STRUMENTI INFORMATICI IN ANTICHISTICA*

I. Premessa: dal libro al libro elettronico

I.1. Testi e Ipertesti

"Rivoluzione cibernetica, telematica, multimediale" sono annunciate ormai quotidianamente dai *media*. "È davvero finita la nostra secolare civiltà della carta?" ci interroga il sottotitolo di un inserto speciale di "Repubblica" ("Rapporto Information Technology. Il signor Gutenberg e il CD Rom", del 19/9/1995) dedicato al nuovo orizzonte dell'editoria, il libro su dischetto e CD-ROM: dizionari, enciclopedie, ed anche libri di testo e romanzi, forniti magari di immagini e di suoni richiamabili con un semplice click del mouse. I bambini di un domani sempre più vicino, anzi di oggi, potranno ascoltare e insieme vedere le immagini delle fiabe non più grazie ai libretti accompagnati da dischi "microsolco originale di Walt Disney" o da cassette, ma inserendo un CD-ROM nel lettore: i titoli sono i più vari, dalle fiabe, appunto, ai miti greci, alle enciclopedie generali e specializzate, di musica, arte, ecc. Per l'antichità classica si può ricordare ad esempio *Perseus 2.0. Interactive Sources and Studies on Ancient Greece*, che contiene le opere di 31 autori greci (tra cui Omero, Esiodo, Pindaro, Bacchilide, Aristofane e i tre tragici maggiori, Erodoto, Tucidide, Pausania, Apollodoro, Plutarco, Demostene, Lisia, Isocrate e gli oratori minori, Platone ed Aristotele), spesso accompagnati da traduzioni e note, 25.000 immagini di opere architettoniche, di sculture, vasi e siti archeologici, riprese anche dal satellite Landsat (a costituire assieme alle mappe un vero e proprio atlante), l'*Intermediate Greek-English Lexikon* di Liddell e Scott, una storia della Grecia nel V secolo, una enciclopedia di antichità classiche. Come spiega E.Mylonas, già direttore del progetto, "tutti i materiali contenuti in *Perseus* sono strettamente collegati, cosicché il lettore di un'opera teatrale, se incontra il nome di Eracle, può facilmente trovare occorrenze dello stesso nome in altri testi, vedere raffigurazioni del personaggio o una carta geografica su cui sono indicati i luoghi più importanti per quel personaggio, consultare un dizionario, leggere la voce Eracle sull'enciclopedia"[1]. Interattività è proprio questo, sfruttando i rimandi interni presenti nel testo (anzi nell'ipertesto, data la sua architettura), crearsi dei propri percorsi di lettura attraverso lo scritto, e, laddove ci siano, le

* F.Citti ha scritto i capitoli I e III e il paragrafo II.1; A.Tabarroni i paragrafi II.2-5; T.Del Vecchio il capitolo IV.

1 Cito da M.Lana, *L'uso del computer nell'analisi dei testi. Scienze umane e nuove tecnologie*, Milano, 1994, 189; per una recente descrizione del progetto *Perseus*, v. ancora E.Mylonas - G.Crane - K.Morell - D.N.Smith, *'The Perseus Project'. Data in the Electronic Age*, in J.Solomon (ed.), *Accessing Antiquity. The Computerization of Classical Studies*, Tucson-London 1993, 133-56 (oltre alle pagine dedicate in rete [su cui v. § IV,1-2], reperibili via http://www.yale.edu/yup/Perseus2.html; ci si può fare un'idea più precisa del suo contenuto tramite la Perseus Project Home Page, via http://www.perseus.tufts.edu).

immagini e i suoni, per cui il testo elettronico in realtà non è dato una volta per tutte, ma si crea (o meglio si ridispone) in base alla curiosità del lettore, che si serve a suo piacimento delle possibilità offertegli dall'ipertesto, entro i limiti che esso, pur sempre, gli impone. L'ipertesto poi, potrebbe essere in futuro un punto di approdo di quello che è il lavoro più tipico del filologo, il commento dei testi: le note, un tempo in margine al testo stesso, e quindi spostate in calce, o divenute autonome, forse un giorno saranno trasformate in formato elettronico, così che "cliccando" sulle parole commentate, evidenziate in qualche modo nel testo, verrebbero visualizzate le note, con i passi paralleli, e la possibilità di ampliarne il contesto e di passare di rimando in rimando, da un'opera all'altra, il che conferirebbe all'intertestualità una dimensione immediatamente percepibile, o potrebbe semplificare il raffronto di varianti molto diverse di un testo, difficilmente traducibile in un apparato tradizionale (si pensi ad esempio alle *Grazie* foscoliane)[2]. È evidente che tutto questo interessa il filologo da un punto di vista pratico, perché avrà a sua disposizione nuove e potenti banche dati testuali e bibliografiche, ma anche da un punto di vista teorico, perché ogni mutamento delle tecniche della comunicazione, letteraria o no, ovvero della produzione, della diffusione e della ricezione del testo riguarda la filologia.

I.2. Testo elettronico e neo-oralità

La scrittura prima, la stampa poi, hanno prodotto - come è ben noto - un profondo mutamento non solo sulla prassi letteraria e, più in generale, sulla prassi discorsiva, ma anche su quella cognitiva e mnestica. Ma non solo, come osservava W.J.Ong, "il passaggio dall'oralità alla scrittura e da questa all'elaborazione elettronica comporta un mutamento nelle strutture sociali, economiche, politiche, religiose, ecc."[3]. La società orale infatti tende all'estroversione, grazie anche alla funzione socializzante del suono, e alla trasmissione orale-aurale, quella della stampa ha sviluppato il senso del privato e della coscienza personale, col diffondersi della lettura silenziosa che ha definitivamente allontanato il suono dalle parole, relegandole sul piano visivo. A differenza del "testo" orale, che si rinnova ad ogni performance, la scrittura e la stampa isolano il lettore da altri possibili interlocutori (se non metaforici, come i grandi del passato che dialogavano, dalla pagina scritta, col Machiavelli), dando l'idea di un prodotto "chiuso", finito, in uno stato di autonomia e completezza, che si ripete uguale in tutte le copie in cui è stampato. "Il testo stampato dovrebbe rappresentare le parole di un autore in forma definitiva o 'finale' poiché la stampa è a proprio agio solo con le cose

[2] Cf. Lana, cit., 38 s., e soprattutto M.Gori, *L'edizione ipertestuale della 'Famiglia dell'antiquario'*, in C.Leonardi - M.Morelli - F.Santi (edd.), *Macchine per leggere. Tradizioni e nuove tecnologie per comprendere i testi*, Spoleto 1993, 201-11; J.D.Bolter, *Hypertext and the Classical Commentary*, in J.Solomon (ed.), cit., 157-71.

[3] W.J.Ong, *Oralità e scrittura. Le tecnologie della parola*, Bologna 1986 [ed. or. *Orality and Literacy. The Technologizing of the World*, London and New York 1982], 21.

definitive, finite", scrive epigraficamente Ong[4]. Il testo elettronico, l'ipertesto cui si è accennato sopra, vuole essere l'esatto contrario, un testo aperto e percorribile in maniera ogni volta diversa. Questo è vero, ma non è forse qui la differenza più profonda, giacché del libro stampato ogni lettura, ci insegna Borges, è una riscrittura: l'era dell'elettronica, del multimediale, come ha colto ancora una volta con grande lucidità Ong, è un'era di "oralità di ritorno", o meglio, giacché è permanentemente basata sulla scrittura e sulla stampa, di "oralità secondaria", e presenta "sorprendenti somiglianze con quella più antica per la sua mistica partecipatoria, per il senso della comunità, per la concentrazione sul momento presente e persino per l'utilizzazione delle formule"[5]. Sono caratteristiche, queste, che fanno immediatamente pensare ad un nuovo fenomeno insorgente, e anch'esso significativo sul piano della scrittura: i gruppi di discussione e le cosiddette riviste elettroniche, cui si accede via Internet, la rete che rende veramente concreta ed evidente l'idea di "villaggio universale" costituito dai suoi fruitori, che comunicano grazie ad un codice speciale ed altamente formalizzato, quello dei linguaggi e dei software elettronici. Sono ormai numerosi i gruppi di discussione, detti anche liste a tema per gli umanisti, e per i classicisti in particolare, come si vedrà, e le riviste elettroniche[6], che, solitamente non consentono un intervento libero, ma prevedono la pubblicazione di articoli, non sempre destinati poi alla stampa, e che, talora, vengono aggiornati in redazioni successive: se è evidente che sul piano pratico si pone ormai il problema della conservazione e della citazione di questo materiale altamente "deperibile" perché legato al contingente e quindi fisicamente eliminato, altrettanto chiaro è che ci troviamo in un campo neutro in cui oralità e scrittura si toccano, giacché si tratta di scrittura elettronica, trasmessa via cavo, ma soprattutto perché non ha il segno della permanenza.

I.3. Morte del libro?

Più di trent'anni fa scriveva il grande codicologo e critico testuale A.Dain, con misurata preveggenza e non senza malinconia: "Viendra peut-être un jour où l'âge de l'imprimerie finira à son tour: photographies, microsillons, bandes magnétiques et autres enregistrements du son se substitueront à l'usage des lettres; livres manuscrits et livres imprimés ne seront plus alors que des pièces archéologiques. L'ère du livre, née à Alexandrie vers le IIIe siècle avant J.-C., serait close définitivement. Nous ne voulons pas le croire"[7]. Di morte del libro, poi, negli anni ottanta, parlava tra gli altri Lucia Cesarini Martinelli, in una introduzione divulgativa alla

[4] Id., 187, ma per il concetto di "punto di vista fisso" legato alla stampa, v. già M.McLuhan, *La galassia Gutenberg. Nascita dell'uomo tipografico*, Roma 1976 (ed. or. *The Gutenberg Galaxy. The Making of Typographic man*, Toronto 1962), 176 ss.

[5] Id., 191.

[6] Cf. infra, § IV.2.

[7] A.Dain, *Les manuscrits*, Paris 1975[3] [1949[1]], 14.

filologia: "in futuro è il libro stesso, nella forma a noi familiare, che rischia di estinguersi"[8]; oggi, ebbri di entusiasmo un po' fatuo, i nuovi paladini della cibernetica ne decretano la morte.

L'annuncio pare, per il momento, quanto meno affrettato: il libro elettronico non è autosufficiente, necessita di un personal computer, per quanto portatile, se su dischetto, di un lettore apposito ad esso collegato se su CD-ROM: è perciò poco trasportabile, è consultabile solo in luoghi attrezzati allo scopo. Questo aprirebbe un altro discorso, sull'idea di spazio della lettura, di biblioteca, quale l'abbiamo concepita fino a oggi, ma soprattutto prima della diffusione del tascabile che si può portare con sé, anche in viaggio, e per di più è di grande tiratura, economico, e quindi accessibile a un pubblico più vasto, anche nella dimensione del privato. L'uso di materiale multimediale, destinato non solo alla semplice lettura, ma anche all'ascolto, comporta spazi diversi, oltre che strumenti diversi da quelli tradizionali, e d'altra parte anche dal punto di vista architettonico la storia delle biblioteche avrebbe molto da insegnare su libri, lettori, ed anche sulle politiche culturali.

Paradossalmente il passaggio dal libro al testo elettronico riporterebbe ad un modo di sfogliare il testo in parte analogo a quello abbandonato nel passaggio dal rotolo al codice e che gli inglesi indicano con il medesimo verbo, "scroll", poiché proprio come nel rotolo (che pure si svolge orizzontalmente, e presenta lunghe colonne parallele di testo) non si sfoglia una pagina dopo l'altra, ma si fanno passare attraverso il video porzioni successive di testo che si svolge verticalmente: è questa ormai una prassi divenuta abituale per chi scrive con il più normale dei programmi di videoscrittura, ma certo la migliore consultabilità del codice rispetto al rotolo, e quindi del libro rispetto al video - con tutti i problemi di refresh e di stabilità dei caratteri, e di fastidio alla vista che esso può causare - resta indubbio. Ma, come insegna ancora la storia del libro, le motivazioni pratiche, quali migliore leggibilità e maneggevolezza sono solo "motivazioni che hanno potuto giocare un ruolo complementare, ora più ora meno importante; ed invece nel valutare l'avvento del codice è da considerare, piuttosto, il rapporto tra produzione libraria e pubblico, intendendo per pubblico i destinatari concreti del libro, inseriti nel loro contesto politico, economico-sociale, culturale, in una parola storico nel quale operavano"[9].

Lanciarsi in previsioni sul futuro del libro è quindi quanto mai complicato e improbabile, visto che tecnologia e orizzonti della cultura travalicano spesso la fantasia; comunque mi sentirei di condividere il parere di Ong che "l'elettronica non sta uccidendo il libro stampato, ma anzi ne incrementa la produzione [...] Il nuovo mezzo rafforza l'antico, ma naturalmente lo trasforma"[10]: è probabile, come afferma tra gli altri Umberto

[8] L.Cesarini Martinelli, *La filologia*, Roma 1984, 142.

[9] G.Cavallo, *Libro e pubblico alla fine del mondo antico*, in *Libri, editori e pubblico nel mondo antico. Guida storica e critica*, Roma-Bari 1984, 83.

[10] Ong, cit., 191.

Eco[11], che le enciclopedie, i libri di consultazione, ingombranti e costosi possano essere sostituiti da dischi multimediali, di più rapida consultazione, come ormai sta avvenendo per i cataloghi delle grandi biblioteche, come quella del Congresso[12]; ma i libri da leggere, i romanzi, i classici, resteranno sulla carta, senza subire troppo la concorrenza del prodotto elettronico indirizzato ad un mercato in parte diverso da quello dei lettori che sono sempre stati, e in Italia in particolare, un'élite "aristocratica". Come della carta non si possa fare a meno, è dimostrato da alcune iniziative di editori che producono libri a basso costo o freeware, acquisibili via rete, e quindi stampabili con stampanti laser e rilegabili artigianalmente, risparmiando ovviamente sui costi di produzione, magazzino, distribuzione, nonché abbassando i rischi di impresa. Lo scenario paventato da alcuni[13], di una drastica potatura nella tradizione, nel trasferimento dei testi dal libro al supporto magnetico, e la successiva distruzione dei libri (con il conseguente formarsi di nuovi archetipi, elettronici appunto, per la nostra tradizione), in modo analogo a quanto avvenuto nel passaggio dal rotolo al *codex*, che ha selezionato la letteratura classica, pare per ora molto immaginario, e d'altra parte esiste già una selezione naturale per la fin troppo abbondante pubblicazione che non trova riscontro in una eguale richiesta. Più concrete invece le ipotesi di applicazioni nel campo della didattica: da tempo esistono programmi per lo studio delle lingue con esercizi di verifica e autovalutazione, anche per le lingue classiche[14]. Di recente è stato commercializzato anche un dischetto dedicato all'*Eutifrone* di Platone, concepito con l'idea di far rivivere il dialogo a chi si trova al di là del video non come spettatore, ma "come dialogante effettivo", "sostituto di Eutifrone e interlocutore diretto di Platone"[15], che, in una versione precedente, ha suscitato l'interesse degli studiosi di alcune scuole magistrali e liceali: nella didattica in effetti non si può non tenere conto delle possibilità di presentazione della materia che i nuovi mezzi offrono, e del fascino che hanno verso il pubblico degli studenti, riuscendo a suscitarne un coinvolgimento attivo. Altrettanto evidente è l'utilità del testo elettronico, ancora più di quello registrato da una voce su nastro, per i non vedenti, che finalmente hanno la possibilità di accedere a una grande quantità di testi, e di riprodurli con stampanti braille a costi non proibitivi, ovvero di ascoltarli tramite un sintetizzatore vocalico, anche se resta, come segnala un recente articolo sulle «Transactions of the American Philological Association»[16], il

[11] Cf. l'inserto di "La Repubblica", "Rapporto Information Technology. Il signor Gutenberg e il CD Rom", cit., 15.

[12] Cf. infra, § III.2.5 (ed anche IV.2).

[13] Cf. ad es. Cesarini Martinelli, cit., 146.

