

Filologia ieri, oggi ... e domani

di LORENZO PERILLI, Roma

*scire debet quid petat ille, qui sagittam vult mittere,
et tunc dirigere ac moderari manu telum: errant
consilia nostra, quia non habent quo dirigantur.
Ignoranti, quem portum petat, nullus suus ventus
est. (Seneca, Ep. Lucil. 8,71,3)*

Nozioni liminari

Per poter discorrere senza eccessiva approssimazione di filologia *sub specie informaticae*, si dovrà anteporre qualche definizione. Per *filologia*, si intende la filologia classica, con la quale le altre, seppure talora di malavoglia, sono in debito – non foss’altro che per ragioni cronologiche, essendo nata ad Alessandria nel III secolo a. C., ma poi anche per fondamentali ragioni di metodo; per *critica testuale*, ovvero *ecdótica*, si intende quello snodo centrale della filologia che ha per oggetto la *edizione critica* dei testi, vale a dire la ricostruzione, sulla base dei materiali a noi pervenuti per il tramite di manoscritti, papiri e citazioni da parte di altri autori, di un testo il più vicino possibile all’originale; per *informatica*, si intende la disciplina del trattamento delle informazioni, che può proficuamente fare a meno della sua invadente propaggine operativa, il calcolatore, quando deve definire i suoi metodi, e recuperarne a posteriori le enormi, ma vincolanti, potenzialità; per *tecniche* di indagine si intende l’applicazione del *computer*, e non della informatica, ai dati disponibili, in opposizione alla riflessione *metodologica e disciplinare*, che consiste nel prendere consapevolezza dell’oggetto d’indagine, degli obiettivi, dei principi, degli strumenti. Infine, si vorrà considerare che la eventuale *interazione* tra le due prospettive deve attuarsi considerando la loro *disomogeneità* concettuale, metodica, sostanziale. Per dire insomma, che sono due cose costitutivamente diverse, forzate a dialogare e reciprocamente supportarsi (o sopportarsi?). Con risultati di volta in volta diversissimi tra loro.

Da principio

Gli studi classici, con sorpresa dei non addetti ai lavori, non solo hanno rappresentato un campo particolarmente fertile per le nuove tecnologie, ma hanno svolto il ruolo di battistrada. Vi è stato profuso grande impegno, entusiasmo, capacità.

Gli studi sulle applicazioni dei calcolatori „elettrici“ ai testi greci risalgono già agli anni '50. Tra tutti i campi di studio, la prima raccolta di testi destinata a studiosi è stata il *Thesaurus Linguae Graecae* realizzato in California, contenente i testi della letteratura greca dalle origini al 1453. Il progetto prese le mosse il primo di luglio del 1972, iniziò a diffondere i suoi dati nel 1977, ed è stato poi reso più ampiamente accessibile agli studiosi, senza smettere di aggiornarsi: ha avuto un ruolo pilota. In quegli stessi anni, David W. Packard, figlio del fondatore della Hewlett-Packard e una delle figure di spicco dell'informatica applicata agli studi classici, arrivava a sviluppare, dopo un approfondimento tecnico e teorico durato anni, un'intera stazione di lavoro destinata alle esigenze degli antichisti e all'uso del TLG, chiamata *Ibycus*, del tutto nuova e autonoma, dal sistema operativo al linguaggio di programmazione. Alcuni esemplari sono tuttora in uso, ed è questo un caso raro quanto significativo. Lo stesso si può dire per l'oculatezza di chi in quegli anni decise di adottare il sistema UNIX per gli studi classici, e TUSTEP, il complesso sistema modulare tedesco, di recente adottato anche dal team dell'Accademia delle Scienze di Berlino per il Corpus dei testi medici greci e latini.

La memorizzazione dei testi greci fu dunque precoce, ma poneva problemi di enorme complessità per quell'epoca, quali difficilmente i più giovani potrebbero immaginare. Non è possibile qui entrare nei dettagli, ma almeno un aspetto, che si sarebbe rivelato decisivo e che solo oggi comincia a trovare soluzione, va sottolineato: quello della codifica. Come memorizzare in formato digitale i testi greci, usando il codice ASCII dei calcolatori? Concepito per l'inglese, esso concedeva all'epoca meno di 100 segni, comprendenti lettere maiuscole, minuscole, numeri e segni di interpunzione, e non prevedeva lettere dotate di accento né di segni diacritici, di cui è invece ricca la lingua greca. E come memorizzare fenomeni complessi come una lacuna, una rasura, una integrazione congetturale, una espunzione, i simboli musicali, editoriali, astrologici, le diverse dimensioni dei caratteri, per non parlare della miriade di indicazioni necessarie per un testo papiraceo? Si ricordi, anche, che i caratteri greci non erano visualizzabili sui monitor normalmente disponibili, ma solo su costosissimi terminali di cui quasi nessuno aveva disponibilità (un'eccezione era Harvard). La soluzione escogitata da Packard (già professore di greco) e adottata per il TLG fu quella di elaborare un nuovo codice, detto *Beta Code*, che ricorrendo a combinazioni particolari dei caratteri ASCII riusciva a rappresentare centinaia di fenomeni diversi. Un evento di grande rilievo, che anticipava esigenze, e in parte soluzioni, che solo negli ultimi anni sono state affrontate in modo sistematico, in particolare con la *Text Encoding Initiative* (TEI) e le sue ricadute, i linguaggi per la descrizione e rappresentazione dei testi (markup).

Una cosa Packard aveva saputo chiarire con lungimirante anticipo, un elemento chiave per ogni seria interazione dell'informatica con gli studi classici: il fatto che mentre i sistemi operativi e il software potevano avere vita breve, determinante restava il modo in cui i dati, vale a dire i testi, venivano strutturati e digitalizzati. Questo è un principio che per lungo tempo non fu compreso appieno, e che tuttora non è apprezzato adeguatamente da molti „mestieranti“.

Linee di tendenza

Ad oggi, si possono distinguere almeno tre principali linee d'azione:

1. di grande difficoltà metodologica, la critica del testo;
2. raccolte di materiali (banche e basi di dati), a volte meditate, spesso approssimative, e in tempi recenti frequentemente dettate da esigenze commerciali;
3. strumenti di analisi del testo, dalla ricerca di „parole“ (*patterns*) a Indici e Concordanze, analisi linguistiche, stilistiche, statistiche.

