### DIALETTOLOGIA E INFORMATICA (\*)

#### A. Orientamenti delle ricerche in atto

La ricerca dialettale degli ultimi 25 anni ha trovato nel calcolatore, in Italia e all'estero, un valido strumento per lo studio di testi dialettali e per indagini di tipo geolinguistico e sociolinguistico.

Dai primi anni del '60 (¹) — epoca in cui inizia l'uso del calcolatore in questo campo — le ricerche condotte con l'ausilio dell'informatica sono andate rapidamente estendendosi e, soprattutto in alcuni paesi, sono ormai abbastanza numerose. Non esiste ad oggi una rassegna completa delle attività svolte: i soli repertori di cui siamo a conoscenza sono generalmente parziali, limitati per lo più a una singola area linguistica di indagine e/o ad un particolare tipo di applicazione computazionale (²). Manca un framework generale di riferimento, all'interno del quale collocare e classificare le diverse esperienze. Inoltre, molto spesso, nelle pubblicazioni che illustrano i vari progetti, le metodologie e le modalità di uso del calcolatore trovano spazio così ridotto che, a volte, devono essere inferite dai risultati delle elaborazioni o dalla loro descrizione.

La sola procedura possibile ci appare perciò quella di fornire un rapido excursus, corredato di una traccia bibliografica, su quanto si va sperimentando in questo settore.

Ci limitiamo ad accennare alla ricerca condotta su testi dialettali, letterari e non, in quanto essa fa generalmente ricorso a metodi e tecniche da tempo in uso, negli spogli elettronici di testi, per la produzione di concordanze, indici alfabetici (diretti ed inversi), liste di frequenza, rimari (³). Diamo invece una considerazione più ampia alle applicazioni informatiche nell'ambito della geografia linguistica e della sociolinguistica in quanto individuate, sul piano computazionale, da una propria tipica fisionomia.

<sup>(\*)</sup> La parte A del lavoro è stata scritta da Simonetta Montemagni, la parte B da Antonio Zampolli.

# A.1. Le ricerche geolinguistiche e sociolinguistiche con il calcolatore: il problema della codifica dei dati

La geografia linguistica e la sociolinguistica, in quanto rivolte l'una alla variazione diatopica e l'altra alla variazione diastratica, si trovano a condividere problematiche e soluzioni in rapporto all'uso del calcolatore. Entrambe hanno a che fare con grandi insiemi di dati linguistici — generalmente raccolti mediante inchieste sul campo — da organizzare, classificare e confrontare secondo variabili fonetiche, morfologiche, sintattiche, lessicali da una parte, geografiche, socioculturali dall'altra.

Le caratteristiche quantitative di questi corpora, che implicano spesso un alto numero di entrate (dell'ordine del milione e più), sono da considerare uno dei principali fattori che hanno provocato il ricorso alla elaborazione automatica.

Le caratteristiche qualitative costituiscono un ulteriore importante fattore di indirizzo alla scelta computazionale. Una prima e fondamentale suddivisione delle variabili sulla cui base devono essere ordinati i materiali raccolti, si opera tra il versante linguistico e quello extralinguistico. La presenza di variabili di quest'ultimo tipo, alle quali correlare i dati dialettali raccolti e da cui dedurre l'estensione e la distribuzione dei fenomeni linguistici in una dimensione diatopica e/o diastratica, caratterizza e specifica questo tipo di ricerche nella prospettiva informatica.

Ciò costituisce il quadro metodologico d'insieme in rapporto all'uso del calcolatore. Infatti, con questo duplice ordine di motivazioni viene generalmente "giustificato" il ricorso alla elaborazione elettronica (cfr. per es. Rubin 1969; Hockey 1980).

All'interno di questo schema globale di riferimento, ogni ricerca viene inserita con i propri metodi di raccolta dei dati, con i propri livelli di analisi linguistica, in breve con le proprie finalità. È in rapporto a questi aspetti che nell'ambito di ogni ricerca devono essere stabiliti, in forma leggibile per la macchina, i criteri di codifica dei materiali, linguistici e non, ai fini delle operazioni di ordinamento e di retrieval. Questo comporta la selezione del livello di descrizione e di analisi, ma anche la scelta delle variabili extralinguistiche da associare ad ogni unità linguistica: in altri termini, la traduzione in codice dei fattori pertinenti, adottati nell'ambito di ciascuna ricerca, per la discriminazione di dialetti geografici e/o sociali.

La codifica è particolarmente delicata in rapporto al livello dell'analisi linguistica: qui le scelte operate hanno un risvolto sul piano teorico in quanto è l'operazione di codifica che isola e predispone per le analisi le unità elementari di elaborazione (fonemi, forme lessicali, tipi lessicali, costrutti sintattici o altro). La scelta dei parametri linguistici in base ai quali effettuare l'interrogazione del corpus orienta l'adozione di specifici criteri di codifica: all'indicazione

teorica di massima apertura, cui corrisponde la predisposizione di tutti i possibili approcci in rapporto alla tipologia linguistica del corpus di indagine (Uskup & Al-Azzawi 1972), si affiancano le soluzioni praticamente adottate. Queste possono sintetizzarsi nei punti:

- 1) codifica della trascrizione fonetica;
- 2) rappresentazione ortografica;
- 3) simbolizzazione delle singole entrate mediante numeri, lettere o altro.

Accanto a queste soluzioni si presentano, per gli stessi materiali, casi di codifiche parallele, finalizzate a diversi livelli di analisi (Shuy 1966; Wood 1967).

I livelli di analisi linguistica maggiormente indagati, nell'ambito delle ricerche censite, appaiono quello lessicale e, seppur in misura leggermente minore, quello fonetico-fonologico; i livelli morfologico e sintattico sono affrontati invece in un esiguo numero di casi. Il metodo di raccolta generalmente adottato nelle ricerche esaminate è il questionario, al quale si affiancano sporadici casi di registrazione di parlato spontaneo (per es. Wood 1977). Le soluzioni di codifica, anche se strettamente correlate al tipo di elaborazioni computazionali di cui i materiali codificati costituiscono l'oggetto, sono d'altra parte conseguenza immediata delle tecniche di raccolta dei dati.

Particolare attenzione merita la codifica della trascrizione fonetica. Tutte le ricerche considerate hanno come oggetto di studio la variazione diatopica e/o diastratica della lingua parlata. Già l'oralità come oggetto di studio pone il problema della sua "fissazione" in forma scritta; usando il calcolatore si pone l'ulteriore problema di rendere questa forma scritta leggibile per la macchina e nello stesso tempo suscettibile di tutte le possibili indagini automatiche sul piano fonetico. Nessun calcolatore è in grado di manipolare i simboli IPA o di qualsiasi altro sistema di trascrizione fonetica senza averli preventivamente tradotti. A livello fonetico perciò la codifica consiste nella conversione dei segni della grafia fonetica in forma leggibile per la macchina: la corrispondenza è generalmente ottenuta usando lettere seguite da simboli addizionali. La codifica della trascrizione fonetica è dunque il problema con cui quasi ogni progetto ha dovuto confrontarsi, sia per quanto riguarda l'input, sia, ma più raramente, per quanto riguarda l'output (4). Tra le applicazioni osservate si rilevano codifiche più o meno sofisticate: specifici contributi in proposito sono offerti da Berger (1974), Keil (1974), Hartman (1981), Uskup (1970).

## A.2. Gli indirizzi di uso del calcolatore

Una valutazione delle possibilità di uso del calcolatore deve ignorare le caratteristiche e le finalità specifiche delle singole ricerche, peraltro legate a realtà linguistiche diverse; trova invece nell'analisi e nel confronto di alcune

elaborazioni effettuate, scelte tra le più significative, un'occasione di verifica. Da un punto di vista computazionale, queste elaborazioni possono distinguersi in tre indirizzi in cui il calcolatore è utilizzato per:

- l'ordinamento, la classificazione, la selezione dei dati linguistici raccolti
- l'applicazione di metodi statistici
- la rappresentazione cartografica dei dati raccolti o dei risultati delle analisi.

È analizzando questi settori di applicazione computazionale, integrati spesso all'interno della singola ricerca, che è possibile focalizzare le diverse modalità di impiego del calcolatore nei campi che ci siamo proposti di considerare.

### A.2.1. L'ordinamento, la classificazione e la selezione dei dati linguistici

È il settore di applicazione computazionale che caratterizza la maggior parte delle ricerche censite. Fin dai primi tentativi di trattamento automatico in ricerche geolinguistiche o sociolinguistiche sono emerse le possibilità del calcolatore quale "rapido strumento di ri-classificazione dei dati da ogni punto di vista" (è in questi termini che si esprime Weinreich, 1964).

Disponendo all'interno di una Banca Dati dell'insieme dei materiali dialettali, corredato dai relativi riferimenti extralinguistici, gli obiettivi che possono essere raggiunti con questo tipo di elaborazioni sono:

- 1) la stampa o riproduzione dell'intero corpus di indagine secondo particolari criteri;
- 2) la ricerca e la selezione di dati in base a determinate variabili o combinazioni di variabili per rispondere ad interrogativi linguistici di vario tipo.

Per quanto attiene al primo punto, nell'ambito di ogni ricerca ha un significato centrale la scelta del tipo di ordinamento da attribuire all'insieme dei materiali linguistici raccolti: corrisponde, nella pratica, a fornire le chiavi di lettura del corpus, aprendo per questo particolari prospettive di analisi. Rispetto alla procedura seguita tradizionalmente in modo manuale, sul piano metodologico i vantaggi del trattamento automatico sono essenzialmente:

- la possibilità di classificare parallelamente i materiali raccolti secondo più ordinamenti (invece che in momenti successivi);
- la possibilità di formulare criteri di ordinamento più complessi, manipolando nello stesso tempo più variabili ognuna delle quali assume un peso diverso.

