

Una questione metodologica: la modellizzazione

Il trasferimento di un *corpus* documentario su un computer pone allo studioso, al di là dei problemi pratici, numerosi quesiti teorici¹.

La teoria e la prassi: quante riflessioni sono state oziosamente proposte su questo antichissimo rapporto dialettico, ma quale rivoluzione radicale e dirompente ha prodotto l'informatica, per la quale il detto è fatto, l'algoritmo è la propria realizzazione e per conseguenza, chi non teorizza, o teorizza male, produce danni².

La citazione, tratta dalla bella prefazione di Tito Orlandi al saggio di Raul Mordenti sulla critica informatica dei testi, è un monito per lo storico ad interrogarsi – a fronte dei cambiamenti in atto nel processo di raccolta, decodifica, interpretazione ed esegesi delle fonti mediate dall'utilizzo del *medium* informatico – sui problemi che vengono posti all'epistemologia della storia in relazione alla metodologia e alla strumentazione adottate. Le implicazioni dell'uso del computer nella raccolta, nell'analisi e nell'elaborazione di dati storici sono infatti metodologiche, e non costituiscono un momento autonomo, privo di interferenze con il canale informativo e con la costruzione digitale di una fonte storica. Nel caso della codifica digitale infatti, il dato vincola l'interpretazione così come l'interpretazione vincola il dato.

In verità, il problema affrontato dall'informatica storica nel trasporre l'informazione contenuta in fonti documentarie antiche in forme più congeniali agli strumenti concettuali di interpretazione e analisi dello studioso moderno «è per molti aspetti lo stesso problema che affronta la critica testuale nell'editare un testo a tradizione manoscritta in forme più

¹ Trovandosi al crocevia degli scambi tematici più disparati, l'Informatica Umanistica si trova oggi quasi vocata a ricomporre una relazione perduta tra i due mondi del sapere. Ne costituisce anzi, il punto di contatto, poiché la sua essenza è principalmente metodologica e non applicativa. In questa direzione, il primo testo teorico che si è interessato al rapporto tra filologia e informatica è stato *La critique des textes et son automatisisation* di Dom Jacques Froger, il quale ha mostrato come il calcolatore sia solamente uno strumento al servizio del filologo e non un suo sostituto, cfr. D.J. FROGER, *La critique des textes et son automatisisation*, Paris, Dunod 1968 (Initiation aux nouveautes de la science, 7); ma una svolta fondamentale si è avuta al Convegno Internazionale di Parigi su *La pratique des ordinateurs dans la critique des textes* del 1978 dove, grazie anche alla partecipazione di Cesare Segre, il computer è stato presentato come un generatore di problemi, soprattutto nel campo della trascrizione e della codifica editiva. I partecipanti a questo convegno, che avevano usato procedimenti informatici in ecdotica, cercarono di fare il punto sulla situazione e sulle prospettive in questo campo, evidenziando una crisi non tecnica ma teorica: una delle ragioni, la diffidenza dei filologi tradizionalis, cfr. *La pratique des ordinateurs dans la critique des textes*. Colloque international du CNRS, nr. 579 (Paris, 29-31 mars 1978), eds. J. IRIGOIN, Paris, Editions du CNRS 1979 (Colloques internationaux du Centre national de la recherche scientifique).

² T. ORLANDI Prefazione a R. MORDENTI, *Informatica e critica dei testi*, Roma, Bulzoni 2001 (Informatica e discipline umanistiche, 10), p. 8. Il computer e il digitale sono indubbiamente contraddistinti da una ambiguità di fondo, data dalla loro natura ibrida, che li mostra al tempo stesso come strumenti e come ambiente nel quale siamo immersi, come oggetti e come memoria, come materia e come informazione: un intreccio tra *technè* e *d'episteme* che crea un disagio cognitivo.

familiari al lettore moderno»³. Trascrivere e codificare hanno sempre messo lo storico di fronte a delle scelte fondative. In ambiente informatico tuttavia, si assiste ad un passaggio successivo: concepire un progetto editoriale del tutto sganciato, almeno virtualmente, dalla tecnologia della stampa, consente – e al tempo stesso costringe – a ripensare radicalmente la nozione stessa di testo. Se infatti,

la macchina legge, si tratta evidentemente anche di scrivere per la macchina e, più in generale, di tenere conto della macchina soprattutto quando si produce quella forma fortissima e – in senso proprio fondamentale – di scrittura che è l'edizione di un testo⁴.