Circa il primo campo d'azione, quello dell'ecdotica, la riflessione circa le possibilità di introdurre procedure automatizzate risale addirittura agli anni '60 del Novecento, e vi si intrecciarono contributi di antichisti come di medievisti e di filologi romani o moderni, non sempre contribuendo a maggiore chiarezza, giacché le esigenze e le prospettive hanno rivelato divergenze difficilmente componibili. Ne ripareremo. Il secondo campo è in stretta correlazione con il terzo, che ne rappresenta per così dire una implementazione. Conviene considerare dunque inizialmente le attività di raccolta di materiali.

Disponiamo di repertori di ogni genere. Quasi tutto ciò che poteva essere trasferito in digitale in tempi ragionevoli vi è stato trasferito. Tutti i testi greci pervenuti su manoscritti e la gran parte dei testi del latino classico, la patrologia latina, la letteratura cristiana antica, le iscrizioni, i papiri documentari, le bibliografie degli studi classici, enciclopedie dell'antichità classica, il *Thesaurus Linguae Latinae* (si tratta in questo caso non di testi, come per il TLG, ma dell'omonimo lessico, da oltre un secolo in corso di realizzazione presso l'Accademia Bavarese delle Scienze); è stata infranta infine la barriera del dizionario di riferimento, il *Greek-English Lexicon* di Liddell e Scott, del quale l'editore, la Oxford University Press, era sempre parsa restia a perdere il controllo diretto. Non è possibile elencare tutti i progetti: *online* (e alla fine di questo contributo) sono reperibili risorse che forniscono indicazioni più precise.

Banche e basi di dati sono un punto di partenza indispensabile. Oggetto del nostro lavoro sono i testi, e oggetto del nostro lavoro in modalità „digitale“ sono i testi in forma digitale. Fin qui, nessun dubbio. Il problema è: come dovremmo utilizzare queste raccolte? E come devono essere strutturate, per dare un contributo qualitativo, e non solo quantitativo, agli studi? Un criterio fondante dell'informatica prevedeva che essa si basasse, per il piano operativo, su una interazione tra dati e algoritmi; più precisamente, tra la struttura dei dati, e gli algoritmi che ad essi si prevede di applicare. E voleva, tale criterio, che non fosse corretto assumere decisioni sulla struttura dei dati, senza conoscere gli algoritmi che su di essi dovranno operare, né, viceversa, ignorare la struttura dei dati quando si elabori un algoritmo. Significa, con un esempio, che al momento di decidere i criteri per codificare un testo e trasferirlo in forma digitale, è necessario aver compreso almeno: *a)* quali sono gli elementi che costituiscono quel testo; *b)* quali elementi vanno ritenuti portatori di significato; *c)* se e come tali elementi vanno resi „riconoscibili“ nel testo nella sua forma digitale; *d)* a quale uso, o almeno a quale uso principale, si prevede che quei materiali siano destinati (il che equivale al rapporto tra dati e algoritmi).

L'uso comune alle raccolte di testi in formato digitale è invece da un lato che tali testi sono la riproduzione speculare di ciò che appare sulla pagina stampata, della quale anzi sono spesso una rappresentazione inadeguata; dall'altro che di essi si fa quello stesso uso che era consentito dalle edizioni a stampa: ricerca dei cosiddetti *patterns*, cioè di sequenze di caratteri, secondo criteri più o meno elaborati, o analisi linguistiche e stilistiche, su base quantitativa, nonché realizzazione di indici e concordanze. Infine, in molti casi, con l'eccezione del TLG, non esiste uno standard, ma sistemi diversi e spesso di tipo proprietario (come per i testi latini). Così stando le cose, non meriterebbe occuparsi della questione. Quanto alla ricerca di parole, indici, concordanze e soprattutto lessici tradizionali svolgevano egregiamente il proprio compito, che peraltro nel caso dei lessici resta notevolmente superiore a quello consentito da qualsiasi strumento „elettronico“. Studi stilometrici, ad esempio sulla lingua di Platone, risalgono alla fine dell' 800, non c'era bisogno di aspettare l'informatica; progressi nei risultati, non a caso, sembrano assai limitati.

Un esempio, per intenderci. Nel saggio di qualche anno fa di un noto esperto, da analisi statistiche di testi latini risulta che nei prosatori i verbi subordinati ricorrono con frequenza maggiore rispetto ai poeti. Spicca in particolare il fatto che Lucrezio risulti classificato tra i prosatori, piuttosto che tra i poeti. Resta da chiedersi: non bastava leggere Lucrezio per saperlo? Chiunque abbia sufficiente familiarità con il *De rerum natura* ne constata i peculiari caratteri stilistici; e chi non abbia quella familiarità, può scegliersi un altro mestiere. Né ci

voleva un calcolatore per dirci che i poeti fanno meno uso di subordinate rispetto ai prosatori.

Ogni tanto c'è poi chi si avventura ad annunciare, accade in Italia, ma anche in Francia, di aver realizzato il *Lessico* di un qualche autore, magari impegnativo, così dimostrando di ignorare persino che cosa sia un lessico e quale sia la differenza che lo separa da un indice e da una concordanza – strumenti che la filologia classica fin dall'antichità ha affinato in maniera straordinaria –, spacciando per tale quello che è per l'appunto soltanto un indice lemmatizzato, tanto più superfluo da un punto di vista informatico in quanto i software esistenti consentono a chiunque lo voglia di realizzarsi in pochi minuti il proprio, e in alcuni casi anche di lemmatizzarlo in buona parte, cioè di ricondurre ogni forma flessa alla forma di base, quella registrata come lemma nei dizionari.

Certo nessuno, tanto meno chi scrive (che regolarmente se ne serve), negherà né la comodità, né la velocità e la esaustività con cui gli strumenti elettronici permettono di realizzare ricerche di questo tipo: si fanno meglio e più rapidamente, su testi di autori differenti allo stesso tempo, combinando parole diverse anche distanti l'una dall'altra, ma certo ciò non richiede che si stia qui a scriverne, basterebbe limitarsi a constatare il dato di fatto. È il *grado zero* dell'applicazione dell'informatica alla filologia classica. Troppo spesso ancora accade di leggere di progetti spesso ambiziosi, magari „rivoluzionari“, di assistere a brillanti presentazioni di nuove applicazioni o risorse, finendo col chiedersi: ma che cosa ce ne dovremmo fare? Troppo spesso, siamo ancora di fronte a soluzioni in cerca di problemi.