Gli ordinamenti così ottenuti possono divenire base per le interpretazioni linguistiche e punto di partenza per definire i modi e gli indirizzi di un migliore sfruttamento dei dati.

Le possibili configurazioni che l'intero corpus di indagine può assumere

sono molteplici, da valutarsi caso per caso. La difficoltà di individuare una tipologia dettagliata delle diverse applicazioni effettuate può risolversi con una classificazione basata sulla chiave primaria dell'ordinamento. Ogni item raccolto può così essere ordinato secondo:

- l'ordine alfabetico, diretto, inverso, anche partendo da collocazioni interne alla parola (per quest'ultimo caso si veda ad es. Francis & Svartvik & Rubin 1969);
  - la frequenza con cui occorre nel corpus di indagine;
  - lo stato, la regione, la località in cui è stato reperito;
- la classe di età, il sesso, la fascia socioculturale di pertinenza degli informatori che lo hanno testimoniato;
- la domanda di cui costituisce risposta, quando l'acquisizione dei dati sia avvenuta mediante questionario.

Queste sono soltanto le più frequenti, tra le possibili chiavi di ordinamento.

Può essere qui opportuna l'esemplificazione di alcune combinazioni di questi criteri di ordinamento. Concentriamoci sulla chiave primaria "domanda del questionario": il risultato è una sorta di indice che raccoglie, seguendo l'ordine del questionario, le risposte item per item. Ulteriori criteri di ordinamento, che valgono per le risposte alla singola domanda, possono essere la successione alfabetica (Atwood 1962) oppure la frequenza dell'item nell'area di indagine e/o in singole sezioni di questa (Wood 1971a). Ma anche la chiave primaria, non cumulata ad altre, può dare luogo a diversi risultati: si confrontino a questo proposito il Data Summary ed il Response List, due esempi di ordinamento della stessa grande mole di dati raccolta per il Dictionary of American Regional English (Henderson 1974).

Il Data Summary raccoglie tutte le informazioni reperite attraverso le inchieste sul campo: ad ogni domanda sono associate le risposte ottenute, con relativa frequenza, e ad ogni risposta si accompagnano le coordinate che la collocano nelle dimensioni geografica e socioculturale.

Il Response List, uno strumento di documentazione più agile rispetto al precedente, procede dalla definizione — il "concetto" implicato dalla domanda — alla parola, raccogliendo per ogni domanda le risposte ottenute, corredate della relativa frequenza.

Così, partendo dalla stessa chiave primaria di ordinamento, si è constatata la varietà delle possibili configurazioni dei dati.

Sul piano operativo, i vantaggi dell'elaborazione informatica sull'intero corpus dei materiali dialettali sono notevoli in rapporto alla fase di pubblicazione. Una netta riduzione di costi e di tempi è data dalla possibilità di predisporre, attraverso particolari programmi, il formato di stampa dell'output finale; per cui, invece delle tecniche di stampa convenzionali, viene riprodotto direttamente l'output del calcolatore, sia esso in forma di tabulato, di carta

geolinguistica o altro. Su questo aspetto si soffermano, più o meno insistentemente, molti autori, dato che costi e tempi di pubblicazione di un atlante linguistico sono spesso proibitivi (<sup>5</sup>).

Ancora più importanti le potenzialità offerte dal trattamento automatico per l'accesso, la manipolazione e l'interrogazione dei dati. Dopo le operazioni effettuate sull'insieme dei materiali fin qui descritte, la Banca Dati non è certamente ridotta in utilità e funzione (6): essa permette infatti di derivare dall'insieme dei materiali raccolti liste e confronti incrociati di ogni tipo, ottenuti manipolando variabili linguistiche, geografiche, socioculturali. Tre tipi di variabili che, diversamente combinate in un quadro integrato, contribuiscono a delimitare in modo sempre più preciso i confini tra varietà linguistiche differenziate — in una dimensione e diatopica e diastratica — nonché, in certi casi, a individuare rapporti che non emergerebbero facilmente mediante un confronto manuale dei dati.

In questa ottica, tesi a valutare le interferenze tra più variabili, si sono sviluppati numerosi studi. La loro casistica, considerato l'alto numero di possibilità combinatorie, è estremamente ampia e difficile da chiudere in una classificazione, in quanto correlata non solo alla tipologia di ogni indagine, ma anche agli obiettivi del singolo ricercatore. Si può tuttavia proporre una prima fondamentale suddivisione: da un lato combinazioni di sole variabili linguistiche, dall'altro lato di variabili linguistiche ed extralinguistiche.

Ci limitiamo qui ad esporre pochi esempi, non generalizzabili, tutti effettuati sui materiali del Dictionary of American Regional English o D.A.R.E. (7): il primo è da ricondursi al primo tipo di combinazione, gli altri al secondo.

Hirshberg (1981), nell'ambito di un'analisi contrastiva della morfologia regionale, dichiara di aver operato procedendo su due binari:

- a) attraverso l'analisi ed il confronto dei risultati a più domande tra i quali si osservi l'occorrenza di un dato fatto morfologico;
- b) attraverso la ricostruzione di un quadro sinottico dell'estensione relativa a quello stesso fatto morfologico.

Questi due binari, l'uno volto all'articolazione dello spazio dialettale in rapporto a un dato fenomeno linguistico, l'altro all'estensione di quello stesso fenomeno, non sarebbero stati percorribili parallelamente — è lo stesso Hirshberg che lo ammette — senza l'aiuto del calcolatore, considerata l'enorme mole dei dati.

Henderson (1975) studia la distribuzione che caratterizza ogni item raccolto: generalizzata o limitata a specifiche aree geografiche o a particolari soggetti "sociali". Il passo successivo consiste nell'individuazione delle singole variabili geografiche o sociali rispetto alle quali lo stesso item varia e nel confronto della distribuzione geografico-sociale dei diversi item. La risposta ad ognuno di questi interrogativi è rappresentata in carte.

Cassidy (1977) focalizza la propria attenzione sulla variazione areale correlata alla variazione sociale. A tal fine per ogni risposta, codificata e dunque recuperabile in base alla distribuzione geografica (località, area urbana o rurale, ecc.) e socio-culturale (sesso, età, razza, grado di istruzione dell'informatore), è possibile ricostruire il quadro generale rispetto a ciascuna variabile (costruzione di carte per località, sesso, età, ecc.), e anche confrontare e incrociare i singoli quadri.

È soprattutto rispetto ad elaborazioni di questo tipo che si può parlare di un miglior rendimento scientifico dei dati raccolti, in quanto vengono espresse potenzialità altrimenti non esplorate o che sarebbero emerse solo dopo immani sforzi: "because computers do permit almost endless matching and counting, the abundance of possible comparisons should allow researchers to draw conclusions about language that have eluded them before" (Wood 1970a: 108).

## A.2.2. L'applicazione di metodi statistici

L'adozione di procedure automatiche ha portato non solo ad un raffinamento delle tecniche descrittive e di analisi qualitativa, ma anche allo studio e alla definizione di metodi statistici per sintetizzare in termini quantitativi il quadro della variazione dialettale.

Il ricorso a procedimenti statistici nell'interpretazione della realtà dialettale si è registrato prima dell'introduzione dei calcolatori nel campo della geografia linguistica e della sociolinguistica (8). Tuttavia queste ricerche sono state profondamente rinnovate e rafforzate nel momento in cui il trattamento informatico dei dati dialettali ha permesso di accrescere l'estensione e la varietà dei materiali sottoposti all'analisi.

Una delle prime e globali applicazioni di metodi statistici mediante calcolatore è quella effettuata da Houck (1967) alla fine degli anni '60 sul dialetto della Johnson County nello Stato dello Jowa (USA). Le procedure statistiche adottate spaziano da quelle più semplici quali distribuzioni di frequenza, assolute e relative, coefficienti di correlazione, a tecniche più raffinate e complete quali l'analisi fattoriale e della varianza. Sulla base di questi metodi statistici, Houck arriva a concludere che il dialetto della Johnson County, considerato sul piano lessicale, non presenta differenze significative in rapporto né a fattori geografici né socioeconomici (età, sesso, grado di istruzione, professione degli informatori).

Una tecnica statistica che ha trovato applicazioni nel campo dialettale è la cluster analysis, metodo che ha le proprie origini nelle scienze biologiche, dove è tipico il problema del raggruppamento delle specie nei generi e dei generi nelle famiglie. Applicare questa tecnica ai dati raccolti nell'ambito di un atlante linguistico significa misurare la similarità o dissimilarità tra i diversi punti di inchiesta che vengono così raccolti in gruppi "omologhi" sulla base di nume-

rose variabili. È quanto fatto, ad esempio, da Shaw (1974) su un sottoinsieme di dati del Survey of English Dialects (S.E.D.) (9).

L'approccio quantitativo ha utilizzato non solo procedimenti statistici di altre scienze, ma ha portato a sviluppare una vera e propria dialettologia su base quantitativa: la dialettometria, che si è dotata di originali tecniche di analisi con finalità specifiche.

