

Atlanti regionali: aspetti metodologici, linguistici e etnografici

Atti del XV Convegno del C.S.D.I.
(Palermo 7-11 ottobre 1985)

ESTRATTO

19

Consiglio Nazionale delle Ricerche
Centro di Studio per la Dialettologia Italiana

Pacini Editore

ALBERTO A. SOBRERO, MARIA TERESA ROMANELLO,
ANTONIO ZAMPOLLI

PER UN ATLANTE MODULARE: IL NADIR *

1. *Premessa*

«Una generale rimeditazione» e «un decisivo rinnovamento dei principi e dei metodi della geografia linguistica» auspicava già qualche anno fa Corrado Grassi, in relazione allo spostamento di obiettivi dagli atlanti linguistici nazionali agli atlanti linguistici regionali (Grassi 1983). Tale mutamento infatti non è soltanto il risultato di modificazioni radicali della situazione linguistica da investigare (le premesse esterne alla geografia linguistica), ma è uno degli effetti delle nuove assunzioni generali della scienza dialettologica.

In particolare, qualunque ricerca che, rifiutando apriorismi neogrammaticali, voglia tener conto del legame esistente fra lingua e sentimento linguistico dei parlanti, tra modificazioni socio-economiche e tecnologiche e mutamenti linguistici, insomma che consideri lo spazio linguistico come un dato storico della comunità in esame, deve anche rinnovare — quanto meno — gli strumenti per la raccolta e l'organizzazione dei dati.

Nel progetto che esporremo qualcuno potrà vedere qualcosa di più: un mutamento radicale di oggetto, di obiettivo, di metodi. In realtà, chiunque abbia dimestichezza con i testi di Gilliéron, Jaberg, Terracini, riconoscerà invece che si tratta semplicemente di alcuni sviluppi — diremmo — 'naturali' della metodologia geolinguistica di base *iuxta propria principia*: quella metodologia di salda matrice trans- e cisalpina che è stata via via schematizzata, impoverita e, persino, ridotta a tautologia in alcuni recenti sviluppi (o involuppi?) che sembrerebbero voler ridurre la geografia linguistica a mera diatopia (Grassi 1980).

Un tentativo come quello che esporremo, in quanto ambizioso e immodesto tentativo di riproporre, in vesti aggiornate, metodologie ispirate alla tradizione geolinguistica 'classica', non vuol rimettere in discussione problemi epistemologici e teorici generali né statuti disciplinari, e neppure opzioni metodologiche fondamentali. E tuttavia, prima di farlo diventare itinerario di lavoro, riteniamo che sia utile, corretto, proficuo aprire una discussione intorno a quella zona delicatissima della ricerca che va dalla metodologia alla tecnica e da questa alla strumentazione.

Cercheremo perciò, in questa sede, di dare prima un'idea — anche

se sommaria — della ‘filosofia’ del Nuovo Atlante dei Dialetti e dell’Italiano per Regioni (NADIR), e in particolare del NADIR Salento (§ 2.)⁽¹⁾; descriveremo poi alcune innovazioni tecniche che di quella ‘filosofia dell’atlante’ sono la diretta conseguenza (§ 3.), soffermandoci in particolare su alcune tecniche di elicitazione (§ 4.) e sull’organizzazione di uno strumento particolare, costituito da uno specifico modello di *Linguistic Data Base* previsto per il NADIR (§ 5.).

2. Metodologie integrate per un atlante modulare

Anzitutto, perché una struttura *modulare*? E in che senso? Lo scopo è quello di rispondere a due ordini di esigenze: uno *interno* e uno *esterno*. Un atlante non può più essere uno strumento statico di rilevamento di dati, fissati in numero chiuso in una sincronia determinata. Perché oggi più nessuno si illude che per ‘spiegare’, ma anche per descrivere adeguatamente una parlata, o un rapporto fra parlate, o la distribuzione di un fatto linguistico, sia sufficiente un solo rilevamento attraverso un questionario — sia pure ricco, e molto ricco — somministrato durante un’intervista. Per raggiungere un obiettivo di tipo descrittivo-esplicativo l’atlante deve fornire un *corpus* molto ampio, ottenuto da informatori diversi in condizioni diverse e — se possibile — incrementabile con materiali raccolti sia nella stessa che in altra sincronia: parliamo in questo caso di *i n c r e m e n t o i n t e r n o*.

Lo ‘spessore’ diacronico e diastratico che si può così ottenere aderisce meglio dei tradizionali rilievi monodici alla complessa realtà sociolinguistica del punto. Si può realizzare attraverso una struttura ‘modulare’, a blocchi collegabili in serie, in modo che sia possibile aggiungere dati nuovi a quelli già raccolti, catalogati, assegnati, senza modificarli.

La stessa tecnica può consentire il collegamento in serie di atlanti regionali di aree diverse. In questo caso — che chiameremo dell’*i n c r e m e n t o e s t e r n o* — il massimo del rendimento sarà dato dal collegamento fra atlanti concepiti in modo simile o ben coordinato (con questionari e tecniche di elicitazione parzialmente sovrapponibili). Ma anche collegamenti — e quindi, per l’utenza, assemblaggi — di materiali raccolti con tecniche del tutto o in parte diverse, purché effettuati con le necessarie cautele, possono riuscire molto utili, nell’ottica euristica integrata della quale parleremo più avanti.

Quali dati dovranno essere raccolti? Con quale criterio dovranno essere selezionati? La risposta a queste domande richiede un’esplicitazio-

ne preliminare: bisogna prima chiarire in quali orizzonti teorico-metodologici ci muoviamo, con quali scopi e quali prospettive.

Le ricerche sociolinguistiche — ma anche psico- ed etnolinguistiche — più recenti sono concordi nel considerare oggetto di indagine non produzioni specifiche del singolo dialetto (o lingua, o varietà di lingua o di dialetto) ma il repertorio linguistico di una comunità, o quanto meno una varietà specifica in relazione al repertorio della comunità: là dove coesistono due o più varietà, la reciproca distribuzione, ma anche i singoli caratteri strutturali, sono studiati sia in relazione al *contesto* linguistico e comunicativo, che in relazione a variabili extralinguistiche significative: a questo studio è affidato, oggi, il quadro della variazione dialettale del punto. L'atlante, dunque, è il fuoco in cui convergono le due grandi dimensioni odierne della ricerca dialettologica: la geolinguistica e la sociolinguistica.

Indagine geolinguistica non significa soltanto variazione diatopica (Grassi 1980): la concezione dinamica del punto linguistico (Terracini) conduce a criteri ben precisi nella scelta delle località da esplorare, inducendo a proseguire lungo la linea tracciata dall'atlante nazionale, con criteri di scelta non definiti in astratto ma modellati sulle risultanze che emergono dall'analisi delle principali dinamiche socioeconomiche e delle principali emergenze storiche dell'area: identificazione delle 'aree forti' e delle loro zone d'influenza, valutazione del vettore economico e delle dinamiche sociali delle località in esame, analisi di lungo periodo della storia dell'area.

