

Antonino Pennisi

PIANO DI INFORMATIZZAZIONE  
DELL'ATLANTE LINGUISTICO  
DELLA SICILIA (ALS)  
FONDAMENTI, STRATEGIE, PROGETTO

Estratto da  
*Materiali e ricerche*  
*dell'Atlante linguistico della Sicilia*, vol. I  
a cura di Giovanni Ruffino



CENTRO DI STUDI FILOLOGICI E LINGUISTICI SICILIANI

## PIANO DI INFORMATIZZAZIONE DELL'ATLANTE LINGUISTICO DELLA SICILIA (ALS). FONDAMENTI, STRATEGIE, PROGETTO.

### Premessa.\*

Nell'attuale quadro della problematica geolinguistica, macrosocio-linguistica e dialettologica un peso sempre crescente va assumendo il livello quantitativo dei *corpora* linguistici, delle loro classificazioni, delle loro analisi e la velocità di elaborazione, rappresentazione e pubblicazione dei risultati. Dal numero dei testi, dei soggetti, delle prestazioni, dei modelli di analisi, dipende non solo l'accresciuta adeguatezza descrittiva della teoria ma anche la prospettiva di ulteriori analisi qualitative. Nella rapidità di realizzazione delle ricerche si concretizza, invece, la possibilità di fornire agli studiosi strumenti di conoscenza in "tempo reale", o, per lo meno, in tempi sufficienti a garantire la veridicità delle informazioni in periodi, come quello che ci si appresta a vivere alle soglie del Duemila, di vertiginoso succedersi dei cambiamenti linguistici.

È questo - probabilmente - il motivo per cui nell'ultimo decennio si è molto sviluppato l'interesse verso l'informatica applicata alle modellizzazioni linguistiche che, appunto, si pone lo specifico compito di tradurre le possibilità aperte dall'elaborazione della quantità e della velocità in mozioni dedicate al sondaggio qualitativo.

La bibliografia accumulata in tale periodo costituisce già un vasto capitolo della più generale bibliografia linguistica e, soprattutto, essa

---

\* Al piano di informatizzazione dell'ALS aveva accennato Giovanni Ruffino nel quadro della sua relazione sull'*Atlante linguistico della Sicilia*, tenuta nell'ottobre del 1989 al XVII Congresso di linguistica e filologia romanza, a Santiago de Compostela. Una prima bozza di questo contributo è stata presentata nel novembre del 1989 alla Commissione Scientifica dell'ALS e poco dopo al Consiglio direttivo del Centro di Studi filologici e linguistici siciliani. Ringrazio particolarmente i professori Rosanna Sornicola e Alberto Varvaro per il contributo di idee e per gli stimoli da loro ricevuti in quelle occasioni.

Una seconda bozza organica è stata discussa in sede seminariale nel maggio del 1990 con il prof. Radtke e i suoi collaboratori, durante un soggiorno di scambi culturali tra una parte del gruppo di lavoro dell'ALS e il Romanisches Seminar dell'Università di Heidelberg. Durante il medesimo soggiorno è stata anche proficuamente avviata una discussione con alcuni specialisti dell'Istituto di Germanistica dell'Università di Marburg (in particolare con H. Händler, responsabile della progettazione dei grandi e importantissimi impianti di atlantistica informatica; con il prof. H. Thun dell'Università di Mainz, che ha illustrato le problematiche dell'*Atlas Linguistico-Etnografico del Uruguay* da lui diretto; infine con il prof. K. Mattheier dell'Università di Heidelberg, che ci ha fornito preziosi suggerimenti).

tende a diventarne uno dei settori privilegiati<sup>1</sup>. In essa, mentre risulta adeguato lo sforzo di sfruttare i tradizionali strumenti della macroinformatica legata ai grandi sistemi di sviluppo (laddove, naturalmente, questi già preesistano e siano affidati a tecnici esperti in *mainframes* e nel software di gestione ordinario), sembra tuttavia insufficiente l'esplorazione dei mezzi specifici della microinformatica, che, al contrario, dovrebbe costituire la sede strumentale più adatta alla sperimentazione linguistica e dialettologica, in particolare per i programmi di sviluppo di tipo regionale e/o locale che richiedono spesso soluzioni *ad hoc* e, non di rado, una specifica "inventiva informatica" negata ad hardware poco flessibile come quello dei grandi centri di calcolo. Sebbene per tale situazione possano essere avocate numerose giustificazioni (costo delle apparecchiature in relazione ai bilanci di istituti o dipartimenti di scienze del linguaggio, scarsità "istituzionale" di tecnici competenti, relativa "novità" del settore, etc.), non si può negare il clima di generale stagnazione della ricerca dialettologica e geolinguistica rispetto agli impetuosi sviluppi della tecnologia microinformatica aziendale e/o scientifica, alla sperimentazione in I.A. e, in campo specificamente umanista, alla lessicografia computerizzata.

In particolare il settore dei DBMS (*Data-Base Management Systems*) ha conosciuto momenti di autentico "salto" tecnologico grazie alle ricerche realizzate, sia nel campo dell'hardware sia in quello del software, e oggi ampiamente sfruttate dalla microinformatica personale.

Nel settore dell'hardware si è assistito soprattutto ad un incremento sbalorditivo delle quantità di memoria (RAM e di massa) disponibili, alla progettazione e realizzazione di nuovi microprocessori che incrementano in modo vertiginoso la rapidità di esecuzione dell'elaborazione sui *personal computers*, alla produzione di periferiche di input e di output che hanno permesso di portare al livello di analisi discreta (digitale) segnali continui (analogici) come le immagini o i suoni (compresi i suoni vocali).

Nel settore del software (anche volendo trascurare l'implementazione massiccia di procedure statistico-matematiche rivolte specificamente alla dialettometria<sup>2</sup>) alcuni concetti generali - quali, ad

<sup>1</sup> Riportiamo in Bibliografia i migliori studi a noi noti sull'argomento, suggerendo nelle note le indicazioni specifiche per ogni singola sezione. Per una discussione più generale sulle prospettive teoriche aperte da questi lavori, ulteriori riferimenti saranno contenuti nella relazione che verrà pubblicata negli Atti del Convegno su "Atlanti linguistici italiani e romanzi: esperienze a confronto" (Palermo, 3-7 ottobre 1990).

<sup>2</sup> Cfr. in particolare i lavori di Goebel (1974, 1982, 1983, 1984a, 1984b, 1986, 1987); Houck 1974; Shaw 1974; una rassegna di studi utile a questi scopi in Lo Faro 1989.

es., quello di relazionalità, elaborato da Codd<sup>3</sup> - sembrano aver determinato una radicale svolta nel dibattito sulla costituzione di banche-dati di grandi dimensioni implementate anche sui *personal computers*.

A fronte di questo rapido sviluppo, ed escludendo qualche rara eccezione, anche i contributi più recenti dell'informatica geolinguistica - quando non sono viziati da pregiudizi ormai del tutto anacronistici o da ingenue idealizzazioni degli strumenti di elaborazione dei dati - risultano comunque inadeguati e tendono a ripetere schemi di modellizzazione legati ad una fase pionieristica dell'informatica personale distribuita su scala industriale<sup>4</sup>.

Da qui il cauto tentativo di esportare nel piano programmatico di informatizzazione dell'ALS - ormai uscito dalla sua fase propriamente progettuale ed entrato nel vivo della realizzazione concreta e, quindi, della sua implementazione elettronica - le più recenti affermazioni nel campo della microinformatica personale. Di questo tentativo si cercherà qui di enucleare la problematica teorica di partenza e di esplicitare le premesse tecniche e i dettagli realizzativi rimandando ad altra sede le discussioni generali sulle possibilità aperte da quest'impostazione alle proiezioni speculative nei settori della ricerca geo-sociolinguistica e dialettologica<sup>5</sup>.

Il lavoro si articolerà in tre parti. Nella prima introdurrò le ragioni della domanda linguistica posta dal progetto ALS così come è andato delineandosi a partire dal 1984; nella seconda tratterò brevemente della problematica generale delle banche-dati nel ristretto ambito che può interessare la formazione di un archivio geolinguistico complesso; nella terza parte, infine, cercherò di mostrare come tutte le questioni tracciate preliminarmente troveranno una loro applicazione nella costituenda banca-dati dell'ALS.

---

<sup>3</sup> Cfr. Codd (1970, 1972, 1974, 1987). Vedi anche i lavori di Date 1984; Chen 1977; Aho-Beeri-Ullman 1977.

<sup>4</sup> Nel settore delle applicazioni microinformatiche l'uso dei p.c. in chiave geolinguistica è stato prospettato da Zampolli-Sobrero-Romanello 1986; Zampolli-Montemagni 1987; Geisler 1990a. Studi importanti sulla informatizzazione geolinguistica in generale sono quelli di Putscke, Naumann, Alvar-Nuño, Bauer (tutti citati in bibliografia). Acute osservazioni in Sobrero 1986; Händler (1977, 1983, molto utili anche per gli aspetti tecnici trattati); Cassidy 1969; Francis-Svartvik-Rubin 1969; Grassi 1970; Jones 1978; Shuy 1977; Weinreich 1973 e, più in generale, nei resoconti dell'informatizzazione di importanti progetti geolinguistici moderni (indicati in bibliografia).

<sup>5</sup> Ulteriori riflessioni spero di fornirle già nella mia relazione al corso del prossimo Convegno di Palermo sugli atlanti linguistici, ricordato alla nota 1.

### *1. La domanda teorica posta dal dibattito sull' ALS.*

L'idea di un *Atlante Linguistico della Sicilia* è stata formulata da Giovanni Ruffino, per la prima volta, nel 1984 e si è sviluppata in occasione di alcuni convegni, tavole rotonde, dibattiti, corsi di formazione per rilevatori, culminati in quattro importanti interventi di Varvaro 1986, di Stehl 1989 e dello stesso Ruffino (Congresso di dialettologia di Lugano, 1988; Congresso di linguistica e filologia romanza di Santiago de Compostela, 1989, entrambi in stampa) che, assieme ai contributi precedenti, hanno fissato le coordinate teoriche attorno a cui il progetto è oggi avviato<sup>6</sup>.

La ricchezza della discussione e la complessità ed articolazione dei problemi affrontati (di cui rende conto Ruffino in questo stesso volume) hanno permesso alla Sezione informatica da me diretta di acquisire tutta una serie di informazioni e considerazioni che sono poi divenute determinanti per la costruzione della piattaforma analitica su cui è stato costruito il software di gestione dell'ALS.

Di questa ampia discussione è qui impossibile stendere un resoconto dettagliato, ci limiteremo quindi a riassumere quella che può essere definita la domanda teorico-linguistica essenziale posta agli analisti e programmatori del progetto di informatizzazione dell'ALS.

#### **1.1. Variabilità.**

L'*Atlante Linguistico della Sicilia* nasce nelle intenzioni come uno strumento di integrazione tra una dialettologia con scopi principalmente descrittivi ed una geolinguistica, di matrice sociolinguistica, a vocazione prevalentemente variazionista. Su questo terreno di incontro trovano nel progetto una loro precisa collocazione sia la rilevazione di tipo diatopico sia la rilevazione di tipo diastratico e diafasico. Collocazione che, sebbene sia ancor oggi in fase di focalizzazione nel tentativo di evitare ogni appiattimento verso un'accezione puramente "sommatoria" della nozione di "repertorio" e, contemporaneamente, ogni confusionaria e inapplicabile mistione metodologica nel concreto lavoro di rilevamento empirico dei dati, è ormai definitivamente acquisita nei quadri progettuali dell'ALS. Lo scopo - lucidamente enucleato da Stehl - di "concepire la dialettologia regionale come una disciplina integrale che consideri tutti gli aspetti della variazione linguistica" (1990: 120) necessita ancora, tuttavia, di una serie di precisazioni e puntualizzazioni tecniche che riguardano sia la teoria che la pratica dialettologico-sociolinguistica specificamente

---

<sup>6</sup> Per una più attenta disamina della storia del dibattito sull'ALS, cfr. Ruffino in stampa b.

legate alla realizzazione degli atlanti, e, per ciò che ci riguarda più da vicino, le possibilità di informatizzarli.

Da un punto di vista teorico i problemi suscitati nel dibattito sull'ALS riguardano sia la dimensione delle sovrapposizioni diacroniche sia i livelli di incroci, contatti e mistioni linguistiche sincroniche nel concreto dell'ideologia e della prassi linguistica dei parlanti siciliani attuali. Se, come acutamente ha scritto Varvaro 1986 - e giustamente hanno ripreso Stehl 1990 e Ruffino (in stampa b) - "l'Atlante debba rispecchiare il modo di parlare di oggi, con tutto il suo sovrapporsi o contrapporsi di nuovo e di antico, di dinamico e di conservatore, di locale e di sovralocale", è evidente che lo strumento documentario debba porsi innanzitutto come repertorio comparativo tra rilevazioni precedenti e rilevazioni *in progress* e poi come luogo di individuazione, analisi, descrizione e definizione di macro-entità linguistiche di tipo geografico, sociologico, funzionale. Non solo: tale strumento non può esaurire il proprio compito fornendo criteri puramente tipologici, ma deve anche permettere di spiegare i flussi interattivi che intercorrono tra le classi individuate e che giustificano poi, in ultima analisi, le risposte date dagli intervistati a qualsiasi richiesta.

Parallelamente a questa esigenza variazionistica relativa al soggetto intervistato è emersa l'esigenza di approfondire la disamina della struttura del punto linguistico soggetto a sondaggio. Rispetto alle consuete variabili di tipo diatopico-geografico, nel progetto ALS un sempre maggior peso è stato attribuito - già in Ruffino (in stampa a), poi in Ruffino (in stampa b), infine nell'elaborazione di D'Agostino (in questo stesso volume) - ad una sorta di radiografia sociometrica dei luoghi e delle macro-aree individuate. La consueta dicotomia tra centri rurali presumibilmente "conservativi" e centri urbani presumibilmente "innovativi" - nel cui equilibrio, secondo Ruffino, è da ricercare l'auspicata "integrazione" tra dialettologia e sociolinguistica - è stata ritenuta insufficiente. La caratterizzazione aprioristica del punto come "espansivo" o "recessivo", l'identificazione del suo grado di "vitalità", la percezione quasi "intuitiva" del suo livello di "dinamicità" o "regressività", è stata sottoposta ad un processo di ridefinizione sulla base di una serie complessa e dettagliatissima di indici macrostatistici elaborati sulla base di informazioni rilevate dai principali istituti pubblici e privati di studi sociali. Questo tipo di lavoro ha permesso di rafforzare la già robusta rete di identificazione "anagrafica" dei punti ottenuta con i tradizionali strumenti di rilevazione delle caratteristiche geografiche e demografiche<sup>7</sup>.

---

<sup>7</sup> Cfr. Ruffino, nelle già citate comunicazioni ai Congressi di Lugano e Santiago de Compostela, aveva proposto e perfezionato una rete di punti di osservazione differenziati per vocazione (etnografici, marinari, "di base" e variazionali). Sulla base di que-

Da un punto di vista pratico, tuttavia, il pieno raggiungimento di una dimensione integrata del dato linguistico, sia per ciò che riguarda i soggetti e le loro prestazioni, sia per ciò che riguarda l' "anatomia" del punto linguistico, mal si coniuga con la struttura tradizionale di un atlante, trovando il suo luogo elettivo nella ampia monografia descrittiva di un singolo punto o area linguistica. Anche ampliando i confini dell'attuale concezione dello strumento-atlante, resta comunque il fatto che esso non può rinunciare totalmente all'intento topografico e alla rappresentazione cartografica del dato. Quando questo dato contiene una sua analisi interna (cioè incorpora in sé la differenziazione variazionale, sia essa di tipo diacronico, diastratico, diafasico, sociometrico) l'architettura cartografica complessiva, come *sintesi dell'intero processo di gestione delle informazioni*, deve completamente cambiare e trovare nuove strade rappresentative.

In realtà le esigenze maturate attorno al problema della variabilità linguistica spingono alla riformulazione degli scopi generali di un atlante linguistico e lo riconfigurano come un *centro di produzioni multimediali che esplorino in tutte le possibili dimensioni - soggetto, punto, classi di fenomeni - la rappresentazione e le risultanze delle interrogazioni poste*. Tali produzioni non dovrebbero rappresentare solo un nuovo modo di illustrare analisi tradizionali ma, al contrario, dovrebbero costituire una piattaforma conoscitiva che, accanto alle altre di più consolidata e collaudata metodologia, sposti il baricentro della speculazione teorica e, soprattutto, solleci la convalida e addirittura l'avanzamento stesso di nuove ipotesi interpretative<sup>8</sup>.

## 1.2. Livelli d'analisi.

Al problema degli scopi generali e al tipo di vocazione teorica di un atlante integrato si affianca, arricchendolo, la tematica relativa al livello linguistico che l'analisi deve battere. Si tratta, in altri termini, di fissare un rapporto tra gli obbiettivi descrittivi (storia dei parlanti, storia della lingua, descrizione sincronica e diacronica delle variazioni, descrizione delle macro-entità geografiche, sociali e funzionali, individuazione e proiezione delle tendenze, etc.) e gli oggetti di descrizione (fonetica, morfologia, sintassi, lessico e semantica, prosodia, etc.).

---

sto reticolo di punti si è messa a fuoco un'ottica "integrata" in equilibrio tra ricerca monodimensionale e ricerca pluridimensionale dell'ALS.

<sup>8</sup> Su questo proposito sembra convergere anche l'opinione di Sobrero che già aveva avanzato (nel 1983, contributo pubblicato poi nel 1986) a più riprese ipotesi analoghe per il suo NADIR, anche in relazione al progetto e agli scopi di un piano di informatizzazione.

Su questo punto è possibile persino rimettere in discussione le ipotesi generali sulle esigenze poste da una ricerca di tipo integrato. Gli esempi calzanti che Stehl 1990 ha tratto da Coseriu 1981 sulla descrizione dei processi evolutivi dei dialetti in gerarchie primarie, secondarie e terziarie, causato dal continuo cambiamento nei poli di comparazione e confronto, ad esempio, richiedono, per permettere una felice e utile applicazione dello schema, una precisazione del livello d'analisi. Se, continuando l'esempio di Stehl, l'oggetto di comparazione e raffronto dei dialetti siciliani è sempre stato il latino volgare mentre quello del siciliano contemporaneo - cioè del livello più innovativo - sarebbe l'italiano standard (producendo l'italiano regionale), diventa allora indispensabile precisare le coordinate descrittive di tale processo sia a livello fonetico che a quello morfologico o sintattico. Allo stesso modo la riprova di una pragmatica dell'interferenza linguistica che produrrebbe proiezioni e tendenze considerando solo una dimensione dei fenomeni di *code-switchings* o *code-mixings* (per non dire della definizione stessa di queste nozioni ancora in cerca di una precisa identificazione<sup>9</sup>), evocata dallo stesso Stehl, è impossibile senza una preventiva ricognizione e riconoscimento del valore probativo dei livelli di analisi linguistica e del loro grado di integrazione.

Un altro esempio può essere fornito dallo specifico della microinterazione di tipo diafasico. È impossibile proiettare dati statistici e rappresentativi generali - come un Atlante non può non fare - senza considerare in questo aspetto della variazione il livello prosodico, poiché è sempre possibile che fenomeni anche assai rilevanti possano restare "impermeabili" al livello d'analisi lessicale, morfosintattica e fonetica. Non è un caso che uno dei pochi atlanti che attualmente persegue scopi di ricognizione diafasica - e cioè l'*Atlas Lingüistico-Etnografico del Uruguay* diretto da Harald Thun - abbia percorso anche la via del sondaggio prosodico<sup>10</sup>. Lo stesso progetto di Thun, che, non essendo legato ad una tradizione conservativa, presenta una spiccata vocazione variazionista, considera anche un ampliamento del numero dei livelli possibili di analisi, includendovene uno definito "epistemologico", che

<sup>9</sup> Di grande interesse si preannuncia a tale riguardo la ricerca di Giovanna Alfonzetti che sta tentando una rigorosa definizione del problema proprio in riferimento al siciliano. Un primo consuntivo di questo tentativo in Alfonzetti (in stampa).

<sup>10</sup> Questo è uno degli aspetti più interessanti della presentazione dell'Atlante uruguayano che ci è stata fatta nel corso degli incontri seminariali tenutisi a Heidelberg, Marburg e Mainz (vedi la nota iniziale a questo lavoro). Purtroppo, per motivi di tempo, la descrizione della metodologia di rilevazione e analisi del dato prosodico è rimasta, in quella sede, alquanto vaga. Ci si aspetta da Thun, e da altri sociolinguisti impegnati in progetti geolinguistici, una precisa proposta di lavoro in questa direzione.



individua le variazioni "ideologiche" dell'oggetto-lingua dell'intervistato.

A questo proposito l'ALS - come ha più volte osservato Ruffino - può vantare una serie di privilegi di partenza che lo pongono nelle condizioni ideali per realizzare una più completa penetrazione interpretativa dei dati linguistici *stricto sensu*. Già Stehl 1990 scriveva, infatti: "un atlante regionale, per poter considerare la dimensione della pragmatica diacronica della variazione, dovrebbe o potersi appoggiare su inchieste precedenti, condotte sistematicamente a lungo termine nei centri più dinamici dello spazio geolinguistico da indagare, o dovrebbe prevedere la ripetizione di apposite inchieste dopo un arco di tempo significativo per la modificazione dell'atteggiamento variazionale nelle comunità linguistiche precedentemente indagate" (pp.142-3). Riprendendo e ampliando quest'osservazione si può dire che l'ALS non solo si trova nelle condizioni ideali di poter usufruire della massiccia ricostruzione diacronica avviata con le precedenti rilevazioni relative all'area siciliana (e cioè AIS, ALI, ALM, CDI) e con i contributi del *Vocabolario Siciliano* di G. Piccitto e G. Tropea - oltre che con una cospicua quantità di lavori monografici<sup>11</sup> - ma, soprattutto, esso può contare - sul versante, appunto, del livello "ideologico" o "epistemologico" (per usare le parole di Thun) - su quell'imponente opera di presondaggio "autovalutivo" costituita dall'Osservatorio Linguistico Siciliano (OLS), diretto da Franco Lo Piparo, ormai giunto alla fase di pubblicazione dei primi dati e materiali.

Infine, sempre attorno al problema dei livelli di osservazione del dato linguistico, è stato posto, dalla componente scientifica dell'ALS che sta sviluppando le prospettive etnografiche, il problema dei rapporti e dell'integrazione tra il repertorio della cultura materiale e il repertorio linguistico. Problema che Varvaro 1986 considera teoricamente risolto, inglobando la "competenza che riguarda la cultura materiale" nella "competenza comunicativa globale", ma che, in termini concreti di correlazione tra informazioni semantico-lessicali e informazioni etnografico-descrittive (principalmente iconiche) pone non pochi problemi di organizzazione a un sistema informatico elettronico<sup>12</sup>.

Su questo e altri punti relativi ai livelli di analisi e all'utilizzazione integrata dei materiali fonetici, morfosintattici, lessicali, prosodici e autovalutativi precedenti all'interno di un corpus atlantistico sorgono comunque una serie di problematiche specifiche che la Sezione informatica dell'ALS ha dovuto affrontare, non rinunciando all'obbiettivo

<sup>11</sup> Particolarmente quelli di G. Piccitto, A. Leone, G. Tropea, A.G. Mocciaro, G. Ruffino, S.C. Trovato (indicati, assieme ad altri, in Ruffino 1989).

<sup>12</sup> Cfr., tra i pochissimi lavori dedicati a quest'argomento, Cox (1972, 1973 e 1977).

di partenza di sfruttare il più a fondo possibile le potenzialità dei più moderni strumenti di informatica personale<sup>13</sup>.  
Torneremo in seguito su quest'argomento.

### 1.3. Rappresentazioni.

Un terzo ramo di problematiche teorico-pratiche poste dal dibattito sull'ALS riguarda le esigenze rappresentative dei dati soggetti a ricognizione sistematica. Si tratta in realtà di un elemento di riflessione non molto approfondito sino a questo momento, che, tuttavia, con l'approssimarsi della rilevazione empirica e dell'organizzazione della archiviazione non può più essere rinviato.

