

LE BANCHE DATI DELL'EPIGRAFIA: ESPERIENZE E PROSPETTIVE

È fatto ormai acquisito che il metodo di lavoro condiziona lo svolgimento e le conclusioni del lavoro medesimo: in particolare, l'uso delle metodologie informatiche nelle discipline umanistiche ha introdotto la necessità di procedere, in maniera diversa da quella tradizionale, all'identificazione ed alla organizzazione preliminare dei dati nell'ottica di una ricerca finalizzata alla creazione di grandi *corpora* su supporto magnetico, in breve nell'ottica della gestione *IRS* (*Information Retrieval System*); l'organizzazione delle ricerche scaturita da questa nuova analisi preliminare ha consentito un approccio a studi di più ampio respiro, offrendo talvolta la possibilità di individuare fenomeni che, per la molteplicità e la complessità delle cause che li determinano o li influenzano, sfuggivano all'indagine, e come conseguenza presentavano l'opportunità persino di una rilettura in chiave diversa di lavori precedenti. La microinformatica non ha quindi costituito un semplice rinnovamento sul piano tecnologico, ma ha posto gli studiosi davanti al problema concreto di ridurre a gruppi omogenei i fatti osservabili, e di costituire un percorso logico tra tutte le informazioni in possesso, offrendo così la possibilità di raggiungere risultati di straordinaria importanza.

Anche in epigrafia si è sentita la necessità di una innovazione metodologica, e da lungo tempo si fa uso di queste (non più) nuove tecnologie: esse si sono rivelate un potente strumento di ausilio allo studioso per le caratteristiche stesse della disciplina epigrafica, cioè in particolare per il numero dei testi da analizzare e per la loro dispersione in varie raccolte, purtroppo non più aggiornate né complete (si consideri che i rinvenimenti successivi alla pubblicazione del *CIL*, che si verificano per le iscrizioni con una certa continuità e regolarità, hanno a tutt'oggi globalmente ben più che raddoppiato il numero dei documenti noti ad esempio per la *regio VIII*, ma in certi casi per più volte questo numero è stato moltiplicato); una delle prime esperienze fu condotta da E. G. Jory e D. W. Moore presso l'Università di Western Australia, e si concluse con la pubblicazione dei 10 volumi degli *Indices vocabulorum nominibus propriis inclusis*, relativi al VI volume del *CIL* editi nel 1974-75 sotto l'auspicio dell'Accademia delle Scienze di Berlino. Tale lavoro fu forse non del tutto compreso nel mondo scientifico al suo apparire, ma ha poi rivelato nell'uso (che si può dire quotidiano, per gli specialisti) notevoli spunti innovativi nella messa a punto dei procedimenti di analisi della documentazione: si tratta di una indicizzazione nel formato *KWIC* (*Key Word In Context*) delle quasi 40.000 iscrizioni romane; la parola chiave risulta al centro della pagina, in un contesto di circa 90 caratteri, con il

riferimento, sulla destra, del numero dell'iscrizione e del numero di linea (per una presentazione di tale esperienza cfr. JORY 1975).

Da allora, importanti iniziative sono state condotte in vari paesi del mondo (ad esempio Stati Uniti, Francia, Gran Bretagna, Germania, Italia, Austria, Svizzera), utilizzando diversi sistemi di calcolo, e numerose realizzazioni si sono concretizzate, muovendosi fundamentalmente in tre direzioni:

1. creazione di banche dati delle iscrizioni, viste globalmente sotto gli aspetti monumentali e testuali;
2. realizzazione di archivi testuali;
3. realizzazione di indici lessicali.

Intendiamo per banca dati, semplificando, una raccolta omogenea di informazioni memorizzate su un supporto magnetico, ed accessibili mediante l'elaboratore; ogni unità informativa prende il nome di record, e ciascun record (identificato da un numero di accesso) contiene tutte le informazioni previste, suddivise in campi diversi: sostituisce cioè la nostra tradizionale scheda epigrafica.