La dialettometria, configurandosi come una branca dell'analisi tipologica (Goebl 1985: 210), tende ad una classificazione delle varietà dialettali — quali risultano da ricerche di tipo geolinguistico — attraverso la rilevazione e la misurazione di concordanze e divergenze linguistiche tra i diversi punti di inchiesta. Le ricerche che ricorrono ad analisi di tipo dialettometrico per la classificazione dei materiali registrati in atlanti linguistici (spesso differenziate per l'adozione di tecniche indipendenti ed originali) presentano problemi che, ad oggi, vengono affrontati e risolti quasi esclusivamente con il calcolatore.

Nell'ambito di queste ricerche ci limitiamo qui ad accennare ai lavori dialettometrici di Goebl, chiara espressione di quanto il calcolatore costituisca uno strumento di lavoro centrale ed ormai irrinunciabile. Il punto di partenza di ogni analisi dialettometrica è una matrice bidimensionale (punti di inchiesta per carte dell'atlante) che raccoglie i dati registrati nell'atlante linguistico in forma codificata; ad ogni carta di atlante possono corrispondere più matrici di dati, a seconda che la codifica — nient'altro che una tipizzazione dei dati originali — abbia seguito un criterio lessicale, morfologico o di altro tipo. Costruita la matrice dei dati, l'analisi è condotta mediante procedure tassonometriche: calcolo di indici di similarità, generazione di matrici di similarità, costruzione di carte di similarità sono alcuni degli strumenti di analisi più frequentemente adottati da Goebl al fine, ad esempio, di calcolare la reciproca distanza che separa alcuni punti dell'Italia settentrionale (Goebl 1978).

# A.2.3. La rappresentazione cartografica

La maggior parte delle elaborazioni finora considerate trova una sintesi visiva nella rappresentazione cartografica. L'apporto del calcolatore rispetto a questo centrale momento di ogni ricerca geolinguistica ha determinato, oltre a indubbi vantaggi sul piano operativo (si riducono i tempi di realizzazione e stampa delle carte, i costi di produzione, ecc.), novità non trascurabili su quello metodologico.

Nell'ambito delle ricerche geolinguistiche condotte con l'ausilio del calcolatore, la carta assume una nuova valenza: il disporre del corpus dialettale memorizzato in una Banca Dati consente la compilazione automatica di carte non solo al momento finale ma anche nel corso della stessa ricerca. In breve, è possibile sperimentare più tipi di carte, individuando quelle che di volta in volta si presentano come le più interessanti o le più significative: tutto questo aggiunge una nuova dimensione allo studio del dialettologo per il quale la carta non si configura più soltanto come elaborato finale ma anche come centrale strumento di lavoro (cfr. Rubin 1970).

La carta, in questa prospettiva, assicura l'immediata intelligibilità delle ricerche e delle analisi condotte sul corpus di indagine. Analisi che, divenendo sempre più complete ed articolate per l'estensione dei dati e il numero delle variabili considerate, hanno richiesto la messa a punto di nuove tecniche di visualizzazione della variazione dialettale. Di queste ne accenniamo alcune, tra le più significative, focalizzando l'attenzione sui tipi di variabili linguistiche che permettono di inserire nella rappresentazione cartografica:

- Stereogrammi: il ricorso a questo tipo di rappresentazione nello spazio tridimensionale permette, per ogni punto di inchiesta (individuato sul piano dalle relative coordinate), di riportare sulla terza dimensione valori numerici che, ad esempio, si riferiscano alla frequenza di uso di uno o più item dialettali (Southard 1983). La costruzione di queste carte tridimensionali è estremamente difficile senza l'ausilio del calcolatore;
- Glottogrammi: particolare tipo di visualizzazione messo a punto da Ogino (1982, 1984) nell'ambito del Generalized Linguistic Atlas Printing System (G.L.A.P.S.) (10) al fine di inserire la variazione sociolinguistica nella rappresentazione cartografica. Agli assi cartesiani sono associate rispettivamente una variabile socioculturale ed una geografica; le coordinate individuanti ogni risposta (rappresentata mediante simboli) la qualificano, quindi, in questa duplice prospettiva.
- Carte coropletiche: utilizzate da Goebl per visualizzare le distribuzioni di similarità tra i diversi punti di inchiesta rispetto a un punto di riferimento. Attorno alle singole località, attraverso adeguate costruzioni geometriche, vengono costruite aree disgiunte tra loro e non sovrapposte, il cui tratteggio simbolizza le misurazioni ad esse connesse. Ciò al fine di evitare una rappresentazione per punti e di ottenerne invece una per superfici continue giustapposte (cfr. per es. Goebl 1982).
- Le carte del D.A.R.E.: il sistema di rappresentazione cartografica adottato nell'ambito del D.A.R.E. fa ricorso a tecniche di deformazione dei confini geografici in funzione della densità delle comunità oggetto di inchiesta in ciascuno Stato degli U.S.A.: vengono comunque mantenute la forma e la posizione reciproca degli Stati. Ciò permette di far risaltare con chiarezza la distribuzione delle risposte tra le diverse comunità (cfr. Cassidy 1977; Lance & Slemons 1976).

Tra le numerose e varie possibilità di rappresentazione cartografica offerte dal calcolatore, gli esempi riportati hanno la sola funzione di fare intravedere nuove prospettive metodologiche e di analisi conseguenti all'adozione di procedure automatiche anche nell'ambito della cartografia linguistica.

#### B. Prospettive di sviluppo

I progressi metodologici della linguistica computazionale, la recente evoluzione tecnologica, la tendenza crescente alla cooperazione nazionale e internazionale, suggeriscono alcune considerazioni sulle prospettive di ricerca e di sviluppo organizzativo delle nostre discipline.

#### B.1. Analisi linguistica

I materiali raccolti con le diverse forme di indagine tradizionalmente adoperate dalle nostre discipline consentono, in linea di principio, ricerche ai vari livelli del sistema linguistico: fonologico, morfofonologico, lessicale, sintattico, ecc.

Le elaborazioni tradizionali dei materiali hanno necessariamente come oggetto la rappresentazione delle unità linguistiche che occorrono nel corpus: si conteggiano le frequenze di fonemi, tipi lessicali, costrutti sintattici; si pubblicano contesti relativi a lemmi, forme, categorie grammaticali; si mettono in carta forme o sintagmi corrispondenti a una determinata unità lessicale; ecc.

È ovvio che un corpus di dati analizzati a vari livelli linguistici permette ricerche e interrogazioni molto più utili e interessanti.

Analizzare un corpus richiede, sul piano metodologico, che si disponga di un quadro teorico capace di fornire un insieme adeguato di criteri espliciti per definire le unità proprie di ciascun livello, una procedura per riconoscerne le occorrenze, un formalismo per rappresentarle. Sul piano operativo, la analisi deve essere eseguibile in tempi e con costi accettabili.

È evidente che la situazione è diversa per i diversi livelli linguistici. Non a caso gli spogli di testi condotti fino ad oggi sono eseguiti per lo più a livello lessicale. Le operazioni di analisi in questo caso risultano relativamente semplici e possono essere addirittura, in un certo senso, omesse, perché la parola, così come è definita per il calcolatore (una sequenza di caratteri tra due spazi), può costituire, in mancanza di altro, una approssimazione accettabile — almeno operativamente — delle unità lessicali oggetto delle ricerche (11).

La elaborazione di un testo parlato, sia esso costituito dalla registrazione di una conversazione o dalle risposte a un questionario, inizia necessariamente con la sua trascrizione, la quale coincide con la analisi, dal momento che costruisce concretamente il corpus rappresentandone direttamente le unità fonologiche che lo costituiscono. Se invece si elaborano testi scritti della letteratura dialettale o regionale, il riconoscimento e la rappresentazione delle unità fonologiche costituisce una fase a se stante, e la possibilità di (semi)automatizzarla dipende strettamente dal tipo di lingua o di dialetto specifico (Zampolli 1960 e 1968).

Ben diversa è la situazione per altri livelli linguistici. Si considerino, a titolo di esempio, i problemi connessi al livello sintattico:

- esiste una molteplicità di teorie in base alle quali operare la descrizione sintattica;
- nessuna o pochissime delle teorie più recenti è stata sviluppata al punto da consentire la descrizione esaustiva, non di frasi o esempi isolati, ma di un corpus 'reale' in linguaggio naturale;
- la rappresentazione della struttura sintattica, associata ai componenti di un enunciato, può essere estremamente complessa, non fosse altro che dal punto di vista della laboriosità della notazione (12);
- il tempo e il costo di esecuzione della analisi sintattica di un corpus sono estremamente elevati: prova ne sia il fatto che non esistono a nostra conoscenza, nel mondo, se non rarissimi esempi di corpora analizzati a questo livello;
- in considerazione dei tempi e dei costi necessari per analizzare un corpus a livello sintattico, è inevitabile chiedersi se sia o no possibile analizzarlo in modo per così dire teoricamente 'neutrale': in modo tale, cioè, che le categorie adoperate possano essere utilizzate anche da ricercatori che operano nell'ambito di teorie diverse.

Considerazioni analoghe possono essere richiamate, in maggiore o minore misura, per altri livelli linguistici.

Quanto abbiamo detto, suggerisce alcune considerazioni di principio.

a) Per il momento, è prudente supporre che la analisi linguistica associata ai materiali non possa essere concepita come una struttura definita una volta per tutte. Il corpus deve essere "arricchito" con analisi condotte progressivamente, ed eventualmente incrementabili e modificabili dinamicamente. Non si propone di rimandarne la utilizzazione. Si auspica, invece, che i dati raccolti vengano resi immediatamente disponibili senza attendere il compimento delle analisi, e anzi rendendo disponibili anche queste, mano a mano che vengono eseguite.

