

FORMALIZZAZIONE DEI DATI, SEMIOTICA E COMUNICAZIONE

1. INTRODUZIONE

È necessario premettere che la prospettiva nella quale si inserisce questo saggio è quella di un ipotetico libro di informatica umanistica; esso quindi cerca di indagare i problemi relativi alla formalizzazione dei dati archeologici tenendo presenti le convergenze di metodo fra le applicazioni archeologiche e quelle delle altre discipline umanistiche.

Quei problemi si trovano invece generalmente inquadrati all'interno della metodologia archeologica, onde l'interesse per l'efficacia relativa al successivo corso delle analisi archeologiche prevale sull'interesse per la correttezza teorica dal punto di vista informatico delle soluzioni proposte. Ovviamente quest'ultima prospettiva è del tutto legittima, ma forse tende a mantenere certe soluzioni tradizionali, a spese di un rinnovamento che l'informatica può portare o almeno suggerire.

Il problema della formalizzazione dei dati (che per concisione chiameremo della "codifica") si riduce spesso, anche per l'archeologo più attento, ad un punto di partenza intuitivo, da elaborare al meglio sulla base delle generiche conoscenze del singolo studioso, piuttosto che costituire un procedimento complesso e delicato, da studiarsi in maniera interdisciplinare, cioè tenendo conto dei risultati e dei relativi impliciti suggerimenti di una serie di discipline non archeologiche, ma non per questo trascurabili in ambiente archeologico. Sembra questo uno dei motivi per i quali il problema della codifica in archeologia è stato a lungo trascurato, come notavo in un precedente contributo a cui questo si riallaccia¹ e del quale si terrà conto per evitare ripetizioni. Osserverò che intorno alla codifica dei testi si discute molto, e molto si è operato con le proposte ormai diventate standard, quali quella nota con la sigla SGML, di cui la Text Encoding Initiative rappresenta una applicazione considerata una pietra miliare nel campo della codifica a livello sistematico e dunque scientifico dei testi².

¹ T. ORLANDI, *Sulla codifica delle fonti archeologiche*, «Archeologia e Calcolatori», 4 (1993), 27-38. Cf. D.T. BARNARD, C.A. FRASER, G.M. LOGAN, *Generalized Markup for Literary Texts*, «Literary and Linguistic Computing. Journal of the Association for Literary and Linguistic Computing», 3 (1988) 1, 26-31. D.T. BARNARD, R. HAYTER, M. KARABABA, G.M. LOGAN, J. MCFADDEN, *SGML-Based Markup for Literary Texts: Two Problems and Some Solutions*, «Computers and the Humanities», 22 (1988) 4, 265-276.

² M. BRYAN, *SGML. An Author's Guide to the Standard Generalized Markup Language*, Wokingham (UK), 1988, xvii-364 p.: Addison-Wesley. P.M.W. ROBINSON, *The Transcription of Primary Textual Sources using SGML*, (Office for Humanities Communication Publications), Oxford, 1993: Office for Humanities Communication. C.F. GOLDFARB, Y. RUBINSKY, *The SGML Handbook*, Oxford, 1992 (ristampa), xxiv-663 p.: Clarendon Press. E. VAN HERWIJNEN, *Practical SGML*, Dordrecht-Boston-London, 1990,

Per l'archeologia non conosciamo nulla di simile, se si eccettuano le proposte di *thesauri* di nomenclatura che tuttavia soffrono della debolezza cui accennavo sopra.

È interessante notare peraltro che sul reciproco versante, quello della parte propriamente informatica, dell'informatica teorica e non applicata, il problema della codifica è altrettanto poco discusso, ovvero è discusso solo per quanto riguarda i "character set", cioè il rapporto utente-macchina, e non per quanto riguarda la rappresentazione dell'informazione, cioè la formalizzazione dei dati. Questo è sintomatico, credo, di un imbarazzo ad affrontare i confini fra l'informatica come scienza e il mondo "reale" a cui essa si lega, perché questo in sostanza è il compito della codifica. Occorre per lo meno una distinzione chiara dei vari ambiti e livelli in cui si attua la codifica.

