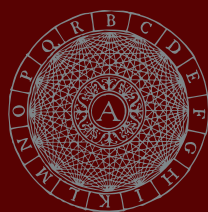


ILIESI digitale
Memorie

ADA RUSSO
MICHELA TARDELLA

THE *OF* PHILO
THESAURUS OF PHILOSOPHY



ILIESI
CNR

Istituto per il Lessico Intellettuale Europeo e Storia delle Idee

2017



ILIESI digitale Memorie

3

Secondo le norme dell'ILIESI tutti i contributi pubblicati nella collana sono sottoposti a un processo di *peer review* che ne attesta la validità scientifica

Assistente editoriale
Maria Cristina Dalfino
Progetto grafico
Silvestro Caligiuri

ISSN
2283-5571
DOI
10.19283/ILIESI-MM.03

ADA RUSSO

MICHELA TARDELLA

THEOPHILO – THESAURUS OF PHILOSOPHY

Abstract

This paper aims at presenting TheofPhilo – Thesaurus of Philosophy, a multilingual philosophical thesaurus developed in the framework of the European project *Agora. Scholarly Open Access Resource in European Philosophy* and the CNR/MIUR project *SM@RTINFRA-SSHCH. Smart Integrated infrastructure of Social Sciences, Humanities and Cultural Heritage*. In this paper, we will describe the methodological approaches and the workflow followed in order to build the thesaurus. We will thus show the work processes, and the criteria adopted to select the relevant philosophical terms; then we will present the lexicographical sources used to select both the interlinguistic equivalents and their intralinguistic related terms (antonyms, synonyms and so on). We conclude our paper with a description of the searching and browsing options (by topics, by semantic annotations, etc.) and of the tools we used. We will finally discuss the tests carried out and some critical aspects emerging from the work.

Keywords:

Philosophical terminology, Thesauri, Ontologies, Multilingualism, Information retrieval

Parole chiave:

Terminologia filosofica, Tesauri, Ontologie, Multilinguismo, Information retrieval

1. INTRODUZIONE*

TheofPhilo – *Thesaurus of Philosophy*, è un *thesaurus* filosofico multilingue, la cui realizzazione è stata avviata dall'ILIESI nel corso del progetto *Agora. Scholarly Open Access Resource in European Philosophy*.¹ Concepito come sistema di chiavi d'accesso multilingue agli archivi digitali, il lavoro è stato ulteriormente sviluppato come contributo al progetto *SM@RTINFRA-SSHCH. Smart Integrated infrastructure (for the data ecosystem) of Social Sciences, Humanities and Cultural Heritage*², collocandosi pienamente all'interno dei Work Packages 2 (*Access services to infrastructure*) e 4 (*Common research*).

Tra le attività sperimentali condotte nel corso del progetto *Agora*, si è elaborata un'ontologia per la rappresentazione della conoscenza contenuta nei testi filosofici pubblicati nel portale *Daphnet* (<http://www.daphnet.org/>),³ rispetto alla quale *TheofPhilo* costituisce il popolamento della classe dei soggetti filosofici.

Una definizione piuttosto chiara di ontologia è stata formulata dallo scienziato computazionale Thomas Gruber:

* Questo lavoro è stato concepito in tutte le sue parti in stretta collaborazione dalle due autrici. Tuttavia, le responsabilità delle singole sezioni vanno così attribuite: Ada Russo §§ 4-5-6, Michela Tardella §§ 1-2-3-7-8.

¹ Il progetto triennale *Agora. Scholarly Open Access Research in European Philosophy*, CIP-ICT PSP Grant agreement n. 270904, nasce nel 2011 con l'obiettivo principale di favorire la diffusione dei risultati delle ricerche europee in ambito umanistico e sviluppare nuovi paradigmi per la loro pubblicazione in Open Access (responsabile del progetto: Antonio Lamarra). Cfr. Lamarra-Marras 2013 e <<http://www.netseven.it/projet/agora>>, ultimo accesso novembre 2016.

² Il progetto premiale del CNR *SM@RTINFRA-SSHCH* nasce con l'obiettivo generale di promuovere e sviluppare un sistema coordinato di nodi nazionali di importanti infrastrutture di ricerca europee operanti nel settore Social Sciences, Humanities And Cultural Heritage. La struttura *SM@RTINFRA-SSHCH* permetterà di razionalizzare le risorse in modo efficace ed efficiente, creare ponti che consentano di superare il divario tra esperti di discipline diverse, risorse strumentali e digitali e metodi di analisi dei dati, da un lato, e gli utenti, dall'altro, favorire l'attrazione di finanziamenti esterni a valere in particolare su programmi di ricerca europei, rafforzare il ruolo dell'Italia come leader europeo in questo settore (responsabile del progetto: Riccardo Pozzo).

³ *Daphnet. Digital Archives of Philosophical Texts on the NET* è un portale dell'ILIESI che dà accesso a un insieme di piattaforme digitali dedicate ad autori e testi di particolare rilievo per la storia del pensiero filosofico e scientifico. Nato all'interno del progetto triennale *eContentplus Discovery 2006-2009* (cfr. D'Iorio 2009 e <<http://www.netseven.it/projet/discovery>>, ultimo accesso novembre 2016), il portale, sotto la direzione scientifica di Antonio Lamarra, è stato ulteriormente arricchito durante il già citato progetto *Agora*.

an explicit specification of a conceptualization. The term is borrowed from philosophy, where an ontology is a systematic account of Existence. For knowledge-based systems, what “exists” is exactly that which can be represented. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge.⁴

Come emerge da questa concisa ma efficace definizione, un’ontologia può essere considerata un modello di rappresentazione della conoscenza, particolarmente raffinato, che implica da parte dell’umanista un’attività interpretativa di selezione e organizzazione delle entità appartenenti all’universo del discorso (dominio) che intende analizzare, avvalendosi degli strumenti e delle metodologie digitali. Affinché tale specificazione concettuale potesse essere modellizzata, nel corso dei progetti *Discovery* e *Agora* si è lavorato all’individuazione delle entità, del sistema di classi di appartenenza e delle relazioni fondamentali attraverso cui strutturare i rapporti tra loro intercorrenti.

Dal punto di vista dell’implementazione si è scelto di utilizzare il software *Pundit*.⁵ Ideato nel sempre più vasto contesto delle *Semantic Web technologies* per la classificazione, la gestione e l’arricchimento semantico dei testi, il software permette di produrre, secondo i principi del modello RDF (*Resource Description Framework*), annotazioni semantiche altamente granulari e *machine-processable*. Come nota Pierre Andrews:

Ontology-based annotations or “semantic annotations” describe a resource with respect to a formal conceptual model, allowing meaning-bearing links between structured and unstructured data (such as an ontology and a text). This empowers a

⁴ Cfr. Gruber 1993, p. 199.

⁵ Pundit è uno strumento di annotazione semantica *open source* che permette agli utenti di creare dati strutturati per l’annotazione di pagine Web. Pundit non solo consente agli utenti di annotare le pagine Web in vari modi (commenti, citazioni, segnalibri), ma anche di convertire queste annotazioni in dati semanticamente strutturati che possono essere successivamente integrati al “Web semantico”. Pundit è stato progettato e sviluppato dalla società Net7 a Pisa; cfr. il sito <<http://thepund.it>>, ultimo accesso novembre 2016. Per una presentazione del software *Pundit* si veda Grassi *et al.* 2012; un quadro piuttosto ampio relativo agli studi sulle annotazioni semantiche è tracciato in Andrews-Zaihrayeu-Pane 2012; Signore 2002; Uren *et al.* 2006; per ulteriori informazioni sul modello RDF si rimanda alle fonti online reperibili al sito <<http://www.w3.org/RDF/>>, ultimo accesso novembre 2016. Il gruppo di ricerca dell’ILIESI ha svolto un training d’uso del software prendendo parte a varie attività seminariali e al workshop “Open Platforms for digital humanities. State of the art and perspectives”, tenutosi a Cortona il 17 e 18 gennaio 2013.

whole new range of retrieval techniques, which can be based on the knowledge schema expressed in the ontology, benefit from reasoning, co-occurrence of annotation or entities in the same resource or context, as well as combine this with unstructured data specific types of retrieval, such as full text search (FTS) in information retrieval (IR).⁶

L'ontologia filosofica, sulla scorta del lavoro fatto durante il precedente progetto *Discovery*, si componeva delle seguenti classi, ciascuna variamente articolata al proprio interno in *sub-classes* e debitamente popolata:

- *Person (Philosophers e Scholars)*;
- *Relevant Concept (Philosophical Subject e Relevant Event)*;
- *Relevant Subject (Geographical Entity, Philosophical Theme, Philosophical School, Quotation e Title)*;
- *Source (Secondary Source e Primary Source)*.