Si è già detto che un atlante che punti all'ambizioso progetto di integrare in un unico corpus la tradizionale descrizione diatopica con gli aspetti variazionistici, sia relativi ai soggetti che ai punti e ai livelli, deve essere capace di riformulare *in toto* i propri apparati rappresentativi e sondare i terreni, ancora molto impervi, della multimedialità. Per necessità di chiarezza cercheremo qui di introdurre una serie di distinzioni propedeutiche che agevolino questo compito.

La prima distinzione è relativa ai mezzi con cui possono essere rappresentati i fenomeni analizzati.

Partendo dal basso essi possono essere così schematizzati:

- a) Ascolto
- b) Scrittura
- c) Proiezione statistica
- d) Proiezione cartografica.

L'ascolto è fornito da tutti i materiali raccolti durante le interviste e registrati in bobine o cassette magnetiche. Questo tradizionalissimo tipo di ricognizione presenta una serie di vantaggi e svantaggi che ne condizionano l'uso, comunque ancor oggi ritenuto il più importante in assoluto. I vantaggi di questo tipo di rappresentazione sono evidenti: immediata e diretta fruizione del dato; possibilità di valutazione approfondita delle sfumature fonetiche; incondizionata reperibilità dei contesti di parlato in cui è pronunciata la risposta analizzata. Sugli svantaggi, tuttavia, non si è ancora discusso a fondo, dando per scontata l'impossibilità di trovare sostituti per la parte fonetica altrettanto fedeli quanto la registrazione magnetica. Un nastro magnetico, tuttavia, resta pur sempre un mezzo assai macchinoso per l'analisi: richiede, infatti, l'uso di apparecchiature complesse di riproduzione; la percorribilità sequenziale ne limita fortemente le possibilità di selezione interna; la per-

---

<sup>13</sup> Una prima organica precisazione di tali questioni in Ruffino (in stampa b).

cettibilità è legata inscindibilmente alla qualità del registratore, del nastro e delle condizioni di registrazione; le possibilità - infine - di manipolazione (rallentamento, velocizzazione, amplificazioni, etc.) e di analisi (spettrogramma, sonogramma, etc.) sono davvero minime e richiedono apparecchiature costosissime. Tutta questa somma di motivi, unitamente alla possibilità di accedere ad una strumentazione elettronica d'avanguardia, ci ha portati all'idea di affiancare alla registrazione magnetica la digitalizzazione elettronica del suono attraverso cui è possibile sfruttare appieno le potenzialità conoscitive dell'ascolto. Torneremo in seguito sui dettagli realizzativi e implementativi di questa importante riforma della struttura rappresentativa degli atlanti.

La scrittura è considerata, per importanza, il secondo metodo tradizionale di rappresentazione del dato linguistico raccolto. Tutti gli atlanti prodotti sino ad oggi sono basati sulla pubblicazione di liste scritte o di trascrizioni fonetiche su carta (nella stragrande maggioranza ripetuti poi su uno sfondo cartografico). Il momento della trascrizione e, soprattutto, della sua fruizione, è normalmente scisso dal supporto acustico. Diciamo che l'integrazione tra pronuncia effettiva (magneticamente registrata) e forma scritta è limitata al momento della trascrizione da parte del rilevatore o del supervisore. Da quel momento in poi la trascrizione fissata è riportata su liste o carte in modo definitivo e l'utente di un atlante deve accettare incondizionatamente il tipo di identificazione del suono che è stato formulato da chi ha operato la trascrizione. Oltre a ciò la fruizione scritta della forma è mediata dal tipo di alfabeto fonetico utilizzato (con i relativi problemi di standardizzazione che comporta). Gli inconvenienti cui si può andare incontro seguendo questa via sono numerosi: impossibilità di controllo effettivo della corrispondenza tra suono e trascrizione; discrasie tra diversi protocolli di trascrizione; difficile leggibilità per utenti non specialistici della trascrizione fonetica. È questo il motivo per cui in una forma rappresentativa rinnovata degli atlanti linguistici il ricorso alla microinformatica più avanzata è assolutamente indispensabile. Quest'ultima, infatti - e ne vedremo in seguito le specifiche realizzative -, permette di definire dalle fondamenta un sistema di trascrizione fonetica completa e direttamente visibile su schermo o su stampa; di affiancare (per utenti non specialistici) una forma trascritta ortograficamente (in modo automatico); e, soprattutto, di collegare la forma trascritta al suono attraverso il rapporto tra digitalizzazione della pronuncia e trascrizione fonetica derivata.

La proiezione statistica dei dati è stata sinora appannaggio delle ricerche sociolinguistiche, e, in particolare, di quelle più orientate sul versante dialettometrico<sup>14</sup>. Essa richiede preventivamente una ridu-

<sup>14</sup> Cfr., in particolare, oltre ai lavori di Goebel, Houck e Shaw, già citati alla nota 2, anche i contributi di Festinger-Schachter-Back 1969; Philips 1978; Roishoven 1984;

zione del dato linguistico a varianza numerica. La creazione di strumenti complessi come le matrici di similarità (che elaborano dei criteri quantitativi per riconoscere "parentele" linguistiche) o l'applicazione dell'analisi matriciale ai più sofisticati concetti della dialettometria gobeliana, sono sino ad ora rimasti fuori dal campo dell'applicazione all'atlantistica linguistica. Questi strumenti, naturalmente, richiedono un atteggiamento interpretativo assai cauto e spesso si rivelano facilmente falsificabili. E tuttavia un robusto apporto dialettometrico all'atlantistica linguistica può rivelarsi decisivo per districarsi all'interno di archivi di materiali linguistici stratificati. Più in generale si può considerare imprescindibile l'applicazione di strumenti di rappresentazione di statistica e grafica statistica (non necessariamente sofisticati come quelli citati) per la creazione di nuovi tipi di cartografia fondati sul concetto di ripartizione del fenomeno osservato su scale varianti e in relazione ai luoghi. Un primo esempio assai stimolante di questo tipo di applicazione lo possiamo ritrovare nel recente Friebertshäuer-Dingeldein 1989, informaticamente organizzato da Harald Händler, in cui l'importantissima scuola geoinformatica di Marburgo ha prodotto una splendida serie di carte sui diversi gradi di competenza linguistica, dei rapporti tra dialetto, tedesco standard e tedesco "medio", sui movimenti e le variazioni linguistiche relative alla regione dell'alto e basso Reno.

Restando nel campo della rappresentazione cartografica - prodotto "ultimo" nel tradizionale *iter* atlantistico - occorre distinguere tra cartografia legata ad un modello di analisi monodimensionale e cartografia proiettata sul versante variazionistico.

La prima è identificata da Stehl 1990 "con la dialettologia geolinguistica tradizionale con le sue ramificazioni verso una geolinguistica quantitativa. Compito di questo indirizzo della linguistica variazionale sarà nel futuro non soltanto concepire e compilare atlanti regionali dei dialetti primari e dei dialetti terziari, ma anche sviluppare i metodi computazionali per dei nuovi atlanti nazionali ampiamente sorretti da procedimenti statistici". In questi casi la produzione del dato cartografico è limitata ad un problema tecnico relativamente semplice: la proiezione della trascrizione (o di una sua simbolizzazione) su due assi di coordinate cartesiane ricavati sulla base di una mappa di sfondo che ne individua la localizzazione geografica<sup>15</sup>.

---

Schütz-Wenker 1966; Thomas (1977 e 1980); Wood 1970; Zampolli 1979; e Zeisel 1969.

<sup>15</sup> Su questo versante si collocano quasi tutti i principali lavori sulla cartografia elettronica. Cfr., oltre ai fondamentali Händler (1977, 1983, 1977); Handler-Naumann 1977; Naumann (1976, 1977a e b); Putschke (1975, 1977a, b,c, 1983); anche Bershin 1977; Brook 1970; Eickmans (1979 e 1981); Francis 1969; Kelle 1977;

Assai più complesso - ed è il caso dell'ALS - il processo di cartografazione relativo alla realizzazione di uno "strumento che, alle soglie del Duemila, documenti, attraverso un'ottica integrata, la condizione del dialetto (dallo stadio più arcaico sino ai livelli più spinti di italianizzazione), le attuali linee di tendenza e le dinamiche del cambiamento in chiave areale e diastratica" (Ruffino, in stampa b). Se aggiungiamo, inoltre, la possibilità di incursioni nelle dimensioni diafasiche e autovalutative (ritenute, come abbiamo visto, assai importanti) e attraversando l'osservazione dei livelli fonetici, prosodici, morfosintattici e lessicali, avremo un quadro cartografico di una complessità sino ad ora sconosciuta<sup>16</sup>.

Di fronte a questo ambizioso obiettivo l'unica possibilità di riuscita è lo stretto collegamento della cartografazione alla gestione elettronica dei dati in una banca-dati unitaria che fornisca diversi tipi di strumenti-variabili pienamente e incondizionatamente combinabili tra loro. A questo proposito occorrerà distinguere tra:

- a) Dato
- b) Selezione (fenomeno)
- c) Background.

Il dato da distribuire corrisponderà al mezzo utilizzato: cioè sarà possibile selezionare sulla base della forma acustica sonora (fonema, parola, sintagma), della forma trascritta (ortograficamente o foneticamente), o della forma numerica graficizzata (istogrammi, barre, torte, etc.).

La selezione da proiettare sarà costituita dal fenomeno "ritagliato" sulla base della varianza diacronica (fonti e anni di rilevamento), diatopica (punto, configurazione amministrativa, configurazione geografica, configurazione sociometrica), diastratica (età, sesso, istruzione), diafasica (valutazione e autovalutazione della tipologia del dato e delle situazioni di inchiesta), autovalutativa (ideologica, per la rete di punti con questionario autovalutativo a parte<sup>17</sup>), e delle loro combinazioni.

Il background sarà costituito dal tipo di cartina (geografica, economica, demografica, sociologica, etc.) che farà da base per lo sfondo.

Questi tre insiemi verranno unificati nell'oggetto-carta la cui realizzazione materiale in un manufatto utilizzabile dipenderà dal tipo di

Lance 1977; Lance-Slemons 1976; Southard 1983; Thomas 1975; Tokugawa-Yamamoto 1967; Wood (1961 e 1972).

<sup>16</sup> Oltre ad alcuni spunti contenuti nei testi citati alla nota precedente, interessanti i tentativi di produrre carte fonologiche - Dean 1965; Francis-Svartvik-Rubin 1969 -, semantico-lessicali - Dees 1980 -, e statistico-dialettometriche - Frieberthäuer-Dingeldein 1989.

<sup>17</sup> Cfr. Ruffino (in stampa a, b).

supporto sul quale riversare l'output finale. Quest'ultimo potrà essere un:

- a) output videografico
- b) output cartaceo
- c) output magneto-ottico.

Nel primo caso la cartografazione elettronica può risolvere quasi tutti i problemi di restrizioni materiali. Un grande schermo (doppio A4 affiancato) potrà contenere sia il background, sia la selezione nella sua forma di dato nudo (trascrizione della forma, caratteristiche diastratiche o diatopiche o diafasiche, grafico statistico di ripartizioni autovalutative, etc.) o manipolato (simbolizzato). Anche nel caso di una selezione che per ogni punto presenti una variabilità di forme multiple (sia con un lemma principale e lemmi secondari, sia nel caso di lemmi tutti paritari) si potrà ricorrere a stratagemmi di interfaccia grafica che permettano di vedere sempre tutte le forme (ad esempio tramite un menù *pull-down* o a tendina che srotoli su tocco del *mouse* un elenco di esiti normalmente contenuti all'interno di una sola stringa visibile)<sup>18</sup>. Inoltre il sistema a video permetterà l'integrazione con il dato acustico attraverso l'associazione di un suono digitalizzato corrispondente alla trascrizione (legata a sua volta ai dati nudi che contengono l'indicazione del tipo di varianza). Al semplice tocco di mouse nella regione di associazione di forma grafica e risorsa sonora digitalizzata la carta potrà "pronunziare" l'esatta pronuncia e l'eventuale contesto sintagmatico e/o frasale. Nel caso di simbologie, infine, l'output a video potrà variare il significato attraverso la colorazione.

Il supporto cartaceo presenta al contrario numerose limitazioni. Innanzitutto l'impossibilità di contenere - a meno che non si disponga di uscite su un *plotter* molto spazioso - tutte le forme trascritte foneticamente e i dati nudi collegati. In secondo luogo l'insufficienza materiale a contenere delle selezioni multiple di forme trascritte e dati nudi collegati (esiti multipli per singoli punti). Infine l'impossibilità di ricorrere (a meno di spese straordinariamente elevate) all'uso del colore come elemento di differenziazione significativa. È quindi giocoforza in questo caso utilizzare una simbologia relativa all'insieme di forme trascritte e dati nudi collegati, con gli inconvenienti interpretativi che si conoscono<sup>19</sup>.

---

<sup>18</sup> Per la spiegazione di questi termini (menù a tendina, *mouse*, etc...) cfr. la seconda e terza parte di questo lavoro.

<sup>19</sup> La simbolizzazione è quasi inevitabile laddove i lemmi selezionati si aggirano nell'ordine delle centinaia. Questa pratica - usata in moltissimi atlanti, informatizzati o meno - come quello del Giappone, o l'*Atlas Linguarum Europae*, o, più di recente, l'*Atlante linguistico della Colombia* - mortifica le possibilità di utilizzazione finale

Infine il supporto magneto-ottico è il più flessibile potendo contenere il dato sotto forma di *file* (soggetto sempre a elaborazioni successive). La memorizzazione magneto-ottica delle carte (che naturalmente corrisponderebbero alle più complesse carte-video) presenta comunque l'inconveniente di dover ad un certo punto essere trasferita, per l'uso, o su video o su carta, ripresentando così i problemi già discussi.

Strettamente legato a queste problematiche di rappresentazione è il problema, molto più importante, del tipo di scopo che si vuole raggiungere con una cartografazione elettronica di un atlante così fortemente segnato dalla prevalenza della variabilità come criterio di distinzione teorico-pratica. Come dicevamo prima non bisogna concepire la ristrutturazione del mezzo-atlante come una semplice riformulazione del *look* rappresentativo. L'atlante elettronico, così come questo progetto lo concepisce, sfrutta sempre le possibilità rappresentative multimediali in stretta relazione con la funzionalità speculativa che tali rappresentazioni devono avere per l'utente. Cambiando l'utente e lo scopo le rappresentazioni cambiano di concerto. Un esempio calzante è costituito dal fatto che lo sfruttamento di archivi elettronici è possibile anche durante lo svolgimento stesso dell'analisi, fornendo sempre un quadro dinamico e aggiornato dei dati immessi sino a un certo punto. Occorre, quindi, prevedere un output qualitativamente anche meno elevato ma molto utile per la determinazione e la comprensione di tendenze in atto e di fenomeni che si stanno "inseguendo", ed un output finale adatto ad una pubblicazione definitiva, che può richiedere procedimenti, strumentazioni e elaborazioni molto più lunghe e sofisticate.

Da questo punto di vista occorre formulare un'ultima, utile differenziazione tra atlanti che nascono informatizzati già in fase di progetto ed atlanti che "subiscono" un'informatizzazione a posteriori. I primi possono godere appieno di una architettura modulare che permette di selezionare gli scopi sulla base della domanda teorico-pratica iniziale (ed è il caso del NADIR o, appunto, dell'ALS). I secondi potranno arrivare ad una relativa facilitazione e velocizzazione dei procedimenti tradizionali di stampa (*desktop publishing*) e gestione elettronica degli archivi e dei loro output visuali (*desktop presentation*).

#### 1.4. L' ipotesi di lavoro.

Sulla base di quanto abbiamo sin qui brevemente illustrato, riassumiamo schematicamente le esigenze teorico-pratiche poste dal dibattito

---

rapida e intuitiva, poichè la percezione della unitarietà del fenomeno cercato viene soffocata dalla pluralità di forme da decodificare con le apposite legende. Sarebbe opportuno limitare questa pratica - laddove possibile - ai fenomeni di tipo fonetico, morfo-sintattico, autovalutativo o prosodico, cercando di evitarne il ricorso almeno nelle carte di tipo lessicale.

sull'ALS (per il quadro completo dei problemi, si veda Ruffino in questo stesso volume), che sono:

- 1) necessità di integrare la tradizionale descrizione diatopica con la dimensione diacronica, diastratica, diafasica e "ideologica" (autovalutativa) della ricerca geolinguistica;
- 2) necessità di scandagliare, in profondità più o meno rilevante, i livelli di analisi fonetica, prosodica, morfosintattica e lessicale;
- 3) necessità di elaborare un sistema di individuazione dei processi di unificazione e/o diversificazione di macroentità dialettali;
- 4) necessità di fornire una base programmatica e sistematica per la ricognizione dei flussi storico-sociologici e sincronico-pragmatici dei movimenti interni alla macro-comunità siciliana e alle realtà areali più focalizzabili;
- 5) necessità di produrre nuove forme rappresentative dell'indagine geolinguistica e, in particolare, atlantistica, capaci di affiancare agli strumenti tradizionali (comprese le monografie "dedicate") un pacchetto completo di strumenti multimediali adatti ad illustrare nelle modalità più opportune descrizioni, proiezioni e tendenze dei fenomeni linguistici sulla base delle ipotesi avanzate e delle ricerche empiriche realizzate.

Su questa mozione dei *desiderata*, e dopo un'ampia discussione di mediazione tra le esigenze teorico-pratiche della Commissione scientifica e le possibilità e opportunità concrete di formalizzazione proposte dalla Sezione informatica dell'ALS<sup>20</sup>, il piano di informatizzazione si è concentrato nell'elaborazione di una piattaforma programmatica che permetta le seguenti operazioni:

- 1) immissione, recupero e manipolazione del materiale registrato nelle interviste a livello
  - a) fonetico-ortografico
  - b) fonetico-trascrittoria
  - c) fonetico-vocale
  - d) lessicale
  - e) morfo-sintattico
  - f) prosodico
  
- 2) immissione, recupero e manipolazione delle informazioni concernenti le variabili sociolinguistiche selezionate dei soggetti intervistati, ed in particolare:
  - a) sesso
  - b) età

---

<sup>20</sup> Discussione che si è mantenuta tenuta costante negli anni 1988-1990.



- c) istruzione
  - d) professione
  - e) profilo comunicativo
  - f) autocoscienza linguistica ed autovalutazione del sentimento linguistico e delle abilità comunicative
- 3) immissione, recupero e manipolazione delle informazioni concernenti le variabili sociometriche relative ai punti linguistici toccati dall'indagine, e cioè:
- a) popolazione
  - b) indici commerciali
  - c) indici di urbanizzazione
  - d) indici di consumo
  - e) indici di espansività
  - f) indici culturali
  - g) combinazioni statistiche degli indici
- 4) immissione, recupero e manipolazione delle informazioni concernenti le variabili geografiche relative ai punti linguistici toccati dall'indagine, e cioè:
- a) comune
  - b) provincia
  - c) tipologia delle localizzazioni geografiche
  - d) configurazione amministrativa
  - e) configurazione macro-economica
  - f) configurazioni areali, demografiche, di comunicazione
- 5) immissione, recupero e manipolazione delle informazioni concernenti le variabili geolinguistico-dialettologiche relative ai soggetti intervistati, e cioè:
- a) rilevazione diacronica delle fonti
  - b) valutazione tecnica qualitativa delle fonti
  - c) tipologia delle fonti
  - d) tipologia linguistica dell'intervistato
  - e) tipologia delle domande
  - f) tipologia articolata delle interrogazioni semasiologiche
  - g) tipologia morfologico-sintattica
  - h) tipologia semantica articolata
  - i) tipologia delle risposte
  - l) valutazione della vitalità linguistica
  - m) autovalutazione della vitalità linguistica
  - n) valutazione dell'uso comunicativo
  - o) autovalutazione dell'uso comunicativo
  - p) valutazione della forma espressiva
  - q) autovalutazione della forma espressiva

- r) valutazioni e annotazioni extralinguistiche
  - r1) relative alle condizioni materiali dell'intervista
  - r2) relative al comportamento dell'intervistato
  - r3) relative alla pronuncia
  - r4) relative alla definizione tecnica della risposta
- s) valutazioni e annotazioni sulla varianza fonetico-morfologico-sintattica

6) immissione, recupero e manipolazione delle informazioni concernenti i reperti etnografici (fotografie, illustrazioni, disegni, notizie, storia, cronologia, etc.)

7) realizzazione di un sistema di cartografazione automatizzato che riesca a produrre in output non solo mappe tradizionali ma tutta una serie di carte speciali derivanti da:

- a) manipolazione di tutte le variabili contenute negli archivi, compreso ogni possibile incrocio e combinatoria
- b) intercambiabilità dei background con ampia scelta delle cartine di sfondo per la proiezione dei dati.

Per la realizzazione di un sistema informatico che tenga conto di questa imponente lista di obiettivi, e perseguendo lo scopo di superare i limiti di flessibilità posti dai sistemi che operano sulla base di piattaforme hardware e software dei grandi centri di calcolo, si è scelto di operare utilizzando una serie di apparecchiature microinformatiche tecnologicamente avanzate che permettano di soddisfare alcuni requisiti ritenuti pregiudiziali. E più precisamente<sup>21</sup>:

Per l'hardware:

1) adozione di unità centrali gestite da microprocessori capaci di realizzare velocità rilevanti di elaborazione-dati come:

- a) Motorola 68000 con clock a 16 Mhz
- b) Motorola 68030 con clock a 25 Mhz
- c) Motorola 68030 con clock a 33 Mhz
- d) Motorola 68030 con clock a 40 Mhz (di prossima acquisizione)
- e) Motorola 68040 con clock a 50 Mhz (di prossima acquisizione)

2) uso di coprocessori matematici schiavi (*slave*) per sgravare il sistema da oneri calcolistici e incrementarne la velocità di esecuzione

---

<sup>21</sup> Per una chiarificazione dei termini tecnici qui adottati si rimanda alla seconda e terza parte di questo lavoro.

- 3) memoria RAM di base di 5MB estensibile a 16 GB
- 4) memoria ROM di 256 e 512 Kb contenenti tutte le routines grafiche e di gestione del sistema non gravanti sul S.O. (Sistema Operativo)
- 5) parola di 32 bit piena
- 6) porte di comunicazione con periferiche di tipo ad altissima velocità (RS 422, RS442, SCSI con *transfer-rate* sino a 2MB per secondo)
- 7) *cache-memory* e memoria paginata
- 8) 6 slot di espansione per qualsiasi scheda video o di altro tipo più una serie di porte dedicate NUBUS per collegamenti con dispositivi di conversione analogico-digitale
- 9) memoria di massa e di *backup* magnetiche e ottico-magnetiche sino a 1.2 GB
- 10) periferiche di input di tipo unitario (*mouse*) e digitali (*scanner* a 300 dpi), oltre le normali tastiere
- 11) periferiche di output ad alta definizione (stampante laser) e *plotter* (di futura acquisizione)

Per il software:

- 1) sistemi operativi a interfaccia grafica e orientati all'utente (interfaccia amichevole) con ampio uso di standardizzazione di dispositivi software di input-output come:
  - a) menù a tendina (*pull down menù*)
  - b) bottoni di conferma o opzione o operativi
  - c) finestre scorrevoli e multiple
  - d) gestione standardizzata dei font con integrale applicazione del WYSIWYG
- 2) sistemi operativi adatti al *multitasking* e al collegamento con grandi centri di elaborazione e calcolo (UNIX)
- 3) generatori di applicativi D-B totalmente relazionali e capaci di sfruttare appieno l'interfaccia grafica del S.O.
- 4) linguaggi di programmazione capaci di sfruttare appieno l'interfaccia grafica del S.O.

- 5) applicativi per la costruzione apposita di alfabeti fonetici
- 6) programmi di campionamento sonoro e digitalizzazione di segnali acustici continui
- 7) programmi di acquisizione di immagini e digitalizzazione di segnali grafici continui
- 8) software di elaborazione lessicografica e statistica di testi di qualsiasi mole ed anche in trascrizione fonetica
- 9) software grafico per qualsiasi manipolazione in B&N e a colori delle cartografie
- 10) programmi di calcolo, statistica, grafica statistica ed elaborazione numerica adatti a applicazioni dialettometriche
- 11) applicativi di comunicazione telematica e/o di reti omogenee o miste
- 12) programmi di videoscrittura con integrale applicazione del WYSIWYG anche ai font di trascrizione fonetica appositamente costruiti.

