

SULLA CODIFICA DELLE FONTI ARCHEOLOGICHE *

1. LA PROSPETTIVA INFORMATICO-METODOLOGICA

Il valore di questo contributo e delle osservazioni che verranno esposte è fortemente legato al riconoscimento di una disciplina che si occupi non delle singole concrete applicazioni dell'informatica nei differenti settori degli studi umanistici, ma dei problemi che tali applicazioni hanno in comune. In altri termini, si accetta la possibilità che di tali problemi si occupi non uno specialista di informatica, e nemmeno uno specialista di un singolo settore umanistico con competenze informatiche, ma uno studioso di estrazione umanistica che prenda in considerazione quanto vi è di comune fra le metodologie dei singoli settori, dal punto di vista delle applicazioni dell'informatica.

Questo può lasciare inizialmente perplessi, perché posizioni di questo tipo nascondono spesso vaghe e poco utili generalizzazioni. Ritengo tuttavia che vi sia più di un motivo per accettare l'ipotesi proposta. La breve storia dell'informatica umanistica è costruita in larga parte di scambi e rapporti fra studiosi e tecnici di specializzazioni anche molto lontane, che devono capire i reciproci fondamenti. Un filologo non sarà più lontano dall'archeologia di un matematico o di uno statistico o di un ingegnere elettronico; e, se ha riflettuto sulle caratteristiche generali dell'informatica, può avere qualche buona idea anche su discipline di non diretta competenza.

Ad ogni modo il mio inserimento nel dibattito di questo convegno può essere pienamente giustificato soltanto se si accetta che le applicazioni umanistiche dell'informatica finiscano per determinare una sfera di interessi comuni, che produce una comune metodologia; e dunque che la metodologia dell'informatica umanistica si intrecci con le metodologie specifiche delle singole discipline. A sua volta questo presuppone che si accetti che l'informatica non sia semplicemente un insieme di tecnologie, ma una serie di metodologie, e dunque una disciplina con un proprio fondamento teorico, che deve in qualche modo interagire con quello proprio delle singole discipline umanistiche.

Cercherò di darne un esempio, prendendo in considerazione i problemi della codifica della documentazione archeologica, non senza aver aggiunto che molto di quanto dirò dovrebbe valere, con i dovuti aggiustamenti, anche per altre discipline. È necessario tuttavia impostare il ragionamento partendo dai fonda-

* Relazione presentata in «*Archeologia e Calcolatori*». *Incontro di studio sui metodi e le prospettive della ricerca*, Roma, Consiglio Nazionale delle Ricerche, 23 gennaio 1992.

menti dell'archeologia, per vedere in quali settori e in quale modo le metodologie informatiche possano efficacemente integrarsi in tale disciplina. Fortunatamente, pubblicazioni recenti (unite alla personale frequentazione con colleghi archeologi) hanno reso possibile questa operazione. Mi riferisco soprattutto al volume storico-metodologico di Bruce G. Trigger (TRIGGER 1989), a quello metodologico fondamentale di Jean-Claude Gardin (GARDIN 1980), a quello specifico di F. Djindjian (DJINDJIAN 1991); e naturalmente alle prime annate di «Archeologia e Calcolatori», che sono ricche di contributi metodologici.

Se si considera il lavoro dell'archeologo dal punto di vista dell'informatica, non mutano ovviamente i fondamenti della disciplina, ma muta la prospettiva. Ora, secondo questa prospettiva, si può sintetizzare il lavoro dell'archeologo come classificazione, valutazione e spiegazione dei documenti che egli seleziona in quanto utili ai fini della propria ricerca. Detto in termini più informatici, l'archeologia sottopone a procedimenti analitici e sintetici i dati desunti dall'informazione di carattere appunto "archeologico", cioè quella che proviene da artefatti e non da manoscritti¹.

