

Alcune note di ecdotica digitale nella prospettiva della filologia collaborativa e cooperativa

Federico Boschetti

(Istituto di Linguistica Computazionale "A. Zampolli", CNR-ILC, Pisa, Italia)

Abstract In the digital age, classical philology is becoming more and more collaborative and cooperative. Even if these terms are apparently synonyms, the distinction adopted by pedagogues in the last decades is useful to describe different ways to organise collective works. Collaboration aims at the realisation of a final product, which is the result of negotiations among the involved parts. Cooperation aims at the creation of independent subproducts, which will be coordinated to achieve the agreed final result, but which can also be assembled in unpredictable new ways by third parties. In order to integrate these approaches in the domain of classical philology, first of all it is necessary to clarify the needs of a typical philologist, not just the requirements of the user of a specific project. For this reason, the last part of the article is devoted to a case study that takes into account the exploration of the secondary literature, often neglected by digital philologists but crucial to the ecdotic best practices.

Sommario 1 Introduzione. – 2 Integrazione fra filologia collaborativa e filologia cooperativa. – 3 La specializzazione nell'ambito della filologia cooperativa. – 4 La peculiarità della filologia cooperativa rispetto alla linguistica computazionale. – 5 Dall'interscambio all'interoperabilità. – 6 Il coordinamento dei mezzi minimi per l'esercizio della filologia cooperativa. – 7 Discussione di un caso d'uso. – 8 Conclusione.

Keywords Classical philology. Collaborative and cooperative philology. Digital philology. Ecdotics. Perseus Digital Library.

1 Introduzione

L'edizione scientifica digitale si sta progressivamente differenziando dall'edizione a stampa, in quanto si configura sempre più come un insieme di viste su un processo aperto anziché come un singolo prodotto in sé concluso. Nell'era della cosiddetta quarta rivoluzione (Roncaglia 2010), il trasferimento dal supporto cartaceo ai supporti digitali aumenta inevitabilmente il rischio di una deriva testuale e per questo la riflessione teorica sulle pratiche editoriali deve accompagnare sia l'opera di digitalizzazione delle risorse già esistenti, sia la creazione di risorse digitali native (Pierazzo 2014).

2 Integrazione fra filologia collaborativa e filologia cooperativa

Negli studi filologici, l'evoluzione della rete telematica in social network ha comportato l'apertura della scrivania del singolo studioso alla comunità dei pari non più soltanto per fruire del prodotto finale, licenziato dall'autore o dall'editore, ma anche per intervenire in corso d'opera. La filologia nell'era digitale è chiamata quindi a essere collaborativa e cooperativa (per la distinzione fra i due termini, si veda Kozar 2010). La collaborazione prevede che una comunità con obiettivi condivisi coordini i propri sforzi per la realizzazione di un prodotto, come ad esempio l'edizione critica di un'opera antica. La cooperazione invece permette ai membri o ai gruppi di diverse comunità di perseguire i propri obiettivi specifici, usando i servizi offerti dagli altri e offrendo in cambio i propri, come ad esempio il servizio di analisi metrica e il servizio di lemmatizzazione che, combinati da terze parti, possono produrre l'analisi formulare, anche se non pianificata inizialmente dai fornitori dei servizi originari.

La filologia cooperativa promuove quindi il disaccoppiamento delle funzionalità offerte dagli obiettivi specifici che devono essere raggiunti. Questo comporta il passaggio dallo sviluppo di applicazioni monolitiche, nate per un singolo progetto, alla creazione da un lato di archivi digitali codificati in modo standard e annotabili da terze parti e dall'altro allo sviluppo di librerie di componenti e di servizi web, che possono essere riusati per nuovi progetti di ricerca (Boschetti, Del Grosso 2015). L'approccio collaborativo e l'approccio cooperativo non sono mutualmente esclusivi, ma anzi si integrano reciprocamente. Infatti, le risorse digitali o gli strumenti computazionali realizzati in seno ad un progetto collaborativo amplificano la loro utilità quando sono impiegati all'interno di altri progetti realizzati da terze parti.

Ad esempio, sul versante delle risorse, la cooperazione fra i progetti collaborativi *Musisque Deoque* e *Perseus Project* permette agli studiosi di fruire dei testi latini rinascimentali sia attraverso l'interfaccia originaria che attraverso l'interfaccia della *Perseus Digital Library*. L'interfaccia originaria (<http://www.mqd.it>) è nata per lo studio dell'intertestualità e mette a disposizione un motore di ricerca avanzato e un sistema di visualizzazione che facilita la comparazione dei testi. D'altro canto, l'interfaccia della *Perseus Digital Library* (<http://www.perseus.tufts.edu>) mostra la sua massima efficacia per lo studio linguistico, grazie all'associazione di ciascuna parola alla relativa analisi morfologica e alla relativa voce lessicale proveniente da dizionari bilingui consultabili online. A loro volta, i testi della *Perseus Digital Library* sono accessibili anche attraverso *Perseus Under PhiloLogic*, grazie all'interfaccia sviluppata presso l'Università di Chicago (<http://perseus.uchicago.edu>), che potenzia ulteriormente i servizi di visualizzazione delle analisi linguistiche.