In altre parole, un testo che può e deve essere letto da un computer deve essere strutturato come testo *Machine Readable Form*: un concetto assai diverso, e assai più impegnativo, di un testo semplicemente digitalizzato o scannerizzato, e che porta con sé modifiche capitali, gravide di conseguenze ecdotiche e operative. La codifica di un testo su supporto digitale, come ogni processo di rappresentazione che coinvolge il computer, è mediata da un apposito linguaggio formale, vincolato da rigide regole sintattiche che richiedono la risoluzione di ogni ambiguità. Un linguaggio formale è costituito da un insieme organizzato di simboli che permettono di riscrivere, formalizzandolo, l'universo che si vuole descrivere, e da un insieme coerente di regole che governano le modalità di combinazione dei vari simboli e le produzioni che possono essere generate all'interno del linguaggio: è quindi un codice che, per le sue caratteristiche di non ambiguità e precisione sintattica, può essere utilizzato per affidare la realizzazione di un determinato compito ad un procedimento automatizzato. Da ciò discende la sua natura sostanzialmente epistemologica: la rappresentazione informatica come processo semiotico che implica dei processi interpretativi.

Nei paragrafi precedenti si è più volte sottolineato come la codifica sia un processo assai più complesso della semplice e meccanica correlazione biunivoca di strutture rappresentazionali, e che impone un'approfondita riflessione sul problema testuale⁵. Se il computer è un manipolatore di simboli, il rapporto che con esso intrattiene il codificatore,

³ D. BUZZETTI, "Historical software" e filologia. Due recenti proposte teoriche di Manfred Thaller, in *Schede Umanistiche*, 3 (1993) 2, pp. 181-190:182.

⁴ R. MORDENTI, *Filologia e computer*, in *Macchine per leggere. Tradizioni e nuove tecnologie per comprendere i testi* cit., pp. 53-68:55-56.

⁵ Cfr. F. CIOTTI, *Testo, rappresentazione e computer. Contributi per una teoria della codifica informatica dei testi*, in *Internet e le Muse. La rivoluzione digitale nella cultura umanistica*, a cura di P. NEROZZI BELLMAN, Milano, Mimesis 1997 (Eterotropie), pp. 221-250.

qualora la codifica sia intesa come processo interpretativo piuttosto che riproduttivo, costituisce un linguaggio teorico, usato per dare forma consapevole a una teoria ontologica del testo, cioè a una determinata concezione sulla struttura del testo. Ponendosi la domanda “*What is text, really?*”, Steven DeRose e i coautori dell’omonimo saggio hanno risposto definendo l’ontologia formale della sua rappresentazione: un testo è *OHCO* (*Ordered Hierarchy of Content Objects*) – una gerarchia ordinata di parti dotate di contenuto – e ha una struttura a grafo che ne organizza gli insiemi come oggetti astratti, ordinati in modo sequenziale e lineare e rappresentati dalle segmentazioni successive e contingue del discorso⁶. Con la definizione, è evidente, muta la stessa visione unitaria del testo, che viene atomizzato – per così dire – in una serie ampia, anche se non necessariamente inesauribile, di costituenti informativi, ciascuno dei quali acquista, o spesso riacquista, valenze diverse⁷. Secondo questa interpretazione inoltre, la scrittura stessa viene intesa come un processo, un’attività nella quale poter descrivere tutto quello che si è fatto, spiegando la relazione tra gli elementi che la compongono e, parallelamente, chiarendo le procedure utilizzate per la costruzione di un modello dei contenuti. Il criterio è valido sempre e comunque, dalla scrittura in senso lato all’edizione di una fonte storica. La trascrizione di un manoscritto, che nella configurazione tradizionale del processo filologico non ha mai costituito un momento scientificamente forte, ma al contrario era ritenuto quasi un preliminare meccanico e noioso dell’attività editiva vera e propria, nelle procedure ecdotiche fondate su un sistema informatico diventa il perno decisivo: non solo perché – di fatto – è l’atto più costoso in termini di tempo/uomo, ma soprattutto perché configura l’aspetto cruciale della codifica. Da un punto di vista teorico, la trascrizione come codifica costringe a interrogarsi analiticamente su cosa si sta codificando, o ricodificando, o decodificando; a scomporre questo qualcosa in elementi discreti, a ordinare in modo sequenziale le operazioni effettuate, evitando ambiguità, contraddizioni, ridondanze e costringendo – soprattutto – a formulare tutto ciò in modo rigoroso⁸. In sostanza, si tratta di fornire un’adeguata quantità di informazione sull’informazione, trasformando il file di un testo codificato in uno strumento