La stessa concezione conduce a rilevamenti multipli nella stessa località: per mettere in luce la qualità e la forza delle correnti che lo attraversano, e per scoprirne l'individualità e l'unità, ma anche per studiare l'ampiezza delle oscillazioni 'libere' e la dinamica particolarismo-standardizzazione.

Con questo passiamo alla dimensione sociolinguistica, che con la prima strettamente si intreccia. Accettando — come accettiamo — il passaggio dall'obiettivo 'dialetto' all'obiettivo allargato 'repertorio linguistico', che abbraccia un continuum provvisoriamente definibile attraverso i suoi estremi *dialetto* e *lingua*, il centro dell'interesse si sposta — o meglio si estende — dai rapporti sintagmatici e frasali (all'interno di una varietà) ai problemi di uso dei codici e di organizzazione del testo nel parlato quotidiano; dalla fonetica e grammatica del dialetto al cambio di codice, alle strategie del discorso e più in generale all'etnografia del parlare (Gumperz-Hymes 1972). Il compito centrale, a questo punto, diventa quello di fornire dati adeguati, cioè un corpus tale da consentire

la descrizione delle varietà sociolinguistiche e lo studio del rapporto tra variazione linguistica e fattori extralinguistici da una parte, delle strategie conversazionali dall'altra (Berruto 1983).

Questa è la delimitazione — si fa per dire — di campo. Ma, all'interno del campo, quali scelte compiere? Qui il discorso si fa ancora più complesso, e ogni scelta 'tecnica' rischia di diventare anche irreversibile scelta di campo, fra i diversi metodi dell'analisi sociolinguistica. Quale approccio — e quindi quali strumenti — privilegiare?

Per noi si tratta di fornire gli strumenti per descrivere, come dicevamo, varietà — o frammenti di varietà — sociolinguistiche di più repertori. Facendo un rapido giro d'orizzonte, vediamo che gli approcci, o meglio le tecniche di descrizione di varietà sociolinguistiche sono oggi sostanzialmente tre (Dittmar *in stampa*):

1) la *g r a m m a t i c a d i v a r i e t à*, eccellente soprattutto per la descrizione sintattica. Si presta all'analisi di un *corpus* ampio e differenziato, spazialmente distribuito in modo complesso, grazie alla sua base probabilistica e all'attenzione per una circoscrizione non pregiudiziale dello 'spazio di varietà'. Allo stato attuale non sembra invece tecnica vantaggiosa per l'analisi fonetica, a causa della necessità di formulare regole acontestuali — almeno nella formulazione di Klein-Dittmar (1979).

2) Consentono di colmare questa lacuna le *r e g o l e v a r i a b i l i* (Labov 1972, Cedergren-Sankoff 1974), che sono perfettamente in grado di misurare le 'variazioni interne', cioè le variazioni dipendenti dal contesto — dalla vocale, o consonante, o dal sintagma che precede o che segue — in termini non solo di enunciazione ma anche di misurazione (quantitativi). Senza contare che, ad esempio nel modello log-lineare di Van Hout (1984), è possibile indurre regole molto sofisticate sia sulle interazioni contestuali della realizzazione della variabile, sia sul grado di significatività della relazione fra regole e fattori extralinguistici.

Allo stato attuale, l'uso di questa tecnica è però affidato, in gran parte, a matematici — come Sankoff — che ne sfruttano le potenzialità matematico-statistiche molto più sul versante 'interno' — cioè con attenzione grammaticalistica ai rapporti contestuali — che sul versante 'esterno', cioè sulla componente sociopragmatica e stilistica. Qualora, in futuro, i problemi di discretezza e campionamento del versante sociologico e stilistico si avviassero a soluzione, con l'uso delle regole variabili sarebbe possibile, fra l'altro, un'identificazione pressoché automatica di *indicatori, markers e stereotipi*. C'è però un'esigenza: quella di disporre di molti dati *b e n c l a s s i f i c a t i*, sia sul piano linguistico (sintagmatico) che sul piano sociolinguistico (assegnazione alle variabili).

3) Le stesse esigenze emergono dal terzo approccio organizzativo- euristico: quello delle *s c a l e d i i m p l i c a z i o n e*. Strumento sul quale avanziamo però riserve circa la sua concreta utilità in un atlante, in particolare per l'area salentina.

Queste riserve si basano su due considerazioni (si veda anche Berruto 1980, 205). In primo luogo, come s'è già accennato, nel repertorio linguistico salentino è assai difficile riconoscere la presenza di varietà ben compartimentalizzate e funzionalmente differenziate in modo tale da costituire un *gradatum* (Sobrero *in stampa c*), mentre sembra più produttivo il concetto di *continuum* (2). Però il continuum è poco adatto ad essere descritto con le scale di implicazione: come appare evidente in Sobrero-Romanello 1981, è difficile trovare fenomeni discreti (quanto meno, le categorie non sono a due valori ma a tre o quattro) e, soprattutto, è difficile disporre tratti e regole scalarmente, secondo il criterio dell'implicazione. In secondo luogo, per definizione, i dati — le variabili linguistiche, gli eventuali gradata, le qualità sociologiche — in un atlante sono, devono essere, numerosissimi. Il che, come avverte lo stesso De Camp (1973) rende particolarmente difficile l'uso delle scale.

L'indagine di Dittmar e Schlobinski (*in stampa*) su Berlino ha dimostrato ampiamente che nessuno dei tre strumenti, dal punto di vista descrittivo, è in assoluto più efficace (o potente) degli altri. Si neutralizzano a vicenda anche i due aforismi, di David Sankoff, quando sostiene che 'analisi qualitativa' è un eufemismo per 'dati insufficienti', e di Wolfgang Klein quando viceversa osserva che «più dati abbiamo, meno conosciamo» (Klein *in stampa*).

In realtà si va facendo strada l'opinione — che noi condividiamo pienamente, soprattutto pensando a un atlante linguistico — secondo la quale si tratta di strumenti, per così dire, complementari (Tuson *in stampa*) che saranno scelti, di volta in volta, secondo: a) l'obiettivo, b) la competenza specifica del ricercatore, c) le caratteristiche della comunità linguistica in esame. La 'copertura' teorica a un'integrazione oculata fra metodologie di tradizione anche diversa — e dunque non solo quelle qui accennate, ma anche altre — è del resto ampia: riferiamo solo il giudizio di Suzanne Romaine, secondo la quale «una scienza che insista sul possesso di un solo metodo e di un solo tipo di risultati accettabili non è scienza ma ideologia» (Romaine 1982, 283).

È vero che un atlante non è uno strumento 'neutrale' di raccolta dati; ma è anche vero che deve offrire alla comunità dei ricercatori di oggi — e, non dimentichiamolo, di domani — i dati più attendibili, nel modo meglio utilizzabile con gli strumenti tecnici e metodologici a

disposizione, e con quelli ragionevolmente prevedibili.