Tornando al terreno dell'archeologia, la grande eccezione è rappresentata da Gardin¹, che fin dall'inizio della sua opera, negli anni '60, intuì che la codifica era uno dei momenti essenziali delle procedure informatiche; ma le sue sintesi anche più recenti non sono state discusse, almeno con l'ampiezza che avrebbero meritato. Vorrei registrare qui le osservazioni di F. Djindjian⁴ che accenna a più riprese alla codifica (nella sua terminologia *non* «codage», ma «description» [de vestiges...]), ma evita di approfondire i problemi, passando subito all'utilizzazione delle descrizioni senza soffermarsi sulla loro validità (validazione). Il passaggio più significativo è a p. 330, nell'ambito della formalizzazione dei ragionamenti, in cui si tratta dei tentativi di utilizzare paradigmi strutturali e semiotici. Ma anche qui si trovano solo poche righe.

Il sostanziale silenzio che ha accompagnato le teorizzazioni e le proposte di Gardin è a mio parere un cattivo segnale della coscienza metodologica degli archeologi. Si noterà in effetti che in una seconda fase Gardin si è rivolto ai testi, sia pure per problemi metodologici parzialmente diversi, trovando in questo campo una maggiore attenzione.

Il mio tentativo è quello di procedere oltre, sulla via indicata da Gardin,

xviii-307 p.: Kluwer Academic Publishers. J.M. SMITH, *The Standard Generalized Markup Language (SGML) for Humanities Publishing*, «Literary and Linguistic Computing. Journal of the Association for Literary and Linguistic Computing», 2 (1987) 3, 171-175. R.C. COVER, N. DUNCAN, D.T. BARNARD, *The Progress of SGML (Standard Generalized Markup Language): Extracts from a Comprehensive Bibliography*, «Literary and Linguistic Computing. Journal of the Association for Literary and Linguistic Computing», 6 (1991) 3, 197-209. C.M. SPERBERG-MCQUEEN, *Text in the Electronic Age: Textual Study and Text Encoding, with Examples from Medieval Texts*, «Literary and Linguistic Computing. Journal of the Association for Literary and Linguistic Computing», 6 (1991) 1, 34-46.

¹ J.-C. GARDIN, *Archaeological Constructs, An Aspect of Theoretical Archaeology*, Cambridge-Paris, Cambridge University Press-Éditions de la Maison des Sciences de l'Homme, 1980. Id., *Les analyses de discours*, (Collection Zethos), Neuchâtel-Paris, Delachaux et Niestlé, 1974 - Ed. it.: *Le analisi dei discorsi*, trad. di Marina Pisaturo, Napoli, Liguori Editore. Id., *Une archéologie théorique*, (Esprit critique), Paris, Hachette, 1980. Id., *Le calcul et la raison. Essais de formalisation du discours savant*, (Recherches d'histoire et de sciences sociales / Studies in History and the Social Sciences, 46), Paris, Éditions de l'École des Hautes Études en Sciences Sociales.

⁴ *Méthodes pour l'archéologie*, Paris 1991, Armand Colin.

e a tal fine vorrei puntualizzare gli argomenti sostanziali che egli prende in considerazione nel suo libro *Archaeological Constructs*⁵, perché essi formano la base di partenza per le mie osservazioni, mentre i successivi contributi sul ragionamento archeologico escono dal campo che intendo trattare.

Si ricorderà che Gardin denomina *compilations* tutto il lavoro di riunione e classificazione delle fonti archeologiche precedente al lavoro di giudizio storico del loro significato. Egli sostiene a ragione (p. 31) che l'apparente diversità delle compilations si può ricondurre ad un fine comune e dunque ad una unità di sostanza. La diversità delle compilations deriva dalla varietà della forma con cui sono presentate le entità materiali (foto, disegni, descrizioni), ma anche dal contenuto testuale delle presentazioni. Vi sono vari modi di esprimersi all'interno delle forme di presentazione scelte. Inoltre alcune diversità derivano dal materiale stesso: oggetti greci *vs* oggetti maya, etc.

Ne consegue (p. 32) che il linguaggio è comunque sempre necessario, per accompagnare gli oggetti rappresentati graficamente con informazioni indispensabili alla loro migliore comprensione. Se da un lato (diremmo noi) la codifica presuppone sempre apprezzamenti soggettivi, la scienza, afferma Gardin ampliando il campo della sua trattazione (p. 33), presuppone un linguaggio. Le compilations sono dunque costruzioni simboliche (p. 34), che servono come una retrieval machine, al centro fra i materiali e le domande poste dagli studiosi circa i materiali (p. 35).