⁶ S.J. DEROSE, D.D. DURAND, E. MYLONAS, A.H. RENEAR, *What is Text, Really?*, in *Journal of Computing in Higher Education*, 2 (1990) 1, pp. 3-26.

⁷ Per queste interpretazioni v. soprattutto T. ORLANDI, *Ripartiamo dai diasistemi*, in *I nuovi orizzonti della filologia. Ecdotica, critica testuale, editoria scientifica e mezzi informatici elettronici*. Atti del Convegno Internazionale in collaborazione con l’Associazione Internazionale per gli Studi di Lingua e Letteratura Italiana (Roma, 27-29 maggio 1998), Roma, Accademia Nazionale dei Lincei 1999 (Atti dei Convegni lincei, 151), pp. 87-101.

⁸ Per queste considerazioni cfr. R. MORDENTI, *Informatica e critica dei testi* cit., pp. 29-30.

euristico che obbliga, affinché risulti efficace, a dichiarare chiaramente le scelte, le pratiche e la metodologia seguite nel lavoro ermeneutico.

La versatilità dei linguaggi di marcatura nel generare definizioni di diverse tipologie di documenti ha prodotto la convinzione diffusa che tali dispositivi potessero costituire uno strumento neutro, attraverso il quale rappresentare la conoscenza implicita delle fonti, dagli ordini commerciali ai manoscritti latini. Nella grande maggioranza dei casi gli umanisti hanno considerato la codifica un problema tecnico, un argomento ozioso – e noioso – del tutto privo di conseguenze ed effetti sulle loro discipline. Niente di più sbagliato:

gli strumenti di ricerca e le descrizioni veicolano sempre ideologie, visioni del mondo e soprattutto della storia; tendono insomma a trasmettere un'immagine della documentazione che contiene nuclei di interpretazione storiografica più o meno forti, ma comunque in grado di condizionare il modo in cui essa, e il processo della sua produzione, vengono percepiti dal ricercatore. Qualsiasi ipotesi di strutturazione dell'informazione contiene, quindi, implicitamente, se non esplicitamente, una proposta di attribuzione di senso all'informazione stessa: ciò vale ancora di più quando questa venga inserita in una cornice tecnologica⁹.

D'altro canto, chi ha pienamente compreso la non neutralità di una codifica selettiva, ha mosso però una feroce critica alla soggettività delle scelte del codificatore, partendo dalla considerazione che, se è vero che ogni indagine storica opera – naturalmente – una selezione finalizzata dei documenti disponibili e delle informazioni in essi contenute¹⁰, privilegiandone alcune in base alle esigenze poste dalla ricerca, una selezione suggerita o in qualche modo predeterminata dal computer costituisca invece un problema di metodo. L'argomentazione in realtà è mal posta e a conti fatti i presupposti da cui parte si rivelano errati. Non diversamente da ogni altro strumento di lavoro, il computer non necessariamente condiziona la professionalità di chi se ne serve:

