

Recensione: Zaccarello, Michelangelo (cur). 2019. *Teoria e forme del testo digitale*. Roma: Carocci

Alessia Luvisotto

Università Ca' Foscari Venezia
alessialuvisotto.97@gmail.com

Abstract

Il volume *Teoria e forme del testo digitale* (Carocci, 2019) di Michelangelo Zaccarello raccoglie in traduzione dieci saggi degli autori più influenti sul tema della testualità digitale, a cui segue una postfazione di H. Wayne Storey. L'introduzione al volume di Zaccarello passa in rassegna le problematiche derivanti dall'avvento del nuovo mezzo digitale che saranno discusse nelle pagine successive; esse non si legano solamente al tema della rappresentazione e della codifica testuale, ma interessano anche il mondo della letteratura born digital e dei diritti d'autore, in un agile viaggio alla riscoperta dell'importanza della filologia.

Teoria e forme del testo digitale (Carocci, 2019) edited by Michelangelo Zaccarello collects a translation of ten essays by the most influential authors on the theme of digital textuality, followed by an afterword by H. Wayne Storey. The introduction to Zaccarello's book examines the problems arising from the advent of the new digital medium, which will be discussed in the following pages; they are not only linked to the theme of representation and text encoding, but also affect the world of born digital literature and copyright, in an agile journey to rediscover the importance of philology.

Obiettivi e contesto

Pubblicato per Carocci nel 2019, *Teoria e forme del testo digitale* di Michelangelo Zaccarello nasce dalla selezione ad opera dell'autore di dieci saggi tradotti da Greta Mazzaggio in cui risuonano le voci più esperte ed autorevoli in materia di *editorial theory* e *textual scholarship*.¹ Essi sono introdotti da un contributo inedito di Zaccarello che condensa le riflessioni derivanti

1 Nel licenziare questa mia recensione desidero ringraziare Marco Sartor e Irene Mamprin per i preziosi consigli e le proficue discussioni che hanno preceduto la stesura di questo contributo.

dall'incontro con le problematiche emerse dall'avvento del nuovo mezzo digitale, e sono conclusi da una postfazione a cura di Wayne Storey. La frequentazione della Society of Textual Scholarship e una certa dimestichezza con i temi propri dalla tradizione filologica anglo-americana dei testi a stampa conseguente la partecipazione, presso la Scuola Normale Superiore di Pisa, al seminario diretto da Alfredo Stussi, dedicatario del volume, sembrano aver portato Zaccarello ad una conclusione tutt'altro che pacifica. L'obiettivo del manuale, in questo senso, si spiega in relazione ad un'urgente necessità di dialogo sugli effetti derivanti dal passaggio del libro dalla forma cartacea a quella elettronica. Nello specifico, il volume ha l'ambizione di sopperire ad un'allarmante mancanza di studi relativi ad alcuni aspetti della testualità digitale che – sebbene meno noti – riguardano non solo la nostra esperienza di lettura ma anche la conservazione del patrimonio culturale mondiale e, di conseguenza, della nostra memoria. Nonostante si rivolga senz'altro ad un pubblico specialistico che abbia consolidati i temi propri della filologia e dei campi affini (ecdotica, bibliografia e codicologia), le riflessioni ivi contenute sembrano voler interpellare e scuotere le coscienze di qualsiasi lettore comune, interessato ad un qualsivoglia ambito del sapere, affinché si interroghi in termini di qualità ed esattezza sulla grande quantità di dati reperibili nel web. Tra le molteplici considerazioni proposte dai saggi è possibile individuarne alcune di carattere “problematico”, altre di “miglioramento”: se da una parte, infatti, ogni contributo sembra rivolto alla messa in luce degli aspetti più pericolosi ed inquietanti del nuovo mezzo digitale e delle potenze economiche che dietro ad esso si celano, dall'altra emerge una grande fiducia e si delinea uno squarcio sulle grandi potenzialità del mondo analogico che, se messe al vaglio della razionalità e sensibilità umane, potrebbero portare a grandi passi avanti nel progresso scientifico.