Le relazioni di interconnessione tra le entità che popolano le classi appena elencate, sono invece state strutturate come segue:

- *Peoples' Properties (Authors, Editors, Scholars)*;⁷
- *Properties of the Documents*;⁸
- *Properties of the Texts (Relations text-to-text,⁹ Relations text-to-subject)*.

Le più rilevanti ai fini dello sviluppo e dell'uso del *thesaurus*, sono chiaramente le relazioni *text-to-subject*, poiché riguardano la connessione tra testi e soggetti filosoficamente rilevanti. Si è per questo lavorato alla definizione di predicati RDF che permettessero di fare soggettazioni fondate non soltanto sulle occorrenze dei termini, ma anche sull'interpretazione dei testi rispetto ai concetti filosofici. Tali relazioni sono:

⁶ Andrews-Zaihrayeu-Pane 2012, pp. 7-8.

⁷ *hasBirthDay/isBirthDayOf; hasDeathDay/isDeathDayOf; hasFirstName/isFirstNameOf; hasPatronymic/isPatronymicOf; hasPlaceOfBirth/isPlaceOfBirthOf; hasPlaceOfDeath/isPlaceOfDeath; hasSurname/isSurnameOf; isAuthorOf/has Author; isEditorOf/hasEditor; isTranslatorOf/hasTranslator.*

⁸ *hasBeenPublishedIn/isPublicationDateOf; hasbeenPublishedAt/isPublicationPlaceOf; hasFirstEditionDate/isFirstEditionDateOf; hasTitle/isTitleOf; isEditedIn/containsEditionOf; isFacSimile/hasFacSimile; isPartOf/hasPart; isPublishedBy/isPublisherOf; isTextTranscription/hasTextTranscription; isTranslationOf/hasTranslation; isWrittenIn/isTheLanguageOf.*

⁹ *QuotesDirectly/IsDirectlyQuotedBy; QuotesIndirectly/IsIndirectlyQuotedBy; Refutes/IsRefutedBy; ArguesFor/IsArguedForBy; Explains/IsExplainedBy; Criticizes/IsCriticizedBy; AgreesWith/IsAgreedWithBy; Interprets/IsInterpretedBy; IsSimilarTo/IsSimilarTo; MakesARefrenceTo/IsReferencedBy; MakesAnInternalReferenceTo/IsInternallyReferencedBy.*

1. Definition – *Defines/IsDefinedBy*;
2. Indirect Definition – *IndirectlyDefines/IsIndirectlyDefinedBy*;
3. Extensional Instantiation – *IsAnExtensionalInstanceOf/Is ExtensionallyInstantiatedBy*;
4. Intensional Instantiation – *IsAnIntensionalInstanceOf/Is IntensionallyInstantiatedBy*;
5. Dealings – *DealsWith/IsDealtWithBy*.

L'attività scientifica di annotazione semantica *text-to-subject* ha implicato un'ampia riflessione sulla terminologia filosofica e sulla relativa lessicalizzazione nelle lingue di cultura presenti nel portale *Daphnet*, cioè italiano, latino, greco, francese e inglese. Costruito su questi presupposti teorico-progettuali, *TheofPhilo* non contiene definizioni dei termini, in quanto il suo specifico fine è soggettare le opere e garantire un efficace *Information Retrieval* (IR) e non costituire un dizionario filosofico.¹⁰

2. METODOLOGIA: PRIMA FASE

Dal punto di vista metodologico, nella pianificazione del lavoro si è proceduto seguendo un approccio di tipo semasiologico, muovendo cioè dall'analisi e individuazione delle forme significanti per sviluppare successivamente un'indagine semantica.

La terminologia rilevante è stata estratta manualmente, facendo riferimento agli indici di due opere scelte per la sperimentazione dell'attività di annotazione all'interno del portale *Daphnet*, segnatamente: i *Lineamenti Pirroniani* di Sesto Empirico e i *Nouveaux Essais* di Gottfried Wilhelm Leibniz.¹¹ Al termine di questa prima operazione, i termini ammontavano a 210 in lingua greca e 689 in lingua francese, implementati nel software Pundit come liste separate e utilizzabili nella costruzione delle triple RDF.

¹⁰ Cfr. Lamarra-Tardella 2014.

¹¹ Il lavoro di selezione dei termini è stato condotto da Francesco Verde, sotto la supervisione di Emidio Spinelli, per quanto concerne l'opera sestana (ed. di riferimento: H. Mutschmann-J. Mau, *Sexti Empirici Opera recensuit* H. M., coll. BT, Leipzig 1912-1954, 4 voll.), da Antonio Lamarra e Andrea Costa per i *Nuovi Saggi* di G. W. Leibniz (ed. di riferimento: G. W. Leibniz, *Nouveaux Essais*, [NE], A. Robinet-H. Schepers (eds.), *Sämtliche Schriften und Briefe*, ser. 6: *Philosophische Schriften*, vol. 6, 1962).

Parallelamente si è proceduto, utilizzando il programma *Excel*, all'allineamento delle due liste di termini¹² secondo la relazione di equivalenza interlinguistica e, in un secondo momento, all'individuazione dei traduenti latini, italiani e inglesi. Per la selezione di questi ultimi si è fatto ricorso alle seguenti fonti lessicografiche, perché redatte secondo un sistema di entrate multilingue:

1. N. Abbagnano, *Dizionario di Filosofia*, Torino, 1998;
2. A. Lalande, *Vocabulaire technique et critique de la philosophie*, Paris, 1983;
3. S. Maso, *Lingua Philosophica Graeca*, Milano-Udine, 2010.

A questi volumi sono inoltre stati aggiunti i seguenti testi:

1. A. Bailly, *Dictionnaire Grec-Français*, Édition revue par L. Séchan et P. Chantraine, Paris, 1950;
2. H. G. Liddell & R. Scott, *Greek-English Lexicon*, Rev. by H. S. Jones, Oxford, 1968;
3. T. Sanesi, *Vocabolario Italiano-Greco*, Pistoia-Siena, 1916;
4. L. Rocci, *Vocabolario Greco-Italiano*, Perugia, 1993.

Al termine di questa operazione di individuazione dei traduenti, il totale dei termini nelle cinque lingue ammontava a circa 5.000.

3. METODOLOGIA: SECONDA FASE

In una seconda fase, il gruppo di lavoro ha optato per una riduzione della quantità dei termini, finalizzata alla sperimentazione di alcune specifiche funzionalità di ricerca associabili al tesoro e ad un suo eventuale rilascio, in modalità *Open Access*, per la ricerca non più limitata alla piattaforma *Daphnet*, ma estendibile all'insieme degli archivi testuali dell'Istituto.¹³

Si è dunque deciso di selezionare un numero ridotto di termini latini secondo la seguente procedura:

- 1) estrazione, dai circa 10.000 termini registrati nella *Banca dati dei testi filosofici dell'età moderna*,¹⁴ delle entrate più frequenti [indice di frequenza compreso tra le 1.598 e le 100 occorrenze];

¹² A questa fase del lavoro ha partecipato anche Francesco Verde.

¹³ Cfr. Liburdi-Russo-Tardella 2015.

¹⁴ *La Banca dati dei testi filosofici dell'età moderna* (<<http://www.iliesi.cnr.it/bancadati>>, ultimo accesso novembre 2016), considerata la banca dati istituzionale dell'ILIESI, nasce agli inizi degli anni Ottanta come archivio del materiale documentario per la

2) organizzazione delle entrate estratte in famiglie lessicali¹⁵ secondo i seguenti criteri:

- a) inclusione dei termini che, pur avendo un numero di occorrenze inferiore a 1, appartengono alla stessa famiglia lessicale dei termini con numero di occorrenze superiore a 100;
- b) esclusione dei termini derivati per prefissazione (per es. “coexistentia”>“existentia”), al fine di evitare un aumento eccessivo dei termini; si fa eccezione per gli antonimi su base morfologica (es. “inexistentia”>“existentia”), che permettono l’arricchimento della famiglia e la strutturazione dei lemmi secondo le relazioni semantiche interlessemiche.

Per la ricostruzione delle famiglie lessicali latine, si è fatto ricorso al *Lexicon totius latinitatis* (Padova 1940) di Egidio Forcellini.

3) Eliminazione di quelle famiglie che non includessero almeno un termine già presente nella versione originaria di TheofPhilo.

4) Controllo degli equivalenti interlinguistici (attraverso il lavoro fatto nell’esperimento precedente) e indicazione precisa della fonte dalla quale sono stati estratti (in particolare i dizionari filosofici multilingue di Nicola Abbagnano, André Lalande e Stefano Maso).

5) Strutturazione in famiglie lessicali degli equivalenti interlinguistici verificati.