⁹ S. VITALI, "Archivi on line": qualche riflessione metodologica, in Workshop su Archivi storici e archivi digitali tra ricerca e comunicazione, Firenze, Dipartimento di Studi Storici e Geografici dell'Università di Firenze, 2000; disponibile on line all'indirizzo:<http://www.storia.unifi.it/storinforma/Ws/archivi/ws-archivi-alia.htm#Materiali>. Lo stesso Vitali ha sottolineato come, con la codifica elettronica, non si parla più di estrarre dalle fonti dati quantitativi da sottoporre all'elaborazione, ma di esplorare le strutture informative presenti nelle fonti e riorganizzarle, recuperarle, aggregarle secondo i punti di vista suggeriti dalle ipotesi di ricerca dello storico, attivando ed evidenziando quelle connessioni prima sconosciute o scarsamente evidenti, cfr. S. VITALI, *Passato digitale. Le fonti dello storico nell'era del computer* cit., p. 31.

¹⁰ «La soggettività dello studioso è un presupposto ineliminabile dello studio storico, così come la soggettività lo è di qualsiasi forma o attività di conoscenza», G. GALASSO, *Nient'altro che storia. Saggi di teoria e metodologia della storia* cit., p. 115.

a condizione di conoscerlo, padroneggiarlo e soprattutto non attribuirgli prerogative improprie, non c'è alcuna fondata ragione perché esso alteri la qualità di un problema¹¹.

Quello che è nuovo nell'uso dello strumento informatico consiste anzi nel fatto che i passaggi in cui si evita l'intervento umano – escludendo quelli di ordine puramente pratico – sono quelli che in passato venivano esplicitati per lo più mediante operazioni mentali non dichiarate. Citando Fernand Braudel: «il programma del programmatore: questo è quel che mi interessa»¹². Paradossalmente, lo storico dimostrava il proprio interesse verso quella soggettività che tanta parte della critica ha fustigato, verso quella competenza dello studioso come meta-struttura, non infondata o arbitraria, proponendo – quasi inconsapevolmente – la metodologia più appropriata in ambiente informatico: motivare ogni decisione, illustrarla in modo dettagliato e in tutti i suoi aspetti, conseguendo per questa via un'attendibilità documentabile e forse maggiore rispetto a casi pretenziosamente oggettivi.

Fin qui, i nodi teorici di un antico problema: quello della delimitazione pragmatica del processo di memorizzazione e trasposizione di un testo, in nessun caso – e tanto meno nel caso informatico – riducibile ad una semplice e meccanica trasposizione. Di fatto, quanto giace dietro le quinte dell'edizione ha una struttura definita anzitutto dalla necessità, da un lato, di scomporre l'informazione in elementi enucleabili e dall'altro, di mantenere i contenuti quanto più possibile indipendenti da qualsiasi forma specifica. Il duplice profilo dei documenti – forma e contenuto – che in ambiente cartaceo mantiene un legame profondo, tanto che quando si agisce sull'uno si modifica anche l'altro, in ambiente elettronico si scinde inesorabilmente, creando inedite possibilità applicative. Al pari di altri supporti tradizionali, quello cartaceo realizza infatti un'integrazione profonda tra forma e contenuto: le strutture logiche interne ai documenti sono esplicitate grazie alle caratteristiche grafiche o, ancora più astrattamente, in virtù del linguaggio semiotico adottato, permettendo la comprensione dei vari livelli di significato che il documento stesso propone. Una caratteristica del genere è il fondamento su cui si basa, ad esempio, una disciplina di lunga tradizione, la diplomatica, dove l'analisi delle forme permette di individuare le strutture e viceversa, la presenza di strutture è roborata dall'uso di determinate forme. Il documento elettronico permette invece una gestione separata di forma

¹¹ T. DETTI, *Lo storico e il computer: approssimazioni*, in *Storia & Computer. Alla ricerca del passato con l'informatica* cit., pp. 83-104:104.