I problemi teoretici che le caratterizzano sono:

a. criteri di selezione del materiale

1. origine o collocazione del materiale
2. tipologia del materiale

(Il primo criterio è puro espediente per identificare il materiale di cui si tratta. Il secondo viene a coincidere con le varie discipline archeologiche).

b. scelta del linguaggio rappresentativo

(Esso non può essere sostituito dalla pattern recognition automatica. Gardin si sofferma molto sui tipi di linguaggio: naturale, specialistico, scientifico, che in sostanza rappresentano gradi della formalizzazione.)

c. valutazione della funzionalità della compilation (p. 60).

In sintesi, l'intuizione metodica fondamentale di Gardin consiste nell'evidenziare il fatto che la codifica si basa su un linguaggio derivato a sua volta dallo studio di materiale codificato. Questa osservazione è passibile di interessanti sviluppi, che sono utilmente introdotti da un altro contributo teorico anch'esso purtroppo trascurato da chi studia la metodologia delle applicazioni informatiche in archeologia.

⁵ Citato sopra. Purtroppo non ho potuto vedere PEEBLES-GARDIN, *Representations in Archaeology*, 1992.

Si tratta del lavoro che Gordon Childe condusse verso la fine della sua lunga carriera di studioso a proposito di ciò che potremmo chiamare la «conoscenza archeologica», cioè circa la possibilità di enucleare una particolare epistemologia che aiutasse a chiarire le caratteristiche del lavoro archeologico, e della sua individualità rispetto al lavoro storico inteso in senso generale. Il fatto che molte delle vedute, spesso differenti, esposte da Childe nel corso dei suoi studi sono ormai oggi considerate superate, comunque non corrette rispetto ad una attenta valutazione dei documenti, anche perché spesso viziata da pregiudizi tipici della sua epoca, non dovrebbe tuttavia far passar sopra al suo estremo lavoro epistemologico, perché esso è il frutto di una conoscenza molto vasta e di una riflessione non partigiana sul lavoro archeologico. Childe comprese in quel tardo momento il valore della riflessione epistemologica, proprio perché si avvide dell'essenzialità dell'impatto della metodologia sulla qualità delle interpretazioni.

Si ricorderà come gli anni '50 abbiano portato ad una più matura considerazione dell'archeologia come di una disciplina che cerca di trarre dai documenti su cui si basa delle informazioni di tipo storico, ivi compresi i settori antropologico, sociologico, economico, etc. Come nota giustamente Trigger⁷ le proposte di Binford in questo senso erano per alcuni versi forzatamente polemiche, e oscuravano una linea di continuità con l'evoluzione metodologica già presente nella migliore archeologia dagli anni '30 in avanti. Inoltre l'accento messo sull'antropologia derivava dal tipo di archeologia (preo proto-storica) tipica dell'ambiente americano; in altri ambienti questo valeva (come abbiamo detto) per la politica, la sociologia, l'economia, etc. Tuttavia la maggiore consapevolezza storica raggiunta aveva come conseguenza la necessità di formulare una più coerente epistemologia archeologica. Le proposte di Childe, meno settoriali di altre (che giungevano immediatamente alla proposta di una visione evolutiva dell'oggetto stesso della ricerca, le culture che avevano prodotto i reperti archeologici), mi sembrano di notevole interesse, se spogliate di alcune ingenuità filosofiche peraltro naturali.

Childe afferma (p. 15): «Per meritare il suo nome ... la conoscenza deve essere comunicabile e in tal senso collettiva»; e poco oltre (p. 20): «Strettamente parlando, il termine conoscenza dovrebbe essere riservato all'informazione comunicabile in qualche modo.» Questo inserisce il problema della conoscenza, e dunque della codifica, del dato archeologico nell'ambito della comunicazione, e perciò stesso della semiotica.

In seguito Childe inserisce in questo ambiente il comportamento delle culture dell'antichità (p. 22): «Biologicamente tutti i meccanismi che controllano e dirigono il comportamento di ogni organismo ... sono sopravvissuti

⁶ V.G. CHILDE, *Society and Knowledge*, 1956, trad. ital. *Società e conoscenza*, [Milano] Mondadori, il Saggiatore, 1962.

⁷ TRIGGER, *A History of the Archaeological Thought*, Cambridge, 1989, 240.

perché hanno posto i loro possessori in condizione di sopravvivere e di moltiplicarsi. La conoscenza comunicabile è l'ultimo nel tempo e il più valido di questi meccanismi», affermando poi che la trasmissione della conoscenza avviene attraverso gli strumenti che le stesse popolazioni lasciano come testimonianza (dato archeologico) della loro cultura.

