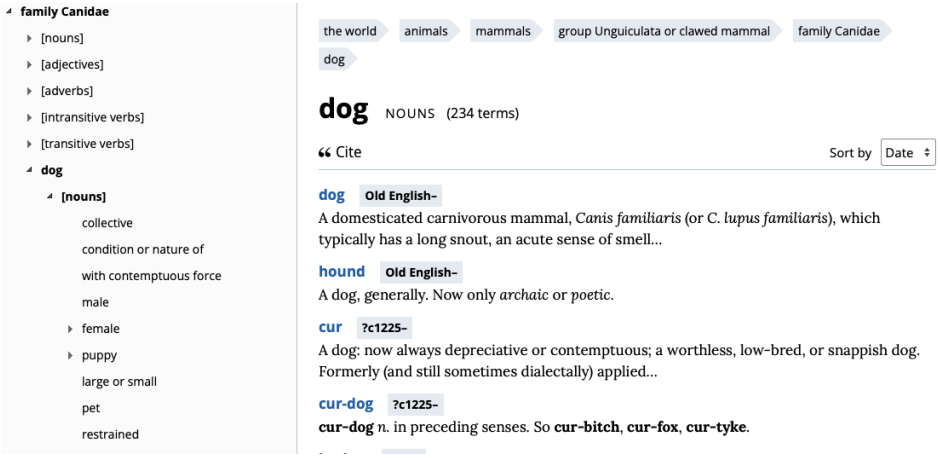

Figura 6 Il concetto DOG nell'Historical Thesaurus dell'OED.

In conclusione, anche la lessicografia storica sembra aver superato la fase iniziale di mera riproduzione digitale dei dizionari tradizionali, a favore di una strutturazione più raffinata dei

[34] https://ht.ac.uk.

dati e dell'ibridazione con altre risorse. Esemplare a tal proposito è la prima parte del titolo con cui Sabine Tittel presenta il progetto ALMA: *Ceci n'est pas un dictionnaire* [66].[35]

La rappresentazione dettagliata delle informazioni semantiche (sincroniche e diacroniche) e l'approccio onomasiologico sono tenuti in particolare considerazione da quei progetti che hanno adottato il paradigma Linked Data e l'utiizzo delle tecnologie del Web Semantico, come la maggior parte delle opere native digitali menzionate in Tabella 1[36] che hanno fatto anche della semantica diacronica (in particolare, dell'onomasiologia) il loro punto di forza.

3. Web Semantico e lessicografia

Ricollegandoci ai vantaggi delineati nella sezione precedente, nei due paragrafi che seguono forniremo brevi cenni introduttivi al Web Semantico e al paradigma (Linguistic) Linked Data, con l'obiettivo di chiarire e rendere più espliciti i benefici derivanti dall'integrazione tra lessicografia e semantica all'interno di tali contesti. Presenteremo infine il modello ontologico OntoLex-Lemon che rappresenta il vocabolario standard per la descrizione delle risorse lessicografiche in ambito Linked Data.

3.1 Cenni su Web Semantico e Linked Data

Il Web Semantico, concepito da Tim Berners-Lee sul finire degli anni Novanta, si basa sulla rappresentazione del significato intensionale di qualsiasi risorsa presente sul Web mediante metadati semantici [21]. Di ogni risorsa, identificata univocamente con un URI (*Uniform Resource Identifier*), vengono definite proprietà a relazioni in metalinguaggio RDF (*Resource Description Framework*)[37], il quale permette di compiere asserzioni in forma di tripla (soggetto – predicato – oggetto) sulle risorse. SPARQL[38] è il linguaggio d'interrogazione standard per le risorse RDF.

In questo contesto di rappresentazione della conoscenza, vengono modellate delle concettualizzazioni di dominio – le *ontologie* – che permettono di fare asserzioni a proposito delle risorse in un dato ambito di conoscenza. Nel contesto del Web Semantico si inserisce il paradigma Linked Data (LD), il quale adotta una serie di principi per la pubblicazione e la interconnessione dei dati strutturati sul Web [10].

Il Web Semantico si configura come una grafo di dati interconnessi tra loro, semanticamente definiti mediante le asserzioni rese possibili dalle ontologie. Questa concezione della rappresentazione delle informazioni viene accolta con favore anche in ambito linguistico dove vengono sviluppate ontologie per la descrizione dei cosiddetti Linguistic Linked Data (LLD).[39] Infatti, molti settori della linguistica, soprattutto quelli basati su metodologie empiriche, sono responsabili della creazione di un gran numero di dataset e risorse che, tuttavia, non sono

[35] L'ibridazione è esplicitamente menzionata anche da Elton Prifti [56] per la digitalizzazione del LEI.

[36] ALMA, DiTMAO, DiaMANT, LiLa, ItAnt e MHDBDB. Il titolo utilizzato da Tittel per ALMA (*Ceci n'est pas un dictionnaire*) si presta bene anche a descrivere reti linguistiche e di conoscenza come LiLa e ItAnt, in cui l'aspetto lessicografico riveste un ruolo centrale, ma non esaurisce la natura né gli obiettivi di questi progetti.

[37] https://www.w3.org/TR/rdf11-concepts/.

[38] https://www.w3.org/TR/sparql11-query/.

[39] Lo stato dell'arte dei LLD, aggiornato al 2022, si legge in [40].

interoperabili tra di loro a causa della diversità dei sistemi di rappresentazione ([20]: 29). Applicati in ambito linguistico, gli strumenti del Web Semantico risultano quindi di grande importanza per il conseguimento di una maggiore uniformità nella modellazione dei dati e di una garantita interoperabilità tra le risorse.

3.2 Il modello OntoLex-Lemon

Il concetto di base del Web Semantico è assimilabile a quello della *semantic network*. Non a caso, Umberto Eco, descrivendo la transizione dal modello semantico dizionariale al modello semantico enciclopedico nella cultura occidentale, considera le ontologie informatiche come la miglior messa in atto della rete semantica ideata da Quillian [24]. A partire dai primi anni del nuovo millennio sono state condotte indagini atte a verificare la possibilità di utilizzare le tecnologie del Web Semantico in lessicografia, proprio come nei decenni precedenti la lessicografia digitale aveva valutato le potenzialità del modello reticolare nella compilazione dei dizionari ([25]).

Nel 2011 John McCrae, Dennis Sphor e Philipp Cimiano presentano il modello lemon (*Lexicon Model for Ontologies*) «that supports the sharing of terminological and lexicon resources on the Semantic Web as well as their linking to the existing semantic representations provided by ontologies» ([47]: 245). Lo scopo principale di lemon è la creazione di lessici che possano essere usati per la lessicalizzazione di ontologie esistenti.

Nel 2016, in seguito a una fase di sviluppo del modello del 2011, Cimiano, McCrae e Paul Buitelaar pubblicano la specifica finale di OntoLex-Lemon, versione definitiva composta da una serie di moduli per la rappresentazione delle risorse lessicali [19].

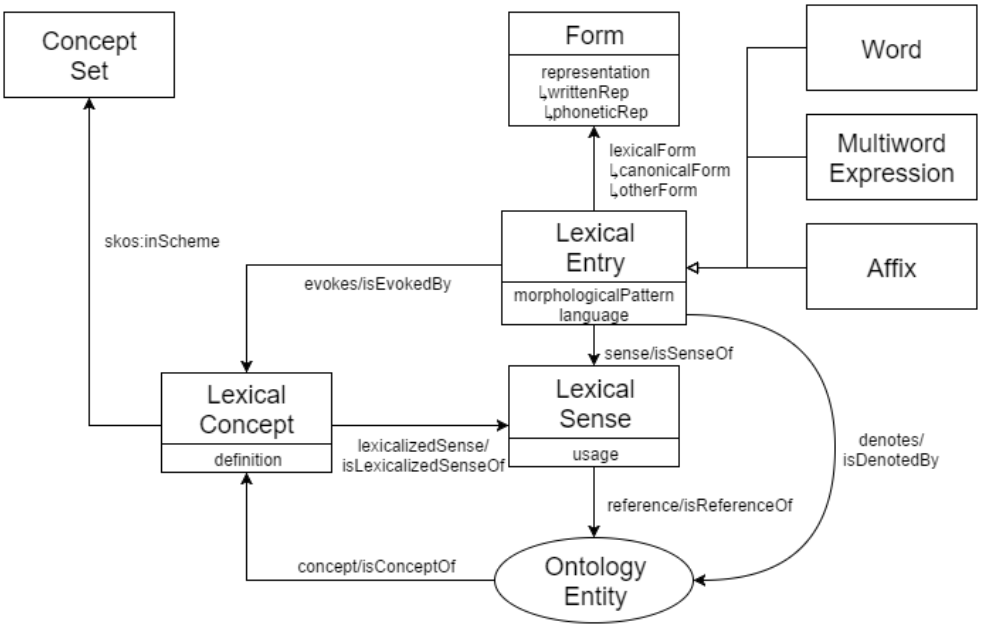

Figura 7 Il modulo core di OntoLex-lemon.

Il modulo core (Figura 7) mostra le possibilità di modellazione sia delle informazioni formali che di quelle semantiche relative a un'entrata lessicale (LexicalEntry) che, con discreta approssimazione, può essere assimilata al concetto saussuriano di segno linguistico. La modellazione delle informazioni semantiche risulta teoricamente flessibile e adattabile a diversi modelli semiotici. La classe LexicalSense rappresenta i significati lessicali dell'entrata, mentre LexicalConcept modella un'astrazione mentale, un concetto che può essere lessicalizzato dai significati lessicali.[40] Le classi LexicalEntry, LexicalSense e LexicalConcept possono legarsi direttamente a un'entità ontologica (Ontology Entity) (Listing 1).

Gli altri moduli sono synsem, decomp, vartrans, lime e lexicog. Quest'ultimo è stato sviluppato in una fase successiva per venire incontro a esigenze di modellazione lessicografica che non venivano soddisfatte dal modello iniziale. Infatti, il modulo lexicog viene presentato come complementare a core per la modellazione di risorse lessicografiche già esistenti e per la costruzione di risorse lessicografiche native LD.[41]

```
ex:entry a ontolex:LexicalEntry ;
ontolex:canonicalForm ex:entry_lemma ;
ontolex:sense ex:entry_sense1.

ex:entry_sense1 a ontolex:LexicalSense ;
ontolex:isSenseOf ex:entry ;
ontolex:isLexicalizedSenseOf ex:entry_sense1_concept .

ex:entry_concept a ontolex:LexicalConcept ;
skos:definition "Definizione" ;
ontolex:isConceptOf ex:ontology_entity .
```

Listing 1 Modellazione della semantica di una LexicalEntry (ex:entry) in OntoLex-lemon.