Per questo nel disegno di un atlante linguistico oggi, pur operando scelte ed esclusioni, a nostro avviso bisogna accedere a una visione 'integrata' delle metodiche socio- e geolinguistiche di maggiore potenza esplicativa.

3. *I punti, gli informatori, il corpus*

Scendiamo ora al livello della tecnica: riprendiamo per alcuni punti cruciali le esigenze emerse da questa prima riflessione metodologica, e presentiamo le *proposte* conseguenti.

S c e l t a d e i p u n t i

Bisogna tenere conto: a) delle dinamiche socioeconomiche recenti e — ove possibile — meno recenti dell'area, utilizzando le analisi anche raffinate spesso disponibili in seguito a ricerche condotte da economisti, sociologi e persino urbanisti; b) dei modelli storici soggiacenti, della loro rilevanza e del loro rapporto di continuità/frattura con il quadro attuale; in una parola, delle loro proprietà vettoriali applicate alla sincronia. I risultati delle nostre analisi per il Salento, contenenti anche criteri specifici per la scelta delle 30 località dell'area sono stati esposti in Sobrero-Romanello-Tempesta *in stampa*.

S c e l t a d e g l i i n f o r m a t o r i

Per motivi diversi (l'uno teorico, l'altro pratico) bisogna scartare i due estremi della rappresentatività secondo criteri sociologici (rapporto statisticamente significativo campione/universo) e dell'informatore unico. Ma l'esperienza dimostra che, scartati questi, qualunque altro criterio concretamente applicabile è teoricamente insoddisfacente, e viceversa. D'altra parte, concordiamo con le critiche fatte, ad esempio, da Suzanne Romaine (1982, cap. V) alle tecniche di campionamento dello stesso Labov. Converrà perciò uscire da questa direttrice 'laboviana' per utilizzare stimoli diversi.

Un'ottima cartina di tornasole è costituita, a nostro avviso, dall'*indice di integrazione in rete*, che in un recente sondaggio compiuto in Salento, messo in relazione con l'analisi di punti cruciali del comportamento linguistico, si è rivelato come un'interessante chiave di accesso al 'sentimento linguistico' del parlante (Sobrero *in stampa b*). L'uso del

concetto di 'rete sociale' — per l'occasione rettificato, rispetto al classico Milroy 1980 — comporta, fra l'altro, l'abbandono della caccia all'informatore NORM (*nonmobile, older, rural, male*), in favore di una più rappresentativa scelta di informatori distribuiti secondo qualità sociologiche, ma anche secondo competenze linguistiche diverse: diventano importanti in questa ottica anche i parlanti non perfettamente *in-rete*, e quelli che sono stati definiti (Dorian 1977) *semi-speakers*.

S c e l t a d e l l e p r o d u z i o n i l i n g u i s t i c h e

Bisogna puntare alla massima rappresentatività del campione testato nei confronti delle produzioni effettive della comunità di parlanti: e dunque privilegiare il *parlato spontaneo*, in condizioni comunicative abituali (con tutto il corredo di selezione, cambio, sostituzione di codice; interferenze, infiltrazioni, ibridazioni ecc. ecc.). Il problema delle metodologie di approccio al parlato è stato trattato, e le conseguenti opzioni nelle tecniche di elicitazione sono state illustrate ampiamente in Sobrero-Romanello-Tempesta *in stampa*. Qui ricordiamo soltanto che gran parte dei materiali consisterà in: a) rilevamenti 'in situazione', con tecniche vicine a quella dell'osservazione partecipante; b) etnotesti, di prevalente interesse etnografico; c) interviste a tema.

Infine (o, se si preferisce, in primo luogo): letture 'tradizionali', etimologiche, geolinguistiche e sociolinguistiche convergono sulla necessità fondamentale di disporre anche di materiale facilmente *ordinabile* e *paragonabile*. E questo si può fare solo con la somministrazione di questionari 'tradizionali' di tipo fonetico-fonemico, morfosintattico, lessicale, su nozioni generalmente condivise presso una comunità anche più ampia di quella investigata.

È senza dubbio quest'ultimo il punto più interessante e di più facile 'aggancio' del modulo NADIR con altri eventuali moduli esterni, cioè con altri atlanti regionali passati, presenti e futuri.

Questo mi sembra un punto importantissimo. Se continueremo a procedere ciascuno per la propria strada renderemo inutilmente e assurdamente complicato il lavoro dell'utente dei nostri futuri atlanti linguistici, il quale dovrà districarsi fra metodi, questionari, tecniche di raccolta e di rappresentazione, alfabeti, soluzioni grafiche e informatiche tutte diverse fra loro. La modularità del NADIR, soprattutto per il materiale ordinabile e paragonabile, va in questa direzione; ma bisogna fare ulteriori passi avanti, pensando soprattutto ai questionari e ai sistemi di trascrizione.

Riteniamo che questa sia la sede adatta per avviare un discorso con-

creto di collaborazione fra le diverse iniziative in corso, tale che — pur salvaguardando le specificità concettuali delle diverse iniziative — renda i dati accessibili in modo omogeneo. E ci auguriamo che, partendo da questo Convegno, si raggiunga presto un protocollo d'intesa su alcuni standard di raccolta, di trascrizione e di interrogazione dei dati.

4. *I questionari*

Si è già detto che è necessario disporre di materiale facilmente ordinabile e paragonabile, e che ciò si può solo ottenere con la somministrazione di questionari 'tradizionali'. Partendo da questa esigenza, al momento di elaborare il progetto del NADIR ci siamo posti doverosamente il problema metodologico della verifica del questionario tradizionale come strumento affidabile per lo stimolo e la raccolta di dati dialettali, in una situazione come quella salentina, tutt'altro che rara in Italia.

Avevamo previsto un «questionario dialettale orientato» (3). Prima di metterlo a punto, abbiamo eseguito, nella nostra area, un esperimento che aveva un duplice obiettivo: 1) verificare, in pre-inchiesta, una breve sezione del questionario; 2) confrontare la resa tecnica di diversi tipi di questionari per l'inchiesta dialettale, con lo scopo di trarne, ovviamente, conseguenze operative per l'atlante.

In questa sede, offriamo alla discussione alcuni dati relativi a questo secondo obiettivo.

Si è scelto di operare per ora sul livello lessicale, selezionando, tra alcune sfere semantiche (parti del giorno; parentela; amore, nascita, matrimonio, morte; arnesi da cucina) voci (o referenti) fortemente indiziati come segnalatori di una fase di destrutturazione (o ristrutturazione) lessicale (o semantica).

Si sono quindi costruiti tre brevi questionari diversi, con i medesimi tipi lessicali. Il primo (n. 1) a struttura mista: costituito, cioè, in parte da *d o m a n d e d i r e t t e* (del tipo «come si dice in dialetto *grandine?*») e in parte da *d o m a n d e i n d i r e t t e* (del tipo «che cosa è per te una bambina che porti al battesimo o alla cresima?») alcune delle quali corredate da foto (specie per i referenti appartenenti alla sfera 'arnesi da cucina'). Una struttura mista di questo tipo, come è noto, è legata alla necessità di ridurre quanto più è possibile l'ambiguità dei quesiti o il rischio di risposte 'obbligate' o ricalcate sulla domanda.

