

ESTRATTO

SLI

SOCIETÀ DI LINGUISTICA ITALIANA

DIECI ANNI
DI LINGUISTICA ITALIANA
(1965-1975)

a cura di

DANIELE GAMBARARA e PAOLO RAMAT

BULZONI ROMA 1977

E. 1. 1. 90

A. ZAMPOLLI

Trattamento automatico di dati linguistici e linguistica quantitativa

1. PREMESSA

Come G. Lepschy ha giustamente rilevato, questo settore è stato spesso considerato come «staccato dal corpo centrale della linguistica». (Lepschy 1966 p. 189). Sembra pertanto opportuno premettere alcune precisazioni terminologiche e alcuni cenni storici sul suo sviluppo in Italia prima del 1965.

1.1. *Premesse terminologiche.* Termini come *linguistica computazionale*, *linguistica matematica*, *linguistica quantitativa*, *linguistica algebrica*, sono stati spesso usati indifferentemente. In effetti, la parziale sovrapposizione delle aree da essi designate può spiegare questa incertezza terminologica. Non si vuole qui definire l'area pertinente a ciascun termine, (cfr. Zampolli 1977C) ma solo chiarire l'uso che ne viene fatto, e di conseguenza l'area che la presente rassegna intende coprire. Nel titolo ho preferito usare il termine *trattamento automatico di dati linguistici* (d'ora in poi LDP, dall'espressione *linguistic data processing*) anziché quello più comunemente usato di *linguistica computazionale* (*computational linguistics*), perché esso si propone come più generale, inglobando tutti gli usi teorici e applicati di tecniche computazionali o di calcolatori nella elaborazione (*processing*) dei linguaggi naturali: dalla elaborazione di modelli fonologici, morfologici, sintattici, semantici, pragmatici, ecc., alla formalizzazione, rappresentazione e calcolo delle strutture linguistiche (esempi tipici sono i sistemi per il *parser* sintattico o semantico automatico) alla elaborazione con i calcolatori di testi, corpora, e in genere banche di dati linguistici, sovente di grandi dimensioni (esempi tipici sono la produzione di concordanze a scopi lessicografici o la ricerca di parole «chiave» nella documentazione automatica). *Linguistica matematica*, come bene avvertiva G. Lepschy nell'opera citata, è un'«etichetta che ha almeno due valori distinti. Uno di «linguistica quantitativa», in cui entrano considera-

zioni numeriche, e che potremo chiamare in senso lato « statistica linguistica »; l'altro di linguistica per così dire « algebrica » o « algoritmica » (...); in questo secondo senso il termine « matematica » è usato come nell'espressione « logica matematica » (...). La grammatica trasformazionale (...) è un esempio di linguistica algebrica in questo senso non quantitativo». (Lepschy 1966 p. 191).

Sono evidenti, come del resto traspare immediatamente dal termine « algoritmica », le strettissime relazioni tra la linguistica algebrica e il LDP. Per quanto concerne la presente rassegna diremo solo che un importante elemento di differenziazione è, rappresentato dal fatto che il LDP si interessa non solo alla formalizzazione delle descrizioni linguistiche ai vari livelli, ma anche alle procedure e strategie che permettono di calcolarle, e cioè di assegnarle « meccanicamente ».

Sembra chiaro che la linguistica algebrica rientri nel capitolo sulle formalizzazioni in linguistica 1965-1975 di questo volume. La presente rassegna è perciò limitata al LDP e alla linguistica quantitativa.

1.2. *Cenni storici.* Mentre l'esempio del lavoro pionieristico di R. BUSA (1951) nel settore degli spogli elettronici (il suo primo esperimento risale al 1949 e nel 1953 egli fondava il CAAL — Centro per l'Automazione dell'Analisi Linguistica — di Gallarate) veniva seguito, a cavallo del 1960, da Centri di molti altri paesi europei, (tra questi, l'Istituto di Lingua e Letteratura italiana dell'Università di Utrecht lavorava allo spoglio di testi italiani — cfr. Alinei 1968), in Italia invece, nel periodo 1950-1965, oltre alla attività del CAAL, si registrano due soli progetti: la « omogeneizzazione degli indici del Frank e dello Spanke » (Roncaglia, Pacifico, 1962), e la tesi di Zampolli (1960), che costituisce il primo esempio di spoglio elettronico di un testo italiano (*Veglia d'Armi* di Diego Fabbri). Condotta a livello fonemico, morfologico e lessicale, si proponeva essenzialmente indagini di statistica linguistica (Busa et alii 1962, Tagliavini 1962, Zampolli 1968A).