Per l’italiano è stato utilizzato il *Vocabolario Treccani*; per il francese le risorse digitali rese disponibili dal *Centre National de Ressources Textuelles et Lexicales*;¹⁶ per l’inglese l’*Oxford Dictionary*;¹⁷ per il greco ancora una volta il *Greek-English Lexicon* (Oxford, 1968) di H. G. Liddell & R. Scott.

redazione di un *Lessico filosofico dei secoli XVII e XVIII*; l’archivio ha acquistato nel tempo un’autonomia sempre crescente, che ne ha fatto uno strumento finalizzato a svolgere una vasta gamma di ricerche lessicografiche e terminologiche sulla letteratura filosofica e scientifica dal Cinquecento al Settecento. Cfr. Russo 1985, pp. 321-333. Sui criteri di strutturazione della terminologia presente nella banca dati si veda Spinosa 1986, pp. 17-30.

¹⁵ Assumiamo in questa sede la definizione di “famiglia lessicale” indicata nel *Dizionario di linguistica* diretto da Beccaria, ovvero: “Insieme di parole che hanno in comune la stessa base o radice” (*ad vocem*). Rispetto alla definizione delle famiglie si è ovviamente tenuto conto “del fenomeno del *suppletivismo* o polimorfia lessicale, per cui una parte di derivati ha una base e un’altra parte ha un’altra base [...] o una forma leggermente diversa” (*ibid.*). I termini *deus* e *divinus* costituiscono, nel nostro corpus, un esempio del fenomeno appena descritto.

¹⁶ Cfr. <<http://www.cnrtl.fr/etymologie/>>, ultimo accesso novembre 2016.

¹⁷ Cfr. <<http://www.oxfordlearnersdictionaries.com>>, ultimo accesso novembre 2016.

6) Inserimento di alcuni termini composti¹⁸ o di entrate polirematiche¹⁹ in lingua latina, scelti tra quelli selezionati durante la lemmatizzazione condotta in Istituto negli anni passati. Ciascuno di questi termini si forma dalla composizione di due termini semplici inclusi nella lista originale.²⁰

Allo stato attuale dei lavori, la terminologia risulta ripartita secondo le quantità riportate nella tabella 1

	Theof Philo	Test	Semplici	Composte	Con Famiglia Lessicale	Senza Famiglia Lessicale	Famiglie Lessicali
Latino	1.211	251	139	112	95	156	28
Italiano	946	127	125	2	71	56	27
Francese	959	123	120	3	62	61	25
Inglese	1.167	170	163	7	51	119	20
Greco	980	137	122	15	0	137	0
Totale	5.263	808	669	139	279	529	100

Tab.1: Quantità e ripartizione dei termini allo stato attuale.

4. L'INTERFACCIA DI THEOPHILO

L'interfaccia di TheofPhilo, attraverso la quale è possibile navigare nel database terminologico e accedere ai testi degli archivi digitali,

¹⁸ Un composto è una parola complessa costituita dall'unione di due parole (forme libere) in una unità di significato, come ad esempio, all'interno del nostro *thesaurus*, il termine *natura dei*. Si tratta generalmente di atomi sintattici, cioè di complessi che non possono essere modificati al loro interno e che non ammettono riferimento pronominale al costituente non testa (intendendo con "testa" il costituente iperonimico che determina la categoria semantica, la categoria sintattica, il genere e la classe flessiva del composto stesso, nel nostro caso il formante *natura*). Questi criteri aiutano a distinguere i composti dai sintagmi, sebbene tale distinzione sia spesso piuttosto problematica. I composti si distinguono inoltre dalle polirematiche perché il significato di queste non è compositivo e, inoltre, il complesso polirematico non è iponimo rispetto alla testa (vedi n. 7). Nel caso dei composti, invece, il complesso è iponimo della testa (*natura dei* è infatti iponimo di *natura*).

¹⁹ Per riferirsi a questa tipologia di combinazioni di parole, in ambito linguistico vengono utilizzati prevalentemente tre termini tecnici: in lessicografia troviamo (*unità*) *polirematica* (termine che utilizziamo nel presente lavoro); in morfologia abbiamo *composto sintagmatico* e, da ultimo, *unità lessicale superiore* o *lessema complesso*. Se con la seconda denominazione si intende sottolineare, delle polirematiche, le analogie con i composti, la terza ne mette piuttosto in evidenza l'unitarietà lessico-semantica. L'individuazione e l'analisi di queste formazioni sono operazioni complesse, perché le formazioni stesse si collocano in un'area del lessico difficilmente circoscrivibile perché posizionata tra i composti e i sintagmi.

²⁰ Ad esempio, l'entrata *affectus animi* ha come formanti i due monorematici *affectus* e *animus*, inclusi tra le entrate inizialmente selezionate.

presenta una schermata dedicata alla scheda informativa delle entrate, illustrata nella figura 1. Nella sezione superiore sono indicate le informazioni relative al singolo termine e alle relazioni che esso intrattiene con gli altri termini appartenenti alla stessa lingua.

The screenshot displays the 'TheofPhilo' Thesaurus of Philosophy interface. The header features the site name and logo, along with the ILIESI CNR logo. A navigation menu on the left includes 'Lista entrate', 'Cerca', 'Inserisci entrata', and 'Home', with a UK flag icon below. The main content area shows a search bar with 'Cerca:' and 'Lingua:' fields, and 'Invia' and 'RESET' buttons. The entry for 'corporeus' is displayed, including its family (F46) and related terms like 'corpus', 'corporeus', and 'corporeae'. A section for 'Traduzioni' lists translations in French, Greek, English, and Italian. At the bottom, there are buttons for 'Modifica entrata' and 'Cancella entrata'.

Fig. 1: L'interfaccia di TheofPhilo (scheda del termine latino *corporeus*)

Vengono dunque visualizzati:

- a. l'entrata;
- b. la famiglia lessicale di appartenenza (relazione individuata per tutte le lingue);

- c. i termini composti o le entrate polirematiche di cui il termine è un formante;
- d. gli eventuali antonimi;
- e. i pulsanti di ricerca nei testi.

Nella sezione inferiore della scheda sono invece riportate le relazioni interlinguistiche, cioè le traduzioni registrate, per ciascuna entrata, incrociando le soluzioni adottate dai dizionari filosofici multilingue e dai dizionari di lingua sopra elencati.

5. FUNZIONALITÀ E RISULTATI DELLA RICERCA

Come è già stato detto, uno degli scopi di questo lavoro era testare l'efficacia e la funzionalità di TheofPhilo come sistema di chiavi di accesso ai testi filosofici. Per l'esperimento sono stati considerati i testi presenti in due archivi dell'ILIESI: la *Banca dati dei testi filosofici dell'età moderna*, che è anche l'archivio che ha contribuito con le sue parole chiave alla formazione di TheofPhilo, e gli archivi *Ancient* e *Modern Source* della piattaforma *Daphnet*, in particolare per i testi che sono stati annotati semanticamente.

La ricerca sui testi può essere lanciata dalla scheda descritta nel paragrafo precedente; per ogni entrata sono disponibili diverse opzioni di interrogazione.

5.1 Ricerca nei testi con parole chiave

La ricerca per parole chiave si avvale dei risultati dell'attività critica di soggettazione e lemmatizzazione condotta dai ricercatori dell'ILIESI fin dai primi tempi della sua costituzione, e che ha riguardato soprattutto i testi in lingua latina. L'intento di tale attività era quello di segnalare la presenza di termini particolarmente significativi in un contesto preso in esame,²¹ indicando per tale contesto il lemma

²¹ La zona *lemmi* dei documenti presenti nella *Banca dati dei testi filosofici dell'età moderna* contiene i descrittori, ovvero i termini scelti come parole chiave, lemmatizzati e eventualmente organizzati secondo sottolemmi o sintagmi. "Secondo le norme del LIE, già in fase di selezione e schedatura il contesto è descritto attraverso i lemmi che costituiscono le parole chiave. [...] il LIE si è sempre impegnato a elaborare e offrire materiale lemmatizzato, e quest'operazione appare necessaria ove si voglia oltrepassare il limite puramente materiale nell'analisi e nello

normalizzato riferito al termine rilevante che si intendeva segnalare. Attraverso la ricerca di questi lemmi, che sono stati registrati tra le entrate di TheofPhilo, si ritrovano i documenti che rappresentano contesti d'uso selezionati e pregnanti.