[14] La maggior parte dei materiali disponibili è in lingua inglese (più di rado in francese): rinvio al capitolo relativo di A.Cristofori, *Rassegna degli strumenti informatici per lo studio dell'Antichità classica* (http://ecno01.cineca.it/dipartim/stoant/rassegna1/intro.html)

[15] Platone, *L'Eutifrone*, a c. di L.Rossetti, Roma 1995, 9.

[16] G.Forsythe, *Electronic Technology and Blindness: a New Solution to an Old Problem*, TAPA 124, 1994, 333-35.

problema del formato: non tutti i testi dei CD sono facilmente esportabili e maneggiabili con un programma normale di scrittura, ma numerosi, anche per proteggere il copyright, sono compressi e non consentono le medesime operazioni.

I.4. Il manoscritto elettronico e i problemi di composizione e trasmissione

Per ora, comunque, si deve prendere atto che la composizione elettronica ha ormai del tutto soppiantato i vecchi, scomodi, ma probabilmente più durevoli "piombi", e questo con ovvie conseguenze sulla tipologia dell'errore della tradizione a stampa. È la nozione stessa di manoscritto che varia da epoca ad epoca, come osservava ancora il Dain: una cosa sono i codici manoscritti, una cosa i manoscritti preparatori del *Discours de la condition de l'homme* di Pascal, e quelli delle lettere di Napoleone all'imperatrice Maria-Luisa; "la notion du manuscrit sera encore différente s'il s'agit des rédactions autographes des poèmes de Victor Hugo, voire des dactylogrammes de nos romanciers modernes - car le manuscrit moderne se tape à la machine et on l'appelle, je crois, *typescript*"[17]. Ancora differenti i manoscritti odierni, giacché spesso gli autori, quando addirittura non confezionano il libro "camera ready", pronto per la sola riproduzione fotografica, consegnano un testo elettronico, su dischetto, per evitare la fatica di correggere le bozze, o meglio per alleviarla: a parte gli errori sistematici di transcodifica dal formato del word processor a quello di stampa, soprattutto per i caratteri speciali (e per i font, greco ed ebraico, per fare un esempio), a volte non mancano tuttavia errori inopinati, come la scomparsa di un corsivo tra le seconde e le terze bozze, o l'abbassamento del maiuscoletto a tondo, per l'errore di un tipografo maldestro che ha eliminato il carattere di controllo del corpo di stampa.

Il problema è comunque ancora più complesso e non riguarda solo la fase della edizione a stampa del libro, perché ormai la scrittura a macchina (al computer) "devient pour beaucoup l'écriture normale"[18]. Ogni età ha il suo immaginario della scrittura: grandi giornalisti come Brera sono legati alla portatile Olivetti Lettera 22, un modello che ha segnato un'epoca, anche nel design; per l'antichità classica il poeta che scrive è associato a stilo e tavoletta, nella sua Recanati ci possiamo figurare il Leopardi intento all'opera tra le sudate carte, mentre gli autografi dei *Canti*, le edizioni a stampa, nonché le correzioni sull'edizione a stampa dello Starita, la cosiddetta Starita corretta (pubblicati dal Moroncini[19] e ora da Peruzzi e De Robertis[20]), ci danno

[17] Dain, cit., 12.

[18] Id., ibid.

[19] G.Leopardi, *I canti*, edizione critica ad opera di F.Moroncini, Bologna I-II, 1927 (rist. 1961 e 1978).

[20] G.Leopardi, *Canti*, edizione critica di E.Peruzzi, con la riproduzione degli autografi, Milano 1981, G.Leopardi, *Canti*, edizione critica e autografi a cura di D.De Robertis, Milano I-II, 1984.

un'idea del continuo lavoro di lima, e possiamo talora seguire il processo generativo di un componimento, dalle prime riflessioni, nello *Zibaldone*, alla sua realizzazione e continua revisione (si pensi anche agli abbozzi italiani e latini del Pascoli conservati dalle carte di Castelvecchio). Ben diversa e ancora una volta più effimera e legata all'istante, la nostra civiltà ci propone lo scrittore intento a scrivere davanti ad un video, tagliando, cancellando e ricucendo, senza che di tutto ciò resti traccia: la variantistica del futuro dovrà cercare tra le stampe di prova, o ritrovare nei dischetti eventuali copie intermedie conservate, se non cercare di recuperare i dati nelle tracce via via cancellate. All'era del computer dobbiamo anche gli instant books, prodotti veloci e di consumo, di taglio soprattutto giornalistico, buttati sul mercato sull'onda di una notizia o nell'infuriare di uno scandalo, composti dall'oggi al domani, mettendo insieme pezzi già scritti e conservati nel computer, e quindi rapidamente stampati, consegnando all'editore un testo già in formato elettronico, senza bisogno di una complessa fase di composizione.

II. Discipline umanistiche, filologia e testi elettronici

II.1. Il filologo dalla macchina per scrivere al computer

Ma allora, concretamente, a cosa può servire il computer per il filologo[21], custode della tradizione, con il suo mestiere di retroguardia e conservazione, con tutta l'ambivalenza del termine? Il primo impiego, ovvio, è quello adombrato sin qui: sostituire la macchina per scrivere, consentendo di introdurre facilmente nei testi nuove correzioni, di numerare automaticamente le note, e fornire alla fine una stampa di gradevole lettura, grazie ai corpi di varia dimensione e ai vari stili (grassetto, corsivo, e così via). A parte programmi specificamente realizzati per i classicisti, come *Nota Bene*, *Multi-Lingual Scholar*, o il vecchio e ormai datato *ChiWriter*, che consentono di scrivere in greco classico (con accenti e spiriti), ebraico, russo, arabo, di utilizzare i vari segni diacritici necessari nelle edizioni e nelle trascrizioni fonetiche, e di gestire vari livelli di note, ovvero tre mantisse di apparato critico, sono numerosissimi i pacchetti di fonts di tutte le lingue e tipi, scalabili in tutte le dimensioni, per chi utilizza il sistema operativo Macintosh, o Windows, con i vari programmi di scrittura[22]. Indici dei nomi, dei luoghi e delle cose notevoli sono ormai realizzabili con una certa

[21] Come guide introduttive si potranno consultare Lana, cit. (il capitolo 8, "I testi scritti in greco e latino", 175-93, è consultabile anche in rete, col titolo *Strumenti informatici per le lingue classiche*, sul numero 1.1, 1995 della rivista elettronica Arachnion, [http://www.cisi.unito.it/arachne/num2/lana2.html]), T.Orlandi, *Informatica umanistica*, Roma 1990, L.Perilli, *Filologia computazionale*, Roma 1979, Solomon (ed.), cit., nonché le rassegne on line, e quindi continuamente aggiornate di J.Ruebel, *Repositories of Classical Texts or Publications* (http://studentweb.tulane.edu/~rwoods/classics.html), e soprattutto la *Rassegna degli strumenti informatici* di A.Cristofori, cit. supra, che contiene un capitolo espressamente dedicato alle guide.

[22] Cf. Lana, cit., 175 s., e la rassegna contenuta nella *Software Directory for the Classics* (http://www.centaursystems.com/soft_dir.html).

semplicità (soprattutto se concepiti sin dalla fase iniziale di realizzazione di un libro), marcando i termini che si vogliono far comparire in indice, e facendo elaborare il prodotto finale, ordinato alfabeticamente e con le caratteristiche grafiche desiderate, dalla macchina stessa. Specifici programmi di database, o integrati nel programma di scrittura (come il *Textbase* di *Nota Bene*) consentono di ottenere agevolmente una bibliografia ordinata nel modo voluto: alfabeticamente o per anno, con le citazioni all'americana, o per intero, creando una banca dati sempre aggiornabile, da cui estrarre ogni volta i titoli via via necessari. Altre caratteristiche, come la numerazione automatica dei paragrafi, con l'aggiornamento ugualmente automatico dei rinvii interni ogni volta che, nel corso della revisione, dovessero mutare, costituiscono un notevole ausilio nella redazione di un libro, sin dalla sua gestazione. Come si è visto ormai tutte le case editrici preferiscono, se addirittura non richiedono i testi in formato elettronico, così da ridurre le spese di stampa, ma non manca chi arriva sino alla fase della vera e propria impaginazione e stampa del modello "camera ready", da cui l'editore riprodurrà le varie copie[23].

Fermarsi qui, tuttavia, è forse un po' poco anche per il filologo più tradizionalista che, peraltro, anche se si limita a sfogliare indici e concordanze stampate su carta o riprodotte su microfiches (come nel caso del *Corpus Christianorum*), si deve già dichiarare largamente debitore dell'informatica che ha dato un notevolissimo impulso alla produzione di questi materiali: la realizzazione delle *Concordanze* di San Tommaso (oggi edite sia a stampa, in 56 volumi, contenenti, oltre al testo, gli indici e le concordanze complete, sia su CD-ROM[24]) da parte di padre Busa, a partire dal 1946, è infatti il primo tentativo noto di applicazione delle tecnologie informatiche alle discipline umanistiche. Prima di passare comunque ad una rapida presentazione di alcuni possibili usi del computer per l'analisi dei dati e dei testi antichi, si impone una ulteriore riflessione sulla particolare natura del testo elettronico, che non coincide semplicemente con quello su carta che riproduce.

II.2. Natura del testo elettronico

L'avvento del calcolatore nell'ambito delle discipline umanistiche ha già portato con sé molte novità, modificato vecchie abitudini, introdotto nuove prospettive, nuovi comportamenti ed anche (occorre riconoscerlo)

[23] Si vedano ad es. le edizioni di Semonide di E.Pellizer-G.Tedeschi, Roma 1990, o Lisia, *Contro Eratostene*, a c. di G.Avezzù, Padova 1992; in generale v. W.Ott, *Automatic Composition for critical Editions*, in *La pratique des ordinateurs dans la critique des textes. Colloques Internationaux du Centre National de la Recherche Scientifique, Paris 29-31 mars 1978*, Paris 1979, 237-40.

[24] Cf. *Index Thomisticus. Sancti Thomae Aquinatis operum omnium indices et concordantiae* ... digessit R.Busa, Stuttgart-Bad Canstatt 1974-1980 (a stampa) e (in CD) *Thomae Aquinatis Opera Omnia cum hypertextibus in CDROM auctore R.Busa*, Milano, Editoria Elettronica Editel 1992.

nuovi problemi. Sin dai primi tempi, inoltre, è apparso chiaro (ed è stato forse sin troppo enfatizzato) che questa *novitas* avrebbe indotto anche un'evoluzione (se non proprio una rivoluzione) metodologica nell'ambito delle scienze umane[25]. Quale fosse tuttavia la direzione o il profilo di questa evoluzione metodologica non risultava chiaro allora e ancora oggi, nonostante abbiano ormai raggiunto un'età decennale i primi contributi in questo settore nelle varie discipline, il quadro non appare del tutto perspicuo. Non che ciò possa causare troppa meraviglia: spesso i risultati del mutamento storico risultano pienamente percepibili soltanto al termine di esso, mentre chi vi prende parte ne ha solamente una intuizione confusa, viziata in genere da una prospettiva per certi versi distorta.

Ugualmente, sembra già possibile individuare sin d'ora perlomeno un nodo concettuale che svolge, e certamente svolgerà ancora in futuro, un ruolo di catalizzatore in questo processo evolutivo: il nodo che stringe insieme il nuovo concetto di testo elettronico con quello, assai più familiare (e pur sempre problematico per l'umanista), di testo in generale. Ogni disciplina umanistica in cui si è intrapresa la sperimentazione delle nuove tecnologie nel campo della ricerca ha dovuto (e sempre più dovrà d'ora in poi) considerare in via preliminare il problema del trasferimento dell'informazione testuale contenuta nelle fonti e della sua trasformazione in accordo con le esigenze del nuovo *medium* elettronico. Si tratta di rappresentare l'informazione testuale in un formato che la macchina sia in grado di riconoscere e di gestire (*machine readable form*, MRF) ed è quindi essenzialmente un problema di trasferimento dell'informazione da un sistema di rappresentazione ad un altro.

Spesso questo passaggio cruciale è stato paragonato a quello che ha accompagnato il momento epocale dell'introduzione della stampa. Senza dubbio, l'analogia riveste una qualche efficacia, ma, a mio modo di vedere, possiede soprattutto un valore evocativo, non una reale rilevanza euristica[26]. Occorre tenere presente infatti che le differenze risultano molto più

[25] Tra gli studiosi italiani che hanno dedicato maggiore attenzione alle conseguenze metodologiche dell'introduzione delle nuove teconologie nell'ambito delle discipline umanistiche è senz'altro doveroso ricordare almeno il nome di Tito Orlandi. Dei suoi diversi contributi su questo argomento, scanditi su di un arco ormai decennale, si vedano almeno T.Orlandi, *Problemi di codifica e trattamento informatico in campo filologico*, in G.Savoca (ed.), *Lessicografia, filologia e critica*, Firenze 1986, 69-82; Id, *Informatica umanistica. Riflessioni storiche e metodologiche, con due esempi*, in G.Gigliozzi (ed.), *Studi di codifica e trattamento automatico di testi*, Roma, 1987, 1-38; Id., *Informatica umanistica*, cit.; Id., *L'ambiente Unix e le applicazioni umanistiche*, Archeologia e calcolatori, 1, 1990, 237-51; Id., *Per l'informatica nella Facoltà di Lettere*, Roma 1990; Id, *Informatica umanistica: realizzazioni e prospettive*, in *Calcolatori e scienze umane*, Milano 1992, 1-22; Id.(ed.), *Discipline umanistiche e informatica. Il problema dell'integrazione (Roma, 8 ottobre 1991)*, Roma 1993; Id., *Sulla codifica delle fonti archeologiche*, Archeologia e calcolatori, 4, 1993, 27-38; Id. *Alla base dell'analisi dei testi: il problema della codifica* in M.Riciardi (ed.), *Scrivere, comunicare, apprendere con le nuove tecnologie*, Torino 1995, 65-86.

[26] Più appropriato è invece richiamare, come fa W.J.Ong, il processo che condusse dall'oralità alla scrittura; cf. supra, § I.2.

significative delle analogie. Si pensi solo a questo: nel caso dell'invenzione della stampa i testi manoscritti furono trasformati in testi stampati, e, pur con i notevoli mutamenti che questo passaggio comportò, essi tuttavia non mutarono la loro natura di testi *scritti*. Ora invece testi scritti debbono essere trasformati in testi *digitali*: non è soltanto il supporto materiale e il sistema di produzione del testo a cambiare, ma anche e soprattutto il sistema di rappresentazione dell'informazione testuale.

II.3. Dal testo scritto al testo digitale

A tal proposito, non si deve restare fuorviati dalla considerazione che il computer è in grado di rappresentare l'informazione testuale anche sotto forma di scrittura (e per giunta in maniera molto efficace da quando si sono diffusi gli ambienti grafici cosiddetti "a finestre"). Anche se molti, e in specie chi si occupa di scienze umane, si avvicinano per la prima volta ad un computer attratti dalla sua efficienza come sistema di scrittura, ciò non toglie che questa forma di rappresentazione dell'informazione testuale non sia quella primaria per il computer, bensì sia ottenuta come prodotto di un'elaborazione. Del resto è facile rendersene conto: per operare su di un testo rappresentato sullo schermo di un computer noi dobbiamo immettere comandi (attraverso la tastiera o il dispositivo di puntamento o *mouse*), non possiamo invece intervenire direttamente sullo schermo (con cancellazioni o altro) così come faremmo sul foglio di carta.

Dunque il testo elettronico non è in sé un testo scritto, e il passaggio dalla carta al computer implica un mutamento del sistema di rappresentazione. In effetti, considerato da un punto di vista genericamente astratto, l'elaboratore è una macchina che tratta con due sole categorie di oggetti: dati e istruzioni. Entrambe queste categorie sono rappresentate all'interno del sistema in forma di valori numerici secondo la numerazione binaria; in questo ambito, allora, "testo" non è nient'altro che un particolare tipo di dato, quel tipo di dato che appare codificato in forma di sequenze di caratteri, cioè di particolari combinazioni di valori binari. Dunque il trasferimento dell'informazione testuale in un formato che la macchina sia in grado di elaborare è un processo di *digitalizzazione*: si tratta cioé di far corrispondere a combinazioni di valori binari i caratteri alfabetici o in generale i grafemi delle lingue storiche. Ciò può avvenire soltanto grazie all'utilizzazione di un codice, vale a dire di un insieme (convenzionale, ma non arbitrario) di regole di corrispondenza tra caratteri alfanumerici e valori binari; dunque il processo in questione è un processo di *codifica*. La codifica delle fonti è quindi, in ultima analisi, la condizione necessaria e la chiave d'accesso ad ogni tipo di applicazione dell'informatica alle discipline umanistiche[27].