Un secondo esempio, questa volta sul versante degli strumenti e dei servizi, è l'accesso tramite *web service* alle analisi prodotte dallo strumento di

lemmatizzazione latina sviluppato presso l'Istituto di Linguistica Computazionale "A. Zampolli" (end point: <http://cophilab.ilc.cnr.it:8000/aucepsws/morphanalysis/j/> seguito dalla forma flessa latina da analizzare, ad es. cano). Nel fornire questo servizio, si è seguita la strada dell'*Alpheios Project*, che ha creato dei web services per la lemmatizzazione delle lingue storiche (<https://github.com/PerseusDL/morpheus-wrapper>) a partire dall'originario sistema di analisi morfologica *Morpheus* (Crane 1991).

3 La specializzazione nell'ambito della filologia cooperativa

Nella fase pionieristica e artigianale dello sviluppo di software per le *Digital Humanities*, il filologo per lo più interagiva con un singolo programmatore oppure creava autonomamente i propri programmi, facendo convergere tutte le competenze necessarie in una o al massimo due figure professionali. Nella fase matura delle *Digital Humanities*, dove più individui e più gruppi di diverse comunità interagiscono per la creazione di prodotti digitali, la specializzazione si rende necessaria. In particolare, è bene tener distinte le competenze richieste per l'elaborazione di modelli formali (ad esempio, per la modellazione della struttura dati più efficaci a rappresentare la variantistica), per la creazione di risorse digitali (ad esempio, l'archivio di *Musisque Deoque*), per lo sviluppo di strumenti computazionali (ad esempio, gli algoritmi di *Tesserae* per lo studio dell'intertestualità: <http://tesserae.caset.buffalo.edu>) e per la costruzione di infrastrutture (ad esempio, la *Perseus Digital Library*). La filologia cooperativa è costituita quindi dai seguenti sottodomini: formale, digitale, computazionale e infrastrutturale (o *e-philology*). I modelli formali degli oggetti e dei processi coinvolti nella disciplina filologica prescindono dalle contingenze delle tecnologie disponibili a implementarli e dirigono le modalità di codifica digitale, di analisi computazionale e di gestione infrastrutturale.

Un esempio dovrebbe chiarire e anticipare quanto sarà esposto nelle sezioni seguenti. I manoscritti da un lato e le collazioni dall'altro sono rispettivamente oggetti e processi dell'ambito filologico che devono essere formalizzati a prescindere dal livello tecnologico raggiunto per acquisire immagini ad alta definizione o dallo specifico linguaggio di programmazione con cui è implementato un algoritmo per la collazione automatica. A livello formale, un manoscritto deve essere descritto tramite le relazioni delle sue parti costituenti e la collazione deve essere descritta tramite la sequenza delle operazioni elementari di cui è composta. La codifica digitale del manoscritto, la implementazione dell'algoritmo per la collazione automatica di più manoscritti e la creazione dell'infrastruttura web che permette l'interazione fra oggetti e processi si devono conformare a tali modelli astratti.

4 La peculiarità della filologia cooperativa rispetto alla linguistica computazionale

Fin dalle origini con padre Busa, la ricerca filologica in ambito digitale ha mutuato strumenti e metodi dalla linguistica computazionale. Se da un lato questo ha permesso di analizzare testi di interesse storico e letterario sfruttando i risultati delle ricerche in campo linguistico, dall'altro ha frenato lo sviluppo di strumenti e metodi propri dell'ambito filologico.

La parte della linguistica che più si avvicina agli interessi del filologo è la *Corpus Linguistics* (McEnery, Wilson 1996), in quanto quest'ultima concentra la sua attenzione sui fenomeni della lingua che si possono estrarre da collezioni di testi reali e non da insiemi di esempi fittizi. Alla sua natura descrittiva, la *Corpus Linguistics* aggiunge inoltre la componente quantitativa, per cui i fenomeni non vengono soltanto individuati, ma anche conteggiati e pesati, garantendo un alto livello di oggettività e di falsificabilità ai risultati raggiunti.

