Un'ontologia per il DiTMAO (Dictionnaire de Termes Médico-botaniques de l'Ancien Occitan)

Section 16 - Projets en cours; ressources et outils nouveaux

Damiana Luzzi

Il Progetto DiTMAO[1] ha fra gli obiettivi quello di organizzare e rendere accessibili sul Web, in formato digitale, i testi relativi all'antica terminologia medico-farmaceutica in occitano[2] per la redazione di un dizionario terminologico. Le voci entrano in un duplice sistema di classificazione logico-semantica: quello aderente al periodo medievale ove furono usate e quello coerente con l'uso attuale. Per raggiungere tali scopi vengono adoperate le tecnologie del Semantic Web (ontologie) e una suite di applicazioni software di ultima generazione dedicata al trattamento di testi, immagini e lessici (Reperio[3]). In questo contributo, si prende in esame l'approccio metodologico che ha portato alla scelta dell'ontologia, intesa in senso informatico[4] e la descrizione dell'ontologia risultante. La riflessione ha preso avvio dall'analisi delle informazioni in formato cartaceo:

- il glossario [1] dei termini medici in antico occitano,
- i termini medici in antico occitano scritti in alfabeto ebraico e che includono spiegazioni del loro significato in ebraico, arabo e latino [2],
- i manoscritti[5], ovvero le fonti primarie,
- eventuali edizioni critiche a stampa.

Materiali eterogeni, quindi, in lingue e alfabeti diversi. A questa complessità si aggiunge quella della creazione di un sistema di relazioni tra le due classificazioni temporalmente distanti: medioevo e età contemporanea.

L'ontologia [3] è ciò che riesce a rispondere a tali esigenze. Il primo passo per la definizione di classi, proprietà e regole dello schema ontologico è stato l'analisi dei materiali e la successiva verifica dell'esistenza di standard, riconosciuti a livello internazionale, idonei allo scopo. Sono stati individuati due standard internazionali:

- FRBR-oo [4]
- LMF [5].

I test di popolazione dello schema ontologico, eseguiti per la verifica dell'applicabilità di tali standard, hanno evidenziato come tali standard non fossero sufficienti a rappresentare il dominio di conoscenza relativo all'ambito medico-farmaceutico rappresentato nei testi del corpus. È stato necessario, quindi, estendere lo schema ontologico, pur mantenendo la compatibilità con gli standard di partenza. Un processo iterativo che, grazie a test successivi di popolamento dello schema ontologico[6], ha portato a modifiche e al raffinamento di quest'ultimo, fino a soddisfare pienamente i requisiti posti.

Dalla rappresentazione, anche solo sintetica in UML[7] (fig. 1) delle classi principali dello schema e dalle loro relazioni, si evince come sia possibile catalogare lemmi, sottolemmi, sinonimi e varianti (morfologiche, grafico fonetiche, ecc.), indicando l'alfabeto, la categoria grammaticale, il numero, il significato, la lingua, il nome scientifico, l'eventuale corrispettivo in un'altra lingua antica, il periodo nel quale ogni singola voce (lemma, sottolemma e/o variante) era in uso e in quali documenti risulti attestata[8]. Grazie alla relazione con la classe OldOccitanicMedicalTerninology e relative sottoclassi, è possibile attribuire ai lemmi un valore semantico (ad esempio: aloe è una pianta e appartiene alla famiglia delle aloeacee). La classe OldOccitanicMedicalTerminology racchiude classi e relazioni che descrivono la classificazione medievale dei termini medici e botanici a partire da tre sottoclassi

---

[1] Progetto di collaborazione fra la Freie Universität Berlin - Institut für Romanische Philologie, l'Universität zu Köln - Martin-Buber-Institut für Judaistik, l'Università degli Studi di Pisa - Dipartimento di Lingue e letterature romanze, l'Istituto di Linguistica Computazionale "Antonio Zampolli" CNR e la Fondazione Rinascimento Digitale.