Il principale obiettivo del processo di costituzione di un qualsiasi sistema informativo è il reperimento selettivo di particolari informazioni, cioè la funzione di *Information Retrieval*: essa si può svolgere secondo virtualmente infiniti algoritmi di ricerca dei documenti, ma non può prescindere dalla struttura del record, dalle voci insomma in cui è articolata la nostra scheda; è questo il motivo per cui, nella fase di progettazione di una banca dati, è metodologicamente importante tenere conto degli obiettivi d'indagine.

Passo successivo alla definizione delle singole voci di una scheda è quello relativo allo sviluppo della terminologia e della sintassi da adottarsi per registrare le informazioni nei vari campi, poiché l'uso di differenti termini per esprimere il medesimo concetto (sinonimia) o viceversa del medesimo termine per esprimere concetti diversi (polisemia) impedirebbero al sistema di restituire all'utente risposte precise e complete alle interrogazioni fatte: è indispensabile insomma definire uno standard terminologico, che verrà utilizzato nella fase di registrazione dei dati, ma che interesserà anche la fase successiva di interrogazione.

I diversi progetti di banche dati epigrafiche in corso di attuazione hanno risolto variamente questi problemi, di cui hanno dovuto necessariamente tenere conto.

Mi preme segnalare, tra le realizzazioni attualmente in corso di implementazione, P.E.T.R.A.E. (Programme d'Enregistrement, de TRaitement et d'Analyse en Epigraphie), nato da una collaborazione tra le Università di Bordeaux III, Besançon e Liegi, che registra ora numerose adesioni a livello internazionale, tra cui quella del Centro « Bartolomeo Borghesi » dell'Università di Bologna; tale

programma utilizza come *hardware* il sistema *Macintosh*, e come *software* *4^e Dimension*. Finalità di P.E.TR.A.E. è la costituzione di una banca dati epigrafica (comprendente sia le iscrizioni latine che quelle greche), che potrà in futuro essere anche distribuita *on line*. La struttura relazionale degli archivi è stata progettata partendo dalla constatazione che ogni testo è forzatamente associato ad un supporto monumentale (mentre un unico supporto può portare contemporaneamente numerosi testi): nell'ottica di evitare quindi inutili duplicazioni della descrizione del medesimo supporto si è stabilito di considerare tali schede come unità di base gerarchica dell'archivio, attribuendo loro un numero di riferimento formato da 2 cifre che corrispondono al codice della zona di provenienza (es.: Italia = 20), seguite da altre 2 indicanti il codice della regione antica (es.: *Aemilia* = 08), da 3 relative al codice dell'entità (intendendo per entità una città antica ed il suo territorio) e da 5 corrispondenti al numero d'inventario; questo riferimento numerico serve di collegamento con gli altri archivi.

Viene quindi definito ed analizzato in primo luogo il supporto monumentale, con una articolazione in campi formattati e a testo libero che permettono una precisa catalogazione di ogni suo aspetto; con la medesima precisione si puntualizzano poi gli aspetti del testo in esame. Non scendo in ulteriori particolari sulla struttura delle schede, ma tengo ad evidenziare come sia stato considerato fondamentale il poter ottenere una descrizione semplice e contemporaneamente articolata in maniera sufficiente da soddisfare le esigenze legittime dei professionisti della disciplina. Inoltre, la struttura consente l'acquisizione delle informazioni in maniera rapida e semplice, e la banca dati è aggiornabile in ogni momento (ovviamente da parte del personale autorizzato). Per quanto riguarda la trascrizione dei testi, si utilizzano *font* di caratteri creati appositamente, che pongono a disposizione tutti i segni diacritici usuali per l'epigrafista secondo il sistema di Leyda: greco puntato, latino puntato, nessi, soprilineature, interpunzioni di forma particolare, ed altro ancora.