Articolazione e contenuto: problemi e assi di miglioramento

Marcatura

«La precisione richiesta dalla codifica del computer [...] ha posto le basi per la nostra esigenza di capire più a fondo cosa significa “testo”» (Eggert, p. 59). Questa breve asserzione mostra come l'avvento del computer abbia stimolato alcune domande sui principi fondanti delle discipline umanistiche e della critica testuale, mettendone in dubbio alcuni assiomi, tra cui quello di ritenere l'ultima volontà d'autore «criterio di scelta del testo da pubblicare» (McGann, p. 38). Indagare i caratteri del testo «liberi dalle limitazioni delle pagine rettangolari da rilegare insieme in un libro» (Hockey, p. 44) significa porsi nuove domande sulla sua natura e sul modo migliore per rappresentarla. Nonostante il grande entusiasmo nei confronti dell'ipertesto, sul piano pratico si rileva una mancanza di applicazioni che possano qualitativamente superare quella che è la classica riproduzione su carta: lo standard di marcatura del testo elaborato a partire dalla Text Encoding Initiative (TEI), infatti, parte dal presupposto che l'«e-text venga considerato un luogo logico piuttosto che materiale» ([1]) imponendo una concezione gerarchica e piramidale, sebbene i testi di area umanistica richiedano, al contrario, l'analisi di elementi sovrapposti. Benché la marcatura testuale si riveli un'operazione intrinsecamente interpretativa dell'editore, essa è necessaria affinché il testo possa essere interrogato ed utilizzato in modo produttivo: da qui

l'invito di Susan Hockey allo studio e alla pratica della TEI con lo scopo di «organizzarsi [...] in vista della futura disponibilità di strumenti migliori» (p. 51). La risposta di Paul Eggert a tale sfida è promettente e fiduciosa: intendendo l'edizione come una sorta di “cantiere testuale”, la tecnica dal nome di Just-In-Time-Markup (JITM) permette di arricchire una trascrizione di base (la forma meno opinabile possibile) attraverso l'applicazione di tag da parte di utenti esterni, marcatura che sarà sempre separata dalla trascrizione “grezza” per garantirne la correttezza. Questo approccio potrebbe incentivare un lavoro sui testi collaborativo e progressivo e costituire «la base strutturale di una nuova metodologia di comunicazione e pubblicazione accademica» (p. 69). Una proposta simile viene avanzata anche da Peter Shillingsburg ([2]) che, interrogandosi su quale sia il livello di marcatura che permetta l'impiego di un testo a più scopi, argomenta la necessità di raccogliere separatamente i vari tag aggiuntivi, dal momento che «usare *inline* marcature complesse di analisi e annotazione rende molto difficile la collaborazione» (p. 81) e si metterebbe a repentaglio la correttezza del testo ogniqualvolta esso fosse reso disponibile a ulteriore modifica o marcatura.

OCR e crowdsourcing

Se la natura interpretativa dell'operazione di markup può rivelarsi problematica, non meno pericolosa risulta la conversione di file contenenti immagini di pagine scannerizzate in testi attraverso l'uso di software di riconoscimento dei caratteri (Optical Character Recognition, OCR). In questo senso, i contributi di Diana Kichut e Paul Conway mirano a rendere il lettore consapevole della possibile scarsa qualità degli e-book digitalizzati conservati all'interno di archivi o biblioteche digitali. Le ricerche della prima studiosa si concentrano su una particolare consuetudine editoriale: la maggioranza dei testi digitalizzati derivanti da tecnologia OCR è pubblicata, nei differenti archivi, nella sua forma “grezza”, senza alcun tipo di correzione. La qualità dipende da un «gran numero di variabili, tra cui la risoluzione spaziale dell'immagine, il contrasto testo-sfondo, la [...] diversa forma dei caratteri» (p. 151). In condizioni ottimali l'OCR può raggiungere tassi di precisione del 97-99%: sebbene tale valore possa sembrare promettente, in realtà indica un range di 1-3 errori ogni 100 caratteri, mentre una trascrizione a quattro mani assicurerebbe un'accuratezza del 99,997% (o virtualmente senza errori). Data l'evidenza della necessità di una qualche forma di revisione, Kichut propone una nuova forma di cooperazione tra uomo e tecnologia suggerendo un rimedio valido ed economico alle debolezze dell'OCR: il *crowdsourcing*, un sistema che si giova della disponibilità dei volontari di Distributed Proofreader per la correzione di bozze. Ma non solo: l'idea è quella di un ambiente collaborativo in cui l'insieme dei file di testo “grezzi” possano divenire una «vasta area di collaudo digitale dove volontari o imprese commerciali possano avventurarsi per migliorare e rilanciare lavori imperfetti come futuri testi autorevoli» (p. 164). Conway, riprendendo uno studio di Stvilia basato sul modello di un grande progetto ancora in corso, ossia “Validating quality in large-scale digitization” ([3]), mette in luce, attraverso una serie di tabelle ed esemplificazioni grafiche, gli errori più comuni in Google Books e Internet Archive. Proponendo una gerarchia di tali errori dal più al meno grave, conclude come la loro incidenza e, conseguentemente, la quantità di “cattivi libri” in Google sia nettamente superiore rispetto ad Internet Archive. In questo frangente l'auspicio dell'autore è, da una parte, quello di giungere alla formulazione di algoritmi automatici