[2] I testi editi a Berlino e a Colonia sono in ebraico, ma contengono molti termini medici in antico occitano scritti utilizzando l'alfabeto ebraico e che spesso includono spiegazioni del loro significato in ebraico, arabo e latino.

[3] Reperio è un ambiente di lavoro collaborativo, modulare, personalizzabile e utilizzabile on line, che fornisce gli strumenti a supporto dell'intero ciclo di vita della ricerca negli studi umanistici (editor di ontologie, gestione e editing di testi e immagini, immissione di annotazione e commenti, confronto di più testimoni, ecc.). Reperio è sviluppato dalla Fondazione Rinascimento Digitale (Firenze) in collaborazione con l'Istituto di Linguistica Computazionale "Antonio Zampolli" – CNR (Pisa), http://www.reperio.it/.

[4] Una "specificazione esplicita di una concettualizzazione", Gruber ,T. R.: Translation Approach to Portable Ontology Specifications. In Knowledge Acquisition. 5 (2), 199--220 (1993).

[5] Cfr. Corradini Bozzi, M. S.: Ricettari medico-farmaceutici medievali nella Francia meridionale. In Studi dell'Accademia di Scienze e Lettere "La Colombaria", 1. Leo Olschki Editore, Firenze (1997).

[6] Per la realizzazione dell'ontologia e la sua popolazione è stato utilizzato il modulo Ontology Editor disponibile in Reperio.

[7] UML (Unified Modeling Language). Nello schema i nomi delle classi e delle relazioni, per convenzione, sono scritti in inglese. Viene utilizzato il francese come lingua del progetto per dati, interfaccia di popolazione dell'ontologia e consultazione.

[8] Ad esempio: *aloe* è un lemma in antico occitano e alfabeto latino, è un nome di genere maschile, la traduzione francese è *aloès*, ha variante in alfabeto ebraico *'LW'N,* si trova in altre lingue come l'aramaico *'YLWW',* latino medievale *aloe*, ecc., il nome scientifico è *aloe vera*, ha sottolemma *lin d'aloe* (che a sua volta viene descritto), è attestata nel MS. Princeton Garrett (P ric.) e ha riferimenti bibliografici, ad esempio, in Wartburg, Von W.: Französischen Etymologisches Wörterbuch, p. 25, 345b. Bonn, Leipzig, Tübingen, Basel (1922-1987).

principali: AnimalWorld (mondo animale), VegetalWorld (mondo vegetale), MineralWorld (mondo minerale). Il legame tra la "superclasse" OldOccitanicMedicalTerminology e la "superclasse" ModernMedicalTerminology rende esplicita la capacità dello schema di esprimere la relazione diacronica fra i termini.