P.E.TR.A.E. consente inoltre di indicizzare filologicamente (in maniera semiautomatica, quindi con un intervento interattivo dello studioso) tutte le parole dei singoli testi, secondo il programma elaborato presso il CIPL (*Centre Informatique de Philosophie et Lettres*) di Liegi. La banca dati mette a disposizione anche un modulo di schedatura prosopografica. Il linguaggio di interrogazione, cioè l'insieme dei comandi a disposizione dell'utente per effettuare la ricerca delle informazioni e la selezione dei documenti, consente di registrare le proprie strategie di ricerca, ed anche i set di documenti selezionati, per utilizzarli di nuovo in sedute successive di lavoro.

Sono attualmente in corso di realizzazione vari *corpora* regionali: Rodi e le isole rodie, Caria, Lesbo, Bulgaria, Egitto, Aquitania, regioni del Nord della penisola iberica, regioni del Nord dell'Italia.

Altro esperimento di schedatura epigrafica (anche qui è in corso l'implementazione) è quello realizzato presso il Dipartimento di Scienze Storiche, Archeologiche e Antropologiche dell'Antichità dell'Università di Roma « La Sapienza », per iniziativa del prof. Silvio Panciera (sulla scheda utilizzata, si veda DI STEFANO MANZELLA 1984; PANCIERA, DI STEFANO MANZELLA 1989).

Questa ricerca, nata per la realizzazione del Supplemento a *CIL*, VI, non si proponeva aprioristicamente un'applicazione generalizzata: a me pare invece che si tratti di procedura snella e agevole di catalogazione, attraverso la quale è possibile svolgere numerose e proficue indagini analitiche, e che assicura, oltre alla correttezza operativa, una notevole facilità di uso anche agli utenti non specialisti di informatica.

Con questo sistema è già stato schedato quasi tutto il materiale urbano non incluso in *CIL*, VI, conservato nel Museo Nazionale Romano delle Terme, ed il procedimento di indicizzazione automatica dei testi è stato già sperimentato su numerose raccolte.

Alla medesima necessità di mettere a disposizione uno strumento contemporaneamente flessibile e versatile di catalogazione epigrafica risponde il progetto proposto da G. Alföldy e M. Clauss sotto *ORACLE*. Senza entrare in una dettagliata descrizione di carattere tecnico, la strutturazione di questo archivio prevede una chiara organizzazione dei dati; l'accesso è semplificato dalla costituzione di menu e maschere interattive, e notevolmente flessibili sono le griglie predisposte per la raccolta della documentazione, nonché per l'elaborazione e la selezione dei dati.

Altre applicazioni si propongono la realizzazione di schedari mirati: ad esempio, ad Atene il *KERA* ha avviato un progetto di banca dati documentaria per la Macedonia (saranno analizzate più di 5.000 iscrizioni in pietra mediante un programma che gira su IBM), il *Freiburger Seminar für Alte Geschichte* ha in corso un progetto sull'epigrafia greca, l'Università di Oldenburg sta lavorando sull'epigrafia delle Gallie. Sentita è evidentemente, nel mondo scientifico, la necessità di procedere ad una analisi mediante elaboratore della documentazione epigrafica, e si può ancora ricordare, a titolo esemplificativo, il lavoro di R. W. Mathisen, teso a creare, in campo prosopografico, un archivio per la tarda antichità, che potrà in seguito essere sottoposto ad analisi statistiche o quantitative (mediante pacchetti già predisposti), e quello dell'Università di Melbourne, che ha in programma la ricostruzione dei gruppi familiari dell'Attica.