2. LA CODIFICA IN ARCHEOLOGIA

L'attenzione degli studiosi interessati alla metodologia si concentra soprattutto sul trattamento dei dati, e sulla conseguente formalizzazione del ragionamento archeologico. Meno sul passaggio iniziale di definizione e descrizione, o rappresentazione, dei dati, cioè quello che intendiamo appunto per codifica, e che viene in certo modo ritenuto ovvio, e quasi banale. Si nota per esempio che nella storia dell'archeologia lo spazio dato al metodo di descrizione degli oggetti è quasi nullo rispetto a quello dato ai vari metodi di interpretazione².

Fra coloro che dedicano attenzione al problema della codifica sono Gardin e Djindjian. Il primo (che fa della rappresentazione, o "compilation", uno dei

¹ Djindjian (1991, cap. 15: Des méthodes aux formalisations des raisonnements en archéologie, pp. 325-340) parla di « formalisations des raisonnements », che coincide con quello che intendo con "punto di vista informatico". Egli tuttavia prende in considerazione i metodi, piuttosto che l'oggetto (parla infatti genericamente di « reconstruire le passé », p. 326) e quindi può affermare che « l'archéologie n'existe pas, ou n'existe pas encore... il existe des archéologies » che via via si rifanno al concetto tradizionale di « archéologie des antiquités », o alle scienze naturali, o all'antropologia, o allo strutturalismo, etc. A mio avviso, nel distinguere i diversi ambiti applicativi, l'informatica (umanistica) deve piuttosto prendere in considerazione gli oggetti della ricerca. I metodi finiscono per essere comuni e "trasversali".

² Trigger (1989) sostanzialmente lo trascura. In effetti ancora l'archeologia informatizzata non sembra rientrare nell'orizzonte di una "storia", e solo con l'avvento delle ricerche aiutate dal calcolatore si è avvertita la necessità di formalizzare i linguaggi descrittivi. Cfr. per esempio i lavori pionieristici di J.-Cl. GARDIN e collaboratori (serie dei *Codes pour l'analyse...*, Paris, CNRS, 1976-78).

due momenti essenziali dell'attività archeologica) sottolinea come non sia sufficiente una riproduzione che diremmo di tipo analogico, ma sia necessaria per la scienza una rappresentazione simbolica mediante un linguaggio. Egli discute a lungo quale tipo di linguaggio debba essere utilizzato a questo scopo (GARDIN 1980, cap. 3: *The Analysis of Compilations*).

Djindjian ritiene riduttivo questo punto di vista, valido cioè soprattutto nella prospettiva della realizzazione di banche dati, e propone invece come codifica la scelta di quelle caratteristiche che consentano un determinato tipo di analisi: preferibilmente quantitativa, ma eventualmente anche morfologica, seriale, etc... (DJINDJIAN 1991, 99-100).

Da parte nostra, pensiamo che una riflessione sulla codifica vada fatta partendo da presupposti più generali, non solo di ambito archeologico, ma genericamente informatico.

Diremo intanto che per l'informatica intesa come tecnologia³, (che pure è fondamentale, perché i procedimenti informatici sono oggi presi in considerazioni in quanto consentono l'uso dei computer per aiutare lo studioso in alcuni importanti momenti della sua ricerca) la codifica è un momento necessario per poter "scrivere" i dati su supporto magnetico, e quindi gestirli per mezzo dei computer. Il computer diventerà il principale veicolo di diffusione scientifica, e il principale ausilio nel lavoro quotidiano dello studioso. Dunque non se ne può prescindere.

In questo senso la codifica consiste semplicemente nella possibilità di trasferire all'interno di una memoria magnetica i dati da sottoporre agli appositi programmi di un computer, così come sono già preparati sulla carta. Questo passaggio (che propriamente noi chiameremmo di "transcodifica", cioè di puro passaggio da un codice all'altro) si risolve semplicemente nell'uso corretto della tastiera, che oggi è lo strumento naturale per l'*input* dei dati. Anche questo semplice passaggio non è esente da problemi, perché vi sono molti tipi di tastiera, che offrono molti tipi di corrispondenza fra i simboli riportati sui tasti e le sequenze magnetiche (o sequenze di bit) prodotte nella memoria del computer. Ma non è di questi problemi che intendiamo trattare in questa sede.