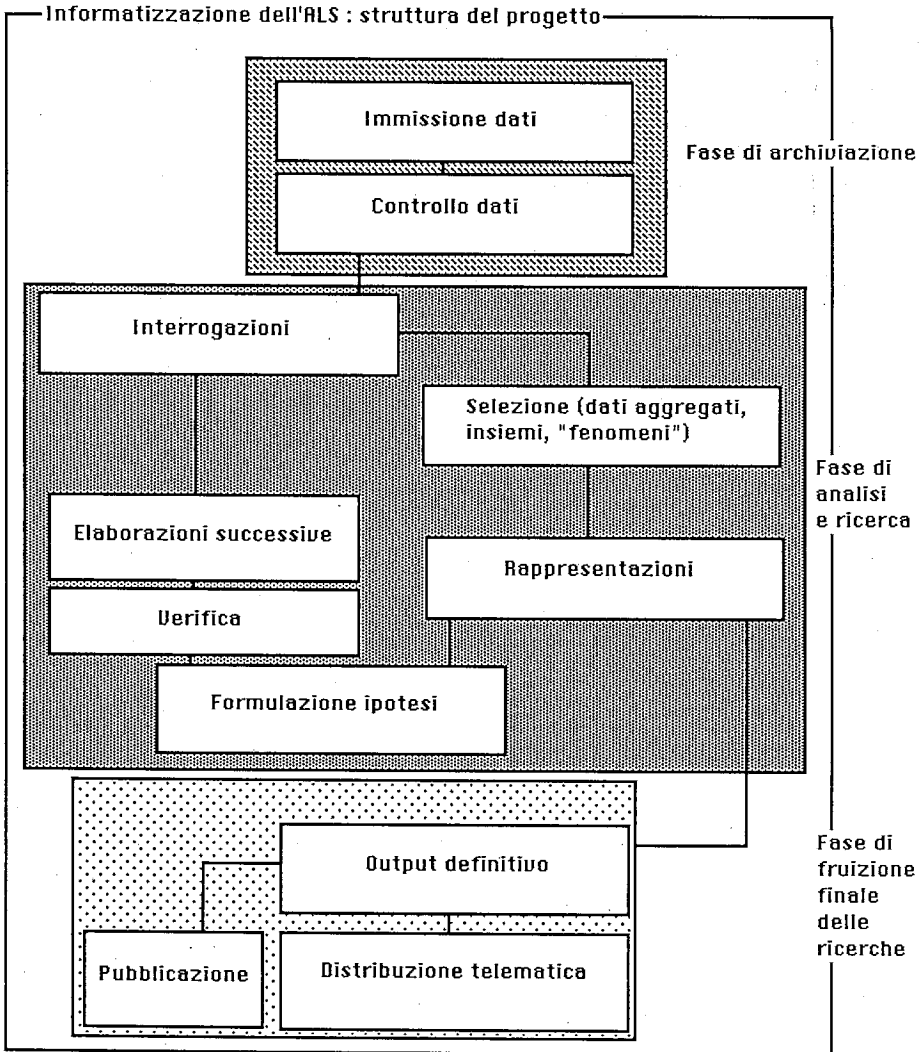
L'integrazione completa tra scopi e strumenti utilizzati ha permesso di risolvere alcuni problemi tradizionali (trascrizione fonetica, recupero incrociato delle informazioni, valutazione di configurazioni complessive della "vitalità" linguistica del punto, valutazione del "sentimento linguistico del parlante", configurazione degli usi e dei repertori linguistici e sociolinguistici, etc.) e di proporre alcune soluzioni del tutto nuove.

Tra queste ultime è importante ricordare *l'implementazione della risposta fonetica con integrale riproduzione del suono della risposta e del contesto in cui essa viene pronunciata e una cartografazione interamente variazionale* con risposte in trascrizione fonetica e con la possibilità di generare cartine speciali.

È possibile, a questo punto, delineare una scansione articolata dei diversi livelli realizzativi del progetto. Essi sono riassumibili in tre fasi principali:

- 1) fase di archiviazione
- 2) fase di analisi e ricerca
- 3) fase di fruizione finale delle ricerche

Queste tre fasi sono connesse nello schema semplificato dell'intero progetto (tavola n° 1).



tav. 1

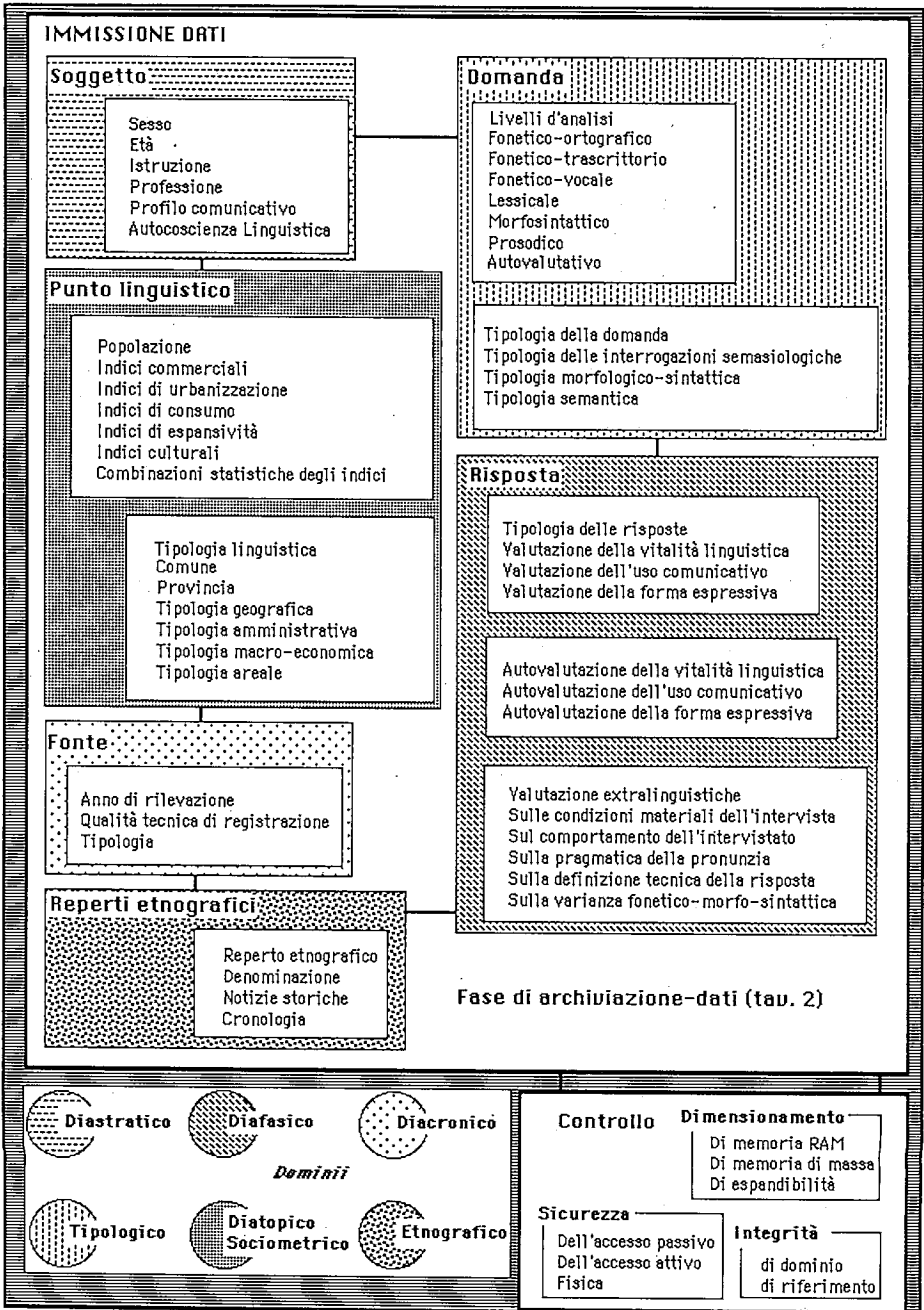
La prima fase comprende le problematiche relative all'immissione dei dati grezzi all'interno di una Banca-Dati (su cui verterà la discussione della seconda e terza parte di questo lavoro) e sul loro controllo. Si tratta in primo luogo di realizzare un *frame* complessivo per l'organizzazione delle informazioni funzionalizzate alla soddisfazione informatica degli scopi variazionisti dell'ALS. Ad ogni dominio previsto - diatopico, diastratico, diacronico, diafasico, sociometrico, etnografico - corrisponde un'area tematica dedicata alla memorizzazione (in seguito vedremo che ciò significherà creare un sottoarchivio separato e specifico di una banca-dati relazionale per qualsiasi dominio).

L'ipotesi diastratica, ad esempio, si concretizza informaticamente nella possibilità di organizzare profili esaurienti dei soggetti campionati considerandone il sesso, l'età, l'istruzione, la professione, la predisposizione comunicativa - come le competenze attive e/o passive del dialetto e dell'italiano, le capacità metalinguistiche, etc. - e, per quanto riguarda i soggetti del reticolo di punti più specificamente variazionista, la valutazione integrata del grado di autocoscienza linguistica, affidato ad un questionario specifico, ancora in fase di collaudo (cfr. Ruffino e Stel in questo stesso volume) che rielabora lo schema adottato dall'OLS. Ad ognuna delle informazioni, opportunamente quantificate in numeri, stringhe alfanumeriche, scelte sì/no, etc., verrà associato un campo (argomento su cui torneremo nella seconda e terza parte di questo lavoro) memorizzato attraverso una codificazione atta al facile recupero del dato. A questi profili diastratici vanno affiancati non solo i profili di tutti gli altri domini ma anche un sistema articolato delle tipologie di interrogazione (privilegiando ora il dato fonetico - con trascrizione ortografica o fonetica più la pronuncia associata -, ora quello morfosintattico - con valori numerici relativi ad una classificazione tipologica delle categorie grammaticali o dei fenomeni<sup>22</sup> -, ora quello lessicale - aree semantiche con strutture "a fisarmonica"<sup>23</sup> siglate -, ora quello prosodico - curve opportunamente simbolizzate e/o numericamente trascritte -, ora quello autovalutativo - nel caso già ricordato di inchieste su punti a vocazione prevalentemente variazionista -, etc.). Speciali informazioni sono immagazzinate nel dominio etnografico: si tratta delle illustrazioni o delle foto digitalizzate tramite scanner. Qualsiasi dato immesso verrà sottoposto a una serie di controlli e *tests* di congruenza (relazionale, di dominio, di riferimento, etc.), di legalità (accesso consentito) e di dimensionamento<sup>24</sup>. Nella tavola n°2 si può vedere lo schema dettagliato di questa fase.

<sup>22</sup> Anch'essa attualmente ancora in fase di discussione preliminare.

<sup>23</sup> Per una discussione sulla loro articolazione cfr. Ruffino (in questo stesso volume). La nozione di liste "a fisarmonica" è approfondita nella terza parte di questo lavoro.

<sup>24</sup> Tutti questi argomenti sono ampiamente discussi nella seconda e terza parte di questo lavoro.



La seconda fase gestirà il sistema delle interrogazioni della Banca-Dati e in essa andranno specificate tutte le ipotesi di selezione dei fenomeni, delle loro modalità rappresentative, del processo di formulazione delle ipotesi che - attraverso un momento di verifica empirica e di elaborazioni successive - porterà alla convalida o alla falsificazione delle ipotesi specificamente "inseguite" e alla scelta dei fenomeni da illustrare.

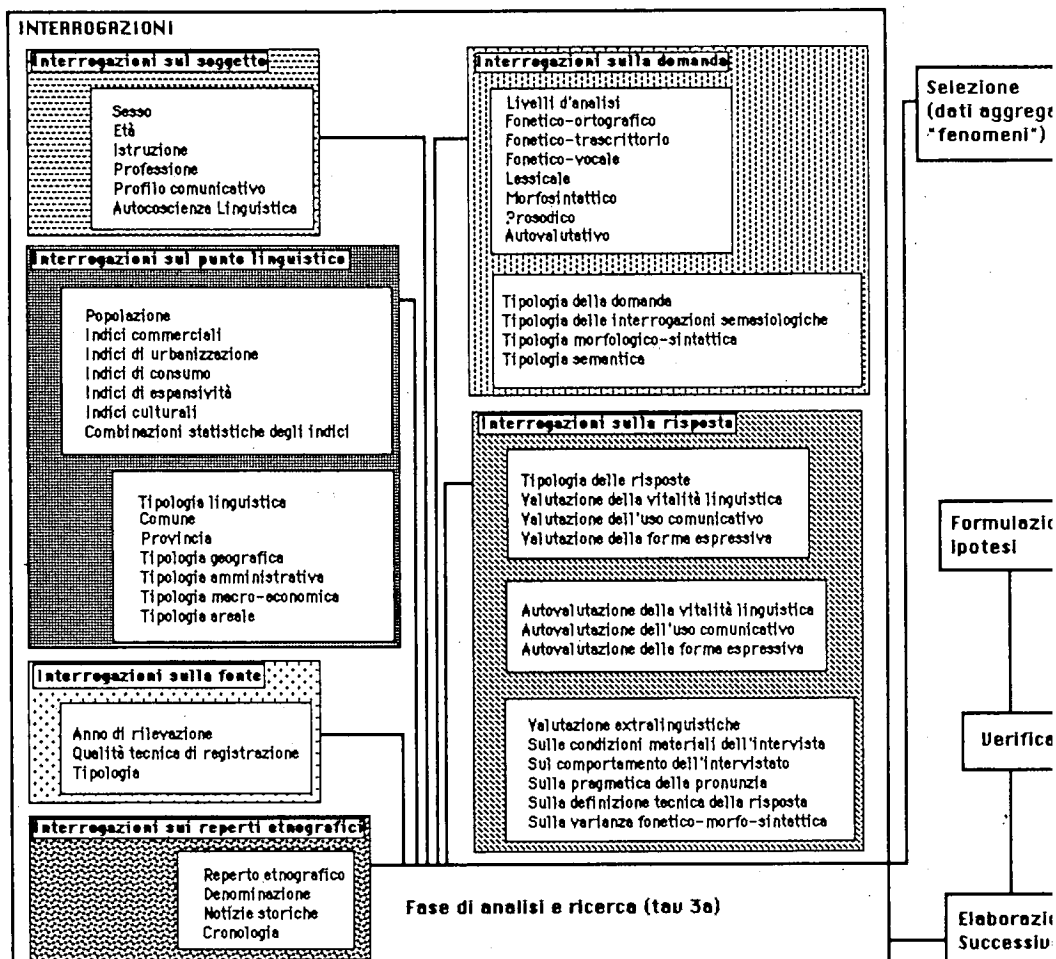
Nel sistema di interrogazioni e rappresentazioni l'obiettivo primario resta quello di ottenere sia in input che in output l'avverarsi di quella integrazione pluridimensionale auspicata dalla piattaforma teorica dell'ALS.

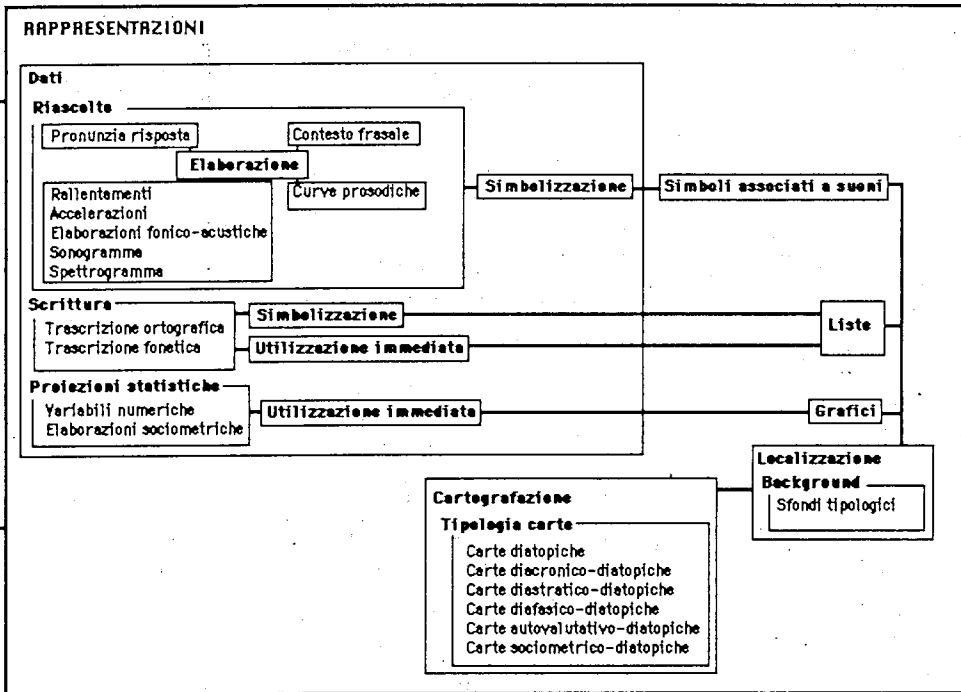
Ciò significherà nella pratica che qualsiasi *frame* di archiviazione organizzata delle informazioni può generare un criterio di ricerca o autonomo o connesso con un altro (o più d'uno). La selezione in input costituisce, a questo punto, il dato aggregato, o il "fenomeno" ricercato in fase di input (di ricerca). Questo si manifesterà inizialmente come un sottoinsieme di schede estratte dal totale che in quel momento contiene la Banca-Dati e potrà essere semplicemente "sfogliato" (per esaminarne gli esiti in trascrizione), oppure "riascoltato" (pronuncia dell'item o del contesto sintagmatico o frasale), o, ancora, trasportato su proiezioni quantitative (statistico-dialettometriche) sotto forma di grafico, o, infine, proiettato su una cartina.

La raffigurazione cartografica in output "trasporterà" con sé i dati relativi ai domini consultati in input, producendo così, anche in uscita, il ritorno delle informazioni collegate. Ciò non costituirà ancora un risultato definitivo. Il fenomeno selezionato sarà sottoposto alla fase di speculazione ipotetica, verificato empiricamente ed eventualmente sottoposto a elaborazioni ulteriori che potranno essere eseguite attraverso la macchina (procedure "dedicate", su cui torneremo in seguito), o attraverso l'osservazione diretta "umana". Solo quando il fenomeno si è ormai delineato in tutta la sua nettezza si potrà passare alla terza fase (vedi lo schema nella tavola n° 3a).

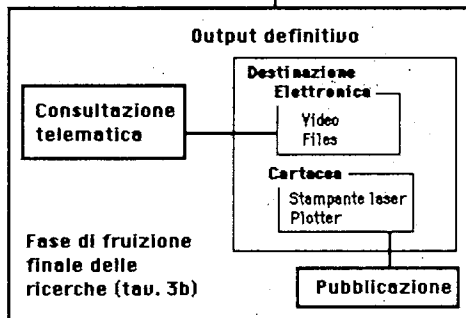
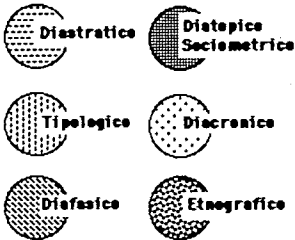
L'illustrazione definitiva dei fenomeni scelti sarà lo scopo specifico della terza fase. In essa il sistema deve essere in grado di produrre manufatti definitivi per la consultazione o la stampa. La consultazione è pensata come uno strumento di informazione telematica, ovvero come struttura informativa disponibile in una rete di collegamento che metta in grado chiunque di usufruire dei dati ALS attraverso un video-terminale. La stampa, invece, resta un problema aperto. Si tratta di produrre un oggetto finale dalle caratteristiche funzionali ed estetiche quanto più sofisticato possibile, e ciò non sempre può essere realizzato con semplici strumenti personali (stampante laser, plötter). È prevedibile, quindi, uno sviluppo a parte di questa fase, sviluppo in cui gli strumenti del *desktop publishing* personale vanno integrati con grandi strumenti tipografici professionali. Si riportano nelle tavole nn° 3b e 4 rispettivamente lo schema dettagliato della terza fase e di tutto il progetto complessivo.



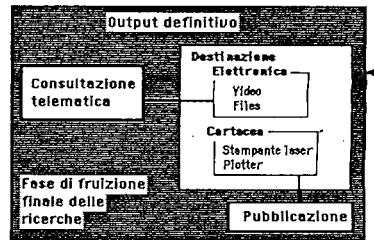
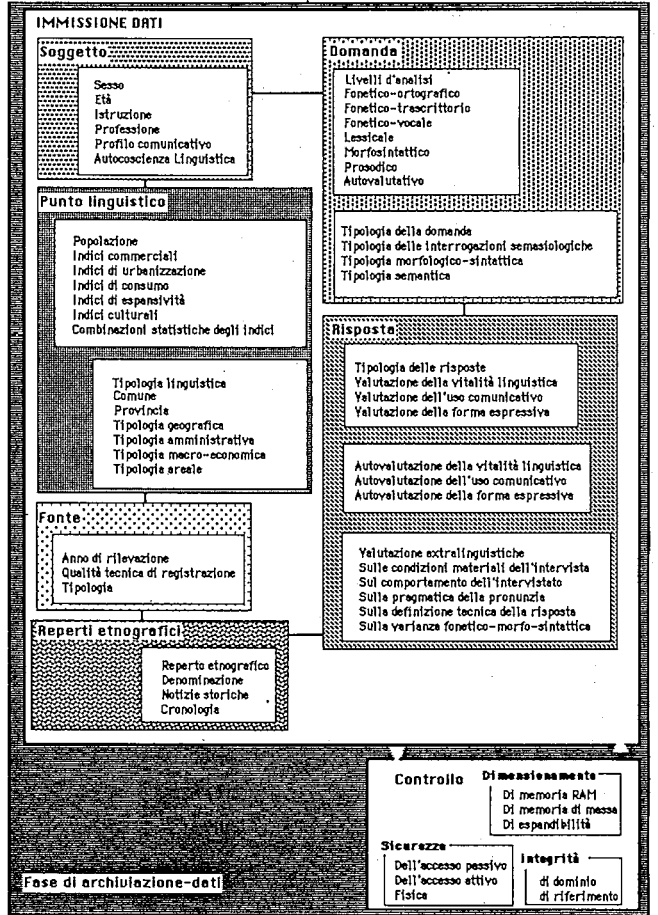


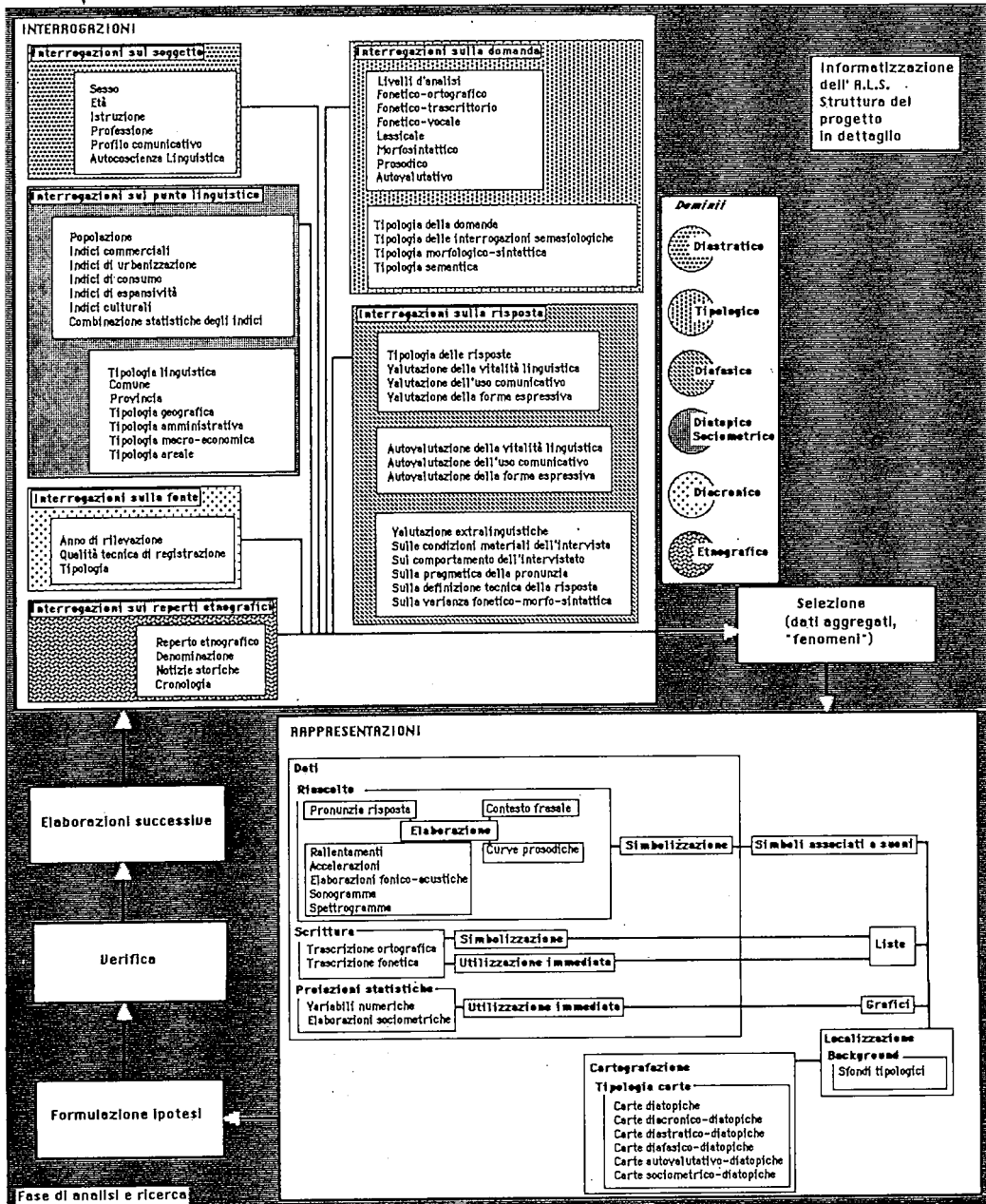


*Dominii*



tau. 4





## 2. *Banche-Dati: problemi, risorse, potenzialità.*

Nella sua accezione più tecnica lo scopo principale del piano di informatizzazione dell'ALS è la realizzazione di una banca-dati centralizzata, ad architettura relazionale, contenente la schedatura completa dei materiali linguistici e sociolinguistici riguardanti la Sicilia in relazione a:

- precedenti rilevazioni geolinguistiche
- materiali non ancora trascritti (indagini sociolinguistiche specifiche, tesi di laurea, rilevazioni parziali, lavori in corso di stampa, etc.);
- le rilevazioni previste dal progetto ALS.