È possibile che l'esame progressivo dei materiali raccolti possa fornire elementi per la modificazione dei criteri teorici e metodologici di analisi.

In altre parole, non sembra consigliabile mantenere rigidamente la successione tradizionale delle operazioni: raccolta dei dati; classificazione ed analisi; pubblicazione; utilizzazione. Sembra piuttosto che ci si debba orientare verso un framework operativo che consenta, non solo ai raccoglitori, ma all'intera comunità degli studiosi, di accedere ai dati il più presto possibile, non appena raccolti.

b) Le analisi linguistiche potranno essere effettuate progressivamente, non solo dai raccoglitori, ma anche dagli utilizzatori. L'utente assumerebbe così un ruolo attivo. L'analisi e la descrizione del corpus potrebbero diventare il risultato progressivo di uno sforzo cooperativo, ed evolvere con l'evolvere

delle conoscenze. La base di dati dovrebbe essere considerata, se ci è consentita la metafora, anche come una 'banca': non solo come una fonte dalla quale si prelevano dei dati, ma anche come un archivio comune nel quale ciascuno deposita le proprie analisi, le quali divengono immediatamente disponibili agli altri, contribuendo così al graduale arricchimento di un patrimonio comune.

- c) Sembra auspicabile che si giunga a stabilire, mediante accordi di collaborazione scientifica e tecnica, un insieme di criteri che assicurino la compatibilità non solo della trascrizione e registrazione in machine-readable form dei materiali, ma anche della loro descrizione ed analisi.
- d) La collaborazione tra linguisti e linguisti computazionali si è rinsaldata, in questi ultimi tempi, attorno ad alcuni temi comuni, ed in particolare:
- Confrontare le diverse teorie per valutare la possibilità di descrivere testi ed unità linguistiche in modo "neutro", o "politeorico". Basi di dati che, per le loro dimensioni, richiedono costi e tempi di costruzione considerevoli (in particolare corpora e lessici mono e plurilingui), dovrebbero essere (ri)utilizzabili in diversi frameworks teorici e computazionali, e non solo nella prospettiva di una scuola o di una applicazione determinata (13).
- Studiare e implementare metodi, componenti, procedure che facilitino e, al limite, automatizzino la analisi di testi ai diversi livelli linguistici: lessici automatici, analizzatori sintattici (di sintagmi o di frasi), sistemi per la disambiguazione assistita dal calcolatore, ecc.

Sembra auspicabile che le attività computazionali nel settore della dialettologia, della geografia linguistica, della sociolinguistica, si associno, più che per il passato, a questa collaborazione. Esse non solo possono derivarne contributi metodologici e operativi, ma possono contribuire ad arricchire la tipologia dei problemi comuni con esigenze specifiche.

# B.2. Organizzazione dei materiali in basi di dati

I progetti in corso, con o senza l'ausilio del calcolatore, di cui abbiamo conoscenza diretta, o che abbiamo trovato citati nella letteratura esaminata, sono caratterizzati, come si è detto, dalla tecnica adottata per la raccolta dei materiali.

Si elaborano testi dialettali pubblicati come opere letterarie, corrispondenze private, raccolte lessicografiche, risposte (parole singole, sintagmi complessi, intere frasi o periodi) a questionari, conversazioni registrate, spontanee o elicitate in vario modo, risposte a tests che presentano scelte tra elenchi di forme linguistiche alternative; ecc.

Talora della stessa area si occupano progetti diversi, ciascuno con un diverso metodo di raccolta.

Recentemente, alcuni progetti hanno fatto, o si propongono di fare, ri-

corso contemporaneamente a diversi tipi di raccolta.

Paradigmatico ci sembra il caso del progetto NADIR, promosso da A. Sobrero (1985).

Ogni tipo di raccolta fornisce una propria classe di dati (che per semplicità chiameremo genere): testi, lessici, elenchi di parole, coppie domanda-risposta, ecc.) caratterizzata da strutture proprie, che richiedono metodi di rappresentazione e procedure di elaborazione spesso assai diverse.

Le tecniche tradizionali hanno, per così dire, "istituzionalizzato" forme specifiche e distinte di documentazione e diffusione per ogni genere: concordanze e frequenze per i testi; indici e carte per le risposte a domande; statistiche per i tests a scelta multipla; ecc.

I materiali ottenuti con le diverse forme di rilevamento dovrebbero invece essere considerati complementari tra loro, ed integrati in una base di dati. Lo studioso dovrebbe poter rintracciare la documentazione del fenomeno o dell'insieme dei fenomeni che lo interessano nell'intero corpus dei materiali disponibili per una data area. In altre parole, egli dovrebbe poter accedere contemporaneamente ai risultati dei diversi tipi di inchiesta, anche se ciascuno di essi richiede, per la propria struttura, una diversa tecnica computazionale di accesso.

Per questa via, si entra direttamente nel settore, in rapida espansione, delle basi di dati linguistiche "multigeneri" e "multifunzionali", che vengono oggi proposte dalla linguistica computazionale.

"We use the expression Linguistic Data Base (LDB) to define a set of linguistic data of different types (not only texts and dictionaries, but also, for example, the results of socio-linguistic studies, linguistic models, bibliographic data, ecc.) finalized for the interactive utilization by multiple categories of potential users, stored, structured and linked for this purpose, and associated to specialized software modules for access, interrogation and on-line processing". (Zampolli 1983: 263).

Il disegno di modelli di LDB è oggi materia di ricerca e di sviluppo presso alcuni dei maggiori centri specializzati nel settore della linguistica computazionale.

Le ricerche mirano ad integrare conoscenze e tecniche sviluppate in settori diversi: lessicologia computazionale, text-processing, statistica linguistica, trattamento delle immagini, parsers, ecc. Riassumere sia pur brevemente le acquisizioni o — per così dire — lo stato dell'arte raggiunto nei diversi settori è qui ovviamente impossibile. Si rinvia ad alcune opere di sintesi (Zampolli 1975, Hockey 1980, Grishman 1986).

La concezione generale e la progettazione dettagliata di un sistema informativo di questo tipo per il settore dialettale costituisce un importante tema di ricerca. Sembra necessaria una fase iniziale di studio, intesa ad esaminare la riutilizzabilità di componenti e metodi approntati o in corso di approntamento

in altri settori della linguistica computazionale. Non si può ignorare che Centri specializzati, i quali hanno raccolto grossi archivi di testi in machine-readable form, stanno sperimentando metodologie per rendere tali corpora disponibili alla comunità nazionale e internazionale dei ricercatori mediante la consultazione interattiva. I loro sforzi sono concentrati principalmente in due direzioni:

- trovare una struttura per la memorizzazione dei dati testuali che costituisca un compromesso ottimale tra lo spazio di memoria occupato (e quindi il costo di conservazione on-line dei dati) e il tempo di accesso ai dati (e quindi il costo della singola interrogazione da parte dell'utilizzatore)
- "immaginare" e realizzare una serie di funzioni di accesso, che consentano all'utente della LDB di interrogare in vario modo, a seconda dei propri interessi scientifici e professionali, i testi. Per esempio: cercare nel corpus, o in suoi sottoinsiemi definiti a piacere (per genere, datazione, ecc.), una parola, una sequenza di lettere in una posizione determinata (all'inizio, in fine, all'interno di parola, ecc.), sequenze contigue di due o più parole, cooccorrenze di due o più parole in un dato intervallo del testo (stessa frase, o stesso paragrafo, o a una distanza massima prestabilita, ecc.). La risposta del sistema è costituita di solito dalla frequenza dell'elemento cercato, dall'index locorum delle sue occorrenze, o dai contesti costruiti con criteri e lunghezze definiti di volta in volta dall'utente (Picchi, Calzolari 1986).

Queste strategie di accesso dovrebbero però essere arricchite per tenere conto di caratteristiche specifiche delle basi di dati dialettali.

Il punto, a nostro avviso, cruciale, sta nel disegnare una struttura che metta in relazione informazioni e materiali appartenenti a generi e a varietà linguistiche diverse (lingua nazionale, regionale, varietà dialettali).

In un primo tempo possono essere utilizzate soluzioni piuttosto semplici: ad esempio, si possono creare funzioni di interrogazione che ricerchino una stringa di caratteri contemporaneamente nei diversi generi memorizzati separatamente nella base dati.

La soluzione cui, secondo noi, si deve tendere, dovrebbe consistere nel creare una struttura centrale nella quale siano elencate, livello per livello, le unità linguistiche proprie di ciascuna varietà linguistica, accompagnate da descrizioni delle loro proprietà sintagmatiche e relazioni paradigmatiche. A queste unità dovrebbero essere collegate, per mezzo di un opportuno sistema di rinvii, le rispettive occorrenze nei corpora di diverso genere presenti nella base dati. Le unità delle diverse varietà linguistiche potrebbero poi essere messe in corrispondenza per mezzo di relazioni bilaterali, analoghe, per così dire, a quelle che collegano i "translational equivalents" in un dizionario bilingue o multilingue. Oppure, indirettamente, mediante la riconduzione a "tipi" astratti o alle unità linguistiche appartenenti ad una varietà prescelta come "pivot", o nodo centrale della rete di collegamento (per es. la lingua nazionale) (14).

Su dati così organizzati, si potranno applicare le metodologie con cui si riesce, oggi, a potenziare le strategie di ricerca dell'utente. Si tratta per lo più di strumenti linguistico-computazionali che hanno la funzione di "fonti di conoscenza" complementari alle conoscenze dell'utente: dizionari di macchina, mono e bilingui, regole grammaticali e parsers locali, ecc. (15).