Prima ancora esiste un problema teorico della codifica, che è sempre stato conosciuto, ma di solito è stato trascurato perché risolto alla buona, in quanto precedentemente il messaggio archeologico era rivolto sempre a intelligenze umane, che compivano implicitamente grossi passaggi di interpretazione. Con il supporto cartaceo questo era consentito, perché i dati su carta possono essere letti solo da intelligenze umane, che compiono su di essi una serie di elaborazio-

³ Sulla distinzione fra informatica come tecnologia e "informatica teorica" cfr. ORLANDI 1992, 1-22 (sopr. 11-16).

ni ovvie, e inconse, prima di entrare nel merito scientifico. I dati su supporto magnetico vengono invece analizzati da un computer, che non andrà mai oltre ciò che è perfettamente esplicito nei dati stessi.

Si può dire che, come si può benissimo usare una lingua senza esplicitarne le regole grammaticali, ma se si vuole stabilire la correttezza di un testo che si pretende espresso in quella lingua, bisognerà chiarirne la struttura; così si può usare efficacemente della codifica, senza teorizzarla, solo finché non sorga la necessità di discutere se un certo lavoro di codifica sia stato fatto correttamente o no.

3. DEFINIZIONE DELLA CODIFICA

Definiremo dunque in senso tecnico la codifica come il procedimento per mezzo del quale i dati che compongono una informazione vengono espressi mediante un codice (possiamo anche chiamarlo alfabeto) in modo da poter costituire un messaggio, cioè da poter essere trasmessi fra soggetti, che non sono necessariamente persone umane (ORLANDI 1990, cap. 2).

Si comprende subito che, definita in questo modo così generale, la codifica da un lato presuppone un accordo almeno di massima su concetti niente affatto semplici o banali: informazione, messaggio, codice, trasmissione, etc.; dall'altro una serie di operazioni all'apparenza intuitive, ma in realtà assai problematiche: scelta del codice, verifica della sua correttezza, dichiarazione della correlazione fra codice e dati; effettuazione operativa della codifica, etc.

Per quanto riguarda il primo punto, una riflessione sui concetti che entrano in gioco in una teoria della codifica ci porta direttamente al settore studiato dalla semiotica, disciplina se si vuole controversa per aver in qualche caso dato luogo a tediose e apparentemente inconcludenti elucubrazioni, ma i cui problemi non si possono trascurare, soprattutto nell'ambito di una attività il cui compito è quello di trattare informazione. Ad ogni modo, non possiamo farne qui più di un accenno (per una discussione più ampia cfr. ADAMO 1987).

Il secondo punto, quello operativo, riguarda invece da vicino il nostro tema, e su questo intendiamo proporre una serie di considerazioni. Una prima, di carattere preliminare, è che in questo modo il procedimento della codifica non riguarda solo un trasferimento di segni, cioè l'utilizzazione di un codice al posto di un altro (su ciò cfr. sotto: vari tipi di operazioni di codifica...), ma anche la valutazione dei dati che si intendono codificare. Infatti, per verificare la validità e la congruità di una codifica, occorrerà avere un'idea ben precisa non solo del codice, ma anche dei dati; e occorrerà che questi dati siano analizzabili come unità ben specificate.

4. DAL CONTINUO AL DISCRETO

La codifica presuppone l'individuazione consapevole e meditata di elementi discreti in un universo continuo, quale si può definire l'oggetto di uno studio. Questo modo di esprimersi è tipicamente informatico (dell'informatica teorica), ma il procedimento è ben presente nelle riflessioni storiche e metodologiche sull'archeologia, anche se i diversi autori lo definiscono in modo più convenzionale. Così lo storico dell'archeologia (TRIGGER 1989, 196) nota come l'evoluzione scientifica dell'archeologia (culture-historical approach, opposto all'antiquaria e all'evoluzionismo) porta come conseguenza anche una maggiore attenzione alla definizione degli ambienti da cui provengono i reperti, e alla considerazione di tutti i reperti, non soltanto di alcuni.