Talora Ingannevoli Aspettative

Quello che invece ci si aspetta, è andare oltre la *tecnicizzazione e il mero perfezionamento esecutivo di attività tradizionali*, per ottenere risultati che non sarebbe stato possibile raggiungere altrimenti. Impostare diversamente i problemi, ma innanzitutto riconsiderare la propria disciplina e i suoi tratti fondamentali alla luce del confronto con i criteri, le peculiarità e le esigenze della informatica come scienza del trattamento delle informazioni, che in quanto tale può e inizialmente dovrebbe prescindere dal collegamento con la sua protesi meccanica.

Fin d'ora si può sottolineare un dato. Una banca dati come quella del TLG mette a disposizione di chiunque, in rete o nello spazio di un disco di pochi centimetri di diametro, l'intera letteratura greca, ivi compresi testi che sarebbe difficile reperire anche nelle biblioteche più fornite. Nel corso di un lavoro, è

possibile eseguire in pochi istanti controlli e verifiche che un tempo avrebbero richiesto lunghe ricerche. Tuttavia, di questo affascinante mondo digitale l'utente perde rapidamente di vista gli inganni. Si legge spesso, che „il tale termine ricorre x volte nella letteratura greca“, basandosi sul TLG: ignorando che molti testi – tutti quelli di autori tramandati in frammenti di tradizione indiretta – sono memorizzati due o anche più volte nella banca dati, sotto il nome dell'autore (poniamo, Democrito) e sotto il nome del testimone che ne tramanda il frammento (poniamo, Stobeo). Indagini, dunque, i cui risultati sono inattendibili. Ancora: il Codice Beta utilizzato nel TLG è estremamente vasto, permette la codifica di ogni fenomeno rilevante presente nel testo a stampa; tuttavia, non solo il testo a stampa presenta caratteristiche incompatibili con i principi dell'informatica – e quindi il criterio del testo digitale come *mirror image* di quello stampato comporta di per sé difficoltà di vario genere –, ma, per sovrammercato, i software esistenti per l'uso di quei testi semplificano drasticamente, ignorando gran parte dei codici previsti dal TLG (generalmente all'insaputa dell'utente). Tale semplificazione si deve a un puro criterio di economia nel lavoro del programmatore, che altrimenti dovrebbe esplicitare tutte le centinaia di combinazioni previste, e ciò renderebbe il suo lavoro più arduo, anche considerato che almeno la metà delle applicazioni più diffuse nascono per iniziativa di singoli studiosi o di piccoli gruppi. (Ma quelle realizzate da grandi editori, e vendute a cifre sproporzionate, non sono migliori – tutt'altro.)

Andiamo oltre. Una volta memorizzato, il testo appare all'utente, contro ogni aspettativa, non come un mondo „aperto“, ma come qualcosa di definitivo e immutabile. Si smarrisce la dimensione non più che propositiva di ogni edizione, alla quale non a caso si accompagna un apparato detto critico, con materiali che non sono né un abbellimento né degli *addenda*, ma parte integrante del testo stesso. Rappresentano la dimensione „virtuale“, che è essenziale a ogni testo classico. Non sapere se un termine si deve a congettura o se invece è probabile che sia una interpolazione, inficia alla base qualsiasi indagine lessicologica. Noi sappiamo che la lezione accolta nel testo non è altro che il risultato della scelta dell'editore, e che questa spesso ha la stessa autorità della variante relegata in apparato: ma nessuna banca dati ha mai memorizzato le varianti ai testi antichi, sia pur solo quelle riportate negli apparati critici. E chi si impone, ormai, dopo aver consultato un testo digitale, di recarsi in una biblioteca a verificare l'apparato dell'edizione a stampa? Il testo critico è pura ipotesi di lavoro: in cui il testo, come luogo del reale, trova compimento e senso solo grazie al rapporto con la virtualità degli apparati. Spogliato di questa dimensione, il testo, isolato, diventa un fossile, dando una sensazione di fissità che contraddice ciò che esso in realtà è.

Di nuovo un esempio. Testo édito per la prima volta solo di recente, il *De propriis placitis* di Galeno viene ricostruito dall'editore, V. Nutton, con una mescolanza di brani in greco, latino, ebraico. A un certo punto (58,2), si legge *dixit Pictagoras*. In apparato si apprende però che la lezione è erronea: l'originale recava il nome di Protagora. Per questioni connesse alla tradizione, l'editore preferisce non correggere. Una volta memorizzato in formato digitale, il testo recherà dunque il solo nome di Pitagora, e inutilmente si cercherà la presenza di quel Protagora, che pure Galeno voleva. Le ricerche automatiche daranno un risultato fuorviante, per chi rinunci allo scrupolo (tutto filologico) di una verifica eseguita caso per caso.

E chi mai si prende la briga di verificare, mediante consultazione dell'apposito *file* o del volume a stampa, di quale edizione di un certo autore il TLG abbia memorizzato il testo? Giacché non si può ignorare che se l'edizione di partenza è inattendibile – e ce ne sono numerose, soprattutto tra quelle ottocentesche –, inattendibili sono anche i risultati della ricerca. E tuttavia, l'esperienza insegna che queste situazioni si ripetono sistematicamente, che tutto è messo sullo stesso piano.

