

Guido Milanese, *Filologia, letteratura, computer. Idee e strumenti per l'informatica umanistica*, Milano, Vita & Pensiero, 2020, pp. XI, 344.

Questo saggio presenta con grande efficacia e capacità espositiva l'impiego delle tecniche informatiche nel settore umanistico, con particolare riguardo all'analisi linguistica intesa come elemento principe della filologia. Per impiegare le parole di Tito Orlandi nella *Presentazione* (p. 3), «[q]uello che lo distingue da trattati per alcuni versi similari è il suo essere concepito come esposizione sistematica dei problemi per le persone colte interessate, ma anche e forse soprattutto come manuale ampio ed esauriente rivolto all'insegnamento e all'apprendimento universitario».

Il libro si compone di sette parti. La prima si pone il fine di inquadrare le problematiche di fondo dell'intero saggio: nel *Capitolo 1* "Orientamenti" (pp. 9-13) viene definito il campo di indagine consistente nel rapporto tra teoria e prassi nel settore denominato *informatica umanistica*, sottolineando la necessità di passare da un'impostazione ermeneutica ed epistemologica alla sua implementazione pratica e fattuale, senza la quale sarebbe impossibile ogni reale progresso scientifico. In tal senso l'impiego del computer può e, in un certo senso, deve ridefinire il lavoro del filologo alla luce degli apporti della nuova tecnologia. Inoltre, in fase introduttiva viene da subito dichiarata la preferenza dell'autore per il *free software*, argomento approfondito nei capitoli 4 e 5.

Il *Capitolo 2* "Filologi, letteratura, tecnologia. Uno sguardo storico" (pp. 15-48) costituisce una breve storia della trasmissione del sapere. Sin dall'antichità mesopotamica e vicino-orientale sono state redatte liste di parole, elemento recepito anche dal mondo greco e romano ove si svilupparono glossari di forme e dimensioni diverse, antenati dei nostri vocabolari. Si deve poi alla cerchia culturale della Biblioteca di Alessandria d'Egitto la produzione di schede bibliografiche, i cosiddetti

Πίνακες di Callimaco, riguardanti i testi posseduti nella prima grande biblioteca del mondo. Recepita a Roma, questa sistemazione organica e rigorosa del sapere condusse alla stesura di opere generali di argomento linguistico-scientifico-bibliografico (e.g. la *Naturalis historia* di Plinio il Vecchio, le *Noctes Atticae* di Aulo Gellio, i *Saturnalia* di Macrobio...) e di opere lessicografiche come il *De verborum significatu* di Verrio Flacco, la *Compendiosa doctrina* di Nonio Marcello o le *Elementarium doctrinae* di Papia. A questo punto dell'esposizione è introdotto un concetto fondamentale per comprendere non solo la storia della trasmissione della cultura ma anche il senso dell'opera recensita: il salto tecnologico (pp. 22-27) fa sì che l'introduzione di una nuova tecnologia porti ad un cambio radicale del «modo di operare, e quindi di pensare, in un certo campo» (p. 23). Un primo esempio, legato a quanto sinora esposto, è costituito dal passaggio dallo schedario bibliografico in formato cartaceo, ben diffuso ancora sino al principio del Terzo Millennio, al database relazionale online che consente di individuare le stesse informazioni presenti in altri schedari analoghi (pp. 22-29). Un altro esempio, sul quale l'autore si dilunga, è legato alle modalità di trasmissione dei testi (pp. 29-36): nel corso della storia dell'umanità, le prime forme di scrittura di carattere pittografico-ideografico, particolarmente complesse data l'enorme quantità di grafemi, sono sostituite dapprima da caratteri di natura sillabica, quindi dalla scrittura alfabetica, il cui successo 'globale' dipende dalla sua economicità in termini di apprendimento dei caratteri da parte sia dei lettori sia degli scrittori. Ma il passaggio da una forma di scrittura all'altra presuppone una riflessione ermeneutica mediante la quale si attua una corrispondenza precisa tra valore fonetico e rappresentazione grafica, alla base del salto tecnologico. Anche il supporto scrittoria non è indifferente alla trasmissione delle informazioni: se nell'antichità greco-romana il libro si presentava nella forma del rotolo papiraceo, il passaggio al codice ha consentito una migliore maneggevolezza tale da rendere più agile l'accesso ai contenuti; il passaggio poi dalla scrittura manuale alla stampa a caratteri mobili, inventata da Johannes Gutenberg, porta con sé il vantaggio di una tiratura potenzialmente illimitata unita alla riduzione notevole dei costi di