Sabine Tittel [67] ha descritto le analogie tra il modulo core di OntoLex-Lemon e il trapezio semiotico nella versione di Andreas Blank ([11]: 102) in cui vengono rappresentate graficamente le relazioni tra segno linguistico e realtà extralinguistica (Figura 8).[42]

[40] LexicalConcept può essere inteso come un concetto esterno a una lingua particolare che permette l'accesso onomasiologico ai lessici modellati. In questo si comporta come l'*außereinzelsprachlicher Begriff* ('concetto esterno a una lingua particolare') utilizzato dai linguisti tedeschi per indicare il carattere extralinguistico del concetto, in quanto entità astratta non necessariamente legata a una singola lingua, a differenza del significato lessicale che, nella visione strutturalista, ha un valore esclusivamente intralinguistico (cf. [31], [32]).

[41] https://www.w3.org/2019/09/lexicog/.

[42] Il modello semiotico distingue il livello linguistico (*einzelsprachlich*) da quello extralinguistico (*außersprachlich*), così come distingue il livello astratto (*abstrakt*) da quello concreto (*konkret*). La composizione del segno linguistico (*Zeichen*) riprende la partizione strutturalista tra piano dell'espressione (*Zeichenausdruck*) e piano del contenuto (*Zeicheninhalt*). A livello concreto il segno linguistico si manifesta come sostanza fonica (*Lautung*). A livello astratto il segno si lega a contenuti extralinguistici (*Designat*), i quali si collegano a referenti della realtà concreta (*Referent*).

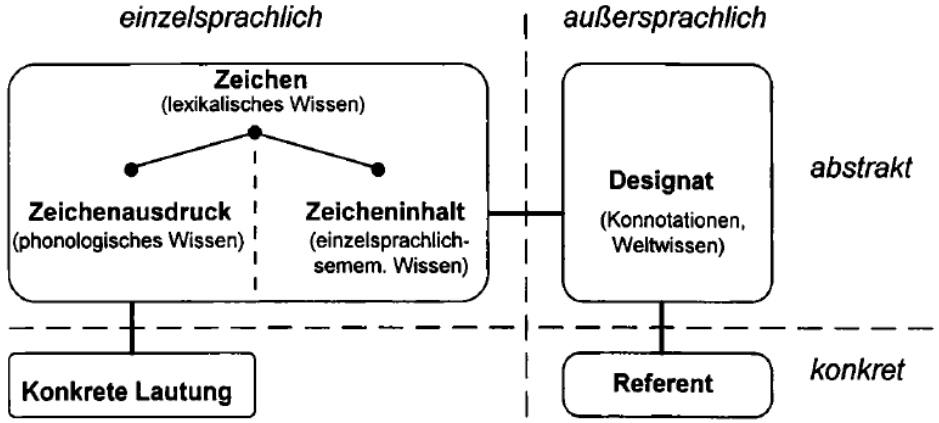

Figura 8 Il modello semiotico di Andreas Blank ([11]: 102).

OntoLex-Lemon ha visto negli anni un crescente utilizzo in ambito lessicografico, dove il paradigma Linked Oped Data (LOD) permette ai dizionari "to move beyond their confines of a fixed set of pages, and instead to be more richly linked and connected to other resources, including corpora and other lexico-conceptual resources such as encyclopedias and thesauri" ([20]: 57).

Proprio in virtù della sua flessibilità e adattabilità, il modello è diventato presto lo standard della lessicografia LOD, arrivando ad attirare l'interesse della lessicografia storica.

4. Lessicografia storica e Linked Data: un esempio di modellazione

Per quanto riguarda la lessicografia storica LD, si individuano due direttrici che si intrecciano frequentemente e si alimentano a vicenda. Da una parte si segnala lo sviluppo di modelli che espandono il modello OntoLex-Lemon per la rappresentazione dei dati storico-linguistici, dall'altra si nota l'utilizzo di OntoLex-Lemon e delle sue espansioni per la compilazione di nuove risorse lessicografiche.

Nella parte che segue forniremo un esempio di modellazione con OntoLex-Lemon e altri modelli ontologici sviluppati per la descrizione degli aspetti semantici ed etimologici (HisThes, lemonDIA, SemShift e lemonEty).

Nello specifico, abbiamo modellato i lessemi *cane* del *Lessico Etimologico Italiano* (LEI X, 858), *chien* del *Französisches Etymologisches Wörterbuch* (FEW II, 191) e *gato* del *Diccionario histórico de la lengua española* (DHLE)[43]. I file RDF contenenti informazioni lessicografiche, serializzati in formato Turtle[44], sono stati caricati in una repository di GraphDB[45], database RDF sviluppato da

[43] https://www.rae.es/dhle/gato.

[44] https://www.w3.org/TR/turtle/.

[45] https://graphdb.ontotext.com.

Ontotext per la gestione di dati semantici e LD. Questo metodo ha permesso di sfruttare le capacità di interrogazione SPARQL, nonché le funzionalità di inferenza offerte dal sistema, facilitando la gestione e l'interconnessione dei dati lessicografici all'interno del framework del Web Semantico.[46] Inoltre, l'ambiente GraphDB fornisce strumenti per la visualizzazione interattiva dei grafi RDF che sono stati utilizzati per generare le figure riportate nei prossimi paragrafi.[47]

4.1 Modellazione delle entrate lessicali e mappatura semantica

Le informazioni formali e semantiche relative a ogni entrata dei dizionari-fonte sono state rappresentate mediante il modulo core di OntoLex-Lemon,[48] per cui i diversi lessemi (*cane*, *chien*, *gato*) sono stati modellati come LexicalEntry. Mediante la proprietà lexicalForm ogni entrata è stata associata a due forme lessicali, rispettive realizzazioni del numero grammaticale, alle quali corrispondono altrettante realizzazioni grafematiche. La forma al singolare, che può essere considerata il lemma dizionariale, viene collegata all'entrata tramite la proprietà canonicalForm, mentre si usa otherForm per il plurale. L'aspetto grafematico di ogni forma è espresso da un valore letterale associato alla proprietà ontolex:writtenRep.

Le informazioni di POS e genere delle LexicalEntry così come quelle di numero delle singole forme sono state espresse mediante l'utilizzo di LexInfo[49] (Listing 2), un'ontologia di categorie linguistiche per la rappresentazione delle risorse lessicali (cf. [20]: 151). Nel nostro esempio abbiamo utilizzato le proprietà lexinfo:partOfSpeech e lexinfo:gender per associare le istanze di LexicalEntry rispettivamente alle classi Noun e Masculine di LexInfo. Inoltre, abbiamo impiegato lexinfo:grammaticalNumber per collegare canonicalForm e otherForm al rispettivo numero grammaticale.

[46] Il lessico modellato nel presente articolo non è attualmente accessibile tramite un endpoint SPARQL pubblico, in quanto alcuni materiali sono ancora in fase di revisione. L'accesso sarà reso disponibile al termine di tale processo.

[47] In questa sede non ci occupiamo degli strumenti per la compilazione lessicografica LD, per i quali rimandiamo a [27].

[48] In accordo con le indicazioni fornite dagli sviluppatori del modulo lexicog, non abbiamo ritenuto necessario il ricorso al modulo lessicografico per la modellazione di informazioni che possono essere codificate con core. (cf. https://www.w3.org/2019/09/lexicog/#aim-and-scope)

[49] https://lexinfo.net.

```
# LexicalEntry "cane"
lei:cane a ontolex:LexicalEntry ;
lexinfo:partOfSpeech lexinfo:noun ;
lexinfo:gender lexinfo:masculine ;
ontolex:canonicalForm lei:cane_form ;
ontolex:otherForm lei:cani_form ;
ontolex:sense lei:cane_sense ;
ontolex:sense lei:cane_sense2.

# Canonical form
lei:cane_form a ontolex:Form ;
ontolex:writtenRep "cane"@it ;
lexinfo:grammaticalNumber lexinfo:singular .

# Other form (plural)
lei:cani_form a ontolex:Form ;
ontolex:writtenRep "cani"@it ;
lexinfo:grammaticalNumber lexinfo:plural .
```

Listing 2 Modellazione della LexicalEntry lei:cane e delle sue forme.

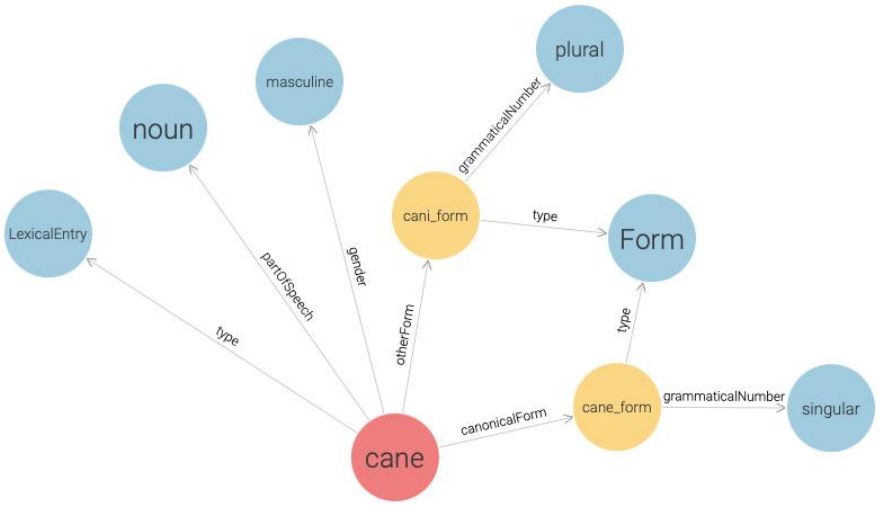

Figura 9 Rappresentazione delle informazioni morfologiche.

Come mostrato in Figura 7, il modulo core di OntoLex-Lemon offre diverse soluzioni per la modellazione delle informazioni semantiche legate a un lessema. Alcuni progetti utilizzano

risorse esterne come DBpedia[50], Wikidata[51] o WordNet per la rappresentazione di tale dominio. Tuttavia, risorse di questo tipo sono state spesso valutate come insufficienti per la lessicografia storica dove, soprattutto negli ultimi anni, sono stati presi in considerazione sia la creazione di ontologie di dominio relative a determinati stadi linguistico-culturali, sia il riutilizzo di strumenti propri della lessicografia onomasiologica tradizionale.

Andrea Bozzi e Damiana Luzzi evidenziano come il *Dictionnaire de Termes Médico-botaniques de l'Ancien Occitan* (DiTMAO) abbia abbandonato l'idea di una mappatura tramite WordNet, preferendo lo sviluppo di un'ontologia di dominio specifica per la rappresentazione della "classificazione medievale dei termini medici e botanici" ([13]: 61) dell'occitano antico. Sabine Tittel e Christian Chiarcos [68], dopo aver illustrato la mappatura semantica di lessemi del francese antico con le entrate di DBpedia, sottolineano il cosiddetto *Historical Semantic Gap*, ovvero la difficoltà di mettere in corrispondenza risorse linguistiche relative a fasi storiche della lingua con sistemi concettuali pensati per applicazioni e dati contemporanei. Su queste considerazioni si fonda uno degli obiettivi principali del progetto *Wissensnetze in der mittelalterlichen Romania* (ALMA), il quale mira allo sviluppo di ontologie storicizzate per la rappresentazione della specificità storica dei sistemi concettuali medievali, anche a partire da sistemi preesistenti [58].

