DOI: 10.26314/GARR-Conf22-proceedings-07

Il Laboratorio di Filologia Collaborativa e Cooperativa (CoPhiLab) del CNR-ILC: dati, strumenti, servizi e infrastrutture

Federico Boschetti[1,2], Cosimo Burgassi[1], Riccardo Del Gratta[1], Angelo Mario Del Grosso[1], Elisa Guadagnini[1], Ouafae Nahli[1], Simone Zenzaro[1]

[1] Istituto di Linguistica computazionale "A. Zampolli" - CNR, [2]VeDPH - DSU - Università Ca' Foscari Venezia

Abstract. Questo contributo illustra le attività e le risorse del Laboratorio di Filologia Collaborativa e Cooperativa (CoPhiLab) dell'Istituto di Linguistica Computazionale "A. Zampolli" del Consiglio Nazionale delle Ricerche (CNR-ILC), con particolare attenzione all'uso delle infrastrutture di ricerca nazionali e internazionali.

Keywords. Filologia Computazionale, Modelli Formali, Lingua Araba, Domain-Specific Languages, Ingegneria del Software

Introduzione

Questo contributo illustra le attività e le risorse del Laboratorio di Filologia Collaborativa e Cooperativa (CoPhiLab) dell'Istituto di Linguistica Computazionale "A. Zampolli" del Consiglio Nazionale delle Ricerche (CNR-ILC). Dopo una breve presentazione del laboratorio e la sua collocazione nel contesto italiano delle Digital Humanities (DH), sono descritte le linee di ricerca del CoPhiLab e la loro applicazione alla didattica e alla formazione. Tra i diversi servizi infrastrutturali offerti dal laboratorio, sono descritti più approfonditamente quelli proposti all'interno di CLARIN, che consistono da un lato nel fornire risorse e tecnologie al repository nazionale ILC4CLARIN, dall'altro lato nella condivisione di conoscenze e buone pratiche grazie al nuovo Knowledge Centre di CLARIN dedicato alle DH, il DiPText-KC, costituito in collaborazione con il Venice Centre for Digital and Public Humanities (VeDPH), che afferisce al Dipartimento di Studi Umanistici dell'Università Ca' Foscari Venezia.

1. Presentazione del laboratorio

La filologia è lo studio dei testi in prospettiva storica: in questo senso, essa può indagare le diverse fasi del costituirsi del testo, dalle prime stesure dell'autore alla pubblicazione (non necessariamente a stampa), la storia della trasmissione - che in filologia si chiama "tradizione" - del testo, o ancora la sua ricezione, vale a dire le diverse interpretazioni che si danno del testo con il passare del tempo e il mutare dei contesti (culturali e linguistici). La filologia digitale impiega i principi e le tecniche dell'informatica e si avvale dei vantaggi che ne derivano. Uno dei principali riguarda, a nostro avviso, l'aspetto collaborativo: la filologia digitale, infatti, implementa strumenti che permettono l'accesso simultaneo di più

Conferenza GARR 2022 - Condivisioni - Selected Papers
Palermo, 18-20 maggio 2022 - ISBN: 978-88-946629-1-7 - Creative Commons CC-BY

persone alla visualizzazione e alla modifica di risorse condivise. Accanto a questo si pone l'aspetto cooperativo, che consiste nello sviluppo di risorse che garantiscano l'interoperabilità sintattica e semantica tra sistemi diversi.
Il CoPhiLab nasce nel 2013, con diverse finalità: 1) costruire modelli teorici per rappresentare la storia della tradizione testuale, sui due piani della variantistica e delle interpretazioni del testo; 2) sviluppare strumenti di analisi e di fruizione delle edizioni scientifiche digitali; 3) implementare basi di dati linguistici (sincronici e diacronici) e testuali (con varianti manoscritte e ricostruzioni congetturali).