A sua volta, Albert Spaulding nota che il concetto di manufatto « fornisce la classe di entità con cui si confronta l'archeologia, gli oggetti o i resti di oggetti che mostrano gli attributi di una attività socialmente orientata. (...) La possibilità di riconoscere i manufatti implica la capacità di riconoscere quegli attributi dei manufatti che rappresentano una modificazione umana coerente, cosicché noi siamo condotti ad una categoria fondamentale, l'insieme degli attributi culturalmente significativi, riuniti insieme dal loro esistere in un singolo oggetto » (SPAULDING 1971, 4).

Il teorico Djindjian si sofferma a più riprese sul carattere anche di scelta che deve avere l'operazione di codifica, e in particolare nota che « È indispensabile sottolineare che le descrizioni libere di un reperto rendono inefficace ogni tentativo di costruzione tipologica. La ricerca di una ipotetica esaustività della descrizione, ottenuta moltiplicando le variabili descrittive, è un'illusione pericolosa che occorre denunciare energicamente. Ne risulta generalmente una classificazione in cui l'aggiunta o la soppressione di una variabile può modificare le classi, impedendo ogni validità del risultato e ogni interpretazione delle classi » (DJINDJIAN 1991, 76).

Questo vale per il rapporto immediato fra la codifica e la sua utilizzazione, non per il problema in sé della correttezza della codifica.

Da parte sua Gardin afferma: « Ciò che distingue una 'compilation' scientifica dalla prosa di un collezionista è che essa fa uso di un linguaggio rappresentativo che si suppone superiore, sia teoreticamente per l'accumulazione della conoscenza scientifica, sia praticamente, per la gestione dell'informazione. (...) Una pura collezione di oggetti o riproduzioni (foto, disegni, etc.) non costituisce una 'compilation' scientifica »⁴.

⁴ GARDIN 1980, 38. In modo più "informatico" diremmo che si tratta dell'opposizione fra continuo e discreto e del relativo passaggio dall'uno all'altro, che permette la comunicazione

5. LA PERDITA D'INFORMAZIONE

Non c'è dubbio, ad ogni modo, che nel passaggio dal continuo della realtà al discreto della codifica, intendendo per tale la prima presa di contatto dello studioso con la realtà che diventa oggetto del suo studio, si verifica necessariamente una perdita di informazione nei riguardi del messaggio originale. Questa perdita di informazione dovrà da un lato essere contenuta nel minimo indispensabile; ma dall'altro essere attuata tenendo conto dei fini per cui si agisce.

C'è innanzi tutto un fine immediato, cioè un'analisi o una serie di analisi che si sono già progettate nel momento in cui si codifica il materiale studiato⁵. C'è d'altra parte la necessità che sia possibile sfruttare il materiale codificato per altre analisi da parte anche di altri studiosi. Per questo mi sembra che sia più corretto, in termini di generale metodo informatico, tendere fin dall'inizio ad una completezza intrinseca della codifica, considerando gli oggetti per se stessi, e non solo nel contesto di una specifica ricerca; nemmeno della differenza, che pure sembra così originaria, fra gli aspetti da codificare in vista dell'inserzione in una banca dati, e quelli in vista di un'analisi quantitativa.

6. VARI TIPI DI CODIFICA

Normalmente si considera banale la codifica, perché si pensa ad una operazione puramente macchinistica, cioè che presuppone già un'informazione codificata (generalmente in linguaggio naturale, talora in linguaggio tecnico formalizzato), e che quindi consista semplicemente nella trasposizione in un codice memorizzabile. In un primo tempo per codifica si intendeva addirittura un procedimento di semplificazione del documento, in modo da estrarne solo alcuni elementi fondamentali, che potessero essere memorizzati sfruttando il minimo spazio di supporto magnetico possibile. Questa visione è stata superata da tempo a causa del progresso negli apparati di memorizzazione (IGARTUA 1991, 78).