di correzione delle immagini; dall'altra, quello che il progetto si possa concludere con l'ideazione di uno «strumento efficiente di revisione di specifici volumi e la valutazione di quest'ultimi in termini di maggiore o minore presenza di errori di rilievo» (p. 194).

Le forme del testo digitale

Un tema che percorre in modo più o meno evidente l'intero volume è quello della condizione materiale del testo digitale (per cui cf. i contributi di Eggert e Kirschenbaum). Sebbene la digitalizzazione implichi un cambiamento del medium attraverso cui viene raffigurato il testo, è necessario distinguere tra materialità e fisicità per non cadere in quello che Kirschenbaum definisce «equivoco tattile» (p. 62). Contro l'argomento antirealista di Pichler, per cui «i testi non sono mai uguali a sé stessi, anche se l'illusione che lo siano è tradizionalmente alimentata dagli standard più o meno trasparenti di impaginazione» (p. 64), Eggert connette il concetto di *opera* a quello di *lavoro* che si svolge intorno ad esso e alla dialettica tra aspetto materiale/fisico e di significato. Lo studio di Matthew Kirschenbaum approfondisce il tema della letteratura born digital e del rapporto dei singoli autori con la videoscrittura, tecnica che Heim vede in profonda discontinuità «con la storia della resistenza materiale, questo tira a molla fra corpi doloranti e superfici bianche, fra strumenti e tracce scritte» (p. 89). In relazione ad una concezione “fisica” della scrittura cara a molti autori si spiega l'ampio impiego di WordStar,² che aderisce ad un «modello di composizione derivato dalla scrittura manuale» (p. 87) in parte dovuto alla disposizione dei comandi della tastiera. In generale, però, si pone il problema del nuovo approccio da adottare in relazione ad una letteratura che non si esprime più tramite manoscritti, ma si compone all'interno di programmi che non permettono, ad esempio, la visualizzazione di varianti d'autore. La conclusione invita all'approfondimento di una scienza che faccia i conti con «le caratteristiche dei diversi sistemi operativi, delle varie versioni dei software, degli apparati hardware, come anche delle relative tecnologie di stampa» (p. 94).

Copyright ed “edizioni sociali”

Uno dei pregi del volume di Zaccarello sta nell'affrontare non solo le tematiche proprie del settore informatico ed umanistico, ma anche questioni di ambito giuridico e legale che, sebbene possano sembrare distanti dalla riflessione sul testo, si rivelano centrali in relazione al suo trasferimento dal mezzo cartaceo a quello analogico. Spicca in questo frangente il contributo di Maurizio Borghi e Stavroula Karapapa, il cui oggetto d'indagine è il diritto d'autore in relazione

2 WordStar 4.0 è stato un programma rilasciato nel 1987, poco prima che Windows soppiantasse DOS come sistema operativo più usato da Microsoft. Particolarmente amato da molti romanzieri del genere fantasy o fantascienza, esso è ispirato «a ciò che oggi chiameremo una falla: Rob Barnaby aveva preso la finestra di comando (che permetteva al programmatore di correggere la codifica) e aveva esteso il display all'intera pagina di testo. Da tale soluzione, il passo era breve per raggiungere il modello WYSIWYG (What You See Is What You Get) [...]. Ciò che stava sullo schermo era quello che sarebbe stato sulla pagina da stampare. WordStar aveva già le dotazioni pionieristiche come la divisione automatica delle parole fra le righe, la giustificazione del testo e la visualizzazione delle interruzioni di pagine» (pp. 86-87).