2. Linee di ricerca

Le principali linee di ricerca del laboratorio sono:
1) la formalizzazione di entità, proprietà e relazioni del dominio degli studi filologici. Per definire in maniera non ambigua i concetti chiave della filologia (le varianti, la ricostruzione delle relazioni fra le fonti, etc.) sviluppiamo una definizione matematica di documento e di operazioni possibili su di esso, descritti da tipi di dati astratti (ADT) e all'interno di un paradigma evoluzionistico;
2) lo sviluppo di strumenti per l'annotazione collaborativa di testi letterari e documentari tramite Domain-Specific Languages (DSL). Ciò permette di coadiuvare i filologi per l'elaborazione di linguaggi formali che siano vicini alle pratiche tradizionali ma concisi, machine actionable e automaticamente convertibili in formati standard, come XML-TEI;
3) la progettazione e lo sviluppo di modelli, servizi e software di linguistica e filologia computazionale, mediante i principi del Domain-Driven Design e l'elaborazione automatica dei testi, che permettano di acquisire, produrre, analizzare, annotare, pubblicare, fruire e interrogare risorse storico-letterarie;
4) lo studio di testi in lingua araba, grazie alla creazione di risorse lessicali digitali, ottenute estraendo le informazioni da dizionari di lingua araba classica, e morfologiche, ottenute definendo pattern di parole arricchiti di informazioni morfosintattiche;
5) lo studio di testi italiani di diverse epoche storiche, mediante l'allestimento di edizioni digitali e banche dati testuali annotate;
6) l'Optical Character Recognition e l'Handwritten Text Recognition per l'acquisizione di testo dalle immagini digitali di documenti a stampa o manoscritti.
La maggior parte delle risorse prodotte sono catalogate su CLARIN e rese disponibili con licenze aperte, per poter essere fruite e interrogate mediante le tecnologie del Web (in particolare del Web semantico).

3. Ricadute sulla didattica e sulla formazione

Gli strumenti descritti, nati inizialmente per la ricerca, sono stati anche adattati per scopi didattici in diversi progetti: ricordiamo, tra gli altri, EuporiaEdu, Voci dall'Inferno, A lexical corpus-based model of Contemporary Written Arabic, Postille di Bassani.
EuporiaEdu, che ha coinvolto alcuni licei classici di Benevento, Siracusa e Pisa, consiste nella semplificazione della piattaforma web di annotazione di testi letterari tramite DSL: gli studenti sono stati coinvolti nella realizzazione di un linguaggio formale, quanto più vi-

cino possibile alle loro consuetudini di analisi linguistica, per l'annotazione di figure retoriche e per l'analisi morfologica dei testi greci previsti dal programma scolastico. Il processo di formalizzazione ha aiutato gli studenti, sotto la guida delle insegnanti, a potenziare il ragionamento induttivo; l'uso della piattaforma web ha inoltre favorito la collaborazione, anche durante il difficile periodo della pandemia.
Voci dall'Inferno è un progetto di ricerca coordinato da M. Riccucci (Università di Pisa) in collaborazione con il CoPhiLab e CLARIN-IT: esso si propone la costituzione di un corpus digitale delle testimonianze non letterarie dei sopravvissuti all'Olocausto (interviste, diari, lettere) e uno studio linguistico-filologico, mirato specificamente ad analizzare la presenza di lessico dantesco per descrivere la propria esperienza nei Lager. Il lavoro di rappresentazione digitale delle testimonianze coinvolge studenti per tirocini curriculari e tesi di laurea.
A lexical corpus-based model of Contemporary Written Arabic mira a creare una risorsa lessicografica per l'arabo scritto contemporaneo: essa potrà contribuire efficacemente alla produzione di materiali didattici e di apprendimento, soprattutto nel quadro dell'insegnamento della lingua araba nelle scuole superiori italiane. In questo modo, tra gli esiti del progetto si attende anche un impatto sociale positivo, nei termini dell'inclusione delle comunità arabofone in Italia. Nell'ambito del progetto sono attive collaborazioni nazionali (Roma Tre e IULM) e internazionali con l'Università di Fes e l'Università di Tetouan (Marocco).
Postille di Bassani è un progetto dottorale condotto da A. Siciliano che consiste nell'edizione delle postille vergate da G. Bassani durante i suoi studi: per la gestione del problema, assai complesso, della rappresentazione dei fenomeni di genesi testuale risultano straordinariamente efficaci le possibilità offerte dalla filologia digitale.

4. Servizi infrastrutturali

Tra le applicazioni web sviluppate dal CoPhiLab menzioniamo un sistema di allineamento statistico di testi paralleli in lingua ebraica, la piattaforma di analisi del testo Omega e l'ambiente di editing collaborativo CoPhiEditor. Grazie alla collaborazione con il VeDPH, il CoPhiLab ospita il progetto Cretan Institutional Inscription di I. Vagionakis e contribuisce al progetto Musisque Deoque, coordinato da P. Mastandrea.