#### B.3. Coordinamento, collaborazione, scambi

a) In molti paesi, sono in corso progetti che si occupano di aree più o meno strettamente interrelate, ed è evidente che la comunità dei ricercatori, nei limiti del possibile, dovrebbe poter accedere contemporaneamente ai dati raccolti separatamente dai diversi progetti.

Affinché ciò sia concretamente possibile, sembra necessario stabilire alcuni fondamentali accordi metodologici fra i vari progetti e centri di ricerca.

Questi dovrebbero avere come oggetto almeno una certa "normalizzazione" della rappresentazione dei dati, linguistici ed extralinguistici, della strutturazione della base di dati, dei criteri di analisi, del disegno delle procedure di interrogazione e dei programmi di elaborazione.

Non ho usato il termine "standards", perché l'esperienza in campi analoghi, quali la raccolta di corpora di testi letterari o la costruzione di basi di dati lessicali, ha mostrato che fattori di diversa natura ostacolano di fatto la creazione e la adozione di standards rigidamente definiti da parte di progetti in corso da tempo, i quali non possono o non vogliono, anche per ragioni economiche, assumersi il carico di mutare, "in medias res", i propri sistemi di codifica.

Riteniamo, tuttavia, che non si possa, in questi casi, rinunciare a uno sforzo comune per definire un metalinguaggio che consenta di descrivere univocamente e chiaramente le caratteristiche di ciascun sistema. Tale descrizione è assolutamente necessaria per studiare delle procedure che consentano di compiere, sulla rappresentazione dei dati, le trasformazioni formali necessarie per renderli omogenei, almeno per quanto concerne le operazioni di consultazione dell'utente.

Nel caso di progetti che stanno per iniziare, o che ancora non hanno avviato la memorizzazione dei dati, ci sembra invece importante una cooperazione, almeno per estese aree geografiche o linguistiche, che conduca al disegno e alla adozione di veri e propri standards. Il progetto CHILDES, che si propone di unificare in una unica base di dati campioni di linguaggio infantile raccolti in diversi paesi, può costituire un esempio e nel contempo una dimostrazione della fattibilità di un coordinamento di questo tipo. Il progetto ha messo a punto un sistema per la trascrizione e la analisi di conversazioni il quale, accanto a un nucleo fondamentale di norme standardizzate, consente ai

singoli progetti l'aggiunta e l'inserimento di norme particolari per trattare fatti ad essi peculiari.

La adozione di uno standard comune consente, oltre alla interscambiabilità dei dati e al collegamento delle basi di dati in un unico "network" di consultazione, la riutilizzabilità di procedure e programmi messi a punto da uno qualsiasi dei membri del network, con benefici economici e risparmio di risorse facilmente immaginabili. Condizione indispensabile ci sembra non solo la consultazione e la informazione reciproca, ma anche la creazione di un gruppo di coordinamento che si incarichi di disegnare le metodologie e di realizzare gli strumenti computazionali di base. Questo gruppo dovrebbe includere, accanto agli specialisti responsabili di progetti, anche esperti di linguistica computazionale.

b) È un fatto purtroppo noto che la utilizzazione dei dati, secondo le tecniche tradizionali, è possibile in genere solo molto tempo dopo la raccolta. Ne risulta fra l'altro che spesso la situazione rispecchiata al momento delle inchieste non è più attuale al momento della utilizzazione. Le tecniche tradizionali di documentazione e diffusione dei dati richiedono una lunga fase di elaborazione come premessa alla pubblicazione.

Ci sembra invece, come si è detto, che i materiali raccolti dovrebbero essere messi a disposizione della comunità degli studiosi il più presto possibile, senza attendere il completamento della loro elaborazione. Non si vuole negare la utilità della pubblicazione tradizionale dei risultati. Si afferma invece che la possibilità di accedere tempestivamente ai dati, a mano a mano che vengono registrati in machine-readable form, consente di beneficiare della flessibilità dei linguaggi d'interrogazione, che consentono di definire, di volta in volta, a ciascun ricercatore combinazioni e incroci di parametri extralinguistici e di categorie linguistiche, a seconda degli obiettivi specifici della propria ricerca, ed eventualmente di ridefinirli ricorsivamente fino a trovare la configurazione desiderata. Le tecniche di documentazione tradizionali invece, pur utilissime, rendono accessibili i dati solo pubblicandoli ordinati definitivamente secondo criteri scelti una volta per tutte (16).

Il compito di rendere rapidamente accessibili i dati raccolti in machinereadable form a una comunità di studiosi distribuiti in località diverse, è indubbiamente facilitato dallo sviluppo di nuove tecnologie, che rendono disponibili, presso il singolo ricercatore, e a costi rapidamente decrescenti, personal computers dotati di potenza di calcolo più che adeguata, collegati a supporti, anch'essi a basso costo e di dimensioni ridottissime, che consentono di memorizzare e gestire basi di dati di dimensioni largamente superiori a quelle tipicamente richieste dalle nostre ricerche.

Fino ad oggi, lo scenario organizzativo di una base di dati di questo tipo prevedeva la memorizzazione dei dati nelle memorie di massa di un grosso centro di calcolo (main-frame), dotato di apparecchiature capaci di mettere a disposizione grandi quantità di dati per la consultazione on-line. Gli utilizzatori accedevano ai dati mediante una rete di terminali collegati telefonicamente al main-frame, anche da località lontane. Questa organizzazione può essere oggi ancora da preferire durante il periodo di raccolta. Se la trascrizione avviene direttamente su supporto magnetico (per es., per mezzo di un PC), i dati possono essere consultati immediatamente a distanza, a mano a mano che la raccolta procede.

Una volta però che una quantità adeguata di dati sia stata trascritta, questi possono essere trasferiti facilmente su una memoria compatta scrivibile direttamente dal PC, mentre gli eventuali aggiornamenti possono essere registrati su supporti a bassissimo costo, riscrivibili più volte, quali ad esempio i floppy disks.

Quando il corpus raccolto abbia raggiunto una certa dimensione e un certo grado di completezza e stabilità, esso potrebbe essere distribuito su compact-disk (CD-ROM). Un compact disk di 12 centimetri contiene più di 600 milioni di caratteri. La spesa più elevata (circa 15 milioni) è quella per la 'masterizzazione', e cioè per il trasferimento dei dati sulla matrice, dalla quale può essere "prodotto" un numero qualsivoglia di copie, con una spesa irrisoria (dell'ordine di dieci dollari per ciascuna copia).

Poiché un compact-disk può contenere, tipicamente, 15-20 corpora dialettali, la spesa di masterizzazione potrebbe essere suddivisa tra vari progetti.

Una stazione di lavoro ideale per un ricercatore (dialettologo, linguista, o umanista in genere) può essere costituita, nel quadro tecnologico attuale, da un PC di medie dimensioni, da un lettore di compact-disk, da una stampante che permetta di programmare i caratteri necessari. Eventualmente, ma non necessariamente, il PC potrebbe essere collegato a una rete di "posta elettronica" e a un "main-frame", per consentire il "dialogo" con gli altri ricercatori in tempo reale nel corso della utilizzazione dei materiali. Il costo di una tale stazione di lavoro rientra certamente nei limiti delle possibilità individuali di molti ricercatori.

Il punto cruciale va riconosciuto invece nella disponibilità di basi di dati di qualità adeguata ed organizzate opportunamente (<sup>17</sup>), e nello sviluppo di programmi di elaborazione e di metodi di accesso multiplo ai dati, adatti alle esigenze dei ricercatori umanisti.

La creazione di questi programmi e metodi richiede uno sforzo per analizzare e descrivere le operazioni e le interazioni con i dati che costituiscono il know-how comune a ciascuna categoria specifica di ricercatori.

#### NOTE

- (1) Tra i primi contributi apparsi si segnalano: Atwood (1962), Weinreich (1964), Wood (1962, 1963).
  - (2) Ci riferiamo in modo particolare a:
- Schneider & Viereck (1984) che raccoglie e descrive brevemente 30 progetti di ricerca limitati alla dialettologia inglese, americana e britannica;
- Hockey (1980) che nel capitolo 4, dal titolo Vocabulary Studies, Collocations and Dialectology, descrive, inserendoli in un quadro complessivo, i contributi di Shaw (1974), Lance & Slemons (1976), Rubin (1970), Crawford & Jones (1976).
- Goebl (1983), Händler (1983), Putschke (1983) che ricompongono le varie esperienze di dialettologia computazionale in base al filone di ricerca nel quale si inquadrano (rispettivamente analisi dialettometriche, definizione di un sistema informativo dialettale, cartografazione automatica).
- (3) Per una sintetica rassegna degli studi su testi dialettali condotti con l'ausilio del calcolatore si rinvia a Cortelazzo, 1984.
- (4) Nel caso della rappresentazione a livello di output la soluzione sta nel tornare ai segni della grafia fonetica di partenza: cfr. Churchhouse & Hockey (1971), Crawford & Jones (1976).
- (5) Si segnalano in particolare i contributi di Wood (1971a) e Uskup & Al-Azzawi (1972) che, come risulta dagli stessi titoli Why not the computer as editor? e Editing and printing a dialect atlas by computer si occupano in modo particolare di questo problema.
- (6) Si riporta, a titolo del tutto indicativo, quanto afferma Hartman (1969: 197) in proposito: "it is Cassidy's hope that the information in the DARE files will not be left to gather dust after the pubblication of the dictionary".
- (7) La Banca Dati del Dictionary of American Regional English o D.A.R.E. è costituita da 2.500.000 di risposte da parte di 2.752 informatori ai quali è stato somministrato un questionario di 1.847 domande (prevalentemente di tipo onomasiologico) in 1.002 comunità degli Stati Uniti tra il 1965 e il 1970. Gli informatori variano per età, sesso, razza, grado di istruzione, tipo di comunità (la tipologia delle comunità prevede cinque classi, tra rurale e urbana). Per una descrizione del progetto nelle sue finalità, nei suoi metodi e nei primi risultati si rinvia a Cassidy (1969, 1977a, 1977b), Henderson (1974, 1975), Hirshberg (1981).
- (8) Si vedano, ad esempio, i seguenti contributi: Reed, D.W., Spicer, J.L., Correlation methods of comparing idiolects in a transition area, "Language", 28, 1952, pp. 210-215; Seguy, J., La dialectométrie Dans l'Atlas Linguistique de la Gascogne, "Revue de linguistique romane", 37, 1973, pp. 1-24.
- (°) Orton, H., Halliday, W., Survey of English Dialects, The Basic Material, E.J. Arnold, Leeds, 1969-71.
- (10) Per una caratterizzazione del progetto si rinvia a Ogino (1982, 1984) e Ogino & Sibata (1977).
- (11) Ovviamente, non si mette in discussione in alcun modo la utilità di lemmatizzare, qualora sia possibile sul piano economico, il corpus (cfr. Zampolli 1983).
- (12) Si pensi, per esempio, alla complessità dell'eventuale uso di parentesi incassate per linearizzare la rappresentazione di una struttura sintattica.
- (13) Per l'analisi di testi italiani scritti sono già disponibili, o potranno essere disponibili in futuro, parsers sintattici interamente o parzialmente automatici, i quali potrebbero ridurre conside-