produzione dei libri stessi. Infine, lo sviluppo di nuove tecnologie dall'Ottocento in avanti rese possibile la collazione in forma più agile di un numero elevatissimo di testimoni di una certa opera: un esempio, particolarmente caro all'autore (pp. 36-41), è costituito dalla filologia musicale iniziata a Solesmes grazie alle riproduzioni fotografiche le quali consentirono ai monaci del cenobio francese, senza allontanarsi da esso, di confrontare un altissimo numero di manoscritti dotati di notazione musicale così da produrre testi liturgici più prossimi alle fonti medievali. L'avvento dei computer (pp. 42-48) poi permise l'automatizzazione di alcune operazioni meccaniche, in particolare la *recensio* dei testimoni, senza però dimenticare che «[i]l salto di qualità si avrà quando si comprenderà che il computer non interessa all'umanista per fare più agevolmente le stesse cose che si potevano fare in epoca pre-informatica, ma per ripensare la propria attività con nuovi paradigmi» (p. 43).

La *Parte II* intitolata "Durata e accessibilità dell'informazione" va a toccare una questione di capitale importanza per chi ha la necessità di trasmettere i contenuti della propria disciplina o del proprio lavoro alle generazioni successive. Il titolo "La pergamena e il floppy disk" del *Capitolo 3* (pp. 51-71) pone il serio problema dell'invecchiamento del supporto che assicura la conservazione delle informazioni in relazione alla definizione di salto tecnologico data in precedenza. Nel caso dei supporti elettronici l'obsolescenza è legata sia all'hardware, che nel corso della seconda metà del Novecento e nei primi decenni di questo secolo ha subito un'evoluzione rapidissima, sia al software, il quale, venendo continuamente sviluppato, spesso non risulta più accessibile nel passare da una versione più vecchia a una più recente.

Una prima risposta alle questioni appena enunciate viene dall'impiego del *free-open software* la cui 'paternità' può essere attribuita al matematico e informatico statunitense Richard Stallman (1953-) che lo definì con queste parole citate a p. 74: «Free software is a matter of the users' freedom to run, copy, distribute, study, change, and improve the software». La libertà cui si fa riferimento è dunque legata all'accessibilità del codice con cui il programma è scritto: tale accessibilità distingue

dunque software liberi da quelli proprietari. Nel *Capitolo 4* “Free-Open Software” (pp. 73-89), oltre a presentare la definizione appena richiamata, viene mostrato nella pratica in che modo questo genere di programmi venga distribuito, a cominciare dall’esempio di GNU-Linux, il sistema operativo che utilizza esclusivamente software libero. Proprio per la sua natura aperta, questo genere di programma risponde alla problematica di fondo legata all’obsolescenza dei documenti. Il successivo *Capitolo 5* “Cicerone, i monopoli e l’open source” (pp. 91-94) approfondisce la questione della proprietà intellettuale, strettamente connessa alla concezione che si ha del software (p. 94): «da una parte c’è chi considera l’informatica un tipo di produzione come qualunque altro, e che quindi considera un programma per computer come un qualunque altro prodotto [...]. Dall’altra parte c’è la concezione di coloro che considerano il software come un prodotto dell’ingegno umano paragonabile a una creazione artistica». Sebbene l’autore ritenga che queste due visioni dell’informatica possano essere conciliabili, tuttavia pensa che la seconda possibilità sia preferibile soprattutto nel campo della ricerca scientifica e universitaria in quanto «il software libero fornisce non solo gli strumenti necessari, ma anche di qualità in genere non inferiore, e spesso anche migliore, rispetto a quelli commerciali e (soprattutto) continuamente modificabili e adattabili alle proprie necessità» (p. 94).