alla “digitalizzazione di massa” intesa come «conversione in formato digitale su scala industriale di opere protette da copyright» (p. 95). La creazione di una sorta di “biblioteca universale” comprendente tutto lo scibile è idealmente l’ambizione di tale proposito, eppure le diverse operazioni ad esso collegate hanno uno statuto di liceità incerto, come dimostra il caso di Google Books, oggetto di cause legali in Europa e Stati Uniti sin dal 2005. La conversione del patrimonio librario al digitale si sviluppa secondo modalità diverse che prevedono primariamente la digitalizzazione di opere open access, quindi di quelle autorizzate dai titolari dei relativi diritti ed infine di testi coperti da diritti commerciali a gestione collettiva o rispondenti a specifici provvedimenti legislativi. Google, avendo concluso degli accordi con le più grandi biblioteche mondiali, ha ottenuto la licenza per scansionare interamente le loro collezioni, pur non avendo conseguito alcun tipo di autorizzazione da parte dei titolari dei diritti o dei soggetti commerciali ad essi delegati. Il punto critico si individua nello scontro tra libertà di accesso alle risorse del nostro patrimonio culturale – elemento auspicabile per garantirne la sopravvivenza – e la sua relativa limitazione a causa del diritto d’autore, sebbene esso sia «un riconoscimento legale del rapporto comunicativo fra autori e loro pubblico come espressione cruciale dell’aggregazione umana» (p. 105). Il saggio si interroga poi sui temi della conservazione, dell’accessibilità e dell’analisi computazionale della tradizione libraria nel mondo analogico, decretando l’inevitabile impossibilità di rispettare il copyright per ogni singolo testo, operazione che si rivelerebbe troppo dispendiosa economicamente o di fatto impraticabile. Il caso di Europeana, promosso dalla Commissione Europea nel 2006, ne è l’esempio: l’iniziativa procede lentamente, soprattutto per le opere recenti, proprio perché «la frammentazione e complessità del sistema attuale di licenze ostacola la digitalizzazione di gran parte del patrimonio culturale d’Europa» ([4]). La seconda conclusione riguarda l’allarmante possibilità di un monopolio sulla cultura esercitato da intermediari sempre più grandi e potenti: centralizzare il sapere in mani di pochi potrebbe promuovere un certo tipo di logiche non democratiche che minerebbero la libertà non solo d’accesso, ma anche di pensiero dell’utente.

La descrizione di uno scenario inquietante in questo contributo cede il passo, in quello successivo di Peter Robinson, ad una riflessione sul testo inteso nella sua dimensione collettiva, all’interno della quale spicca il nome di McGann e, specificatamente, del Rossetti Archive, ovvero sia un progetto condotto attraverso metodologie editoriali ispirate per l’appunto al “testo sociale”. Al suddetto lavoro si sarebbe ispirato Ray Siemens nell’ideazione di una “edizione sociale” prodotta in modo collaborativo e trasparente da una certa collettività di utenti. Si tratterebbe di un nuovo procedimento ecdotico che «privilegia una nuova rete di comunicazione scientifica, dove si rifuggono le strutture tradizionali, gerarchiche e sostenute dall’establishment accademico, per fondarsi invece sui valori generati dalla comunità» ([5]). La maggior critica mossa a tale ideazione si fonda sulla constatazione della mancanza dell’elemento che è proprio dell’edizione critica “tradizionale”, ossia l’autorevolezza del curatore da cui scaturiscono determinate decisioni che, secondo Robinson, non possono essere di “consenso popolare”, né derivare dall’«autorevolezza della collettività» (p. 131). In ogni caso, sebbene la retorica del saggio di Siemens del 2012 sulla natura e le modalità dell’«edizione sociale» sia senz’altro accattivante, il prodotto finale, concretizzatosi nella pubblicazione del Manoscritto del Devonshire nel 2011 su Wikibooks, non può essere definito tale, dal momento che tutte le fasi editoriali sono state condotte chiaramente

sotto stretta supervisione di Siemens. In sintesi, il filo rosso di tutti i saggi parrebbe essere proprio una chiamata alla collaborazione internazionale con il fine di migliorare sia l'accesso al nostro patrimonio librario, sia la tecnica ecdotica, in nome della creazione di una nuova comunità “digitale” interconnessa e consapevole.