5.2 Ricerca nei testi per forme

Con la ricerca per forme i documenti vengono individuati attraverso le occorrenze nei testi di tutte le forme associate all'entrata di TheofPhilo.²² La ricerca può produrre un output con un alto livello di "rumore", cioè un numero di documenti poco significativi rispetto alla ricerca effettuata partendo dall'entrata. I risultati prodotti possono essere poco significativi sia perché i contesti ottenuti non sempre si riferiscono a passi rilevanti rispetto all'attestazione del termine filosofico; sia perché la procedura di ricerca delle forme, non potendo disambiguare i termini omografi, può produrre risultati non sempre esattamente congruenti con la ricerca effettuata. Al contrario, la ricerca potrebbe in alcuni casi non produrre tutti i documenti significativi. Avrebbe quindi un basso indice di *recall*,²³

studio statistico. Le norme di lemmatizzazione sono molte e in qualche modo sempre *in fieri*, ma si indicano le più importanti perché significative per la nomenclatura delle entrate del *Lessico filosofico*: i sostantivi sono riportati sotto la forma del lemma al nominativo; i verbi alla prima persona dell'indicativo presente; gli aggettivi al nominativo maschile singolare; i comparativi e i superlativi sono stati ricondotti al positivo, salvo i casi di aggettivi fortemente irregolari. I casi più difficili da classificare sono quelli, noti, dei participi passati e dei participi presenti: a seconda dei contesti, possono avere valore verbale, aggettivale o addirittura sostantivale. Seguendo regole precise, si tende a classificare i participi con valore verbale sotto la voce la cui entrata è costituita dal verbo all'indicativo; e a classificare la forma con valore aggettivale sotto la voce la cui entrata è costituita dalla forma del participio presente o passato", *Lessico filosofico dei secoli XVII e XVIII*, Sezione latina, p. XIII. Da notare che mentre i verbi presenti nelle entrate di TheofPhilo sono riportati all'infinito, i lemmi chiave presenti nei documenti della *Banca dati dei testi filosofici dell'età moderna* dell'ILIESI sono, come si è visto, riportati alla prima persona del presente indicativo.

²² Per indicizzare i documenti e per effettuare le *query*, sia in latino che nelle altre lingue, non sono stati usati algoritmi di stemming (ad esempio un'implementazione dell'algoritmo di Schinke) o lemmatizzatori (pensiamo a LemLat, l'analizzatore morfologico automatico della lingua latina sviluppato presso l'Istituto di Linguistica Computazionale del CNR a Pisa <<http://www.ilc.cnr.it/>>, ultimo accesso novembre 2016) ma, definita una stringa radice, si è fatto uso di caratteri jolly per mascherare le terminazioni che differenziano le diverse forme, declinate nelle singole lingue.

²³ Ricordiamo qui la definizione di due concetti relativi all'Information Retrieval (IR): *precision*: percentuale di documenti restituiti (in risposta ad una *query*) che sono rilevanti; *recall*: percentuale di documenti rilevanti che sono restituiti.

perché, ad esempio, una determinata entrata presenta all'interno dei testi specifiche varianti grafiche che non sono note.²⁴

Per entrambe le modalità di ricerca appena descritte, una volta lanciata la richiesta è possibile, in modo non esclusivo, a) visualizzare un diagramma che raffigura le occorrenze del termine nei testi e negli autori della *Banca dati* dell'ILIESI (fig. 2);



Fig. 2: Output grafico (termine latino *idea*)

²⁴ La possibile perdita di specificità dei risultati ottenuti a seguito di tale “esplosione” di forme per la query, dipende dal possibile reperimento di omografi, cioè parole che hanno significato diverso ma grafia identica: ad esempio, nel latino le forme *animis* (dativo/ablativo plurale di *anima* e *animus*), *malum* (nel significato di “mela”, di “male” o “malvagio”), *actus* come sostantivo e come participio passato da *ago*, ecc.; inoltre può darsi il caso di varianti non previste di una stessa forma, come *imaginatio* per *imaginatio*, *adfectus* per *affectus*, *accommodatum* per *accomodatum*, etc.

b) accedere direttamente ai testi secondo la strutturazione autore/opera (fig. 3);



The screenshot shows the 'Thesaurus of Philosophy' interface. At the top, the title 'Thesaurus of Philosophy' is displayed in a serif font, with the 'ILIESI CNR' logo to the right. Below the header, the search results are presented. The main heading is 'Risultati della ricerca «idea»' followed by a bar chart icon and a document icon. Underneath, it states 'Banca dati dei testi filosofici dell'età moderna — Documenti trovati: 1636' with a document icon. The results are listed as a series of numbered folders, each containing a philosopher's name and a count of documents. The list includes: 1 — BACON 8, 2 — BAUMGARTEN 26, 3 — BRUNO 9, 4 — CAMPANELLA 15, 5 — CLAUBERG 59, 6 — DESCARTES 209, 7 — GALILEI 2, a sub-section 'Ju 2' with '22' and '110', 8 — GASSENDI 1, 9 — GENOVESI 199, 10 — GEULINCX 23, 11 — GRAVESANDE 107, 12 — HERBERT 2, 13 — HOBBS 29, 14 — KANT 23, 15 — KEPLER 20, 16 — KOMENSKY 8, 17 — LEIBNIZ 21, 18 — MORE 42, 19 — NEWTON 1, 20 — PUFENDORF 1, 21 — SPINOZA 205, 22 — TSCHIRNHAUS 4, 23 — VICO 47, and 24 — WOLFF 575.

Thesaurus of Philosophy **ILIESI**
CNR

Risultati della ricerca «idea»  

Banca dati dei testi filosofici dell'età moderna — Documenti trovati: 1636 — 

-  1 — BACON 8
-  2 — BAUMGARTEN 26
-  3 — BRUNO 9
-  4 — CAMPANELLA 15
-  5 — CLAUBERG 59
-  6 — DESCARTES 209
-  7 — GALILEI 2
-  Ju 2
 -  22
 -  110
-  8 — GASSENDI 1
-  9 — GENOVESI 199
-  10 — GEULINCX 23
-  11 — GRAVESANDE 107
-  12 — HERBERT 2
-  13 — HOBBS 29
-  14 — KANT 23
-  15 — KEPLER 20
-  16 — KOMENSKY 8
-  17 — LEIBNIZ 21
-  18 — MORE 42
-  19 — NEWTON 1
-  20 — PUFENDORF 1
-  21 — SPINOZA 205
-  22 — TSCHIRNHAUS 4
-  23 — VICO 47
-  24 — WOLFF 575

Fig. 3: Output lista (termine latino *idea*)

c) visualizzare i risultati ottenuti tramite concordanze di tre righe dei termini cercati (fig. 4).

(Bacon, NO, 160). Vai alla prima occorrenza.

Non leve quiddam interest inter humanae mentis idola et divinae mentis **ideas**; hoc est, inter placita quaedam inania et veras signaturas atque impressiones factas in creaturis, prout

(Bacon, NO, 205-206). Vai alla prima occorrenza.

nisi tantummodo a Platone, qui ad excutiendas definitiones et **ideas**, hac certe forma inductionis aliquatenus utitur. Verum [./] ad hujus inductionis, sive demonstrationis, instructionem bonam

(Bacon, NO, 217-218). Vai alla prima occorrenza.

(id quod superius diximus) quantum intersit inter humanae mentis idola, et divinae mentis **ideas**. Illa enim nihil aliud sunt quam abstractiones ad placitum: hae autem sunt vera signacula

Fig. 4: Output concordanze (termine latino *idea*)

5.3 Ricerca nei testi con annotazioni semantiche

Si è testata, con esito positivo, la possibilità di accedere ai documenti annotati semanticamente con il software Pundit. Sono stati annotati alcuni termini che occorrono negli archivi *Ancient Source* e *Modern Source* del portale *Daphnet*.²⁵ L'output prodotto da tale ricerca indica l'identificativo dei notebook contenenti le annotazioni semantiche di ciascun ricercatore²⁶ e delle annotazioni riferite al termine cercato, l'autore e la sigla dell'opera, il tipo di relazione associato al termine e un link al contesto annotato (fig. 5).

²⁵ Claudio Buccolini e Cristina Marras sono stati coinvolti nella sperimentazione in quanto studiosi, rispettivamente, di R. Descartes e G. W. Leibniz. La sperimentazione stessa è stata inoltre resa possibile grazie alla preziosa collaborazione di Romeo Zitarosa, informatico della società Net7 di Pisa.

²⁶ Le annotazioni vengono raccolte in un notebook personale pubblico. Il sito che li raccoglie tutti è al link <<http://ask.thepund.it/>>, ultimo accesso novembre 2016.



The screenshot shows the 'Thesaurus of Philosophy' interface. At the top, the title 'Thesaurus of Philosophy' is displayed in a serif font, and the logo for 'ILIESI CNR' is in the top right corner. Below the header, the main content area has a light yellow background and is titled 'Risultati della ricerca «idée»'. Underneath, it states 'Piattaforma Daphnet - Annotazioni semantiche — Documenti trovati: 5'. There are five entries, each consisting of a notebook ID and annotation ID, followed by a 'Vedi:' link and a description of the relationship type (e.g., 'DealsWith' or 'Defines').