Ma ancora usuale è purtroppo il punto di vista di chi considera la codifica come il puro passaggio da una situazione descrittiva adatta al veicolo cartaceo (in sostanza il linguaggio naturale) al suo corrispondente in codice binario, adat-

scientifica, e in particolare l'analisi informatica. Tuttavia esempi come quello del progetto AR-COS (KRINZINGER, SCHICK, TEEGEN 1990) indicano come sia difficile esplicitare il confine fra i due sistemi.

⁵ Mentre per Djindjian (1991, 76 e 99) la codifica è sempre in vista di una analisi quantitativa, nella visione di Guimier-Sorbets può avere un valore documentario in se stessa, preliminarmente ai procedimenti di analisi. Cfr. GUIMIER-SORBETS 1986; inoltre cfr. nota 9. L'accento è posto tuttavia sull'information retrieval con linguaggio naturale. Più difficile è stabilire quanto questo approccio libero alla codifica (che in verità è quello che preferiremmo, a certe condizioni) possa poi dar luogo ad analisi automatiche.

to al supporto magnetico. Si dimentica che in tal caso i procedimenti automatici riguarderanno non il materiale documentario in sé, ma la sua descrizione; e non si sottoporrà tale descrizione alla necessaria critica analitica, che consenta di evitare errori e confusioni nella valutazione finale dei risultati del procedimento automatico. Scrive a questo proposito José E. Igartua: « A questo livello la codifica significa definire il significato delle parole e numeri che si inseriscono nel computer, in modo che la manipolazione dei simboli operata dalla macchina corrisponda ad una manipolazione dell'informazione che sia utile allo storico. Incoerenze e ambiguità che rimangano dopo che i dati sono stati resi "machine-readable" inficeranno l'analisi successiva e sarà difficile da rimediare » (IGARTUA 1991, 75).

Inoltre non si tiene conto del fatto che non c'è un rapporto di necessità e dipendenza fra quello che possiamo memorizzare su supporto magnetico, per poi sottoporre a procedimento automatico, e quello che è rappresentato sulla tastiera per mezzo della quale vengono compiute le operazioni di input (lettere, numeri, parentesi, etc.). Il valore attribuito alle sequenze di bit che vengono memorizzate può anche non coincidere con quello previsto dalle convenzioni normalmente utilizzate (per esempio il codice ASCII), anche se è più comodo accedere ai modelli di codice previsti dai costruttori delle macchine. Tali modelli sono lontani dall'essere completi, e comunque non sono aderenti alle esigenze della ricerca umanistica, in particolare archeologica. In realtà, salvo il fatto che i dati devono essere discreti, la loro scelta e la scelta di come rappresentarli dipende esclusivamente dallo studioso.

C'è inoltre una possibile confusione fra il momento della scelta di un codice e quello della individuazione della corrispondenza biunivoca fra i simboli ottenuti con il codice scelto e gli elementi della realtà che vengono rappresentati da quei simboli. Ci sono da un lato vari passaggi, per mezzo dei quali l'alfabeto dà origine a unità più complesse; dall'altro decisioni soggettive, mediante le quali si attua la corrispondenza fra gli elementi da codificare e le unità complesse⁶.

Tutto questo va visto in rapporto ai differenti *piani di rappresentazione* della realtà. Infatti i procedimenti informatici possono essere applicati alla realtà, come si presenta direttamente allo studioso; ovvero a rappresentazioni di tale realtà. Sarà perciò possibile distinguere:

1. *Rappresentazione diretta*: è quella che avviene mediante immagini assunte co-

⁶ ORLANDI 1990. Su questa problematica si fonda il lavoro del prestigioso gruppo internazionale della Text Encoding Initiative, volto a formalizzare un "linguaggio di descrizione" dei testi che prevede preliminarmente la scelta accurata degli elementi da descrivere, e quindi codificare.

me tali (fotografia, riprese televisive, disegni, etc.) ovvero mediante banche dati formate per mezzo di inchieste dirette, cioè con questionari a cui rispondono i diretti interessati. La rappresentazione per immagini è naturalmente usata pesantemente in archeologia, e sarà opportuno citare il progetto ARCOS, perché frutto anche di una riflessione approfondita sui risvolti in fase di ricerca (KRINZINGER, SCHICK, TEEGEN 1990).