A Bologna è stata realizzata una procedura per l'esecuzione degli spogli lessicali, che non provvede alla lemmatizzazione (disponibile nel sistema P.E.TR.A.E., ma con l'intervento interattivo dello studioso), e restituisce indici delle forme utilizzate, organizzate in modo da elencare contemporaneamente sia parole singole che parole nel contesto (individuate, queste ultime, sulla base

di alcuni criteri-guida che hanno aprioristicamente organizzato e selezionato gli esponenti utili ad una corretta indicizzazione delle fonti epigrafiche); si è ritenuto che l'analisi del linguaggio così elaborato fornisca elementi utili ad enucleare dati sociali, istituzionali, economici, di storia militare e religiosa, tutti evidentemente di supporto nell'interpretazione e nella ricostruzione culturale delle comunità antiche, ed hanno influito in senso positivo nella scelta del sistema la sua totale automazione e la velocità di esecuzione: questi fattori consentono di indicizzare senza difficoltà né perdita di tempo qualunque selezione di testi memorizzati si ritenga utile in un particolare momento della ricerca. Si parte da un archivio di testi redatto mediante *WS* nel modo *non document* (ad evitare che il programma inserisca suoi caratteri speciali che renderebbero l'archivio non compatibile con altri programmi), tenendo presente che, per l'individuazione da parte della macchina dei formulari che si vogliono riportati negli indici, è fondamentale il concetto di "parola" (possiamo definirla come una sequenza di caratteri alfanumerici o speciali compresa tra due spaziature); diventa così perspicuo che, se si opera una sostituzione delle spaziature con un carattere speciale, automaticamente la "sequenza di parole" diviene per la macchina una unica "parola".

In conseguenza di ciò è stata adottata, nella fase di codifica, la sostituzione delle spaziature con caratteri speciali diversi:

- * si usa per separare parole che verranno tutte restituite negli indici come esponente chiave, in contesto con le successive;
- \$ separa parole che compaiono negli indici una sola volta, in catena, ordinate sotto l'esponente alfabetico della prima;
- serve perché il programma ricongiunga parole che proseguono sulla riga successiva;
- £ collega i formulari relativi agli anni di vita;
- % collega i formulari relativi agli anni di milizia;
- @ collega i formulari relativi agli anni di matrimonio.

Il testo viene poi trascritto per linee successive, con le abbreviazioni sciolte. Ovviamente, ogni trascrizione deve essere preceduta (e quindi individuata) da un'etichetta: ad esempio, per *CIL*, *XI*, *150*, si è scelto di usare 1100150, ma il criterio non è tassativo (inderogabile invece, per il funzionamento del programma, l'uso esclusivo di caratteri numerici). Al *file* di testo così creato si assegna un'estensione *.TXT*, per consentirgli di comunicare con il generatore di archivi. A questo punto, si manda in esecuzione, sotto *DBIII*, una serie di procedure, che trascrivono i testi successivamente in vari archivi, elaborando di volta in

1100135
C(aio)*Didio*Concordiano
v(ixit) a(nnis)VI m(ensibus)VIII d(iebus)XX
C(aius)*Didius*Celer n(auta) et lunia*_
Marcellina parentes
110137
C(aius)*Iul(ius)*Mygdonius
generi*Parthus
natus*ingenuus*capt(us)*_
pubis*aetate*dat(us)*in*terra(m)*Romana(m)
qui*dum*factus*_
cives*R(omanus) iuvente*fato*co_
locavi*arkam dum*esse(m)
annor(um)L peti(i)*usq(ue)*a*pub_
ertate*senectae*meae*pervenire
nunc*recipe*me*saxe*libens
tecum*cura*solutus*ero
1100138
D(is)\$M(anibus)
Iuliae*Primae
v(ixit) annisXI m(ensibus)III d(iebus)XVII
C(aius)*Iulius*Crescens
n(auta)
f(iliae)*carissimae

Fig. 1 — Esempio di introduzione, con codifica, di testi (sotto Wordstar)