Trasposizioni digitali

Elaborazioni e risultati sofisticati richiedono codifiche sofisticate. Di un testo, e del suo apparato, va codificata in modo esplicito non solo la superficie espressiva, ma anche la *conoscenza inespressa*, o *implicita*. Tutto quello, cioè, che il lettore, e soprattutto il lettore specialista, *intende* nel testo, nei segnali che esso reca, senza che debba essere esplicitato. Perché i dati raccolti diventino davvero significativi e passibili di ulteriore trattamento, perché una banca dati risulti davvero funzionale, è necessario che siano adeguatamente compresi, esplicitati e implementati i problemi di formalizzazione dei dati. La questione si fa complessa quando si cerchi ad esempio di dare del concetto di variante, e di variante significativa, una definizione univoca e universale, tale da prestare il concetto stesso alla sua indispensabile algoritmizzazione: che è come dire, che mentre un algoritmo richiede non ambiguità, non ridondanza, finitezza, invarianza per esecuzione, nella filologia troviamo l'esatto contrario – concetti non univoci, conoscenza implicita difficile o impossibile da esplicitare in un modo che sia valido sempre e in ogni caso (come si rappresenterà in modo univoco una lettera dubbia, se è dubbia?). È singolare che il primo riferimento della nozione di *algoritmo* alla filologia risalga a inizio Novecento (cf. G. Pasquali, *Filologia e storia*, Firenze 1920, 52). Il termine non aveva, va da sé, il significato oggi familiare, ma indicava il ricorso a un linguaggio „cifrato“, fatto di abbreviazioni, simboli, notazioni, quali caratterizzano qualsiasi studio scien-

tifico, e in particolar modo gli apparati che accompagnano le edizioni dei testi, fornendo quegli elementi che non trovano posto all'interno del testo stesso – sigle di manoscritti, abbreviazioni di titoli di riviste, e quant'altro. Indicava cioè proprio quel tratto che si è definito della *conoscenza implicita*.

È, questo, l'aspetto forse più difficile del rapporto con i nuovi metodi e le nuove tecnologie, l'elemento che più ne marca la distanza e che ricorre continuamente. Un esempio, il più elementare: se incontra delle parentesi quadre all'interno dell'edizione di un testo greco, il lettore esperto sa che esse indicano che la parola seclusa deve essere espunta. Questo, però, vale per una edizione moderna. Se si ha sott'occhio un'edizione ottocentesca, la parentesi quadra può significare l'esatto contrario: integrazione congetturale, o anche lezione di dubbia autenticità. Se il testo è un lessico o si tratta di scoli, la parentesi quadra alla fine di una parola indica che la parola che precede è un lemma. Se si tratta di un testo papiraceo, la parentesi quadra può essere isolata o in coppia, indicare vari tipi di lacune di tipo meccanico (difetto del supporto materiale), ovvero congettura che integra una lacuna. Tutto questo, i lettori di un testo classico lo sanno; o comunque sta a loro preoccuparsi di acquisire le competenze necessarie ad intendere. Il testo digitale dovrebbe, per essere corretto e utile, esplicitare di volta in volta il significato di quella parentesi, e non limitarsi a rappresentarla in quanto tale. Solo così si potrà, ad esempio, individuare automaticamente i termini di un lessico che fungono da lemma; o le integrazioni congetturali presenti in un testo; e via elencando.

Chi voglia passare a un testo in formato digitale per un trattamento non banale, deve essere in grado di *intendere* il proprio oggetto, in tutti gli aspetti significativi, e deve quindi disporre di un adeguato sistema per rappresentarlo. Infine, ed è essenziale, non può accontentarsi di utilizzare i testi digitali disponibili, che sono sempre a un livello di codifica primordiale, ma deve ripartire daccapo e predisporre una propria codifica adeguata agli scopi che si prefigge, operazione estremamente dispendiosa in termini di tempo e impegno. Infine, anzi preliminarmente, costui dovrà porsi la domanda fondamentale: se davvero ne valga la pena. Se cioè i possibili risultati giustifichino tanto lavoro, poiché anche l'economia del rapporto costi benefici deve essere tenuta in considerazione, soprattutto nel caso del singolo studioso.

Edizioni (a)critiche?

L'edizione critica è non solo la fase culminante del lavoro del filologo, ma anche la più complessa. Anche in questo campo, si hanno riflessi evidenti di quanto si è detto finora. Chi scrive trova in una osservazione di un filologo

acuto come Sebastiano Timpanaro un condivisibile giudizio: egli osservava, nel suo libro sul metodo del Lachmann (48n.), che il ricorso al metodo del Quentin, da parte dei teorici dell'automazione, è una „triste necessità“. Dom Henri Quentin scriveva la sua opera principale nel 1926, e riteneva praticabile, e auspicabile, l'esame comparato dei manoscritti di un'opera e insomma la costituzione del testo procedendo *sine interpretatione*. Musica, evidentemente, per chi negli anni '60 si riprometteva di studiare *La critique des textes et son automatisaton*, o per chi ancora alla soglia di questo nuovo millennio bucinava (ignoro se accada tuttora) di „rinascite“ o persino „resurrezioni“ del metodo del Lachmann, sotto l'egida della computazione. Una strada che non conduce da nessuna parte, come ormai dovrebbe essere chiaro: per difetto di riflessione. Non sarà stato un caso, se i metodi del Quentin furono giudicati già da Pasquali „unilaterali e primitivi“, definiti da Alphonse Dain nel 1949, con allusione cinematografica, „la grande illusione“, dichiarati „sterili“ da Timpanaro. La „triste necessità“ di ricorrervi si deve al fatto che solo tali metodi sono apparsi adatti ai nuovi approcci, e quindi sono stati comunque adottati – in mancanza d'altro.

Piuttosto che riflettere sul se, oltre che sul come i nuovi metodi e i nuovi strumenti potessero essere posti in relazione con la critica del testo, si è andati alla ricerca di quelle teorie che si presentassero come più adatte a una traduzione in pratiche operative „automatiche“. Se l'interpretazione è soggettiva, si è pensato che rinunciarvi avrebbe dato oggettività. L'errore di prospettiva è evidente. Fu dovuto a un problema ben preciso: la costituzione di un testo (sia ribadito: almeno di un testo dell'antichità classica) si sottrae *per principio* alla automazione del processo, poiché tutto ciò che più conta in ogni decisione è demandato non a una valutazione *quantitativa* o *comparativa*, ma al *giudizio* dell'editore. Già uno dei padri della moderna computazione, Alan Turing (in *Computing machinery and intelligence*, 1950), chiariva un presupposto fondamentale: perché la procedura sia corretta, e lo siano i risultati, ci sono domande che a una macchina non si possono porre. La macchina, per quanto sofisticata, resta servile per definizione, ed è in grado di affrontare efficacemente solo problemi formalizzati e posti in termini adeguati. E il problema della costituzione del testo a partire da uno o più manoscritti, legato com'è a valutazioni occasionali e alla mutevolezza della interpretazione, non si lascia formalizzare in termini correttamente informatici. Saper rinunciare è importante quanto cercare nuove soluzioni.