2. *Rappresentazione indiretta*: in questo caso il punto di partenza è rappresentato da documenti scritti (epigrafici, archivistici, letterari), i quali vengono codificati. La corrispondenza è dunque col documento, non con la realtà rappresentata dal documento. Questo significa che nel valutare i risultati dei procedimenti informatici occorrerà tener conto sia dell'interpretazione di chi ha redatto i documenti, sia di chi li codifica.

3. *Rappresentazione doppiamente indiretta*: quando il documento su cui si lavora è un'opera storiografica, da cui si ricavano notizie da codificare, si opera una prima sintesi o scelta degli argomenti, oltre a codificare e interpretare il documento.

Per l'*archeologia* è opportuno introdurre una distinzione ulteriore, che riguarda la volontarietà e la consapevolezza del messaggio. Posto che con la codifica si perde sicuramente una parte dell'informazione contenuta in un messaggio (o comunque nella "cosa" che prendiamo in considerazione come trasmettitrice di un messaggio) mi sembra che uno dei criteri, forse il principale, per cercare di ridurre al minimo quella perdita, sia quello dell'aderenza alla volontà della sorgente del messaggio.

A questo punto, l'*archeologia* si trova di fronte a due possibilità che devono essere distinte. Da un lato, è per esempio possibile immaginare che chi ha costruito un edificio sacro abbia avuto effettivamente intenzione di trasmettere messaggi che riguardassero la sua concezione della religione o anche una sua concezione estetica; d'altro lato chi ha fabbricato un certo tipo di ceramica non ha sicuramente avuto intenzione di farci conoscere il tipo di società che essa presuppone⁷.

7. LA CODIFICA ALL'INTERNO DEL PROCESSO ARCHEOLOGICO

La codifica è, per un verso, il momento iniziale di un procedimento di ricerca, in particolare quello attuato mediante metodi anche informatici. Ma, sotto

⁷ TRIGGER 1989, 19-21. La discussione a p. 297-99 sulle teorie di Binford, sul significato della relazione fra tecnologia, organizzazione sociale, ideologia si può considerare basata sul valore documentario dato a questo proposito ad oggetti che non sono nati propriamente per questo.

un altro aspetto, rappresenta anche (e forse principalmente) il momento conclusivo di un procedimento precedente, che ha condotto ad individuare: i confini del materiale documentario che si vuole prendere in considerazione; le caratteristiche di quel materiale che determinano il suo interesse per la ricerca; la scelta di un linguaggio per esprimere i due punti precedenti; il modo di esprimere in tale linguaggio l'identificazione del materiale documentario e le sue caratteristiche.

Risulta da questo che una codifica non sarà mai definitiva, perché i risultati del procedimento di ricerca che ha dato origine alla codifica è prevedibile che siano tali da far mutare almeno in parte la codifica, o almeno a rendere necessaria la codifica di ulteriori documenti o caratteristiche⁸.

Dunque la struttura della codifica deve essere aperta, in modo che si possa intervenire apportando aggiunte o modifiche. In questo senso la codifica comprende anche l'aspetto di struttura dei dati, non solo quello della loro individuazione in quanto oggetti singoli; e sembra opportuno vederla come una struttura relazionale. La struttura relazionale dei dati è da un lato un modo più funzionale di gestire una banca dati; ma per quanto ci riguarda qui, essa, vista da chi la disegna nella fattispecie singola, è un altro dei modi di intendere la codifica.

Il sistema relazionale è il migliore, perché la realtà è più vicina ad una struttura relazionale che non ad una struttura gerarchica. La struttura dunque di una banca dati relazionale riprodurrà in modo più fedele ciascuno degli oggetti della ricerca, anche perché ne manterrà i rapporti con gli altri oggetti su un piano parallelo, e non verticale, così come nella realtà vi sono piuttosto relazioni multiple sullo stesso piano che relazioni di dipendenza di un oggetto da un altro.