Ritorno alla filologia

Non sorprende che dopo un simile *tour de force* Zaccarello concluda il suo volume inserendo un elaborato dal titolo eloquente, *Ritorno alla filologia. La memoria del passato nel contesto digitale* di McGann. Nel collegare i nuovi problemi dell'era digitale a quelli tradizionali della filologia e nel ricordo dei grandi nomi della classicità come Erodoto, Platone e Tolomeo sembra delinearsi un certo bisogno di stabilità con il fine di rallentare quell'orientamento sempre rivolto “in avanti” tipico delle scienze dure, per promuoverne uno “all'indietro”, orientato verso quello che è il tema principe di tutta l'indagine: la conservazione del nostro patrimonio culturale. In questo senso si spiega la necessità di un ritorno a quell'«infrastruttura di conoscenza plasmata dalla filologia [...] di un sapere rivolto a conservare il ricordo concreto dell'importanza della memoria» (pp. 205-6). Sintetica e incisiva, la postfazione di Hayne Storey individua nella mancanza di un «regolamento legale e materiale» (p. 121) per il mondo digitale – sulla falsa riga di quello stabilito dalla Repubblica di Venezia alle origini della stampa per ogni aspetto della nuova industria – il motivo dello scetticismo di molti nei confronti del nuovo mezzo. Il cambio di paradigma auspicato da Wayne Storey è chiaro: il pensare filologico non può rimanere ancorato al *print thinking*, ovvero ai meccanismi della stampa, ma dovrà rivolgersi ad un *digital thinking* che sarà, forse, in grado di proporre nuove modalità di presentazione del testo attraverso un lavoro interattivo ed interdisciplinare.

Conclusioni

L'umanista italofono (digitale o meno) che percorra le pagine di questo volume potrà apprezzare la scorrevolezza e la “narratività” della lettura, mai disturbata da un eccessivo uso di termini inglesi, riportati tra parentesi solamente se semanticamente rilevanti. Ciononostante, l'imbarazzo di fronte a tematiche del tutto trascurate dalla critica italiana e la conseguente mancanza di una terminologia adeguata alla materia è dichiarata fin da subito da Greta Mazzaggio nella sua *Nota alla traduzione*. Nel rivolgersi programmaticamente ad autori anglo-americani, spicca di certo la citazione di *Filologia dei testi a stampa* di Pasquale Stoppelli ([6]) a cui l'intero libro si ispira. Un *unicum* che sembra denunciare, implicitamente, un marcato ritardo di ricerche e studi nel nostro paese. Questo motivo parrebbe spiegare l'esclusione dal volume di una serie di argomenti di assoluta centralità per il filologo italiano, come una discussione sulle modalità di applicazione degli strumenti digitali a codici manoscritti antichi. Tuttavia, la ricchezza di temi proposti dai vari contributi pare incoraggiare lo studioso a volgere lo sguardo verso campi di ricerca ancora inesplorati. Per questo motivo *Teoria e forme del testo digitale* può essere considerato un ottimo punto di partenza per una nuova consapevolezza tutta italiana, affinché accolga gli spunti provenienti dall'estero e li approfondisca alla luce dei problemi specifici della nostra tradizione.

References

- [1] DeRose, Steven J., David G. Durand, Elli Mylonas, and Allen H. Renear. 1990. "What Is Text, Really?". *Journal of Computing in Higher Education* 1, no. 2: 3–26. <https://doi.org/10.1007/BF02941632>.
- [2] Shillingsburg, Peter L. 2018. *Textuality and Knowledge: Essays*. Pennsylvania: Penn State Univ. Press.
- [3] Conway, Paul. 2010. "Preservation in the Age of Google: Digitization, Digital Preservation, and Dilemmas". *Library Quarterly* 80: 61-79.
- [4] Comunicazione della Commissione Europea "A Digital Agenda for Europe". http://europa.eu/european-union/file/digital-agenda-europe_it.
- [5] Siemens, R., M. Timney, C. Leitch, C. Koolen, A. Garnett, and with the ETCL, INKE, and PKP Research Groups. 2012. "Toward Modeling the Social Edition: An Approach to Understanding the Electronic Scholarly Edition in the Context of New and Emerging Social Media." *Literary and Linguistic Computing* 27, no. 4 (December 1): 445–61. <https://doi.org/10.1093/lc/fqs013>.
- [6] Stoppelli, Pasquale. 1987. *Filologia dei testi a stampa*. Bologna: Il Mulino.