a pubertate senectae meae pervenire	CIL, XI, 137
aetate dat(us) in terra(m) Romana(m)	CIL, XI, 137
anni di vita	
a(nnis) VI m(ensibus) VIII d(iebus) XX	CIL, XI, 135
annis XI m(ensibus) III d(iebus) XVII	CIL, XI, 138
annor(um) L	CIL, XI, 137
arkam	CIL, XI, 137
C(aio) Didio Concordiano	CIL, XI, 135
C(aius) Didius Celer	CIL, XI, 135
C(aius) Iulius Crescens	CIL, XI, 138
C(aius) Iul(ius) Mygdonius	CIL, XI, 137
capt(us) pubis aetate dat(us) in terra(m) Romana(m)	CIL, XI, 137
carissimae	CIL, XI, 138
Celer	CIL, XI, 135
cives R(omanus)	CIL, XI, 137
colocavi arkam	CIL, XI, 137
Concordiano	CIL, XI, 135
Crescens	CIL, XI, 138
cura solutus ero	CIL, XI, 137
dat(us) in terra(m) Romana(m)	CIL, XI, 137
Didio Concordiano	CIL, XI, 135

Fig. 2 a-b — Esempio di indicizzazione.

Didius Celer	CIL, XI, 135
D(is) M(anibus)	CIL, XI, 138
dum esse(m)	CIL, XI, 137
dum factus cives R(omanus)	CIL, XI, 137
ero	CIL, XI, 137
esse(m)	CIL, XI, 137
et	CIL, XI, 135
factus cives R(omanus)	CIL, XI, 137
fato colocavi arkam	CIL, XI, 137
f(iliae) carissimae	CIL, XI, 138
generi Parthus	CIL, XI, 137
in terra(m) Romana(m)	CIL, XI, 137
ingenuus capt(us) pubis aetate dat(us) in terra(m) Romana(m)	CIL, XI, 137
Iuliae Primae	CIL, XI, 138
Iulius Crescens	CIL, XI, 138
Iul(ius) Mygdonius	CIL, XI, 137
Iunia Marcellina	CIL, XI, 135
iuvente fato colocavi arkam	CIL, XI, 137
libens	CIL, XI, 137
Marcellina	CIL, XI, 135
me saxe libens	CIL, XI, 137
meae pervenire	CIL, XI, 137
Mygdonius	CIL, XI, 137
n(auta)	CIL, XI, 135; CIL, XI, 138
natus ingenuus capt(us) pubis aetate dat(us) in terra(m) Romana(m)	CIL, XI, 137
nunc recipe me saxe libens	CIL, XI, 137
parentes	CIL, XI, 135
Parthus	CIL, XI, 137
pervenire	CIL, XI, 137
peti(i) usq(ue) a pubertate senectae meae pervenire	CIL, XI, 137
Primae	CIL, XI, 138
pubertate senectae meae pervenire	CIL, XI, 137
pubis aetate dat(us) in terra(m) Romana(m)	CIL, XI, 137
qui dum factus cives R(omanus)	CIL, XI, 137
recipe me saxe libens	CIL, XI, 137
Romana(m)	CIL, XI, 137
R(omanus)	CIL, XI, 137
saxe libens	CIL, XI, 137
senectae meae pervenire	CIL, XI, 137
solutus ero	CIL, XI, 137
tecum cura solutus ero	CIL, XI, 137
terra(m) Romana(m)	CIL, XI, 137
usq(ue) a pubertate senectae meae pervenire	CIL, XI, 137
v(ixit)	CIL, XI, 135; CIL, XI, 138

volta i dati contenuti, fino ad ottenere un archivio strutturato come segue:

campo 1	c, 15	BIBLIOGRAFIA
campo 2	c, 70	PAROLA
campo 3	c, 70	ORDINE

in cui il campo BIBLIOGRAFIA contiene il rimando bibliografico principale (restituito ora per simboli alfanumerici: non più quindi 1100150, ma correttamente *CIL, XI, 150*), il campo PAROLA contiene quanto comparirà negli indici (sono scomparsi i caratteri speciali, sostituiti dalle spaziature), il campo ORDINE contiene l'elemento chiave per l'ordinamento alfabetico (elaborato automaticamente dal programma, con l'eliminazione quindi dei segni diacritici). L'*output* è organizzato di nuovo sotto *WS*.