Risultati della ricerca «idée»

Piattaforma Daphnet - Annotazioni semantiche — Documenti trovati: 5

Notebook: 8868775a, annotazione da22f23a.
[Vedi:](#) Descartes, DesMed_fr,11[1]. Tipo relazione = DealsWith

Notebook: 8868775a, annotazione e16ab32d.
[Vedi:](#) Descartes, DesMed_fr,29[2]. Tipo relazione = Defines

Notebook: 8868775a, annotazione ce71ba22.
[Vedi:](#) Descartes, DesMed_fr,29[3]. Tipo relazione = Defines

Notebook: 8868775a, annotazione c45a2072.
[Vedi:](#) Descartes, DesMed_fr,29[3]. Tipo relazione = DealsWith

Notebook: 8868775a, annotazione f48b9390.
[Vedi:](#) Descartes, DesMed_fr,45[2]. Tipo relazione = DealsWith

Fig. 5: Output annotazioni (termine francese *idée*)

Una volta giunti al contesto è possibile, come in tutti i documenti presenti negli archivi *Ancient* e *Modern Source*, visualizzare, attraverso una funzionalità dell'archivio legata al tool Pundit, tutte le annotazioni presenti nel documento.

Se si procedesse ad una attività di annotazione estesa a tutti i testi e nelle varie lingue, questa opzione sarebbe certamente una delle più efficaci per l'IR, perché implica un'attività interpretativa del testo, a fronte, ovviamente, di un consistente impegno riguardante il lavoro di annotazione.

5.4 Ricerca nei testi con parole topic modeling

Molto spesso una *query* lanciata su termini molto frequenti produce come risultato un numero elevato di documenti, difficilmente analizzabile. Nella ricerca di soluzioni che permettessero di ottenere una sintesi dei risultati, è stato effettuato un secondo test, utilizzando un'applicazione comunemente usata per il *data mining*, Mallet –

MAchine Learning for LanguagE Toolkit.²⁷ Questa opzione permette di ricercare l'entrata di TheofPhilo all'interno di aggregazioni (*topics*) di altri termini che risultano statisticamente correlati.

I *topics* corrispondono ad argomenti probabilisticamente significativi presenti nei documenti. L'obiettivo di un simile approccio è da una parte raffinare la ricerca (che da semplicemente terminologica potrebbe essere semanticamente orientata); dall'altra dare indicazioni per la soggettazione dei testi nel caso di inserimento di nuove annotazioni semantiche nei documenti. Il risultato che offre tale ricerca (fig. 6) consiste in: a) una lista di *topics*, in ordine decrescente di significatività, in cui per ogni *topic* vengono indicate le venti parole più frequenti, che formano delle nuvole in cui ogni singolo termine è cliccabile e dà luogo ad una ricerca nei testi; b) i documenti in cui è maggiormente significativa la

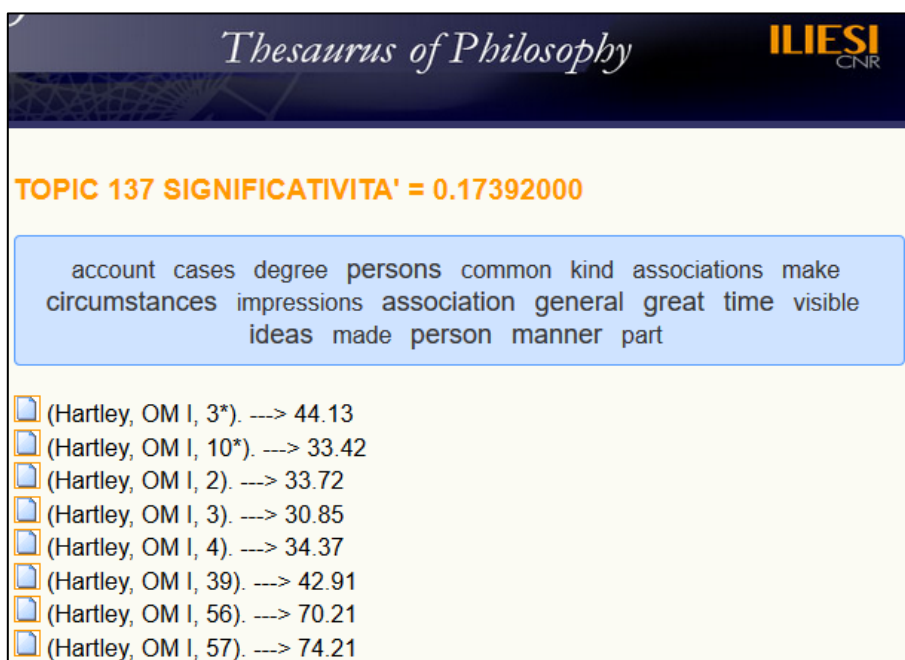


Fig. 6: Output topics (termine inglese *idea*)

²⁷ MALLET è un software *Open Source* scritto da Andrew McCallum, con il contributo di numerosi studenti e del personale dell'Università del Massachusetts Amherst e dell'Università della Pennsylvania (cfr. McCallum 2002). MALLET implementa algoritmi specifici per l'analisi del *topic modeling*. Il *topic modeling* fornisce un modo semplice per analizzare grandi volumi di testo non strutturato. Un *topic* è costituito da un insieme di parole che compaiono spesso insieme. Usando semplici regole di campionamento probabilistico, i *topic models* possono connettere parole con significati simili e distinguere gli usi di parole con più significati (<<http://mallet.cs.umass.edu/topics.php>>, ultimo accesso novembre 2016).

presenza di un determinato *topic*, e che sono anch'essi cliccabili per visualizzare il testo.

Un ulteriore uso del software potrebbe consistere nell'individuazione dei prestiti linguistici non adattati. MALLEET, infatti, elabora *clusters* che includono termini chiave per l'individuazione di argomenti rilevanti, prescindendo dalla lingua d'appartenenza. La presenza di questa tipologia di prestiti è a nostro avviso particolarmente significativa, in quanto testimone di importanti processi di *translatio*, dell'attività dei traduttori e delle scelte lessicali da essi compiute al fine di veicolare e diffondere il sapere filosofico.

In un sistema di *IR* assistito attraverso un *thesaurus*, la *query* effettuata dall'utente viene espansa, cioè riformulata attraverso l'acquisizione di termini che risultano essere in relazione con quelli ricercati dall'utente. Il sistema di entrate fino ad ora implementato può essere consultato attraverso una gamma di relazioni piuttosto articolata a partire da ciascuna delle singole entrate; abbiamo però anche visto che il sistema non sempre garantisce una buona *recall* dell'informazione, a causa di problemi soprattutto di tipo morfologico. Esso è inoltre ancora carente sul piano della *precision*, a causa della mancanza di una strutturazione semantica più fine, orientata cioè all'implementazione delle relazioni semantiche interlessemiche e/o alla definizione delle relazioni che intercorrono tra le accezioni proprie di ciascun termine e i relativi traducenti. Questi ultimi aspetti, se sviluppati, avrebbero un ruolo decisivo, da un lato per il potenziamento della precisione del sistema di interrogazione, dall'altro per una riflessione propriamente filosofico-linguistica sul tema della traduzione nella storia delle idee scientifiche e filosofiche.

6. TECNOLOGIA

TheofPhilo è stato progettato e sviluppato in ambiente PHP/MySQL: nella fig. 7 è riprodotto il modello concettuale dei dati. Oltre alle interfacce di ricerca, precedentemente illustrate, nel sistema sono state sviluppate le comuni funzionalità di inserimento e aggiornamento dei dati. Sono in fase di studio soluzioni per l'esportazione dei dati in formato RDF, in termini di triple del tipo *Soggetto – Predicato – Oggetto*.