[27] Questo punto è ribadito con particolare insistenza, tra gli altri, da G.Gigliozzi (ed.), *Studi di codifica*, cit.; T.Orlandi, *Alla base dell'analisi dei testi*, cit.; F.Ciotti, *Il testo elettronico: memorizzazione, codifica ed edizione informatica del testo* in Leonardi - Morelli - Santi (edd.), cit.,

Sin qui tuttavia si potrebbe pensare che un ambito di problemi come quello delineato esuli dalle competenze e dagli interessi degli studiosi di scienze umane: è ovvio infatti che il compito di stabilire un codice per la rappresentazione dell'informazione testuale all'interno di un computer non spetta all'umanista, bensì all'informatico. L'umanista, ragionando per così dire a lume di naso, dovrebbe semplicemente preoccuparsi di applicare le regole di codifica predisposte dagli informatici per riprodurre le proprie fonti in MRF. Meglio ancora: si potrebbe essere portati a ritenere che il lavoro dell'umanista dovrebbe aver inizio solo *dopo* che le fonti siano state trasferite "all'interno del computer", attraverso le tecniche più adeguate e nel modo più accurato possibile.

Questa concezione sarebbe giustificata se il trasferimento dei testi da un sistema di rappresentazione ad un altro fosse un'operazione neutra, esente da costi in termini, per così dire, di entropia informativa. Ma proprio le scienze umane, e in primo luogo la filologia e la semiotica, ci insegnano che non è così. Ogni codifica dell'informazione in un nuovo *medium* implica una diversa segmentazione della materia significante, sia essa il continuo sonoro o lo spazio grafico o la serie infinita dei numeri binari, e questa a sua volta porta con sé una ridefinizione dei tratti pertinenti del contenuto informativo, che quindi non rimane del tutto invariante nel passaggio da un sistema semiotico ad un altro. Questo passaggio non avviene dunque secondo le modalità del calco o dell'impronta con cui un originale imprime la propria rappresentazione su di una diversa materia, bensì piuttosto come la proiezione cartografica di una superficie curva su una superficie piana, in cui non tutti i punti della prima hanno uno ed un solo corrispondente nella seconda. Non a caso l'espressione idiomatica che si usa in informatica in questo caso è *mapping*.

Fuori di metafora, se ne conclude che nel passaggio dalla carta al computer l'inevitabile perdita di informazione testuale richiede di essere minimizzata, e comunque preventivata, con consapevolezza metodologica e quindi occorre che il processo sia controllato non solo da chi è esperto del *medium*, ma anche da chi conosce i testi, cioè, nel caso delle fonti storiche, dallo studioso di scienze umane.

II.4. Codifica primaria e secondaria

La necessità di un controllo sulla fase di codifica emerge ancor più chiaramente quando si consideri che un solo livello di codifica non è certo sufficiente per riprodurre in maniera adeguata il contenuto informativo di una fonte storica. È esperienza comune, per il ricercatore che si occupa del patrimonio testuale del passato, che un testo non è informativo solo per quello che esplicitamente *dice*, ma anche per quello che, intenzionalmente o meno,

213-30 e Id., *Testi elettronici e banche dati testuali: problemi teorici e tecnologie*, Schede Umanistiche n.s. 1995, 2, 147-78.

esibisce di sé. Ad esempio, a chi conosce le regole di composizione epistolare in un particolare ambiente storico può risultare chiaro dalla sola disposizione sulla pagina quale sia il nome del destinatario e quale quello del mittente di una lettera, anche se ciò non fosse dichiarato in modo esplicito. Per citare un ulteriore esempio, di portata ancor più generale, si pensi alla parte di testo che si suole identificare come *explicit* in filologia. Essa non è identificata come tale da null'altro che dalla sua collocazione finale, e non da una precisa indicazione. Anzi in genere nei vari testimoni è a sua volta seguita da una nota finale, il *colophon*, spesso introdotta dalla formula "Explicit", che non fa parte del testo così come lo ha redatto l'autore e che può recare, oltre alla paternità e al titolo dell'opera, anche indicazioni relative alle circostanze di copia, come il nome del copista e la datazione. Eppure la distinzione tra *explicit* e colofone, benché non dichiarata nel testo stesso, è immediatamente riconoscibile da un esperto che prenda in esame il documento in questione.

I due casi menzionati, a titolo soltanto esemplificativo, dovrebbero valere a ricordare come per "catturare" tutta l'informazione testuale contenuta in un documento non sia sufficiente riprodurne la mera sequenza dei caratteri, ivi compresi gli spazi bianchi tra una parola e l'altra. A volte, anzi spesso nel caso delle fonti storiche (che come tali costituiscono documenti testuali unici e, per così dire, autoreferenziali), anche la disposizione tipografica o la *mise en page* di un testo contribuiscono a fornire informazioni direttamente rilevanti per la comprensione del testo stesso[28]. Dunque alla codifica che si può definire primaria, che riproduce in combinazioni di valori binari la sequenza dei caratteri alfanumerici di un testo, si dovrà aggiungere una "seconda codifica" che ci metta in grado di recuperare tutta l'informazione che un testo non *dichiara* esplicitamente, ma che altrettanto esplicitamente ed univocamente *esibisce* o dimostra.

Di questo aspetto del problema della costituzione del testo elettronico si è già da tempo occupato quel settore dell'informatica che cura la predisposizione dei sistemi di scrittura, vale a dire il settore del *text processing* o del *word processing*. Ma le soluzioni sin qui più note e diffuse, vale a dire quelle commerciali che si trovano praticamente realizzate nei più comuni programmi di videoscrittura, per quanto indubbiamente efficaci, hanno un difetto sostanziale dal punto di vista delle scienze umane: esse pongono una seria ipoteca tecnologica e commerciale sul processo di riproduzione elettronica delle fonti[29].

[28] Occorre fare attenzione a distinguere in questo caso l'informazione *testuale*, cioè desumibile direttamente dal testo stesso attraverso la decifrazione del suo contenuto secondo le regole convenzionali che ne hanno guidato la composizione (regole della lingua, del genere letterario ecc.), da tutta la restante informazione che dal medesimo testo si può trarre per via di inferenza sulla base di conoscenze fattuali *esterne* al testo, come ad es. dal fatto che una lettera reca una firma e una datazione cronica e topica si può derivare l'informazione che il suo autore in quel giorno si trovava in quel luogo.

[29] Sul problema di quella che viene qui definita "codifica secondaria" dei testi cf. J.S.Coombs - A.H.Renear - S.J.DeRose, *Markup Systems and the Future of Scholarly Text Processing*,

Il problema del trattamento e del recupero di quella parte dell'informazione testuale che si aggiunge alla pura sequenza dei caratteri è infatti risolto dai programmi di videoscrittura attraverso la realizzazione di un sistema di marcatura del testo (*markup language*) che contiene le indicazioni necessarie per la gestione o la riproduzione dell'impaginazione e, in generale, delle caratteristiche tipografiche del testo. A partire dall'introduzione dei programmi di scrittura "interattivi" (in cui cioè è possibile vedere immediatamente sullo schermo il risultato dei comandi di impaginazione), l'immissione dei codici di marcatura è divenuta completamente "trasparente" per l'utente, che viene messo in condizione di controllare la disposizione del testo direttamente attraverso istruzioni impartite mediante combinazioni di tasti o selezionate sullo schermo mediante l'azione del puntatore. Ciò che si ottiene, tuttavia, quando si immettono tali comandi è di aggiungere alla sequenza dei caratteri (ottenuta attraverso la "codifica primaria" che fa corrispondere valori binari ai caratteri reali) una "seconda codifica" contenente le istruzioni che sono interpretabili dal modulo di visualizzazione del programma per disporre il testo sullo schermo secondo le indicazioni dell'utente e dal modulo di stampa per ottenere gli stesi effetti sul foglio di carta.

Questo tipo di codifica secondaria ha tre caratteristiche che la rendono inadatta agli scopi che si presumono tipici di un trattamento elettronico delle fonti: a) essa è *procedurale*, consiste cioè di indicazioni che debbono essere eseguite come istruzioni dal programma per ottenere un determinato effetto di presentazione del testo; b) essa è dunque *specifica* di un determinato programma, cioè è significativa solo se l'elaborazione è compiuta dallo stesso programma che ha generato la codifica, e infine c) essa è relativa alle caratteristiche di *presentazione* del testo, e non alla sua composizione intrinseca. In conseguenza di (c), allora, noi non possiamo con questo tipo di codifica registrare l'informazione che una determinata parte del testo riporta, ad es., il titolo, ma solo che essa deve essere, poniamo, realizzata in grassetto e corpo 12. Ancor più limitativo è che in conseguenza di (b), come è esperienza comune di molti semplici utenti, uno stesso testo composto da un programma di videoscrittura non può essere completamente riprodotto con l'ausilio di un diverso programma.

D'altro canto, sia (b) che (c) sono dirette conseguenze di (a), cioè della scelta di realizzare un sistema di marcatura mescolando insieme dati ed istruzioni, ed è proprio partendo dal rovesciamento di questa opzione che è

Communications of the Association for Computing Machinery, 30, 1987, 933-947; S.J.DeRose - D.Durand - E.Mylonas - A.H.Renear, *What is Text, Really?*, Journal of Computing in Higher Education, 1.2, 1990, 3-26; C.M.Sperberg McQueen, *Text in the Electronic Age: Textual Study and Text Encoding, with Examples from Medieval Texts*, LLC, 6.1, 1991, 35-41; S.J. DeRose, *Markup Systems in the Present*, in G.Landow - P.Delany (edd.), *The Digital Word: Text-Based Computing in the Humanities*, Cambridge 1993, 119-35 e T.Orlandi, *Teoria e prassi della codifica dei manoscritti*, in *Gli Zibaldoni di Boccaccio: memoria, scrittura, riscrittura*, di prossima pubblicazione, ma accessibile come http://rmcisadu.let.uniroma1.it/~orlandi/encod.html

stato possibile progettare una soluzione alternativa al problema della "codifica secondaria" del testo, che sia in grado di neutralizzare gli inconvenienti dei sistemi commerciali. Questa soluzione è stata realizzata sotto forma di uno standard internazionale che definisce le regole sintattiche per la costruzione di linguaggi di marcatura di tipo non procedurale e specifico, ma dichiarativo e generico. Lo standard in questione — lo SGML (Standard Generalized Markup Language - ISO 8879:1986) — offre la possibilità di definire in maniera formale un sistema di codici (o etichette, *tags*) attraverso i quali si possono descrivere tutte le caratteristiche proprie di un testo e di tutte le sue parti componenti[30]. Non avendo carattere procedurale, cioè non consistendo di istruzioni, ma di dichiarazioni, SGML non è legato ad alcun tipo di hardware o di software ed è quindi indipendente dalle incognite dello sviluppo tecnologico del settore. Inoltre, pur potendo essere utilizzato anche per descrivere le caratteristiche di presentazione del testo, esso è stato disegnato esplicitamente allo scopo di descrivere un testo a partire dalla sua struttura compositiva, cioè attraverso l'individuazione e la definizione delle parti del testo e delle relazioni che intercorrono tra queste parti.

Da questo punto di vista l'applicazione di SGML ad un determinato genere di testi, ad es. ad una tipologia di fonti, richiede esplicitamente la partecipazione, in fase di progettazione dello specifico sistema di codifica da utilizzare, di esperti delle fonti in questione, che siano in grado di indicarne la struttura in base alle regole di composizione storicamente accertate per quel tipo di testi ed in funzione delle esigenze di analisi scientifica cui le fonti potranno essere sottoposte (edizioni critiche, analisi stilistiche, rilevamenti statistici, raccolta di dati prosopografici o di natura economica, e quant'altro). Per questo già da un quinquennio è stata stabilita una commissione internazionale che, attraverso la consultazione di comitati di esperti dei vari settori delle scienze umane, ha curato la redazione di un corpo di raccomandazioni ("Guidelines") da seguire per l'applicazione della codifica SGML ai testi di interesse umanistico. Questa iniziativa — denominata TEI (Text Encoding Initiative) — ha così posto in essere un quadro di riferimento internazionale all'interno del quale le singole campagne e i singoli progetti di codifica delle fonti potranno utilmente coordinarsi in maniera tale che il

[30] Per un'introduzione allo SGML si vedano International Organization for Standardization. 1986. ISO 8879. *Information Processing - Text and Office Information Systems - Standard Generalized Markup Language*; M.Bryan, *SGML: An Author's Guide to the Standard Generalized Markup Language*, Wokingham - Reading - New York 1988; Ch.Goldfarb, *The SGML Handbook*, Oxford 1990; L.Burnard, *What is SGML and how does it help?*, in D.Greenstein (ed.), *Modelling Historical Data*, St. Katharinen 1991, 65-79; E.Van Herwijnen, *Practical SGML*, Dordrecht 1994[2]. Le informazioni più aggiornate si possono trovare alla pagina WWW http://www.sil.org/sgml/sgml.html

lavoro impiegato per trasferire un determinato insieme di testi sia poi immediatamente fruibile dall'intera comunità degli studiosi interessati[31].

II.5. Filologia del testo elettronico

Può forse cominciare a farsi strada nella mente del lettore il sospetto che, concepita secondo le modalità e le caratteristiche sopra descritte, l'operazione di codifica di una fonte per il suo trasferimento nel *medium* elettronico sia un'operazione in linea di principio filologica almeno altrettanto che informatica. In fondo, si tratta pur sempre di rappresentare in nuovo *medium* un testo, ricostruendone la forma originale e dando conto di tutte le vicissitudini storiche che tale forma ha conosciuto. Codificare un testo per immetterlo nel computer è dunque in ogni caso fare un'edizione elettronica: si pone quindi il problema di stabilire le condizioni e i requisiti affinché un'edizione elettronica possa definirsi critica. Dovrebbe risultare evidente da quanto si è detto sin qui che il concetto di "edizione elettronica critica" non può che porsi come un ampliamento, e non come un superamento, del concetto di "edizione critica" tradizionale. Ciò significa che i criteri tradizionali di criticità non possono in ogni caso essere disattesi, ma semmai che ad essi se ne aggiungono di nuovi: ad esempio, che la scelta e l'applicazione del sistema di codifica secondaria sia tale da fornire una rappresentazione adeguata della struttura compositiva del testo in questione e da rendere possibili tutte quelle analisi e operazioni sul testo che la comunità degli studiosi giudica sensate e desiderabili.

Tuttavia, già a questo livello di generica indicazione di alcune possibili linee di sviluppo della nozione di testo elettronico, è possibile intravvedere alcuni nodi metodologici di non facile soluzione: se infatti da un lato, come si è visto, la codifica è uno strumento di rappresentazione dei *risultati* di uno sforzo di ricostruzione e di analisi critica del testo, dall'altro essa si pone certamente anche come una *condizione* preliminare da soddisfare ancor prima di poter cominciare il lavoro critico. In altri termini, la codifica del testo, se deve essere adeguata, ne presuppone la piena comprensione, ma quest'ultima si pone come un obbiettivo dello sforzo critico che, almeno idealmente, è sempre *in progress*. Si può considerare questo problema come l'ennesima riproposizione della condizione epistemologica permanente della ricerca storica: quella del circolo ermeneutico, ma si può anche affrontare la questione da un punto di vista pratico e richiedere ad esempio che il sistema di codifica sia dinamico, cioè modificabile e rivedibile a seconda degli stadi

[31] Cfr. L.Burnard - C.M.Sperberg-McQueen, *TEI Lite: An Introduction to Text Encoding for Interchange*, (TEI Document No: TEI U 5), [http://www.uic.edu/orgs/tei/intros/teiu5.html] disponibile anche come L.Burnard - C.M.Sperberg-McQueen, *Encoding for Interchange: An Introduction to the TEI*, TEI (08.07.1994); P.Robinson, *The Transcription of Primary Textual Sources Using SGML*, Oxford, Office for Humanities Communication Publications, 1994 e C.M. Sperberg - McQueen - L.Burnard, *Guidelines for Electronic Text Encoding and Interchange (TEI P3)*, Chicago, Ill. - Oxford, ACH-ACL-ALLC 1994.

della ricerca. Per questo motivo, se in ogni caso esso rispecchia le ipotesi di lavoro che il ricercatore di volta in volta considera valide sul testo, dovrà allora essere sempre ben chiara la distinzione tra ciò che è codifica e ciò che è dato, il che equivale a dire tra il dato e la sua interpretazione[32].