Tuttavia il linguista non è interessato alla materialità e ai contesti originari delle fonti primarie da cui i suoi testi derivano, non è interessato cioè agli originari dispositivi di comunicazione (per il concetto, si veda Lamé 2015) che hanno veicolato i testi. Inoltre per il linguista il testo è dato, non è esso stesso oggetto di discussione ed è quindi un punto di partenza, non di arrivo. Le collezioni di testi che interessano il linguista provengono da edizioni singole, scelte con criteri assiologici. Infine, i molteplici livelli di analisi (morfosintattica, semantica, tematica, ecc.) applicati al testo prevedono una sola interpretazione e, nel caso in cui più annotatori o più sistemi automatici di analisi linguistica forniscano dati divergenti, il risultato finale viene armonizzato da un supervisore umano o da un sistema di valutazione dell'analisi più probabile.

Per il filologo, al contrario, questi tre punti di articolazione sono sostanziali per la sua disciplina: molteplicità di dispositivi di comunicazione che veicolano una molteplicità di varianti testuali, su cui insiste una molteplicità di interpretazioni distribuite su diversi livelli di analisi. Di conseguenza, le molteplicità che interessano al filologo non possono essere affrontate solo a livello di codifica dei testi, ma vanno affrontate anche a livello di strumenti di analisi e di infrastrutture telematiche.

5 Dall'interscambio all'interoperabilità

Le risorse digitali, e in particolare i corpora testuali, sono la materia prima della filologia cooperativa. È comprensibile quindi che l'attenzione degli umanisti digitali si sia concentrata soprattutto sull'acquisizione e la codifica delle risorse, dove si sono raggiunti da un lato alti livelli di accuratezza del testo tramite avanzati sistemi di OCR, e dall'altro elevati livelli

di standardizzazione della codifica grazie alla *Text Encoding Initiative*.

Tuttavia, la modellazione dei dati e la standardizzazione dei formati, pur necessaria all'interscambio delle risorse, non garantisce l'interoperabilità fra i processi che gestiscono i dati (Bauman 2011). Se si pensa ad esempio alla codifica di un apparato critico secondo le molteplici opzioni offerte da TEI (<http://www.tei-c.org>), si può notare quanto sia facile, grazie ad appositi fogli di stile, visualizzare le notizie d'apparato in modo simile a un'edizione critica a stampa, ma quanto sia invece difficile eseguire operazioni più complesse, come l'allineamento delle singole parole di una variante sulle parole del testo accolto dall'editore. Attualmente, al di là delle trasformazioni eseguibili tramite fogli di stile da un documento a un altro documento, la gestione dei dati codificati in TEI è per lo più affidata agli sviluppatori di applicazioni specifiche.

Per questo motivo la comunità delle *Digital Humanities* si sta orientando verso forme più consapevoli di interoperabilità, e il progetto europeo *Interedition* è uno dei testimoni più rappresentativi di questa tendenza (Haentjens Dekker et al. 2015, 452):

Interoperability is the key term within the framework of [...] Interedition, whose aim is 'to encourage the creators of tools for textual scholarship to make their functionality available to others, and to promote communication between scholars so that we can raise awareness of innovative working methods'. The tools developed by Interedition's 'Prototyping' working group were tested by other research teams, which formulate strategic recommendations.

6 Il coordinamento dei mezzi minimi per l'esercizio della filologia cooperativa

Come si è visto, la filologia cooperativa si proietta su uno scenario dove le crescenti risorse digitali sono interoperabili e dove gli strumenti computazionali complessi sono costruiti tramite l'assemblaggio di componenti semplici e la concatenazione di servizi web.

A livello delle risorse, come più volte rimarcato da Crane (si veda, ad esempio, Crane et al. 2009), il punto di partenza per la cooperazione è la condivisione di uno schema citazionale (messo a disposizione dai *Canonical Text Services*, CTS) in grado di identificare in modo inequivocabile le unità testuali della fonte primaria a qualsiasi livello di granularità (libro, capitolo, parola, singolo carattere). In questo modo, le molteplici annotazioni (non solo metriche, morfosintattiche, semantiche, ma anche storiche od estetiche, ad esempio) che insistono su una particolare versione del testo, sono totalmente disaccoppiate dal documento che rappresenta la fonte primaria e fanno ad esso riferimento solo attraverso lo schema citazionale,

che rende ragione anche della specifica versione del documento. Dato che ciò che viene citato è indipendente da ciò che lo cita, la quantità di informazioni associate e portate da terze parti può crescere in modo indefinito.

A livello degli strumenti, i servizi di analisi linguistica tramite *web services* stanno rivelando la loro efficacia, mentre ancora molto deve essere fatto per la creazione di librerie di componenti software con cui costruire i *web services* stessi, le *web*, *desktop*, e *mobile applications* senza dover reimplementare tutte le funzionalità con tecnologie sempre diverse per le diverse destinazioni.

Il paradosso che rallenta maggiormente lo sviluppo della filologia cooperativa è la scarsa comunicazione fra la comunità delle *Digital Humanities* e la comunità dei filologi tradizionali. I contatti non sistematici fra le due comunità sono la causa di analisi dei requisiti e specifiche parziali, legate a visioni basate sulle esigenze di singoli progetti anziché sulle ampie possibilità offerte dal dominio della filologia nel suo complesso.