5. CoPhiLab e CLARIN

Il laboratorio ha trovato la propria collocazione naturale all'interno dell'infrastruttura CLARIN posto che, come ha dichiarato M. Monachini, coordinatrice nazionale di CLARIN-IT , in un recente intervento al GARR, "L'utente tipo di CLARIN è lo studioso delle Scienze Umane e Sociali, il linguista, lo storico, il filologo, il filosofo, il letterato che voglia analizzare fonti testuali, ma anche il linguista computazionale". L'infrastruttura CLARIN rende disponibili per i ricercatori in scienze umane e sociali (SSH) numerose risorse e strumenti linguistici prodotti da diversi Paesi europei. CLARIN, diventata ERIC (European Research Infrastructure Consortium) nel 2012, è la prima Infrastruttura di Ricerca pensata da studiosi delle SSH per studiosi delle SSH. Nel suo manifesto è chiaro l'obiettivo di creare e mantenere una Infrastruttura di Ricerca FAIR . I centri CLARIN assegnano un identificativo persistente (PID) alle risorse rendendole accessibili, anche con metodi di

autenticazione e autorizzazione. Le risorse descritte sono interoperabili grazie al set di metadati standard (CMDI). Infine, CLARIN ha tra le sue linee guida quella di gestire le differenti versioni dei materiali descritti in modo da rendere le analisi linguistiche replicabili nel lungo termine. Al consorzio nazionale italiano CLARIN-IT afferiscono il centro nazionale ILC4CLARIN e il centro di ricerche CLARIN Eurac (ERCC). Dalla collaborazione del VeDPH con il CNR-ILC e ILC4CLARIN, è nato un nuovo Knowledge Centre , il Digital and Public Textual Scholarship K-Centre (DiPText-KC), che si avvale delle competenze specifiche dei suoi membri nel settore delle DH per favorire il trasferimento tecnologico e di conoscenze.

6. Conclusione

Il CoPhiLab contribuisce alla crescita delle Digital Humanities in Italia nel solco della tradizione del CNR-ILC e in sinergia con gli altri luoghi della ricerca, universitari e del CNR. Il laboratorio promuove l'interoperabilità tra le risorse prodotte dai suoi membri e quelle già esistenti e disponibili nelle infrastrutture di ricerca, facilita la conservazione a lungo termine delle risorse tramite il repository di ILC4CLARIN e contribuisce alla valorizzazione del patrimonio culturale tramite progetti ICT. Inoltre, il laboratorio supporta attivamente la formazione di studenti e giovani ricercatori su temi di filologia digitale.

Riferimenti bibliografici

F. Boschetti, G. Mugelli (2021). "Il metodo Euporia per creare nuovi archivi digitali sulla tragedia greca". FuturoClassico FCl, (7), 83-113.

D. Del Fante, F. Frontini, M. Monachini, V. Quochi (2021). "CLARIN-IT: An Overview on the Italian CLARIN Consortium After Six Years of Activity," 18th Italian Research Conference on Digital Libraries (IRCDL 2022).

R. Del Gratta, F. Boschetti, L. Bambaci, F. Sarnari (2020). "Approaching document analysis with a formal model", 6th IEEE Congress on Information Science and Technology (CiSt), pp. 208-214, doi: 10.1109/CiSt49399.2021.9357202.

A. M. Del Grosso, D. F. Fihri, M. el Mohajir, O. Nahli, A. Tonazzini (2020). "Digital safeguard of laminated historical manuscripts: the treatise 'Poem in Rajaz on medicine' as a case study" 6th IEEE Congress on Information Science and Technology (CiSt), pp. 192-197, doi: 10.1109/CiSt49399.2021.9357192.

O. Nahli, A. M. Del Grosso (2020). "Creating Arabic Lexical Resources in TEI: A Schema for Discontinuous Morphology Encoding", 6th IEEE Congress on Information Science and Technology (CiSt), pp. 178-187, doi: 10.1109/CiSt49399.2021.9357273.

I. Vagionakis, R. Del Gratta, F. Boschetti, P. Baroni, A. M. Del Grosso, T. Mancinelli, e M. Monachini (2022). "Selected Papers from the CLARIN Annual Conference 2021". In 'Cretan Institutional Inscriptions' Meets CLARIN-IT, a cura di F. de Jong e M. Monachini, 189. CLARIN ERIC. https://doi.org/10.3384/9789179294441.