Il secondo questionario (n. 2) è invece una lista di frasi in italiano, costruite attorno ai tipi di cui si vuole il corrispettivo dialettale, delle

quali si richiede *t r a d u z i o n e*. Per ridurre (visto che non si possono annullare) le note perplessità su questo tipo di questionario (Cortelazzo 1969, 119), si è badato a variare la posizione dei tipi selezionati all'interno delle frasi, a creare fuochi multipli, introducendo più tipi nella stessa frase, e, in fase di somministrazione, a riordinare le domande in modo tale da attenuare nell'informatore la possibilità di immediato riconoscimento delle sfere semantiche in esame. La costruzione di questo tipo di questionario ha richiesto particolare attenzione, per la difficoltà di reperire in italiano il corrispettivo di alcuni dei tipi dialettali attesi, riducendo al minimo le ambiguità legate alle polivalenze, alle sinonimie, agli spostamenti semantici, alle ipergeneralizzazioni, specialmente per la sfera relativa agli arnesi da cucina, o alle parti del giorno.

In sostanza la fase di costruzione di questionari di tal genere pone al ricercatore tutti i problemi che sottendono l'operazione del tradurre, cioè dello stabilire equivalenze.

Il questionario n. 2 si rivelerà però quello di più agevole somministrazione in fase di inchiesta.

Il terzo questionario (n. 3) è solo uno schema 'aperto' per l'*i n t e r v i s t a g u i d a t a*. Il raccoglitore, che ha adattato lo schema alle singole situazioni di inchiesta, ha utilizzato prevalentemente il dialetto perché più 'naturale' con l'informatore prescelto: in conseguenza della condizione di spontaneità in cui si è realizzata l'inchiesta il comportamento del raccoglitore ha subito i fenomeni di *code switching* normali nel parlato.

Con questo esperimento, dunque, sono state separate e trattate distintamente situazioni-inchiesta, o meglio tecniche di elicitazione, che la tradizione dialettologica vuole generalmente compresenti.

Per questo, i questionari così predisposti sono stati somministrati, nella medesima località e dal medesimo raccoglitore, a tre informatori diversi, ma omogenei per caratteristiche sociali e personali: di età media (dai quaranta ai sessanta anni) dello stesso livello di istruzione (superiore a Lecce, elementare nei centri della provincia).

L'opportunità di reperire informatori omogenei all'interno del punto ha prevalso su quella di stabilire a priori caratteristiche valide per gli informatori di tutta l'area: l'obiettivo primario non era tanto la comparabilità tra punto e punto, quanto tra le diverse tecniche, in ogni punto e in tutta l'area. È così risultata una situazione profondamente marcata: i centri agricoli hanno consentito la selezione di informatori secondo variabili, per così dire, 'classiche'; in città si è privilegiato l'informatore con istruzione elevata e per il quale è difficile escludere periodi

più o meno lunghi di allontanamento dal luogo di origine. Ma questo riapre il discorso, a cui già si è accennato (§ 3), della necessità di rivedere i criteri tradizionali di selezione degli informatori.

Le inchieste sono state eseguite a: San Vito dei Normanni, Mesagne (due inchieste), San Pietro Vernotico, Lecce (due inchieste), Calimera, Parabita.

Come si vede, in tal modo si è saggiata, seppur attraverso pochi punti, ciascuna delle tre aree del salentino (secondo la classificazione tradizionale): al salentino settentrionale appartengono S. Vito, Mesagne, San Pietro, della provincia di Brindisi, a quello centrale Lecce e Calimera (che, come è noto, è uno dei centri dell'area grica); a quello meridionale Parabita. Secondo il nostro tentativo di integrazione dei dati linguistici noti con i risultati delle più recenti indagini economiche ⁽⁴⁾ Parabita e Calimera rientrano nelle zone 'emergenti' individuate a sud di Lecce (le due fasce Lecce-Casarano, Lecce-Soletto), mentre i tre centri a nord di Lecce sono particolarmente idonei a comprovare la soluzione di continuità, sotto il profilo economico, tra i centri della provincia di Brindisi e il capoluogo.

Riassumo ora brevemente alcuni dei dati più significativi emersi dall'indagine, relativamente alla 'resa' delle tre diverse tecniche:

Tabella n. 1 - *Percentuale risposte ottenute*

Quest. n.	S. Vito N.	Mesagne	S. Pietro V.	Lecce	Calimera	Parabita	Media gen.
1	98,2	98,1	96,5	76	95	74	89,63
2	100	97	100	92	95	96	96,6
3	82,5	55,2	80,7	50,8	59,6	35	60,63

Non tutte le domande hanno ottenuto risposta; però la resa dei questionari è stata diversa. In relazione al numero di domande soddisfatte, il questionario n. 1 ha dato un rendimento piuttosto alto, anche se non ottimale: i valori più bassi si segnalano a Lecce e a Parabita. Il questionario n. 2 (traduzione di frasi) ha dato valori generalmente più alti, anche a Lecce, dove comunque si registra ancora il numero più alto di assenza di risposta.

L'intervista guidata (n. 3) ha dato invece valori discontinui, però in ogni località più bassi di quelli registrati con le altre tecniche. Certamente questi valori si spiegano con le caratteristiche peculiari della tecni-

ca, la cui resa dipende da fattori diversi, tra cui l'abilità del raccoglitore, la struttura conversazionale, l'organizzazione testuale, ecc.

Una lettura di tipo meramente quantitativo confermerebbe quindi una maggiore facilità di somministrazione del questionario n. 2, insieme ad una generale difficoltà a ottenere dati dialettali a Lecce, dove del resto si è registrato il più alto numero di risposte esitanti (per 1: 23%, per 2: 11%).

È opportuno però integrare queste indicazioni con alcuni dati sul rendimento dei questionari in rapporto all'obiettivo proposto, la raccolta di dati lessicali di tipo dialettale; è opportuno cioè chiederci quante risposte diagnosticabili come sicuramente dialettali sono state ottenute mediante le tre tecniche:

Tabella n. 2 - *Percentuale risposte dialettali* (calcolata sul numero di risposte ottenute)

Quest.	S. Vito N.	Mesagne	S. Pietro V.	Lecce	Calimera	Parabita	Media gen.
1	75	83,1	90	64,3	64,8	50	73,2
2	85	69,3	92	55,7	77,7	74,5	72,8
3	89,3	92,06	95,6	43,4	82	75	81,3

In tal senso, nessuno dei tre questionari offre una resa del cento per cento.

È certo però che in tutti i centri della provincia (tranne che a Lecce) il questionario che ha fornito il maggior numero di risposte sicuramente dialettali è stato il terzo, cioè quello costituito dall'inchiesta guidata. I valori non seguono quest'andamento, e sono piuttosto vicini tra loro, a Lecce, che conferma lo status tipico della situazione urbana, fornendo in generale l'indice più basso di dialettalità.