Secondo i principi di onomasiologia presentati da Andreas Blank e Peter Koch [12], la classificazione semantica del lessico deve basarsi su reti concettuali compilate a partire dall'analisi semasiologica, poiché proprio questo tipo di analisi mette in evidenza i concetti lessicalizzati dai segni linguistici anche dal punto di vista diacronico; tali concetti possono essere sistemati e utilizzati in prospettiva onomasiologica per la classificazione semantica del lessico. Gli autori ammettono la possibilità di riutilizzare sistemi nati dall'analisi semasiologica di una lingua per la classificazione di un'altra lingua: gli studi tipologici evidenziano la natura limitata della varianza concettuale interlinguale e, almeno per quanto riguarda le aree linguistiche romanze e germaniche (focus del nostro interesse nel presente contributo), si può parlare di una certa affinità dei sistemi semantici (cf. [22]: 85, [8]: 124).

Anche alla luce di queste basi teoriche, alcuni lessicografi hanno valutato il riutilizzo di sistemi concettuali presenti in altre opere e la loro conversione in RDF.

Tittel, Frances Gillis-Webber e Alessandro Alfredo Nannini [69] hanno presentato la conversione in RDF del *Begriffssystem* di Rudolf Hallig e Walther von Wartburg, a partire da una preesistente versione in formato XML. Il sistema concettuale viene convertito sia in base di conoscenza SKOS[52] che in ontologia OWL.[53] Da un lato il contributo mette in evidenza la fattibilità di una tale conversione, mentre dall'altro si concentra sulle carenze del *Begriffssystem*, sottolineando la necessità di integrare il sistema con concetti storicizzati al fine di superare il già menzionato *Historical Semantic Gap*.

Sanders Stolk ha convertito il *Thesaurus of Old English* in risorsa RDF utilizzando *lemon tree*, un'estensione di SKOS sviluppata appositamente per la lessicografia onomasiologica [65]. Il sistema è stato quindi riutilizzato da progetti lessicologici e lessicografici relativi ad altre varietà linguistiche ([55], [72], [39], [23]).

[50] https://www.dbpedia.org.

[51] https://www.wikidata.org.

[52] https://www.w3.org/TR/skos-reference/.

[53] https://www.w3.org/TR/owl-features/.

Mariafrancesca Giuliani e Itziar Molina Sangüesa [28], nell'ambito di una collaborazione tra il DHLE [63] e TLIO [51], hanno modellato una sezione dell'*Historical Thesaurus of English* in OWL e hanno utilizzato l'ontologia per la mappatura semantica del lessico medico presente nei due dizionari, dimostrando l'efficacia di tale metodo.

In maniera analoga viene svolta la mappatura semantica del *Lessico Etimologico Italiano* (LEI), i cui materiali editi, attualmente disponibili in formato XML, vengono collegati alle classi di HisThes, un'ontologia in OWL che si ispira parzialmente al sistema concettuale dell'HTE. In HisThes i concetti vengono modellati come classi relazionate gerarchicamente tra loro mediante l'assioma SubClassOf. A ogni classe vengono associate *labels* in inglese, italiano e francese attraverso la proprietà rdfs:label (Listing 3).

```
# Classe n_Dog in HisThes
histhes:n_Dog rdf:type owl:Class ;
          rdfs:subClassOf <http://www.semanticweb.org
/HistoThesLEI#01_05_19_05_03_01_Dog> ;
          rdfs:label "Cane"@it ,
                     "Chien"@fr ,
                     "Dog"@en .
```

Listing 3 Modellazione della classe n_Dog in HisThes

Le limitazioni del sistema ad albero di partenza vengono superate mediante l'utilizzo di altri assiomi OWL (EquivalentClasses e DisjointClasses), così come dei *logical class constructors* che permettono di esprimere relazioni complesse basate sugli operatori logici AND, OR e NOT.[54]

La testata efficacia del sistema inglese come base concettuale per la classificazione di altre lingue germaniche e romanze deriva anche dal fatto che i suoi concetti emergono dall'analisi semasiologica del lessico inglese nella sua massima ampiezza diacronica. La considerazione di diversi stadi cronologici è responsabile della presenza di concetti o campi concettuali legati a determinati ambiti socio-culturali sia del passato che, in virtù dei fenomeni di contatto, esterni all'area anglofona.[55]

La mappatura mediante un sistema utilizzato da altre risorse può consentire di cogliere analogie e differenze tra le strutture semantiche di lingue diverse. Inoltre, il formato digitale permette di integrare facilmente il sistema originario con concetti o campi concettuali lessicalizzati dalle varietà italoromanze in analisi.

In accordo con Tittel ([67]: 127) abbiamo utilizzato le diverse possibilità offerte da OntoLex-Lemon per modellare la semantica del nostro lessico secondo il trapezio semiotico di Andreas Blank e, sulla base delle argomentazioni teoriche e pratiche appena presentate, ci siamo serviti delle classi di HisThes per mappare il materiale lessicografico.

L'entrata, assimilabile al segno linguistico, è composta da una parte formale e da una parte semantica intralinguistica che viene modellata con la classe LexicalSense. Quest'ultimo elemento lessicalizza un'astrazione mentale – un concetto – che viene espresso con LexicalConcept,

[54] Cf. https://www.w3.org/TR/owl2-primer/.

[55] In uno studio del 2023 [50] ho dimostrato l'efficacia del sistema HTE e dell'ontologia HisThes per la mappatura degli etimi latini presenti nel LEI.

sottoclasse di SKOS:Concept, che mediante la proprietà skos:definition può contenere una defizione in valori letterali.

L'entrata lessicale può essere collegata a un'entità ontologica esterna in diversi modi. Nel nostro caso abbiamo messo in relazione il LexicalConcept, già collegato al significato lessicale, con la rispettiva classe ontologica di HisThes utilizzando la proprietà isConceptOf. Questa modellazione, legando il LexicalSense al LexicalConcept e quest'ultimo all'entità ontologica, riprende esplicitamente il modello teorico di riferimento. Più in dettaglio, si intende rappresentare il processo di significazione per cui il segno linguistico non si riferisce direttamente alla realtà ontologica, bensì ai concetti in quanto rappresentazioni mentali di quella ([73]: CXII).[56]

[56] A tal proposito, Coseriu ([22]: 187) richiama la distinzione scolastica tra *conceptus* e *res*, così come quella operata da Wilhelm von Humboldt tra *innere Sprachform* e *Gegenstand*. In buona sostanza, il segno linguistico lessicalizza un concetto che è, a sua volta, una rappresentazione mentale della realtà extralinguistica. La modellazione qui proposta permette di porre l'accento sui processi di significazione delle diverse comunità linguistiche, laddove il legame diretto tra segno linguistico e istanza ontologica registrerebbe soltanto un processo di nomenclatura che, secondo gli studi di semantica, non è rappresentativo delle effettive modalità di produzione linguistica.

```
# LexicalSense 1 di di "cane"
lei:cane_sense a ontolex:LexicalSense ;
ontolex:isSenseOf lei:cane ;
ontolex:isLexicalizedSenseOf lei:cane_concept .

# LexicalConcept 1 di "cane"
lei:cane_concept a ontolex:LexicalConcept ;
skos:definition "gènere di mammìferi carnìvori
appartenente alla famiglia cànidi; ne esiste un grande
nùmero di razze, classificate, secondo l'uso a cui sono
destinate (Canis familiaris)."@it ;
ontolex:isConceptOf histhes:n_Dog .

# LexicalSense 1 di "chien"
few:chien_sense a ontolex:LexicalSense ;
ontolex:isSenseOf few:chien ;
ontolex:isLexicalizedSenseOf few:chien_concept .

# LexicalConcept 1 di "chien"
few:chien_concept a ontolex:LexicalConcept ;
skos:definition "canis domesticus"@la ;
ontolex:isConceptOf histhes:n_Dog .

# LexicalSense 1 di "gato"
rae:gato_sense a ontolex:LexicalSense ;
ontolex:isSenseOf rae:gato ;
ontolex:isLexicalizedSenseOf rae:gato_concept .

# LexicalConcept 1 di "gato"
rae:gato_concept a ontolex:LexicalConcept ;
skos:definition "Mamífero carnívoro de la familia de los
félidos, en particular el doméstico, de hasta 70
centímetros de longitud, de cabeza redondeada, hocico
triangular con bigotes, orejas puntiagudas y lengua
áspera, cuerpo generalmente cubierto de pelo de color
negro, blanco, gris, marrón o amarillento, garras con uñas
retráctiles y cola larga. Nombre científico: Felis
silvestris catus."@es ;
skos:prefLabel "gato"@it ;
ontolex:isConceptOf histhes:n_Felis_domesticus_cat .
```

Listing 4 Modellazione delle informazioni semantiche e mappatura con HisThes.

Nel Listing 4 l'entrata italiana *cane* risulta collegata a un LexicalSense che è senso lessicalizzato del LexicalConcept definito 'gènere di mammìferi carnìvori appartenente alla famiglia cànidi; ne esiste un grande nùmero di razze, classificate, secondo l'uso a cui sono destinate (Canis familiaris)' (LEI X, 860). Il LexicalConcept rimanda, tramite la proprietà isConceptOf, alla classe

n_Dog di Histhes che rappresenta la classe di un'entità che viene etichettata come *dog* in inglese, *cane* in italiano e *chien* in francese.

La Figura 10 mostra il grafo con le informazioni semantiche delle entrate *chien* e *cane*, dove l'entità ontologica costituisce il raccordo tra due significati di lessemi diversi.