¹² F. BRAUDEL, *Scritti sulla storia*, Milano, Bompiani 1973 (Tascabili Bompiani, 256), p. 27.

e contenuto: i caratteri estrinseci, compresa la forma o il tipo di materiale, sono in gran parte una funzione del software. Non si tratta cioè di attributi permanenti del documento medesimo, ma solo delle specifiche opzioni di visualizzazione, di stampa o di stile selezionati di volta in volta sulla base delle esigenze specifiche, separati quindi dal contenuto informativo e dal contesto del documento.



Struttura di un documento in ambiente cartaceo e struttura di un documento in ambiente elettronico

La questione va però affrontata anche dal punto di vista pratico: come concretizzare i presupposti metodologici fin qui analizzati? Come esplicitare le scelte, gli interventi soggettivi, le metastrutture realizzate nell'ambito di un'edizione digitale? Come sciogliere quegli aspetti problematici, che riguardano – ad esempio – la funzione interpretativa dei linguaggi di marcatura? La soluzione prospettata dalla codifica elettronica passa per due termini fondamentali e strettamente intrecciati tra di loro: astrazione e modellizzazione.

Nell'introdurre i due concetti, si legga in prima istanza la semplice ma efficace enunciazione di Rosenblueth e Wiener, fondatori della cibernetica:

Nessuna parte costitutiva dell'universo è talmente semplice da lasciarsi afferrare e controllare senza astrazione. Quest'ultima consiste nel prendere la porzione dell'universo considerata e nel sostituirla con un modello avente struttura analoga, ma più semplice¹³.

La modellizzazione è dunque un processo di astrazione¹⁴ virtualmente interminabile ma in grado di individuare, all'interno dei testi, le due macrocategorie di struttura e contenuto.

¹³ A. ROSENBLUETH, N. WIENER, *The role of the Models in Science*, in *Philosophical Science*, 12 (1945), pp. 316-321:320.

¹⁴ Le astrazioni più comuni sono: la classificazione, un processo astrattivo mediante il quale si estraggono le caratteristiche comuni – ritenute rilevanti – di entità diverse e si definisce una classe superiore che le rappresenta;

Costruire un modello per memorizzare testi significa quindi da un lato, individuare correttamente ed esaustivamente i dati pertinenti della fonte, ovvero tutti gli elementi significanti del testo, anche quelli che ad una normale lettura non vengono considerati rilevanti perché immersi nella sfera dell'inconsapevole (sistema semantico del codice); e dall'altro, progettare un sistema di correlazioni biunivoche tra tali elementi e unità espressive discrete, compatibili con la natura fisica del canale e del destinatario, in grado di rappresentare nel passaggio dalla forma cartacea a quella elettronica tutte le informazioni codificate in modo trasparente. È in queste due attività che lo scrupoloso rispetto della fonte, presupposto ineludibile dell'approccio storico-filologico, si fonde con il dichiarato intervento soggettivo dello studioso. Il modello è una visione astratta di una realtà complessa e come tale soggetto a inevitabili nonché opportune semplificazioni: il livello di analiticità adottato deve essere costantemente riferito agli obiettivi e ai profili d'indagine, onde evitare dettagli irrilevanti per la corretta comprensione degli oggetti documentari, delle loro caratteristiche e delle loro interazioni. Tuttavia, l'elemento discrezionale non inficia il rigore di una metodologia che affonda le sue radici nei linguaggi formali, e che pertanto riduce drasticamente sia le possibilità di descrizione incoerente sia le ambiguità e le ridondanze del linguaggio naturale. Costruire un modello significa inoltre organizzare, partendo da un insieme di dati e di osservazioni sperimentali, uno schema assiomatico all'interno del quale, per mezzo di una serie di regole e simboli stabiliti, è possibile dedurre una serie di proprietà che sono vere nel modello. Non ha quindi senso chiedersi se un modello sia vero o falso in senso assoluto: è uno strumento di comodo, utile per l'interpretazione ma soggetto a modificarsi con l'accrescersi delle conoscenze e delle variabili in atto durante la ricerca.