7 Discussione di un caso d'uso

Non vi è ancora sufficiente interoperabilità fra gli strumenti creati per lo studio delle fonti primarie dirette (ad esempio, tramite collazioni automatiche), lo studio delle fonti indirette (ad esempio, tramite estrazione di citazioni e allusioni) e, infine, lo studio della storia della tradizione (ad esempio, tramite lo studio della fortuna di varianti antiche).

Il caso d'uso scelto prende in considerazione un ulteriore aspetto di cui l'editore deve tener conto, che sta emergendo anche grazie alla grande disponibilità di fonti secondarie digitalizzate (commentari, articoli, monografie, ecc.): la storia degli studi filologici.

Per esempio, al v. 135 dei *Persiani* di Eschilo la variante ἀκροπενθεῖς è attestata da tutti i codici tranne Y, che reca ἄβροπενθεῖς. È ovvio che, per una tradizione complessa come quella eschilea, l'apparato digitale non si può limitare alle sole informazioni derivanti dalla collazione di manoscritti. La variante ἄβροπενθεῖς è accolta dalla maggior parte degli editori dopo Paley non per il valore di Y, ma per le ragioni di ordine idiomatico, stilistico, paleografico e di evidenza di uno *scholium* (che contiene il verbo ἀβρύνεσθαι) addotte dallo stesso Paley, che non conosceva Y e che quindi propose ἄβροπενθεῖς per congettura. Di più: lo studio approfondito del *modus operandi* di Paley (Richardson 1957) ci informa che la sua congettura non nasce dall'evidenza estrinseca dello *scholium*, ma che in esso si rafforza semmai l'ipotesi già formulata grazie alle altre ragioni, intrinseche.

Un'edizione scientifica digitale deve coordinare i mezzi per codificare anche questo tipo di informazioni provenienti dalla letteratura secondaria (il mesotesto di J. Cummings, menzionato in Pierazzo 2015) e i mezzi per reperire e visualizzare queste informazioni.

8 Conclusione

In questo breve contributo si è cercato di evidenziare quali siano i principi della filologia cooperativa e come il concetto di interoperabilità non si possa limitare alla codifica delle risorse digitali, ma debba prendere in considerazione gli strumenti computazionali e le infrastrutture telematiche che insistono su tali risorse.

Bibliografia

- Bauman, Syd (2011). «Interchange vs. Interoperability». *Proceedings of Balisage = The Markup Conference 2011* (Montréal, 2nd-5th August 2011). Balisage Series on Markup Technologies 7. URL <https://www.balisage.net/Proceedings/vol7/cover.html> (2017-11-27).
- Boschetti, Federico; Del Grosso, Angelo Mario (2015). «TeiCoPhiLib. A Library of Components for the Domain of Collaborative Philology» [online]. *Journal of the Text Encoding Initiative*, 8. URL <https://jtei.revues.org/1285> (2017-10-17).
- Crane, Gregory (1991). «Generating and Parsing Classical Greek». *Literary and Linguistic Computing*, 6(4), 243-5.
- Crane, Gregory; Seales, Brent; Terras, Melissa (2009). «Cyberinfrastructure for Classical Philology» [online]. *Digital Humanities Quarterly*, 3(1). URL <http://www.digitalhumanities.org/dhq/vol/3/1/000023/000023.html> (2017-10-17).
- Haentjens Dekker, Ronald et al. (2015). «Computer-Supported Collation of Modern Manuscripts. Collatex and the Beckett Digital Manuscript Project». *Digital Scholarship in the Humanities*, 30(3), 452-70.
- Kozar, Olga (2010). «Towards Better Group Work. Seeing the Difference between Cooperation and Collaboration». *English Teaching Forum*, 48(2), 16-23.
- Lamé, Marion (2015). «Primary Sources of Information, Digitization Processes and Dispositive Analysis». Tomasi, Francesca; Rosselli Del Turco, Roberto; Tammara, Anna Maria (a cura di), *Proceedings of the Third AIUCD Annual = Conference on Humanities and their Methods in the Digital Ecosystem*. New York: ACM, 1-5.
- McEnery, Tony; Wilson, Andrew (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Pierazzo, Elena (2015). *Digital Scholarly Editing. Theories, Models and Methods*. Farnham (Surrey): Ashgate.
- Roncaglia, Gino (2010). *La quarta rivoluzione. Sei lezioni sul futuro del libro*. Roma-Bari: Laterza.
- Richardson, L.J.D. (1957). «Aeschylus Persae 135. ἀβροπενθεϊς Paley (codd. ἀκροπενθεϊς)». *Hermathena*, 89, 73-5.