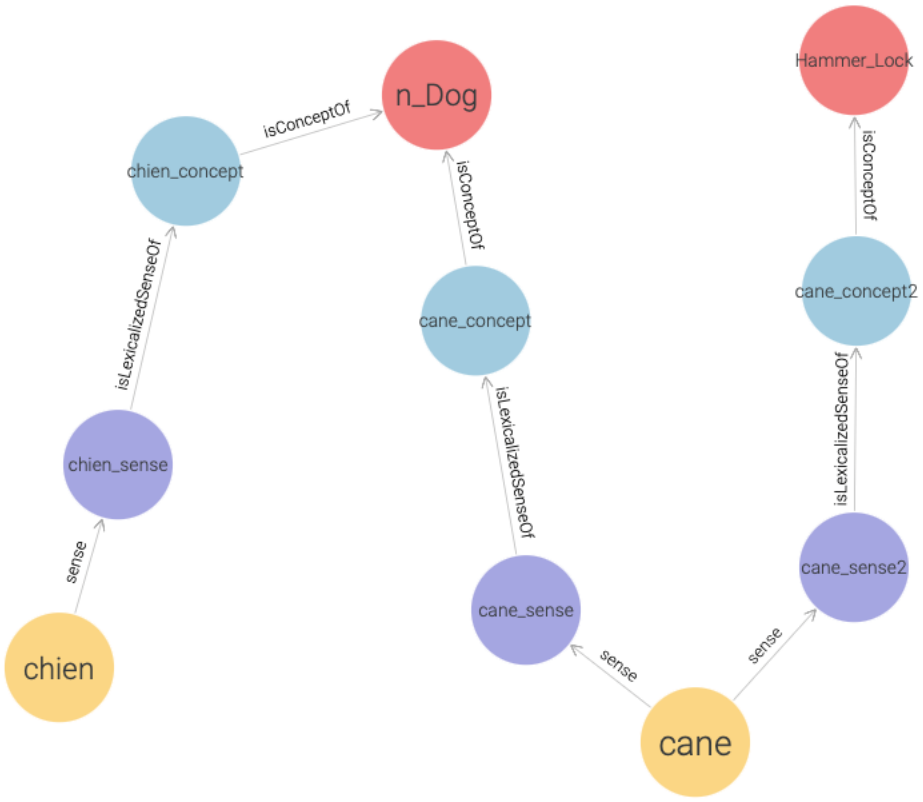

Figura 10 Rappresentazione delle informazioni semantiche mappate con HisThes.

La mappatura di risorse RDF facilita l'interconnessione semantica tra i dati e, come mostrato nel grafo, la visualizzazione sinottica dell'articolazione dei significati delle diverse entrate, così come dei loro legami con la realtà extralinguistica.

4.2 Modellazione del cambiamento semantico

Per la modellazione dei cambiamenti semantici abbiamo integrato due modelli ontologici. Il primo, chiamato *lemonDIA*, viene presentato in [37] come un'estensione di OntoLex-Lemon per la rappresentazione della dimensione temporale in cui si verifica il cambiamento semantico di una LexicalEntry. In un lavoro successivo [38] vengono definite la classe SemanticShift per la modellazione del cambiamento semantico e le proprietà shiftSource e shiftTarget per collegare il cambiamento ai due LexicalSense coinvolti, mentre in [36] viene introdotta la proprietà senseShiftsTo per mettere in relazione il senso base con il senso che ne costituisce lo sviluppo (cf. anche [41]).

Il secondo modello utiizzato è SemShift, sviluppato da Tittel ([67]: 413)[57] per la rappresentazione delle modalità del cambiamento semantico. L'ontologia si basa su considerazioni teoriche che si lasciano ricondurre ai lavori di semantica lessicale di Andreas Blank. Nella teoria del cambiamento semantico di Blank [11] le relazioni associative tra concetti occupano un ruolo fondamentale. I percorsi (o modalità) del cambiamento semantico dipendono dalle associazioni concettuali che coinvolgono i parlanti; tali associazioni possono dar vita a un'innovazione semantica che produce e lessicalizza un nuovo significato di un lessema già esistente. Ad esempio, lessema inglese *mouse* sviluppa un significato informatico ('mouse del computer') a partire dal significato animale ('topo'). Un'associazione di somiglianza tra le caratteristiche salienti dei due designati (l'animale e il dispositivo) sta alla base di un cambiamento semantico di tipo metaforico ([11]: 150).

SemShift presenta un ampio set di proprietà per la rappresentazione dei percorsi del cambiamento semantico tra due LexicalSense. Inoltre, Tittel definisce la proprietà associativeRelation che ha come dominio LexicalSense e come range un gruppo di classi che esprimono le relazioni associative che motivano l'innovazione semantica (MetaphoricalSimilarity; ConceptualContiguity; TaxonomicSuperordination; TaxonomicSubordination).

Per rappresentare il fenomeno del cambiamento in maniera dettagliata abbiamo optato per l'utilizzo congiunto dei due modelli dal momento che lemonDIA esprime gli elementi coinvolti nel cambiamento e la direzione senso primario > senso secondario,[58] mentre SemShift rende esplicite le motivazioni e i percorsi di tale sviluppo.

Accanto al significato collegato al LexicalConcept definito in relazione all'animale, l'entrata *cane* presenta un significato metaforico (cane_sense2) collegato al concetto cane_concept2 così definito dal LEI: 'nelle armi da fuoco, òrgano ricurvo, a forma di martelletto, che scatta e percuote le càpsule per lo sparo, grilletto; nelle armi da fuoco antiche, il ferro che conteneva la pietra focaia' (Listing 5).

```
# LexicalConcept 2 di cane
lei:cane_concept2 a ontolex:LexicalConcept ;
skos:definition "nelle armi da fuoco, òrgano ricurvo, a
forma di martelletto, che scatta e percuote le càpsule per
lo sparo, grilletto; nelle armi da fuoco antiche, il ferro
che conteneva la pietra focaia."@it ;
skos:prefLabel "cane delle armi antiche"@it ;
ontolex:isConceptOf histhes:Hammer_Lock .
```

Listing 5 LexicalConcept associato al secondo significato di *cane*.

Dal primo significato della voce spagnola *gato*, procede per sineddoche il significato alimentare 'Carne de gato, generalmente doméstico', legato all'usanza di mangiare la carne del felino in questione (Listing 6).

[57] https://github.com/SabineTittel/Historische-lexikalische-Semantik-und-Linked-Data/blob/main/semshift.owl.

[58] In questa sede non viene presa in considerazione la modellazione della dimensione temporale, sebbene rappresenti uno degli aspetti più interessanti di lemonDIA.

```
# LexicalConcept 2 di gato
rae:gato_concept2 a ontolex:LexicalConcept ;
skos:definition "Carne de gato, generalmente
doméstico."@es ;
skos:prefLabel "Carne de gato"@es ;
ontolex:isConceptOf histhes:Cat_Flesh .
```

Listing 6 LexicalConcept associato al secondo significato di *gato*.

lemonDIA permette di modellare la genesi di un significato mediante la proprietà senseShiftsTo che collega il primo significato al secondo, mentre il cambiamento viene modellato con la classe SemanticShift che si lega con shiftSource e shiftTarget rispettivamente al primo e al secondo significato (Listing 7).

```
# Lexical Sense 1
lei:cane_sense a ontolex:LexicalSense ;
ontolex:isSenseOf lei:cane ;
ontolex:isLexicalizedSenseOf lei:cane_concept ;
lemondia:senseShiftsTo lei:cane_sense2 .

lei:shift_cane a lemondia:SemanticShift ;
lemondia:shiftSource lei:cane_sense ;
lemondia:shiftTarget lei:cane_sense2 ;
rdfs:comment "Il secondo significato si sviluppa per
metafora in virtù della somiglianza dei referenti" .
```

Listing 7 Modellazione del cambiamento semantico con lemonDIA.

SemShift permette invece di esprimere il processo di cambiamento che lega l'innovazione (sense2) al significato da cui essa procede (sense1) e la relazione associativa che motiva tale processo. Nel caso di *cane* un'associazione di somiglianza (MetaphoricalSimilarity) motiva un processo metaforico (metaphor), mentre un'associazione di contiguità (ConceptualContiguity) produce il cambiamento per sineddoche (synechdoche) di *gato* (Listing 8 e 9).

```
# LexicalSense 2 di cane
lei:cane_sense2 a ontolex:LexicalSense ;
ontolex:isLexicalizedSenseOf lei:cane_concept2 ;
ontolex:isSenseOf lei:cane ;
semshift:metaphor lei:cane_sense ;
semshift:associateRelation
semshift:MetaphoricalSimilarity .
```

Listing 8 Modellazione del secondo significato di *cane* con SemShift.

```
# LexicalSense 2 di gato
rae:gato_sense2 a ontolex:LexicalSense ;
ontolex:isLexicalizedSenseOf rae:gato_concept2 ;
ontolex:isSenseOf rae:gato ;
semshift:synechdoche rae:gato_sense ;
semshift:associateRelation semshift:ConceptualContiguity .
```

Listing 9 Modellazione del secondo significato di *gato* con SemShift.

La figura 11 mostra la rappresentazione grafica del cambiamento semantico di *gato* modellato con le due ontologie. Come si vede, il grafo favorisce una consultazione visivamente dinamica del passaggio semantico, rendendo evidente il processo cognitivo che ne sta alla base.

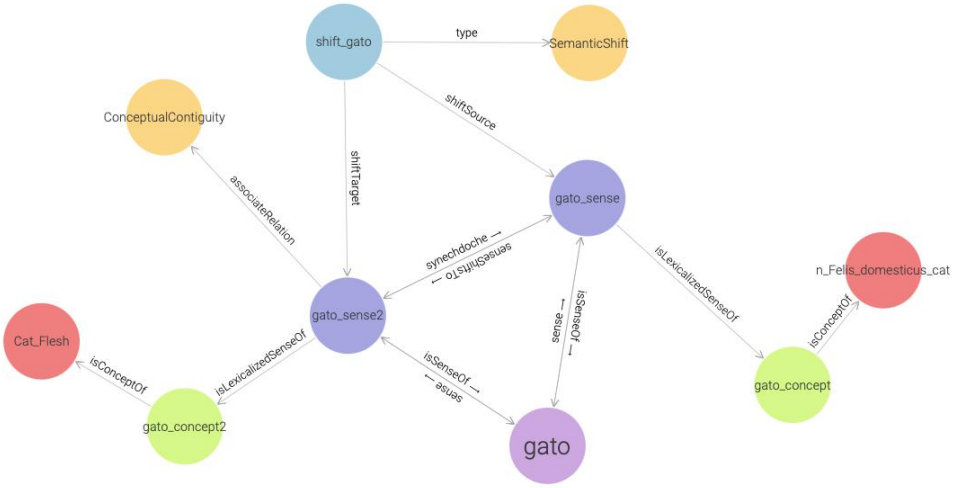

Figura 11 Grafo del cambiamento semantico di *gato*.