Finalmente, a proposito del rapporto fra quella che per brevità chiamo codifica dei dati archeologici, ed il loro oggetto, Childe afferma (p. 110): «Se il mondo esterno segue uno schema, la riproduzione ideale che noi chiamiamo conoscenza deve necessariamente seguire a sua volta uno schema, comunque ridotto o semplificato esso sia. ... Un mondo delle idee deve per forza avere una base simbolica, e la conoscenza, essendo comunicabile, deve essere esprimibile. Uno schema ideale deve quindi essere uno schema di simboli.»

È interessante come questo modo di ragionare (ammetto che la mia interpretazione del pensiero di Childe è in qualche modo forzata, diciamo così, a fin di bene) ci conduca direttamente sul terreno dell'informatica, la quale alla sua radice consente di produrre (e di gestire) strutture di simboli. Le strutture sono generate dai programmi, ma i simboli? Qui si torna al mondo teorico di Gardin che abbiamo sopra delineato.

2. NECESSITÀ DELLA CODIFICA. CRITERI DI CORRETTEZZA

Gardin, come abbiamo visto, dà il giusto valore ai procedimenti di codifica, ma prende in considerazione soltanto l'effetto di essa nei riguardi di una eventuale interrogazione di una banca dati o di una analisi dei dati, mentre la valutazione deve dipendere dalla correttezza in sé della codifica. Gardin parte giustamente dai linguaggi di descrizione, ma non prende in considerazione la possibilità di inserire gli oggetti stessi, studiati dall'archeologia, come parte di un processo integrale di analisi archeologica.

Diremo allora che un'altra ragione dell'evasività nei riguardi della codifica, che accomuna gli studiosi di archeologia e quelli di informatica, oltre a quanto abbiamo osservato sopra, consiste nella complessità dell'intrico fra segno, significato, designato, etc., che sembra caratterizzare soprattutto l'archeologia, mentre in realtà è tipico anche dell'analisi dei testi, ma normalmente non viene preso in considerazione. È abbastanza apparente in archeologia che gli elementi presi in considerazione per sviluppare l'osservazione archeologica sono di per sé semplici (= oggetti materiali del tutto normali), ma quello che conta nel prenderli in considerazione è un valore che viene immediatamente aggiunto ad essi (tipologia, struttura, forma, localizzazione, etc.) e dunque solo l'unione di questi due elementi forma la base del ragionamento scientifico, e tale unione è facilmente espressa, anche per le necessità computazionali, da un simbolo (p.es. una parola chiave o un numero). Le cose non stanno così.

La questione che ci sta a cuore consiste soprattutto nella possibilità di teorizzare la correttezza della codifica e da questo punto di vista desideriamo sottolineare l'importanza della teoria, sia a livello archeologico (metodologia), sia a livello informatico. Per quanto riguarda quest'ultimo, occorre distinguere e identificare bene i diversi usi (applicazioni) dell'informatica (del computer).

La necessità di una rappresentazione non puramente analogica o immediata dell'oggetto (fotografia, disegno, etc.; cf. le acute osservazioni in proposito di Gardin) è comune ai procedimenti informatici ma anche non informatici in archeologia, e questo ha portato a qualche confusione. Infatti, se è vero che i principi di rappresentazione-codifica possono essere gli stessi (Gardin non tratta propriamente la codifica informatica), nelle applicazioni informatiche essi hanno una necessità tecnica immediata e una necessità di formalizzazione spinta, che possono condurre a semplificazioni e anche a fraintendimenti sul ruolo che la codifica ha all'interno dell'informatica. Occorre chiarire la funzione del simbolo nella computazione, per poter stabilire quali valori possono essere associati con quel simbolo.

Esporrò su questo punto, brevemente, il mio punto di vista, che è stato svolto ampiamente in altri contributi⁸. Per le sue particolari caratteristiche, lo strumento informatico può essere utilizzato come il puro simulatore di uno strumento non informatico (macchina da scrivere, macchina tipografica, calcolatore, macchina fotografica, televisione, etc.). È soltanto quando si usa lo strumento informatico in quanto tale, cioè in quanto motore di automazione in procedure che normalmente vengono definite "intelligenti", che interviene la necessità di rendersi ben conto di quale sia la sua natura, e quali i presupposti di una buona utilizzazione.