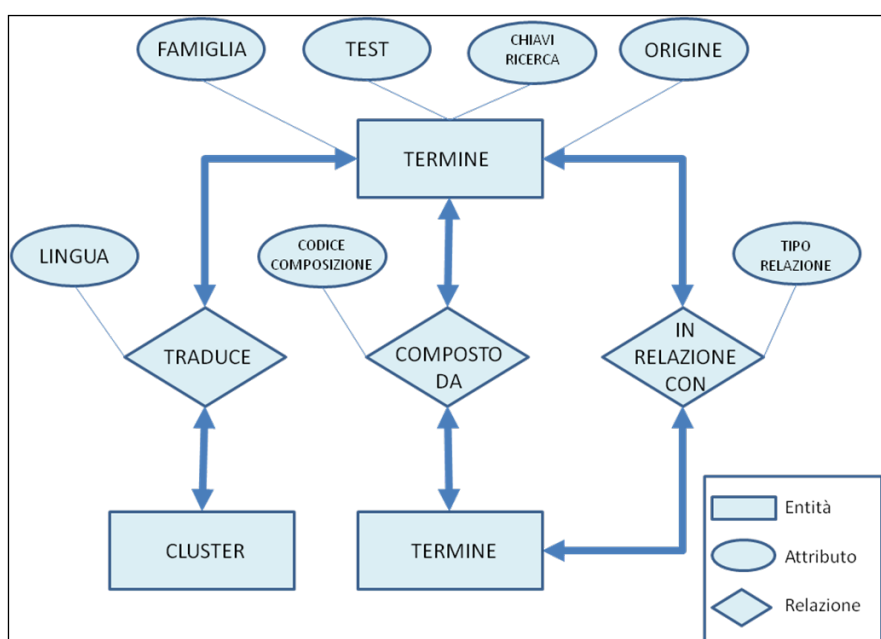


Fig. 7: TheofPhilo. Modello concettuale dei dati

Uno specifico test per la gestione di TheofPhilo è stato effettuato utilizzando il pacchetto TemaTres²⁸ (v. 1.81), applicazione web *open source* per la gestione delle rappresentazioni formali della conoscenza (tesauri, tassonomie e vocabolari multilingue). Il software, anch'esso sviluppato in ambiente PHP/MySQL, utilizza le relazioni standard per la costruzione di tesauri, offrendo inoltre la possibilità di esportare i metadati in diversi schemi (tra cui SKOS-RDF²⁹). Alcuni fattori, come la difficoltà nella gestione degli id-set e l'inesistenza di un *thesaurus* per il latino, hanno però indotto a

²⁸ Cfr. <<http://www.vocabularyserver.com/>>, ultimo accesso novembre 2016.

²⁹ Cfr. <<http://www.w3.org/2004/02/skos>>, ultimo accesso novembre 2016.

proseguire il test senza adottare, almeno per il momento, tale software.

Le interfacce di ricerca nei testi sono state sviluppate in PHP; i programmi di accesso ai documenti contenuti nell'archivio *Banca dati* sono stati invece sviluppati in Java, attraverso l'uso di JavaServer Pages. Le interfacce per l'interrogazione delle annotazioni semantiche prodotte dal software Pundit richiamano specifiche API (Application Programming Interface) pubbliche che permettono di recuperare i metadati delle annotazioni di un determinato notebook pubblico generato dal tool.

Per quanto riguarda l'uso del software Mallet per il topic modeling, il programma è stato usato su tutti i documenti di uno stesso ambito linguistico, e i topics elaborati sono stati messi in relazione con i documenti in tabelle MySQL, in cui vengono successivamente ricercati i termini presenti in TheofPhilo.

7. QUESTIONI APERTE E POSSIBILI ESTENSIONI

Attualmente è possibile navigare nel *thesaurus* multilingue ed avere un quadro esaustivo delle relazioni interlinguistiche di traduzione, attraverso l'acquisizione di ciascuna delle opzioni indicate nelle fonti lessicografiche utilizzate. Come è noto, a livello interlinguistico possono essere individuati tre tipi di relazioni, che costituiscono articolazioni interne della più estesa relazione di equivalenza, segnatamente:

- i) "equivalenza esatta", riscontrabile nei casi in cui un termine nella lingua *target* è semanticamente equivalente al termine della lingua *source*, coprendone l'intero (eventuale) spettro di significati;
- ii) "equivalenza inesatta" o "quasi sinonimia", che si verifica quando un termine della lingua *target* non copre integralmente il significato del termine della lingua *source* preso in esame, ma ne lessicalizza soltanto parte delle accezioni, per cui risulta che i due termini sono pienamente interscambiabili soltanto in alcuni contesti;
- iii) "non-equivalenza", che troviamo quando la lingua *target* non presenta termini semanticamente corrispondenti.

TheofPhilo presenta numerosi esempi per ognuna di queste classi. In relazione alla prima, come si può vedere nella fig. 8, il termine greco *ἄνθρωπος* risulta associato a quattro equivalenti interlinguistici (lat. *homo*; it. *uomo*; ingl. *man*; fr. *homme*), che ne coprono l'intera gamma di significati.

The screenshot shows the 'Thesaurus of Philosophy' interface. At the top, there is a search bar with 'Cerca:' and a language dropdown menu set to 'Lingua:'. Below the search bar are 'Invia' and 'RESET' buttons. The main content area displays the entry for the Greek word *ἄνθρωπος*. It shows 'Entrata: ἄνθρωπος', 'Lingua: GRECO', and a note: 'Note: Greco tratto da ABB./MASO'. Below this, there is a section 'Cerca testi con' with buttons for 'Parole chiave', 'Topic', 'Forme', and 'Annotazioni'. A horizontal line separates this from the 'Traduzioni' section, which is labeled 'CLUSTER N. 424'. The translations are listed as follows: FRANCESE: [homme](#), INGLESE: [man](#), ITALIANO: [uomo](#), and LATINO: [homo](#). At the bottom of the entry, there are buttons for 'Modifica entrata' and 'Cancella entrata'.

Fig. 8: Equivalenti del termine greco *ἄνθρωπος*

Lo stesso accade se prendiamo in considerazione come lingua *source* il francese, esplorando gli equivalenti del termine *homme* (fig. 9).

The screenshot shows the 'Thesaurus of Philosophy' interface. At the top, there is a search bar with 'Cerca:' and 'Lingua:' fields, and buttons for 'Invia' and 'RESET'. Below the search bar, the entry for 'homme' is displayed. It includes the following information:

- Entrata:** homme
- Lingua:** FRANCESE
- Note:** Greco tratto da ABB.
- Famiglia:** F261
- Links to [homme](#) and [humain](#).

Below this, there is a section 'Cerca testi con' with buttons for 'Parole chiave', 'Topic', 'Forme', and 'Annotazioni'. A horizontal line separates this from the 'Traduzioni' section, which is titled 'CLUSTER N. 424'. The translations are listed as follows:

- GRECO:** [ἄνθρωπος](#)
- INGLESE:** [man](#)
- ITALIANO:** [uomo](#)
- LATINO:** [homo](#)

At the bottom left of the entry, there are buttons for 'Modifica entrata' and 'Cancella entrata'.

Fig. 9: Equivalenti del termine francese *homme*

Se consideriamo invece gli equivalenti del termine latino *homo*, la situazione si complica, poiché esso risulta associato, in base alle fonti utilizzate, a due *cluster* terminologici.

Questi *cluster* individuano due distinte accezioni del termine (fig. 10). Il primo (fr. *homme*; gr. *ἄνθρωπος*; ingl. *man*; it. *uomo*) fa riferimento alle tre possibili definizioni di uomo indicate da Abbagnano: i) definizioni, di natura religiosa e teologica, che si avvalgono del raffronto tra uomo e Dio; ii) definizioni che esprimono una caratteristica o una capacità propria dell'uomo (prima tra tutte, "l'Uomo è animale ragionevole"); iii) definizioni che esprimono, come propria dell'uomo, la capacità di autoprogettarsi (Abbagnano 1998: *ad vocem*).

The screenshot shows the 'Thesaurus of Philosophy' interface. At the top, there is a search bar with 'Cerca:' and 'Lingua:' fields, and buttons for 'Invia' and 'RESET'. The main content area displays the entry for 'homo' with the following details:

- Entrata:** homo
- Lingua:** LATINO
- Note:** MASO/ABBAGNANO
- Famiglia:** F91
 - [homo](#)
 - [humanus](#)
 - [inhumanus](#)
- Composti/Polirematiche:**
 - [natura hominis](#)
 - [natura rerum et hominis](#)
 - [potentia hominis](#)

Below this, there is a section 'Cerca testi con' with buttons for 'Parole chiave', 'Topic', 'Forme', and 'Annotazioni'. The 'Traduzioni' section is divided into two clusters:

- CLUSTER N. 424:**
 - FRANCESE:** [homme](#)
 - GRECO:** [ἄνθρωπος](#)
 - INGLESE:** [man](#)
 - ITALIANO:** [uomo](#)
- CLUSTER N. 839:**
 - FRANCESE:** [sujet](#)
 - GRECO:** [τίς](#)
 - INGLESE:** [subject](#)
 - ITALIANO:** [soggetto](#)

At the bottom, there are buttons for 'Modifica entrata' and 'Cancella entrata'.

Fig. 10: Equivalenti del termine latino *homo*

Il secondo (fr. *sujet*; gr. *τίς*; ingl. *subject*; it. *soggetto*) è invece espressione di un ulteriore significato che la tradizione latina ha inteso denominare facendo ricorso al termine *homo*, in relazione al pronome greco *τίς* (in opposizione ad *ἄνθρωπος*), cioè il riferimento al singolo individuo e non al genere umano (Maso 2010; Liddell-Scott 1968).