D'altra parte c'è sempre un momento di crisi nella concezione di una struttura relazionale, ed è quello della scelta di ciò che si considera soggetto, e di ciò che si considera attributo. A meno che non si facciano tabelle di due soli campi (identificativo e soggetto), quelli che per certi fini di studio sono considerati attributi per altri fini andranno considerati soggetti.

Intendendo tuttavia il sistema relazionale come una costruzione teorica, è lecito chiedersi se le realizzazioni informatiche che ne vengono date rappresentino una situazione ottimale nella prassi. Si può allora dire che è indubbiamente vero che il sistema relazionale rappresenta un progresso radicale rispetto agli altri sistemi precedentemente immaginati; ma che in ambito umanistico, e in particolare archeologico e storico, si fa strada la convinzione che esso possa rimanere sullo sfondo come organizzazione mentale dello studioso, da attuare al momento in cui serve, ed essere sostituito a livello di codifica da una procedura

⁸ La codifica si inserisce così nel processo che torna su se stesso ben delineato da Gardin (1980, cap. 5), e poi da Djindjian (1991, 334).

ancora più libera e aderente alla realtà come si presenta allo studioso⁹.

Tale procedura si articola in due momenti fondamentali:

- (1) Codifica rappresentativa al massimo del documento. La codifica non è basata su ciò che sembri utile estrarre dal documento (questo si farà poi tramite i Tag), ma solo sulle caratteristiche intrinseche del documento, cioè sul messaggio "autentico" che chi ha prodotto il documento voleva trasmettere. Essa è attuata mediante file del tutto liberi, nei quali la struttura è determinata "a posteriori" da annotazioni (Tag) inserite diciamo così fuori testo, come un meta-linguaggio descrittivo.
- (2) Trattamento del documento in modo da estrarre come "secondo passaggio" le notizie utili per lo studioso, nella particolare indagine che sta conducendo.

La codifica riveste dunque un'importanza assai maggiore di quella che di solito le si concede. C'è addirittura la possibilità (che riteniamo da evitare) che essa diventi totalizzante, cioè che riunisca in sé anche gli altri passaggi del trattamento automatico (l'analisi dei rapporti logici fra i dati, e la verifica della correttezza teorica delle procedure di analisi e di sintesi).

In effetti occorre sempre considerare con attenzione l'elemento di *petitio principii* insito nella codifica, cioè quel fenomeno per cui un certo materiale, codificato in un certo modo, darà necessariamente un certo tipo di risultati una volta sottoposto all'analisi automatica. Spesso questo fenomeno è sottovalutato, perché fra la codifica del materiale e i risultati che si ottengono non c'è più alcun intervento umano, che generalmente comporta un ripensamento successivo dei problemi. Questa è del resto una tipica caratteristica dei procedimenti automatici¹⁰.

8. LA CODIFICA COME SINTESI DEL CONOSCIUTO

In conclusione, la codifica rappresenta la sintesi formale di quanto conosciamo dell'oggetto che vogliamo studiare. Essa deriva in prima istanza, cioè al momento della prima conoscenza che si ha dell'oggetto, soprattutto dalla dialettica fra le caratteristiche intrinseche dell'oggetto (forma, dimensioni, materiale...) e

⁹ Per questo rimando alla pubblicazione (in corso) di un seminario tenuto presso l'Accademia dei Lincei su « Strumenti Informatici nelle Discipline Umanistiche: il Problema dell'Integrazione » (5 ottobre 1991). Si vedranno soprattutto i contributi di Manfred Thaller e Anne-Marie Guimier-Sorbets.

¹⁰ È interessante notare come Djindjian (1991, cap. 15) rifugga dal parlare di informatica e automatismi quando delinea il suo sistema formalizzato. Probabilmente egli non intende automatizzarlo, ma dal nostro punto di vista formalizzazione e automazione (logicismo) in sostanza coincidono.

la competenza archeologica dello studioso. Il confronto fra questi due elementi viene attuato con procedimenti sostanzialmente intuitivi, che determinano quali aspetti dell'oggetto prendere in considerazione e quindi codificare.