Nel caso qui prospettato – quello di una fonte dalle forti connotazioni storico-diplomatiche – modellizzare ha inoltre significato:

- individuare un prototipo dei documenti conforme agli obiettivi dell'indagine e derivato dalle esigenze analitiche dell'interprete;
- scegliere un linguaggio funzionale con cui implementare tale schema.

l'aggregazione, procedimento attraverso il quale si definisce invece una classe di oggetti integrando le caratteristiche di ognuna delle classi subordinate in una nuova classe, distinta da quelle; infine, la generalizzazione, procedimento mediante il quale si giunge alla definizione di una classe come unione di un insieme di classi che ereditano le caratteristiche della classe superiore. Le astrazioni sono quindi il meccanismo fondamentale di rappresentazione dei modelli, siano essi riferiti ai dati o alle funzioni.

Scontato dire la modellizzazione relativa al codice Vat.Lat.3880 è stata attività delicata, fortemente interpretativa e intenzionale perché – più che soluzioni tecnologiche – ha riguardato le strutture semantiche del testo, l'individuazione e l'utilizzo delle sue componenti logiche e la definizione di sotto-sistemi efficaci ed efficienti, nel segno di una codifica pragmaticamente legata alle esigenze poste dalla ricerca. Il problema cruciale è stato infatti chiedersi quali fossero i caratteri irrinunciabili della fonte, quelli da non perdere nel processo di codificazione digitale. È nel cuore di questi interrogativi che «si nascondono nodi concettuali di notevole spessore problematico, che hanno a che fare con la natura dei testi, con quella dei documenti nonché con l'operazione epistemologica che trasforma gli uni e gli altri in fonti storiche»¹⁵. Delineatasi come operazione che ha richiesto una progettualità forte e complessa, la modellizzazione ha quindi avuto come scopo precipuo la creazione di percorsi molteplici ma comunque preordinati e in quanto tale, ha posto come primo problema la selezione di un linguaggio adeguato a rappresentare l'organizzazione strutturale della fonte e i suoi elementi singolarizzanti. Centrale è risultata, inoltre, l'analisi del livello descrittivo dei metadati stabiliti durante la modellizzazione, in relazione agli oggetti informativi da trattare: la granularità¹⁶, proprietà dei documenti digitali come di quelli analogici, il cui livello non è una grandezza oggettiva e immutabile, ma può variare a seconda del punto di vista del suo fruitore o creatore, del contesto in cui è inserito e degli strumenti con cui è gestito. Da qui, in conseguenza, la discrezionalità più volte dichiarata e

¹⁵ S. VITALI, *Passato digitale. Le fonti dello storico nell'era del computer* cit., p. 56.