In questa direzione, la natura del *medium* elettronico risulta certamente favorevole: la principale caratteristica della gestione dell'informazione in formato digitale, rispetto al mondo tradizionale della carta stampata e del libro, risiede proprio nella sua intrinseca dinamicità e versatilità. Già da più parti è stato indicato, ad es., come questa caratteristica possa essere d'aiuto per il filologo elettronico nel trattamento dei casi normalmente considerati devianti dalla stemmatica tradizionale, come quelli dei testi fluidi o fortemente interpolati[33]. La capacità di elaborazione delle informazioni propria dei sistemi informatici rende agevole la possibilità di rappresentare efficacemente e dinamicamente stati evolutivi o alternativi del testo. Varianti ed accidenti di trasmissione di incidenza maggiore possono essere mostrati in primo piano in tutto il loro valore di alternativa rispetto al testo tradito, e non meramente relegati in apparato o in appendice. Il ruolo giocato dalla tradizione indiretta nella costituzione del testo può essere pienamente valorizzato attraverso la realizzazione di una rete di collegamenti immediatamente attivabili alle testimonianze rilevanti[34]. Lo stesso può valere per il valore da attribuire alle fonti cui attinge l'autore nel comporre la sua opera. La stessa disposizione materiale del testo nei vari testimoni può essere efficacemente mostrata attraverso immagini[35].

Tutte queste possibilità — ed altre ancora: il pensiero corre subito alle nuove parole d'ordine della gestione elettronica dell'informazione, ipertesto e multimedialità, e alla mitologia culturale che se ne è subito impossessata —

[32] È questo un punto di vista propugnato soprattutto da Manfred Thaller con la sua proposta di un "source-oriented data processing". Cf. M.Thaller, *Automation on Parnassus. CLIO - A Databank Oriented System for Historians*, in: Historical Social Research. Historische Sozialforschung, 15, 1980, 40-65; Id., *The Need for a Theory of Historical Computing*, in P.Denley - S.Fogelvik - C.Harvey (edd.), *History and Computing II*, Manchester 1989, 2-11; Id., *The Historical Workstation Project*, CHum, 25, 1991, 2-3, 149-62; Id., *Historical Information Science: Is There Such a Thing? New Comments on an Old Idea*, in Orlandi (ed.), *Discipline umanistiche e informatica*, cit., 51-86; Id., *The Processing of Manuscripts*, in M. Thaller (ed.), *Images and Manuscripts in Historical Computing*, St. Katharinen 1992, 41-72; Id., Κλειω. *A Database System*, St. Katharinen 1993; Id., *What is Source-oriented Data Processing? What is Historical Computer Science?*, Historical Methods, 28, 1995.

[33] Per i problemi relativi, v. comunque più oltre, § III.2.3.

[34] A.Cristofori (al quale direttamente - e alla sua *Rassegna* indirettamente - siamo debitori di preziose precisazioni e integrazioni) ci segnala a questo proposito un'edizione critica on line dell'*Eneide* virgiliana, costruita in modo da poter consultare le lezioni di quattro manoscritti e il testo critico fondato sulla loro base (http://ccat.sas.upenn.edu:80/~sbb).

[35] Cfr. in particolare su questo punto D.Buzzetti - A.Tabarroni, *Informatica e critica del testo: il caso di una tradizione 'fluida'*, Schede Umanistiche, 1991, 2, 185-93 e D.Buzzetti, *Image Processing and the Study of Manuscript Textual Traditions*, in J.Fikfak - G.Jaritz (ed.), *Images Processing in History: Towards Open Systems*, St. Katharinen 1993.

sono da un lato come una promessa di nuovi orizzonti per lo studioso di scienze umane, dall'altro però gli pongono una sfida, in termini di rigore e di coerenza scientifica, che è soprattutto di natura metodologica e che, in quanto tale, già di per sé costituisce una garanzia di rinnovamento cui sarebbe imperdonabile tentare di sottrarsi.

III. L'analisi dei dati: percorsi di ricerca

III.1. L'analisi quantitativa

Già dagli inizi degli anni '60 la linguistica ha intravisto le grandi possibilità di verifica ed esame dei dati che sono consentite dall'informatica, sviluppando un metodo d'indagine che, d'altra parte, era già proprio della filologia tradizionale: "Non si vede perché il metodo statistico, che si è dimostrato suscettibile di svolgimenti originali nell'ambito di molte discipline umane quali la psicologia sperimentale, l'economia politica ecc., non possa dare buoni frutti se applicato al linguaggio, tenendo presente, da una parte, che la statistica fornisce il metodo che ci consente di analizzare e di interpretare le enumerazioni, dall'altra parte, che nel linguaggio non tutto è suscettibile di una *analisi* e di una *interpretazione* fondata su di un metodo matematico, ma pure che nemmeno tutto si sottrae ad esso", scriveva nel 1961 il glottologo L.Heilmann[36], ben conscio delle possibilità, ma anche dei limiti del mezzo informatico, la cui indagine è sempre basata su "dati che abbiano fra loro dei rapporti logici definibili da qualche tipo di logica formale, e li gestisce unicamente mediante quel tipo di logica"[37], mentre linguaggio, codice letterario e tradizione testuale non sono sempre uniformabili ad una logica (si pensi agli studi sul lapsus freudiano nella teoria dell'errore, con le conseguenze che ciò ha su un'ipotesi di collazione meccanica)[38].

L'analisi statistica dei dati, incentivata dai tentativi di traduzione automatica, ha finito tuttavia per diventare una scienza autonoma, l'analisi dei testi assistita dai computer, con applicazioni soprattutto nel campo delle lingue vive. Si tratta di un tipo di ricerca complessa e attualmente in crisi, per le critiche che le vengono rivolte sia dall'esterno, che dall'interno: *The Quantitative Analysis of Text. Why has it failed?* si interroga significativamente D.L.Clayman[39], spiegando poi che con questa frase intende che "it has not

[36] L.Heilmann, *Statistica linguistica e critica del testo,* in *Studi e problemi di critica testuale. Convegno di studi di Filologia italiana nel Centenario della Commissione per i testi di Lingua (7-9 aprile 1960)*, Bologna 1961 [rist. anast. 1982], 173 s. [rist. in *Letteratura e filologia* a cura di B.Basile, Bologna 1977, 88].

[37] Orlandi, *Informatica umanistica*, cit., 109.

[38] Cf. infra, § III.2.3.

[39] D.L.Clayman, *The Quantitative Analysis of Text: Why has it failed?* in S.Cacaly-G.Losfeld (edd.), *Sciences historiques, sciences du passé et nouvelles technologies d'information. Bilan et évaluation. Actes du Congrès International de Lille (16-17-18 mars 1989)*, Lille 1990, 37-47.

produced results that other scholars consider to be reliable or useful. That is not to say that our work so far has been in error, or in vain, but rather that we have been not yet succeeded in convincing very many people to believe in it"[40]. Mentre dunque gli specialisti della materia si lamentano di sentirsi emarginati (anche accademicamente) e confinati su alcune specifiche riviste (come «Computer and Humanities», o «Literary and Linguistic Computing»), specularmente viene loro rinfacciato il linguaggio eccessivamente tecnico ed esoterico: "the results have been expressed in a complex jargon that makes them unintelligible to most working historians"[41], commentava ad esempio J.Popkins in una rassegna degli studi recenti sulla rivoluzione francese, e quindi esaminando anche alcune analisi del Centre de Lexicologie Politique de Saint Cloud. Il problema della competenza esiste indubbiamente, osservava per le lingue classiche Évrard che "dans les études quantitatives, il arrive trop souvent que le philologue ou le linguiste ne soit pas assez versé en statistique et, de ce fait, s'aventure dans des applications inacceptable. Trop souvent, en particulier, on considère comme équivalent telle propriété quantitative (homogénéité) et tel trait qualitatif (authenticité, identité d'auteur)"[42]. Il mezzo informatico tuttavia è strumento utilissimo per lo studio del linguaggio, perché in grado di elaborare grandi masse di dati. Il problema, come si è detto, è quello dell'uso dei dati: alcune ingenuità non certo veniali hanno finito per screditare agli occhi dei filologi più tradizionalisti l'intero ricorso a queste tecnologie. Valgano per tutti un paio di esempi: Évrard stesso ricorda severamente ("de tels travaux sont une honte et l'on devrait les boycotter"[43]) come qualcuno, avendo prodotto una concordanza non lemmatizzata, in cui non ha distinto gli omografi, ha stampato poi ugualmente alla fine del volume una lista di frequenze in cui, ad esempio, si apprende quante sono le ricorrenze di *quam*, senza poter dividere il relativo dall'avverbio e dalla congiunzione.

Quanto i metodi statistici siano pericolosi per chi non possegga adeguate competenze, lo dimostra ancora un articolo di J.B.Booth, apparso nel 1976 sulla «Classical Quarterly»[44]: un paio d'anni prima M.Pope[45] aveva

[40] Clayman, cit., 41.

[41] J.Popkins, *Recent West German Work on the French Revolution*, JMH 59, 1987, 748, ricordato da M.Olsen, *Signs, Symbols and Discourses: A New Direction for Computer-Aided Literature Studies*, CHum 27.5-6, 1993-94, 310. Si veda quanto diceva già, con la consueta arguzia A.Dain, cit., 178, a proposito di un tentativo di classificazione matematica fondata sul calcolo combinatorio di manoscritti dei *Vangeli* (per cui v. infra), pubblicata nel 1958 da Mgr. de Solages: "Brillante réussite! Mais n'entre pas qui veut dans ce domaine interdit aux philologues moyens que nous sommes presque tous. *Lasciate ogni speranza*".

[42] E.Évrard, *Informatique appliquée aux langues et aux textes. Comment? Pourquoi?*, in Cacaly-Losfeld (edd.), cit., 24. Si veda anche, per il dibattito sull'identificazione dell'autore dell'*Historia Augusta*, D.Sansone, *The Computer and the 'Historia Augusta'*, JRS 80, 1990, 174-77.

[43] Id., ibid.

[44] N.B.Booth, *Zeus Hypsistos Megistos: an Argument for Enclitic που in Aeschylus, Agamemnon 182*, CQ 26, 1976, 220-28.

[45] M.Pope, *Merciful Heavens*, JHS 94, 1974, 100-13.

difeso la lezione βιαίως del codice **M** nell'*Agamennone* di Eschilo, v. 182 (dove Turnèbe corregge in βίαιος), portando, tra gli altri argomenti, quello della scarsa probabilità statistica di un mutamento di o in ω, accompagnato dallo spostamento di accento[46]. A parte alcuni minori errori di calcolo, Booth rilevava che la formula che consente di stimare la probabilità della corruzione è fondata su una costante che deve ritornare in tutti i casi esaminati, mentre "there is far higher probability of omicron being corrupted to omega in a case where the new text still makes a grammatically correct word than when it doesn't; there is even higher probability of omicron going to omega when the new text is not only grammatically correct but also makes something like sense in the context"[47]. Se ne conclude che è praticamente impossibile quantificare le probabilità di questa corruzione nel testo, "and second, that arguments of this kind should be used with extreme circumspection and regard for the individual features of the individual case"[48].

Dunque, tenendo conto dell'alta specializzazione e della competenza necessaria per questo tipo di studi, che, fra l'altro ora tendono a indirizzarsi verso la direzione ambiziosa della filosofia del linguaggio e dell'analisi delle strutture cognitive[49], in prospettiva di un'applicazione nell'intelligenza artificiale, pare più opportuno tenere fuori dall'ambito di questa introduzione gli studi più propriamente statistici, e considerare in che modo il filologo, senza snaturare il suo lavoro, e senza avere bisogno di competenze che solitamente non ha, possa servirsi dei nuovi strumenti informatici, con un'ottica immediata e pragmatica, anche se qualcuno potrà considerarla riduttiva[50].

III.2. Banche dati testuali per il filologo classico e medievale

III.2.0. Premessa

Come si diceva il primo strumento di lavoro per il filologo sono i testi, e numerosissime sono ormai le banche testuali, come ad esempio il CD-ROM del *Thesaurus Linguae Graecae*, che contiene quasi tutti i testi dalle origini al 600 d.C., e sta colmando il periodo compreso tra il 600 e il 1450[51], con un

[46] Pope pare dimenticare tuttavia che nell'antico alfabeto attico in cui scriveva Eschilo non si distingueva o da ω, né tanto meno erano indicati accenti e spiriti, dunque βίαιος e βιαίως hanno la stessa probabilità: sarà la considerazione del contesto e dell'uso eschileo a dirimere il problema.

[47] Booth, cit., 223.

[48] Id., ibid.

[49] Cf. C.Henry, *The Surface of Language and Humanities Computing*, CHum 27.5-6, 1993-94, 315-22, S.N.Matsuba, *Finding the Range: Linguistic Analysis and Its Role in Computer-Assisted Literary Study*, CHum 27.5-6, 331-40 (e in generale l'intero fascicolo della rivista, uscito col titolo *A New Direction for Literary Studies*, curato da P.A.Fortier).

[50] Un'introduzione alle indagini stilometriche, basate sulle frequenze, si può comunque trovare in Lana, cit., 75 ss.; cf. dello stesso *Studiare lo stile dei testi per mezzo delle frequenze*, Arachnion 1.2, 1995 (http://www.cisi.unito.it/arachne/num2/lana2.html).

[51] Cf. T.Brunner, *Thesaurus Linguae Graecae*, in Cacaly - Losfeld (edd.), cit., 54, e la sua "Preface" a L.Berkowitz - K.A.Squitier, *Thesaurus Linguae Graecae. Canon of Greek Authors*

indice che consente rapidamente di verificare in quale autore si trovano le varie forme, senza effettuare una ricerca su tutto il CD: si tratta di uno strumento non immune da errori o inconvenienti, come quelli inevitabili di digitazione da parte dei moderni "scribi" che hanno trascritto il testo dal supporto cartaceo a quello elettronico, o quelli dovuti ai progressi della ricerca che ha portato al costituirsi di nuove edizioni critiche, ma ormai indispensabile per qualsiasi ricerca di lessico, come rilevano ad esempio i redattori del *Diccionario Griego-Español*[52], ed anche sui singoli autori, mentre già si è detto dei testi contenuti nel videodisco *Perseus 2.0.*[53]. A queste due banche dati si affiancano le raccolte di papiri ed *ostraka,* nonché di iscrizioni contenute nel cosiddetto *PHI6*[54] ed un gran numero di autori o di opere singole, spesso con traduzione, o con bibliografia, reperibili nella rete[55]. Gli autori latini di età classica sono poi disponibili in vario formato[56], mentre

and Works, Oxford 1990[3], X, che costituisce una guida indispensabile al CD, indicandone contenuto, sigle ed edizioni impiegate: v. ora anche gli interventi degli stessi Brunner e Berkowitz nel volume miscellaneo curato da Solomon, cit., 10-61. Per consultare il *TLG*, come il *PHI5* e *PHI6* (su cui v. infra), sono in commercio programmi di editing e di interrogazione, per cui v. Lana, cit., 185-88, T.Del Vecchio, *CD-ROM Latin Texts Large Archives*, in Papers on Grammar 4, 1994, 4,1994, 85-98; C.Schäfer, *Computer und Antike Texte. Wortrecherche, Konkordanz und Indexerstellung mit Volltextdatenbanken*, St. Katharinen 1993; informazioni sul contenuto, sul software (Dos, Windows e Mac) e sulle modalità di ordinazione sono disponibili anche via Internet (gopher://tlg.cwis.uci.edu:7011/1, o, meglio: http://www.tlg.uci.edu/~tlg).