Altro è utilizzare i nuovi strumenti per fornire una serie di supporti, questi sì efficacissimi, al lavoro dell'editore. Collazione automatica dei testimoni in ogni fase del lavoro, e confronto tra i materiali di tradizione diretta e indiretta, banche dati, immagini digitali dei manoscritti che supplementino il solo messaggio „testuale“ o meglio „verbale“, di per sé insufficiente, indici, analisi degli er-

ori e delle divergenze dei codici per tipologie o per collocazione geografica, e così via. Tutto quello che può essere di *supporto alla decisione*, e su cui non è qui il caso di dilungarsi; ogni procedimento passibile di essere modellato in termini matematici, può essere riprodotto con la massima efficacia da una macchina.

Dalla critica al Do-It-Yourself

Non entro nel dettaglio di ciò che riguarda le cosiddette „edizioni elettroniche“, che si vorrebbero destinate a soppiantare la tradizionale edizione critica. La definizione più suggestiva le chiama edizioni fluide, cooperative, distribuite. Per un'altra nozione di recente diffusione si usa l'acronimo DIY. Sta per Do-It-Yourself, si riferisce ai commenti di impostazione ipertestuale, ma esprime con efficacia il principio sotteso a questo tipo di iniziative. L'idea è che sia opportuno rinunciare ai tentativi di ricostruire uno e un solo testo per ciascun autore, il vagheggiato „originale“, sostituendo a questo una raccolta di materiali. Non si offre il testo ricostruito da un editore, ma un testo base, accompagnato da trascrizioni di manoscritti, collazioni, immagini, stemma dei codici ottenuto automaticamente, dati statistici, annotazioni, traduzioni, *et cetera multa*. A quel punto, ogni lettore potrà diventare editore. Non c'è più una struttura gerarchica, in cui l'editore offre il proprio testo al lettore, ma una comunità di editori/lettori che lavora sugli stessi materiali, aggiungendo del proprio – le proprie considerazioni, congetture, ipotesi, annotazioni. Edizione, come si diceva, cooperativa, distribuita. DIY.

Non è chi non veda la suggestione di un approccio del genere. Se riferita all'antichità classica (la nozione è nata in riferimento a testi medioevali e posteriori), l'errore consiste in: *a)* credere che questo tipo di lavoro *sostituisca* una edizione critica, piuttosto che *integrarla*; *b)* ritenere che il lettore sia in grado, e voglia, farsi editore: non è così, il lettore chiede all'editore non solo i materiali, ma la sua esperienza e competenza, e non una falsa democratizzazione che è in realtà livellamento verso il basso, e puro demandare ad altri il lavoro più impegnativo, quella della scelta e della decisione. È privatizzazione, non democratizzazione della conoscenza; *c)* credere che l'edizione tradizionale sia un testo statico, immutabile, rispetto alla fluidità che si pretende di introdurre: non è così, statico è il supporto che veicola il testo (la pagina stampata), non il testo veicolato, che invece tutti gli studiosi „che abbiano riflettuto un poco sul proprio mestiere“ (Pasquali) hanno sempre giudicato aperto, provvisorio, mutevole – basterebbe leggerli; *d)* fraintendere il ruolo dell'editore: egli si trasforma in impiegato di un catasto inesistente, il suo lavoro in quello di una società di servizi; *e)* quanto all'entusiasmo suscitato da sistemi in cui, ad esempio, semplicemente passando con un localizzatore (insomma, un „mouse“) su

una parola sullo schermo, ne appaiono le varianti e i manoscritti che le testimoniano, in serie orizzontali e/o verticali, basterà ricordare che si tratta dei medesimi risultati che il sistema TUSTEP consentiva già oltre vent'anni fa, sia pure con minor scintillio nella modalità grafica; f) nel rivendicare, infine, di voler fornire alberi genealogici, indicazioni visualizzabili in vario modo sullo schermo, magari diversificate con l'uso di colori, sigle di codici, varianti, etc., si perde di vista un elemento essenziale: che l'essenza, e l'utilità, del lavoro di un editore risiede nel *togliere*, non nell'*aggiungere*, nella scelta, nella selezione, non nell'*accumulo*. *Non multa, sed multum!*

Emerge allora un tema decisivo, quello del rapporto tra quantità e qualità. Non è poi nuovo. Lo si può definire ricorrendo nuovamente a Pasquali, che nel 1920 concordava in ciò (caso più che raro!) con Ettore Romagnoli: „il mineralogo che raccogliesse e catalogasse uno per uno tutti i ciottolini della ghiaia di un fiume, sarebbe un rotondissimo imbecille“.

Un difficile termine tedesco (*longa nomina, contortuplicata habemus*: era Plauto, al v. 708 della *Persa*) usato nei primi trattati sulla tecnica dell'edizione definisce ciò che una edizione e un apparato non devono essere: *Schuttablagerungsstätte*. L'italiano ne possiede un espressivo corrispondente: „discarica di rifiuti“. È ciò a cui conduce la rinuncia a valutare, selezionare, scartare, e invece raccogliere indistintamente tutti i materiali che la tradizione offre, fornirli al lettore senza vaglio adeguato. Una *contradictio in adiecto*, se in gioco sono edizioni *critiche*, giacché la parola, già nel greco *krisis*, indica precisamente la decisione, la scelta, il giudizio: tutto il contrario del soffocante accumulo, più o meno organizzato ma comunque deresponsabilizzante.

Siamo lontani dal raggiungimento di risultati davvero nuovi, come è stato recentemente riconosciuto anche dai fautori a ogni costo delle edizioni elettroniche: fino ad oggi, ciò che si è realizzato nel campo delle edizioni grazie alle nuove tecnologie non è andato al di là di ciò che la stampa ha offerto, né in termini di prodotto, né in termini di processo (il modo in cui il prodotto è realizzato). Sono cambiate presentazione e distribuzione dei risultati: non la produzione. Una cosa sia detta con chiarezza: si è spesso di fronte a iniziative interessanti, e a loro modo utili. Che tuttavia non possono essere spacciate per ridefinizione delle edizioni critiche, né qualcuno può partire di lì per venirci a insegnare „what is critical editing“, come recitava il titolo di un articolo un decennio fa.