### 2.1. Definizioni preliminari.

**Banca-dati (*Data Base*).** Si intende per Banca-dati (da ora DB) un modello formalizzato di riproduzione (o simulazione) di uno stato reale attraverso la costruzione di un un archivio (o un insieme di archivi) contenente informazioni memorizzate su supporti magnetici (o magneto-ottici) e organizzate in strutture adatte a qualsiasi tipo di operazioni gestionali (ricerca, ordinamento, aggiornamento, cancellazione, produzione di output, etc.) o elaborative (manipolazione dati, elaborazione di strutture primitive in strutture derivate, etc.).

**Archivi (*Files*).** Un archivio è un insieme di schede (records) coordinate e gestite da opportuni formati di input, elaborazione ed output. In una DB complessa è quasi sempre necessario ricorrere a più archivi collegati tra loro tramite un modulo di coordinamento generale. È molto importante considerare la problematica della formazione degli archivi in due fasi distinte. La prima nella quale si definiscono le caratteristiche specifiche dell'autonomia, funzionalità ed esaustività dei singoli archivi, e la seconda che ha per unico oggetto di studio le problematiche della connessione tra archivi diversi (su questo concetto torneremo nel §. 2.2.).

**Schede (*Records*).** Una scheda raccoglie tutti i campi previsti per fornire un'informazione precisa ed esauriente sul singolo soggetto di analisi. Le informazioni richieste da una scheda sono permanentemente memorizzate nell'archivio e serviranno da base per tutte le operazioni di gestione della DB. Una scheda può essere rimossa in qualsiasi momento dall'archivio ed il sistema gestionale provvederà alle cancellazioni e alla riorganizzazione delle informazioni complessive di tutto l'archivio. Sebbene la scheda fisica contenga sempre tutti i campi dell'archivio, la visualizzazione (logica) dei campi necessari ad una data operazione dipende sempre dal formato di visualizzazione della scheda (e possono esistere molti formati diversi per ogni scheda). La struttura della scheda può essere modificata anche nel corso di immisione dei dati nel tempo. Tuttavia tutte le operazioni di modifica e ag-

giornamento della struttura dei records sono soggetti a forti restrizioni per salvaguardare la sicurezza e l'integrità della DB (cfr. §. 2.3.).

**Campi (*Fields*).** Un campo è un contenitore singolo di entità d'informazione condizionato da vincoli di immissione e rigorosamente tipologizzato al fine di ottimizzarne gli usi gestionali e/o elaborativi. La tipologia dei campi - cioè la dichiarazione di specializzazione funzionale del dato immesso - varia da soggetto a soggetto. Generalmente qualunque applicazione che produca DB prevede almeno un tipo alfanumerico, numerico (in svariate formattazioni), calcolato, testuale, booleano, grafico. Nel caso dei sistemi tecnologicamente più avanzati un campo può contenere anche delle *resources* di tipo fotografico o sonoro (in entrambi i casi sotto forma di digitalizzazione di segnali analogici, cfr. §. 3.3.4.). Qualsiasi tipo di campo prevede una serie di attributi obbligatori e/o opzionali che determinano i domini di controllo dell'input, le compatibilità dei dati per particolari tipi di elaborazione, la congruenza semantico-sintattica (cioè la possibilità stessa di "esistenza" all'interno di un sistema di regole altamente formalizzato). Gli attributi dei campi non vanno confusi con la dichiarazione dei tipi di campo. Tratteremo la funzione dei tipi e degli attributi nel §. 2.3.

**Formato (*Format / Template*).** Si intende con formato una maschera variabile di insiemi (o selezioni) di campi, oggetti procedurali, strumenti grafici, e della loro architettura dispositiva. Come abbiamo già detto è sempre possibile avere più formati per un intero insieme o sottoinsieme di campi, ma la generazione di un formato deve essere sempre giustificato dalla necessità di realizzare una particolare operazione gestionale o elaborativa del DB. Approfondiremo la nozione di formato nel quadro più generale dei problemi di funzionalità ed economicità dell'interfaccia sistema-utente nel §. 2.3.

**Input.** Il termine input (all'interno di un DB) designa tutto ciò che concerne i formati relativi all'immissione dei dati all'interno dei campi. I dispositivi di ingresso dei dati possono avere diverse fonti produttive: digitazioni manuali, letture automatiche attraverso periferiche e applicativi dedicati, digitalizzazione di sequenze analogiche come fotografie, immagini, illustrazioni o suoni, elaborazioni di output primitivi come input per la formazione di nuovi classi di risultati, etc.

**Output.** Anche per l'output si può parlare di una classe di formati "specializzata". La funzione specializzata è in questo caso l'ottimizzazione, la produzione e la presentazione dei processi elaborativi. L'output può essere riversato su video, su periferiche di uscita grafica (stampanti, plotter), di comunicazione (modem) o sonoro-vocaliche (canali HI-FI, casse acustiche, amplificatori, piastre di registrazione, etc.). Qualsiasi output, infine, può essere riversato su files interni o esterni al DB. Anche in questo caso un output può costituire l'input primitivo per processi di elaborazione di nuovi risultati (derivati).

**Elaborazione (*Process / Data Processing*).** La parola designa l'insieme delle procedure che applicano ai dati di input le manipolazioni previste da un insieme di *script* programmatici, cioè da una serie di istruzioni formulate in un apposito linguaggio di programmazione. Il risultato di un'elaborazione è sempre un output che produce nuove informazioni o ristrutturazioni di informazioni precedentemente immesse. Il processo di elaborazione che traduce classi di input in classi di output va considerato come un insieme di stati successivi, un processo continuo il cui grado di adeguatezza esplicativa (quindi di finezza conoscitiva) aumenta in relazione al numero dei passaggi. L'accrescersi dell'adeguatezza esplicativa definisce in ultima analisi il grado di veridicità del modello formalizzato di uno stato reale. In altre parole: il numero e la complessità dei passaggi fornisce un sempre maggiore avvicinamento della simulazione alla realtà effettiva (ammesso che esista). La potenza di un DB può essere sempre misurata dalla quantità e qualità dei processi elaborativi che è capace di generare.

## 2.2. Relazionalità.

La nascita del concetto di relazionalità nell'ambito del dibattito teorico sui DBMS si deve principalmente alle ricerche di Codd (1970, 1972, 1974, 1987). Da tali lavori emerge in modo chiaro la necessità di ottimizzare i sistemi di gestione delle informazioni eliminando tutte le ridondanze interne e concentrandosi, di conseguenza, non più sulla fase di immissione dei dati, quanto sulla progettazione delle procedure di elaborazione.

A questo scopo Codd escogitò la progettazione di sistemi relazionali che sostituissero ai dati fisici delle relazioni logiche tra valori di archivi diversi. Un esempio potrà chiarire l'importanza delle innovazioni introdotte da Codd nell'ambito della progettazione dei DBMS.

Supponiamo che si debba costruire un DB che centralizzi la gestione complessiva di una grossa biblioteca universitaria comprendente, oltre al modulo generale delle classificazioni e collocazioni, un archivio per l'ordinamento dei libri, un altro per la gestione del prestito, un altro ancora per le pratiche amministrative. Ognuno di questi archivi sarà formato da un'insieme di campi alfanumerici (es., Autore, Titolo, Editore, etc.) e numerici (es. Anno, Numero di pagine, Numero di volumi, Prezzo, etc.). L'insieme di questi campi sarà soggetto a operazioni di gestione (ricerca per autore, per prezzo, per editore, ordinamento per anno, città, titolo, etc.) e/o di elaborazione (ad esempio applicare a tutti i prezzi una variazione dovuta all'aumento dell'IVA; quindi programmare uno script che estragga il valore numerico del prezzo, lo moltiplichi per una variabile che contenga la percentuale prescritta dalla legge per l'IVA corrente, riscrivere su tutte le schede il nuovo prezzo auto-

maticamente). Per realizzare tutte queste operazioni occorrerà operare separatamente su tutti gli archivi (applicare per es. la ricerca e il ricalcolo dell'IVA prima sull'archivio delle classificazioni, poi su quello delle schede di ordinazione, poi su quello amministrativo, etc.). Ciò naturalmente porterà ad un aggravio di lavoro e al rischio di commettere errori.

In un'architettura relazionale dei DB si parte da un principio-chiave: pur conservando ogni archivio un'autonoma vita propria (e quindi proprie possibilità di modificazioni strutturali), esiste sempre un insieme di informazioni comuni a tutti gli archivi; sarà pertanto sempre possibile realizzare le operazioni richieste attraverso una funzione comune applicata ai dati condivisi. Come realizzare ciò? La risposta di Codd e dei suoi collaboratori (i cui lavori sono tuttora in fase di elaborazione di un modello standard definitivo) fu quella di concepire un modulo unitario che "apra" dei canali di comunicazione tra archivi diversi considerando poi il tutto una struttura unitaria.

Per es., i campi Autore, Titolo, Anno, Editore saranno uguali sia nell'Archivio delle classificazioni che in quello amministrativo o in quello di prestito. Basterà quindi "prelevare" il dato da un sottoarchivio comune per riempire automaticamente tali campi a partire da qualsiasi modulo-Archivio si voglia. Tale sottoarchivio comune sarà caratterizzato da un codice unitario evocabile da uno qualsiasi degli archivi interessati.

Nella pratica accadrà che ogni qualvolta si compili una scheda globale, digitando il codice comune, i campi di tutti gli archivi saranno riempiti con le informazioni in distribuzione condivisa.

I vantaggi di questa impostazione sono numerosi. Innanzitutto si risparmierà tempo e si acquisirà un maggior controllo degli errori e una più generale rapidità di tutte le operazioni sul DB nel suo complesso. In secondo luogo potrà essere sempre attuata una modifica degli *archivi-schiavi* (cioè dei sotto-archivi di un modello relazionale) senza ricorrere all'aggiornamento di tutte le schede dell'archivio centrale che si "automodificherà" senza alcun intervento umano. In terzo luogo sarà facile elaborare procedure "creative" sempre più complesse aumentando la potenza del DB. Infine si favorirà una maggiore "specializzazione" dei singoli archivi che, non essendo gravati dai balzelli della ripetitività, potranno liberare spazi di memoria enormi riservandola a informazioni più dettagliate (maggior numero di campi).

A questi vantaggi si aggiungono (tenendo conto degli ultimi sviluppi dell'informatica condivisa, quindi dei sistemi multi-utente) quelli derivanti dalla facilitazione degli accessi multipli ad uno stesso DB.

Rendendo, infine, il concetto di relazionalità un oggetto di studio a sé stante sarà possibile immaginare sempre nuove utilizzazioni speculative del DB.



### 2.3. Integrità, Sicurezza, Dimensionamento, Espandibilità, Fruibilità

Una componente fondamentale della progettazione dei DB, molto spesso trascurato dalla microinformatica applicata a utilizzazioni non strettamente economico-finanziarie, è quel complesso di risposte inerenti non tanto la problematica logica della base-dati quanto la sua organizzazione funzionale. Passiamo in rassegna i problemi principali.

**Integrità.** In tale ambito ricadono i problemi relativi alla possibilità di evitare errori gestionali/procedurali, o almeno limitare - ove essi avvengano comunque - le loro conseguenze e la loro espansione a macchia d'olio. Per tale motivo occorre programmare una serie di procedure di controllo del dato immesso o di recupero e isolamento dei dati "malati". In tale ambito si possono distinguere due generi di regole-base:

- a) regole di integrità di dominio
- b) regole di integrità di riferimento.

La prima categoria comprende tutti i problemi che scaturiscono dalla gestione del tipo e dell'attributo del campo (cfr. §.2.1.) considerato a sé stante. Esempi tipici sono costituiti:

a.1) dall'obbligatorietà del tipo di dato e dal controllo della sua funzionalità gestionale/procedurale (es. un tipo numerico può essere ordinato per cardinalità ma non alfabeticamente; viceversa un tipo alfanumerico può essere ordinato solo alfabeticamente; se un dato che contiene numeri viene posto in un campo di tipo alfanumerico il suo ordinamento risulterà alfabético con conseguenze disastrose per il DB);

a.2) dal controllo dell'intervallo dei valori assumibili (cioè dalla dichiarazione del *range* di valori legali). Nell'attributo del campo andrà quindi specificato l'arco dell'escursione prevedibile (da...a...) e non saranno ammessi valori diversi (se non su richiesta esplicita);

a.3) dal controllo dell'unicità del dato (nei campi a scelta binaria, ad esempio, si può fissare un attributo che impedisca una scelta diversa da "campo vuoto" / "campo pieno"; nei campi contenenti i codici di relazione si eviterà di selezionare più di una scheda di riferimento per record, etc.);

a.4) dalla restrizione dei valori di input a quelli presenti in una lista (più o meno lunga) prefissata e/o modificabile;

a.5) dal controllo della "completezza" della scheda nei suoi campi contrassegnati da dichiarazioni di obbligatorietà. In altri termini è essenziale che alcuni campi (quelli su cui si dovrà operare una ricerca esaustiva all'interno del corpus) siano vincolati dall'obbligo del riempimento tramite procedure che arrestino l'accettazione di immissione della scheda se essa risulterà incompleta all'analisi finale di "legalità".

Questa precauzione permetterà l'estrazione di output statisticamente validi assicurando sempre un universo di riferimento costante delle percentuali tratte dall'interrogazione del DB.

La seconda categoria considera, invece, i problemi che possono scaturire dal controllo della congruità dei valori distribuiti in più archivi o, in altri termini, dai problemi posti dalla compatibilità dei valori messi in relazione tra loro. Tra questi i più importanti sono:

b.1) il controllo del riferimento comune ad un dato condiviso (o che si crede tale). Si tratta di un controllo di fondamentale importanza poiché se viene evocato un codice comune a cui non corrisponde, in uno degli archivi interessati dal legame di relazione, un valore uguale, si immetteranno automaticamente valori errati. In questo caso è essenziale la costruzione di procedure che avvertano l'utente dell'assenza del dato che sarebbe stato "chiamato" da un archivio (in quel momento principale) rispetto ad un altro (in quel momento ausiliario). Una procedura ancor più accorta dovrebbe rendere possibile la correzione automatica di questo problema, ciò, tuttavia, non è sempre realizzabile.

b.2) il controllo del tipo di dato messo in relazione. Il "canale" di comunicazione attivato dalla dichiarazione di relazione può essere aperto solamente se il campo comune è di tipo uguale (in alcune applicazioni anche di tipo "compatibile"). Generalmente il tipo di dato comune deve essere accompagnato anche dalla dichiarazione di un attributo di indicizzazione di tutti e due i campi connessi. Per indicizzazione si intende un campo il cui ordinamento deve essere reso particolarmente veloce ricorrendo ad un algoritmo che non ordina i dati "fisici" ma un loro "puntatore" logico (generalmente un numero comune). Quest'ultimo tipo di precisazione si rende indispensabile in quei DB, come ad esempio sarà l'ALS, che operano su una mole massiccia di dati, rendendo più rapido il sistema delle relazioni. In tali tipi di DB la velocità di esecuzione delle procedure di relazione - generalmente "invisibili" all'utente - non può essere considerato un dato accessorio ma diventa essenziale al funzionamento della banca-dati nel suo complesso.

b.3) il controllo della completa relazionalità delle procedure applicative. In altri termini un DB può considerarsi propriamente relazionale quando una richiesta di modificazione dovuta all'intervento di una procedura coinvolga *tutti gli archivi connessi* e non solo un suo sottoinsieme. A questo proposito vanno costruite una serie di script programmatici che garantiscano il legame di ritorno e la sua biunivocità. Un esempio, già fatto, può essere costituito dalla modifica dell'incidenza dell'IVA, relativamente al caso della biblioteca universitaria, oppure, sempre, nello stesso ambito, dalla correzione di un nome di autore o titolo errati. Se un DB è veramente relazionale, la

modifica di questi dati dovrà coinvolgere tutte le schede di tutti gli archivi che condividono, in un modo o nell'altro, il campo interessato.

**Sicurezza.** Oltre al controllo della "legalità grammaticale" dei dati immessi (integrità) una buona progettazione di DB deve poter esercitare un controllo sugli accessi possibili e sulla sicurezza dei dati utilizzati per le proiezioni esterne. Questo genere di problemi possono essere riassunti in:

- c) problemi di pertinenza dell'accesso attivo
- d) problemi di pertinenza dell'accesso passivo
- e) problemi inerenti la sicurezza fisica (hardware e software) dei dati.

Al momento della costituzione dell'organigramma complessivo del personale addetto alla gestione e manutenzione di un DB (accesso attivo) è indispensabile fissare limiti precisi nei compiti e nelle funzioni svolte. Una buona programmazione deve saper prevedere eventi futuri con una certa approssimazione alle possibilità reali. Un quadro completo di tali possibilità dovrebbe considerare:

c.1) la competenza totale. Un DB è, come abbiamo visto, un complesso di gestione dati e gestione procedure. Nel caso in cui, per una qualunque ragione, venisse meno l'apporto di chi gestisce le procedure potrebbe essere impossibile la correzione dei dati nella loro generalità, la ristrutturazione di tutto l'insieme degli archivi e delle loro relazioni, la programmazione (questa mai definitiva) degli output producibili. È quindi indispensabile prevedere la formazione di un soggetto che abbia la competenza totale su tutta la banca-dati. All'interno dell'organigramma devono esistere sempre almeno due responsabili della programmazione.

c.2) la competenza esecutiva. Gli addetti all'immissione manuale o automatica dei dati vanno formati in modo differenziato. Una competenza generalizzata sulle modalità di input deve essere comune a tutti. Ogni archivio, tuttavia, richiede una sua specializzazione di immissione (ad esempio un archivio di fonti iconografiche ha procedure di immissione tutte sue proprie ed una routine rigorosamente dedicata) e, di conseguenza, una formazione specifica degli operatori incaricati di gestirlo.

c.3) oltre alla formazione dei diversi livelli del personale addetto al DB occorrerà fissare una gerarchia obbligatoria dei livelli di accesso. Ciò è possibile tramite un sistema di *passwords* differenziate per livello e funzione, onde impedire manipolazioni indesiderate da personale non autorizzato.

d.1) oltre agli operatori può accedere al DB qualsiasi utente autorizzato. Gli utenti passivi devono poter limitare il loro intervento alla sola consultazione e alla produzione di eventuali output. Anche in questo caso va prevista una password di primo livello che disabilita qualsiasi procedura di immissione o, peggio, programmazione del DB.

È opportuno, infine, prevedere un'organizzazione decentrata del lavoro di immissione materiale attraverso la costituzione di nuclei di operatori in posti diversi con un sistema di centralizzazione periodica dei dati immessi.

Una classe di problemi particolarissimi è quella relativa alla sicurezza fisica dei dati. Essa può essere minacciata da:

- e.1) danni all'hardware
- e.2) danni al software
- e.3) intervento doloso sul sistema.

I danni all'hardware possono essere previsti con una certa precisione. I guasti alle unità centrali non possono influenzare la struttura del DB. I più dannosi potrebbero essere i guasti alle periferiche di output e, in particolare, alle memorie di massa in cui sono immagazzinati i dati. A questo tipo di danno - che non riguarda il software di gestione delle memorie di massa ma proprio il dispositivo in sé - si può reagire solamente se si è previsto un adeguato *refresh* periodico dei dati e un regolare *backup* su altri supporti magnetici o magneto-ottici. Una radicale precauzione consiste nel caricare tutta la banca dati su almeno due supporti di massa paralleli collocati in sedi differenti.

I danni al software possono riguardare il funzionamento dei programmi o la distruzione parziale o totale dei documenti di descrizione delle *directory* (o *desktop* nei sistemi MacIntosh, come quello utilizzato dall'ALS) delle memorie di massa. Per il funzionamento dei programmi occorrerà prevedere una fase di collaudo preventivo dettagliato (discuteremo di questo problema più avanti parlando del dimensionamento dei DB) ed una costante manutenzione dell'impianto software. Eventuali errori non previsti potranno essere corretti solo dai creatori degli script di programmazione che dovranno prevedere comunque una possibilità di operazione chirurgica a "cuore aperto", ovvero prevedere possibilità di modificazioni anche a sistema avviato.

Quanto alla distruzione dei documenti di descrizione delle *directory* (o *desktop*) delle memorie di massa vanno approntati dei *tools* - peraltro già esistenti - di manutenzione di tale tipo di software e di recupero "disperato" dei dati grezzi. Si tratta di un livello di intervento a cui non si dovrebbe mai arrivare e che può essere evitato con il sistema di prevenzione già descritto parlando dei danni hardware.

Infine due parole sui danni dolosi che possono essere di due tipi: da virus o da shock fisici deliberati. Sui secondi c'è poco da dire, tranne che la dislocazione su sedi differenti e la "duplicazione" dei sistemi riesce a recuperare anche questa (rara) evenienza. Consigliabile sarebbe poi la "crittografazione" degli archivi per evitare letture indesiderate.

Sulla questione dei virus occorre invece predisporre soprattutto una rete di sistemi preventivi software che blocchino l'insorgere della "malattia". Tali software diagnostici devono operare sempre in linea,

cioè agire mentre gli operatori lavorano al DB. In alcuni casi bisogna prevedere l'arresto immediato del sistema ed un ricondizionamento da affidare ai programmatori. Poiché, tuttavia, è sempre possibile che vengano prodotti virus nuovi e non-diagnosticabili, una norma ferrea dovrà evitare l'immissione di *floppy-disk* o applicazioni o collegamenti in rete che non siano preventivamente e rigorosamente controllati.

**Dimensionamento.** Il dimensionamento di un DB è uno degli argomenti più importanti nella progettazione avanzata di banche-dati. Alle dimensioni di base concorrono alcuni parametri fondamentali:

- a) il numero degli archivi
- b) il numero di schede per archivio
- c) il numero dei campi per ogni scheda
- d) il numero dei formati di input e output
- e) il numero delle procedure e la loro dimensione
- f) il numero delle relazioni tra gli archivi
- g) il numero delle variabili delle procedure

Fattori di grave accrescimento delle dimensioni sono:

- h) la struttura complessiva dell'interfaccia-utente
- i) il tipo di campo
- l) gli attributi dei campi

Per tutto ciò che concerne i punti a-g non è possibile alcun intervento, se non quello di limitare all'indispensabile la progettazione generale. Tuttavia sarebbe un errore progettuale rinunciare a campi su cui poi sorgerebbe la necessità di elaborazioni e ricerche. Lo stesso discorso vale per formati e procedure, dovendo comunque il DB assicurare la massima funzionalità e sicurezza nel recupero dei dati e nella loro leggibilità.