52 Cf. F.R.Adrados - J.Rodríguez Somolinos, *The TLG Data Bank, the DGE and Greek Lexicography*, Emerita 62, 1994, 241-51. Anche per il *TLG* si dovrà mantenere la cautela metodologica di cui si è parlato prima, giacché non solo è incompleto, ma registra il solo testo e non gli apparati: esso affianca, ma non sostituisce i lessici tradizionali, come ricorda in una nota dal titolo provocatorio E.Degani, *Il mostro di Irvine*, Eikasmos 3, 1992, 276 s.

53 Cf. supra, § I.I, e n.1; attualmente è pianificata la realizzazione di un'analoga banca dati per il latino, le cui caratteristiche sono descritte sulla rete (http//www.perseus.tufts.edu/neh.ann.html): dovrà comprendere tra l'altro un'antologia di testi, un'edizione elettronica del Lewis-Short, programmi per leggere il *PHI5*, [vd. n.1], commenti classici, fotografie relative all'arte romana, una bibliografia di traduzioni.

54 Il nome deriva dal "Packard Humanities Institute", 300 Second Street, Los Altos CA 94022, che lo distribuisce in licenza; per una descrizione analitica del contenuto, cf. C.Schäfer, cit., 23-42, e, in generale, per le banche dati epigrafiche, v. M.Fell - C.Schäfer - L.Wierschowski (edd.), *Datenbanken in der Alten Geschichte*, St. Katharinen 1994, 1-49; i papiri sono descritti nella *Checklist of Editions of Greek and Latin Papyri, Ostraka and Tablets*, curata da J.F.Oates (e corrispondente al *Canon* del *TLG*), disponibile anche in rete, presso il *Duke Papyrus Archive* (cf. dello stesso Oates l'intervento in Solomon (ed.), cit., 62-72). È in preparazione un *PHI7*, per le fonti epigrafiche, cf. http://132.236.125.30

55 I testi vanno da Eschine ad Omero (con una trascrizione elettronica della recente edizione di van Thiel; mentre è anche disponibile su disco l'elenco dei papiri di Omero, a c. di F.Sutton), ai tragici, ad Apollonio Rodio e Pindaro, per cui è disponibile anche una bibliografia: cf. gli elenchi nelle rassegne citate di Ruebel e Cristofori.

56 I testi latini dall'età arcaica fino al II sec. d.C., con qualche eccezione più tarda, sono contenuti nel *PHI5* (il numero lo distingue dal *PHI6*, per cui v. supra): un elenco a stampa degli autori in Schäfer, cit., 17-22. Altre raccolte sono *Aureae Latinitatis Bibliotheca* e *Poesis* (quest'ultima dedicata agli autori poetici, notevole per velocità e per la presenza di autori anche tardi, e comunque non compresi in altre banche dati), prodotte dalla casa editrice Zanichelli di Bologna,

sono ugualmente disponibili su CD, *Corpus Christianorum*[57], *Patrologia Latina*[58], e parte dei *Monumenta Germaniae Historica*[59], nonché le fonti di diritto romano e una parte dei *CIL*[60]. Questi testi elettronici sono codificati in modo da contenere i riferimenti necessari per il reperimento dei passi (libri, capitoli, righe) e addirittura, come scrive T.F.Brunner, responsabile del progetto del *TLG*, "each TLG text is an electronic mirror image of the source edition from which it derives"[61]: sono quindi possibili ricerche su un autore (ma in questo caso la ricerca su un indice cartaceo, se disponibile, è altrettanto, se non più comoda), o su un gruppo di autori, o su tutto un *corpus*, ovvero di una radice, di gruppi di parole, o di combinazioni di parole. Si possono ottenere facilmente concordanze, indici ed indici inversi, ovvero, con programmi più sofisticati, ricercare ripetizioni lessicali o confrontare il lessico di più autori. Esistono anche altri tipi di banche dati, lemmatizzate, come quelle prodotte dal LASLA[62], che ad ogni termine associano anche i

mentre testi singoli di autori sono disponibili per altra via, cf. Lana, cit., 179-81, e le indicazioni contenute nelle rassegne indicate sopra, n.21. Per i testi del LASLA, v. infra.

57 Il cosiddetto *CLCLT-2 (Cetedoc Library of Christian Latin Texts)*, 1991[1], 1994[2], distribuito da Brepols, contiene, oltre alla *series latina* e alla *continuatio medievalis*, numerosi autori pubblicati negli *CSEL*, *SC* e *PL*, proponendosi come concorrente della *Patrologia Latina Database*; per una rassegna, v. D.Béguin, *Le CLCLT de Brepols: la littérature latine patristique et médiévale interrogeable sur CD-ROM*, RHT 24, 1994, 485-93.

58 La *Patrologia Latina Database* è prodotta da Chadwick-Healey, Cambridge, in 4 CD, in una versione ancora non completa e che pone ancora qualche problema di consultazione (accanto a quello testuale, della relativa attendibilità del Migne).

59 Si tratta del *eMGH-1*, prodotto da Brepols, che intende progressivamente, con aggiornamenti annuali, completare la raccolta: lo stesso editore produce l'*Archive of Celtic-Latin Literature (ACLL)*, per i testi patristici e medievali latini di area celtica.

60 La *Bibliotheca Iuris Antiqui*, Catania, Libreria Editrice Torre, 1994, consta di tre archivi: il primo, *Fontes*, è una versione di *Romtext*, edizione elettronica delle principali fonti giuridiche romane, realizzata dalla Università di Linz (e disponibile anche per l'installazione su disco rigido, cf. J.Menner, *Die Datenbank ROMTEXT: Quellen des römischen Rechts*, in Fell - Schäfer - Wierschowski (edd.), cit., 160-64), e quindi contiene il *Corpus Iuris Civilis*, leggi e senatoconsulti dalle *FIRA I*, *Collatio*, *Consultatio*, *Fragmenta Vaticana*, le *Institutiones* di Gaio, l'*Epitome Gai*, i *Fragmenta Augustodunensia*, le *Pauli Sententiae*, con l'*Interpretatio* visigotica, l'*Epitome Ulpiani*, l'*Edictum Theoderici* e la *Lex Romana Burgundionum*; le *Constitutiones Sirmondianae*, le *Novellae posttheodosianae* contenute nel *Breviarium Alaricianum*. Il secondo archivio, *Opera*, comprende la bibliografia relativa al diritto romano dal 1950 al 1989, mentre il terzo, *Thesaurus*, è un vocabolario di circa 9000 termini che sono patrimonio specifico delle discipline relative ai diritti dell'antichità, con una classificazione "a catena" che consente di individuare termini legati, sinonimi e contrari. *Epigraph. A database of Roman Inscriptions* (CD a c. di E.J.Jory, prodotto da Perth, The University of Western Australia, disponibile per PC e Mac), contiene il vol. VI del *CIL*, ovvero le iscrizioni di Roma: cf. la descrizione via http://www.arts.uwa.edu.au/Classics/EpiGraph1.html

61 Brunner, cit., in Cacaly - Losfeld (edd.), cit., 53.

62 Per una informazione generale sugli autori e sulle caratteristiche, v. G.Purnelle, *Exploitations philologiques et historiques d'une banque de données de la langue latine*, in Cacaly- Losfeld (edd.), cit., 63-71, J.Denooz, *Le traitement des textes latins, grecs et français au laboratoire d'Analyse Statistique des Langues Anciennes*, Revista Universidad Complutense 25, 1976, 143-67, e, più recentemente, l'elenco di "Bases de données et publications" in RIS 25, 1989, 213-21,

suoi dati grammaticali, morfologici, prosodici: a partire da questi testi ad esempio sarà possibile di un autore produrre una concordanza lemmatizzata, ovvero richiedere di trovare tutte le occorrenze di una data parola, senza dovere indicare tutte le forme che assume nella flessione, il che risulta particolarmente utile anche nelle ricerche combinate. Ma le possibilità non finiscono qui: sarà possibile effettuare ricerche di tipo grammaticale, morfologico e sintattico, nonché di tipo metrico-prosodico (solo in parte effettuabili anche con i testi non codificati). Da tempo erano disponibili dizionari delle lingue moderne, anche di livello scientifico, come il classico *Oxford English Dictionary*, la cui trasformazione in formato elettronico codificato consente complesse ricerche: ora il neo costituito *Consortium for Latin Lexicography* (*CLL*)[63] sta pianificando una edizione elettronica del più importante strumento lessicografico latino, il *Thesaurus linguae Latinae*, che consenta non solo la lettura, ma anche ricerche complesse (all'interno delle varie rubriche dei singoli lemmi, ma anche per autore ed opera): siamo ancora ad uno stadio inziale, mentre è già in distribuzione un indice di tutte le forme latine, attestate negli autori a partire dagli autori arcaici, fino alla patristica, all'età medievale, fino ai nostri giorni, al Vaticano II, il *Cetedoc Index of Latin Forms* (*Thesaurus Formarum totius Latinitatis*)[64].

III.2.1. Ricerche lessicali, morfologiche, sintattiche

Per dare un esempio di queste possibilità ricorderò ad esempio che S.Mellet ha pubblicato una ricerca su *L'imparfait de l'indicatif en latin classique. Temps, aspects, modalité. Étude synchronique dans une perspective énonciative*[65], basando la sua ricerca su un *corpus* di autori codificati dal centro del LASLA: per gli autori considerati ha potuto richiedere la lista dei passi contenenti l'imperfetto, e la stampa con i rispettivi contesti. Il programma di gestione della banca dati ha inoltre elaborato dei dati relativi alla presenza dell'imperfetto indicativo in rapporto agli altri tempi del passato: l'autrice ha potuto così verificare, su una base statistica, differenze

nonché S.Mellet, *Logiciel d'exploitation de la banque de données de textes latins du LASLA*, RIS 30, 1994, 91-108. G.Purnelle, *Opera Latina: la banque de données de textes latins du L.A.S.L.A. sur CD-ROM*, RIS 31, 1995, 173-82 annuncia la imminente distribuzione su disco da parte di Hachette dei testi (77 opere di 18 autori di prosa e poesia) del Laboratorio di Liegi.

63 Una presentazione del *CLL* e del progetto (*The Electronic 'Thesaurus Linguae Latinae'*) è disponibile in rete: (http://www.cs.usask.ca/faculty/devito/e-TLL/index.html), cf., sempre on line, il più articolato progetto di A.De Vito, *Developing an Electronic Thesaurus Linguae Latinae* (http://www.cs.usask.ca/faculty/devito/e-TLL/ach95poster.html).

64 È anch'esso prodotto da Brepols.

65 Louvain 1988, v. anche la rassegna che ne fa Purnelle, cit., 65. G.Purnelle-C.Purnelle Sinnart, *Actes de la table ronde "Linguistique latine et informatique", V^e colloque international de linguistique latine, Louvain-la-Neuve 1989*, RIS 25, 1989, 193-213, presentano una serie di possibili ricerche come ad esempio di sostantivi in *-tudo* in Accio e Pacuvio, ovvero dell'uso del dativo in dipendenza delle forme verbali in *-ndus*, sottolineando la maggiore precisione delle interrogazioni a partire da testi lemmatizzati: per una ricerca sull'avverbio *profecto*, si potranno escludere le ricorrenze dell'omografo participio, e così via.

significative nella frequenza, e quindi nell'uso dell'imperfetto nei vari tipi di enunciato (narrazione o discorso), generi letterari, e nelle opere dei singoli autori.

Non è tanto la fase della elaborazione quindi che viene affidata al computer, quanto quella della raccolta dei dati: con un tono eccessivamente enfatico T.F.Brunner proclama che "in fact, the computer's efficiency is rapidly effecting fundamental changes in the scholarly process itself: whereas, in the past, scholars would spend as much as 75 per cent of their research time compiling raw data and information, leaving only 25 per cent for interpretative and analytical pursuits, there is clear evidence that computer-aided research in producing a reversal in this ratio. This, in turn, can be expected to result in a vast increase in both scholarly quantity and quality[66]". Resta comunque inconfutabile il fatto che la ricerca su banche dati produce in breve tempo una notevole massa di materiale: essa sarà tanto più mirata e precisa quanto più codificata (e quindi ricca di dati) sarà la base di partenza[67]. "In ambiente umanistico la maggior parte del tempo impiegato in un progetto informatizzato è speso nella preparazione dei dati" osservava Orlandi[68], il che è vero se si parte da zero nel preparare un testo su cui operare: nel caso in cui si operi su una banca dati già pronta, si deve tenere conto del lavoro di codifica svolto da chi vende il prodotto: un lavoro di esame forma per forma per inserire tutti i dati, che richiede notevole tempo, per quanto esistano dizionari elettronici in grado di svolgere per molte forme una analisi automatica, impossibile però nel caso degli omografi.

III.2.2. Altre analisi stilistiche (cooccorrenze, ripetizioni, citazioni) e metriche

L'analisi lessicale e stilistica è quella che maggiormente può usufruire dell'aiuto informatico: accanto alle analisi stilometriche, sulla lunghezza di parole, frasi, o sulla frequenza di parole o di tipi di parole, chiaramente è possibile estrarre il materiale per operare ricerche di sinonimia, preparando delle liste di parole da ricercare (per le lingue moderne esistono anche dei dizionari sinonimici in forma elettronica), e quindi ottenere, come si è accennato, le occorrenze semplici, ovvero dati relativi alla frequenza nei diversi autori, generi, o metri. Una ricerca come quella delle co-occorrenze di gruppi di parole in uno stesso verso, come ad esempio di incipit o di clausole dell'esametro, può essere svolta con minore o maggiore precisione a seconda del tipo di banca testuale che si usa: G.Purnelle ha ricercato in questo modo le co-occorrenze metriche nelle opere di Ovidio[69], ovvero ha cercato di individuare i sintagmi di due o tre parole che appaiono assieme in un verso e

66 Brunner, cit., 53.

67 Cf. supra, §§ II.3-4.

68 Orlandi, cit., 114.

69 Cf. Purnelle, cit., 66, e Id., *Répétitions de cooccurrences métriques chez Ovide*, RIS 23, 1987, 135-66.

che ritornano impiegati nelle medesime sedi in altri versi (come ad esempio *ars* 1,474 *Interit adsidua* ***uomer aduncus humo***; *rem.* 172 *Sauciet ut duram* ***uomer aduncus humum***; *ars* 1,280 *Femina cornipedi* ***semper adhinnit equo***; *rem.* 634 *Fortis equus uisae* ***semper adhinnit equae***; *ars* 1,312 *Fertur ut Aonio* ***concita Baccha*** *deo*; *ars* 3,719 *Euolat, ut thyrso* ***concita Baccha***, *uias*), per verificare se, e in quali casi, si tratta di ripetizioni casuali, oppure di effetti determinati dal contesto e dal condizionamento metrico, o se si tratta di vere e proprie formule. Una analoga ricerca condotta su Catone[70] ha portato ad individuare il ritornare di gruppi sintagmatici di sostantivo- aggettivo, o di sostantivo - verbo, o di verbo e complementi, in sostanza una certa recursività nello stile di Catone.