Quello che è nuovo nell'uso di strumenti informatici (e per questo appunto vengono chiamati informatici, per distinguerli da tutti gli altri strumenti, che vengono implicitamente compresi in una medesima categoria, diciamo, di "convenzionali") consiste nel fatto che i passaggi in cui si evita l'intervento umano non sono passaggi di carattere puramente "pratico" e insomma materiale, ma passaggi che si esplicavano finora per lo più mediante operazioni mentali. Questo presuppone che si accetti che l'informatica stessa non sia semplicemente un insieme di tecnologie, ma piuttosto di metodologie, e dunque una disciplina con un proprio fondamento teorico, che deve in qualche modo interagire con le metodologie delle singole discipline umanistiche. L'uso corretto del computer come macchina che realizza procedimenti logici sui dati, e nello stesso tempo funge da mezzo di comunicazio-

⁸ *Informatica Umanistica*, (Studi Superiori NIS, 78), Roma, La Nuova Italia Scientifica, 1990; *Alla base dell'analisi dei testi: il problema della codifica*, in M. RICCIARDI (ed.), *Scrivere comunicare apprendere con le nuove tecnologie*, Torino, 1995, 69-86; *Informatica umanistica: realizzazioni e prospettive*, in AA.VV., *Calcolatori e Scienze Umane*, Milano 1992, 1-22.

ne non passivo fra gli studiosi, richiede la consapevolezza dell'aspetto teorico che sta alla base dei vari tipi di macchina fisica che lo studioso utilizza.

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 dovrà fornire la chiave per intendere correttamente i rapporti fra il computer e le discipline umanistiche alle quali viene applicato.

In realtà, come nota Bolter⁹, il computer, nemmeno quando le simula, è alcuna delle macchine cui accennavo sopra, ma è lo strumento che "governa" quelle macchine. Per questo motivo l'informatica correttamente intesa è la disciplina che studia i principi di funzionamento del computer, ovverossia come modello formale (se si vuole matematico, ma in senso logico) del funzionamento del computer (macchina di Turing, algoritmi); quindi come metodologia del trattamento dei dati sottoposti al computer.

Il computer può essere utilizzato anche con metodi diversi da quello propriamente informatico: questo accade quando ci si avvale della sua capacità di simulare una macchina "non informatica". Quando p.es. si applicano metodi statistici, si dovranno utilizzare le regole della statistica, che poco hanno a che fare con l'informatica intesa secondo quanto ho detto sopra. Invece i procedimenti "logicistici" (cari a Gardin) sono in buona parte equivalenti ai procedimenti informatici in senso proprio (cf. *modelling, simulation, etc.*). Per questo motivo il fatto che un certo tipo di codifica sia sufficiente per mettere in moto procedimenti di analisi statistica non è per nulla equivalente ad affermare che quel tipo di codifica sia teoricamente corretto dal punto di vista dell'informatica.

La codifica è un procedimento essenziale, perché rappresenta lo stadio preliminare essenziale per poter sottoporre i dati a procedimenti informatici. Occorre riconoscere la centralità della codifica in quanto passo preliminare essenziale che determina la possibilità di analizzare dati ed ottenere risultati soddisfacenti. I dati sottoposti a procedimenti informatici devono essere finiti proprio perché la macchina di Turing non può funzionare con dati infiniti. Se le celle possono contenere uno fra infiniti dati il computer si ferma immediatamente¹⁰.

C'è tuttavia un problema della finitezza quantitativa, ma anche uno della finitezza qualitativa. Intendo con ciò la necessità che i singoli oggetti (essi stessi in numero finito) siano riuniti in gruppi caratterizzati da valori qualitativi, siano essi dimensioni, o forme, o strutture, o finalità, etc. Qui l'informatica si ferma, e subentra la metodologia propriamente archeologica.

⁹ *Turing's Man*, Chapel Hill, 1984. Tr. it. *L'uomo di Turing*, Parma, Pratiche, 1985, 15 e 47-52.

¹⁰ Su tutta la questione cf. M. DAVIS, *Mathematical Logic and the Origin of Modern Computers*, in R. HERKEN (ed.), *The Universal Turing Machine. A Half-Century Survey*, II ed., Wien-New York, Springer, 1994, 135-158, con cui concordo pienamente.

C'è uno stadio di sovrapposizione nei problemi di codifica tout-court per la classificazione archeologica anche senza computer, e per l'utilizzazione dell'informatica. Occorre riconoscere e distinguere un doppio passaggio: in una prima fase si utilizza un tipo di linguaggio, fra quelli distinti da Gardin (cf. sopra) in naturale, specialistico, e scientifico. Quindi si pone il problema di codificare quel linguaggio in forma binaria per le applicazioni informatiche.