Con la *Parte III* “Verso il testo informatico” l’argomentazione passa decisamente dalle questioni teoriche a quelle ‘pratiche’, cominciando dal problema della digitalizzazione del testo, operazione preliminare a un suo studio informatico. Il *Capitolo 6* (pp. 97-123), sin dal titolo, si sofferma sulla “Acquisizione del testo” alla cui base sta la digitalizzazione: tale operazione può avvenire o manualmente o attraverso la scansione consistente nella trasformazione di un documento analogico (in questo caso una pagina scritta) in un file grafico che riproduca l’immagine della pagina la cui fedeltà dipende dalla risoluzione legata al numero di pixel. La scansione è operazione preliminare al trasferimento su file testuale vero e proprio mediante il processo OCR (optical character recognition) garantito da una pluralità di programmi, tra i quali è espressamente citato Tesseract. Un’ulteriore possibilità di acquisizione

testuale prevede il ricorso a banche dati, che possono essere tanto ad accesso libero (“generaliste”: Gutenberg Project, Progetto Manuzio, Internet Archive, Google Books, HathiTrust; specializzate: Perseus Digital Library, Packhard Humanities Institute, *Musisque Deoque*, Digital library of late-antique Latin texts) quanto su abbonamento (Library of Latin Texts, *Thesaurus Linguae Graecae*, Literature Online). Una problematica legata alla distribuzione di testi sul web è però connessa con la qualità stessa dei testi: «se un testo di *TLG*, di *LION* o di *Perseus* è da considerarsi affidabile quanto un testo a stampa, lo stesso non si può dire per testi “casuali” collocati in rete senza un intervento professionale, in questi casi il controllo più affidabile resta quello puramente umano» (p. 117). La seconda fondamentale questione legata alla digitalizzazione del testo riguarda la codifica dei caratteri, quel «procedimento che associa un insieme di caratteri grafici a un insieme di altri oggetti» (p. 119): alla fine del Novecento è stato realizzato lo standard Unicode consistente in un’«immensa tabella capace di assegnare in modo univoco un indicatore numerico a ogni carattere di ogni lingua e alfabeto del mondo» la cui «possibilità teorica è di 1.114.112 posizioni (caratteri), dei quali sono attualmente utilizzati circa 137.000 posizioni» (p. 121).

Il *Capitolo 7* “La codifica del testo” (pp. 125-179) arriva quindi al cuore della riflessione consistente nella definizione di quale sia la differenza sostanziale tra testo computerizzato e informatizzato. Il primo altro non è che un testo puro, non diverso da quello che si trova su un foglio con l’unica differenza del supporto che lo veicola – elettronico e non cartaceo – mentre il secondo contiene una pluralità di informazioni connesse a un metalinguaggio che stabilisce la funzione delle diverse parti dello scritto. Nel caso di un testo informatizzato la marcatura è finalizzata a esplicitare l’organizzazione interna del testo fondata sul concetto di *frame*, ossia uno spazio di riferimento che consente di comprendere ciò che per natura propria può assumere molteplici significati (pp. 126-140). Molteplici sono le forme con le quali è possibile inserire delle marcature: tra le più importanti, sono dettagliatamente descritte l’XML (Extensible Markup Language) e l’HTML (Hypertext Markup

Language). Nonostante i codici XML e HTML siano completamente aperti, non mancano problematiche: in relazione al primo, sono notevoli sia la complessità intrinseca sia i vari standard – spesso incompatibili tra loro – che esso ha assunto nel corso degli anni; per quanto concerne il secondo, la coesistenza di semantica e grafica ha fatto sì che il web sia diventato oltremodo caotico, mettendo a rischio la durata dell’informazione. Esistono tuttavia altre forme di marcatura leggera che uniscono in sé immediatezza di lettura e durevolezza dei contenuti: il volume presenta in particolare Markdown, progettato nel 2004 da John Gruber (1973-), che mediante il programma Pandoc può essere convertito in quasi tutti i tipi di file oggi diffusi. Ne consegue che un testo informatico implica necessariamente la codifica attraverso la quale si rappresenta un modello che è la risposta a un’originaria domanda di senso.