Non solo singole citazioni di un autore (che non sono altro che combinazioni di parole) si possono cercare in un altro (o in un *corpus*), come si è visto, ma alcuni programmi consentono di ricercare automaticamente le citazioni di un autore in un altro, per fare un esempio di ricercare i virgilianismi in Ovidio, confrontando fra di loro due autori, e verificando se nel secondo autore ritornano in sequenza all'interno di una frase, o di un verso, o di un rigo di testo, due o più parole che si trovano in sequenza nel primo: una ricerca di questo tipo darà luogo a un materiale estremamente "impuro", tanto più se si lavora con testi non lemmatizzati, in cui è estremamente difficile operare una corretta taratura della ricerca (se confrontare 2 o 3 parole, e di quale lunghezza, all'interno di 1 o 2 vv.), perché un contesto troppo ristretto darà risultati scarsi, mentre uno troppo ampio una quantità di materiale decisamente inutile. Non si può non condividere in gran parte, quindi, lo scetticismo di Cupaiuolo[71] nei confronti di siffatti metodi: è indubbio che alla macchina sfuggirà un gran numero di echi, per quanto ovvi per il filologo, e che molto materiale trovato sia casuale, se non fasullo. Non è detto, comunque, che tra tanto materiale inutile non venga evidenziata anche qualche nuova corrispondenza: allo studioso poi il compito di analizzarla e di giustificarla, per non ridurre tutta la ricerca ad una insulsa elencazione di *loci similes*, trappola in cui è possibile cadere anche senza il ricorso ai mezzi informatici. Qualche risultato di rilievo si ottiene confrontando un autore con se stesso, il che porta ad evidenziare le ripetizioni lessicali, da considerarsi sempre con la cautela che si è vista. Ancora una volta è chiaro che simili ricerche risultano più precise se condotte su testi ben codificati, ma anche con i testi più semplici, come ad esempio quelli del *PHI5*, che pur sempre contengono la codifica della fine dei versi, e utilizzando semplici programmi di ricerca (come *Lbase* o *Lector*), come ha dimostrato T.Del Vecchio[72], è possibile condurre ricerche di una certa complessità, e raccogliere ad esempio

[70] Cf., anche per altre possibilità di indagine, G.Purnelle, *Recherches automatiques de groupes verbaux récurrents et de formules dans les fichiers latins lemmatisés*, RIS 25, 1989, 157-91.

[71] F.Cupaiuolo, *Bibliografia della lingua latina (1949-1991)*, Napoli 1993, 153.

[72] Del Vecchio, cit., 85-98 e Id., *Evidence for zero metrical value of interjections in Plautus*, in corso di stampa su Tema.

il materiale necessario per analizzare la presenza del monosillabo finale nei versi di Plauto e Terenzio. La metrica, come si è ripetuto più volte, è, per la sua natura uno dei campi in cui più facilmente trova applicazione lo strumento informatico, come dimostrano le ricerche in qualche modo pioneristiche di W.Ott sull'esametro latino, che hanno portato a indagini a campione su numerosi autori, da Lucrezio, a Catullo, Orazio, a Virgilio e Stazio[73],

Sono possibili anche analisi metrico stilistiche, come quella condotta da R.Oniga sulla allitterazione in Plauto e Terenzio[74], che, prende le mosse dalla definizione di "allitterazione" come "ripetizione di uno o più fonemi iniziali in due o più parole contigue"[75], ed esclude le cosiddette "pseudo-allitterazioni" costituite, come rilevava già a suo tempo il Ronconi, da particelle grammaticali, congiunzioni, pronomi, e così via[76], ovvero le anafore. L'esame contrastivo tra due autori consente di verificare la bontà del metodo impiegato: il risultato è che la frequenza media tra tutte le commedie per Plauto è di 0,51 allitterazioni per verso, per Terenzio di 0,44. Plauto ha cioè in media "un 30 % di allitterazioni in più rispetto a Terenzio"[77], ma la differenza aumenta se si adotta il criterio della contiguità assoluta e si escludono del tutto congiunzioni e preposizioni. È stato poi possibile verificare che le allitterazioni sono più frequenti nei *cantica* polimetrici e nei versi lunghi, rispetto ai senari, in entrambi gli autori, come del resto aveva visto lo Haffter, nonché individuare i nessi e i fonemi allitteranti più frequenti nei due autori. È questo un esempio in cui i metodi di indagine tradizionale si fondono perfettamente con le potenzialità di elaborazione dei dati consentita dal computer, mentre il linguaggio estremamente chiaro consente anche ai non esperti di cogliere gli esiti della ricerca, a differenza di tanti altri studi, in cui tutto sembra portare ad aride tabelle di dati, e a calcoli numerici, senza una precisa utilità, e senza che questi vengano poi sfruttati per un vero discorso sullo stile dell'autore esaminato.

III.2.3. La critica del testo (collazione automatica e stemmatica)

"Non è ancora venuto il tempo in cui i manoscritti potranno essere collazionati automaticamente; non sono ancora state realizzate macchine che

[73] Cf. ad es. W.Ott, *Metrische Analysen zur Ars Poetica des Horaz*, Göppingen, 1970, *Metrische Analysen zu Vergil Aeneis Buch I, VI, XII,* Tübingen 1973 (che costituiscono i primi tre fascicoli dei *Materialen zu Metrik un Stilistik*), *Metrische Analysen zu Statius Thebais Buch I*, Ibid. 1973: una breve rassegna bibliografica in A.Traina - G.Bernardi Perini, *Propedeutica al latino universitario.* Quarta edizione a c. di C.Marangoni, Bologna 1992, 299 s.

[74] R.Oniga, *L'allitterazione in Plauto e Terenzio: un esperimento di analisi quantitativa*, Lexis 12, 1994, 117-34.

[75] Cf. Id., 119: viene tuttavia ammesso che tra una parola e l'altra si possano trovare parole di ridottissimo corpo fonetico, come ad es. in *PAtriam et PArentes.*

[76] Cf. le osservazioni di metodo in A.Ronconi, *Allitterazione e ritmo,* in *Interpretazioni grammaticali*, Roma 1971[2], 336.

[77] Oniga, cit., 124.

possano far fronte alle varietà che compaiono nella scrittura manuale", scriveva scettico M.L.West nel 1973[78], evidenziando i limiti della generazione automatica di uno stemma codicum, limiti che non paiono essere stati superati, ma che pongono degli interessanti interrogativi su alcuni dei metodi tradizionali di classificazione. Sono da tempo noti i problemi legati alla ricostruzione meccanica fondata sulla applicazione rigorosa del dettato dello stemma, a partire da quelli di datazione dei manoscritti, per arrivare alla soggettività contenuta nella identificazione di varianti ed errori, tenendo presente inoltre che non mancano varianti d'autore per i testi antichi; lo stemma, poi, è una rappresentazione di comodo della genealogia dei manoscritti, ma non sempre conduce alla determinazione delle lezioni dell'archetipo[79]. Il problema si fa ancora più complicato nel caso di contaminazione, tanto che, per quanti sforzi si siano fatti non sembra che ci sia tanto allontanati dalla affermazione di Maas: "gegen Kontamination ist kein Kraut gewachsen"[80], che Pasquali tendeva a rendere meno drastica, notando che "in particolari casi, rimedi si possono escogitare con buon frutto, così come equazioni algebriche in genere insolubili si possono risolvere in casi particolari"[81].

Una via di scampo più generale, a fronte di una tradizione complessa e contaminata come quella della *Vulgata* latina della Bibbia, ha cercato H.Quentin che già nel 1926[82] proponeva di rinunciare alla valutazione degli errori, e di procedere ad un confronto statistico delle lezioni coincidenti nei codici, ponendo praticamente su un asse cartesiano i codici e sull'altro le varianti, in modo che risultassero evidenziate, in base alla "vicinanza" o "lontananza", le famiglie di codici. Questo metodo, basato sulla elaborazione statistica dei dati, sarebbe stato ripreso fra gli altri da J.Froger, per tentare una automatizzazione della recensione, fondata, questa volta, sul disaccordo delle

[78] M.L.West, *Critica del testo e tecnica dell'edizione*, Palermo 1991, [ed. or. *Textual Criticism and Editorial Technique*, Stuttgart 1973, 71] 72.

[79] Si veda ad esempio il differente modo di considerarlo da parte di Ritschl e Lachmann, tratteggiato da S.Timpanaro, *La genesi del metodo del Lachmann*, Padova 1981², 56, "l'uno appassionato ... alla storia della tradizione in quanto tale, e tendente a risolverla, come già il vecchio Wolf, in storia della cultura, l'altro tutto teso allo scopo di liberare il testo dalle interpolazioni tarde e di restituire la forma più antica per noi raggiungibile, ma incurante di quella indagine genealogica che, pure, costituiva un concetto essenziale della ricostruzione dell'archetipo".

[80] "Contro la contaminazione non è stato scoperto un rimedio", cf. P.Maas, *Critica del testo*, Firenze 1980³ [ed. or. *Textkritik*, in A.Gercke-E.Norden, *Einleitung in die Altertumswissenschaft*, I.VII, Leipzig 1927], 62, ma, come osservava G.Pasquali nella *Presentazione* alla edizione italiana, VIII s., "il tedesco scrive, più possente «contro la contaminazione non è ancora cresciuta alcun'erba» e pensa alla farmacologia antica e medievale, che non conosce ancora medicine prodotte chimicamente, e insieme allude parodicamente al proverbio «Gegen den Tod ist kein Kraut gewachsen»."

[81] Id., IX.

[82] H.Quentin, *Essai de critique textuelle (Ecdotique)*, Paris 1926.

varianti[83]. Con un sistema analogo a questo (ma non senza un compromissorio affiancarsi del metodo tradizionale, e della ricostruzione stemmatica operata da Courtney), J.R.C.Martin ha recentemente pubblicato una edizione di Giovenale[84], in cui, dopo aver prodotto uno schema generale di affinità dei codici, il cui andamento evidenzia che non tutte le satire hanno seguito la medesima tradizione, riproduce in fondo ad ogni pagina delle linee che evidenziano per brevi sezioni di testo i gruppi di codici (*manipuli codicum*): "quo minor inter codices est numerus, eo magis vicini inter se cohaerent". Questa linea "non solum particulatim lectiones nonnullas vel novas vel olim dubias confirmare potest, sed etiam passim naturas codicum aut mutabiles aut constantes multo clarius ad locum depingit"[85]. Osservando che in alcuni casi questo sistema risulta approssimativo, fornendo delle false prossimità, J.G.Griffith ha a sua volta proposto una rappresentazione su assi fattoriali, che resta comunque fondata sugli stessi principi statistici[86].

I problemi con cui si scontra l'automazione della stemmatica sono, in parte, quelli di fondo della ecdotica tradizionale, come ad esempio quello della scelta del testo (a stampa o manoscritto) da adoperare come base su cui effettuare la collazione, perché non influisca nella definizione dello stemma: si è provato ad affidare totalmente al computer anche questa fase della recensio, digitando completamente tutti i testimoni e affidando alla macchina il compito del confronto[87]. Quanto ciò sia dispendioso in tempo, e pericoloso, ognun vede, perché banali errori di digitazione possono portare perturbazioni serie nello studio genetico. Ma è il metodo statistico in generale che ha degli indubbi limiti, che ne hanno determinato la messa in discussione, soprattutto quando si ritiene che l'algoritmo sia sufficiente a ricostruire la tradizione sulla base del confronto delle varianti, a prescindere quindi dall'analisi "storica" che si considera quasi immanente allo stemma. Ma lascio qui parlare uno specialista della materia, Gian Pietro Zarri, che, nel convegno già ricordato su *La pratique des ordinateurs dans la critique des textes*, del 1978, sottolineava come, per elaborare una teoria scientifica che permetta la spiegazione di un certo tipo di fenomeni, occorre che possa essere stabilita una precisa definizione di questi fenomeni, completa e non ambigua. Se ciò

[83] Cf. J.Froger, *La critique des textes et son automatisation*, Paris 1968 e *La méthode de Dom Quentin, la méthode des distances et le problème de la contamination*, in *La pratique des ordinateurs*, cit., 13-22; è d'altra parte questa la obiezione che gli muove Timpanaro, cit., 48 n.18: "il metodo del Quentin è impotente dinanzi alla obiezione che solo la coincidenza in errore può indicare parentela tra due mss.: la coincidenza in lezione giusta non prova n u l l a, perché è un fatto di conservazione, che può essere accaduto anche in mss. non imparentati".

[84] D.Iuni Iuvenalis *Satirae*, edidit J.R.C.Martin, Amsterdam 1987.

[85] Id., XXIV.

[86] J.G.Griffith, *Non-stemmatic Classification of Manuscripts by Computer Methods*, in *La pratique des ordinateurs*, cit., 73-86, con osservazioni sulle tradizioni dell'*ars* oraziana, di Giovenale, dei *Vangeli*, e di alcune tragedie eschilee.

[87] Cf. per es. G.P.Zarri, *Linguistica algoritmica e meccanizzazione della «collatio codicum»*, in L&S 3,1968, 32-40 [rist. in *Letteratura e filologia*, cit., 158-67], S.V.F.Waite, *Two programs for comparing texts*, in *La pratique des ordinateurs*, cit., 241-44.

non è possibile, "il faut alors abandoner tout rêve d'interprétation globale des phénomènes, c'est-à-dire de construction d'un 'modèle' explicatif général qui permette d'obtenir les propriétés de ces phénomènes par voie uniquement déductive"[88]. Ci si deve quindi accontentare di un modello "parziale", da verificarsi volta per volta: è il caso della critica testuale in cui "les données de base dont nous disposons, fautes ou variantes, sont, en effet, comme toutes données historiques, incomplètes, et par surcroît, ambigües". È soprattutto l'ambiguità che rende impossibile raggiungere la precisione di una scienza esatta (e di applicare quindi anche una recensione meccanica in senso stretto): ambiguità che incide a livello di definizione del luogo variante, di scelta delle varianti da inserire e di pertinenza delle varianti stesse. Il metodo stesso proposto da Griffith è ambiguo, giacché la "somiglianza" è un concetto estremamente generico[89]; Zarri vi vedeva semplicemente un primo strumento puramente descrittivo, utile preliminarmente, cui non compete la ricostruzione della tradizione: "je suis convaincu qu'un jumelage savant de procédures de taxonomie textuelle et de méthodes 'généalogiques' pourrait être d'une grande utilité surtout dans le cas de traditions très riches, où un *stemma* de la totalité des manuscrits n'a aucun sens et où la statistique pourrait faciliter une répartition en familles; à des méthodes plus précises, plus "historiques", ensuite, le soin de réaliser le *stemma* des représentants de ces familles"[90]. Resta una ulteriore osservazione: considerare i codici alla stregua di contenitori di varianti è estremamente riduttivo, si considera solo il testo, mentre si "ignora lo strumento materiale che di quest'ultimo scandisce la comparsa e la veicolazione, vale a dire il libro-manoscritto"[91], prescindendo dagli aspetti storici (modi, tempi e luoghi) della tradizione, dalle divergenze che insorgono in base alla recezione e alla utilizzazione del testo stesso. "Occorrerebbe un «filologo artificiale» che per ora non abbiamo"[92], il che non significa rinuncia assoluta all'impiego del computer nella critica testuale: esistono infatti programmi più o meno complessi che aiutano il filologo a compiere l'operazione della collazione, consentendo di registrare via via per ogni passo le varianti dei vari codici[93]. Il passo

[88] Id., *Une méthode de dérivation quentinienne pour la constitution semi-automatique de généalogies de manuscrits; premier bilan*, in *La pratique des ordinateurs*, cit., 132.

[89] Cf. inoltre la fondamentale critica metodologica di Timpanaro, cit. supra in n. 71.

[90] Zarri, cit., 136.

[91] G.Cavallo, *Premessa* a *Le strade del testo*, Bari 1987, V, che sottolinea come grande merito di Pasquali fu proprio quello di porre una stretta connessione tra "vicende editoriali e storico/culturali di autori/opere, quindi storia della tradizione manoscritta, e critica testuale" (IV).

[92] Così ancora Timpanaro, cit., 48 n.18, ove, peraltro, non disconosce in casi di tradizione molto complessa e più ricche di varianti che di vere corruttele, la utilità di un ricorso al metodo quantitativo.

[93] Penso ad esempio a *TUSTEP* (sistema Dos e Unix), per cui v. il già cit. W.Ott, *Automatic composition*, e, dello stesso, *A Text Processing System for the Preparation of Critical Editions*, in CHum 13, 1979, 29-35, *TUSTEP. System von Textverarbeitungs-Programm*, Tübingen, 1985, ed ora *Electronic Reading, electronic Publishing and Tools in between. Experiences from supporting humanities computing*, in Leonardi - Morelli - Santi (edd.), cit., 91-99, nonché il

successivo, della scelta del testo, delle varianti da ritenere, viene condotto dal filologo, mentre l'elaboratore aiuta a trovare raggruppati gli elementi dell'apparato, e, con metodi più complessi, a selezionare un apparato positivo o negativo, nonché, a seconda delle codifiche che si sono introdotte, a fare ricerche su codici, tipologie di errore, e così via, come in qualsiasi banca dati: non si affida più all'elaboratore il compito di ricostruire lo stemma, ma ci si serve di esso esattamente nello stesso modo del taccuino (*notebook*), o degli inchiostri di vari colori di West[94], solo che alla fine il lavoro è già pronto per la stampa (con minori probabilità di errore).