- revolmente tempi e costi dell'analisi (Cfr. Cappelli, Ferrari, Moretti, Prodanoff, Stock 1980). Nulla esiste invece per quanto concerne il linguaggio orale e, in particolare, materiali dialettali. Per l'inglese orale, si veda Crystal 1985.
- (14) Zampolli ha promosso, nel 1986, un gruppo di studio cui partecipano studiosi delle Università di Stanford, MIT, Princeton, Heidelberg, Cambridge, Cardiff, in rappresentanza di diverse teorie linguistiche: GB, LFG, CF, GPSG, systemic grammar, dependency grammar. Il gruppo, cui partecipano anche rappresentanti di case editrici lessicografiche e di industrie informatiche, si propone di provare la possibilità di creare dei lessici automatici "neutrali", le cui informazioni sono utilizzabili, con opportune interfacce, da diverse teorie linguistiche e frameworks computazionali.
- (15) Si avrebbe così uno strumento analogo a quello auspicato da sociolinguisti e psicolinguisti che vogliano condurre le ricerche in parallelo su corpora equivalenti di diverse lingue. Essi scelgono una lingua "pivot" (di solito l'inglese, alla quale vengono collegate le parole delle altre lingue), che serve per l'interrogazione di corpora multilingui. L'operazione di ricondurre le voci dialettali a un termine "pivot" italiano, che i ricercatori dell'ALI hanno chiamato "tipicizzazione", è giustificata, ovviamente, solo in quanto strumento di "retrieval" dei dati.
- (16) Presso l'ILC di Pisa alcuni esperimenti hanno dimostrato l'utilità dell'impiego di un lessico automatico per rendere più efficace la interrogazione di un corpus testuale. Se si dispone di un dizionario macchina opportunamente strutturato e corredato di procedure "intelligenti" per il calcolo delle relazioni tra gli elementi del dizionario (lemmi, categorie grammaticali, definizioni, tratti semantici, ecc.), è possibile ricercare nel testo occorrenze e cooccorrenze di forme o di lemmi connessi da relazioni linguistiche di vario tipo.
- (17) Esempi di basi di dati linguistici in CD-ROM sono il Corpus del Thesaurus Linguae Graecae, disponibili anche a Pisa presso l'ILC, e l'Oxford English Dictionary, di prossima distribuzione.

#### BIBLIOGRAFIA

- Agostiniani, L., T. Poggi Salani 1985 "Dalle inchieste all'Atlante: per la costituzione di una banca dati". In: AA.VV., Atlante Lessicale Toscano Presentazione. Firenze, Olschki.
- Agostiniani, L., S. Montemagni, T. Poggi Salani in stampa "Atlante Lessicale Toscano: il lavoro di preedizione e la costituzione della banca dati", Atti del XV° Convegno di Studi Dialettali Italiani "Atlanti regionali: aspetti metodologici, linguistici ed etnografici" (Palermo, 7-11 ottobre 1985).
- Albano Leoni, F. 1970-71-72 Concordanze belliane con lista alfabetica, lista di frequenza, lista inversa e rimario, 3 voll. a numerazione unica. Göteborg, Acta Universitatis Gothoburgensis.
- Allen, H.B. 1981 "Regional Dialects, 1945-1974". American Speech 52: 163-261.
- Altmann, G. 1977 "Zur Ähnlichkeitsmessung in der Dialektologie". Germanistische Linguistik 3-4: 305-310.
- Atwood, E.B. 1962 The Regional Vocabulary of Texas, Austin, Univ. of Texas Press.
- Bart Rossebastiano, A. 1978 "L'edizione dei vocabolari veneto-tedeschi del secolo XV". La Ricerca Dialettale 2: 91-96.

- Bart Rossebastiano, A. (a cura di) 1983 Vocabolari veneto-tedeschi del secolo XV. Savigliano, Ediz. L'Artistica Savigliano.
- Berger, J. 1974 "Computing and Phonology". Association for Literary and Linguistic Computing Bulletin 2,1: 4-6.
- Berger, J. 1980 The Dialect of Holy Island. A Phonological Analysis. European University Studies, Series XIV, Anglo-Saxon Language and Literature. Bern, Lang, 83.
- Bershin, H. 1977a "Was leistet die elektronische Datenverarbeitung in der Dialektologie?". Germanistische Linguistik 3-4: 107-120.
- Bershin, H. 1977b "Probleme der Datenaufbereitung in der Automatischen Sprachkartographie". Germanistische Linguistik 3-4: 159-162.
- Brook, R.J. 1970 Computer Reduction and Retrieval of Dialect Data: a Methodology for Cartographic and Tabular Display of Lexical and Sociolinguistic Variation. Ph. D. Univ. of Jowa.
- Cappelli, A., G. Ferrari, L. Moretti, I. Prodanoff, O. Stock 1980 "Analisi di un testo italiano con un analizzatore ATN". Rivista di Informatica 10,3: 5-13.
- Cassidy, F.G. 1967 "American Regionalism and the Harmless Drudge". Pubblications of the Modern Language Association, 82: 12-19.
- Cassidy, F.G. 1969 "Dialectology and the Electronic Drudge". In: S. Ellis (a cura di) Studies in Linguistics in Honour of Harold Orton on the Occasion of His Seventith Birthday. Leeds Studies in English. N.S.2: 135-143.
- Cassidy, F.G. 1969 "Dictionary of American Regional English". Computers and the Humanities 3: 285.
- Cassidy, F.G. 1977a "Computer-Aided Usage 'Labeling' in a Dictionary". Computers and the Humanities 11: 89-99.
- Cassidy, F.G. 1977b "On-line Mapmaking for the Dictionary of American Regional English". Germanistische Linguistik 3-4: 107-119.
- Churchhouse, R.F., S. Hockey 1971 "The Use of an SC4020 for Output of a Concordance Program". In: *The computer in literary and linguistic research*. Cambridge, Cambridge U.P.: 221-231.
- Clivio, G.P., M. Danesi 1974 Concordanza linguistica dei "Sermoni Subalpini". Torino, Centro Studi Piemontesi.
- Cortelazzo, M. 1984 "La dialettologia quantitativa in Italia", in Goebl 1984: 1-14.
- Crawford, T.D., G.E. Jones 1976 "Automated Concordancing of Welsh Dialects with Output in the IPA". Bulletin of the Board of Celtic Studies 27: 45-50.
- Crystal, D. 1979 Working with LARSP. London, Edward Arnold: 12-14.
- Crystal, D. 1981 Clinical Linguistics. Vienna and New York, Springer: 5 ss.
- Crystal, D. 1982 Profiling Linguistic Disability. Edward Arnold, London.
- Crystal, D., P. Fletcher, M. Garman 1976 The grammatical analysis of language disability. London, Edward Arnold: 9-11.
- Dean, H.O. 1965 "Toward Automated Phonological Mapping". Linguistics 17: 21-35.
- Eickmans, H. 1979 "Automatische Sprachkartographie im Rahmen des Projekts Fränkisher Sprachatlas". Niederdeutsches Wort, 19: 133-164.