L’implementazione pratica dell’analisi filologica e linguistica, che l’autore considera inseparabili come due facce di una stessa medaglia, è approfondita nella *Parte IV*. In particolare, il *Capitolo 8* “Analizzare il testo” (pp. 183-197) si sofferma sull’indagine informatizzata dei *corpora* testuali mediante il programma TextStat che consente di realizzare liste frequenziali di parole, concordanze e indici. Inoltre, TreeTagger permette di lemmatizzare tutte le forme di parola presenti in un determinato testo, fornendo un altro strumento di grande utilità per lo studio della lingua di un determinato autore.

Il successivo *Capitolo 9* “Critica del testo e filologia computazionale” (pp. 199-217) sposta il campo d’indagine verso la preparazione di un’edizione critica informatica, a cominciare dalla prima fase consistente nella *recensio*. Il computer è in grado di confrontare un numero molto alto di testimoni, mostrandone le varianti e gli accordi: i programmi diff e wdiff consentono di operare una collazione semplice solo tra due o tre testimoni, mentre i più potenti TUSTEP e Juxta riescono a gestire il confronto tra un numero molto più elevato; inoltre Juxta è in grado di costruire un elementare apparato critico che potrà essere successivamente perfezionato dallo studioso. Dunque la realizzazione di

un'edizione critica informatica non consiste semplicemente nella 'computerizzazione' di un'edizione critica mantenendo una forma analoga a quelle cartacee, ma nella produzione di un ipertesto nel quale ogni parola consente di accedere a tutta una serie di informazioni aggiuntive che l'editore riterrà opportune (varianti, passi paralleli, riproduzioni di manoscritti e fonti a stampa...): alcuni esempi sono *Musisque Deoque*, il *Corpus rhythmorum musicum* a cura di Francesco Stella o il *De nomine* di Orso Beneventano edito da Paolo Monella.

Nel *Capitolo 10* "Costruire i propri strumenti" (pp. 219-244) l'autore approfondisce ulteriormente gli strumenti di analisi testuale accessibili attraverso i comandi della 'shell bash', tra cui la ricerca di espressioni regolari mediante 'grep' (general regular expression print) o il conteggio delle parole con 'wc' (word count). Il passo successivo riguarda l'impiego di 'minilinguaggi' di UNIX, come 'sed' (stream editor) e 'awk' (dalle iniziali dei cognomi dei suoi autori Alfred Aho, Peter Weinberger e Brian Kernighan), che danno risultati ancora più potenti rispetto a quelli appena definiti; quindi viene introdotto l'uso di alcuni linguaggi di programmazione quali C, Perl, Python, Prolog e Snobol. Chiaramente si tratta di affrontare una complessità sempre crescente: imparare un linguaggio di programmazione consentirebbe al filologo di rispondere con maggiore agio a problematiche molto precise, creando degli strumenti molto potenti che risolvano una questione ben definita.

La *Parte V* "La scrittura scientifica" sposta infine la riflessione sulla redazione di testi finalizzati alla comunicazione del risultato delle proprie ricerche in ambito accademico. Anzitutto, nel *Capitolo 11* "Dal WP alla scrittura strutturata" (pp. 247-254) è presentata la sostanziale differenza esistente tra il normale approccio alla videoscrittura attraverso i programmi di word processing visivi (Word del pacchetto MicrosoftOffice, Writer di OpenOffice e di LibreOffice), nei quali l'attenzione dell'utente si concentra – quasi esclusivamente – sulla resa grafica tralasciando però la struttura del documento, e l'uso dei linguaggi di marcatura come XML o Markdown. A svantaggio dei primi va ricordato come la struttura e la definizione di tag siano aspetti di fondamentale importanza nella redazione di ogni documento; inoltre non si deve

dimenticare che, essendo continuamente sviluppati, essi potrebbero presentare problemi di stabilità e, in ultima analisi, di obsolescenza del software, con il conseguente rischio di condurre alla perdita dei dati nel giro di qualche decennio.