In un certo senso, si potrebbe dire che il lavoro dell'archeologo sta essenzialmente nel determinare il simbolo, cioè individuare gli oggetti e dar loro una caratterizzazione; mentre i principi del ragionamento successivo, vuoi con sistemi statistici o con sistemi logicistici, appartengono ad altre sfere, e non sono propri dell'archeologia se non nella loro applicazione¹¹. La delimitazione dell'area di ricerca di una disciplina e l'individuazione dei dati emergenti dall'area di ricerca definita presuppone una competenza iniziale di carattere pre-logicista e specialistico. La competenza relativa alla determinazione e all'uso dei simboli in campo archeologico è appunto la sostanza della disciplina archeologia nei confronti dell'informatica¹².

Il problema diventa allora quello di determinare in maniera formalizzata i singoli fenomeni che saranno poi l'oggetto (il "significato") dei simboli che li rappresenteranno. Tali simboli possono appartenere in un primo tempo al "codice" del linguaggio archeologico, sia che si scelga un linguaggio naturale, sia un linguaggio specialistico o scientifico. A mio avviso questo deve essere fatto attribuendo funzioni comunicative (di determinate caratteristiche sociali, culturali, etc.) ad oggetti che originariamente avevano funzioni di altro genere (pratiche, religiose, etc.). Saranno queste funzioni a essere espresse facilmente in un tipo di linguaggio. Da questo punto di vista, la codifica varierebbe a seconda della funzione dell'oggetto che essa è chiamata ad esprimere.

Registrerò a questo punto l'affermazione esattamente contraria di Hinge¹³: «The purpose of archaeology is not to create lists, and refine typologies, but rather to take all those lists and ask what do they tell us about the past.» Ma mi sembra evidente che si tratta piuttosto di un punto di vista differente, tanto che subito dopo afferma: «An economic description of the archaeological investigation is that a site is carefully divided into its smallest referable components, = the contexts = which are then reconstructed into a meaningful narrative of past activity, known as the interpreted archive. Thus for a single site the fully interpreted archive is the most integrated account of the information it contained.» Io intendo aggiungere ad un simile quadro alcune determinazioni semiotiche; comunque segnalerò che in Hinge è im-

¹¹ Di nuovo ricordiamo lo spostamento di Gardin verso l'analisi del testo, e quella di Djindjian verso la statistica.

¹² Sul concetto di competenza cf. sotto.

¹³ *New Fusions: Archaeological Information in the Relational Database*, «Archeologia e Calcolatori», 5 (1994), 175-202, p. 176.

portante anche l'osservazione circa il significato del sistema relazionale come modello della realtà.

Considerazioni importanti a questo riguardo si trovano anche negli esponenti della cosiddetta "Cognitive Archaeology", p.es. in un recente saggio di Renfrew¹⁴. L'accento posto sulla mentalità che ha prodotto i reperti archeologici è certo opportuno; ma credo che la discussione sui rapporti fra i dati archeologici, vuoi prodotti umani, vuoi situazioni naturali (cf. in Renfrew, p. 4, il passo sulla divisione "idealistica" fra natura e storia), e la "ancient mind" dovrebbe essere riportata alla discussione circa le competenze, quella dei produttori e quella degli studiosi.

La riduzione a linguaggio e l'uso di un linguaggio porta al terreno della semiotica, e richiede l'inserimento del concetto di *competenza*, essenziale nella valutazione di un linguaggio. Secondo le mie vedute, la competenza (per quanto attiene all'informatica, dunque in ambito *formalizzato*), è l'insieme di regole che hanno guidato un autore a produrre un determinato oggetto (nel senso di oggetto semiotico); p.es. un "testo". Questo sostituisce vantaggiosamente il più banale richiamo al "mondo reale", spesso invocato dall'intelligenza artificiale, perché riconduce husserlianamente all'intenzionalità dell'autore, che comprende in sé la parte soggettiva che organizza e struttura la *sua* visione del mondo reale, con la quale possiamo mettere in relazione (biunivoca) un nostro sistema di codifica.

Sotto un altro aspetto, la competenza è ciò che determina la relazione fra i singoli oggetti; e per converso, ciò che permette di riconoscere la relazione fra gli oggetti studiati, quando essi sono stati posti in determinate relazioni da una competenza "originaria". Ricordando anche le intuizioni di Childe sui simboli e le strutture di simboli, e tenendo presente che l'informatica deve formalizzare, cioè ridurre a simboli, la competenza rappresenta la padronanza teorica della struttura di un codice, vuoi in sé, vuoi nei rapporti fra codice e realtà che esso può essere chiamato a rappresentare.