Il settore di maggiore intervento per limitare la dimensione dello archivio è quello della accorta previsione delle reali necessità riguardanti i campi e l'interfaccia utente.

L'aggravio di carico memoriale aumenta, infatti, con l'attribuzione di particolari tipologie ai campi. I campi numerici e booleiani riducono gli spazi di memoria al minimo; i campi alfanumerici e, soprattutto, testuali, richiedono già ampi spazi di memoria, poiché ogni singolo formato testuale o alfanumerico comprenderà un certo numero di caratteri deciso a priori, sia esso realmente saturo, sia esso vuoto nella scheda-record; infine i campi grafici o foto-grafici, quelli sonoro-vocalici e, comunque, tutti i campi che richiamano informazioni immagazzinate sotto forma di *resources* riducono drasticamente gli spazi di memoria, e il loro uso rende problematica la manipolazione di un grosso numero di schede.

Per contro l'uso di campi testuali, alfanumerici e, a volte, grafici determina una migliore leggibilità complessiva della base-dati. I campi che richiamano informazioni sonoro-vocaliche immagazzinate come *resources* possono risultare di importanza fondamentale per la corretta

valutazione della trascrizione fonetica, della prosodia, del contesto fonetico-sintattico in cui è stato prodotto un dato item.

Occorrerà quindi operare nella fase di dimensionamento tenendo conto dei fattori di equilibrio tra leggibilità e spazio occupato. È comunque escluso che un DB di enormi dimensioni come - vedremo - si prospetta quello dell' ALS, possa ricorrere in maniera massiccia ai tipi di campo più dispendiosi.

Anche in fatto di attributi dei campi il dimensionamento dovrà puntare all'essenziale. In particolare gravano sulla memoria RAM i campi con attributo di indicizzazione (cfr. §. 2.3. Integrità) che rende, tuttavia, molto più veloce una qualsiasi ricerca.

Un ottimo stratagemma per aggirare gli ostacoli relativi al tipo e all'attributo di ogni singolo campo e, inoltre, ad accrescere la sicurezza del DB limitando al massimo gli errori, è quello di creare dei campi numerici (ove sia possibile) aumentando le loro possibilità di facile lettura col ricorso alle liste pre-costituite, che possono anche fungere da legenda.

Questo tipo di soluzione accresce anche una facile interfacciabilità tra i DB e l'utente (su questo punto vedi più avanti il sottoparagrafo riguardante la fruibilità).

**Espandibilità.** Una banca-dati, a rigore, non può dirsi mai completata. È per sua natura una struttura dinamica sia in fase di immissione dei dati, sia in fase di consultazione ed emissione di output. Diversamente dagli archivi cartacei, un archivio elettronico può fornire dati in qualsiasi momento lo si interroghi, rapidamente ed accuratamente. È sempre possibile da un DB ottenere un output che permetta di fissare lo stato dei lavori, un continuo *working in progress* che rende possibile l'individuazione di tendenze, di errori, di fenomeni che possono, ad un certo punto, apparire talmente importanti da riconsiderare i programmi iniziali e modificarli.

Naturalmente tutto ciò sarà reso possibile dalla capacità di progettare un'architettura aperta a nuove possibilità e soluzioni hardware e software, una struttura flessibile nella dichiarazione di campi, tipi, attributi, procedure e relazioni che è sempre possibile trasformare senza perdita di informazioni già faticosamente acquisite.

A questo scopo vanno considerati almeno due ordini di problemi:

- a) l'espandibilità della memoria RAM
- b) l'espandibilità delle memorie di massa
- c) la trasportabilità dei dati in ambienti operativi diversi da quello di partenza.

La memoria RAM è un area di memoria volatile e operativa che permette di trattenere nelle "stanze" dei calcoli, ovvero, nei luoghi dedicati alla elaborazione, tutte le informazioni che necessitano di un trattamento procedurale, per tutto il tempo necessario a portare a termine tali operazioni. Non è detto che ciò sia sempre possibile, anche

partendo da una dotazione iniziale di RAM che è stata stimata sufficiente in fase di progettazione. La memoria RAM è incrementabile tramite delle schede interne alle unità centrali: la capacità massima varia da sistema a sistema. In ultima analisi una progettazione lungimirante deve rendere sempre possibile il passaggio da un sistema ad un altro più potente.

Le memorie di massa (di cui abbiamo parlato già in relazione ai problemi della sicurezza) sono costituiti da tutte quelle periferiche di output che immagazzinano permanentemente i dati e li conservano nella loro forma finale. In attesa (prossima) di sistemi di memorizzazione magneto-ottici, o ottici a lettura e scrittura, i sistemi attuali prevedono l'utilizzazioni di capienti hard-disk a dimensione variabile. Il problema della sopravvenuta insufficienza degli hard-disk è il più facilmente risolvibile all'interno di uno stesso sistema. Il giorno in cui un hard-disk da 100, 200, 300 MB si riveli insufficiente è sempre possibile trasferire in modo semplicissimo il DB in un Hard-disk più potente (sino a 1.2 GB ed oltre). Gli unici problemi seri che riguardano le memorie di massa sono di tipo economico. È sempre necessario, infatti, disporre di almeno due periferiche di massa uguali per rendere possibile il backup dinamico dei dati e la loro certa salvaguardia (cfr. § 2.3. Sicurezza). Soluzioni alternative prevedono una memoria di massa dinamica (hard-disk) accompagnata da un'unità di backup a nastro (lenta e non in-linea, ovvero da caricare su hard-disk). Quest'ultima soluzione, più economica, non sembra oggi godere di molta popolarità.

Quanto al terzo problema, quello della trasportabilità dei dati in ambienti diversi, occorre distinguere tra:

- a) trasportabilità del dato grezzo leggibile da software più potenti all'interno dello stesso sistema;
- b) trasportabilità del dato grezzo in applicativi e sistemi diversi ed esterni a quello di partenza.

Il punto (a) prelude in sostanza alla possibilità di trasferire i dati in pacchetti software che all'inizio della ricerca non erano ancora stati realizzati, oppure in reeleases più avanzate degli stessi programmi.

Si tratta di una precauzione minima poiché lo stato di ricerca e avanzamento nel campo del software per DB è continuo, impetuoso e inarrestabile. Non sarebbe neanche lontanamente immaginabile una banca-dati le cui informazioni fossero leggibili solamente da un particolare generatore di DB o, peggio, una particolare versione di generatori di DB.

Il punto (b) riguarda invece il radicale cambio di sistemi che può intervenire nel caso di eccessiva crescita del DB. Qui la situazione può farsi complicata poiché i diversi sistemi operativi e, soprattutto, le diverse classi di macchina (personal, mini, micro, mainframe) possono adottare formati incompatibili. Al momento della progettazione, quindi, sarà indispensabile tracciare una mappa della traducibilità reale

e possibile dei diversi sistemi nei diversi ambienti operativi, scegliendo, tra l'altro, macchine iniziali capaci di funzionare con sistemi operativi diversi (MS DOS, OS2 e 3, System MAC, UNIX, XENIX, etc.).

Come si vede i rischi che si possono correre nel progettare il dimensionamento dei dati sono notevoli. Siamo probabilmente al problema più importante per la sopravvivenza stessa del DB. È per questo motivo che in fase progettuale va prevista una simulazione di caricamento del DB con dati-civetta per testare la resistenza del sistema hardware e software nel suo complesso e stabilire i punti di rottura (se ce ne saranno) in modo da calcolare quanto tempo sarà possibile lavorare con i materiali a disposizione.

**Fruibilità.** Con questo sgraziato termine indichiamo qui l'ultimo requisito che dovrebbe avere un DB per risultare non solo efficiente e potente ma anche facile da usare. Si può dire che tutta la problematica della fruibilità è riassumibile in due punti:

- a) la realizzazione di un'interfaccia-utente "amichevole"
- b) la possibilità di distribuzione multimediale degli output.

L'interfaccia-utente è costituita da tutto il complesso di strumenti hardware e software che stabiliscono i contatti e determinano la comunicazione tra l'uomo e la macchina o il sistema di organizzazione dei dati. In questi ultimi anni i maggiori centri di ricerca mondiale nel campo della microinformatica distribuita hanno molto lavorato in questo campo di studi. Negli anni ottanta sono nati sistemi concepiti proprio con lo scopo di realizzare una facile interazione uomo-macchina. I sistemi Apple MacIntosh, in particolare, hanno realizzato un sistema di comunicazione fondato sulla filosofia iconica, sull'utilizzo di periferiche di input unitarie (*mouse*), su software di gestione con *menù pull-down*, bottoni, campi di edizione, etc. Tali sistemi, inoltre, hanno rivoluzionato le procedure sintattiche invertendo l'originaria procedura azione-oggetto in oggetto-azione. Così facendo sono scomparse le interminabili sequenze di digitazione per l'interrogazione di DB, sostituite tutte dal concetto operativo: "prima seleziona l'oggetto poi seleziona il comando". L'aver impostato rigide specifiche per gli sviluppatori di software MacIntosh ha determinato il sorgere di applicativi altamente standardizzati su questa filosofia comunicativa. Il risultato è che oggi con tali sistemi i tempi per l'apprendimento delle procedure operativo-gestionali sono stati drasticamente ridotti (anche se è aumentato il carico di difficoltà dei programmatori: quanto più un'interfaccia sarà facile da usare, tanto più, dietro, ci sarà un programma complicato e difficile da realizzare). In questo contesto si è verificata una tendenza concorrenziale all'imitazione dei sistemi software MacIntosh che hanno migliorato anche l'interfaccia-utente di sistemi tradizionali (si pensi a Windows che gira su MS-DOS e OS2).



Anche all'interno dei sistemi che utilizzano interfacce-utenti "amichevoli" esistono, comunque, generatori di DB diversamente strutturati e potenti. Una progettazione oculata deve prevedere una fase di test dei principali pacchetti applicativi. Nel caso di DB complessi è giocoforza orientarsi sui generatori di DB relazionali ad alto contenuto iconico e con ampie possibilità di programmazione *object-oriented*, cioè capaci di uniformarsi alle specifiche procedurali richieste da un utente finale non-specializzato.

L'utilizzazione finale è l'ultimo dei problemi qui trattati. Anche in questo caso è molto importante la facilità di uso per chi dovrà poi utilizzare a scopi scientifici le informazioni immagazzinate e organizzate. Non bisogna tuttavia sottovalutare la potenza del sistema, né sacrificarla alla facilità d'uso. Gli output producibili dal sistema usato devono poter essere manipolati sia su video che su stampanti, plotter o reti telematiche. Un'utilizzazione multimediale della fruizione prevederà:

a) la compattazione dei dati su c.d. rom, cioè dischi ottici a lettura laser che sono capaci di trattenere una enorme quantità di informazioni strutturate (leggibili attraverso poco costosi lettori di c.d. rom);

b) la condivisioni in reti telematiche di diverso tipo. A questo proposito la progettazione del DB deve prevedere un'ampia scelta di protocolli di comunicazione, comprendenti quelli dei maggiori network nazionali e internazionali. La compatibilità in rete dovrebbe approssimarsi il più possibile alla fruizione diretta su qualsiasi DBSM possibile. In mancanza di ciò fornire almeno l'opportunità di lettura grezza (ASCII puro) adatta ad una lettura universale tramite programmazione distribuita;

c) un sistema flessibile di estrazioni autonome di dati convertibili in formati numerici, alfanumerici e grafici.

### 3. La banca-dati dell' ALS.

La banca-dati dell'ALS sarà un DB altamente complesso, formato da una serie plurima di archivi connessi relazionalmente tra loro e gestiti da un modulo centralizzato che permetterà il recupero, la manipolazione e la produzione di output su tutti i dati immessi, siano essi relativi alle inchieste già compiute che a quelle in via di completamento o a quelle programmate nel piano di sviluppo.

#### 3.1. Gli archivi, i campi e le relazioni dell'ALS.

Il DB si articolerà su 7 archivi (ognuno dei quali corrisponde ad uno specifico dominio variazionale) per un totale di 82 campi complessivi, di cui due di tipo speciale che considereremo a parte:

a) **Archivio Punti (dominio diatopico)** nel quale saranno conservate tutte le informazioni di tipo geografico, linguistico, sociologico e gli indici sociometrici relativi alle località in cui si sono svolte o si svolgeranno le inchieste, contenente 13 campi;

b) **Archivio Informatori (dominio diastratico)** che memorizzerà tutte le variabili anagrafiche e sociologiche utili per un'analisi dei soggetti intervistati, contenente 6 campi più un campo speciale di tipo sottoarchivio relativo solo alle inchieste con vocazione variazionale che conterrà il questionario metalinguistico;

c) **Archivio Domande (dominio tipologico dei livelli d'analisi)** in cui saranno immagazzinati tutti gli *items* relativi ai questionari già presentati o da presentare, comprese le domande semasiologiche contenute nei diversi questionari, la tipologia delle domande e le aree semantiche di riferimento suddivise in quattro sezioni più un riassuntivo generale per ogni domanda, contenente 10 campi;

d) **Archivio Fonti (dominio diacronico)** che immagazzinerà le specifiche delle diverse indagini, contenente 4 campi;

e) **Archivio Etnografico (dominio etnologico)** che conterrà tutto il materiale iconografico e fotografico derivante dai reperti etnografici raccolti durante il rilevamento, contenente 3 campi;

f) **Archivio Carte (dominio delle rappresentazioni)** che conterrà tutti gli output testuali richiesti sotto forma di liste e tutte le cartografazioni, in trascrizione ortografica o fonetica e/o in proiezione sociometrico-sociolinguistica prodotte a partire da qualsiasi campo o incrocio di campi selezionati sull'insieme del DB, contenente 3 campi;

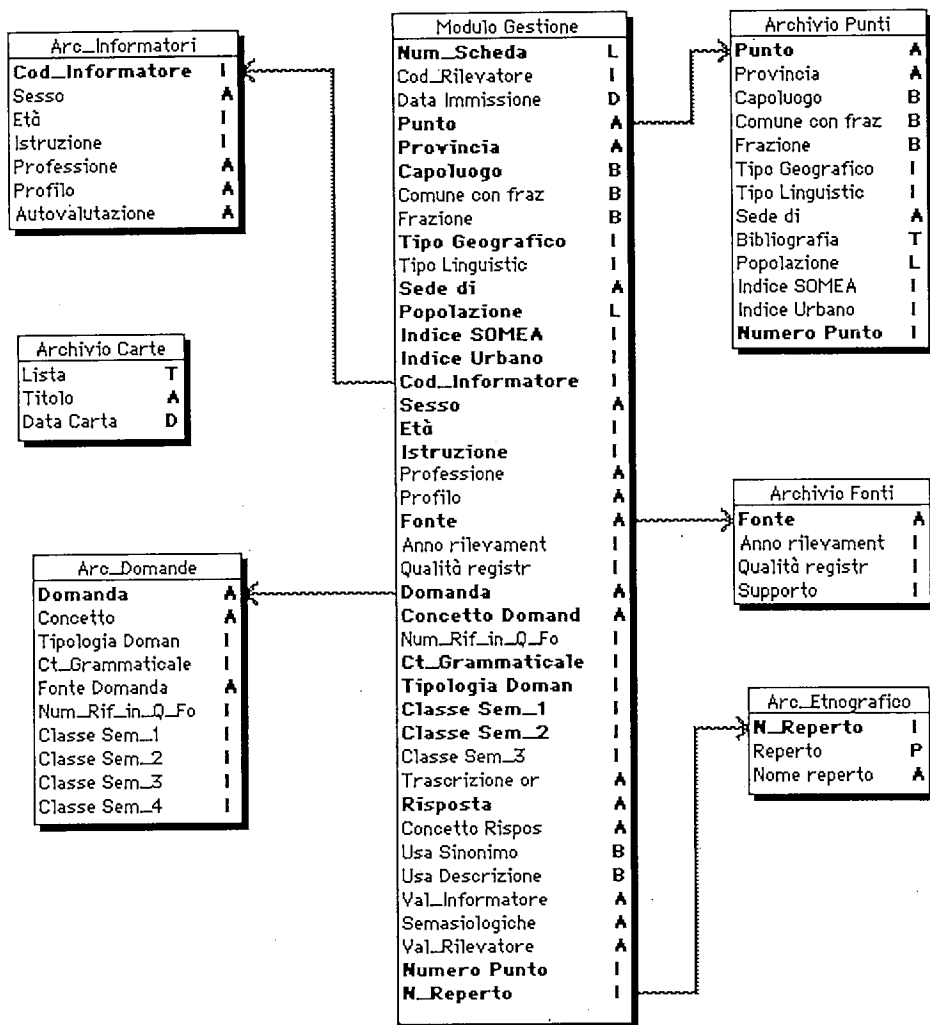
h) **Archivio (Modulo) di Gestione (dominio delle selezioni e delle opzioni diafasiche)** che costituisce lo schema centralizzato e unificante contenente le singole schede: il corpus generale soggetto a interrogazione complessiva e i campi speciali per la visualizzazione dei reperti etnografici collegati, la pronuncia della risposta e del contesto linguistico interessato, di complessivi 41 campi distribuiti in quattro pagine di formato di immissione e consultazione più un formato generale di tipo lista.

### Le relazioni dell'ALS.

Il modulo di gestione sarà connesso direttamente, attraverso alcuni campi specifici (con letture "a variabile"), a cinque archivi. La direzione delle relazioni sono:

- a) Archivio Punti ' Archivio (Modulo) Gestione (tramite il campo alfanumerico PUNTO-PUNTO)
- b) Archivio Informatori ' Archivio (Modulo) Gestione (tramite il campo numerico COD\_INFORMATORE-COD\_INFORMATORE)
- c) Archivio Domande ' Archivio (Modulo) Gestione (tramite il campo alfanumerico DOMANDA-DOMANDA)
- d) Archivio Fonti ' Archivio (Modulo) Gestione (tramite il campo alfanumerico FONTE-FONTE)
- e) Archivio Etnografico ' Archivio (Modulo) Gestione (tramite il campo numerico N\_REPERTO-N\_REPERTO).

Indirettamente il modulo di gestione è connesso anche all'Archivio Carte. Una procedura di esportazione e importazione filtrata trasmette i dati elaborati dal modulo di gestione alle liste di esiti. Le liste di esiti saranno processati da un programma che li considererà vettori a più dimensioni. Ad ogni dimensione del vettore di dati corrisponderà una specifica graficizzazione su cartine di diverso tipo (economiche, sociologiche, sociometriche, geografiche, tipologiche, linguistiche, demografiche, etc. cfr. §. 3.3.3.). I campi di ogni archivio hanno tutti una specifica tipologia e una fitta rete di attributi che garantiscono la piena compatibilità del Modulo di gestione. *Ad ogni modifica degli archivi-schiavi (sotto-archivi) corrisponderà una modifica automatica - senza intervento umano - del modulo centrale di gestione* (nell'Appendice 1 si troverà una schematizzazione completa della struttura degli archivi, dei campi con tutta la tipologia, gli attributi e le loro relazioni. In figura, invece, la struttura nuda dell'ALS).



### 3.2. I formati dell'ALS

Per ogni archivio il DB dell'ALS potrà contare su almeno 2 formati, uno di input ed uno di output.

Le maschere di introduzione dei formati di input sono contraddistinti - per una più facile identificabilità - da strutture e colori diversi.

Tra i tipi di campi scelti alcuni si rivelano particolarmente funzionali come i campi-testo scorrevoli (finestra con ascensore) e quelli booleani (sotto forma di bottoni di opzione validati attraverso un contrassegno con il *mouse*, un dispositivo di input unificato a valore binario). Ecco sotto un esempio di quest'ultimo tipo di campo:

**Capoluogo**

Si

No



I campi che richiedono la formulazione di numeri o stringhe alfanumeriche di rappresentazioni simboliche sono poste sotto forma di una lista di caselle barrabili in cui si visualizza per esteso la domanda, anch'essa validata poi attraverso un contrassegno con il *mouse*. Ecco un esempio di questa simulazione grafica di scelta alfanumerica:

**Sede di**

a = Museo etnografico

b = Fiera

c = Porto

d = Ospedali pubblici

e = Scuole superiori

f = Uffici giudiziari

g = Uffici finanziari

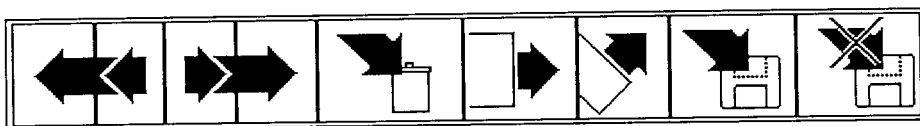
h = Diocesi

z = Risposta nulla



Alcuni campi alfanumerici (come quello in figura) potranno utilizzare l'alfabeto fonetico appositamente realizzato (vedi Appendice 5), sia direttamente che attraverso una traduzione automatica di caratteri standard generata da una speciale procedura di conversione ortografico-fonetica.

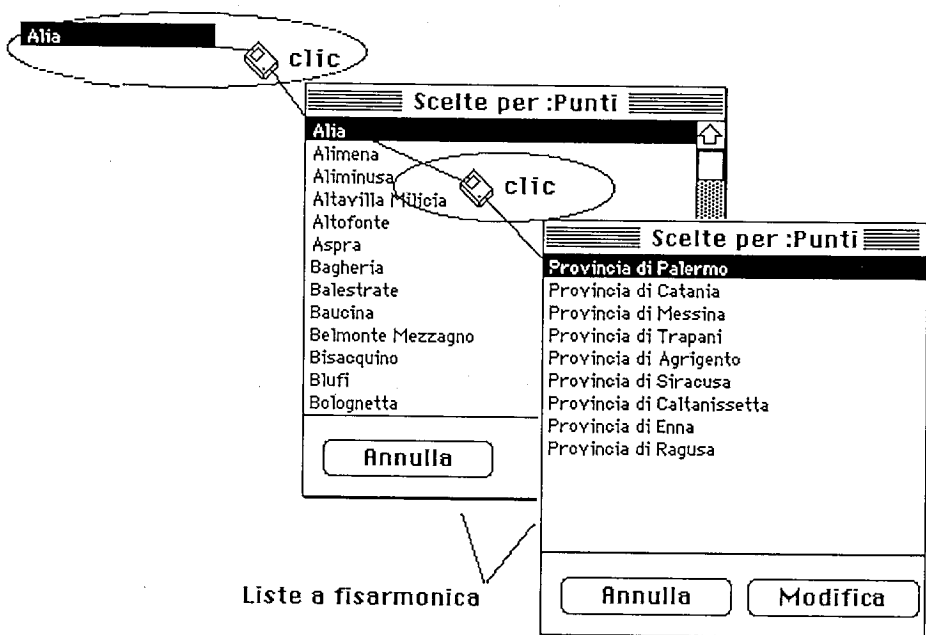
Le procedure standard di gestione della scheda (immissione, rifiuto, validazione o conferma, distruzione, passaggio da una pagina di formato all'altra, scorrimento delle schede singole o dalla testa alla coda dell'archivio e viceversa) sono affidate a *script* programmatici interfacciati all'utente sotto forma di icone dalle funzioni facilmente individuabili.