- Eickmans, H. 1981 "Mogelijkheden en grenzen van de automatische verwaardiging von taalkaarten" taal en tongval 33, 3-4: 162-178.
- Fossat, J.L. 1977 "Vers un traitement automatique des données dialectologiques en dialectométrie". Germanistische Linguistik 3-4: 311-334.
- Francis, W.N. 1968 (1969) "Modal Daren't and Durstn't in Dialectal English". In S. Ellis (a cura di) Studies in Honour of Harold Orton on the Occasion of His Seventieth Birthday. Leeds Studies in English 1969, N.S.2: 145-163.
- Francis, W.N. 1969 "Computer Production of Dialect Maps". Computers and the Humanities 3: 290.
- Francis, W.N. 1970 "English Dialectology". Computers and the Humanities 4: 332-333.
- Francis, W.N. 1977 'International Colloquium on Automatic Dialect Mapping: a Report'. Computers and the Humanities 11: 339-340.
- Francis, W.N., J. Svartvik, G.M. Rubin 1969 Computer-Produced Representation of Dialectal Variation: Initial Fricatives in Southern British English. International Conference on Computational Linguistics, Preprint n° 52. Stockholm.
- Goebl, H. 1976 "La dialectométrie appliquée a l'A.L.F. (Normandie)". In: A. Varvaro (a cura di) Atti del XIV° Congresso Internazionale di Linguistica e Filologia Romanza, vol. II°. Napoli, Macchiaroli and Benjamin: 165-195.
- Goebl, H. 1977 "Rätoromanisch versus Hochitalienisch versus Oberitalienisch. Dialektometrische Beobachtungen innerhalb eines Diasystems". Ladinia 1: 39-71.
- Goebl, H. 1978 "Analyse dialectométrique de qualques points de l'AIS (italien standard, valdotain, provencal alpin, turinois, milanais)". In: P. Clivio, G. Gasca (a cura di) Lingue e dialetti nell'arco alpino occidentale. Atti del Convegno internazionale di Torino (12-14 aprile 1976). Torino, Centro Studi Piemontesi: 282-294.
- Goebl, H. 1980 "Dialektgeographie + Numerische Taxonomie = Dialektometrie. Anhand rätoramanischer und oberitalienischer Dialektmaterialien (AIS)". Ladinia 4: 31-95.
- Goebl, H. 1981 "Eléments d'analyse dialectométrique (avec application à l'AIS)". Revue de linguistique romane 14: 349-420.
- Goebl, H. 1982 "Atlas, matrices et similarités: petit aperçu dialectométrique". Computers and the Humanities 16: 69-84.
- Goebl, H. 1983 "Ansätze zu einer computativen Dialektometrie". In W. Besch, H. Knop, W. Putschke, H.E. Wiegand *Dialektologie*. Berlino: 778-792.
- Goebl, H. (a cura di) 1984 Dialectology. Bochum Brockmeyer.
- Goebl, H. 1984 "Lineamenti di dialettometria". In M. Cortelazzo Guida ai dialetti veneti VI. Padova, CLEUP: 7-53.
- Goebl, H. 1985 "Coup d'oeil dialectométrique sur les Tableaux phonétiques des patois suisses romands (TPPSR)". Vox Romanica 44: 189-233.
- Goebl, H., N. Winterleitner 1971 "Projekt einer sprachstatistischen Auswertung von in Sprachatlanten gespeicherter linguistischer Information mit Hilfe elektronischer Rechenanlagen". Linguistische Berichte 14: 60-61.
- Grassi, C. 1973 "Perspectives de l'emploi de l'élaborateur électronique en géographie linguistique et en dialectologie". In: A. Zampolli (a cura di) Linguistiça Matematica e Calcolatori. Atti

- del Convegno e della prima Scuola Internazionale (Pisa 16/VIII-6/XIX 1970). Firenze, Olschki: 233-239.
- Grassi, C., A. Genre, L. Massobrio 1977 "Die Anwendung der elektronischen Datenverarbeitung bei der Redaktion und dem Druck des Italienischen Sprachatlasses (ALI)". Germanistische Linguistik 3-4: 69-88.
- Grootaers, W.A. 1970 "Nouvelles méthodes et nouveaux problèmes: le nouvel Atlas Linguistique du Japon". In: Actes du Xe Congrès International des Linguistes (Bucarest, 28 agosto-2 settembre 1967) II°: 113-118.
- Hartman, J.W. 1969 "Some Preliminary Findings from DARE". American Speech 44: 191-199.
- Hartman, S.L. 1981 "A Universal Alphabet for Experiments in Comparative Phonology". Computers and the Humanities 15: 75-82.
- Händler, H. 1977a "Zum Dokumentations und Speicherungsproblem im Rahmen der Automatischen Sprachkartographie". Germanistische Linguistik 3-4: 163-180.
- Händler, H. 1977b "Eine Methode zur Automatischen Herstellung von Isoglossen". Germanistische Linguistik 3-4: 259-274.
- Händler, H. 1983 "Eptwürfe zu dialektalen Informationssystemen". In W. Besch, H. Knop, W. Putschke, H.E. Wiegand *Dialektologie*. Berlino: 792-806.
- Händler, H., C.L. Naumann 1976 "Zur Automatisierung der Isoglossenfindung". Germanistische Linguistik 3-4: 121-159.
- Henderson, M.M.T. 1974 "Of Matters Lexicographical: Processing Data for DARE-current Practices". American Speech 49: 119-122.
- Henderson, M.M.T. 1975 "Use of an Interactive Program in Analyzing Data for a Dialect Dictionary". Computers and the Humanities 9: 105-113.
- Hirshberg, J. 1981 "Regional Morphology in American English: Evidence from DARE". American Speech 56: 33-52.
- Hockey, S. 1980 A Guide to Computer Applications in the Humanities. London, Duckworth.
- Houck, C.L. 1967 "A Computerized Statistical Methodology for Linguistic Geography: a Pilot Study". Folia Linguistica 1: 80-95.
- Houck, C.L. 1969 A Statistical and Computerized Methodology for Analyzing Dialect Materials. Ph.D., Univ. of Iowa. Dissertation abstract reprint in Computers and the Humanities, 8 (1974): 120-121.
- Jakob, K., B. Kelle 1984 "Ein Verfahren zur quantitativen Bestimmung von 'Isoglossen'". In Goebl 1984: 136-163.
- Jones (Sargent), V.M. 1978a Some Problems in the Computation of Sociolinguistic Data. Ph. D., Univ. of Newcastle-upon-Tyne.
- Jones (Sargent), V.M. 1978b "The Tyneside Linguistic Survey an Approach to Data Processing in Sociolinguistics". Statistical Methods in Linguistics Quaterly Journal 2: 5-23.
- Keil, G.C. 1972 "IMPAC (Indexing, Map Plotting, and Analysis by Computer)". Computers and the Humanities 7: 116.
- Keil, G.C. 1974 "Narrow Phonetic Transcription on the Computer: Taking the Phone off the Hook". Computers and the Humanities 8: 217-229.
- Kelle, B. 1976 Das Belegmaterial des Südwestdeutschen Sprachatlas Aufbereitung und Auswer-

- tung mit Hilfe der Elektronischen Datenverarbeitung. Friburg.
- Kelle, B. 1977 "Datenaufbereitung und Automatische Kartierung beim Südwestdeutschen Sprachatlas (SSA)". Germanistische Linguistik 3-4: 89-106.
- Lance, D.M. 1977 "Determining Dialect Boundaries in the United States by Means of Automatic Cartography". Germanistische Linguistik 3-4: 289-303.
- Lance, D.M., S.V. Slemons 1976 "The Use of the Computer in Plotting the Geographical Distribution of Dialect Items". Computers and the Humanities 10: 221-229.
- Linn, M.D. 1982 "A Statistical Model for Classifying Dialect Speakers". In: H.J. Warkentyne (a cura di) Papers from the Fourth International Conference on Methods in Dialectology (1981): 244-253.
- MacWhinney, B. 1987 Transcript analysis. Department of Psychology, Carnegie Mellon University, Pittsburg.
- Martel, P. 1986 "Richesse lexicale et variables sociologiques". In En hommage à Charles Muller. Méthodes quantitatives et informatiques dans l'étude des textes. Genève-Paris, Slatkine-Champion: 599-608.
- Massobrio, L. 1978 "Il progetto per un 'Saggio di carte linguistiche delle inchieste francoprovenzali' dell'Atlante Linguistico Italiano". In: G.P. Clivio, G. Gasca Queirazza Lingue e dialetti nell'arco alpino occidentale. Atti del Convegno internazionale di Torino (12-14 aprile 1976). Torino, Centro Studi Piemontesi: 163-170.
- McDavid, R.I., R.K. O'Cain 1977 "From Fieldworker to Data Retrieval: Some Questions". Germanistische Linguistik 3-4: 11-24.
- Melillo, M. 1981 "La strumentazione elettronica di una cattedra di dialettologia (nella facoltà di lettere a Bari)". Lingua e storia in Puglia 12: 3-16.
- Montemagni S. 1986 Un esperimento di dialettologia computazionale: elaborazione ed analisi delle risposte ad un gruppo di domande dell'Atlante Lessicale Toscano, Tesi di laurea in Dialettologia Italiana (rel. G. Giacomelli), Università di Firenze A.A. 1985-86.
- Naumann, C.L. 1976 "Grundzüge der Sprachkartographie und ihrer Automatisierung". Germanistische Linguistik 1-2: 1-285.
- Naumann, C.L. 1977 "Klassifikation in der Automatischen Sprachkartographie". Germanistische Linguistik 3-4: 181-210.
- Neumann, R. 1977 "Symbolisierungsverfahren im Rahmen der Automatischen Sprachkartographie". Germanistische Linguistik 3-4: 225-244.
- Ogino, T. 1982 "Computer-aided Analysis of Field Survey Data-GLAPS and its Application". In Proceedings of the XIIIth International Congress of Linguists (Tokyo, Aug. 29-Sept. 4) The Hague: 981-985.
- Ogino, T. 1984 "Computer-assisted Analysis of Field Survey Data Glaps and its Application". In Goebl 1984: 254-274.
- Ogino, T., T. Sibata 1977 "Generalization of Procedures Relating to Computer-produced Linguistic Atlates". Germanistische Linguistik 3-4: 55-68.
- Pederson, L. 1981 "The Linguistic Atlas of the Gulf States: interim report four". American Speech 56: 243-259.
- Pellowe, J., G. Nixon, B. Strang, V. McNeany 1972 "A Dynamic Modelling of Linguistic Variation: the Urban (Tyneside) Linguistic Survey". Lingua 30: 1-30.