Ma in un secondo tempo, quando sono stati effettuati degli studi, eventualmente anche con procedimenti automatici, su quell'oggetto, le conclusioni vengono a far parte esse stesse delle nuove competenze archeologiche, e quindi consentono di allargare e complicare il modo con cui viene considerato l'oggetto, e dunque la sua codifica. L'aspetto soggettivo insito in questa situazione non dovrebbe spaventare, perché (se le procedure sono attuate in maniera scientifica) non è più soverchiante di quello che esiste già all'origine. Inoltre, e soprattutto, se i procedimenti di studio sono stati automatizzati, la memoria di tali procedimenti consentirà sempre di avere un controllo su quanto è stato fatto. D'altra parte la convenienza di approfondire i lati teorici del procedimento di codifica deriva dal fatto che esso, nell'ambito dell'automazione, è quello in cui maggiormente entrano elementi soggettivi, mentre gli altri procedimenti dovrebbero obbedire a regole logiche obiettive.

Da ciò consegue una delle caratteristiche più interessanti delle metodologie informatiche, che purtroppo è stata tenuta in poco conto fino a questo momento: la necessità di una completa trasparenza nella diffusione non solo dei risultati delle ricerche, ma anche degli algoritmi utilizzati, dei presupposti metodologici, e dell'esecuzione della codifica.

Un'ultima considerazione deve poi essere dedicata al fatto che non è prevedibile che l'archeologia detta quantitativa sia destinata a sostituire quella che possiamo chiamare "concettuale", e quindi occorrerà mantenere fra i due metodi dei rapporti costruttivi. Tali rapporti potranno soprattutto essere basati sul momento della codifica.

TITO ORLANDI

Centro Interdipartimentale di Servizi per
l'Automazione nelle Discipline Umanistiche
Università di Roma « La Sapienza »

BIBLIOGRAFIA

- ADAMO G. 1987, *La codifica come rappresentazione*, in G. GIGLIOZZI (ed.), *Studi di codifica e trattamento automatico di testi*, Roma, Bulzoni, 39-84.
- DJINDJIAN F. 1991, *Méthodes pour l'archéologie*, Paris, Colin.
- GARDIN J.-Cl. 1980, *Archaeological Constructs*, Cambridge - Paris, Cambridge University Press - Editions de la Maison des Sciences de l'Homme.
- GUIMIER-SORBETS A.-M. 1986, *The Research Centre for Automatic Treatments in Classical Archaeology*, « Computer and the Humanities », 20, 319-321.

- IGARTUA J.E. 1991, *The Computer and the Historian's Work*, « History and Computing », 3.
- KRINZINGER F., SCHICK M., TEEGEN W.R. 1990, *Un calcolatore disegna e registra ceramica antica: il sistema ARCO5-1 negli scavi di Velia, comune di Ascea (Salerno)*, « Archeologia e Calcolatori », 1, 179-210.
- ORLANDI T. 1990, *Informatica Umanistica*, Roma, La Nuova Italia Scientifica.
- ORLANDI T. 1992, *Informatica umanistica, realizzazioni e prospettive*, in AA. VV., *Calcolatori e Scienze Umane*, Milano, Etas Libri, 1-22.
- SPAULDING A.C. 1971, *Some Elements of Quantitative Archaeology*, in F.R. HODSON *et al.* (edd.), *Mathematics in the Archaeological and Historical Sciences*, Edinburgh.
- TRIGGER B.G. 1989, *A History of Archaeological Thought*, Cambridge, Cambridge University Press.

ABSTRACT

The encoding problems are often neglected, in archaeological as in other humanities related research, because of their apparent triviality. Encoding is assumed to be the reproduction in an "electronic alphabet" (be it the ASCII code or a second level language like SGML) of something written on paper. On the contrary, the encoding process begins with the recognition, choice and declaration of the elements of reality which we are going to submit to an electronic process. As a consequence, we must examine very carefully the substance of the logical and formal passages that we undertake. This article tries to show the difference between many kinds of encoding and the significance of encoding in the archaeological disciplines, in comparison with other opinions, mainly by J.-Cl. Gardin and F. Djindjian.