Tutti i formati di input sono caratterizzati da un'immissione guidata del dato. Quasi tutti i campi sono compilabili attraverso un'operazione di cliccaggio (sempre attraverso il *mouse*) in una lista prefissata e condivisa. Un esempio di tale sistema a liste scorrevoli nell'illustrazione sotto:



Come si può vedere, la lista è fissa ma prevede un intervento manuale tramite il bottone MODIFICA. L'introduzione così guidata è esente da errori e garantisce una ricerca statisticamente affidabile. Inoltre il sistema a liste è supportato da una struttura "a fisarmonica" che guida le scelte complesse a indentazione progressiva. Ad es. se è richiesta l'introduzione di un punto linguistico tra i 406 prescelti si passerà prima attraverso il raggruppamento provinciale e poi attraverso la scelta del comune all'interno della lista di quella singola provincia come è mostrato nell'illustrazione seguente:



I formati di input degli Archivi Etnografico e Carte hanno una struttura diversa dagli altri, dovendo prevedere l'immissione di dati grafici. Per la gestione di questi formati è prevista una routine di immissione complessa, suddivisa in varie fasi:

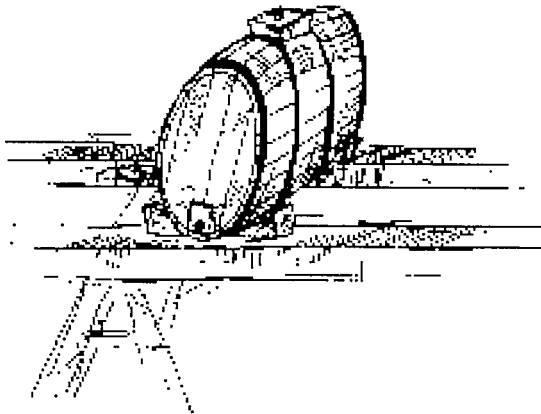
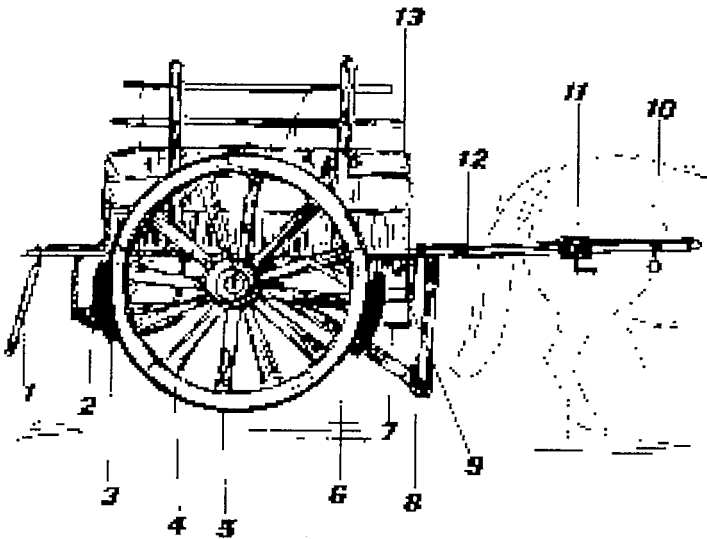
a) scelta del materiale iconografico (che può essere una fotografia, un disegno, un'illustrazione, un grafico, tratti da originali o da libri o altro materiale bibliografico, oppure composti con particolari programmi di disegno libero o tecnico - vedi Appendice 3 contenente tutto il software proposto -);

b) digitalizzazione delle immagini tramite scanner Apple - vedi Appendice 4, hardware della stazione informatica dell'ALS) - a 256 toni di grigio e risoluzione di 300 punti per pollice;

c) formazione di un archivio di incollaggi grezzi su software di conservazione di immagini *on-line*;

d) incollaggio su scheda dell'archivio, nel campo specificato.

Nelle illustrazioni sotto alcuni esempi di immagini-campione digitalizzate:





Per quanto riguarda i formati di output (8 in tutto) è stata scelta un'impostazione colonnare standard per facilitare la consultazione di ampi schedari. Le colonne sono disposte anche esse su finestre scorrevoli e possono gestire i menu di ricerca, ordinamento, inserimento, aggiornamento. Non sono possibili le immissioni o cancellazioni dirette se non su richiesta specifica.

Ai formati di output standard si affiancano tutti quei formati di output delle possibili produzioni dell'ALS: alcuni già disponibili (per esempio la stampa di tabulati e carte), altri in fase di allestimento e altri ancora per il momento neppure prevedibili.

I formati di output potranno tutti essere riversati sulle periferiche in possesso dell'ALS (vedi Appendice 4, hardware della stazione informatica dell'ALS). In particolare su quelle già disponibili:

- a) Stampante Laser in formato A4 Apple SCII
- b) Stampante ad aghi per bozze Imagewriter II
- c) Memorie di massa (Hard-disk Rodime 140,100, 20 MB)

In futuro sarà altamente raccomandabile l'acquisizione di un *plotter* per la stampa accurata di carte a dimensione tabloid o più grandi.

Un elenco completo e una descrizione dei formati più importanti si può vedere nell'Appendice 2 (i formati dell'ALS).

### 3.3. Le procedure dell'ALS.

La gestione di un DB complesso come quello dell'ALS ha richiesto un ingente piano di procedure programmatiche. Alcune di queste sono già state realizzate, altre sono ancora in fase di progettazione o di realizzazione. Elenchiamo di seguito le procedure sinora realizzate.

#### Procedure globali:

- Cartografia
- Creazione Liste
- Variabili Cartine
- Esportazione Concetti
- Esportazioni Risposte
- Esportazione Domande-Risposte
- Esportazione Fonti
- Esportazioni Province
- Esportazione Punti
- Esportazione Sociolinguistica
- Esportazione Sociometrica
- Esportazione Trascrizioni fonetiche

- Inizializzazione script (bottoni)
- Inizializzazione Carte
- Inizializzazione Informatore
- Inizializzazione Rilevatori

Inizializzazione Professioni  
Inizializzazione Profili  
Inizializzazione Semasiologiche

Ricerca Indicizzata  
Ricerca per Punto  
Ricerca per Formula

Ordinamenti

Campionamento vocale  
Elaborazione campionamento digitalizzato  
Relazione campionamento-scheda  
Relazione campionamento-contesto

Gestione Menù a tendina  
Gestione grafica  
Gestione degli interrupts  
Startup  
End

#### Procedure di Archivio

Procedure di inizializzazione Archivio Punti  
Procedure di inizializzazione Archivio Informatori  
Procedure di inizializzazione Archivio Domande  
Procedure di inizializzazione Archivio Fonti  
Procedure di inizializzazione Archivio Gestione  
Procedure di inizializzazione Archivio Carte  
Procedure di inizializzazione Archivio Etnografico

Procedure di stampa Archivio Punti  
Procedure di stampa Archivio Informatori  
Procedure di stampa Archivio Domande  
Procedure di stampa Archivio Fonti  
Procedure di stampa Archivio Gestione  
Procedure di stampa Archivio Carte  
Procedure di stampa Archivio Etnografico

Procedure di formattazione font di caratteri e alfabeto fonetico Archivio Punti  
Procedure di formattazione font di caratteri e alfabeto fonetico Archivio Informatori  
Procedure di formattazione font di caratteri e alfabeto fonetico Archivio Domande  
Procedure di formattazione font di caratteri e alfabeto fonetico Archivio Fonti  
Procedure di formattazione font di caratteri e alfabeto fonetico Archivio Gestione  
Procedure di formattazione font di caratteri e alfabeto fonetico Archivio Carte  
Procedure di formattazione font di caratteri e alfabeto fonetico Archivio Etnografico

Procedure di relazione Gestione-Punti  
Procedure di relazione Gestione-Informatori  
Procedure di relazione Gestione-Domande  
Procedure di relazione Gestione-Fonti  
Procedure di relazione Gestione-Carte  
Procedure di relazione Gestione-Etnografico

### Procedure di formato

Procedure di passaggi pagina interni al formato di Archivio Punti  
 Procedure di passaggi pagina interni al formato di Archivio Informatori  
 Procedure di passaggi pagina interni al formato di Archivio Domande  
 Procedure di passaggi pagina interni al formato di Archivio Fonti  
 Procedure di passaggi pagina interni al formato di Archivio Gestione  
 Procedure di passaggi pagina interni al formato di Archivio Carte  
 Procedure di passaggi pagina interni al formato di Archivio Etnografico

### Script interni ai formati

Script per Selezione sedi  
 Script per Selezione profili  
 Script per Selezione valutazioni rilevatore  
 Script per Selezione valutazione informatore  
 Script per Selezione domande semasiologiche  
 Script per Selezione sinonimi  
 Script per Selezione descrizioni  
 Script per Selezione cartografazione in trascrizione fonetica  
 Script per Selezione cartografazione sociolinguistica  
 Script per Selezione cartografazioni speciali

Data la vastità del pacchetto di procedure qui presentato eviterò di diffondermi in particolari tecnici e in descrizioni approfondite.

Toccherò solo alcuni punti qualificanti.

#### 3.3.1. Ricerca e ordinamento.

Le ricerche, ad esempio, sono state programmate per una massima efficienza del sistema. Sono possibili, evidentemente, tutti i tipi di ricerca per chiave singola o multipla. Le associazioni di chiavi multiple sono gestite da funzioni booleane (e /o / tranne) o aritmetiche (>di, <di, =a, <=di, >=di, contiene). Le ricerche sono possibili a partire dal modulo generale di gestione. Sono possibili anche ricerche produttive, cioè che producono schede nuove automaticamente quando si rispettano certe condizioni. Per tale scopo è stato approntato un editor di ricerche che simula un mini-linguaggio di programmazione con tutte le funzioni principali in linea.

Ecco un esempio dell'uso di questo editor di ricerca complessa:

**Finestra di Ricerca**

[Area di ricerca vuota con scrollbar]

---

<table border="1"> <tr><td>←</td><td>→</td></tr> <tr><td>Num_Scheda</td><td>↑</td></tr> <tr><td>Cod_Rilevatore</td><td>□</td></tr> <tr><td>Data Immissione</td><td>□</td></tr> <tr><td>Punto</td><td>▨</td></tr> <tr><td>Provincia</td><td>□</td></tr> <tr><td>Capoluogo</td><td>↓</td></tr> </table>	←	→	Num_Scheda	↑	Cod_Rilevatore	□	Data Immissione	□	Punto	▨	Provincia	□	Capoluogo	↓	<table border="1"> <tr><td>è uguale a</td><td>↑</td></tr> <tr><td>è diverso da</td><td>□</td></tr> <tr><td>è superiore a</td><td>▨</td></tr> <tr><td>è maggiore o uguale a</td><td>▨</td></tr> <tr><td>è minore di</td><td>▨</td></tr> <tr><td>è minore o uguale di</td><td>▨</td></tr> <tr><td>contiene</td><td>↓</td></tr> </table>	è uguale a	↑	è diverso da	□	è superiore a	▨	è maggiore o uguale a	▨	è minore di	▨	è minore o uguale di	▨	contiene	↓	<table border="1"> <tr><td><input type="radio"/> E anche</td></tr> <tr><td><input type="radio"/> Oppure</td></tr> <tr><td><input type="radio"/> Eccetto</td></tr> </table>	<input type="radio"/> E anche	<input type="radio"/> Oppure	<input type="radio"/> Eccetto
←	→																																
Num_Scheda	↑																																
Cod_Rilevatore	□																																
Data Immissione	□																																
Punto	▨																																
Provincia	□																																
Capoluogo	↓																																
è uguale a	↑																																
è diverso da	□																																
è superiore a	▨																																
è maggiore o uguale a	▨																																
è minore di	▨																																
è minore o uguale di	▨																																
contiene	↓																																
<input type="radio"/> E anche																																	
<input type="radio"/> Oppure																																	
<input type="radio"/> Eccetto																																	

Valore: [ ]

Cerca nella Selezione

Stesso discorso vale per gli ordinamenti, con in più la possibilità di estrarre sottoinsiemi da studiare in analisi trasformandoli in piccole banche-dati dedicate ad uno o più fenomeni dedicati. Gli ordinamenti possono essere dinamici o permanenti.

### 3.3.2. Import-Export dati.

Le procedure di output funzionano pure tramite un editor di specifiche dell'uscita prodotte dalle selezioni di Ricerca e Ordinamento. Esse ricorrono più che ad un linguaggio di programmazione ad una serie di comandi gestibili col mouse che selezionerà sia le periferiche che gli ordini di uscita che le disposizioni grafiche di questi ultimi. Gli output producibili possono essere convertiti - quando sono in forma statisticamente adeguata - in grafici statistici a istogrammi verticali o orizzontali, assi cartesiani, torte, grafici personalizzati riportati a disegni o icone. L'import-export dei dati prodotti dal modulo gestione prevede la programmazione dei codici di fine campo e fine record, variabili da caso a caso sia in lettura che in scrittura. Una particolare procedura ge-

stisce la filtratura dei caratteri ASCII in associazioni arbitrarie. I filtri prodotti possono essere registrati e riutilizzati. Nelle fasi di import ed export possono essere selezionati i campi che dovranno essere processati ed estratti. Questo permette di considerare le liste di esiti come vettori a più file indeterminate ed aperte.

Di ciò si avvantaggeranno molto le procedure cartografiche alcune delle quali sono state già approntate mentre altre sono in fase di progettazione avanzata. Esse saranno gestite da un algoritmo di lettura di liste indefinite ad n.righe per N.colonne. Nelle righe saranno contenuti i campi, nelle colonne i valori. Ad ogni incrocio di riga e colonna corrisponderà un punto grafico localizzato da certe coordinate prestabilite su cartina. La mappatura automatica dei punti favorirà una procedura di tracciamento delle isoglosse con strumenti elettronici e sulla base di una sovrapposizione di modelli cartografici prefissati. Il risultato finale sarà "fotografato" e archiviato nell'Archivio delle Carte.

### 3.3.3. Cartografazione integralmente variazionale.

I vantaggi di un sistema dinamico di cartografazione dovrebbero essere immediatamente evidenti. Il prodotto "cartina" non sarà più il manufatto artigianale definitivo che sancirà i risultati della ricerca. La semplicità con cui sarà possibile ottenerle, la possibilità di "inseguire" fenomeni linguistici di qualsiasi genere, di coglierne gli sviluppi cronologici, di comparare graficamente variazioni diatopiche, diastratiche, diacroniche, diafasiche o autovalutative, di proiettare sulle cartine le tendenze in atto intraviste in un certo momento dello sviluppo della ricerca, sulla base di un interrogatorio parziale e provvisorio del DB, può aprire nuove strade alla geolinguistica. Le procedure già realizzate per la produzione di carte speciali contempla sinora le seguenti possibilità:

- a) carte in distribuzione geografica delle risposte fonetiche (diatopico-fonetiche con trascrizione o riproduzione acustica)
- b) carte in distribuzione geografica delle risposte associate ai punti (diatopico-lessicali in trascrizione ortografica o fonetica)
- c) carte in distribuzione geografica delle risposte associate agli indici sociometrici (diatopico-sociometriche con stringhe alfanumeriche di dati e/o rappresentazioni in grafica statistica)
- d) carte in distribuzione geografica delle risposte associate agli indici sociolinguistici (diatopico-diastratiche con stringhe alfanumeriche di dati e/o rappresentazioni in grafica statistica)
- e) carte in distribuzione geografica delle risposte associate alle fonti e agli anni di rilevamento (diatopico-diacroniche con stringhe numeriche relative alla cronologia)
- f) carte in distribuzione geografica delle risposte autovalutative (diatopico-autovalutative con rappresentazioni in grafica statistica)

g) carte in distribuzione geografica delle condizioni d'inchiesta e dell'uso (diatopico-diafasiche con stringhe numeriche)

Una cartografazione automatizzata così come è stata pensata per l'ALS permette, inoltre, di *non fissare a priori* i criteri e i parametri di ricerca, ma è aperta a tutte le possibili visualizzazioni grafiche di tutti i possibili incroci tra campi. C'è da considerare inoltre che la memorizzazione magnetico-ottica delle carte permette un considerevole risparmio di spazi cartacei, riservando la produzione *delle sole carte prescelte*, in fogli di qualsiasi dimensione tramite uscite di qualità ineguagliabile su plotter multicolori (periferiche di grandi dimensioni che tracciano il grafico servendosi di particolari pennarelli direttamente comandati dal computer).

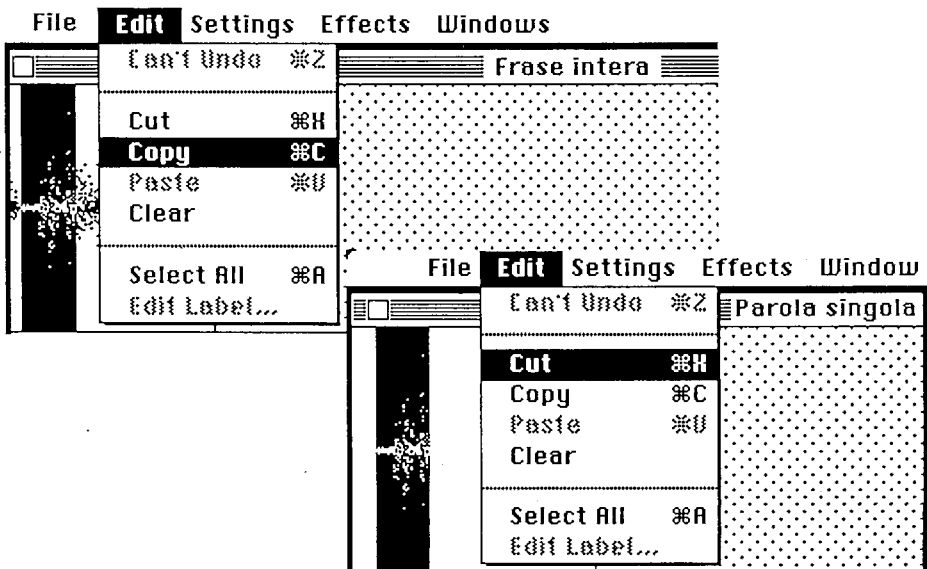
Il completo controllo delle variabili permetterà poi l'applicazione di quella programmazione intensiva cui si accennava nel § 2.1. La finezza degli output continui, che entrano come nuovi input nelle procedure di elaborazione sempre più dettagliata del dato, permetterà un ampliamento degli spazi di ricerca. Un esempio potrebbe essere fornito dall'esperienza dell' OLS in cui, a partire dai dati della trascrizione grezza, si è ottenuta, attraverso un'elaborazione a stati successivi, una quantità di dati fonetici, morfologici e - in seguito - si spera anche sintattici. Questo secondo stadio ha fornito ad un programma di elaborazione statistica le informazioni che hanno permesso di creare tabelle esaustive del fenomeno. A loro volta tali tabelle sono state processate da programmi di individuazione delle costanti e hanno prodotto tabelle sintetiche. Se compariamo i risultati finali alla informazione grezza da cui eravamo partiti ci accorgeremo che la strada compiuta è stata tanta e che non era pensabile all'inizio l'individuazione di fenomeni così ristretti come quelli trovati. Una procedura del genere è sicuramente estensibile all'ALS. Basti pensare alle elaborazioni possibili a partire dalle trascrizioni del parlato libero contenuti in uno dei campi del DB. Naturalmente in fase di progettazione sarà necessario fissare almeno le coordinate di base su cui poi operare. Ad esempio risolvere il problema della trascrizione fonetica, dei segni di evidenziazione per un parsing sintattico, etc.

### 3.3.4. Riascolto pronunce e contesti.

Ci troviamo in questo caso di fronte ad una novità assoluta nel campo dei DB linguistici e in special modo geolinguistici.

Come abbiamo accennato in precedenza è possibile, in particolari sistemi di sviluppo software legati ad un hardware tecnologicamente molto avanzato, predisporre una struttura di campi contenenti *resources* provenienti da un dispositivo di campionamento sonoro. Si tratta di un campionamento eseguito attraverso un convertitore digitale di suoni con entrate dirette per microfoni, piastre di registrazione,

apparecchi stereofonici e riproduttori comuni di cassette magnetiche ed uscite verso un canale predisposto per il *data-processing*. In pratica, attraverso questa strumentazione hardware-software è possibile riportare tutta un'intera intervista all'interno di una "risorsa" che può essere letta, attraverso una specifica procedura del DB, dal campo sonoro predisposto e poi ripetuta col semplice tocco di un pulsante associato o al numero della scheda o a un *range* numerico di schede che derivano tutte da quel reperto vocale (pronuncia del contesto). Con le stesse modalità di associazione è possibile realizzare - nell'ambito dell'uso a video della rappresentazione cartografica - delle "carte parlanti" che illustrano la distribuzione geografica del dato fonetico in ascolto. Le possibilità aperte da questa procedura sono di rilevante importanza. Innanzitutto sarà più facile ed accurata la trascrizione fonetica, potendo riascoltare infinite volte, e direttamente nel modulo del DB, l'item richiesto. In secondo luogo si potrà effettuare un'analisi approfondita del suono, poiché è sempre possibile rallentare o velocizzare il riascolto, alterarne le curve prosodiche, analizzarne gli spettrogrammi e i sonogrammi (in figura il "taglio" di una parola da un'intera frase).



Nella definizione del processo di dislocamento vanno discusse le problematiche relative alle periferiche di massa, cioè va progettata una struttura di *backup* permanente su hard-disk posti nei diversi distaccamenti.

Nel quadro dei problemi riguardanti la sicurezza del sistema va programmato, infine, il piano di protezione del DB. A tale proposito si può sin da ora assicurare una procedura di individuazione dei virus che è già montata sui sistemi operativi. Sia la stazione di Palermo, sia le attrezzature periferiche sono già dotate di software antivirus, che blocca l'accesso a qualsiasi *file* non conosciuto e impedisce, se è il caso, la continuazione dei lavori. La pericolosità di questo software-pirata è tuttavia notevole, poiché la tipologia dei virus è soggetta a continui arricchimenti che vanificherebbero le attuali protezioni. Nel corso di addestramento dei rilevatori si fisseranno, quindi, le norme di profilassi generale e preventiva per evitare la contrazione di "malattie ignote".

### 3.5. Dimensionamento ed espandibilità dell'ALS.

Veniamo, quindi, al problema più preoccupante della progettazione del DB per l'ALS.

Secondo le indicazioni contenute in Ruffino (in stampa b, ma si veda anche Ruffino in questo volume) il totale dei punti di inchiesta dovrebbe essere un massimo di 140. Di questi circa 45 dovrebbero avere vocazione "variazionale", 45 vocazione "etnografica", 15 4 vocazione "peschereccia".

Il numero di informatori per punto è fissato in 6 per ogni punto "normale" e 11 per i punti "variazionali". I punti "etnografici" e "pescherecci" vanno valutati in linea di massima come un solo informatore.

Infine i questionari dovrebbero comprendere un tipo "normale" composto da circa 1000 items, un tipo "etnografico" composto da una serie di items variante da i 1700 dell'ALS e i 7000 dell'ALI. Per le inchieste pescherecce si conviene sul modello ALM di 600 domande circa.

Altri dati saranno costituiti dai materiali di precedenti imprese geolinguistiche e da altri tuttora da definire (cfr. Ruffino in questo stesso volume). Di queste fonti di informazione è difficile calcolare un approssimativo dimensionamento.

Ecco un quadro riepilogativo :

Inchiesta	N° punti	N° Intervistati	N° Items	N° Tot. schede
Intera serie	140	6	1000	840000
Variazionale	45	11	1000	495000
Etnografica	45	1	1000	45000
Peschereccia	15	1	600	9000
Solo Base	35	0	0	0



TOTALE

• 1389000

Se calcoliamo adesso i dati relativi all'occupazione di memoria prevista ci accorgeremo delle difficoltà di dimensionamento del DB.

Il DB è previsto come, abbiamo visto, per un totale di 82 campi distribuiti nei 7 archivi collegati tra loro. Ecco un calcolo massimale su questa base:

Campi	bit*campo	bit	Kb	Mb
82	10	1825320000	1825320	18253,2

Se restringessimo il numero dei campi, almeno quelli su cui poter operare ricerche complete (Modulo di gestione), e dei tipi di campo, la situazione sarebbe questa:

Campi	bit*campo	bit	Kb	Mb
30	4	267120000	267120	2671,2
30	2	133560000	133560	1335,6

Tenendo conto del fatto che le memorie di massa più avanzate (almeno di memorie di massa magnetiche) gestiscono circa 1.200 MB (1.2 GB) in linea, dobbiamo quindi prevedere una soluzione restrittiva che qui viene prospettata sulla base di una tabella realistica:

Inchiesta	N° punti	N° Intervistati	N° Items	N° Tot. schede
Intera serie	80	6	1000	480000
Variazionale	40	11	1000	440000
Etnografica	40	1	1000	40000
Peschereccia	15	1	600	9000
Solo Base	0	0	0	0
<b>TOTALE</b>				<b>969000</b>

Campi	bit*campo	bit	Kb	Mb
82	10	794580000	794580	7945,8
30	4	116280000	116280	1162,8
30	2	58140000	58140	581,4

Solo nelle ultime due ipotesi prospettate è possibile pensare realisticamente ad un contenimento degli spazi di memoria.

Tutto ciò comunque non deve scoraggiarci. Stiamo infatti ipotizzando una situazione estremamente spinta e il cui raggiungimento non potrà essere realizzato prima di una decina di anni. Il discorso sulle periferiche di massa è in continuo movimento ed è sicuramente prevedibile un salto tecnologico enorme già nei prossimi due-tre anni con l'avvento dei cd. rom utilizzabili sia in fase di lettura che in fase di scrittura.