Un motivo si intravede facilmente alla base del diffondersi di progetti e di presunte innovazioni: ed è che in esse possono facilmente trovare spazio quei „mestieranti“ di cui si faceva cenno.

Le strade intraprese

Non sembri, però, che i rilievi critici fin qui adottati esauriscano la questione con un latente invito a lasciar perdere. A volte, sarebbe questa la soluzione ottimale, la più onesta: quando cioè ci si renda conto, che il problema affrontato non è formalizzabile in termini informatici. Ci sono tuttavia molti aspetti su cui vale la pena impegnarsi. Innanzitutto, come si sarà intuito, quelli teorici. Ma anche le ricadute operative.

Delle banche dati, dal TLG in poi, si è già detto. Si tratta di raccolte che consentono di effettuare meglio operazioni tradizionali. Bisognerebbe almeno aggiungere ai testi gli apparati ricorrendo a un sistema di rappresentazione non banale, ma nessuno sa come procedere, giacché qui la semplice riproduzione speculare della pagina stampata ha ancora meno senso. Ideale sarebbe la codifica esplicita e diretta dei manoscritti, come detto – ma questo è un lavoro di là da venire, per l'impegno che richiede, e che consentirebbe di realizzare finalmente anche il repertorio auspicato già da Paul Maas agli inizi del secolo scorso (*Critica del testo*, 18s.) e allora impossibile, al fine di „compilare per epoche, generi letterari, regioni paleografiche un prospetto di tutti gli errori particolari, ordinati secondo le loro specie sul fondamento di testimoni di cui si sia conservato il capostipite (le lezioni particolari dei quali vengono perciò con ragione generalmente taciute nelle edizioni critiche): poi si dovrebbe passare agli errori particolari di quei testimoni di cui è sicuramente ricostruibile l'antenato per mezzo di recensione; solo in ultima linea sarebbero da adoperare quelli il cui antenato è ricostruibile soltanto per mezzo di *selectio* o addirittura soltanto di *divinatio*. Particolarmente desiderata sarebbe una siffatta ricerca per le interpolazioni“. Scrivevo queste stesse cose già e non pochi anni addietro, non faccio che ribadirlo: l'utilità di un tale repertorio sarebbe straordinaria.

Ciò che invece cambia radicalmente rispetto al lavoro tradizionale sono le modalità di rappresentazione e diffusione della conoscenza, e questo anche in progetti che, pur limitandosi a offrire strumenti di servizio, quanto meno non cercano di passare per ciò che non sono. Attenzione viene ad esempio rivolta alle modalità di accesso ai dati. Si vanno sperimentando, è cosa recente, sistemi per la accessibilità delle edizioni critiche, come presso il Corpus Medicorum Graecorum dell'Accademia delle Scienze di Berlino, dove è possibile consultare in via sperimentale, con un sistema interessante, una piccola (per ora)

porzione di un testo greco, con varie modalità a scelta dell'utente, e la possibilità di effettuare ricerche nel testo e negli apparati, secondo categorie (come quelle di congettura, aggiunta, omissione, trasposizione, integrazione marginale, espunzione, etc.). Tra l'Università di Amburgo e l'Aristoteles-Archiv di Berlino un ulteriore progetto si annuncia di notevole portata. Denominato *Teuchos*, fondato da D. Harlfinger e diretto da C. Brockmann, intende costituire un Centro per la ricerca sui manoscritti e testi classici e per la elaborazione di testi in vista della realizzazione di edizioni, e prevede lo sviluppo di specifiche applicazioni (open source). Il sistema trae spunto dalla recente evoluzione del trattamento di manoscritti e papiri mediante strumenti di *imaging* digitale, e intende mettere a disposizione via rete, in un sistema aperto, non solo le immagini stesse, ma tutti i dati risultanti dalla loro descrizione e analisi, collegati tra loro, in vista di trascrizioni, collazioni, stemmi dei codici, raccolta di varianti, traduzioni, commenti.

I sistemi di trattamento delle immagini in questo campo hanno già avuto risultati di un certo interesse nei progetti della „Philological Workstation BAMBI“, presso il CNR di Pisa, e del „Rinascimento Virtuale“, finanziato dalla Unione Europea, per il recupero delle scritture scomparse nei palinsesti; sistemi di *imaging* multispettro hanno anche mostrato le loro potenzialità a Oxford dove sono stati applicati alla decifrazione dei più ardui tra i papiri di Ossirinco e di Ercolano.

Progetti più o meno vasti sono numerosissimi. Tra tanti, spicca decisamente uno dei primi ad aver mostrato ambizioni rilevanti, il progetto PERSEUS dell'Università americana Tufts, grande raccolta di materiali collegati tra loro, testuali e non testuali (qualche centinaio di testi in lingua originale e traduzione, legati in un sistema ipertestuale a commenti, dati archeologici, storici, linguistici, strumenti per l'analisi morfologica del testo, dizionari, compreso il Liddell-Scott – sia pure nella sola edizione del 1940, copyright permettendo –, e quant'altro).

Software ne esistono ormai di ogni genere: dalla ricerca di patterns alle collazioni, a indici e concordanze, alla gestione e impaginazione di edizioni critiche, a software sviluppati nell'ambito delle scienze biologiche (le analisi cladistiche della filogenetica) e utilizzati per la individuazione di rapporti di parentela tra manoscritti, alla lettura di manoscritti e papiri di difficile decifrazione, e poi analisi linguistiche, statistiche, sistemi di lemmatizzazione (uno molto efficiente per il greco gelosamente custodito a Boston) e quant'altro.

Funzioni quasi tutte, ha senso ricordarlo, già presenti e spesso con superiore efficacia – e questo dovrebbe insospettire –, in un sistema nato negli anni '70 come TUSTEP, costituito da moduli autonomi agganciati l'uno all'altro (i risultati dell'uno diventano punto di partenza dell'altro), ognuno programmabile in modo autonomo dall'utente – un sistema di cui si è spesso biasimata la complessità, che però non è altro che la complessità della realtà osservata e fatta oggetto di analisi, e che è possibile semplificare solo a costo di una banalizzazione e perdita di elementi ognuno potenzialmente decisivo.