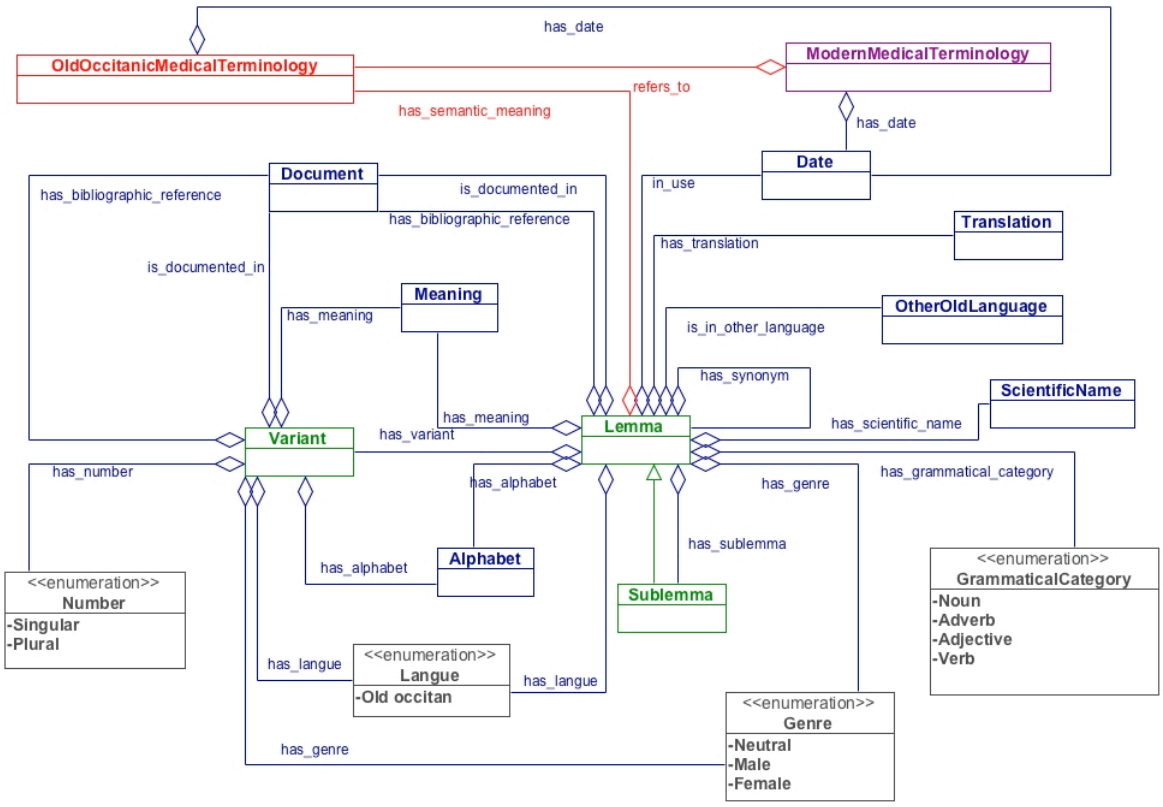

Fig.1. Schema UML

Le informazioni sui lemmi, inserite nello schema ontologico, sono utilizzate per la lemmatizzazione semi-automatica dei testi (trascrizioni e immagini digitali) immessi nel modulo Reperio Text. L'ontologia, collegata in maniera dinamica a Reperio Text, consentirà, inoltre, operazioni di annotazione, anche in maniera collaborativa, su testi e immagini.

Rispetto alla consultazione di concordanze lemmatizzate di impianto tradizionale, la classificazione utilizzata apporta notevoli benefici a livello di maggiore precisione, accuratezza, ricchezza di informazioni e recall nelle operazioni di Information Retrieval/Extraction necessarie agli studi linguistici e filologici relativi alla terminologia attestata dal corpus. Ad esempio, cercando il termine *humores* si otterranno informazioni non solo su quante volte tale termine è presente nel testo o nei testi oggetto della ricerca, ma anche il lemma di riferimento, gli eventuali sottolemmi e varianti, in quali manoscritti ne è attestato l'uso, le traduzioni in altre lingue e il riferimento all'uso attuale.

Riferimenti bibliografici

[1] Corradini Bozzi, M. S.: Glossario. In Corradini Bozzi, M. S.: Ricettari medico-farmaceutici medievali nella Francia meridionale. Studi dell'Accademia di Scienze e Lettere "La Colombaria", 1, 417-- 486. Leo Olschki Editore, Firenze (1997)

[2] Mensching, G.:Listes de synonymes hébraïques-occitanes du domaine médico-botanique au Moyen Âge. In Latry, G. (Eds.): La voix occitane. Actes du VIIIe Congrès Internationale d'Études Occitanes, 2 Bde, Bd. I, S. 509--526. Bordeaux: Presses universitaires (2009)

[3] Staab, S., Studer R. (Eds): Handbook on Ontologies (2ed.). Springer-Verlag, Berlin, Heidelberg (2009)

[4] International Federation of Library Associations (IFLA).: FRBR-oo (Functional Requirements for Bibliographic Records-object oriented). http://archive.ifla.org/VII/s13/wgfrbr/FRBR-CRMdialogue_wg.htm

[5] Lexical Markup Framework (LMF), ISO-24613:2008. http://www.lexicalmarkupframework.org/

[6] National Library of Medicine (NLM – National Institute of Health, United States).: Unified Medical Language Thesaurus (UMLS). http://www.nlm.nih.gov/research/umls/