Ma il resto delle risposte ottenute non sempre è diagnosticabile come sicuramente italiano: generalmente si colloca nella 'zona grigia' del continuum lingua-dialetto (italiano regionale? dialetto italianizzato? italiano con fonetica dialettale?) la cui quantificazione è poco significativa proprio perché i criteri di discriminazione non sono oggettivi (tra i nostri dati: *pomeriggiu*, *vétuvu* 'vedovo', *stasera*, *furmine*, ecc.).

In generale, comunque, i materiali linguistici così ottenuti pongono al ricercatore problemi di classificazione, relativamente al rapporto innovazione-conservazione. Gran parte delle domande d'inchiesta, formulate con le tre tecniche diverse, hanno ottenuto esiti differenti, nella

stessa località. La situazione riguarda tutte le sfere semantiche qui testate. Ma è possibile ipotizzare una relazione tra tecnica d'inchiesta e il carattere innovativo, oppure conservativo, di una risposta? La domanda non è oziosa, visto che è strettamente connessa con gli obiettivi della ricerca, e, nel nostro caso, di un atlante regionale.

Si può tentare una prima risposta analizzando gli esiti delle domande che nella quasi totalità dei punti hanno dato forme differenti con le tre tecniche: dall'esame risulta con chiarezza che i tipi sicuramente conservativi si riscontrano generalmente nelle risposte al questionario n. 1. Ad esempio, è con questa tecnica che si è ottenuto, per 'padre', *sire, tata, attani* ⁽⁵⁾ ~ *papà, padre*; per 'brina' *šilazza* ⁽⁶⁾ ~ *úmutu*; per 'svezzare' *stağğare* ⁽⁷⁾ ~ *llevare lu latte, svezzare, ecc.*

Si noti però che molte di queste risposte sono date con forte esitazione, o addirittura con la dichiarazione esplicita dell'informatore di fare uno sforzo di memoria per richiamare alla propria coscienza linguistica un tratto perduto nell'uso quotidiano. È dunque una tecnica che, richiedendo rigide corrispondenze, tipo a tipo, determina nella memoria linguistica dell'informatore una forte tensione, che può essere vissuta in vario modo, a seconda del rapporto — anche 'ideologico' — che egli ha con il dialetto. Inoltre, data la sua articolazione, il primo questionario è quello che ha registrato il numero più basso di risposte in forma di perifrasi:

Tabella n. 3 - *Percentuale risposte in forma di perifrasi* (calcolata sul numero di risposte ottenute)

Quest.	S. Vito N.	Mesagne	S. Pietro V.	Lecce	Calimera	Parabita	Media gen.
1	0	9,7	7,2	0	5,5	0	4,4
2	8,7	6,3	12,2	6	20,3	12,7	10,02
3	19,1	11	15,2	1,7	17,6	15	12,3

La prima impressione è che la domanda, diretta o indiretta, garantisce un maggior numero di risposte a corrispondenza univoca (da parola a parola) senza inutili dispersioni. Ma è questo che noi cerchiamo? O l'assenza di perifrasi è anch'essa invece indizio della tensione del ricordare termini ormai in disuso, tensione magari sollecitata dalla spinta a reperire il termine tradizionale, pena il deprezzamento del valore-dialetto? Se così fosse, bisognerebbe concludere che l'assenza di perifrasi non ri-

flette il grado di stabilità effettiva del tipo lessicale, ma ormai ne dà un'idea, per così dire, 'ideologizzata'. Il dubbio si affaccia confrontando il 4,4% di media (e addirittura lo zero della metà delle località) per il primo questionario con il 10,02 e il 12,3 dei questionari n. 2 e n. 3.

Il secondo questionario, a lista di frasi, è quello che dà il materiale di più difficile classificazione. Fortemente condizionato dalla capacità dell'informatore a comprendere il meccanismo della traduzione (come ben sanno specialmente gli etnolinguisti, cfr. Cardona 1976, 246) fa registrare molte risposte a prima vista diagnosticabili come tratti innovativi, ma che sono in realtà italianismi da attribuire al condizionamento della domanda proposta. Inoltre, proprio perché riduce al minimo l'incidenza del contesto situazionale, questa tecnica non consente neppure verifiche in competenza passiva. Il fatto che sia stata quella che ha fornito il più alto numero di risposte, e che in nessun caso abbia fatto registrare forti incertezze è soltanto la prova che essa consente, comunque, una risposta. Un altro significativo indizio: ha dato quasi sempre risposte ad unica uscita; la pluralità di esiti, massiccia negli altri questionari qui è ridotta quasi a zero. Anche per questo, ci sembra di poter ipotizzare l'insufficienza di un questionario a traduzione di frasi, almeno per inchieste di tipo lessicale. Se usato come unico strumento, è inadeguato a rappresentare la situazione reale della concorrenza di forme nel punto e nell'area (Cortelazzo 1969, 110). Si confermano dunque, anche nell'analisi dei dati le ipotesi di 'difficoltà' emerse già nella costruzione del questionario per la traduzione.

A questo punto si pone un problema più generale (che vale sia per la tecnica del tipo 1, sia per quella del tipo 2): il questionario, ideato in e per situazioni di rigida diglossia, presenta difficoltà di applicazione là dove i due codici interessati dalla traduzione non sono costantemente, strutturalmente e socialmente differenziati, cioè non sono 'lingue in opposizione' ma sono varietà in co-occorrenza, in situazioni ormai generalizzate di competenza multipla. Il questionario è dunque uno strumento da ripensare?

Certo, a conclusione del nostro esperimento, dobbiamo confermare che offre maggiori garanzie la tecnica dell'intervista guidata: proprio perché meno rigida e più vicina alle condizioni di realizzazione di una situazione 'naturale': ma non tanto da non poter consentire una comparabilità geo-linguistica. Soprattutto, se lo scopo del nostro atlante non è quello di costruire repertori, e quindi non ci interessa sbilanciare le tecniche di elicitazione a favore dell'arcaismo, e se intendiamo raccogliere materiali utilizzabili da più approcci metodologici (8).

Il problema è, a questo punto, quello dell'addestramento dei raccoglitori. Ma questo è altro discorso.

5. *L'elaborazione elettronica dei dati*

Per il NADIR si dovrà dunque raccogliere un corpus ampio, con molti informatori, attraverso tecniche di elicitazione diverse, in situazioni definite da elementi differenti in varia e mutevole combinazione; i dati dovranno essere organizzati in modo comparabile, incrementabile (struttura modulare) e dovranno essere predisposti a utilizzazioni varie, rapide, incrociate, sommate, integrate, di tipo qualitativo e quantitativo, metodologicamente differenziate e possibilmente modificabili.

Il progetto presenta perciò caratteristiche che rendono quanto mai opportuno, e forse indispensabile, il ricorso all'elaborazione elettronica.