III.2.4. I manoscritti, il riconoscimento caratteri e le memorie ottiche

Un'ultima possibilità, che affianca, ma non sostituisce la lettura diretta dei manoscritti, è infine offerta dalle nuove tecnologie delle memorie ottiche[95], che consentono di immagazzinare su disco migliaia di immagini che si possono rielaborare variamente tramite un apposito programma, schiarendo ad esempio i toni di colore o di grigio degli sfondi, aumentando o riducendo i contrasti, per ridurre il disturbo di macchie di inchiostro, o di sfondi ingialliti. Un ulteriore passo, più complesso, consentirà di sezionare il codice nelle sue righe, parole e lettere, e di registrare la trascrizione eseguita dal paleografo, facendo corrispondere ad ogni lettera la sua trascrizione, in modo da potere confrontare fra di loro lettere, parole, e fra di loro manoscritti differenti, per identificarne ad esempio lo scriba, sino ad arrivare, in alcuni casi ad un riconoscimento automatico dei caratteri[96].

III.2.5. La bibliografia

Il campo bibliografico è, infine, quello da cui il filologo, e lo studioso in generale, può fin da subito, in una biblioteca ben attrezzata, trarre un significativo aiuto: cataloghi finora a stampa sono ora disponibili, come si accennava, su CD-ROM, come ad esempio quello della Biblioteca del Congresso, della British Library o della Bibliothèque Nationale de Paris,

capitolo di Perilli, cit., 85-121; disponibile solo per Mac è ora *Collate*, per cui v. ad es. la recensione di E.Johnson, in CHum 27, 5-6, 1993-94, 400-3.

94 West, cit., 68 [ed. or. 66].

95 Oltre al concetto di testo varia così anche quello di lettore, come sottolinea D'A.S.Avalle, *Le procedure informatiche nelle scienze umane*, in Leonardi- Morelli- Santi (edd.), cit., 3-6.

96 Cf. A.Bozzi, *Text editing e text processing: aspetti e problemi di computerizzazione di dati testuali editi ed inediti*, in V.Placella - S.Martelli, *I moderni ausili all'Ecdotica. Atti del Convegno Internazionale di Studi (Fisciano-Vietri sul Mare-Napoli, 27-31 ottobre 1990)*, Napoli 492-94, e, nello stesso volume, P.M.De Biasi, *Manuscrits modernes: les techniques d'analyse scientifique de l'écriture*, 507-17. Il sistema è ovviamente utilizzato anche per le stampe antiche, cf. Ben J.P.Salemans, *Comparing Text Editions with the Aid of the Computer*, CHum 28, 1994-95, 133-39. È forse opportuno notare come sia ora disponibile un repertorio elettronico delle abbreviature medievali latine, *Abbreviationes*, per sistema Mac, distribuito da Brepols.

ovvero dell'Istituto Archeologico Germanico di Roma[97]. Anche repertori bibliografici più tradizionali, come ad esempio la rivista «Gnomon», sono distribuiti sotto forma elettronica, e quindi consultabili per le più varie ricerche[98], mentre sono disponibili su CD-ROM i volumi 47-58 (1976-1987) dell'«Année Philologique», con un efficiente programma di interrogazione (per autore, titolo, anno, rivista, ecc., ma anche in grado di operare sull'abstract, il "full text") e di esportazione dei dati, per Windows e Macintosch, e che verrà completato mediante aggiornamenti di 3/5 volumi per anno [99]. Ma lo strumento più potente sia dal punto bibliotecario che bibliografico, è quello offerto dalla consultazione di banche dati "on line", come appunto cataloghi librari di biblioteche, italiane e straniere, cataloghi di periodici, in modo da reperire facilmente un libro, senza dovere ricorrere agli uffici di amici nelle varie città o nei vari paesi. Unitamente ad essi strumenti continuamente aggiornati, come i TOCS, "Tables of Contents of Journals of Interest to Classicist", che offrono lo spoglio di oltre 140 riviste dal 1992, fino agli ultimi sei mesi[100], ovvero bibliografie specializzate, come quelle sulla donna nell'antichità (il progetto *Diotima*)[101] possono facilitare non di poco la costituzione e il reperimento del materiale su un determinato autore o tema specifico, ma questo è oggetto della prossima sezione.

[97] Cf. *Dyabola. Elektronische Sachkataloge zu den Altertums und Kunstwissenschaften*, pubblicato dal 1990 su CD-ROM, recensito su Gnomon 67, 1995, 551-60.

[98] La *Gnomon Bibliographischen Datenbank* contiene le annate dal 1925 (I) al 1993 (LXVI), ed è arricchita di dati provenienti da altre riviste, cf. J.Malitz, *Gnomon Bibliographische Datenbank. Internationales Informationssystem für die Klassischem Altertumswissenschaften*, in Fell - Schäfer - Wierschowski (edd.), cit., 165 s. Analogamente, ma su un piano più limitato, dal 1992 la casa editrice Erma produce, di pari passo al "Bollettino Novità", *Heracles. Banca Dati Bibliografici del Mondo antico e dell'Arte*, che per il momento contiene 25.000 titoli di volumi in commercio.

[99] Si tratta del *Database of Classical Bibliography*, prodotto a cura della "American Philological Association" (e pubblicato dalla Scholar Press di Atlanta): ne fornisce una breve descrizione il suo curatore, D.Clayman, *The Database of Classical Bibliography*, in Solomon, (ed.), cit., 108-32. Anche l'Università del Québec, che ha nel 1985 prodotto la *Clavis Scriptorum Graecorum et Latinorum* si accinge a trasferire su CD la sua banca dati, che, oltre ad una versione aggiornata della *Clavis* dovrà contenere dossier sugli autori della *Clavis* attingendo a fonti come l'APh, il Fabricius, gli Jahresberichte di Bursian (per il momento è uscita la lettera A, cf. R.La Rue-P.Senay - F.Tremblay, *Un Thesaurus informatisé de bibliographie gréco-latine*, RIS 31, 1995, 85-91).

[100] Sono raggiungibili via ftp://ftp.epas.utoronto.ca/pub/tocs-in, oppure via gopher://gopher.epas.utoronto.ca:70/11/cch/disciplines/classics/cr, oppure ancora tramite il gopher://gopher.lib.Virginia.Edu:70/11/alpha/tocs. La via più semplice per raggiungerli, tuttavia è di usare il *link* (il rinvio) contenuto nel *Repositories of Classical Texts or Publications*, per cui v. supra, n. 21.

[101] *Diotima. Materials for the Study of Women & Gender in the Ancient World* è reperibile via http://www.uky.edu/ArtSciences/Classics/gender.html

IV. La Telematica negli studi classici

IV.1. Premessa

Già le banche dati 'tradizionali', cioè quelle degli anni Ottanta, offrivano delle possibilità di ampio respiro. Allora ci si poteva collegare con enti ed istituti che offrivano dei servizi di *information retrieval* (recupero informazioni) su banche dati bibliografiche e statistiche. Nell'ambito degli studi umanistici la Dialog, ad esempio, disponeva di un archivio bibliografico abbastanza ampio per gli studi di linguistica, molto meno per quelli classici. Ma la scarsa consistenza dell'archivio era solo il male minore, dovendosi scontrare gli utenti con difficoltà di collegamento, rigidità della sintassi interrogativa, e costi assai elevati. Inoltre erano pochissimi i ricercatori in grado di connettersi direttamente, e bisognava allora recarsi in centri di calcolo, ove il personale specializzato effettuava la ricerca. Questo poneva grossi problemi di risorse personali e di accesso ai centri, che solo nelle sedi principali erano presenti in ambito cittadino. Un esempio concreto di quegli anni: una ricerca sull'argomento "metrica plautina", fatta tramite Dialog, dà 16 citazioni e l'abstract di 4 di esse, il tutto con la modica spesa di 90.000 lire! Si pensi inoltre che per far questo si è dovuto ricorrere a una telefonata intercontinentale, perché non si riusciva ad effettuare la chiamata al centro di calcolo di Frascati, che poteva fungere da ponte per gli USA.

Negli anni Novanta abbiamo assistito a una vera rivoluzione che si è manifestata su tre fronti: la diffusione di PC molto efficienti, paragonabili alle grosse e costosissime macchine di dieci anni prima; la creazione, da parte di istituti universitari ed enti pubblici e privati, di archivi bibliografici e di altra natura molto consistenti; la distribuzione capillare delle reti telematiche. Rimaneva ancora il problema della difficoltà del reperimento dei dati e di una sintassi di interrogazione non standardizzata e molto rigida. Inoltre i collegamenti tra le reti erano molto complicati e affidati a grossi centri di calcolo.

Ma in questi anni l'avvento di Internet, che ha dato a tutti la possibilità di collegarsi con il mondo esterno con estrema semplicità, ha modificato sia l'uso che il senso della telematica. Internet è un grande regolatore di traffico, che è in grado di mettere in comunicazione strade di diverse caratteristiche: è come se fosse possibile per chi viaggia in treno usare le autostrade, per chi naviga percorrere le strade ferrate, per chi usa l'auto percorrere fiumi e torrenti. In parole povere questo significa poter esplorare le banche dati di tutti i sistemi collegati (e, ovviamente, aperti a tutti: la questione della democrazia, nella gestione di Internet, si va ponendo proprio in questi mesi), navigando appunto per questo spazio dove le reti sono solo una metafora.

Semplicità di collegamento, standardizzazione e semplificazione della sintassi interrogativa, gratuità dell'accesso agli archivi sono le caratteristiche principali che oggi offre Internet a chi è collegato a una qualunque rete, sia essa una piccola rete dipartimentale o una più ampia rete universitaria. La diffusione delle reti ha offerto il grosso vantaggio di potersi collegare

gratuitamente dai centri di ricerca e dalle Università, che sopportano per questo spese molto basse. I centri che non hanno il collegamento diretto devono al massimo effettuare una chiamata urbana al proprio centro di calcolo. Certo si pone il problema di dotarsi di centri di calcolo efficienti, soprattutto dal punto di vista della professionalità degli addetti.

IV.2. Applicazioni nel campo umanistico

Nell'ambito degli studi classici esistono due grosse possibilità[102]:

a) l'accesso ai cataloghi delle biblioteche collegate;

b) l'accesso a sistemi informativi complessi che sono stati organizzati in questo settore in vari Paesi.

Si mostrerà più avanti quali sono realmente queste possibilità, e quali i problemi pratici che possono insorgere. Alcune osservazioni di carattere generale sono però d'obbligo.

La domanda più consistente di collegamento alle reti viene dagli studenti e dai giovani ricercatori. C'è indubbiamente un problema di carattere generazionale, dovuto alla assoluta mancanza di cultura informatica di chi ha frequentato le nostre università fino agli anni 80. Non si dimentichi che appena dieci anni fa essere in rete e usare la posta elettronica sembrava una perdita di tempo inutile. Mentre da un lato sarebbe utile avvicinare anche i ricercatori "anziani" all'uso degli strumenti informatici, è invece indispensabile che i giovani siano messi in condizione di usare questi strumenti con la stessa disinvoltura con cui siamo stati abituati a usare la penna. Certo alcune cautele si impongono. Immaginiamo ad esempio le reazioni di un giovane studioso davanti a un catalogo cartaceo di una biblioteca di filologia. Alcune informazioni gli sono immediatamente percepibili. Già la struttura dell'edificio, l'arredamento della sala, la conformazione dei mobiletti che contengono i cataloghi, gli fanno capire se si tratta di una vecchia o nuova biblioteca. Gli basta poi aprire un cassetto e consultare le schede, per rendersi conto se si tratta di vecchi fondi o di consistenze recenti. Infatti le schede possono essere o vecchie schede scritte a mano, o fotocopie di vecchie schede, o schede nuove. In questo caso, se le schede nuove sono tutte diverse, si capisce che si tratta di una biblioteca recente, ma cresciuta nel tempo. Se le schede nuove sono tutte uguali, si pensa subito a una biblioteca di recente costituzione. Questi sono solo alcuni esempi che devono far riflettere sugli elementi di differenza qualitativa tra cataloghi cartacei e cataloghi elettronici. Questi ultimi infatti hanno la caratteristica di essere tutti uguali, e non hanno quindi in sé, dunque nel loro essere schede di un catalogo, nessuna caratteristica particolare. Si pensi poi a cosa vuol dire spostarsi fisicamente per andare a consultare una biblioteca. Lasciare Firenze, uscendo dai muri della sua vecchia Biblioteca Nazionale,

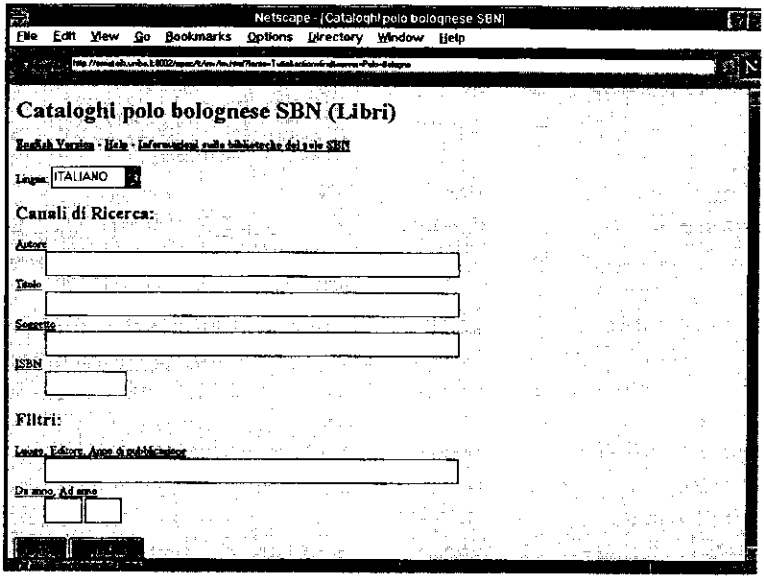
[102] Una panoramica generale su Internet per gli umanisti, in Lana, cit., 217-49; cf. anche i molti dati raccolti dalla già cit. *Rassegna* di A.Cristofori (con i suoi utili *link*).

dopo aver consultato un vecchio codice varroniano, e prendere il treno, per recarsi in Germania, ad Heidelberg, e trovarsi anche lì in una vecchia biblioteca, ma non così vecchia, per vedere un celebre codice plautino, o andare a Bologna, al vecchio Archiginnasio, o a Washington, per perdersi nel mare modernizzato dei tempi della cultura, come a Parigi. E questo è solo l'inizio, riguarda solo la consultazione dei cataloghi. Se si pensa poi all'approccio che si ha con un libro elettronico, dove l'aspetto che esso mostra sul video è determinato esclusivamente dalla quantità di codici che sono stati inseriti nel testo e dalle capacità del 'browser', si capisce subito come si perdano in questo modo una serie di preziose informazioni che si hanno quando si ha un 'vero' libro per le mani. Si va dalla presumibile età del libro, alla sua consistenza, alla possibilità di sfogliarlo rapidamente buttando l'occhio qua e là, alla consultazione con le mani: un dito nell'indice, uno nel terzo capitolo, uno nelle note. È vero che queste operazioni si fanno bene, anzi meglio, con un computer, ma è altrettanto vero che la massa di informazioni che viene assunta nella consultazione manuale contestualmente alla lettura, viene irrimediabilmente persa.