¹⁶ Granulare è un documento digitale articolato in più parti componenti, che consente al lettore piste autonome di lettura e di ricerca. L'analisi del termine, effettuata dal *Grande dizionario della lingua italiana* di Salvatore Battaglia, ne rivela però l'ambiguità di fondo: granulare è definito come ciò che è «ridotto in granuli; composto di granuli; costituito da elementi a forma di granuli (una sostanza). Per estensione: ruvido, scabro (una superficie); granuloso». Sono attestati riferimenti nell'ambito della geologia («struttura granulare o granitica o granitoide»), della biologia («che presenta granulazioni»), dell'elettronica. La serie dei significati si allunga di giorno in giorno, e la grande maggioranza fa riferimento all'ambito informatico, in particolare alla progettazione dei data base, alla programmazione parallela, al calcolo parallelo, all'elaborazione del testo e all'organizzazione delle memorie. Il *Dictionary of Computing* tenta di sintetizzare il significato del termine: «A measure of the size of the segments into which memory is divided for purposes of either memory protection or virtual-memory management». Anche *Whatis.com*, ottimo strumento per muoversi tra le definizioni relative al mondo dell'*information technology* e di Internet, tenta di tirare le fila, con una definizione ampia del termine in grado di abbracciare tutti i diversi settori disciplinari: «Granularity is the relative size, scale, level of detail, or depth of penetration that characterizes an object or activity. It may help to think of it as: which type of 'granule' are we looking at? This term is used in astronomy, photography, physics, linguistics, and fairly often in information technology. It can refer to the level of a hierarchy of objects or actions, to the fineness of detail in a photograph, or to the amount of information that is supplied in describing a person's age. Its meaning is not always immediately clear to those unfamiliar with the context in which it's being used», cfr. <http://whatis.techtarget.com/definition/0,289893,sid9gci212209,00.html>. In ogni caso, in tutte le discipline, la granularità indica il livello di dettaglio, la precisione con la quale si può cogliere un oggetto o un'attività: il problema ad essa connesso è il trattamento di un numero rilevante di dati, che debbono essere resi intellegibili in fase di osservazione o rappresentazione, selezionando un particolare livello di dettaglio. Granularità è dunque il termine che l'informatica adotta per fare riferimento allo spessore di frantumazione dell'informazione.

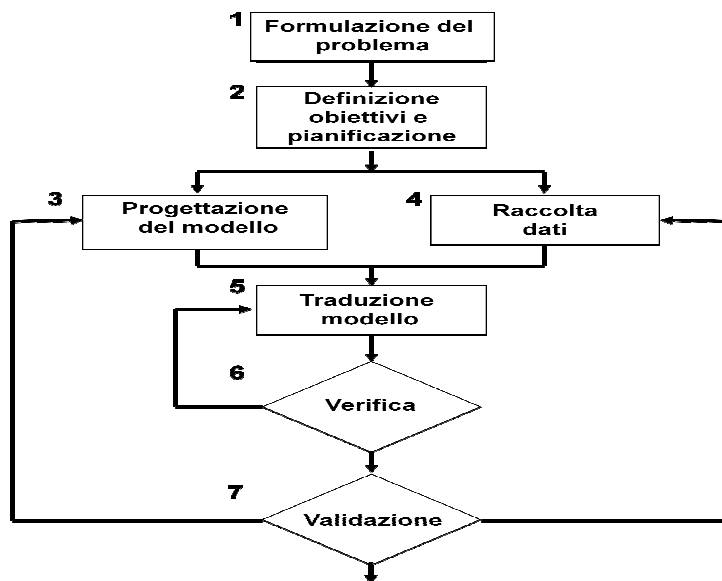
la necessità di una integrazione su base individuale, pur inserite all'interno di un processo che include principi condivisi e una terminologia trasversale ai numerosi standard proposti. Come ha giustamente sottolineato Michele Ansani,

la necessità di una sperimentazione graduale, relativamente indolore, attenta ai linguaggi di codifica ma affatto destabilizzante rispetto alle pratiche tradizionali della critica documentaria, non potrà sortire particolari successi se orientata al collaudo di standard universali e 'chiusi', applicati o presumibilmente applicabili a qualsiasi fenomeno testuale: occorrerà tenere sempre ben presente che un testo giuridico, un testo letterario, un testo documentario, ovvero un *corpus* di testi difficilmente potranno essere ingabbiati entro strutture logico-semantiche o di semplice descrizioni uniformi e ripetitive se non rinunciando a esplicitarne, nelle scelte della codifica, gli elementi legati alle rispettive specificità e storicità¹⁷.

In questo senso la modellizzazione preventiva alla codifica effettiva ha implicato delle scelte e quindi, delle decisioni da prendere in anticipo rispetto alle domande cui il testo informaticamente trattato avrebbe dovuto rispondere una volta codificato. Tuttavia, ed è bene sottolinearlo, la costruzione di un modello non è mai un processo rigidamente predeterminato, ma al contrario, configurabile – di volta in volta – in base alle esigenze, agli obiettivi cognitivi e agli interessi personali del codificatore. La modellizzazione dunque, non è un punto di non ritorno, ma un'operazione potenzialmente infinita, perché infinite sono le domande dello storico.