La tipologia e la natura dei dati raccolti e la varietà delle utilizzazioni cui essi si prestano impongono alcune scelte organizzative. In particolare, i materiali raccolti dovrebbero essere messi a disposizione della comunità degli studiosi il più presto possibile, senza attendere il completamento della loro elaborazione, che può avvenire mediante analisi ed arricchimenti progressivi. La consultazione e l'elaborazione, inoltre, dovrebbero essere possibili secondo parametri (o combinazioni di parametri) diversi, scelti di volta in volta dall'utente a seconda dei propri punti di vista e interessi specifici.

Si considerino, a titolo di esempio, i fatti seguenti:

a) Il progetto prevede l'impiego di diversi tipi di raccolta. Ciascuno di essi produce dati caratterizzati da strutture diverse, che richiedono metodi di rappresentazione e di trattamento specifici. Le tecniche tradizionali offrono forme diverse di documentazione per i testi (concordanze, frequenze, ecc.), per i questionari (carte, indici, ecc.), ecc.

I materiali ottenuti con le diverse forme di rilevamento devono invece essere considerati strettamente integrati e complementari tra loro. Lo studioso dovrebbe poter rintracciare la documentazione del fenomeno o dell'insieme dei fenomeni che lo interessano nell'intero corpus dei materiali disponibili. 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, modi di accesso diversi.

b) I materiali raccolti si prestano ad analisi condotte a livelli linguistici diversi: fonetico, morfosintattico, lessicale, sintattico, semantico, ecc.

Un ulteriore grado di complessità è introdotto dalla compresenza

di testi italiani e dialettali, e dalla necessità di poter rappresentare e rendere interrogabili le relazioni tra le unità linguistiche di due testi (italiano e dialettale) corrispondenti. Il corpus del NADIR, in definitiva, si dovrà presentare come un insieme di rappresentazioni di strutture linguistiche, proiettate sui materiali raccolti, costruite progressivamente.

c) È un fatto purtroppo noto che l'utilizzazione dei dati, secondo le tecniche tradizionali, è possibile in genere solo molto tempo dopo la raccolta. Ne risulta fra l'altro che la situazione rispecchiata al momento delle inchieste non è più attuale al momento dell'utilizzazione. Le tecniche tradizionali di documentazione e diffusione dei dati richiedono una lunga fase di elaborazione come premessa alla pubblicazione. In particolare, alcune analisi richiedono, per loro natura, tempi e costi notevoli.

Per alcuni livelli linguistici, non esistono, e forse non esisteranno per molto tempo ancora, teorie ben specificate ed universalmente accettate per la descrizione linguistica. La descrizione linguistica dei materiali non può quindi essere concepita come una struttura definita una volta per tutte. Il corpus deve essere «arricchito» con analisi condotte progressivamente, ed eventualmente modificabili dinamicamente (§ 2), ed essere nel contempo immediatamente disponibile per l'utilizzazione, senza attendere il compimento delle analisi.

È possibile anzi che l'esame stesso dei materiali raccolti possa progressivamente fornire elementi per la modificazione dei criteri di analisi.

Per tutti questi motivi, non sembra consigliabile mantenere 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. Le analisi linguistiche potranno essere effettuate progressivamente, non solo dai raccoglitori, ma anche dagli utilizzatori.

L'utente assumerebbe anche un ruolo attivo. L'analisi e la descrizione del corpus potrebbe diventare il risultato progressivo di uno sforzo cooperativo, ed evolvere con l'evolvere delle conoscenze.

d) Sui materiali linguistici devono essere proiettate anche le classificazioni extralinguistiche (geografiche, socioculturali, ecc.) del corpus. Lo studioso dovrebbe poter accedere ai dati secondo combinazioni e incroci di parametri extralinguistici e di categorie linguistiche definiti di volta in volta a seconda degli obiettivi specifici della propria ricerca, ed eventualmente ridefiniti ricorsivamente fino a trovare la configurazione desi-

derata. Le tecniche di documentazione tradizionali invece, pur utilissime, rendono accessibili i dati pubblicandoli ordinati definitivamente secondo criteri scelti una volta per tutte.

e) I materiali raccolti nell'ambito del progetto NADIR - Salento dovrebbero essere integrati con quelli disponibili, o in via di raccolta, o raccolti in futuro per altre ragioni.

L'utente deve poter accedere contemporaneamente a materiali raccolti ed elaborati in progetti diversi.

L'utilizzazione stessa del corpus può indicare l'opportunità di arricchirlo con indagini complementari per particolari aree geografiche, strati socioculturali, linguaggi speciali, ecc.

In altre parole, il corpus dei materiali dialettali e regionali viene concepito come un insieme aperto nella dimensione geografica, temporale, e socioculturale.

Affinché ciò sia concretamente realizzabile, è necessario, oltre a stabilire opportuni accordi metodologici tra centri diversi di ricerca, utilizzare metodi che consentano di compiere, sulle rappresentazioni dei dati, le eventuali trasformazioni formali necessarie per rendere omogenei, almeno per quanto concerne le operazioni di consultazione dell'utente, i materiali raccolti ed elaborati in progetti o sottoprogetti diversi.

Non sembra possibile soddisfare tutte queste esigenze con i metodi tradizionali di documentazione e pubblicazione. Sembra invece che la gestione e l'utilizzazione dei materiali raccolti con il progetto NADIR costituisca un caso paradigmatico di «base di dati linguistici» (LDB), così come oggi viene descritta dalla linguistica computazionale: «We use the expression *L i n g u i s t i c D a t a B a s e* (LDB) to define a *s e t o f l i n g u i s t i c d a t a* of different types (not only texts and dictionaries, but also, for example, the results of socio-linguistic studies, linguistic models, bibliographic data, etc.) *f i n a l i z e d* for the interactive utilization by multiple categories of potential users, *s t o r e d*, *s t r u c t u r e d* and *l i n k e d* for this purpose, and *a s s o c i a t e d* to specialized software modules for access, interrogation and on-line processing» (Zampolli, 1983, p. 262).

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 integrano conoscenze e tecniche sviluppate in settori diversi di questa disciplina: lessicologia computazionale, text-processing, statistica linguistica, trattamento delle immagini, parsers, ecc. Riassumere sia pur brevemente le acquisizioni o — per così dire — lo stato del-

l'arte raggiunto nei diversi settori è qui ovviamente impossibile. Si rinvia pertanto alla bibliografia citata. D'altra parte la concezione generale e la progettazione dettagliata del sistema informativo per il NADIR richiedono una fase previa di analisi e di studio, che costituisce una parte fondamentale dell'intero progetto. Per ora sembra possibile solo accennare ad alcune caratteristiche che il sistema dovrà, probabilmente, possedere.

Centri specializzati che 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 è costituita di solito dalla frequenza dell'elemento cercato, dall'index locorum delle sue occorrenze, o dai contesti costruiti con criteri e lunghezza definiti dall'utente.