S. Zenzaro, A. M. Del Grosso, F. Boschetti, e G. Ranocchia (2022). "Verso la definizione di criteri per valutare soluzioni di scholarly editing digitale: il caso d'uso GreekSchools". In Culture Digitali. Intersezioni, Filosofia, Arti e Media. Università del Salento - AIUCD.

https://doi.org/10.6092/unibo/amsacta/6848.
https://www.clarin.eu/
https://www.clarin-it.it/
https://www.go-fair.org/
https://www.clarin.eu/content/cmdi-12
https://ilc4clarin.ilc.cnr.it/
https://clarin.eurac.edu/
https://diptext-kc.clarin-it.it

Autori

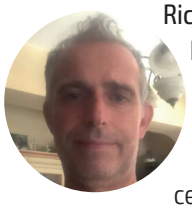
Federico Boschetti federico.boschetti@ilc.cnr.it

Federico Boschetti si è laureato nel 1998 in Lettere Classiche (Università Ca' Foscari Venezia). Presso l'Università di Trento ha conseguito il dottorato di ricerca nel 2005 in Filologia Classica (in cotutela con Lille III) e nel 2010 in Cognitive and Brain Sciences - Language, Interaction, and Computation.
Dal 2011 è ricercatore presso il CNR-ILC.
Interessi di ricerca: Filologia digitale, Filologia collaborativa e cooperativa, OCR/HTR e Semantica distribuzionale applicata a testi antichi.

Cosimo Burgassi cosimo.burgassi@ilc.cnr.it

Cosimo Burgassi è ricercatore all'Istituto di Linguistica Computazionale "Antonio Zampolli" (CNR – Pisa). Ha studiato filologia italiana e romanza a Firenze e a Pisa, ha collaborato come "giovane ricercatore" al progetto DiVo – Dizionario dei Volgarizzamenti (dir. E. Guadagnini e G. Vaccaro) e ha insegnato materie storico-letterarie nella scuola secondaria. Si occupa principalmente di storia del lessico (in particolare di semantica storica) e di testi di traduzione.

Riccardo Del Gratta riccardo.delgratta@ilc.cnr.it



Riccardo Del Gratta è Ricercatore a tempo pieno presso l'Istituto di Linguistica Computazionale del CNR "Antonio Zampolli" di Pisa. I suoi interessi di ricerca spaziano dalla formalizzazione di entità, proprietà e relazioni del dominio degli studi filologici alla integrazione di servizi e risorse all'interno di infrastrutture di ricerca. Dal 2015 è responsabile dei centri italiani di CLARIN.

Angelo Mario Del Grosso angelo.delgrosso@ilc.cnr.it

Angelo Mario Del Grosso è Ingegnere Informatico e dal 2019 è Ricercatore presso l'Istituto di Linguistica Computazionale del CNR "A. Zampolli". I suoi interessi comprendono lo sviluppo di modelli e sistemi avanzati per la linguistica e la filologia computazionale finalizzati alla produzione, rappresentazione, analisi, fruizione e interrogazione di testi sia a stampa sia manoscritti. Dal 2018 è docente esterno di Codifica di Testi presso il Corso di Laurea in Informatica Umanistica dell'Università di Pisa.

Elisa Guadagnini elisa.guadagnini@ilc.cnr.it
Elisa Guadagnini è Primo ricercatore presso l'ILC-CNR. Filologa romanza di formazione, si è occupata a lungo di lessico italiano medievale e dell'eredità dei Classici nel Medioevo romanzo; è stata il P.I. del progetto «DiVo - Dizionario dei Volgarizzamenti» (FIRB 2010). Da qualche anno si occupa prevalentemente di lessicologia storica, lessicografia digitale e Corpus Linguistics. Collabora con il CoPhiLab dal 2021.

Ouafae Nahli ouafae.nahli@ilc.cnr.it
Ouafae Nahli è Ricercatore a tempo pieno presso l'Istituto di Linguistica Computazionale del CNR "Antonio Zampolli" di Pisa. I suoi interessi di ricerca e contributi coprono una serie di aspetti dell'arabo moderno classico e standard, comprese le questioni morfo-sintattiche e lessico-semantiche nell'elaborazione del linguaggio naturale arabo, l'analisi computazionale dei testi letterari e la modellazione lessico-ontologica.

Simone Zenzaro simone.zenzaro@ilc.cnr.it
È assegnista di ricerca presso ILC-CNR di Pisa sul progetto ERC 885222-GreekSchools nell'ambito della papirologia computazionale.
Ha lavorato all'Université de Lausanne sull'edizione digitale e applicazione di tecniche NLP al progetto Le devenir numérique d'un texte fondateur.
Ha lavorato presso la Scuola Normale Superiore su strumenti per l'edizione digitale di manoscritti arabi. I suoi interessi riguardano le Digital Humanities, modelli, servizi e l'applicazione di metodi formali al dominio della filologia.