Si deve distinguere allora un livello della codifica in relazione alla competenza del codificatore da un altro livello, in relazione alla competenza del produttore dell'oggetto. Si può dire che la competenza dello studioso (del codificatore) deve essere una meta-struttura che consente di poter rappresentare con i simboli a disposizione le strutture che possono essere riconosciute negli oggetti che verranno rappresentati da questi simboli.

3. LA SEMIOTICA

Le osservazioni fatte sopra ci hanno condotto alla visione del processo di codifica come parte di processo semiotico; i dati e il loro complesso strut-

¹⁴ *Towards a Cognitive Archaeology*, in C. RENFREW, E. B. W. ZUBROW (eds.), *The Ancient Mind*, Cambridge U. P., 1994, 3-12.

turato come parti di un processo comunicativo. I dati diventano un linguaggio espressivo della competenza iniziale e conseguentemente dei dati da essa individuati. Occorre stabilire una differenza semiotica fra la codifica di un linguaggio e la codifica di dati materiali non linguistici o pre-linguistici.

Poniamo allora il concetto di informazione come di una qualità associata ad un messaggio¹⁵. Dunque occorre partire dal messaggio, il quale va visto nell'ambito del sistema comunicativo. Lasciamo implicito come precedente logicamente il fenomeno della conoscenza, intesa p.es. come una delle attività della coscienza, cioè della base di ogni disciplina spirituale, ovvero filosofica. Questo ci porta a mio avviso in ambito fenomenologico, husserliano, ed è un livello preliminare che può essere trascurato in questa sede. Potremo allora introdurre i concetti di coscienza e volontà della produzione di simboli comunicativi. I simboli linguistici sono espressione della volontà comunicativa. I simboli non linguistici possono essere considerati come l'espressione involontaria di una cultura, ma talora volontaria dell'esplicazione pratica o anche estetica etc. di una cultura. La competenza è ciò che ci assicura che i segni di un certo sistema sono fatti per trasmettere un certo messaggio, e non altri.

La scelta e la caratterizzazione del materiale, dell'oggetto di studio, è da questo punto di vista un problema di codifica: esse realizzano il passaggio dal dato materiale, testimone di una cultura, alla sua espressione linguistica. Rileviamo più competenze in azione: competenza del produttore, competenza del riconoscitore, competenza dell'analista (studioso), competenza del ricevente (altro studioso) che cerca di migliorare l'espressione linguistica dei dati.

Qui entrano in gioco le strutture dei vari livelli di linguaggio adatti ad esprimere il rapporto fra le diverse competenze ed i dati come ad esse si presentano. Esse, come nota Doran¹⁶ (sia pure in contesto diverso) devono offrire la possibilità di riprodurre le strutture dei dati all'interno della codifica:

«Any model will embody certain *structural assumptions* about the target cultural system. These specify the elements or components to be discerned within the system and the relationships to be discerned between them. A good choice of structural assumptions is crucial to the success of the modelling exercise. Structural assumptions express the modeller's view of the target cultural system. These assumptions must themselves be formulated within a symbolic language of formalism (if they are to be subject to mathematical inference or computation) and this symbolic language or formalism will in turn embody a formal conceptual repertoire...» (p. 447-448).

¹⁵ Cf. il fondamentale W. NOTH, *Handbook of Semiotics*, Indiana Univ. Press, 1990, 134: «Information is a semantic property of a message because only meaningful messages can be informative».

¹⁶ *Modelling Cultural Systems*, in F. DJINDJIAN, H. DUCASSE (eds.), *Data Processing and Mathematics Applied to Archaeology*, European University Center = PACT 16, 1987, 447-455.

È qui espresso molto chiaramente il rapporto (che è assai problematico) fra la struttura del codice e la struttura dei dati "reali". Vorrei aggiungere che le culture antiche in se stesse (o la "ancient mind") non sono le strutture dei significati dei segni, cioè delle vestigia di cui i segni sono simboli, ma consistono nella competenza che sta alla base della gestione dei segni da parte dei produttori. Forse la competenza in questo caso rappresenta l'interprete nel senso voluto da Peirce.