Evidentemente, queste strategie di accesso dovranno essere arricchite per tener conto delle caratteristiche specifiche del progetto NADIR: corrispondenze italiano regionale-dialetto; dati extralinguistici; livelli diversi di descrizione (fonetico, lessicale, morfosintattico, ecc.); diverse strutture dei materiali (testi, questionari, ecc.). Proprio per quanto concerne quest'ultimo punto, è da rilevare che le metodologie computazionali consentono di attivare, in risposta ad una domanda dell'utente (per es., nel caso più semplice, presenza di una certa parola o parte di parola), procedure di ricerca che esplorano contemporaneamente testi, liste di parole, dizionari, bibliografie, risposte a questionari, ecc.

Stanno appena iniziando i primi tentativi per potenziare le strategie di ricerca nelle LDB con strumenti linguistico-computazionali aventi la

funzione di 'fonti di conoscenza' complementari a quelli dell'utente. Presso l'ILC di Pisa (9) alcuni esperimenti hanno dimostrato l'utilità dell'impiego di un lessico automatico per rendere più efficace l'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, esplicitamente rappresentate nel dizionario o automaticamente computabili: relazioni di sinonimia, antonimia, iperonimia, iponimia, e in genere di tassonomia naturale, e/o campi semantici, ecc. Potrebbe essere interessante anche esplorare la possibilità di collegare un dizionario di macchina dell'italiano alle voci dialettali. Si avrebbe così uno strumento analogo a quello auspicato da sociolinguisti e psicolinguisti che vogliono 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 (10).

L'operazione di ricondurre le voci dialettali a un termine 'pivot' italiano, che i ricercatori dell'ALI hanno chiamato 'tipicizzazione' (11), è giusticata, ovviamente, solo in quanto strumento di 'retrieval' dei dati.

Non mancano esperienze e punti di riferimento per metodologie di analisi statistica dei dati dialettali, e per la produzione di carte 'personalizzate' (Sobrero 1985), cioè stampate (o visualizzate sullo schermo) dal calcolatore in risposta a parametri linguistici ed extralinguistici formulati di volta in volta dal singolo utente. Per una descrizione dello stato dell'arte, per esempi applicativi e per rinvii bibliografici si vedano per esempio Göbel (1983), Handler (1983), Putschte e Neumann (1983).

Un gruppo di psicologi (Mac Whinney, 1985) ha messo a punto metodologie per collegare le trascrizioni dei materiali linguistici con i dati audiovisivi raccolti sul campo per mezzo della videoregistrazione.

L'avvento delle tecnologie dei CD-ROM (compact-disc read only memory) faciliterà una larga distribuzione dei dati. Fino ad oggi, lo scenario organizzativo di una LDB prevedeva la memorizzazione dei dati nei dischi o nelle memorie di massa di un grosso centro di calcolo (mainframe), dotato di apparecchiature capaci di mettere a disposizione per la consultazione on-line grandi quantità di dati.

Gli utilizzatori accedono ai dati mediante una rete di terminali collegati telefonicamente al mainframe.

Questa organizzazione può essere ancora da preferire durante il periodo di raccolta. Se la trascrizione dei dati videoregistrati avviene direttamente su supporto magnetico, i dati possono essere consultati immediatamente a distanza mediante terminali 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 matrice per compact-disc, dalla quale può essere prodotto un numero qualsivoglia di copie con una spesa a copia irrisoria (dell'ordine di dollari 10 per un disco di 12 cm. di diametro che contiene fino a 600 milioni di caratteri).

Un lettore di compact disc (del costo inferiore a un milione) può essere collegato on-line a un personal computer di qualsiasi tipo. Ogni utente potrà così avere la LDB (o più LDB) a disposizione sul proprio tavolo di lavoro.

Il problema di più difficile e laboriosa soluzione sembra essere l'analisi linguistica dei materiali raccolti.

Gli spogli di testi condotti fino ad oggi nei diversi paesi sono eseguiti per lo più a livello lessicale. La fase di analisi in questo caso risulta relativamente semplice e può essere addirittura, in un certo senso, omissa, perché la parola, così com'è definita per il calcolatore (una sequenza di caratteri tra due spazi), può costituire, in mancanza di altro, un'approssimazione accettabile — almeno operativamente — delle unità lessicali oggetto delle ricerche ⁽¹²⁾.

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 ⁽¹³⁾;

— il tempo e il costo di esecuzione dell'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;

— il costo e lo sforzo richiesto da tali analisi inducono ⁽¹⁴⁾ ovviamente a porre l'interrogativo se sia o no possibile analizzare un corpus 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 sintattiche diverse.

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

D'altra parte, è ovvio che un corpus di dati descritto a vari livelli linguistici permette ricerche e interrogazioni molto più utili e interessanti.

Anche per questi motivi, oltre che per l'ovvia esigenza di rendere operativamente intercambiabili i materiali raccolti nell'ambito di progetti condotti in tempi e in centri diversi, 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 dei materiali ma anche della loro descrizione. 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 progressivo arricchimento di un patrimonio comune.

ABBREVIAZIONI BIBLIOGRAFICHE

Berruto 1980

G. Berruto, *La variabilità sociale della lingua*, Torino

Berruto 1983

G. Berruto, *Per una caratterizzazione del parlato: l'italiano parlato ha un'altra grammatica?* in Holtus-Radtke 1985, 154-184

Besch *et alii* 1983

W. Besch-H. Knop, W. Putschke-H.E. Wiegand (a c. di), *Dialektologie*, Berlino

Calzolari-Picchi 1985

N. Calzolari-E. Picchi, *Textual perspectives through an automatized lexicon*, in «Proceedings of the XII International ALLC Conference (Méthodes quantitatives et informatiques dans l'étude des textes)», Genève

Cappelli *et alii* 1978

A. Cappelli-G. Ferrari-L. Moretti-I. Prodanoff, O. Stock, *Parsing an Italian text with an ATN parser*, Pisa

Cardona 1976

G.C. Cardona, *Introduzione all'etnolinguistica*, Bologna

Cedergren-Sankoff 1974

H.J. Cedergren-D. Sankoff, *Variable rules: performance as a statistical reflection of competence*, «Language» 50, 333-355

Cortelazzo 1969

M. Cortelazzo, *Avviamento critico allo studio della dialettologia italiana*, vol. I: *Problemi e metodi*, Pisa

De Camp 1973

D. De Camp, *What do implicational scales imply?* in C.J.N. Bailey-R.W. Shuy (a c. di), *New ways of analyzing variation in English*, Washington, 141-148

Dittmar *in stampa*

N. Dittmar, *Variatio delectat*, Lecce

Dittmar-Schlobinski *in stampa*

N. Dittmar-P. Schlobinski (a c. di), *Sociolinguistics of Urban Vernacular*, Berlino

Dorian 1977

N.C. Dorian, *The problem of the semi-speaker in language death: evidence from East Sutherland Gaelic*, «Language» 54, 590-609

Göbel 1983

H. Göbel, *Ansätze zu einer computativen Dialektometrie*, in Besch *et alii* 1983, 778-792

Grassi 1980

C. Grassi, *Von der Sprachgeographie zur Soziolinguistik. Vergleich von Erfahrungen und Erlebnissen in der Bundesrepublik und in Italien*, in «Zeitschrift für Dialektologie und Linguistik» 47, 145-159