4.3 lemonEty per la rappresentazione delle informazioni etimologiche

Oltre alle informazioni formali e semantiche abbiamo rappresentato le informazioni etimologiche relative a ciascuna entrata. Gli etimi latini delle entrate romanze sono stati tratti dal LEI e modellati con lemonEty, estensione di OntoLex-Lemon per la codifica delle informazioni etimologiche, nel quale l'etimologia viene rappresentata come ipotesi linguistica sulla storia di un dato elemento lessicale [35].[59]

Etymology è la classe principale di lemonEty e costituisce una reificazione dell'associazione etimologica tra elementi lessicali. Il collegamento tra l'etimologia e l'elemento lessicale è espresso dalla proprietà etymology che ha come range la classe Etymology e come come dominio una LexicalEntry: questa modellazione permette di trattare facilmente quei casi – ben noti agli etimologi – in cui un'entrata può essere associata a diverse ipotesi etimologiche. La proprietà etymon collega Etymology alla classe Etymon, sottoclasse di LexicalEntry per la modellazione

[59] http://lari-datasets.ilc.cnr.it/lemonEty.

dei lessemi che vengono considerati etimi in una data etimologia. Inoltre, il modello offre "classes and properties for specifying an etymological link (EtyLink) between two elements of an OntoLex-Lemon lexicon, one of which is the source (which we specify by etySource), and the other is the target of the etymological link (etyTarget)" [41]: la classe EtyLink costituisce un raccordo tra l'etimo e il suo sviluppo etimologico.[60]

Applicando questa modellazione, gli elementi lessicali del nostro esempio sono stati collegati agli individual di Etymology che, attraverso la proprietà etymology, rimandano agli etimi del LEI modellati come Etymon. Il collegamento tra le entità linguistiche è stato reso anche con l'utilizzo della classe EtyLink che, tramite le proprietà etySource ed etyTarget, costituisce il raccordo tra l'etimo e l'entrata lessicale. Infine, i diversi etimi sono stati collegati alle voci di LiLa con la proprietà rdfs:seeAlso[61] (Listing 10).

```
# Modellazione dell'etimologia di cane
lei:cane lemonety:etymology lei:cane_etymology .

lei:cane_etymology a lemonety:Etymology ;
lemonety:etymon lei:canis_kan_etymon ;
lemonety:hasEtyLink lei:cane_etylink .

lei:canis_kan_etymon a lemonety:Etymon ;
rdfs:seeAlso <https://lila-erc.eu/data/id/lemma/92575> .

lei:cane_etylink a lemonety:EtyLink ;
lemonety:etySource lei:canis_kan_etymon ;
lemonety:etyTarget lei:cane .
```

Listing 10 L'etimologia di cane modellata con lemonEty.

Nel grafo in Figura 12 vengono visualizzati l'etimologia di ogni entrata e gli eventuali rapporti di parentela etimologica che intercorrono all'interno del lessico.

[60] lemonEty viene utilizzato per la modellazione delle informazioni etimologiche del lessico latino nel progetto LiLa [45].

[61] "rdfs:seeAlso is an instance of rdf:Property that is used to indicate a resource that might provide additional information about the subject resource" (https://www.w3.org/TR/rdf-schema/#ch_seealso).

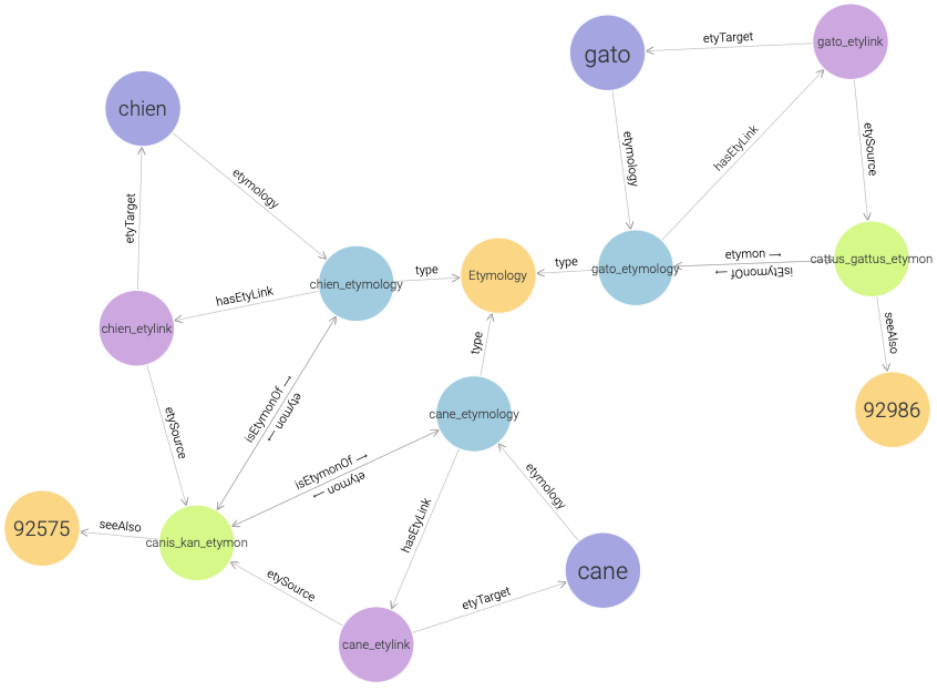

Figura 12 Rappresentazione delle etimologie.

Come si può intuire anche dalla visualizzazione, la modellazione delle etimologie come ipotesi in lemonEty risulta particolarmente interessante poiché permette di collegare più etimologie a una stessa entrata, rendendo accessibili all'utente le diverse posizioni avanzate dagli etimologi in merito a un singolo lessema.

5. Interrogazione del grafo lessicografico

In questa parte del contributo forniamo una serie di esempi volti a illustrare le possibilità d'interrogazione SPARQL del grafo modellato nei paragrafi precedenti, concentrandoci principalmente sull'utilizzo di query per il recupero di informazioni semantiche ed etimologiche.[62]

5.1 Semasiologia e onomasiologia

Per quanto riguarda l'approccio semasiologico, si possono utilizzare query per l'estrazione delle informazioni semantiche associate alle diverse entrate (LexicalSense, LexicalConcept e OntologyEntity) (Tabella 2).

[62] Rimandiamo all'*Appendice* per un esempio di interrogazione del grafo in linguaggio naturale.

?lexicalEntry	**?canonicalFormWrittenRep**	**?sense**	**?definition**	**?ontologyEntity**
few:chien	chien@fr	few:chien_sense	canis domesticus@la	histhes:n_Dog
lei:cane	cane@it	lei:cane_sense	gènere di mammìferi carnìvori [...] (Canis familiaris).@it	histhes:n_Dog
lei:cane	cane@it	lei:cane_sense2	nelle armi da fuoco, òrgano ricurvo, a forma di martelletto [...].@it	histhes:Hammer_Lock
rae:gato	gato@es	gato:sense	Mamífero carnívoro de la familia de los félidos [...]. Nombre científico: Felis silvestris catus.@es	histhes:n_Felis_domesticus_cat
rae:gato	gato@es	rae:gato_sense2	Carne de gato, generalmente doméstico.@es	histhes:Cat_Flesh

Tabella 2 Informazioni semantiche relative alle tre voci.

Sul versante onomasiologico possono essere effettuate diverse query per recuperare il materiale lessicografico a partire dalle informazioni semantiche. Tra queste risultano decisamente utili quelle che prendono in considerazione le classi di HisThes. Il Listing 11 mostra una query per il recupero delle entrate collegate alla classe n_Dog di HisThes; il risultato è illustrato nella Tabella 3.

```
SELECT ?ontologyEntityLabelIT ?lexicalEntry ?sense ?definiti
on
WHERE {
  # Recupera l'entrata lessicale e i suoi sensi
  ?lexicalEntry a ontolex:LexicalEntry ;
                ontolex:sense ?sense .

  # Recupera il concetto lessicale e la definizione
  ?sense ontolex:isLexicalizedSenseOf ?concept .
  ?concept skos:definition ?definition ;
           ontolex:isConceptOf histhes:n_Dog .  # Filtro per
l'entità ontologica histhes:n_Dog

   # Recupera il nome italiano dell'entità ontologica
(histhes:n_dog)
  ?ontologyEntity rdfs:label ?ontologyEntityLabelIT .
  FILTER(?ontologyEntity = histhes:n_Dog)
FILTER(LANG(?ontologyEntityLabelIT) = "it")
}
ORDER BY ?lexicalEntry
```

Listing 11 Query per il recupero delle entrate i cui sensi lessicalizzano dei concetti collegati alla classe n_Dog.

?ontologyEntityLabelIT	**?lexicalEntry**	**?sense**	**?definition**
Cane@it	few:chien	few:chien_sense	canis domesticus@la
Cane@it	lei:cane	lei:cane_sense	gènere di mammìferi carnìvori [...]. (Canis familiaris).@it

Tabella 3 Risultato della query di Listing 11.

L'utilizzo del formato RDF permette di effettuare query più complesse che sfruttano le inferenze ontologiche per il recupero di informazioni implicite nel modello, come la strutturazione di campi semantici. Ad esempio, i significati denotativi di *chien*, *cane* e *gato* possono essere considerati nella loro partecipazione ai campi ANIMALI, QUADRUPEDI, MAMMIFERI. Per questo motivo abbiamo scritto una query che restituisce tutte le entrate lessicali collegate a classi di HisThes che discendono da almeno una classe il cui nome (la label in lingua italiana) contiene una determinata stringa immessa in input.

Nel Listing 12 riportiamo la query per ottenere tutte le entrate che hanno semanticamente a che fare con le classi individuate dalla stringa mammiferi.[63] Il risultato viene presentato nella Tabella 4.

```
SELECT
DISTINCT ?class ?ontologyEntity ?ontologyEntityLabel ?lexi
calEntry ?sense ?definition

WHERE {
  # Recupera l'entrata lessicale e i suoi sensi
  ?lexicalEntry a ontolex:LexicalEntry ;
                ontolex:sense ?sense .

  # Recupera il concetto lessicale e la definizione
  ?sense ontolex:isLexicalizedSenseOf ?concept .
  ?concept skos:definition ?definition .

  # Recupera tutte le entità ontologiche collegate al
concetto con prefisso 'histhes'
  ?ontologyEntity rdfs:label ?ontologyEntityLabel .
  ?concept ontolex:isConceptOf ?ontologyEntity  .
  FILTER(STRSTARTS(STR(?ontologyEntity),
"http://www.semanticweb.org /HistoThesLEI#"))
  FILTER(LANG(?ontologyEntityLabel) = "it")

    # Seleziona le classi ontologiche che sono sottoclassi
di una classe con ["stringa"] nella label italiana

  ?class rdfs:label ?classLabel .
  ?ontologyEntity rdfs:subClassOf ?class .
  FILTER(CONTAINS(LCASE(STR(?classLabel)), "mammiferi"))
  FILTER(LANG(?classLabel) = "it")

}

ORDER BY ?ontologyEntityLabel ?lexicalEntry
```

Listing 12 Query SPARQ per ricerche semantiche mirate.

[63] Utilizziamo DISTINCT come modificatore di SELECT per rimuovere eventuali duplicati dai risultati dal momento che una classe può essere figlia di più classi che abbiano una determinata stringa nella label.

?ontologyEntity	?ontologyEntityLabel	?lexicalEntry	?sense	?definition
histhes:n_Dog	Cane@it	few:chien	few:chien_sense	canis domesticus@la
histhes:n_Dog	Cane@it	lei:cane	lei:gato_sense	gènere di mammìferi carnìvori [...] (Canis familiaris).@it
histhes:n_Felis_domesticus_cat	Gatto@it	rae:gato	rae:gato_sense	Mamífero carnívoro de la familia de los félidos, [...]. Nombre científico: Felis silvestris catus.@es

Tabella 4 Risultato della query di Listing 12.