In altri casi, invece, i gruppi di termini costituiscono lessicalizzazioni diverse di un unico concetto; si tratta dunque di sinonimie o quasi-sinonimie. Riportiamo come esempio le traduzioni relative al greco *έντελέχεια* (fig. 11). Nel primo *cluster* troviamo delle vere e proprie traduzioni nelle altre quattro lingue, il secondo presenta invece una serie di calchi omonimici, una tipologia di prestito molto produttiva nelle lingue speciali, assieme

Thesaurus of Philosophy **ILIESI**
CNR

Cerca: Lingua:

Entrata: έντελέχεια
Lingua: GRECO

Cerca testi con

Traduzioni

————— CLUSTER N. 17

FRANCESE: [acte](#)

INGLESE: [act](#)
[action](#)
[activity](#)

ITALIANO: [atto](#)

LATINO: [actio](#)
[actus](#)

————— CLUSTER N. 336

FRANCESE: [entéléchie](#)

INGLESE: [entelechy](#)

ITALIANO: [entelechia](#)

LATINO: [entelechia](#)

Fig. 11: Equivalenti del termine greco *έντελέχεια*

ai prestiti non integrati e ai calchi sinonimici.³⁰

Indichiamo infine alcune ulteriori possibili estensioni del *thesaurus*. In primo luogo si prevede di acquisire la terminologia tedesca attraverso i dizionari di riferimento e, per i termini in essi non reperibili, ricorrendo eventualmente all'*Historische Wörterbuch der Philosophie* di J. Ritter, K. Gründer e G. Gottfried (Basel, 1971-2007) o alla lista delle concordanze contrastive della *Monadologia* di G. W. Leibniz (Lamarra-Palaia-Pimpinella 2001).

In secondo luogo sarebbe particolarmente utile la compilazione di una scheda terminografica (contenente una descrizione di tipo storico-etimologico, morfologico-grammaticale e semantico) al fine di arricchire il *thesaurus* con informazioni valide per lo studio della terminologia e fruibili dagli utenti degli archivi.³¹

Infine sarebbe possibile implementare la strutturazione dei termini a livello interlinguistico, indicando le relazioni semantiche interlessemiche, al fine di potenziare l'efficacia del *thesaurus* nel reperimento dell'informazione. Ci riferiamo, in particolare, alle relazioni di sinonimia, omonimia, iperonimia, iponimia, co-iponimia e antonimia,³² per la definizione delle quali è necessario individuare precise fonti lessicografiche. Una prima risorsa può essere costituita dai rinvii terminologici reperibili nei dizionari di Nicola Abbagnano e André Lalonde, attraverso l'interpretazione della tipologia di rinvio. Nel primo dizionario, ad esempio, la voce "calcolo combinatorio" è costituita esclusivamente dal rinvio al termine composto "arte combinatoria", che del primo è sinonimo; oppure il rinvio dall'entrata "aristocrazia" a "forme di governo", che della precedente è iperonimo. Utili alla strutturazione delle relazioni semantiche interlessemiche sono inoltre le definizioni relative alle accezioni

³⁰ Si vedano in proposito Renzi *et al.* 2001; Scarpa 2002, p. 32.

³¹ Una prima bozza di questa scheda è stata messa a punto dal Dott. Giancarlo Fedeli dell'ILIESI. Essa contiene un campo *Definizioni dei termini* (proprietà linguistiche; etimo; valori di significato standard e settoriali; uso; calchi), un campo *Definizione delle relazioni* (pseudo gerarchiche – *broader term*, *narrower term*, *related term*; semantiche, inclusi i trofi, omonimia, sineddoche, metonimia, allegoria etc.), un campo *Contesti* (contesti d'uso e contesti storici), cui si aggiunge un campo relativo alle equivalenze interlinguistiche.

³² Le stesse relazioni, pensate per essere implementate con la tecnologia RDF al fine di costruire l'ontologia di dominio, erano state proposte nel corso del progetto *Agora*. I predicati erano infatti i seguenti: *Synonymy: IsSynonymOf-HasSynonym*; *Omonymy: IsOmonymOf-HasOmonym*; *Hyperonymy: IsHyperonymOf-HasHyperonym*; *Hyponymy: IsHyponymOf-HasHyponym*; *Co-Hyponymy: IsCo-HyponymOf-HasCo-Hyponym*; *Antonymy: IsAntonymOf-HasAntonym*.

filosofiche dei termini, reperibili nei dizionari etimologici monolingue. Tutte le relazioni dovrebbero ovviamente essere indicate secondo le normative standard esistenti, come nel caso del già citato SKOS-RDF,³³ per la strutturazione di tesauri multilingue.

8. CONCLUSIONI

In questo lavoro abbiamo riassunto le fasi di elaborazione teorica e di realizzazione di TheofPhilo. *Thesaurus of Philosophy*, il cui sviluppo è stato condotto nel contesto dei due progetti di ricerca *Agora. Scholarly Open Access Resource in European Philosophy* e *SM@RTINFRA-SSHCH. Smart Integrated infrastructure (for the data ecosystem) of Social Sciences, Humanities and Cultural Heritage*.

Abbiamo quindi illustrato il framework entro il quale il tesaurus è stato pensato, cioè la costruzione di un'ontologia per la rappresentazione dei contenuti testuali, autoriali e concettuali propri dei materiali raccolti nel portale *Daphnet*. In relazione all'ontologia, la terminologia strutturata è stata impiegata per il popolamento della classe dei *Philosophical Subject*. Nelle sezioni dedicate alla metodologia abbiamo discusso le fasi di lavoro, dalla selezione dei termini, alle fonti lessicografiche utilizzate per l'individuazione degli equivalenti interlinguistici e per la loro strutturazione intralinguistica. I paragrafi 4 e 5 sono rispettivamente dedicati all'interfaccia dello strumento e alle varie tipologie e funzionalità di ricerca (per lemmi, per *topic*, per forme e per annotazioni semantiche). Nelle ultime due sezioni di questo lavoro sono state invece discusse la tecnologia (testata e adottata) e le possibili estensioni dello strumento, come l'acquisizione del lessico filosofico tedesco, l'inserimento delle definizioni e la strutturazione intralinguistica secondo le relazioni semantiche interlessemiche.

In conclusione vorremmo tentare un bilancio del lavoro svolto, concentrandoci in particolare sul test condotto nella seconda fase del percorso e finalizzato alla verifica della funzionalità di TheofPhilo in quanto sistema di *IR*. Il test ha avuto un esito molto positivo, perché le funzionalità di ricerca implementate, sia di tipo lessicale che di tipo semantico, sono effettivamente utilizzabili e operative all'interno degli

³³ Cfr. paragrafo 6.

archivi. Alcuni aspetti richiedono tuttavia un'ulteriore riflessione, affinché si possa completare lo strumento e renderlo fruibile in modalità OA dalla comunità degli studiosi.

In primo luogo riteniamo necessario applicare all'insieme dei testi, in modo sistematico e omogeneo, le varie funzionalità di ricerca testate, estendendo quindi a tutta la terminologia le quattro modalità (per lemmi, per forme, per *topics* e per annotazioni semantiche). In secondo luogo sarà decisivo implementare le relazioni, tesauriali e semantico-lessicali, affinché le *queries* lanciate dagli utenti possano dare risultati più granulari, siano quindi più sia sul piano della *recall* che su quello della *precision*. Il terzo aspetto concerne invece la necessità di individuare strumenti che permettano di visualizzare in modo sintetico e leggibile gli *output*, che sono in molti casi, soprattutto se si tratta di esiti di ricerca per forme, piuttosto ridondanti.

L'intento che anima il progetto è infatti quello di rendere le piattaforme e gli archivi digitali dell'ILIESI un ambiente di lavoro scientificamente attendibile, dinamico e collaborativo, in grado di favorire le attività di ricerca storico-filosofica sui materiali disponibili. La possibilità di navigare nei testi attraverso un sistema di chiavi di accesso multilingue, costituisce senza dubbio un elemento di originalità e di grande utilità per gli studiosi, sia che essi conoscano, ma soprattutto se non conoscono, le lingue in cui i testi sono stati scritti. Alla terminologia della lingua originale si può effettivamente giungere lanciando *queries* nella propria lingua madre o in un'altra conosciuta. Il *topic modeling*, e ancor di più le annotazioni semantiche, sono tipologie di sperimentazione estremamente interessanti e che saranno certamente potenziate.

La possibilità di fare ricerche di tipo concettuale, in base alle relazioni (quantitative) tra le occorrenze terminologiche raggruppate in *clusters* e in base alle annotazioni al testo fatte da altri ricercatori, si affianca dunque alle ricerche di tipo lessicale, offrendo alla comunità scientifica uno strumento efficace per ricerche diversificate e articolate all'interno dei testi. Aprendo inoltre ad altri studiosi la possibilità di annotare semanticamente i testi, la piattaforma *Daphnet* e gli archivi digitali dell'Istituto diventerebbero un ambiente virtuale ricco e dinamico, ove lo scambio e l'interazione determinerebbero un ampliamento della conoscenza attorno agli autori, ai testi e ai temi in essi discussi.