Con questo sistema è stato indicizzato tutto il materiale epigrafico di provenienza ravennate, considerando compreso sotto la giurisdizione della città di Ravenna anche l'antico Delta padano fino alla supposta linea di confine col territorio di Adria, il ramo settentrionale del Po (area attribuita dal Mommsen, pur se con perplessità, alla *X regio*), con l'esclusione invece delle zone di Cervia, di Campiano, di S. Pietro in Trento e di S. Pietro in Vincoli, queste ultime unanimemente considerate ormai di pertinenza forlivese; parimenti esclusi i testi indicati provenienti da Salona (il lavoro è in corso di stampa nel volume *La storia di Ravenna*, Marsilio ed.).

A Bologna si sta anche procedendo, in collaborazione con l'IBC regionale, all'informatizzazione di tutto il materiale epigrafico di provenienza emiliano-romagnola: non solo quindi testi recuperati nell'area dell'antica *regio VIII*, ma anche in alcune aree limitrofe pertinenti in antico alle *regioni V e X*; in atto è anche la schedatura del materiale proveniente da alcune città della Tunisia (*Lep-tis Minor*, ad esempio).

In chiusura di questa carrellata sulle applicazioni dell'informatica all'epigrafia, non si può fare a meno di ricordare il Colloquio svoltosi a Losanna, sotto il patrocinio dell'*AIEGL*, nei giorni 26 e 27 maggio 1989 sul tema « *Epigraphie et Informatique* », di cui è già disponibile il volume degli Atti; tale Colloquio è servito, come era negli intenti degli organizzatori, non solo a fare un bilancio delle ricerche condotte e in corso, a mettere a confronto una serie di risultati, a presentare un sommario delle problematiche, ma anche a marcare la necessità inderogabile di confrontare con una certa frequenza i risultati ottenuti dalle diverse *équipe*, e soprattutto ad evidenziare la mancanza di un coordinamento a livello internazionale, che agisca nel senso di ridurre gli sprechi sia di energie che di denaro.

In questa ottica, al termine del Colloquio, è stata designata una Commissione di tre membri (proff. P. Ducrey, R. Frei-Stolba, A. Bielman), incaricata di

raccogliere e diffondere l'informazione tra gli studiosi, ed anche di tentare una omogeneizzazione dei sistemi in uso, tale da consentire scambi e comunicazioni tra banche dati, e da evitare una parcellizzazione delle iniziative; il primo passo è l'adozione da parte degli specialisti del medesimo *font* di caratteri per la trascrizione dei testi, ed a questo scopo è stato recentemente distribuito un questionario per una scelta tra varie proposte.

Nella medesima occasione, è stata votata all'unanimità una risoluzione, in base alla quale, già dal 1990, la rivista « Epigraphica » ospita una rubrica *AIEGL*, su *Epigraphie et Informatique*, destinata a concedere spazi a queste problematiche in continua evoluzione, e a consentire una recensione sistematica delle iniziative in atto.

PAOLA GIACOMINI

Dipartimento di Storia Antica
Università di Bologna

BIBLIOGRAFIA

- JORI E. J. 1975, *Problems and prospects for the production of computer compiled indices to epigraphic works*, « Antiquités Africaines », 9, 15-22.
- DI STEFANO MANZELLA I. 1984, *Progetto di una scheda epigrafica computerizzata, 1981/83*, in A. DONATI (ed.), *Il Museo Epigrafico*, Faenza, 159-213.
- PANCIERA S., DI STEFANO MANZELLA I. 1989, *Un esperimento di schedatura epigrafica e di elaborazione automatica dei dati. Aspetti tecnici e scientifici*, in *Epigraphie et Informatique*, Lausanne, 11-16.

ABSTRACT

The dominant mode of working has changed and is changing at an ever-increasing speed, also for the epigraphists: many computer-applications are today available to process epigraphical data; this item provides a rapid survey of the major types of computer-based projects in epigraphy. The procedure currently used at Bologna is here also explained: its purpose is to produce an index of forms in Roman inscriptions. Data entry is still in progress: the data are being entered on IBM computers, and a set of utilities has been written, to allow a two way transfer of data between WS and dBase III.