La risposta che l'autore propone alle criticità appena enunciate consiste nell'adozione di altri sistemi di tipografia digitale, esposti nel *Capitolo 12* "TeX e LaTeX" (pp. 255-267). TeX fu elaborato dal matematico statunitense Donald Knuth (1938-) per rispondere alla necessità di creare una forma di tipografia computerizzata spendibile soprattutto nell'ambito matematico. Le *macro* di LaTeX furono invece sviluppate negli anni '80 dallo scienziato americano Leslie Lamport (1941-): come nel *Capitolo 7*, qui l'attenzione si sposta decisamente dall'aspetto puramente grafico a quello strutturale mediante l'acquisizione di linguaggi di markup. La differenza sostanziale rispetto ai word processing visivi è data dalla compilazione asincrona e dalla memorizzazione di un certo numero di comandi, fatti che costituiscono i principali freni inibitori verso l'adozione di questo tipo di tipografia digitale. Una buona soluzione di compromesso può venire da LyX: senza apprendere alcun comando, l'utente ha la possibilità di concentrarsi sulla struttura semantica lasciando poi alla macchina la composizione grafica del documento nel miglior modo possibile.

L'ultimo aspetto di fondamentale portata per la ricerca scientifica è legato alla gestione delle fonti bibliografiche, di cui si tratta ampiamente nel *Capitolo 13* (pp. 269-285): una volta recepiti i dati provenienti dallo spoglio di cataloghi di biblioteche, indici e archivi, il problema maggiore si pone nei confronti della loro archiviazione. Se l'inserimento di tali dati in un semplice file testuale può sembrare la soluzione più ovvia e meno faticosa, in realtà ci si rende conto che, lievitando il numero delle referenze, più si complica la gestione del materiale: la soluzione che l'autore propone consiste nell'adozione di Jabref, programma che consente l'esportazione delle voci di bibliografia in formato BibTeX, integrabile sia con LaTeX sia con altri sistemi di word processing.

Dal momento che l'informatica umanistica non si esaurisce solo nell'analisi di testi, la *Parte VI* si dedica alle forme non testuali. Il *Capitolo 12* (pp. 289-300) affronta in sequenza: la codifica e l'analisi della musica su pentagramma o tetragramma; la digitalizzazione dei monumenti (reperti archeologici, manoscritti o papiri documentari); la glottodidattica. Conclude l'opera la *Parte VII* "Persone e argomenti" contenente gli elenchi delle persone, del lessico fondamentale, dei siti web a precedere indici e bibliografia.

Nel suo complesso l'opera si fa apprezzare la particolare chiarezza nella presentazione e la ricchezza dell'informazione, che rendono al tempo stesso il volume una lettura istruttiva e uno strumento di consultazione. Particolarmente apprezzabile è la completezza degli argomenti trattati, che danno all'opera un ampio respiro e la rendono adatta sia a chi da sempre ha nutrito interesse nei confronti dell'apporto del computer allo studio umanistico, particolarmente di ambito linguistico-filologico-letterario, sia a chi per la prima volta si trova ad affrontare l'informatica umanistica, soprattutto nell'ambito dei corsi di laurea. Trovo che sia davvero ben studiata l'impostazione globale del volume in cui, dopo ogni capitolo che espone progressivamente la teoria, vengono offerti ulteriori riferimenti e letture che consentano di approfondire gli argomenti via via trattati. Al di là dei singoli contributi in settori specialistici, l'opera ci suggerisce alla fine un ripensamento generale dei paradigmi fondamentali della ricerca umanistica e della complessità della scrittura scientifica.

Alessandro Re
Università degli Studi di Udine
Dipartimento di Lingue e Letterature,
Comunicazione, Formazione e Società
alessandro.re@uniud.it