Segnalo infine un solo ulteriore progetto, che non ha a che fare con i nostri studi ma preannuncia ricadute più ampie e stimolanti riflessioni. Inizialmente chiamato TIDES, avente come fine la „Translingual Information Detection, Extraction, and Summarization“, si è ora articolato nel progetto GALE – Global Autonomous Language Exploitation, e va visto in parallelo con la TREC – Text Retrieval Conference. È un progetto militare americano presso la DARPA (Defense Advanced Research Projects Agency – dove, sia detto per i novizi, ebbe origine la rete Internet), e prevede la comprensione e il trattamento in contemporanea di testi in numerose lingue diverse, e la possibilità di interrogarli tutti insieme con richieste poste in un'unica lingua (l'inglese). Si articola in tre fasi: Trascrizione, Traduzione, Distillazione dei dati. Su scala assai ridotta, una esigenza non dissimile si era presentata anche per i testi classici a chi scrive, ma senza poterle dare seguito, per l'analisi in contemporanea di testi di Galeno tramandati in lingue diverse – greco, latino, arabo, ebraico: si capirà dunque la novità e l'interesse per questa iniziativa.

Infine: grande attenzione viene oggi rivolta alla digitalizzazione di massa di materiali scritti. Da progetti generali come GoogleBooks, ma ancor più da iniziative notevoli come la Open Content Alliance (il cui principio è „Costruire un archivio digitale di contenuto globale per l'accesso universale“), si hanno già ora e si avranno ricadute anche per i testi classici: Gregory Crane, responsabile del progetto PERSEUS, coordina una vasta ricerca focalizzata sul modo di estrarre automaticamente informazioni significative da una biblioteca contenente milioni di volumi in formato digitale. Come sfruttare al meglio milioni e milioni di pagine memorizzate come semplici immagini? E come far sì che i software, ad esempio quelli per le traduzioni automatiche, „imparino“ e affininno le proprie prestazioni sulla base di quanto è già stato codificato in modo esplicito con etichette morfologiche, sintattiche, semantiche che descrivano ciascuna componente del testo? Come si applica questo ai testi classici? Come si ottiene una digitalizzazione automatica corretta delle informazioni contenute in un apparato critico? Le prospettive possono essere entusiasmanti, ma anche deludere se non si sapranno tenere nel dovuto conto le incompatibilità

metodologiche e di principio che troppo spesso si tende a ignorare in nome della immediata fattibilità dei progetti e soddisfazione dei finanziatori.

Un merito va energicamente sottolineato, in ogni e qualsiasi iniziativa in questo campo: esse costituiscono un elemento decisivo nell'ambito di una strategia complessiva per la sopravvivenza e una rinnovata diffusione degli studi classici.

Epilegomeno

Il mondo digitale trasforma ogni suo oggetto in informazione. Questa, è stato scritto, è il contrario della conoscenza, rischia di dissolverne i contenuti. Incolore com'è, ridotta a brandelli decontestualizzati, essa deve trovare nuove coordinate, recuperare significato. „All our knowledge brings us nearer to our ignorance / (...) / where is the knowledge we have lost in information?“ (T.S. Eliot, *Choruses from 'The Rock'*, I 11-16).

Qualche rinvio

Quanto segue è parte informativa integrante dell'argomentazione, e dà più precise coordinate a quanto addotto nel testo (nessuna esaustività, né obiettività).

Un quadro della situazione fino a circa quattordici anni fa, con bibliografia, era fornito da chi scrive in *Filologia computazionale*, Accademia Nazionale dei Lincei, Roma 1995, quindi in *Filologia classica in prospettiva. Machina, ratio et res ipsa*, in *I nuovi orizzonti della filologia*, Atti dei Convegni Lincei 151, Accademia Nazionale dei Lincei, Roma 1999, pp. 133-156.

La collana „Computer und Antike“ dell'editore tedesco Scripta Mercaturae di St. Katharinen comprende 7 volumi, su aspetti applicativi. Si vedano in particolare i voll. 1, su ricerche di patterns, concordanze e indici, e 4, sull'elaborazione elettronica nelle scienze dell'antichità, rispettivamente A. Meier, C. Schäfer, *Computer und antike Texte. Wortrecherche, Konkordanz- und Indexerstellung mit Volltextdatenbanken*, e M. Fell et al., edd., *Machina Computatoria. Zur Anwendung von EDV in den Altertumswissenschaften* [EDV sta per „Elektronische Datenverarbeitung“, trattamento elettronico dei dati]. W. McCarty, *Humanities Computing*, Palgrave, Houndmills – New York 2005, offre riflessioni interessanti anche sull'antichità classica.

Importante il punto della situazione offerto da un Convegno svoltosi nel gennaio 2006 ad Amburgo (nell'ambito dell'iniziativa Teuchos): *Digital Philology – Problems and Perspectives. International conference on the development and applica-*

tion of digital technologies in paleography and scholarly editing (programma e abstract sul sito di Teuchos).

Sulle edizioni digitali, la ricerca riguarda essenzialmente testi medioevali, qui di non diretta pertinenza. Richiamo per tutti, condividendolo, l'approccio di D. Buzzetti, ad esempio in *Ambiguità diacritica e markup. Note sull'edizione critica digitale*, al Convegno di Pavia del marzo 2000, *Soluzioni informatiche e telematiche per la filologia* (disponibile su lettere.unipv.it/diplamm/pubtel/Atti2000/dino_buzzetti.htm). Mostra maggiore consapevolezza rispetto a precedenti interventi dell'autore, sebbene resti per diversi aspetti non condivisibile, P.M. Robinson, *Where we are with electronic scholarly editions, and where we want to be*, „Jahrbuch für Computerphilologie“ 5, 2003, 125-146 (online su computerphilologie.uni-muenchen.de/jg03/robinson.html).

Le guide più aggiornate e complete sono online:

In italiano, la bolognese „Rassegna degli Strumenti Informatici per lo Studio dell'Antichità Classica“, su <http://www.rassegna.unibo.it/index.html>.