L'avvio dell'ALS è senz'altro possibile con gli strumenti attualmente in nostro possesso, ovvero con una batteria di hard-disk che vanno dai 20 ai 140 MB in linea.

Lo stesso discorso vale anche per il software di gestione che - seppure abilitato sin da ora alla gestione di 16.000.000 di records - sarà sicuramente ampliato nel futuro.

D'altro canto il problema non è strettamente informatico. La manipolazione e il dominio completo di tutti i dati dell'ALS (oltre 2.500.000 di schede, senza contare quelle di immagazzinamento degli archivi già esistenti) è *assolutamente impossibile con una gestione manuale*.

In questa fase di progettazione possiamo quindi verificare le possibilità di funzionamento del sistema sulla base di una programmazione biennale o, al massimo triennale. A questo proposito si è già realizzata con ottimi risultati una testatura del sistema immettendo dati simulati per un complesso di 100.000 schede in un hard-disk di 215 MB.

Utilizzando programmi di calcolo potremo volta per volta stabilire la programmazione biennale dei dati da immettere che, comunque, non potranno superare, anche secondo le più spinte previsioni, i 100.000 dati già immessi per simulazione.

Per concludere, il discorso sul dimensionamento dell'ALS dovrebbe far riflettere sulle reali possibilità di realizzare apparati così quantitativamente (oltreché qualitativamente) complessi senza un articolato piano di previsioni temporali e finanziarie.

### 3.6. Fruibilità dell'ALS

Qualche parola conclusiva va spesa sulla facilità di accesso al sistema prescelto.

Diremo innanzitutto che la stazione informatica dell'ALS è fondata sui sistemi Apple MacIntosh, quanto di meglio oggi esiste in fatto di microinformatica personale. Come è stato già detto, un'attenzione particolare è stata dedicata dall'Apple alla creazione di un sistema fondato su un'interfaccia amichevole per l'utente. Essa fa ampio uso di icone, *mouse*, menù a tendina, bottoni, ed è fondata sulla programmazione orientata all'oggetto.

L'architettura della macchina principale (Macintosh IIx) è basata su un processore Motorola 68030 a 25 Mhz, con parola di 32 bit. E' il corrispettivo dei sistemi 386 e 486 dell'IBM con il quale non solo è compatibile a livello di *files* (legge i dischetti IBM) ma è anche totalmente simulabile via software.

Il generatore di applicazioni prescelto è il 4Dimension, un applicativo francese della ACI capace di gestire DB complessi sino a 16.000.000 di records. Tale generatore di applicazioni permette l'uso esteso dell'interfaccia MacIntosh anche nella gestione dei DB.

In parole povere, alla Commissione Scientifica e a tutti gli utenti non specializzati sarà possibile accedere con estrema facilità anche ai dettagli più minuti della banca-dati.

Per quanto riguarda le proiezioni esterne, il sistema è in grado di fornire una serie di output altamente qualificati, sia cartacei che magnetico-ottici. In un prossimo futuro è pensabile una masterizzazione dei dati su cd.rom. E' previsto anche una utilizzazione in rete, attraverso i protocolli di comunicazione più diffusi e, in particolare, sui protocolli di comunicazione UNIX - sistema operativo potentissimo, utilizzato dai principali *mainframes* - già da ora disponibile sul Mac IIx.

## Conclusioni

Il piano di informatizzazione qui presentato mostra, crediamo, quanto complesso sia il lavoro di organizzazione delle informazioni in un settore così complesso come quello della ricerca linguistica in generale e geolinguistica in particolare.

Tutte le possibilità che abbiamo cercato di esplicitare dettagliatamente potrebbero non essere sfruttate a fondo dall'ALS. La messa a punto del piano teorico su cui dovrà muoversi questo atlante sperimentale non è ancora definitiva ed è sempre possibile che gli ultimi ritocchi accentuino o riducano l'interesse per alcuni domini di ricerca.

Restano, inoltre, ancora molte questioni da precisare connesse con le caratteristiche dei questionari, con la possibilità di contemplare una vasta raccolta di reperti di parlato libero, con la eventuale ricognizione di materiale onomastico e toponomastico, etc.<sup>25</sup>

Tutto ciò cui si dovrà rinunciare o, viceversa, l'insieme delle informazioni aggiuntive che ancora devono essere considerate, potranno, tuttavia, trovare una loro sistemazione opportuna nella struttura di questo progetto di informatizzazione che è stato concepito come piattaforma flessibile e modificabile anche durante la sua realizzazione empirica concreta.

Oltre a ciò è sicuramente preventivabile un continuo ricambio e aggiornamento tecnologico sia per quanto riguarda la piattaforma hardware, sia per ciò che concerne la manutenzione e il potenziamento del software<sup>26</sup>.

Ciò che resterà comunque inalterata è la filosofia modulare del progetto, la sua struttura relazionale e il suo intento spiccatamente variazionista nella convinzione che solo entro i quadri di un'impiantistica informatica personale fortemente orientata sul versante dei "progetti dedicati" sarà possibile portare a termine piani di ricerca adeguati al futuro della "società dell'informazione globale" che è ormai un'accertata realtà dei nostri giorni.

---

<sup>25</sup> Opportunità già vagliate in Ruffino (in stampa b).

<sup>26</sup> Tutta l'installazione dell'hardware e del software sono stati realizzati con la massima precisione e con grande competenza tecnica dalla ditta Text Software di Catania - specializzata esclusivamente in informatica per l'area umanistica - a cui vanno tutti i nostri ringraziamenti. Questa ditta ci ha anche assicurato una costante assistenza, manutenzione e aggiornamento dell'impianto di macchine e programmi nei prossimi anni.



## BIBLIOGRAFIA

- Agostiniani, L.- Montemagni, S.- Poggi Salani, T., *Atlante Lessicale Toscano: il lavoro di preedizione e la costituzione della banca dati*, in *Atlanti regionali: aspetti metodologici, linguistici e etnografici* (Atti del XV Convegno del Centro di Studio per la Dialettologia Italiana, Palermo 7-11 ottobre 1985), Pisa s.d., pp. 1-7.
- Aho, A.- Beeri, C.- Ullman, J. D., *The theory of joins in relational databases*, "ACM", 2/3, 1977.
- Alfonzetti, G., *Un' ipotesi sul sistema del code-shifting nel diasistema siciliano*, Comunicazione al Convegno SLI sul tema "Linguistica e pragmatica", Milano, settembre 1990 (in corso di stampa).
- Alvar, M.- Nuño M. P., *Un ejemplo de Atlas Lingüístico automatizado: el ALES*, "Lingüística Española Actual", Instituto de Cooperación Ibero-Americana, Madrid, I/2, 1981, pp. 359-74.
- Batinti, A., *Proposte per la costituzione di una banca dati* (in corso di stampa).
- Bauer, R., *Bericht zum Stand der elektronischen Datenverarbeitung im Projekt ALD I*, 1987 (in corso di stampa).
- Bayer, R.- Mc. Creeight C., *Organization and Maintenance of Large Ordered Indexes*, "Acta Informatica", I/3, 1972.
- Bershin, H., *Was leistet die elektronische Datenverarbeitung in der Dialektologie?*, "Germanistische Linguistik", 3-4, 1977, pp. 119-20.
- Bershin, H., *Problème der Datenaufbereitung in der automatischen Sprachkartographie*, "Germanistische Linguistik", 3-4, 1977, pp. 159-62.
- Brook, R. J., *Computer Reduction and Retrieval of Dialect Data: a Methodology for Cartographic and Tabular Display of Lexical and Sociolinguistic Variation*, Ph. D. Univ. of Iowa, 1970.
- Cassidy, F. G., *Dialectology and the Electronic Drudge*, in S. Ellis (a cura di), *Studies in Linguistics in Honour of Harold Orton on the Occasion of His Seventieth Birthday*, "Leeds Studies in English", N. S., 2, 1969, pp. 135-43.
- Cassidy, F. G., *On-line Mapmaking for the Dictionary of American Regional English*, "Germanistische Linguistik", 3-4, 1977 a, pp. 107-119.
- Cassidy, F. G., *Computer-Aided Usage 'Labeling' in Dictionary*, "Computers and the Humanities", 11, 1977 b, pp. 89-99.

- Chen, P. S., *The entity-relationship model: a basis for enterprise view of data*, Preprints, "NCC", 1977.
- Chen, P. S., *The entity-relationship model: toward a unified view of data*, "ACM", 3, 1977.
- Codd, E. F., *A relational model of data for large shared data banks*, "ACM", 13/6, 1970 .
- Codd, E. F., *Further normalization of data base relation model*, in *Data Base Systems*, Courant Computer Science Symposia Series, 6, Prentice-Hall 1972.
- Codd, E. F., *Recent investigations into relational data base systems*, Proc. IFIP Congress, 1974.
- Codd, E. F., *Codd's 12 rules for relational DBMS*, "The Relational Journal", 1, 1987.
- Coseriu, E., *Los conceptos de 'dialecto', 'nivel', y 'estilo de lengua' y el sentido propio de la dialectologia*, "Linguística Española Actual", 3, 1981, pp. 1-32.
- Cox, H. L., *Elektronische Datenverarbeitung in der thematischen Kartographie. Kartierungstechnik in Ethnologie und Ethnolinguistik*, "Ethnologica Europae", 6, 1972, pp. 108-27.
- Cox, H. L., *Printer und Plotter in der Linguistischen Kartographie. Ein Beitrag zur komputativen Sprachgeographie*, "Leuvense Bijdragen", 62, 1973, pp. 139-54.
- Cox, H. L. - Griffioen, W., *Plotterkarten in der Ethnokartographie*, "Germanistische Linguistik", 3-4, 1977, pp. 121-36.
- D'Agostino, M., *Ipotesi di costruzione di indici sociometrici per l'Atlante linguistico della Sicilia* (in questo volume).
- Date, C. J., *A guide to DB2*, Addison-Wesley 1984.
- Date, C. J., *An introduction to Database Systems: voll. 2*, Addison-Wesley 1985-1986.
- Date, C. J., *Relational Database. Selected Writings*, Addison-Wesley 1986.
- Dean, H. O., *Toward Automated Phonological Mapping*, "Linguistics", 17, 1965, pp. 21-35.
- Dees, A., *Atlas des formes et des constructions des chartes françaises du 13<sup>e</sup> siècle*, Tübingen 1980.
- Delmonte, R. (a cura di), *Linguistica computazionale*, Padova 1983.

- Eickmans, H., *Automatische Sprachkartographie im Rahmen des Projekts 'Fränkischer Sprachatlas'*, "Niederdeutsches Wort", 19, 1979, pp. 133-64.
- Eickmans, H., *Mogelijkheden en grenzen van de automatische verwaardiging von taalkaarten*, "Taal en tongval", 33/3-4, 1981, pp. 162-78.
- Fagin, R., *Multivalued dependencies and a new normal form for relational databases*, "ACM", 2/3, 1977.
- Festinger, L. -Schachter, S. -Bach K., *Analisi matriciale delle strutture di gruppo*, in Boudon, R. - Lazarsfeld, P. (a cura di), *L'analisi empirica nelle scienze sociali*, Bologna 1969.
- Fossat, J. L., *Vers un traitement automatique des données dialectologiques en dialectométrie*, "Germanistische Linguistik", 3-4, 1977, pp. 311-34.
- Francis, W. N., *Computer Production of Dialect Maps*, "Computers and the Humanities", 3, 1969, p. 290.
- Francis, W. N., *International Colloquium on Automatic Dialect Mapping: A Report*, "Computers and the Humanities", 2, 1977, pp. 339-40.
- Francis, W. N.- Svartvik, J.- Rubin, G. M., *Computer-Produced Representation of Dialectical Variation: Initial Fricatives in Southern British English*, in *International Conference on Computational Linguistics*, Preprint n° 52, Stockholm 1969.
- Frierbertshäuser, H. - Dilgendein, H. J., *Wortgeographie der städtischen Alltagssprache in Hessen*, (Graphisch-computative Bearbeitung: H. Handler-W. Putschke), Tübingen 1988.
- Frierbertshäuser, H. - Dilgendein, H. J., *Hessischer Dialektzensus. Statischer Atlas zum Sprachgebrauch*, Tübingen 1989.
- Geisler, H., *Erstellung und Auswertung von Dialektkarten mit Personal Computers*, in AA. VV., *Akten der Sektion Computerlinguistik des XXI. Romanistenstages*, Tübingen 1990a, pp. 1-14 (estratto).
- Geisler, H., *Aufbereitung lexikalischer Information und deren Umsetzung in Sprachkarten (aufgezeigt anhand romanischer Beispiele)*, in Rieger, B. - Schaefer, B. (a cura di), *Lexicon und Lexicographie*, Herbst 1990b, pp. 1-11 (estratto).
- Goebel, H., *La dialectométrie appliquée à l'ALF (Normandie)*, Atti del XIV Congresso Internazionale di Linguistica e Filologia Romanza, a cura di A. Varvaro, Napoli 1974, pp. 165-95.
- Goebel, H., *Atlas, matrices et similarites: petit aperçu dialectométrique*, "Computers and the Humanities", 16, 1982, pp. 69-84.



- Goebel, H., *Problemi e metodi della classificazione geolinguistica*, in Holtus, G. - Metzeltin M. (a cura di), *Linguistica e Dialectologia veneta. Studi offerti a M. Cortelazzo dai colleghi stranieri*, Tübingen 1983, pp. 193-203.
- Goebel, H., *Lineamenti di dialettometria (con applicazione all' AIS)*, "Guida ai dialetti veneti", 6, 1984, pp. 7-53.
- Goebel, H., *Considérations dialectométrique sur le problème de "l'unité rhétoromane (ladine)"*, "Linguistica", 26, 1986, pp. 83-97.
- Goebel, H., *Points chauds de l'analyse dialectométrique: ponderation et visualisation*, "Revue de linguistique romane", 51, 1987.
- Grassi, C., *Perspectives de l'emploi de l'elaborateur électronique en géographie linguistique et en dialectologie*, Atti della Conferenza Internazionale di Linguistica Computazionale, Pisa, Ottobre-Novembre 1970.
- Grassi, C.- Genre, A.- Massobrio, L., *Die Anwendung der elektronischen Datenverarbeitung bei der Redaktion und dem Druck des Italienischen Sprachatlasses (ALI)*, "Germanistische Linguistik", 3-4, 1977, pp. 69-88.
- Händler, H., *Zum Dokumentations und Speicherungsproblem im Rahmen der Automatischen Sprachkartographie*, "Germanistische Linguistik", 3-4, 1977 a, pp. 163-80.
- Händler, H., *Eine Methode zur Automatischen Herstellung von Isoglossen*, "Germanistische Linguistik", 3-4, 1977 b, pp. 259-74.
- Händler, H., *Eptwürfe zu dialektalen Informationssystemen*, in Besch-Knop-Putschke (a cura di), *Wiegand Dialektologie*, Berlin 1983, pp. 792-806.
- Händler, H.- Naumann, C. L., *Zur Automatisierung der Isoglossenbindung*, "Germanistische Linguistik", 3-4, 1977, pp. 121-59.
- Henderson, M. T., *Use of an Interactive Program in Analyzing Data for a Dialect Dictionary*, "Computers and the Humanities", 9, 1975, pp. 105-13.
- Houck, C.L., *A Computerized Statistical Methodology for Linguistic Geography: a Pilot Study*, "Folia Linguistica", 1, 1967, pp. 80-95.
- Houck, C.L., *A Statistical and Computerized Methodology for Analyzing Dialect Materials*, Ph. D., Univ. of Iowa, 1969, Dissertation abstract reprint, "Computers and the Humanities", 8, 1974, pp. 120-21.
- IBM DataBase 2, *SQL Reference*, SC26-4346-0. *Application Programming Guide*, SC26-4293. *Database Planning and Administration Guide*, SC26-4077. *General Information Manual*, GC26-4073-3. *Referential Integrity Usage Guide*, GG24-3312.

- Jones (Sargent), V M., *Some Problems in the Computation of Sociolinguistic Data*, Ph. D. Univ. of Newcastle-upon-Tyne, 1978.
- Jones (Sargent), V M., *The Tyneside Linguistic Survey-an Approach to Data Processing in Sociolinguistics*, "Statistical Methods in Linguistics Quaterly Journal", 2, 1978, pp. 5-23.
- Keil, G. C., *IMPAC (Indexing, Map Plotting, and Analysis by Computer)*, "Computers and the Humanities", 7, 1972, p. 116.
- Keil, G. C., *Narrow Phonetic Transcription on the Computer: Taking the Phone off the Hook*, "Computers and the Humanities", 8, 1974, pp. 217-29.
- Kelle, B., *Das Belegmaterial des Südwestdeutschen Sprachatlas. Aufbereitung und Auswertung mit Hilfe der Elektronischen Datenverarbeitung*, Friburg 1976.
- Kelle, B., *Datenaufbereitung und Automatische Kartierung beim Südwestdeutschen Sprachatlas (SSA)*, "Germanistische Linguistik", 3-4, 1977, pp. 89-106.
- Knuth, D. E., *The art of computer programming*, Vol. III: *Sorting and searching*, Addison-Wesley 1973.
- Lance, D. M., *Determining Dialect Boundaries in the United States by Means of Automatic Cartography*, "Germanistische Linguistik", 1977, 3-4, pp. 289-303.
- Lance M.- Slemmons S., *The Use of the Computer in Plotting the Geographical Distribution of Dialect Items*, "Computers and the Humanities", 10, 1976, pp. 221-29.
- Lo Faro, F., *Metodi quantitativi e informatizzazione nella ricerca dialettologica. Un esempio di dialettologia urbana: Catania*, Tesi di Laurea, Università di Catania, Anno Accademico 1988-89.
- Naumann, C. L., *Grundzüge der Sprachkartographie und ihrer Automatisierung*, "Germanistische Linguistik", 3-4, 1977, pp. 1-285.
- Naumann, C. L., *Klassifikation in der Automatischen Sprachcartographie*, "Germanistische Linguistik", 3-4, 1977, pp. 181-210.
- Naumann, C. L., *Symbolisierungsverfahren im Rahmen der Automatische Sprachkartographie*, "Germanistische Linguistik", 3-4, 1977, pp. 225-244.
- Naumann, C. L. - Putschke, W., *Automatisierung der etymologischen Klassifikation - eine ungenutzte Bearbeitungshilfe im ALE ?*, in AA. VV., *Aspects of Language: Studies in Honour of Mario Alinei*, 1: *Geolinguistics*, Amsterdam 1986, pp. 149-70.
- Nicolas, J. M., *Mutual dependencies and some results on undecomposable relations*, Proc. 4th International Conference on Very Large Data Bases, 1978.

Ogino, T., *Computer-aided Analysis of Field Survey Data GLAPS and its Application*, in *Proceedings of the XIIIth International Congress of Linguistics* (Tokyo), The Hague 1982, pp. 981-95.

Ogino, T.- Sibata, T., *Generalization of Procedures Relating to Computer-produced Linguistic Atlases*, "Germanistische Linguistik", 3-4, 1977, pp. 55-68.

Philps, D., *Projet de traitement automatique de l'Atlas dialectométrique des Pyrénées centrales*, Toulouse-Jeddale 1978.

Philps, D., *Dialectométrie automatique*, in Goebel, H. (a cura di), *Dialectology*, Bochum, Brochmeyer 1984, pp. 275-96.

Pudlatz, H., *Automatische Erzeugung von Isoglossen auf dem Plotte mit Hilfe von Thiessen Poygonen*, "Germanistische Linguistik", 3-4, 1977, pp. 245-58.

Putschke, W., *Planung einer Projektdurchführung: Automatische Kartierung des Atlas Linguarum Europae*, "Germanistische Linguistik", 4, 1972, pp. 547-77.

Putschke, W., *Automatische Kartierung*, in A. Weijnen (a cura di), *Atlas Linguarum Europae. Introduction*, Assen 1975, pp. 103-08.

Putschke, W., (a cura di), *Automatische Sprachkartographie*, Vorträge des Internationalen Kolloquiums zur Automatischen Sprachkartographie, (Marburg vom 11-16 september 1977).

Putschke, W., (a cura di), *Automatische Sprachkartographie: Konzeption, Probleme und Perspektiven*, "Germanistische Linguistik", 3-4, 1977, pp. 25-40.

Putschke, W., *Über ein Computerprogramm zur Herstellung von Sprachkarten*, "Germanistische Linguistik", 1977, 3-4, pp. 45-114.

Putschke, W., -Neumann, R., *Automatische Sprachkartographie*, in Besch-Knop-Putschke (a cura di), *Wiegand Dialektologie*, Berlin 1983, pp. 748-78.

Putschke, W. - Neumann, R., *Atlas Linguarum Europae (ALE). Aus der Sicht der automatischen Kartierung*, in Putschke, W. - Veith, W. H. (a cura di), *Sprachatlant des Deutschen. Laufende Projekte*, Tübingen 1989.

4e Dimension, *Manuale d'uso*, 4 voll.: 1. *Generalità*, 2. *Strutture*, 3. *Utilizzazione*, 4. *Linguaggi di programmazione*, Paris 1989.

Rolshoven, J., *Qualitative und quantitative Dialektologie: Zum Projekt einer automatischen Edition und Auswertung, von Sprachatlanten*, in Goebel, H. (a cura di), *Dialectology*, Bochum, Brochmeyer 1984, pp. 297-313.

Rubin, M., *Computer-Produced Mapping of dialectal Variation*, "Computers and the Humanities", 4, 1970, pp. 241-46.

- Ruffino, G., *Prospettive di lavoro per un atlante linguistico-etnografico della Sicilia*, Palermo 1986.
- Ruffino, G., *La dialettologia siciliana tra consuntivi e programma*, in Holtus, G. - Metzeltin, M. (a cura di), *La dialettologia italiana oggi. Studi offerti a Manlio Cortelazzo*, Tübingen 1989, pp.325-43.
- Ruffino, G., *A proposito del contatto dialetto-lingua: prime ipotesi di lavoro per un atlante linguistico della Sicilia* (in stampa a).
- Ruffino, G., *Linee di discussione e ipotesi di lavoro per l'Atlante linguistico della Sicilia*, Comunicazione al XVII Congresso Internazionale di Linguistica e Filologia Romanza, Santiago de Compostela, settembre 1989 (in stampa b).
- Ruffino, G., *L'Atlante linguistico della Sicilia: impianto progettuale, stato dei lavori, prospettive* (in questo volume).
- Schneider, E. W. -Viereck, W., *The Use of the Computer in American, Canadian and British english Dialectology and Sociolinguistics*, in Goebel, H. (a cura di), *Dialectology*, Bochum, Brochmeyer 1984, pp. 15-60.
- Schopaus, R., *Automatische Herstellung wortgeographischer Karten*, "Niederdeutsches Wort",9, 1969, pp. 97-115.
- Schütz, A. J. -Wenker J., *A Program for the Determination of lexical Similarity between Dialects*, in Garvin-Spolsky (a cura di), *Computation in Linguistics: a Case book*, London Bloomington, 1966, pp. 124-45.
- Shaw, R., *Statistical Analysis of Dialectal Boundaries*, "Computers and the Humanities",8, 1974, pp. 173-77.
- Shuy, R., *An Automatic Retrieval Program for the Linguistic Atlas of the United States and Canada*, in Garvin-Spolsky (a cura di), *Computation in Linguistics: a Case book*, London-Bloomington, 1966, pp. 60-75.
- Shuy, R., *Review of Wood, Vocabulary Change*, "Computers and the Humanities", 7, 1972, pp. 108-09.
- Shuy, R., *Some Problems in Preparing Dialect Data for Processing*, "Germanistische Linguistik",3-4, 1977, pp. 145-58.
- Sobrero, A.A., *Il progetto NADIR*, in Actes du XVII Congrès Internationale de Linguistique et Philologie Romance, Université de Aix en Provence, VI, 1986,pp. 491-504.
- Sobrero, A.A.- Romanello, M.T.- Zampolli, A., *Per un atlante modulare: il NADIR, Atlanti regionali: aspetti metodologici, linguistici e etnografici* (Atti del XV Convegno del Centro di Studio per la Dialettologia Italiana, Palermo 7-11 ottobre 1985), Pisa s.d., pp. 515-37.