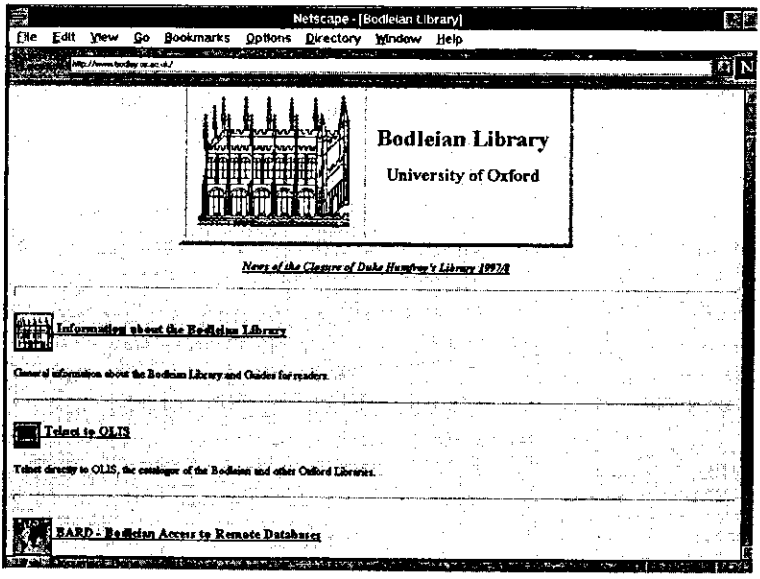
La consultazione a distanza offre però dei vantaggi enormi. Prima di tutto la possibilità di consultare cataloghi in biblioteche sparse in tutto il mondo, senza affrontare lunghi e costosi viaggi. Andare a Washington potrebbe essere, ancora oggi, un problema per molti. Inoltre la rapidità della consultazione non è solo un guadagnar tempo, ma anche da un punto di vista qualitativo offre tutti i vantaggi che derivano dalla possibilità di maneggiare tutta la materia senza dispersioni. Infine, e questa è forse la differenza qualitativamente più importante, l'accesso ai cataloghi elettronici non vuol dire solo consultazione delle schede, ma accesso diretto agli archivi che altrimenti sarebbero inaccessibili. Questo vale sia per gli archivi in cui sono depositate schede informative aggiuntive, sia per quelli appositamente creati come abstracts elettronici, sia per quelli in cui sono presenti indici di riviste e libri di più autori. Non dovendosi più scontrare con un problema di formato obbligato di scheda, gli archivi elettronici possono senza difficoltà contenere una quantità di dati che non disturbano la normale consultazione.

Le biblioteche che hanno messo in linea i propri cataloghi crescono continuamente, anche se ancora molta strada è da fare. La maggior parte dei collegamenti avviene infatti ancora via telnet, costringendo quindi ad usare il proprio PC come un terminale collegato direttamente alla macchina: lo svantaggio è che ogni sistema impone delle modalità di ricerca differente, e bisogna quindi seguire attentamente le istruzioni scritte. Internet ha dato la possibilità di usare delle schede di ricerca standard, le cosiddette *form*, che semplificano enormemente l'interrogazione, ma l'adeguamento comporta per gli amministratori di rete un lavoro non indifferente di transcodifica, che per ora è allo stato iniziale.

Un esempio di *form* è rappresentato dalla scheda predisposta dal CIB di Bologna, di cui si acclude un'immagine:

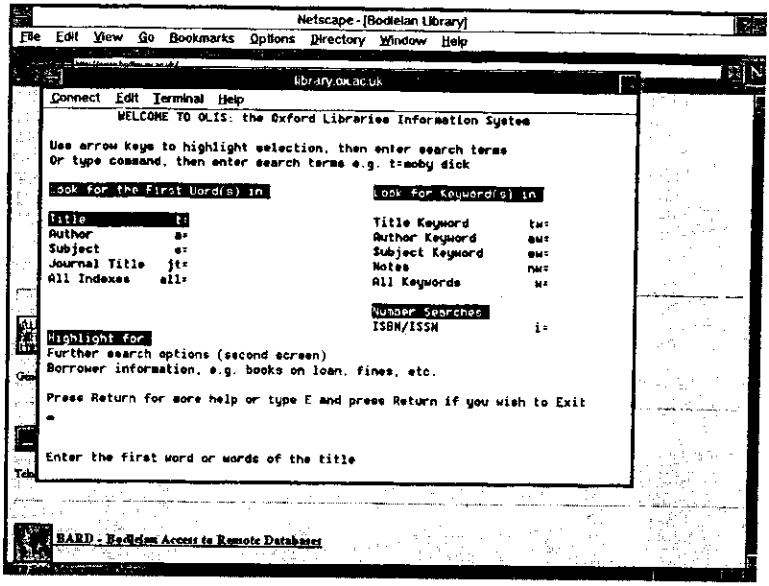


Un altro esempio è dato dalle seguenti immagini, tratta da un collegamento con la *Bodleian Library*[103] di Oxford, delle quali la prima si riferisce a un collegamento di tipo informativo, la seconda invece a un collegamento per la ricerca in Telnet[104].



[103] http://www.bodley.ox.ac.uk

[104] telnet: library.ox.ac.uk



Difficoltà a parte, è importante il fatto che un gran numero di biblioteche sia adesso accessibile dal proprio luogo di lavoro. Con un computer moderno è possibile anche continuare a scrivere il proprio lavoro mentre la ricerca bibliografica va avanti, e poi trasferire i dati direttamente sul proprio foglio. L'integrazione sempre crescente che va delineandosi tra sistemi informatici e telematica renderà ancora più semplice e produttivo l'uso del computer per questi scopi.

Ecco quindi una prima lista delle biblioteche collegate, tanto per dare un'idea delle opportunità offerte da Internet:

[Berkeley Digital Library SunSITE]
Libraries on the Web
Europe
Current update: 09/30/96

* Aachen University of Technology
* Aberdeen University
* Amsterdam Vrije Universiteit
* Arctic Centre, University of Lapland
* Åbo Akademi University
* Århus Public Library
* Aristotle University
* Aston University
* Augsburg University
* Bath University
* Bournemouth University
* Bradford University
* British Library
* British Library of Political and Economic Science
* BYGGDOK
* Cambridge University
* Cambridge University - Department of Genetics
* Catholic University of Eichstätt
* Catholic University of Leuven
* Catholic University of Louvain
* Central Library for Physics in Vienna
* Centre for Mathematics and Computer Science
* Central and State Library of Berlin

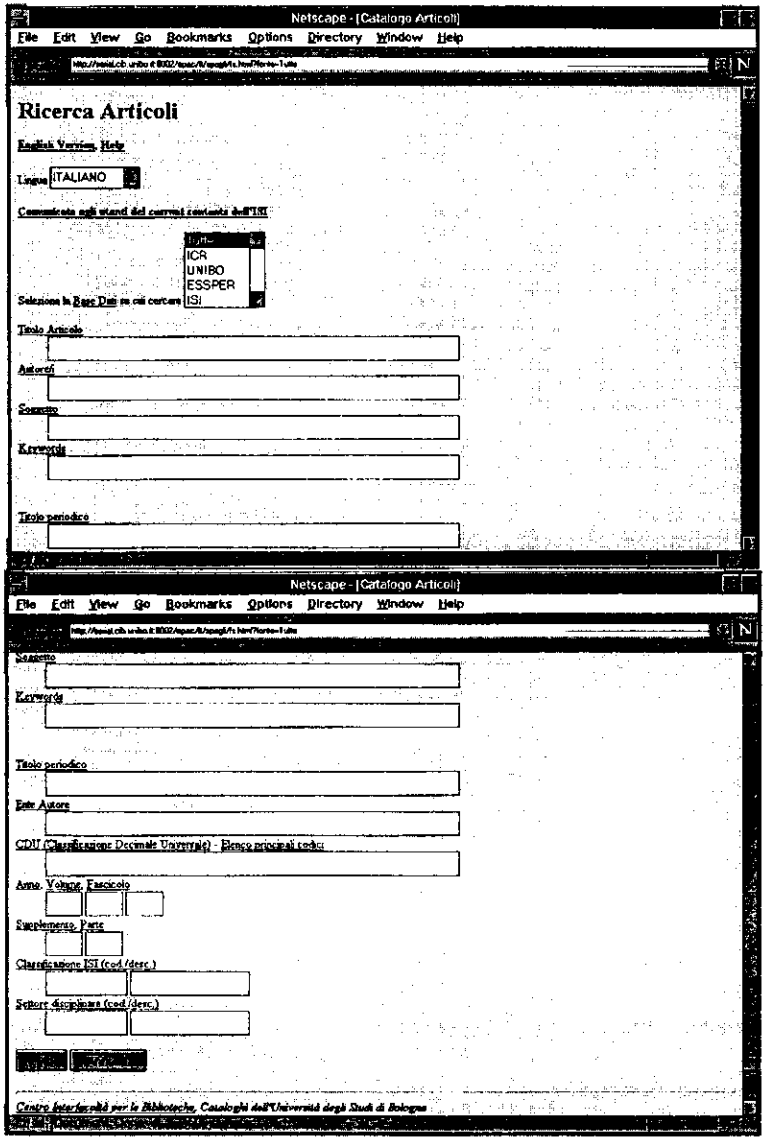
* Centre National de la Recherche Scientifique - Social Sciences - Research Catalog
* Chalmers University of Technology
* Christian Albrechts University of Kiel Technical Faculty
* Cologne Public Library
* Complutense University of Madrid
* Czech National Library
* Danish Centre for Information on Women and Gender
* De Montfort University
* Delft University of Technology
* Dublin City Public Library
* Durham University
* Erasmus University
* European University Institute
* Exeter University
* Finnish National Library of Health Sciences
* Free University of Berlin
* Free University of Brussels
* Fundesco
* German Library Institute
* German National Library
* Glasgow University
* Glasgow Women's Library
* Göteborg University
* Groningen University
* HBZ
* Helsinki City Library
* Helsinki University
* Helsinki University of Technology
* Heriot-Watt University
* Hertfordshire Libraries Arts and information
* INRIA - Research Center of Sophia Antipolis
* Institute and Museum of the History of Science
* Institution of Electrical Engineers
* International Centre for Theoretical Physics
* International Institute of Social History
* IRCAM
* Joensuu University
* Johann Wolfgang Goethe University
* Jönköping College
* Jyväskylä University
* Karolinska Institute
* Keele University
* King's College School of Medicine and Dentistry
* Lajos Kossuth University
* Lancaster University
* Latvia University
* Latvian Academic Library
* Leeds University
* Library Network of Southwest Germany
* Linköping University
* London Business School
* Lund University
* Malmö Public Library
* Milan Polytechnical Institute
* Mönadal Public Library
* National Agricultural Library and Documentation Center of Hungary
* National Agronomic Institute of France
* National and University Library of Slovenia
* National Library of Austria
* National Library of Estonia
* National Library of France
* National Library of Iceland
* National Library of Lithuania
* National Library of Portugal
* National Library of Spain
* National Library of Sweden
* National Library of Switzerland
* National Library of the Netherlands
* National Repository Library of Finland
* National University of Distance Education
* Netherlands Library Service
* Nordic Council for Scientific Information
* Norwegian National Library
* Oslo College
* Oxford University - Radcliffe Science Library and Bodleian Library
* Oulu University
* Polytechnical University of Catalunya
* Roskilde Public Library
* Roskilde University
* Royal Institute of Technology
* Royal Library of Belgium
* Royal Library of Denmark
* Royal Netherlands Academy of Arts and Sciences
* Rudjer Boskovic Institute
* St. Patrick's College - Maynooth
* St. Petersburg State Technical University
* Science Museum, London
* Scuola Normale Superiore
* Sikeborg Library
* Slagelse Public Library
* State and University Library of Aarhus
* State and University Library of Brement
* State and University Library of Göttingen
* State Research Library of Ceske Budejovice
* Szeged University
* Technical College of Darmstadt
* Technical Research Centre of Finland
* Technical University of Braunschweig
* Technical University of Darmstadt
* Technical University of Denmark
* Technical University of Munich
* Technical University of Wroclaw
* Tilburg University
* Trondheim Public Library
* University and Regional Library of Bonn
* University and Regional Library of Münster
* University and Regional Library of Düsseldorf
* University College - Borås
* University College - Galway
* University College - Karlskrona/Ronneby
* University College London
* University Institute of Architecture in Venice
* University Luigi Bocconi
* University of Abertay Dundee
* University of Amsterdam
* University of Angers
* University of Antwerp
* University of Bamberg
* University of Basel

* University of Bayreuth
* University of Berne
* University of Bielefeld
* University of Birmingham
* University of Central Lancashire
* University of Córdoba
* University of Derby
* University of Dortmund
* University of Dundee
* University of East Anglia
* University of Essex, Albert Sloman Library
* University of Freiburg
* University of Gießen
* University of Graz
* University of Hamburg
* University of Hannover
* University of Innsbruck
* University of Karlsruhe
* University of Lapland
* University of Leipzig
* University of Leoben
* University of Liege
* University of Limerick
* University of Linz
* University of London - Royal Holloway
* University of Luleå
* University of Magdeburg Faculty of Medicine
* University of Milan
* University of Minho
* University of Navarra
* University of Oslo
* University of Oslo Informatics Library
* University of Passau
* University of Potsdam
* University of Regensburg
* University of Rome
* University of Seville
* University of Siegen
* University of Stuttgart
* University of Surrey
* University of Tartu
* University of Technology and Agriculture - Bydgoszcz
* University of Trier
* University of Tromsø
* University of Utrecht
* University of Veterinary Medicine
* University of Vienna
* University of Wales - Aberystwyth
* University of Wales - Cardiff
* University of Wales - Swansea
* University of Wolverhampton
* University of Wuppertal
* University of Zaragoza
* University of Zurich
* Uppsala University
* Vilnius University
* Vlaardingen Public Library
* Wageningen Agricultural University
* York University

Oltre ai cataloghi delle biblioteche, nella rete troviamo anche cataloghi generali di riviste, come il grande archivio *Uncover* della *Carl Corporation*[105], l'archivio di «Gnomon»[106], e, per citare uno degli strumenti più importanti, i *Current Contents*[107], un archivio di articoli di rivista con abstracts, che, sviluppatosi in ambito medico-scientifico, si è allargato fino a riviste di filologia classica ed è destinato a diventare l'archivio completo delle riviste pubblicate a livello mondiale.

Con la compilazione di una semplice scheda informativa sarà così possibile avere informazioni su un qualunque argomento trattato negli articoli spogliati. Si veda come esempio la scheda predisposta dal CIB[108] di Bologna per l'interrogazione dell'archivio.

[105] http://ww.carl.org
[106] http://www.gnomon.ku-eichstaett.de
[107] http://www.isinet.com
[108] http://liber.cib.unibo.it



Un altro tipo di archivio che va prendendo piede è l'archivio degli editori, alcuni dei quali stanno rendendo disponibili gli indici dei volumi stampati, com'è il caso dell'editore John Benjamins[109] di Amsterdam. Un altro esempio è l'Oxford University Press[110], che propone, oltre ai servizi

[109] gopher://benjamins.titlenet.com:6400
[110] http://www.oup.co.uk

informativi generali, l'indice delle riviste «Classical Quarterly» e «Classical Review».

Infine, bisogna ricordare il diffondersi delle riviste elettroniche. Intendiamo per riviste elettroniche non solo le riviste che mettono in rete i loro indici e i loro abstracts, ma riviste nate esclusivamente per la rete. Tra queste vale la pena di segnalare «Arachnion»[111], edito da Maurizio Lana ed Emanuele Narducci, «Traditio»[112], e i «Transactions of the American Philological Association»[113].

Molti di questi archivi esistono su disco, e quindi sono consultabili sul proprio personal in ogni momento. Averli in rete vuol dire abbattere i costi per l'acquisto, e renderli accessibili a chiunque, al di là delle caratteristiche tecniche del suo computer. In un domani non troppo lontano, potrebbe essere realtà la diffusione, auspicata da alcuni (si intende grosse case informatiche, come la Digital), di terminali di rete in ogni casa, semplici da usare come un normale televisore dotato di telecomando[114]. Questo terminale costerebbe poco, si diffonderebbe rapidamente, dando a tutti uno strumento di navigazione semplice e potente. Consultare la Library of Congress da casa propria sarà alla portata di qualunque studente di liceo, mentre resterà patrimonio dei più fortunati il poter dire di esserci stati. Fisicamente, si intende, compresa la degustazione di un orribile caffè americano della macchinetta in servizio.

Bologna — Francesco Citti
Bologna — Tommaso Del Vecchio
Udine — Andrea Tabarroni

[111] V. supra, nota 21.

[112] http://ccat.sas.upenn.edu/jod/traditio/traditio.html

[113] gopher://ccat.sas.upenn.edu:70/11/journals/TAPA

[114] È appena uscita la prima macchina di questa generazione, un IBM dal costo di 700 USD.