La mappatura semantica del lessico con un sistema esterno permette un accesso onomasiologico indipendente dalle lingue in analisi. In questo modo, anche grazie alla modellazione dei concetti come nodi tra i significati linguistici e le entità ontologiche, si possono cogliere differenze e analogie tra le strutture semantiche delle singole lingue.

5.2 Semantica storica

Il lessico modellato con SemShift e lemonDIA può essere interrogato in prospettiva diacronica. La Tabella 5 mostra il risultato di una query SPARQL che, per ogni cambiamento semantico, restituisce la definizione associata al primo significato, l'associazione concettuale che ha portato all'innovazione semantica, il processo del cambiamento, la definizione del nuovo significato e un commento in valori letterali.

?SemanticShift	?definitionSource	?semshift	?relation	?definitionTarget	?comment
lei:shift_cane	gènere di mammìferi carnìvori appartenente alla famiglia cànidi [...]	semshift:MetaphoricalSimilarity	semshift:metaphor	nelle armi da fuoco, òrgano ricurvo, a forma di martelletto [...].@it	Il secondo significato si sviluppa per metafora in virtù della somiglianza dei referenti

rae:shift_gato	Mamífero carnívoro de la familia de los félidos, en particular el doméstico [...]	semshift:ConceptualContiguity	semshift:synechdoche	Carne de gato, generalmente doméstico.@es	Il secondo significato si sviluppa per sineddoche

Tabella 5 Risultato di una query per il recupero dei cambiamenti semantici.

Come accennato in precedenza, l'impiego congiunto delle due ontologie per la descrizione dei fenomeni di semantica storica favorisce una consultazione dinamica del materiale e, in vista di un possibile utilizzo su larga scala in lessici di lingue diverse, lascia ben sperare nell'avanzamento degli studi sugli aspetti sociolinguistici e cognitivi connessi ai fenomeni diacronici.

5.3 Informazioni etimologiche

Infine, abbiamo utilizzato SPARQL per il recupero delle informazioni etimologiche modellate con lemonEty. La query del Listing 13 recupera l'entrata, la forma messa a lemma, l'etimo e la voce di LiLa collegata ad esso dalla proprietà rdfs:seeAlso (Tabella 6).

```
SELECT ?lexicalEntry ?lemma ?etySource ?LILAEntry  WHERE {
    ?etymology a lemonety:Etymology .
    ?etymology lemonety:hasEtyLink ?etyLink .
    ?etyLink lemonety:etySource ?etySource .
    ?etySource rdfs:seeAlso ?LILAEntry .
    ?etyLink lemonety:etyTarget ?lexicalEntry .
    ?lexicalEntry ontolex:canonicalForm ?form .
    ?form ontolex:writtenRep ?lemma .
}
```

Listing 13 Query per il recupero delle informazioni etimologiche.

?lexicalEntry	**?lemma**	**?etySource**	**?LILAEntry**
lei:cane	cane	lei:canis_kan_etymon	https://lila-erc.eu/data/id/lemma/92575
few:chien	chien	lei:canis_kan_etymon	https://lila-erc.eu/data/id/lemma/92575
rae:gato	gato	lei:cattus_gattus_etymon	https://lila-erc.eu/data/id/lemma/92986

Tabella 6 Informazioni etimologiche.

La modellazione dell'etimologia permette anche un approccio genetico alternativo nel quale il materiale romanzo viene recuperato a partire dagli etimi. Questa prospettiva risulta particolarmente utile per ottenere raggruppamenti di cognati romanzi ed extraromanzi trattati nelle diverse risorse lessicografiche.

6. Conclusioni

Nel percorso che abbiamo tentato di tracciare risultano evidenti i benefici che la svolta digitale ha apportato alla vocabolaristica.

La linguista Carla Marello notava un duplice effetto della digitalizzazione delle opere lessicografiche.

> "Infine, le versioni elettroniche se, da un lato, esaltano quanto c'è di buono in un dizionario, liberandolo dalla camicia di forza dell'ordine alfabetico, facendolo risaltare anche quando è nascosto tra le pieghe di una definizione lunga e complessa, dall'altro impietosamente rivelano all'occhio attento i difetti di un dizionario, le sue incongruenze, e sono quindi un ottimo strumento di igiene lessicografica, di allenamento per i futuri lessicografi." ([46]: 164)

Nel mondo digitale gli strumenti lessicografici non vengono più concepiti come dizionari tradizionali contraddistinti da modalità di compilazione, accesso e consultazione fisse e raramente sottoposte ad autocritica. I sistemi informatici forniscono risorse nuove e migliori per la rappresentazione delle informazioni linguistiche e per l'accesso ad esse, all'interno di un ambiente che consente l'ibridazione con altre risorse.

La lessicografia storica, una volta abbandonato il legame con il mondo cartaceo, è transitata dall'intenzione di ricreare digitalmente quel mondo al riconoscimento dei vantaggi del superamento delle forme e delle strutture analogiche. Negli ultimi anni abbiamo assistito a un incremento dei progetti che adottano il framework RDF e il paradigma Linked Data, nei quali la disponibilità di vocabolari per la descrizione degli aspetti linguistici e la facilità di integrazione con altre risorse rappresentano vantaggi innegabili.

Il nostro esempio di modellazione ha cercato di evidenziare tali vantaggi mostrando l'utilizzo di strumenti aperti e interoperabili per la compilazione e la consultazione lessicografica.

OntoLex-Lemon permette di modellare le entrate lessicali e di integrarle con ontologie esterne per la rappresentazione dei diversi aspetti linguistici. Sul versante della rappresentazione dei significati mediante il ricorso a risorse esterne, abbiamo segnalato un crescente interesse per la creazione di ontologie storicizzate legate a specifici domini linguistico-culturali, presentando le motivazioni teoriche e pratiche che stanno dietro allo sviluppo di strumenti come come HisThes. Tali strumenti permettono la costruzione di reti in cui i nodi delle entità extralinguistiche offrono l'accesso onomasiologico a materiale eterogeneo, distribuito su più lingue e connotato storicamente.

La realizzazione di modelli ontologici in ottica storico-linguistica rende possibile la rappresentazione formale di fenonemi diacronici come il cambiamento semantico e l'etimologia. Infatti, i modelli lemonDIA, SemShift e lemonEty favoriscono lo sviluppo e la visualizzazione di risorse multidimensionali contenenti le descrizioni di diversi aspetti linguistici.

L'impiego congiunto di lemonDIA e SemShift si è dimostrato efficace per la rappresentazione di dimensioni complementari del cambiamento, con il secondo modello che consente di formalizzare le dinamiche cognitive sottese all'innovazione semantica. In questo modo gli elementi di semantica teorica, spesso ignorati dai dizionari tradizionali, assumono un ruolo di rilievo nelle risorse digitali.

lemonEty, già impiegato con successo in ambito storico-linguistico, ha confermato la propria validità nel nostro piccolo lessico romanzo, con potenzialità di estensione alla realizzazione di risorse etimologiche che coinvolgano più lingue.

Un notevole vantaggio di questi modelli va individuato anche nella possibilità di formalizzare asserzioni diverse – anche contrarie tra loro – a proposito di un medesimo argomento linguistico. Tale possibilità rende visibili i rapporti tra le varie ipotesi scientifiche, anche in un'ottica di riflessione critica sulla storia e sullo stato attuale della ricerca.[64]

Infine, le opzioni di interrogazione e visualizzazione offerte dal formato RDF si configurano come particolarmente promettenti in ambito didattico, in quanto la multidimensionalità degli aspetti linguistici risulta efficacemente rappresentata nelle strutture a grafo.

Appendice: parlare con il grafo

Questa sezione d'appendice presenta un esempio di interrogazione del database lessicografico in linguaggio naturale mediante l'utilizzo del chatbot *Talk to Your Graph*,[65] funzione sperimentale di GraphDB. Il chatbot, che impiega i modelli GPT di OpenAI (gpt-4o nel nostro caso), sfrutta la possibilità di configurare assistenti (*agents*) basati sull'Assistants API di OpenAI.

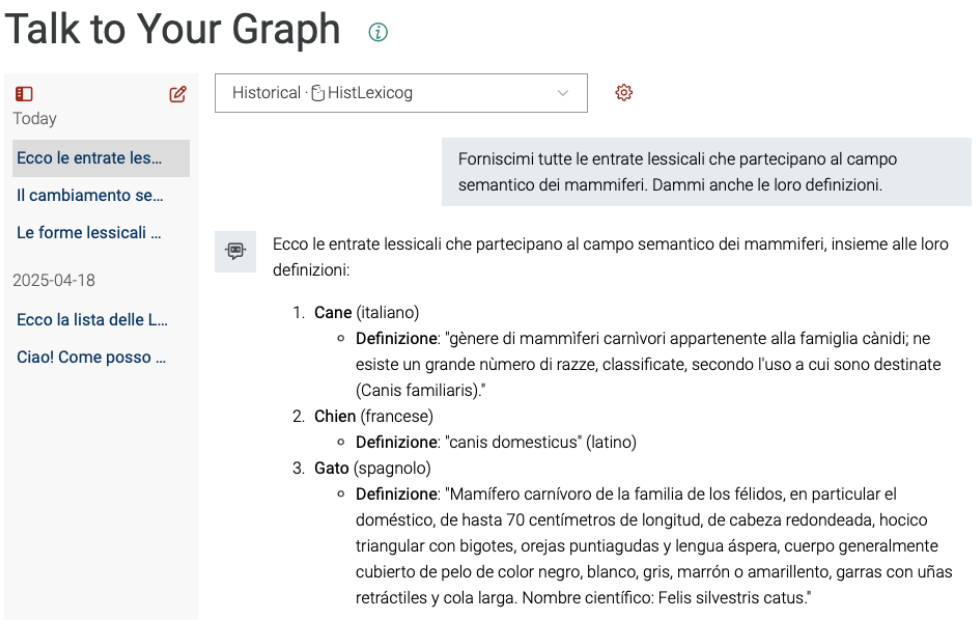

Figura 13 Interrogazione del grafo con Talk to Your Graph

[64] Ci riferiamo ai casi – non esemplificati nel presente contributo – delle etimologie discusse, ma il discorso può essere esteso anche ai cambiamenti semantici e ad altri fenomeni linguistici.

[65] https://graphdb.ontotext.com/documentation/11.0/talk-to-graph.html.

La figura 13 mostra il dialogo con il chatbot che, alla richiesta di restituire le entrate che partecipano al campo semantico dei mammiferi e le loro definizioni, risponde con le informazioni presenti nel lessico.

Bibliografia

[1] Atkins, Sue Beryl T. 1996. «Bilingual Dictionaries: Past, Present and Future». In *Proceedings of the 7th EURALEX International Congress*, a cura di Catalina Röjder Papmehl Martin Gellerstam Jerker Järborg, Sven-Göran Malmgren, Kerstin Norén, Lena Rogström, 515–46. Göteborg, Sweden: Novum Grafiska AB.