Southard, B., *The Linguistic Atlas of Oklahoma and Computer Cartography*, "Journal of English Linguistics", 16, 1983, pp. 65-77.

Sprugnoli, R., *Le basi di dati. Gli strumenti e le tecniche per la gestione automatica delle informazioni*, Roma 1987.

Stehl, T., *Geolinguistica regionale e analisi variazionale. Considerazioni metodologiche per l'Atlante linguistico della Sicilia*, "Bollettino del Centro di studi filologici e linguistici siciliani", 16, 1990, pp. 119-59.

Technology Transfer, *Relational Database '87: Conferenza sullo stato dell'arte*, Codd & Date International, Roma, 6-8 ottobre 1987.

Technology Transfer, *Relational Database '88: Conferenza sullo stato dell'arte*, Codd & Date International, Roma, 24-27 ottobre 1987.

Thomas, A. R., *Dialect Mapping*, "Orbis", 24, 1975, pp. 115-124.

Thomas, A. R., *A Cumulative Matching Technique for Computer Determination of Speech Areas*, "Germanistische Linguistik", 3-4, 1977, pp. 275-88.

Thomas, A. R., *Computer Analysis of a Dialectal Transition Belt*, "Computers and the Humanities", 14, 1980 a, pp. 241-51.

Thomas, A. R., *Areal Analysis of Dialect Data by Computer. A Welsh example*, Cardiff, Wales U. P. 1980 b.

Tokugawa, M. - Yamamoto, T., *An Attempt to Draw a Linguistic Map a Computer*, "Mathematical Linguistics", 40, 1967, pp. 27-30.

Tuailion, G., *Les atlas linguistiques regionaux de France*, "Bollettino dell'Atlante linguistico italiano, III serie, 7, 1983, pp. 68-89.

Tuailion, G., *How the French Dialectal Data Enter the Atlas Linguarum Europae*, "Computers and the Humanities", 20, 1986, pp. 247-52.

Uskup, F. L., *A Method for Automating Dialect Analysis*, ERIC reports ED 037 713, 1970.

Uskup, F. L. - Al-Azawi M. L., *Editing and Printing a Dialect Atlas by Computer*, "American Speech", 3-4, 1972, pp. 203-10.

Varvaro, A., *Osservazioni sul progetto di un atlante linguistico della Sicilia*, "Bollettino del Centro di studi filologici e linguistici siciliani", 15, 1986, pp. 439-44.

Weinreich, U., *Machine Aids in the Compilation of Linguistic Atlases*, "Year Book of the American Philosophical Society", Philadelphia, 1963, pp. 622-25.

Wood, G. R., *Text, Tables, and Maps of a Word Geography of the Interior South*, "Year Book of the American Philosophical Society", Philadelphia, 1961, pp. 602-604.

Wood, G. R., *Dialectology by computer*, International Conference on Computational Linguistics, Preprint n° 19, Stockholm 1969.

Wood, G. R., *On Ways to Examine the Local Language*, "Computers studies in the humanities and verbal behavior", 3/1, 1970 a, pp. 100-10.

Wood, G. R., *Refinements in tabular models of variation in regional american english*, in Atti della Conferenza Internazionale di Linguistica Computazionale, Pisa, ottobre-novembre, 1970 b.

Wood, G. R., *The Computer in Analysis and Plotting*, "American Speech", 3-4, 1972, pp. 195-202.

Zampolli, A., *Trattamento automatico di dati linguistici e linguistica quantitativa*, in *Linguaggi e formalizzazioni*, Roma 1979.

Zampolli, A.- Montemagni, S., *Dialettologia e informatica*, "Rivista Italiana di Dialettologia", XI, 1987, pp. 149-74.

Zeisel, H., *Due esempi di costruzione di un indice*, in Boudon, R.- Lazarsfeld P. (a cura di), *L'analisi empirica nelle scienze sociali*, Bologna 1969.



# Appendice 1

## *Struttura Archivi e Records*





Struttura: Archivio Puntl		
Punto	Alfa 20	Lista Scelte; Indicizzato; Unico; Obbligatorio; Inseribile; Modificabile
Provincia	Alfa 2	Lista Scelte; Obbligatorio; Inseribile; Modificabile
Capoluogo	Booleano	Obbligatorio; Inseribile; Modificabile
Comune con fraz.	Booleano	Obbligatorio; Inseribile; Modificabile
Frazione	Booleano	Obbligatorio; Inseribile; Modificabile
Tipo Geografico	Intero	Lista Scelte; Obbligatorio; Inseribile; Modificabile
Tipo Linguistic	Intero	Lista Scelte; Obbligatorio; Inseribile; Modificabile
Sede di	Alfa 8	Obbligatorio
Bibliografia	Testo	Lista Scelte; Inseribile; Modificabile
Popolazione	Intero Lungo	Obbligatorio; Inseribile; Modificabile
Indice SOMEA	Intero	Obbligatorio; Inseribile; Modificabile
Indice Urbano	Intero	Obbligatorio; Inseribile; Modificabile
Numero Punto	Intero	Indicizzato; Unico; Obbligatorio; Inseribile; Modificabile

Struttura: Arc_Informatori		
Cod_Informatore	Intero	Lista Scelte; Indicizzato; Unico; Obbligatorio; Inseribile; Modificabile
Sesso	Alfa 2	Lista Scelte; Obbligatorio; Inseribile; Modificabile
Età	Intero	Obbligatorio; Inseribile; Modificabile
Istruzione	Intero	Lista Scelte; Obbligatorio; Inseribile; Modificabile
Professione	Alfa 2	Lista Scelte; Obbligatorio; Inseribile; Modificabile
Profilo	Alfa 5	Lista Scelte; Obbligatorio; Modificabile
Autovalutazione	Alfa 2	Inseribile; Modificabile

Struttura: Arc_Domande		
Domanda	Alfa 80	Lista Scelte; Indicizzato; Unico; Obbligatorio; Inseribile; Modificabile
Concetto	Alfa 20	Lista Scelte; Obbligatorio; Inseribile; Modificabile
Tipologia Doman	Intero	Lista Scelte; Obbligatorio; Inseribile; Modificabile
Cl_Grammaticale	Intero	Lista Scelte; Obbligatorio; Inseribile; Modificabile
Fonte Domanda	Alfa 3	Lista Scelte; Obbligatorio; Inseribile; Modificabile
Num_Rif_In_Q_Fo	Intero	Inseribile; Modificabile
Classe Sem_1	Intero	Lista Scelte; Inseribile; Modificabile
Classe Sem_2	Intero	Lista Scelte; Obbligatorio; Inseribile; Modificabile
Classe Sem_3	Intero	Lista Scelte; Obbligatorio; Inseribile; Modificabile
Classe Sem_4	Intero	Lista Scelte; Obbligatorio; Inseribile; Modificabile

Struttura: Archivio Fonti		
Fonte	Alfa 3	Lista Scelte; Indicizzato; Unico; Obbligatorio; Inseribile; Modificabile
Anno rilevament	Intero	Obbligatorio; Inseribile; Modificabile
Qualità registr	Intero	Lista Scelte; Obbligatorio; Inseribile; Modificabile
Supporto	Intero	Lista Scelte; Obbligatorio; Inseribile; Modificabile

Struttura: Modulo Gestione		
Num_Scheda	Intero Lungo	Indicizzato; Unico; Obbligatorio; Inseribile; Modificabile
Cod_Rilevatore	Intero	Lista Scelte; Obbligatorio; Inseribile; Modificabile
Data Immissione	Data	Obbligatorio; Inseribile
Punto	Alfa 20	Lista Scelte; Indicizzato; Obbligatorio; Inseribile; Modificabile
Provincia	Alfa 2	Lista Scelte; Indicizzato; Obbligatorio
Capoluogo	Booleano	Indicizzato; Obbligatorio
Comune con fraz	Booleano	Obbligatorio
Frazione	Booleano	Obbligatorio
Tipo Geografico	Intero	Indicizzato; Obbligatorio
Tipo Linguistic	Intero	Obbligatorio
Sede di	Alfa 8	Indicizzato; Obbligatorio
Popolazione	Intero Lungo	Indicizzato; Obbligatorio
Indice SOMEA	Intero	Indicizzato; Obbligatorio
Indice Urbano	Intero	Indicizzato; Obbligatorio
Cod_Informatore	Intero	Lista Scelte; Indicizzato; Obbligatorio; Inseribile; Modificabile
Sesso	Alfa 2	Lista Scelte; Indicizzato; Obbligatorio
Età	Intero	Indicizzato; Obbligatorio
Istruzione	Intero	Indicizzato; Obbligatorio
Professione	Alfa 2	Lista Scelte; Obbligatorio
Profilo	Alfa 5	Obbligatorio
Fonte	Alfa 3	Lista Scelte; Indicizzato; Obbligatorio; Inseribile
Anno rilevament	Intero	Obbligatorio
Qualità registr	Intero	Lista Scelte; Obbligatorio
Domanda	Alfa 80	Lista Scelte; Indicizzato; Obbligatorio; Inseribile; Modificabile
Concetto Demand	Alfa 20	Lista Scelte; Indicizzato; Obbligatorio; Modificabile
Num_Rif_In_Q_Fo	Intero	Obbligatorio
Ct_Grammaticale	Intero	Lista Scelte; Indicizzato; Obbligatorio
Tipologia Doman	Intero	Lista Scelte; Indicizzato; Obbligatorio
Classe Sem_1	Intero	Lista Scelte; Indicizzato; Obbligatorio
Classe Sem_2	Intero	Lista Scelte; Indicizzato; Obbligatorio
Classe Sem_3	Intero	Lista Scelte; Obbligatorio
Trascrizione or	Alfa 20	Obbligatorio; Inseribile; Modificabile
Risposta	Alfa 80	Indicizzato; Inseribile; Modificabile
Concetto Rispos	Alfa 20	Lista Scelte; Obbligatorio; Inseribile; Modificabile
Usa Sinonimo	Booleano	Inseribile; Modificabile
Usa Descrizione	Booleano	Inseribile; Modificabile
Val_Informatore	Alfa 9	Inseribile; Modificabile
Semasiologiche	Alfa 7	Lista Scelte; Inseribile; Modificabile
Val_Rilevatore	Alfa 25	Inseribile; Modificabile
Numero Punto	Intero	Indicizzato; Inseribile; Modificabile
N_Reperito	Intero	Lista Scelte; Indicizzato; Inseribile; Modificabile

Struttura: Archivio Carte		
Lista	Testo	Modificabile
Titolo	Alfa 20	Lista Scelte; Inseribile; Modificabile
Data Carta	Data	Obbligatorio; Modificabile

Struttura: Arc_Etnografico		
N_Reperito	Intero	Lista Scelte; Indicizzato; Obbligatorio; Inseribile; Modificabile
Reperito	Disegno	Inseribile; Modificabile
Nome reperito	Alfa 20	Lista Scelte; Inseribile; Modificabile

## Appendice 2

*Formati*



*Elenco dei formati*

**Archivio Punti**

Input  
Output

**Archivio Informatori**

Input  
Output

**Archivio Domande**

Input  
Output

**Archivio Fonti**

Input  
Output

**Archivio Gestione**

Input  
Output  
Esportazione Concetti  
Esportazioni risposte  
Esportazione Domande-Risposte  
Esportazione Fonti  
Esportazioni Provincie  
Esportazione Punti  
Esportazione Sociolinguistica  
Esportazione Sociometrica  
Esportazione trascrizioni fonetiche

**Archivio Carte**

Importazione  
Visione  
Carica liste  
Stampa Carte  
Stampa carte in trascrizione fonetica

**Archivio Etnografico**

Input  
Output

Si mostrano nelle pagine seguenti i più importanti formati di Input usati per l'immissione dei dati.

Archivio Punti -	
<b>Punto</b>	<b>Agrigento</b> N° punto 283
<b>Provincia</b>	ag
<b>Capoluogo</b>	<input checked="" type="radio"/> Si <input type="radio"/> No
<b>Comune con frazioni</b>	<input checked="" type="radio"/> Si <input type="radio"/> No
<b>Frazione</b>	<input type="radio"/> Si <input checked="" type="radio"/> No
<b>Tipo Geografico</b>	2
<b>Tipo Linguistico</b>	6
<b>Popolazione</b>	55.347
<b>Indice SOMEA</b>	1
<b>Indice Urbano</b>	1
<b>Sede di</b>	ecdfgh
<input type="checkbox"/> a = Museo etnografico	<b>Bibliografia</b>
<input type="checkbox"/> b = Fiera	ALI OLS
<input type="checkbox"/> c = Porto	
<input type="checkbox"/> d = Ospedali pubblici	
<input type="checkbox"/> e = Scuole superiori	
<input type="checkbox"/> f = Uffici giudiziari	
<input type="checkbox"/> g = Uffici finanziari	
<input type="checkbox"/> h = Diocesi	
<input type="checkbox"/> z = Risposta nulla	

Formato di Input dell'Archivio Punti

Archivio Informatori	
<b>Cod_Informatore</b>	1
<b>Sesso</b>	M
<b>Età</b>	18
<b>Istruzione</b>	7
<b>Professione</b>	E
<b>Profilo</b>	ade
<input type="checkbox"/>	Disponibilità
<input type="checkbox"/>	Competenza dialettale
<input type="checkbox"/>	Competenza etnografica
<input type="checkbox"/>	Perspiciacia
<input type="checkbox"/>	Autocoscienza linguistica
<input type="checkbox"/>	Risposta nulla
<b>Autovalutazione</b>	
<div style="border: 1px solid black; height: 100px;"></div>	

Formato di Input dell'Archivio Informatori



Archivio Domande	
<b>Domanda</b>	<b>mangiatoia</b>
<b>Concetto</b>	mangiatoia
<b>Tipologia Doman</b>	1
<b>Ct_Grammaticale</b>	2
<b>Fonte Domanda</b>	AIS
<b>Num_Rif_in_Q_Fo</b>	1
<b>Classe Sem_1</b>	2
<b>Classe Sem_2</b>	24
<b>Classe Sem_3</b>	244
<b>Classe Sem_4</b>	2455

**Scelte per : Domande**

- mangiatoia**
- carretto
- forchetta
- boga
- porta
- bambino
- barca
- botte
- cernia
- pentole
- strada
- donna
- bambini

**Annulla**      **Modifica**

Formato di Input dell'Archivio Domande

Archivio Fonti	
Fonte	<b>OLS</b>
Anno rilevamento	1984
Qualità registrazione	4
Supporto	2

**Scelte per : Fonti**

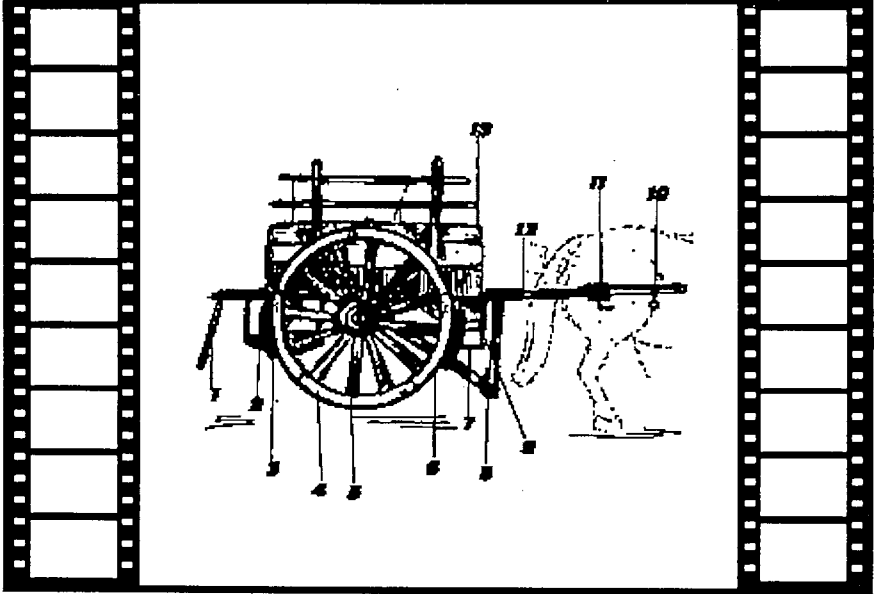
- AIS
- ALI
- ALM
- ALS
- CDI
- OLS
- Altri

**Annulla**      **Modifica**




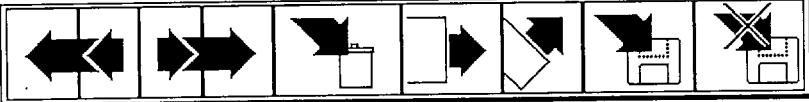
Formato di Input dell'Archivio Fonti

Archivio Etnografico

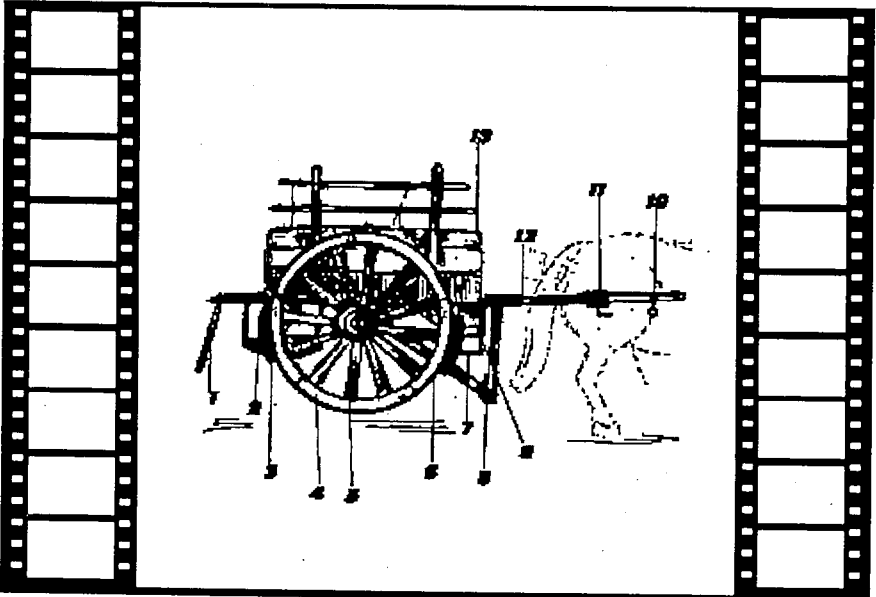
N\_Reperto  Nome reperto



Formato di Input dell'Archivio Etnografico

<b>Modulo Gestione : pagina 1</b>		<b>Dati obbligatori inseribili</b>	
Numero Scheda <b>23</b>			
Codice Rilevatore <b>2</b>		<b>Reperto etnografico</b>	
Codice Informatore <b>5</b>		 	
Fonte AIS	N°Reperto etno. <b>2</b>	<b>Pronuncia Contesto</b>	
Punto linguistico <b>Trapani</b>			
Domanda <b>carretto</b>			
Tr.Fonetica <b>k ā r r è t t u</b>			
Tr.Ortogr. <b>karretto</b>			
Concetto <b>carretto</b>			
			

Formato di Input del Modulo Gestione (pag.1)

Modulo Gestione: pagina 3		Reperto Etnografico	
N_Reperto	<input type="text" value="2"/>	Nome reperto	<input type="text" value="carretto"/>
			





# Appendice 3

*Software*





Tutto il software di supporto al programma da me realizzato per la gestione dell' ALS (descritto in questo lavoro) è stato approntato dalla Text Software, una ditta di Catania (Sicilia, Italia) specializzata esclusivamente nello sviluppo di software in area umanistica.

Eccone un elenco:

Testo. Programma di concordanze in lingue moderne e greco antico  
Testo OCR AutoApprendimento  
Testo OCR Rapido  
TextSoftware Statistica  
TextSoftware GraficaStatistica  
Text.Soft ArchiviazioneDati  
Text Soft Creazione Alfabeti fonetici  
Text.Soft Archiviazione per proiezione  
Text Software: Manutenzione Hard Disk e Antivirus  
Font ALS TextSoftware™  
(completo di 18 formati, stili e dimensioni)

I programmi standard usati sono stati:

Microsoft Word 4 Italiano  
Microsoft Excel 1.51 . Italiano  
DB3 MacIntosh. Ashton Tate  
File Maker IV. Nashoba  
Page Maker III  
Word Perfect  
Mac Spell  
Sil-la  
Poiwer Paint Desk Presentation  
Ready set go 4! Impaginatore

I Data-Base di sviluppo:

4e Dimension (ACI)  
Omnis 5  
Oracle

I Sistemi operativi usati:

System 6.00 e sgg. per MacIntosh  
Unix per Macintosh e IBM  
Xenix per IBM  
OS2 per IBM  
Hypercard e Supercard  
PcSoft Simulazione MAc Iix in IBM PC

I Linguaggi di programmazione:

Think "C"

Think "Pascal"

TMPL Pascal

MPW

Quick Basic

Un software speciale della Farallon™  
è stato usato per la digitalizzazione degli items sonori.

# Appendice 4

## *Hardware*



Stazione attuale di lavoro comprendente:

P.C. MacIntosh II X 5MB RAM

(processore 68030, a 16 Mhz, mouse, 6 slot di espansione,  
porte di comunicazione veloce SCSI,

Connettori NUBUS, Coprocessore matematico 68002)

Hard-disk di backup da 20 a 140 MB

Drive per floppy da 1.4 MB con formattazione MacIntosh o IBM

Scanner Apple 300 dpi, con programmi di lettura ottica dei caratteri

Monitor Colore ad alta visibilità (13 pollici)

Periferiche di stampa ad aghi e laser (Laser SCII)

Driver per lettura di CD. ROM

Rete di collegamento generale mista (UNIX-XENIX-MacIntosh-IBM)

CD-ROM in lettura e scrittura

Unità MacIntosh Plus 1MB

Unità MacIntosh Plus 1MB - hard disk 20 MB

Digitalizzatore di suoni

Apparecchiatura di prossima acquisizione:

P.C. MacIntosh II FX 16MB RAM (processore 68030, a 40 Mhz)

Unità MacIntosh SE 2MB/hard Disk 40MB

Monitor B&N doppio A4 ad alta visibilità (21 pollici)

Plotter a 4 colori

Unità di Backup removibile da 40 MB a cartuccia

Unità di memorizzazione di massa ottico-magnetica da 1.2 GB

Stampante Laser a colori (sino A B3)



# Appendice 5

*Alfabeto fonetico*





L'alfabeto di trascrizione fonetica costruito appositamente per l'ALS è totalmente ispirato ai principi del WISYWYG (cioè l'abbreviazione di "What You See Is What You Get"), finalmente nella sua realizzazione più completa grazie al sistema Apple MacIntosh. Si tratta di un principio secondo cui ciò che si vede sullo schermo corrisponderà esattamente a ciò che la macchina stamperà. Con questo sistema è possibile evitare fastidiose traslitterazioni che affliggono quasi tutti gli attuali sistemi di trascrizione computerizzata dei segni alfabetico-fonetici. L'alfabeto di trascrizione è disponibile in 6 stili (normale **grassetto corsivo spaziato** **grassetto** *spaziato corsivo*) e 3 dimensioni (12, 24, 48 punti).

Nelle pagine che seguono un elenco (non ancora definitivo) dei grafemi utilizzati (per cui si veda Ruffino in questo volume) e l'implementazione nelle tastiere del MacIntosh IIX.

## CONSECUTIVI IN ORDINE DI TASTIERA

a a ā á à â ' à ã \* á , -  
 . / 0 1 2 3 4 5 6 7 8 9 : ; à á á á ä å a  
 a b b č č d đ đ ð e ē é è ê ë ê ê é è é è  
 ě ě ě ě ě ě ě ě ě ě è è è è é é é é é é  
 e e e e ə é f g g ğ ğ γ h h i ī í ì ï ì ì  
 ì ì í ĩ ĩ ĩ ĩ ì ì ì í í í ì ì i i ie j k x  
 χ l l l' l' l m n ñ ñ ñ o o ó ò ò õ õ õ õ  
 õ õ õ õ õ õ õ õ õ õ ò ò ò ò ò ó ó ó ó  
 ó ó ó ó ó ó ö ö p r r r r r r r r r  
 s s s š š ſ ſ t t ð u ū ú ũ ů ů ů ů ů ů ů ů  
 ů ů ů ů ů ů ú ú ú ú ú ú ú ú ú ú ů ů v ŷ ~ ~







Finito di stampare  
dalla Arti Grafiche Siciliane  
Palermo, Ottobre 1990