- Pellowe, J., V.M. Jones 1979 "Tyneside Linguistic Survey". Computers and the Humanities 13: 54-55.
- Peterson, J., Z. Thundiyil 1973 "A Computerized Questionnaire for the Dialectal Survey of the Upper Penisula of Michigan". Annual Meeting Papers: Michigan Linguistic Society 1, 3: 1-32.
- Philps, D. 1984 "Dialectometrie automatique". In Goebl 1984: 275-296.
- Philps, D. 1978 Projet de traitement automatique de l'Atlas dialectométrique des Pyrénées centrales, Toulouse-Jeddale.
- Picchi, E., N. Calzolari 1985 "Textual Perspectives through an automatized lexicon". In Proceedings of the XII International ALLC Conference (Nice).
- Pudlatz, H. 1977 "Automatische Erzeugung von Isoglossen auf dem Plotte mit Hilfe von Thiessen-Poygonen". Germanistische Linguistik 3-4: 245-258.
- Putschke, W. 1969 "Über ein Computerprogramm zur Herstellung von Sprachkarten". Germanistische Linguistik 1: 45-114.
- Putschke, W. 1975 "Automatische Kartierung". In: A. Weijnen Atlas Linguarum Europae. Introduction. Assen: 103-108.
- Putschke, W. (a cura di) 1977a Automatische Sprachkartographie. Vorträge des Internationalen Kolloquiums zur Automatischen Sprachkartographie (Marburg vom 11.-16. september 1977).
- Putschke, W. 1977b "Automatische Sprachkartographie: Konzeption, Probleme und Perspektiven". Germanistische Linguistik 3-4: 25-40.
- Putschke, W., R. Neumann 1983 "Automatische Sprachkartographie". In: W. Besch, H. Knop, W. Putschke, H.E. Wiegand *Dialektologie*. Berlino, De Gruyter: 748-778.
- Rolshoven, J. 1984 "Qualitative und quantitative Dialektologie: Zum Projekt einer automatichen Edition und Auswertung von Sprachatlanten". In Goebl 1984: 297-313.
- Rubin, G.M. 1970 "Computer-Produced Mapping of Dialectal Variation". Computers and the Humanities 4: 241-246.
- Schneider, E.W., W. Viereck 1984 "The Use of the Computer in American, Canadian and British English Dialectology and Sociolinguistics". In Goebl 1984: 15-60.
- Schütz, A.J., J. Wenker 1966 "A Program for the Determination of Lexical Similarity between Dialects". In: P.L. Garvin, B. Spolsky (a cura di) Computation in Linguistics: a Case Book. London Bloomington, Indiana Univ. Press: 124-145.
- Shaw, D. 1974 "Statistical Analysis of Dialectal Boundaries". Computers and the Humanities 8: 173-177.
- Shuy, R.W. 1966 "An Automatic Retrieval Program for the Linguistic Atlas of the United States and Canada. In: P.L. Garvin, B. Spolsky (a cura di) Computation in Linguistics: a Case Book. London Bloomington, Indiana Univ. Press: 60-75.
- Shuy, R.W. 1977 "Some Problems in Preparing Dialect Data for Processing". Germanistische Linguistik 3-4: 145-158.
- Shuy, R.W., L. McCreedy, J. Firsching 1977 "Toward the Description of Areal Norms of Syntax: Some Methodological Suggestion". Germanistische Linguistik 3-4: 367-389.
- Southard, B. 1983 "The Linguistic Atlas of Oklahoma and Computer Cartography". Journal of English Linguistics 16: 65-77.

- Stussi, A., M. Alinei (a cura di) 1973 Spogli elettronici dell'italiano delle origini e del Duecento. II: Forme. 17: Prose veneziane. Bologna, Il Mulino.
- Stussi, A., M. Alinei (a cura di) 1978 Spogli elettronici dell'italiano delle origini del Duecento. II: Grafia. 2: Prose veneziane. Bologna, Il Mulino.
- Thomas, A.R. 1975 "Dialect Mapping". Orbis 24: 115-124.
- Thomas, A.R. 1977 "A Cumulative Matching Technique for Computer Determination of Speech-Areas". Germanistische Linguistik 3-4: 275-288.
- Thomas, A.R. 1980a Areal Analysis of Dialect Data by Computer. A Welsh Example. Cardiff, Wales U.P.
- Thomas, A.R. 1980b "Computer Analysis of a Dialectal Transition Belt". Computers and the Humanities 14: 241-251.
- Tisano, V. 1985 Concordanze lemmatizzate delle poesie in dialetto tursitano di Albino Pierro. Pisa, Servizio Editoriale Universitario.
- Tokugawa, M., T. Yamamoto 1967 "An Attempt to Draw a Linguistic Map with a Computer". Mathematical Linguistics 40: 27-30.
- Tuaillon, G. 1986 "How the French Dialectal Data Enter the Atlas Linguarum Europae". Computers and the Humanities 20: 247-252.
- Uskup, F.L. 1970 A Method for Automating Dialect Analysis, ERIC reports ED 037 713.
- Uskup, F.L., M.L. Al-Azzawi 1972 "Editing and Printing a Dialect Atlas by Computer". American Speech 47, 3-4: 203-210.
- Venetzky, R.L. 1969 "Computer-aided Humanities Research at the University of Wisconsin". Computers and the Humanities 3: 129-138. [On "The Dictionary of American Regional English": 131-133].
- Venetzky, R.L. 1981 "Storage, Retrieval, and Editing of Information for a Dictionary". American Documentation 19: 71-79.
- Viereck, W. 1984 "Der Einsatz des Computers in der amerikanisch-englischen und britisch-englischen Dialektologie und Soziolinguistik". Zeitschrift für Dialektologie und Linguistik 52: 6-30.
- Viereck, W. 1984 "The Presentation and Interpretation of English Dialects: Computer-assisted Projects". In: *Proceedings of the XIIIth International Congress of Linguists* (Tokyo, Aug. 29-Sept. 4 1982), The Hague: 997-1000.
- Warkentyne, H.J. 1971 "Contemporary Canadian English: a Report of the Survey of Canadian English". American Speech 46: 193-199.
- Weinreich, U. 1964 Machine Aids in the Compilation of Linguistic Atlases. Year Book of the American Philosophical Society, 1963. Philadelphia: 622-625.
- Wood, G.R. 1962 Text, Tables, and Maps of a Word Geography of the Interior South. Year Book of the American Philosophical Society, 1961. Philadelphia: 602-604.
- Wood, G.R. 1963 "Dialect contours in the Southern States". American Speech 38: 243-256.
- Wood, G.R. 1967 Sub-regional speech variations in vocabulary, grammar, and pronunciation. Cooperative Research Project No. 3046, Final Report. Edwardsville, Illinois.
- Wood, G.R. 1969 "Dialectology by computer". International Conference on Computational Linguistics, Preprint n° 19. Stockholm.

- Wood, G.R. 1970a "On Ways to Examine the Local Language". Computers studies in the humanities and verbal behavior 3, 1: 100-110.
- Wood, G.R. 1970b "Regional Vocabulary, Syntax, and Usage in English". Computers and the Humanities 5: 93-94.
- Wood, G.R. 1970c "Vocabulary Change in Regional English". Computers and the Humanities 5: 94.
- Wood, G.R. 1971a Vocabulary Change. A Study of Variation in Regional Words in Eight of the Sothern States. Carbondale & Edwardsville, Southern Illinois Univ. Press.
- Wood, G.R. 1971b "Why not the computer as editor?". In: Burghardt, L.H. (a cura di) Dialecto-logy: problems and perspectives. Knoxville (Tenn.), Univ. of Tennessee Press: 41-53.
- Wood, G.R. 1972 "The Computer in Analysis and Plotting". American Speech 47, 3-4: 195-202.
- Wood, G.R. 1977 "Refinements in Tabular Models of Variation in Regional American English". In: Computational and mathematical linguistics. Proceedings of the International Conference on Computational Linguistica (Pisa 27.8.-1.9.1973). Firenze, Olschki Ed., I° Vol.: 343-346.
- Zampolli, A. 1960 Studi di statistica linguistica eseguiti con impianti IBM. Thesi, Padova.
- Zampolli, A. 1968a "Recherche statistique sur la composition phonologique de la langue italienne executée avec un système IBM". In: Les Machines dans la Linguistique. Praga, Academia: 25-34.
- Zampolli, A. 1975 "L'elaborazione elettronica dei dati linguistici: stato delle ricerche e prospettive". In: Colloquio sul tema: le tecniche di classificazione e loro applicazione linguistica (Firenze, 13.12.1972). Roma, Accademia Nazionale dei Lincei: 23-107.
- Zampolli, A. 1983 "Lexicological and Lexicographical Activities at the Instituto di Linguistica Computazionale". In Zampolli, A., A. Cappelli 1983: 237-278.
- Zampolli, A., A. Cappelli (a cura di) 1983 "The Possibilities and Limits of the Computer in producing and publishing Dictionaries". Linguistica Computazionale 3.
- Zampolli, A., A. Sobrero, M.T. Romanello in stampa "Per un atlante 'modulare': il NADIR". Atti del XIV Convegno di Studi Dialettali Italiani, Atlanti Regionali: aspetti metodologici, linguistici ed etnografici (Palermo, 7-11 ottobre 1985).