[2] Baldinger, Kurt. 1952. «Die Gestaltung des wissenschaftlichen Wörterbuchs». *Romanistisches Jahrbuch* 5: 65–94.

[3] Baldinger, Kurt. 1960. «Alphabetisches oder begrifflich gegliedertes Wörterbuch?». *Zeitschrift für romanische Philologie* 76: 521–36. https://doi.org/10.1515/zrph.1960.76.5-6.521.

[4] Baldinger, Kurt 1966. «Semasiologia e onomasiologia». *Alfa: Revista de Lingüística* 9: 7–36.

[5] Baldinger, Kurt, a c. di. 1971. *Dictionnaire étymologique de l'ancien français*. Québec/Tübingen/Berlin: Presses de L'Université Laval/Niemeyer/De Gruyter.

[6] Baldinger, Kurt, a c. di. 1975a. *Dictionnaire onomasiologique de l'ancien gascon (DAG)*. Tübingen/Berlin: De Gruyter.

[7] Baldinger, Kurt. 1975b. *Dictionnaire onomasiologique de l'ancien occitan (DAO)*. Tübingen: Niemeyer.

[8] Baldinger, Kurt. 1980. *Semantic theory: towards a modern semantics*. Oxford: Blackwell.

[9] Battaglia, Salvatore, a c. di. 1961. *Grande dizionario della lingua italiana (GDLI)*. UTET. 21 voll. Torino.

[10] Berners-Lee, Tim. 2006. «Linked Data». https://www.w3.org/DesignIssues/LinkedData.html.

[11] Blank, Andreas. 1997. *Prinzipien des Lexikalischen Bedeutungswandels Am Beispiel der Romanischen Sprachen*. Tübingen: Niemeyer.

[12] Blank, Andreas, e Peter Koch, a c. di. 2003. *Kognitive romanische Onomasiologie und Semasiologie*. Tübingen: Niemeyer. https://doi.org/doi:10.1515/9783110911626.

[13] Bozzi, Andrea, e Damiana Luzzi. 2016. «Un'ontologia per il DiTMAO (Dictionnaire des Termes Médico-botaniques de l'Ancien Occitan)». In *Projets en cours ; ressources et outils nouveaux*, a cura di D. A. Trotter, Andrea Bozzi, e Cédric Fairon, section 16: 55–63. Actes du XXVIIe Congrès international de linguistique et de philologie romanes. Nancy: ATILF. http://www.atilf.fr/cilpr2013/actes/section-16/CILPR-2013-16-Bozzi-Luzzi.pdf.

[14] Buchi, Eva, e Wolfgang Schweickard, a c. di. 2008. *Dictionnaire Étymologique Roman*. Nancy: ATILF. http://www.atilf.fr/DERom.

[15] Calzolari, Nicoletta. 1988. «The dictionary and the thesaurus can be combined». In *Relational models of the lexicon. Representing knowledge in semantic networks*, a cura di Martha Walton Evens, 75–96. New York: Cambridge University Press.

[16] Campos Souto, Mar. 2020. «La ordenación de las acepciones en el NDHE: Bases para el establecimiento del criterio genético». *Verba* 47: 277–307. https://doi.org/10.15304/VERBA.47.6428.

[17] Casares, Julio. 1921. *Nuevo concepto del diccionario de la lengua. Discurso leído en el acto de su recepción por Julio Casares y Sánchez, y contestación de Antonio Maura y Montaner*. Madrid: G. Koehler.

[18] Casares, Julio. 1942. *Diccionario ideológico de la lengua española: desde la idea a la palabra, desde la palabra a la idea*. Barcelona: Gili.

[19] Cimiano, Philipp, John P. McCrae, e Paul Buitelaar. 2016. «Lexicon Model for Ontologies: Community Report, 10 May 2016: Final Community Group Report 10 May 2016». 1 gennaio 2016. https://www.w3.org/2016/05/ontolex/.

[20] Cimiano, Philipp, Christian Chiarcos, John P. McCrae, e Jorge Gracia. 2020. *Linguistic Linked Data : Representation, Generation and Applications*. Cham, SWITZERLAND: Springer International Publishing AG. http://ebookcentral.proquest.com/lib/sulb/detail.action?docID=6012490.

[21] Ciotti, Fabio. 2012. «Web semantico, linked data e studi letterari: verso una nuova convergenza». In *Dall'Informatica umanistica alle culture digitali. Atti del Convegno di studi in memoria di Giuseppe Gigliozzi (Roma, 27-28 ottobre 2011)*, a cura di Gianfranco Crupi e Fabio Ciotti, 243–76. Quaderni Digilab - Casa Editrice Università La Sapienza.

[22] Coseriu, Eugenio. 1977. *Principios de semántica estructural*. Madrid: Gredos.

[23] Depuydt, Katrien, e Jesse de Does. 2021. «Linking the Dictionary of Old Dutch to A Thesaurus of Old English: A First Exploration». *Amsterdamer Beiträge Zur Älteren Germanistik* 81 (3–4): 493–513. https://doi.org/10.1163/18756719-12340240.

[24] Eco, Umberto. 2007. *Dall'albero al labirinto: studi storici sul segno e l'interpretazione*. Milano: Bompiani.

[25] França, Patrícia Cunha. 2010. «Onomasiological dictionaries and ontologies». In *Proceedings of the XIV EURALEX International Congress. 6-10 July 2010*, a cura di Anne Dykstra e Tanneke Schoonheim, 1291–1298. Leeuwarden: Fryske Akademy.

[26] Geeraerts, Dirk. 2010. *Theories of lexical semantics*. Oxford linguistics. Oxford: Oxford University Press.

[27] Giovannetti, Emiliano, Davide Albanesi, Andrea Bellandi, Simone Marchi, Mafalda Papini, e Flavia Sciolette. 2024. «Maia: an Open Collaborative Platform

for Text Annotation, E-Lexicography, and Lexical Linking». *Umanistica Digitale* 8 (18): 27–52. https://doi.org/10.6092/issn.2532-8816/19705.

[28] Giuliani, Mariafrancesca, e Itziar Molina Sangüesa. 2020. «Hacia una taxonomía integrada en la redacción y revisión de diccionarios históricos». *Bollettino dell'Opera del Vocabolario Italiano*, fasc. XXV: 325–374.

[29] Hallig, Rudolf, e Walther von Wartburg. ²1963. *Begriffssystem als Grundlage für die Lexikographie, Versuch eines Ordnungsschemas*. 2ª ed. Berlin: Akademie-Verlag.

[30] Hartmann, Reinhard R. K. 2005. «PURE OR HYBRID? THE DEVELOPMENT OF MIXED DICTIONARY GENRES». *FACTA UNIVERSITATIS Linguistics and Literature* 3: 193–208.

[31] Heger, Klaus. 1964. «Die methodologischen Voraussetzungen von Onomasiologie und begrifflicher Gliederung». *Zeitschrift für romanische Philologie* 80: 486–516. https://doi.org/10.1515/zrph.1964.80.5-6.486.

[32] Hilty, Gerold. 1971. «Bedeutung als Semstruktur». *Vox romanica* 30: 242–263. https://doi.org/10.5169/seals-24604.

[33] Kay, Christian, a c. di. 2009. *Historical thesaurus of the Oxford English dictionary: with additional material from «A Thesaurus of Old English»*. Oxford ; New York: Oxford University Press.

[34] Kay, Martin. 1983. «The Dictionary of the Future and the Future of the Dictionary». In *The possibilities and limits of the computer in producing and publishing dictionaries: proceedings of the European Science Foundation Workshop, Pisa, 1981*, a cura di Antonio Zampolli e Amadeo Cappelli, 161–74. Pisa: Giardini.

[35] Khan, Anas Fahad. 2018. «Towards the Representation of Etymological Data on the Semantic Web». *Information* 9 (12). https://doi.org/10.3390/info9120304.

[36] Khan, Anas Fahad. 2020. «Representing Temporal Information in Lexical Linked Data Resources». In *Proceedings of the 7th Workshop on Linked Data in Linguistics (LDL-2020)*, a cura di Maxim Ionov, John P. McCrae, Christian Chiarcos, Thierry Declerck, Julia Bosque-Gil, e Jorge Gracia, 15–22. Marseille, France: European Language Resources Association. https://aclanthology.org/2020.ldl-1.3/.

[37] Khan, Anas Fahad, Federico Boschetti, e Francesca Frontini. 2014. «Using lemon to Model Lexical Semantic Shift in Diachronic Lexical Resources». In *3rd Workshop on Linked Data in Linguistics: Multilingual Knowledge Resources and Natural Language Processing (LDL2014), Reykjavik, May 27th, 2014*, 50–54.

[38] Khan, Anas Fahad, Javier E. Díaz-Vera, e Monica Monachini. 2016. «Representing Polysemy and Diachronic Lexico-Semantic Data on the Semantic Web». In *Proceedings of the Second International Workshop on Semantic Web for Scientific Heritage co-located with 13th Extended Semantic Web Conference (ESWC 2016), Heraklion, Greece, May 30th, 2016*, a cura di Isabelle Draelants, Catherine Faron-Zucker, Alexandre Monnin, e Arnaud Zucker, 37–45. CEUR Workshop Proceedings. CEUR-WS.org.

[39] Khan, Anas Fahad, Javier E. Díaz-Vera, Francisco Javier Minaya Gómez, Rafael Cruz González, e Monica Monachini. 2021. «Mapping Conceptual Variation through A Thesaurus of Old English and Evoke: Towards a Topical Thesaurus of Old English Emotional Expressions». *Amsterdamer Beiträge Zur Älteren Germanistik* 81 (3–4): 442–56. https://doi.org/10.1163/18756719-12340238.

[40] Khan, Anas Fahad, Christian Chiarcos, Thierry Declerck, Daniela Gifu, Elena González-Blanco García, Jorge Gracia, Maxim Ionov, et al. 2022. «When linguistics meets web technologies. Recent advances in modelling linguistic linked data». *Semantic Web* 13 (6): 987–1050. https://doi.org/10.3233/SW-222859.

[41] Khan, Anas Fahad, e Francesca Frontini. 2024. «Toward a Representation of Semantic Change in Linked Data». *Languages* 9 (6). https://doi.org/10.3390/languages9060215.

[42] Kipfer, Barbara Ann. 1986. «Investigating an onomasiological approach to dictionary material». *Dictionaries: Journal of the Dictionary Society of North America*, fasc. 8, 55–64.

[43] Lyons, John. 1963. «Structural semantics: an analysis of part of the vocabulary of Plato». Oxford: Blackwell.

[44] Maggiore, Marco. 2020. «Considérations sur la structure XML des articles du DERom». In *Dictionnaire Étymologique Roman (DÉRom) 3. Entre idioroman et protoroman*, a cura di Wolfgang Schweickard e Éva Buchi. Berlin-Boston: De Gruyter.