Grassi 1983

C. Grassi, *Spunti di discussione per un 'Atlante linguistico italiano per regioni'* in «Quaderni dell'Atlante lessicale toscano», 1, 27-33

Gumperz-Hymes 1972

J.J. Gumperz-D. Hymes (a c. di), *Directions in Sociolinguistics: the ethnography of communication*, New York

Handler 1983

H. Handler, *Eintwürfe zu dialektalen Informationssystemen*, in Besch *et alii* 1983, 792-806

Holtus-Radtke 1985

G. Holtus-E. Radtke (a c. di), *Gesprochenes Italienisch in Geschichte und Gegenwart*, Tübingen

Van Hout 1984

R. Van Hout, *The need for a theory of choice in sociolinguistics*, «Linguistische Berichte» 90, 39-57

Klein *in stampa*

W. Klein, *The unity of a vernacular. Some remarks on «Berliner Stadtsprache»* in Dittmar-Schlobinski *in stampa*

Klein-Dittmar 1979

W. Klein-N. Dittmar, *Developing Grammars. The acquisition of German Syntax by Foreign Workers*, Berlino

Labov 1972

W. Labov, *Language in the Inner City*, Philadelphia

Milroy 1980

L. Milroy, *Language and social networks*, Londra

Putschte-Neumann 1983

W. Putschte-R. Neumann, *Ansätze zu einer computativen Dialektometrie*, in Besch *et alii* 1983, 748-778

Rohlf's VDS

- G. Rohlfs, *Vocabolario dei dialetti salentini*, Galatina 1976 (ed. orig. 1956-61)
- Romaine 1982
S. Romaine, *Socio-historical linguistics. Its status and methodology*, Cambridge
- Sobrero 1985
A.A. Sobrero, *Per una prima raccolta sistematica di dati sull'italiano parlato in Salento* in Holtus-Radtke 1985, 77-85
- Sobrero *in stampa a*
A.A. Sobrero, *Il progetto NADIR*, in stampa negli Atti del XVII Congresso Internazionale di Linguistica e Filologia Romanza (Aix-en-Provence, 29 agosto-3 settembre 1983)
- Sobrero *in stampa b*
A.A. Sobrero, *Microconvergences in the conversation, between dialect and language*, in corso di pubblicazione in P. Auer-A. Di Luzio (a c. di), *Convergence and Variation*, Berlino
- Sobrero *in stampa c*
A.A. Sobrero, *Italienisch: Regionale Varianten* in «*Lexicon der Romanistischen Linguistik*», n. 274, in corso di pubblicazione, Tübingen
- Sobrero-Romanello 1981
A.A. Sobrero-M.T. Romanello, *L'italiano come si parla in Salento*, Lecce
- Sobrero-Romanello-Tempesta *in stampa*
A.A. Sobrero-M.T. Romanello-I. Tempesta, *I metodi d'inchiesta per l'italiano regionale: osservazioni e proposte dal laboratorio del NADIR Salento*, in corso di pubblicazione in Atti del XVIII Congresso Internazionale della SLI (Padova-Vicenza, settembre 1984), Roma
- Stehl *in stampa*
T. Stehl, *Il problema di un Italiano Regionale in Puglia*, in corso di pubblicazione in Atti del XVIII Congresso Internazionale della SLI (Padova-Vicenza, 14-16 settembre 1984), Roma
- Tuson *in stampa*
A. Tuson, *El repertori lingüistic de la ciutat de Barcelona*, Roma
- Mac Whinney 1985
B. Mac Whinney (a c. di), *Transcript Analysis*, II, 2
- Zampolli 1983
A. Zampolli, *Lexicological and lexicographical activities at the Istituto di Linguistica Computazionale*, in Zampolli-Cappelli 1983, 237-288
- Zampolli-Cappelli 1983
A. Zampolli-A. Cappelli (a c. di), *The possibilities and limits of the Computer in producing and publishing dictionaries*, Pisa

NOTE

(*) I paragrafi 1, 2 e 3 sono di A.A. Sobrero, il paragrafo 4 di M.T. Romanello, il paragrafo 5 di A. Zampolli. I tre autori costituiscono il 'gruppo progetto' del Nuovo Atlante dei Dialetti e dell'Italiano per Regioni (NADIR).

(¹) Aspetti generali e particolari del progetto NADIR sono già stati esposti in Sobrero

ro 1985, Sobrero *in stampa a* e in Sobrero-Romanello-Tempesta *in stampa a*, ai quali si rinvia, a completamento della presente esposizione.

(2) Opposta l'opinione di Stehl *in stampa*.

(3) Per le notizie sui questionari previsti per il NADIR, si rinvia a Sobrero *in stampa a*, Sobrero 1985, Sobrero-Romanello-Tempesta *in stampa*.

(4) Sobrero-Romanello-Tempesta *in stampa*: cfr. in particolare il paragrafo relativo a *I punti d'inchiesta*.

(5) Le forme sono registrate in Rohlf, *VDS* ss.vv.

(6) Cfr. Rohlf, *VDS*, s.v.

(7) Cfr. Rohlf, *VDS*, s.v. *stajare*.

(8) L'utilizzazione sistematica delle tre tecniche ha comunque un vantaggio: consente di individuare le forme tradizionali più salde nella competenza dialettale degli informatori. I casi di uniformità di risposta, ad uscita unica (quando, cioè, lo stesso unico tipo ricorre nei tre informatori), benché poco numerosi, segnalano quanto ancora di unitario resiste nel punto (tra i nostri dati, ad esempio, *tiestu* per 'pentola di creta' a S. Pietro V.; *nazzicare* per 'cullare' a S. Vito, ecc.).

Se poi lo stesso tipo ricorre, nelle stesse condizioni, in più punti, è un fatto che aiuta a capire qualcosa di più della dialettica tra varietà e unità in un'area, anche piccola: fra i tipi raccolti, si segnala in questa prospettiva *krai* per 'domani', presente in tutti i punti, in tutte le inchieste; *lajanaru / lianaru / lakanaru* per 'matterello' da S. Vito a Lecce.

(9) Si vedano per es. Calzolari-Picchi (1985), Zampolli (1983).

(10) Si veda, per questo argomento e per altri a proposito dei quali si citano più avanti lavori di psicolinguisti, Mac Whinney (1985).

(11) Si vedano, su vari numeri del *Bollettino dell'ALI*, gli articoli per il progetto elettronico dell'ALI.

(12) Ovviamente, non si mette in discussione in alcun modo l'utilità di lemmatizzare, qualora sia possibile sul piano economico, il corpus (cfr. Zampolli 1983).

(13) Si pensi, per esempio, alla complessità dell'eventuale uso di parentesi incassate per linearizzare la rappresentazione di una struttura sintattica.

(14) 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 considerevolmente tempi e costi dell'analisi. Diversa è la situazione per quanto concerne il linguaggio orale e, in particolare, materiali dialettali.