Occorre inserire tutti questi problemi nel loro contesto naturale, che a mio avviso è un contesto semiotico. La stessa cosa accade per la critica e l'analisi dei testi; successivamente ci si dovrà occupare di una teoria che unifichi i due campi. In campo archeologico il compito non è banale, perché il punto di partenza è dato da codici che non necessariamente dipendono dalla volontà di chi produce il messaggio. La semiotica tende a non occuparsi di questo genere di codici. Inoltre la semiotica tende ad occuparsi della produzione e trasmissione dei segni/messaggi/segnali, e non del riconoscimento di segni/messaggi/segnali senza conoscere preventivamente il codice¹⁷. Ritengo tuttavia che la riflessione semiotica sia l'unica che dia la possibilità di emettere una teoria coerente, in quanto aiuta a distinguere e riordinare i "livelli" in cui si posizionano i vari problemi. È senz'altro possibile rifiutare la semiotica come teoria sistematica generale entro cui inserire i problemi dell'espressione delle interpretazioni archeologiche etc. Resta il fatto che tali problemi vengono per lo più a coincidere con quelli discussi dalla semiotica, e la conseguenza è che, più che un rifiuto della semiotica, le proposte alternative possano essere ragionevolmente considerate come una diversa teoria, comunque nell'ambito della semiotica. E spesso tale teoria è l'ingenua riproposizione di soluzioni che la semiotica ha già riconosciuto come fallaci.

In sostanza il processo semiotico che propongo è il seguente:

- 1) riconoscimento delle vestigia materiali come messaggi, cioè come portatrici di significati in relazione alla cultura di chi le ha prodotte.
- 2) riduzione delle vestigia materiali a segnali/segni, cioè ad unità discrete portatrici di informazione.
- 3) riconoscimento delle strutture delle vestigia-segni. Tale riconoscimento sarà duplice: in relazione alla competenza di chi le ha prodotte; in relazione alla competenza di chi le studia. Questa dicotomia è data da quanto si diceva, della non volontà del messaggio di chi lo ha comunque inviato.
- 4) individuazione di un codice che consenta di codificare le singole vestigia-segni, e inoltre abbia una struttura interna che possa corrispondere alle due strutture (o anche ad una sola delle due) di cui al punto (3).
- 5) individuazione di un codice binario che consenta di (trans-)codificare su

¹⁷ Cf. però la text-semiotic, НОТН, *cit.*, p. 180 (ruolo del ricevente).

- supporto elettronico le codifiche ottenute al punto (4).
- 6) individuazione delle strutture espresse mediante la codifica del punto (4) [si badi bene, non (5)!], che obbediscano a regole tali da potersi gestire mediante un calcolatore [macchina di Turing].
 - 7) costruzione di un modello computazionale che corrisponda semioticamente alla struttura iniziale delle vestigia/segni, e permetta l'applicazione di procedimenti computazionali al fine di ottenere risultati di tipo sintetico, e non più solo analitico. Qui interviene il concetto di banca dati relazionale (e reti di Petri etc.), che è di portata molto più vasta di quanto non si creda normalmente. Esso rinvia infatti alla scelta delle entità che formano la struttura da studiare, e correlativamente il modello computazionale. Tali entità vengono caratterizzate, nelle tabelle, da un certo numero di attributi, ma è sempre difficile decidere quali attributi porre direttamente nella tabella di una entità, e quali invece trasformare in altre entità.

Volendo configurare quello che io possa vedere come sbocco operativo delle considerazioni sin qui fatte, è possibile emettere una "proposta finale": produrre una minuta di standard nel senso di un linguaggio descrittivo delle descrizioni, parallelo a SGML (o forse coincidente; ricordare che per la musica è stato fatto un linguaggio analogo); e una proposta di vari standard per la descrizione con quel linguaggio delle varie fonti archeologiche, parallelo al TEL.

TITO ORLANDI

C.I.S.A.D.U. = Università di Roma "La Sapienza"
Accademia Nazionale dei Lincei = Roma

ABSTRACT

Formalisation of data, which is the base of encoding procedures, raises a number of problems which should not be solved, as it is usually done, through generic intuition. Starting from Gardin's discussion of the «compilations», as distinct from «explanations», and from Gordon Childe's late epistemological propositions, the formalisation is defined as the production of structures of symbols which perfectly match the archaeological evidence as the scholars sees it. This may be done by means of different types of language; using computers requires modelling techniques. They depend on our appreciation of the evidence itself, and semiotics helps us in distinguishing between material evidence and its symbolic meaning.