In tedesco alcune tra le più ricche, tra cui Kirche, „Katalog der Internet-Ressourcen für die Klassische Philologie aus Berlin“, <http://www.kirke.hu-berlin.de>, nonché, nato a Göttingen ma ora non aggiornato, Athena, „Ausgewählte Tools und Hinweise zur EDV-Nutzung in der Altphilologie“, su <http://www.athena.luehken.info>.

In inglese, presso il sito del TLG, la lista di Electronic Resources for Classicists, <http://www.tlg.uci.edu/index/resources.html>, nonché la sezione specifica del classico Humbul Humanities Hub di Oxford, ora divenuto Intute, su <http://www.intute.ac.uk/artsandhumanities/>.

In francese, la Bibliotheca Classica Selecta di Lovanio, su bcs.fltr.ucl.ac.be/ResElec.html, nonché il Carnet d'adresses sur l'antiquité gréco-romaine, su artela.cnarela.free.fr/signets-latin-grec.htm e <http://www.musagora.education.fr>.

Sempre a Lovanio, una guida allo studio del greco con particolare attenzione per la grammatica e la storia della lingua, è su perswww.kuleuven.be/~u0013314/greekg.htm.

Sul tema Internet per l'insegnamento del greco e del latino, e per un elenco di links, utile è il sito dell'Istituto di Filologia classica di Vienna: <http://kphil.ned.univie.ac.at/node/124677>.

Alla filologia non classica, ma con indicazione di utili risorse, è dedicato Digital Variants: <http://www.selc.ed.ac.uk/italian/digitalvariants/philology/philologyhome.htm>.

Esempi di software per la realizzazione (in realtà il layout) di edizioni critiche e commenti e per la pubblicazione elettronica: presso l'Accademia Austriaca delle Scienze è stato sviluppato il Classical Text Editor, su cui cf. <http://www.oeaw.ac.at/kvk/cte>.

Allo stesso scopo: Critical Edition Typesetter, su <http://www.karas.ch/cet>.

Anche per il sistema TeX (LaTeX), meritoriamente Open Source, sono stati sviluppati pacchetti macro ed eccellenti strumenti specifici per le edizioni critiche, cf. <http://www.ctan.org/tex-archive/macros/latex/contrib/ledmac/>, una descrizione è in <http://www.djdekker.net/ledmac>, nonché (per Apple Mac) <http://www.uoregon.edu/~koch/texshop>.

Caratteri greci per la visualizzazione su web: cf. Unicode Polytonic Greek for the World Wide Web del Consorzio Stoa, su <http://www.stoa.org/unicode>.

Per l'esempio di visualizzazione delle edizioni critiche del Corpus Medicorum Graecorum della Accademia delle Scienze di Berlino: pom.bbaw.de/cm/index2.html.

Per il progetto Teuchos, il sito è <http://www.teuchos.uni-hamburg.de>.

Per il progetto Gale: projects.ldc.upenn.edu/gale; Trec: trec.nist.gov; Tides sembra fermo a <http://www.darpa.mil/darpatech99/Presentations/scripts/ito/itotidesscript.txt>

Per il progetto coordinato da G. Crane, si sono tenuti tra il 2007 e il 2008 incontri negli USA e in Europa, l'ultimo a Berlino nel marzo 2008 in collaborazione con il progetto Teuchos. Cf. <http://www.dlib.org/dlib/march06/crane/03crane.html> nonché <http://www.clir.org/activities/digitalscholar/Nov28final.pdf> e <http://www.digitalhumanities.org/dhq/vol/003/1/000034.html>.

Open Content Alliance: <http://www.opencontentalliance.org>.

TLG: <http://www.tlg.uci.edu>.

PHI: <http://www.packhum.org> (ma il sito dice poco).

PERSEUS: <http://www.perseus.tufts.edu>.

Sui papiri di Oxford cf. www.papyrology.ox.ac.uk. Per i papiri ancora Apis, „Advanced Papyrological Information System“, <http://www.columbia.edu/cu/lweb/projects/digital/apis>, nonché il Duke Papyrus Archive, scriptorium.lib.duke.edu/papyrus, e la Duke Data Bank of Documentary Papyri, disponibile su cd (phi 7) o attraverso il sito del progetto Perseus. La banca dati è in corso di conversione alla codifica tei-sgml. Descrizione: odyssey.lib.duke.edu/papyrus/texts/DDBDP.html.

Iscrizioni greche (ottimo): epigraphy.packhum.org/inscriptions/.

Il sistema modulare TUSTEP può essere utilizzato per la maggior parte delle applicazioni, comprese quelle per cui sono stati già ricordati software specifici, come edizioni critiche, collazioni, indici, etc.: <http://www.zdv.uni-tuebingen.de/static/skripte/tustep>.

LSJ: il lessico di Liddell e Scott è consultabile sul sito del Perseus, oppure attraverso il software Diogenes, e disponibile su cd-rom.

Bibliografie: Gnomon è su <http://www.gnomon.ku-eichstaett.de/Gnomon/ts.html>, l'Année Philologique su <http://www.annee-philologique.com/aph>. Altre bibliografie sono elencate su perswww.kuleuven.be/~u0013314/pinaxonline.html.

TOCS-IN, „Tables of Contents of Journals of Interest to Classicists“, scheda gli interventi apparsi su riviste e su volumi miscelanei (non sempre è completa): <http://www.chass.utoronto.ca/amphoras/tocs.html>.

TEI: <http://www.tei-c.org> (il manuale, in particolare la sezione 19. Critical Apparatus).

BMERR (Bryn Mawr Electronic Resources Review): csanet.org/bmerr.

CNR Pisa: cf. in particolare la sezione „Modelli e strumenti computazionali per la ricerca umanistica, con particolare riguardo alle discipline linguistiche, filologiche e alla lessicografia“, su <http://www.ilc.cnr.it/viewpage.php/sez=ricerca/id=7/vers=ita>.

BAMBI (Better Access to Manuscripts and Browsing of Images – Philological Workstation): <http://journals.tdl.org/jodi/article/viewArticle/10/20>.

(link verificati in data 13.07.2009)

Lorenzo Perilli
Università di Roma Tor Vergata
Dipartimento di Antichità e Tradizione Classica
Via Columbia 1
I-00133 Roma
E-Mail: lorenzo.perilli@uniroma2.it