[45] Mambrini, Francesco, e Marco Passarotti. 2020. «Representing Etymology in the LiLa Knowledge Base of Linguistic Resources for Latin». In *Proceedings of the 2020 Globalex Workshop on Linked Lexicography*, a cura di Ilan Kernerman, Simon Krek, John P. McCrae, Jorge Gracia, Sina Ahmadi, e Besim Kabashi, 20–28. Marseille, France: European Language Resources Association. https://aclanthology.org/2020.globalex-1.3/.

[46] Marello, Carla. 1996. *Le parole dell'italiano: lessico e dizionari*. Bologna: Zanichelli.

[47] McCrae, John, Dennis Spohr, e Philipp Cimiano. 2011. «Linking Lexical Resources and Ontologies on the Semantic Web with Lemon». In *The Semantic Web: Research and Applications*, a cura di Grigoris Antoniou, Marko Grobelnik, Elena Simperl, Bijan Parsia, Dimitris Plexousakis, Pieter De Leenheer, e Jeff Pan, 245–59. Berlin, Heidelberg: Springer Berlin Heidelberg.

[48] Meyer-Lübke, Wilhelm. 1911. *Romanisches etymologisches Wörterbuch*. Heidelberg: Carl Winter.

[49] Nagao, Makoto, Jun'Ichi Tsujii, Y. Ueda, e M. Takiyama. 1982. «An Attempt to Computerize Dictionary Data Bases». In *Lexicography in the electronic age: proceedings of a symposium held in Luxembourg, 7-9 July, 1981*, a cura di J. Goetschalckx e L. Rolling, 51–73. Amsterdam/New York/Oxford: North-Holland Publisher Company.

[50] Nannini, Alessandro A. 2023. «La mappatura semantica del Lessico Etimologico Italiano (LEI). Possibilità, metodi e prospettive». In *Digitale romanistische Sprachwissenschaft: Stand und Perspektiven. Romanistisches Kolloquium, XXXIV*, a cura di Lidia Becker, Julia Kuhn, Christina Ossenkop, Claudia Polzin-Haumann, e Elton Prifti, 119–45. Tübingen: Narr.

[51] Opera del Vocabolario Italiano (OVI) - CNR, a c. di. 1997. *Tesoro della Lingua Italiana delle Origini*. Firenze. http://tlio.ovi.cnr.it/TLIO/.

[52] *Oxford English Dictionary (OED)*. Oxford University Press. www.oed.com.

[53] Pfister, Max. 1982. «Sémantique et lexicologie». In *Congrés Internacional de Lingüística i Filologia Romàniques (Palma de Mallorca, 7-12 d'abrile de 1980). Actes, vol. 1: Sessions Plenàries i Taules Rodones*, a cura di Aina Moll, 278–94. Palma de Mallorca: Editorial Moll.

[54] Pfister, Max. 1983. «Présentation du LEI (Lessico Etimologico Italiano): Possibilités d'Établir des Index Lexicaux et Morphologiques par Ordinateur». In *The possibilities and limits of the computer in producing and publishing dictionaries: proceedings of the European Science Foundation Workshop, Pisa, 1981*, a cura di Antonio Zampolli e Amadeo Cappelli, 187–91. Pisa: Giardini.

[55] Porck, Thijs. 2021. «Onomasiological Profiles of Old English Texts: Analysing the Vocabulary of Beowulf, Andreas and the Old English Martyrology through Linguistic Linked Data». *Amsterdamer Beiträge Zur Älteren Germanistik* 81 (3–4): 359–83. https://doi.org/10.1163/18756719-12340236.

[56] Prifti, Elton. 2022. «Il LEI Digitale. Un resoconto, con particolare attenzione Alla dialettologia». In *Lessicografia Storica Dialettale e Regionale*, edited by Michele Cortelazzo, Silvia Morgana, and Massimo Prada, 293–314. Firenze: Franco Cesati.

[57] Prifti, Elton, e Wolfgang Schweickard, a c. di. 1979. *LEI. Lessico Etimologico Italiano. Fondato da Max Pfister*. Wiesbaden: Reichert. https://lei-digitale.it.

[58] Prifti, Elton, Wolfgang Schweickard, Maria Selig, e Sabine Tittel. 2023. «Sprachdatenbasierte Modellierung von Wissensnetzen in der mittelalterlichen Romania (ALMA): Projektskizze». *Zeitschrift für romanische Philologie* 139: 301–332. https://doi.org/10.1515/zrp-2023-0012.

[59] Pruvost, J. 2000. «Des dictionnaires papier aux dictionnaires électroniques. VIIe journée des dictionnaires (22 mars 2000)». *International journal of lexicography* 13: 187–93. https://doi.org/10.1093/ijl/13.3.187.

[60] Quillian, M. Ross. 1968. «Semantic memory». In *Semantic information processing*, a cura di Marvin Minsky, 216–70. Cambridge, Mass.: MIT Press.

[61] Quillian, M. Ross 1969. «The teachable language comprehender: a simulation program and theory of language». *Commun. ACM* 12 (8): 459–76. https://doi.org/10.1145/363196.363214.

[62] Quochi, Valeria, Andrea Bellandi, Fahad Khan, Michele Mallia, Francesca Murano, Silvia Piccini, Luca Rigobianco, Alessandro Tommasi, e Cesare Zavattari. 2022. «From Inscriptions to Lexicon and Back: A Platform for Editing and Linking the Languages of Ancient Italy». In *LREC 2022 Workshop Language Resources and Evaluation Conference. Second Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA 2022). Proceedings*, 59–67. Marseille: European Language Resources Association. https://aclanthology.org/2022.lt4hala-1.9/.

[63] Real Academia Española, a c. d. 2013. *Diccionario histórico de la lengua española (DHLE)*. https://www.rae.es/dhle/.

[64] Saussure, Ferdinand de. 1983. *Corso di linguistica generale*. Tradotto da Tullio De Mauro. Biblioteca Universale Laterza. Roma/Bari: Laterza.

[65] Stolk, Sander. 2021. «Evoke: Exploring and Extending A Thesaurus of Old English Using a Linked Data Approach». *Amsterdamer Beiträge zur älteren Germanistik* 81 (3–4): 318–58. https://doi.org/10.1163/18756719-12340235.

[66] Tittel, Sabine. 2023. «Ceci n'est pas un dictionnaire. Adding and Extending Lexicographical Data of Medieval Romance Languages to and through a Multilingual Lexico-Ontological Project». In *Electronic lexicography in the 21st century (eLex 2023): Invisible Lexicography. Proceedings of the eLex 2023 conference. Brno, 27–29 June 2023*, a cura di Marek Medved', Michal Měchura, Carole Tiberius, Iztok Kosem, Jelena Kallas, e Miloš Jakubíček, 39–52. Brno: Lexical Computing CZ s.r.o.

[67] Tittel, Sabine. 2024. *Historische Lexikalische Semantik und Linked Data : Modellierung Von Ressourcen der Mittelalterlichen Galloromania Für das Semantic Web.* 1st ed. Beihefte Zur Zeitschrift Für Romanische Philologie Series. Berlin/Boston : Walter de Gruyter GmbH,.

[68] Tittel, Sabine, e Christian Chiarcos. 2018. «Historical Lexicography of Old French and Linked Open Data: Transforming the resources of the Dictionnaire étymologique de l'ancien français with OntoLex-Lemon». In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*, a cura di Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, et al., 58–66. Miyazaki, Japan: European Languages Resources Association (ELRA).

[69] Tittel, Sabine, Frances Gillis-Webber, e Alessandro A. Nannini. 2020. «Towards an Ontology Based on Hallig-Wartburg's Begriffssystem for Historical Linguistic Linked Data». In *Proceedings of the 7th Workshop on Linked Data in Linguistics (LDL-2020)*, 1–10. Marseille, France: European Language Resources Association. https://www.aclweb.org/anthology/2020.ldl-1.1.

[70] Trier, Jost. 1931. *Der deutsche Wortschatz im Sinnbezirk des Verstandes: die Geschichte eines sprachlichen Feldes. Von den Anfängen bis zum Beginn des 13. Jahrhunderts*. Heidelberg: Winter.

[71] Ullmann, Stephen. 1962. *Semantics: an introduction to the science of meaning*. Oxford: Blackwell.

[72] Van Baalen, Amos van. 2021. «Identifying, Categorising and Exploring 'Ælfrician' Vocabulary Using the Dictionary of Old English, A Thesaurus of Old English and Evoke». *Amsterdamer Beiträge Zur Älteren Germanistik* 81 (3–4): 384–441. https://doi.org/10.1163/18756719-12340237.

[73] von Humboldt, Wilhelm. 1836. *Über die Kawi-Sprache auf der Insel Java: nebst einer Einleitung über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwickelung des Menschengeschlechts. 1*. Abhandlungen der Königlichen Akademie der Wissenschaften in Berlin 1832,2. Berlin: Dümmler.

[74] von Wartburg, Walther, a c. di. 1921. *Französisches Etymologisches Wörterbuch*. Bonn/Lipsia/Basel: Klopp. https://lecteur-few.atilf.fr.

[75] von Wartburg, Walther. 1937. «Betrachtungen über die Gliederung des Wortschatzes und die Gestaltung des Wörterbuchs». *Zeitschrift für romanische Philologie (ZrP)* 57:312. https://doi.org/10.1515/zrph.1937.57.2-4.296.

[76] von Wartburg, Walther. ²1962. *Einführung in Problematik und Methodik der Sprachwissenschaft*. A cura di Stephan Ullmann. Seconda edizione. Tübingen: Max Niemeyer.

[77] Weinreich, Uriel. 1967. «Lexicographic definition in descriptive semantics». In *Problems in lexicography*, a cura di Fred D. Householder e Sol Saporta, 25–44. Bloomington: Indiana University Press.

[78] Wierzbicka, Anna. 1985. *Lexicography and conceptual analysis*. Ann Arbor, Mich.: Karoma Publ.

[79] Zampolli, Antonio. 1983. «Lexicological and Lexicographical Activities at the Istituto di Linguistica Computazionale». In *The possibilities and limits of the computer in producing and publishing dictionaries: proceedings of the European Science Foundation Workshop, Pisa, 1981*, a cura di Antonio Zampolli e Amadeo Cappelli, 237–78. Pisa: Giardini.

[80] Zampolli, Antonio, e Amadeo Cappelli, a c. di. 1983. *The possibilities and limits of the computer in producing and publishing dictionaries : proceedings of the European Science Foundation Workshop, Pisa, 1981 / Eds. A. Zampolli, A. Cappelli*. Pisa: Giardini.

[81] Zimmerman, Herald H. 1983. «Multifunctional Dictionary». In *The possibilities and limits of the computer in producing and publishing dictionaries: proceedings of the European Science Foundation Workshop, Pisa, 1981*, a cura di Antonio Zampolli e Amadeo Cappelli, 279–88. Pisa: Giardini.