Riferimenti bibliografici

- Adamo, G.-Della Valle, V., *Le parole del lessico italiano*, Roma, Carocci 2008.
- Andrews P.-Zaihrayeu I.-Pane J., *A Classification of Semantic Annotation Systems*, "Semantic Web. Interoperability, Usability, Applicability", 3, 3, 2012.
- Cortelazzo, M. A., *Lingue speciali. La dimensione verticale*, Padova, Unipress 1994.
- Dardano, M., *Manualetto di linguistica italiana*, Bologna, Zanichelli 1996.
- D'lorio, P., *Discovery D1.8 Final Report*, ECP 2005 CULT 038206, 2009.
- Fattori, M.-Bianchi, M. L. (a cura di), *Lessico filosofico dei secoli XVII e XVIII, Sezione latina*, vol. I, 1, a-aetherius, Firenze, Olschki 1992.
- Grassi, M. et al., *Pundit: Semantically Structured Annotations for Web Contents and Digital Libraries*, in Mitschick, A. et al. (eds), *Semantic Digital Archives 2012: Proceedings of the Second International Workshop on Semantic Digital Archives*, Paphos, CY, September 2012, <<http://ceur-ws.org/Vol-912/paper4.pdf>>, ultimo accesso novembre 2016.
- Grossmann, M.-Rainer, F. (a cura di), *La formazione delle parole in italiano*, Tübingen, Niemeyer 2004.
- Gruber T. R., *A Translation Approach to Portable Ontology Specification*, "Knowledge Acquisition", 5, 2, 1993, pp. 199-220.
- Lamarra, A.-Palaia, R.-Pimpinella, P., *Le prime traduzioni della "Monadologie" di Leibniz (1720-1721): introduzione storico-critica, sinossi dei testi, concordanze contrastive*, Firenze, Olschki 2001.
- Lamarra A.-Marras C., *Scholarly Open Access Research in Philosophy: Limits and Horizons of a European Innovative Project*, *Digital Humanities of Nebraska* 2013, <<http://dh2013.unl.edu/abstracts/ab-316.html>>, ultimo accesso novembre 2016.
- Lamarra A.-Tardella M., *TheofPhilo. A prototype for a Thesaurus of Philosophy*, in *Digital Humanities, Lausanne-Switzerland 2014, Book of Abstracts*, 2014, pp. 235-237. <<http://dharchive.org/paper/DH2014/Paper-636.xml>>, ultimo accesso novembre 2016.
- Leibniz, G. W., *Nouveaux Essais*, in Robinet, A.- Schepers, H. (eds.), *Sämtliche Schriften und Briefe*, ser. 6: *Philosophische Schriften*, vol. 6, Darmstadt/Leipzig/Berlin, Akademie-Verlag 1962.
- Liburdi A.-Russo A.-Tardella M., *Ricerca filosofica e archivi digitali nella diffusione della cultura scientifica*, "Digital Humanities e beni culturali: quale relazione?", Quarto convegno annuale dell'AIUCD, Torino, 17-19 dicembre 2015.
- Magris, M. et al. (a cura di), *Manuale di terminologia*, Milano, Hoepli 2002.
- McCallum, A. K., *MALLET: A Machine Learning for Language Toolkit*, 2002, <<http://mallet.cs.umass.edu>>, ultimo accesso novembre 2016.
- Mutschmann, H.-Mau, J., *Sexti Empirici Opera recensuit* H. M., coll. BT, Leipzig, Teubner 1912-1954, 4 voll.
- Renzi, L. et al. (a cura di), *Grande grammatica italiana di consultazione*, Bologna, Il Mulino 2001.
- Russo, A., *A Data-Base for the Compiling of a Philosophical Dictionary*, in *Computers in Literary and Linguistic Computing – L'ordinateur et les recherches littéraires et linguistiques*, Proceedings of the Eleventh International Conference of the ALLC, Paris-Genève, Champion-Slatkine 1985, pp. 321-333.
- Scarpa, F., *Terminologia e lingue speciali*, in Magris et al. (a cura di), 2002, pp. 27-47.
- Signore, O., *RDF per la rappresentazione della conoscenza*, in *Knowledge Management (V edizione) – Forum Proceedings on the knowledge management in the organizations*, 2002, pp. 35-46 <<http://www.w3c.it/papers/RDF.pdf>>, ultimo accesso novembre 2016.
- Spinosa, G., *Un "Thesaurus strutturato" per la gestione automatica del lemmario latino del Lessico Filosofico dei secoli XVII e XVIII*, "Lexicon Philosophicum", 2, 1986, Roma, Edizioni dell'Ateneo, pp. 17-30.

- Tardella, M.-Marras, C., *Nuovi scenari per la ricerca in filosofia: i testi e gli strumenti del portale Daphnet*, in M. Agosti and F. Tomasi (Eds), *Collaborative Research Practices and Shared Infrastructures for Humanities Computing*, 2nd AIUCD Annual Conference, AIUCD 2013, Padua, Italy, 11-12 December 2013, Padova, CLEUP 2013, pp. 39-54.
- Uren, V. et al., *Semantic Annotation for Knowledge Management: Requirements and a Survey of the State of the Art*, "Web Semantics: Science, Services and Agents on the World Wide Web", 4, 2006, pp. 14-28.

Dizionari e risorse lessicografiche

- Abbagnano, N., *Dizionario di Filosofia*, Torino, UTET 1998.
- Bailly, A., *Dictionnaire Grec-Français*, Édition revue par L. Séchan et P. Chantraine, Paris, Hachette 1950.
- Beccaria, G. L. (diretto da), *Dizionario di linguistica, e di filologia, metrica, retorica*, Torino, Einaudi 1996.
- Forcellini, E., *Lexicon totius latinitatis*, Padova, Typis Seminarii 1940.
- Lalande, A., *Vocabulaire technique et critique de la philosophie*, Paris, PUF 1983.
- Liddell, H. G.- Scott, R., *Greek-English Lexicon*, Rev. by H. S. Jones, Oxford, Clarendon Press 1968.
- Maso, S., *Lingua Philosophica Graeca*, Milano-Udine, Mimesis 2010.
- Ritter, J.-Gründer, K.-Gottfried, G. (eds), *Historische Wörterbuch der Philosophie*, Basel, Schwabe 1971-2007.
- Rocci, L., *Vocabolario Greco-Italiano*, Perugia, Dante Alighieri 1993.
- Sanesi, T., *Vocabolario Italiano-Greco*, Pistoia-Siena, Fratelli Bracali 1916.
- Il Vocabolario Treccani*, Roma, Istituto dell'Enciclopedia Italiana 1997.

Sitografia (ultimo accesso novembre 2016)

- Applicazione web per la rappresentazione della conoscenza TemaTres
<http://www.vocabularyserver.com/>
- Banca dati dei testi filosofici dell'età moderna
<http://www.iliesi.cnr.it/bancadati>
- Daphnet. Digital Archives of PHilosophical text on the NET
<http://www.daphnet.org/>
- Istituto di Linguistica Computazionale-CNR
<http://www.ilc.cnr.it/>
- Istituto per il Lessico Intellettuale Europeo e Storia delle Idee-CNR
<http://www.iliesi.cnr.it/>
- MALLET: A Machine Learning for Language Toolkit
<http://mallet.cs.umass.edu>
- Net7
<http://www.netseven.it/>
- Ontologia filosofica elaborata dall'Università dell'Indiana
<https://inpho.cogs.indiana.edu/>
- ORTOLANG. Outils et Ressources pour un Traitement Optimisé de la LANGue.
<http://www.cnrtl.fr/etymologie/>
- Oxford Dictionary
<http://www.oxfordlearnersdictionaries.com>
- Progetto Agora. Scholarly Open Access Research in European Philosophy
<http://www.netseven.it/project/agora>

Progetto Discovery

<http://www.netseven.it/project/discovery>

Pundit - Semantic web annotator

<http://thepund.it>

Pundit - Semantic web annotator (Notebooks)

<http://ask.thepund.it/>

Resource Description Framework (RDF)

<http://www.w3.org/RDF/>

SKOS Simple Knowledge Organization System

<https://www.w3.org/2004/02/skos>

Thesaurus multilingue pluridisciplinare dell'attività dell'Unione Europea (SKOS/RDF)

<http://eurovoc.europa.eu/drupal/?q=it>



ILIESI digitale Memorie



Istituto per il Lessico Intellettuale Europeo e Storia delle Idee

2017