Tenendo in considerazione, tra le proprietà della metodologia adottata, i criteri di generalità (il modello proposto nasce per poter essere utilizzato indipendentemente dalla fonte in questione e dagli strumenti a disposizione), qualità e facilità d'uso, i passi con cui si è raggiunta la modellizzazione della codifica adottata sono stati sintetizzati nel grafico sottostante:

¹⁷ M. ANSANI, *Un passaggio complicato. L'edizione di fonti storiche in rete*, in *Il documento immateriale. Ricerca storica e nuovi linguaggi*, in *L'Indice dei libri del mese*, Dossier 4 (maggio 2000), a cura di G. ABBATISTA e A. ZORZI, disponibile on line all'indirizzo: <http://lastoria.unipv.it/dossier.index.htm>.



La formulazione del problema, presupponendo una definizione sufficientemente precisa del sistema da analizzare e dei dati disponibili è stata, ovviamente, il punto fondamentale per un corretto sviluppo del modello e del piano di tutto il progetto. Lo schema sinteticamente presentato nella figura mostra però come ciascuno dei problemi, pur riferito ad una procedura unitaria, abbia implicazioni diversificate, e richiami ambiti teorici e relative discussioni che portano in numerose e differenti direzioni. Ma il punto finale del discorso è e resta che la codifica per supporto elettronico non è un procedimento che rimane parallelo ad una parte del processo semiotico di comunicazione, in modo da lasciare invariato nella sostanza tale processo: al contrario, essa potrà lasciare invariato il *risultato* di tale processo, cioè l'esatta comprensione del contenuto del messaggio da parte del ricevente, che è quello che si vuole, o comunque che più importa, soltanto se sarà perfettamente corretta proprio dal punto di vista del processo euristico di modellizzazione.

Nella progettazione dell'attività di modellizzazione, dovendo scegliere tra numerose opportunità, si è preferito comunque optare per il più classico modello *Entità-Relazione (ER)*, che utilizza il simbolismo grafico per costruire dei diagrammi e descrivere in maniera formale la realtà d'interesse, attraverso cinque strutture di rappresentazione cui corrispondono altrettanti simboli:

Costrutti	Rappresentazione grafica
Entità	
Relazione	
Attributo semplice	
Attributo composto	
Cardinalità di relazione	
Cardinalità di attributo	
Identificatore interno	
Identificatore esterno	
Generalizzazione	
Sottoinsieme	

Particolarmente utili sono state, ovviamente, la classe di *entità*, attraverso le quali sono stati individuati e rappresentati gli elementi con proprietà omogenee nel contesto in esame; quella degli *attributi*, che hanno permesso di descrivere le entità e rappresentare le proprietà della classe; infine, la classe delle *relazioni*, che hanno consentito di individuare i legami logici tra le istanze di una o più entità. È chiaro che una simile organizzazione rende *ER* il modello adatto a descrivere la fonte storica in modo indipendente dall'ambiente informatico che la supporta, presentando un livello di astrazione che da un lato, si rivela indispensabile per adeguarsi alla molteplice realtà presentata, tramite l'uso di un formalismo di valenza universale; dall'altro, richiede un processo di esplicitazione per passare dal modello concettuale dei dati, ovvero quello realizzato tramite le strutture che gli sono proprie, al modello logico, che tiene conto invece dell'ambiente operativo nel quale la fonte verrà inserita. Il computer assume aspetti diversissimi e può essere costruito con materiali diversissimi: ciò che sta alla base di questa diversità è un meccanismo intimo, che guida le procedure operative. Sarà dunque questo "meccanismo", che qui è assunto la forma di un modello, a fornire la chiave per intendere correttamente i rapporti fra gli strumenti informatici e le discipline umanistiche alle quali vengono applicati.