Per superare lo scetticismo nel nostro paese, è stato decisivo l'esempio autorevole della Accademia della Crusca, la quale iniziava la serie degli spogli per la costituzione dell'archivio lessicale, previsto per la redazione del *Vocabolario storico*, presso il CNUCE (Centro Nazionale Universitario di Calcolo Elettronico dell'Università degli Studi di Pisa, oggi Istituto del CNR), che, fondato, nel 1965, aveva celebrato la propria inau-

gurazione pubblicando gli *Indici e le Concordanze della Divina Commedia* (Tagliavini 1965). Nello stesso anno, anche il CAAL trasferiva presso il CNUCE le elaborazioni dell'*Index Thomisticus* (Busa, Zampolli 1968).

Nel 1967 l'Accademia Nazionale dei Lincei inseriva in un Convegno sul tema «L'Automazione elettronica e le sue implicazioni scientifiche, tecniche e sociali» una sezione specializzata (Duro, Zampolli 1968; Tagliavini 1968).

Da quel momento i progetti si moltiplicavano rapidamente (Zampolli 1968B), e nel 1968 erano così numerosi che A. Zampolli poteva costituire presso il CNUCE la *Divisione Linguistica*, (d'ora in poi DLC) con un duplice scopo istituzionale: sviluppare in proprio la ricerca nei diversi campi del LDP e in genere della applicazione dei calcolatori alle scienze umane; promuovere la diffusione delle tecniche di elaborazione elettronica e assistere i progetti italiani del settore (Faedo 1973).

Il fatto che fino ad oggi tutti, o quasi, i progetti italiani svolgono la loro componente informatica presso la DLC, ha avuto, come importante conseguenza, sul piano scientifico-tecnico, la adozione di uno schema comune per la registrazione dei dati in *machine readable form*, la quale ha condotto alla costituzione di una grande biblioteca elettronica di testi, registrata ed elaborabile con uniformità di strumenti e criteri tecnico-scientifici. Sul piano pratico, permettendo un radicale abbassamento dei tempi e dei costi, ha consentito la moltiplicazione dei progetti di spoglio, con un ritmo che sembra decisamente superiore a quello di altri paesi.

Della crisi generale che nei primi anni '60 colpì pressoché dovunque la traduzione meccanica, risentirono anche l'unico progetto italiano del settore, quello del Centro di Cibernetica e Attività Linguistiche dell'Università di Milano, diretto da S. Ceccato (Ceccato 1962) e i progetti di «information retrieval» ad esso connessi. I lavori di questo centro ebbero però il merito di aiutare ad identificare correttamente le reali difficoltà della traduzione meccanica anticipando, in un certo senso, le posizioni di quelle correnti del LDP che oggi lavorano nel cosiddetto paradigma dell'«intelligenza artificiale».

Per lo sviluppo della linguistica quantitativa in Italia si veda Tagliavini (1968), ben documentato.

2. TRATTAMENTO AUTOMATICO DI DATI LINGUISTICI

I progetti sono così numerosi che sarebbe impossibile elencarli tutti. Ci limiteremo perciò ad indicare i principali settori di ricerca, portando per ciascuno alcuni esempi tratti, salvo quelli di Utrecht, dalla Banca Nazionale di Dati Linguistici della DLC, che contiene oltre 5000 testi in più di 30 lingue.

2.1. *Archivi lessicali per la compilazione di grandi dizionari storici.* Il progetto più significativo è senz'altro la costituzione degli archivi per il *Tesoro della lingua italiana delle origini* e per il *Vocabolario storico della lingua italiana*, ad opera della Accademia della Crusca.

L'archivio è articolato in tre forme principali. (1) Una biblioteca di nastri magnetici conserva la registrazione dei testi che sono stati tradotti in *machine readable form*. L'insieme di queste registrazioni, conservate presso la DLC, rappresenta un patrimonio di valore incommensurabile, perché costituisce la base di tutta una serie di possibili future elaborazioni. (2) L'Accademia della Crusca conserva le concordanze integrali per forma di tutte le opere sottoposte a spoglio elettronico. (3) L'Archivio delle schede contesto conterrà tutte e solo le occorrenze prescelte per costituire la documentazione di base per la redazione delle voci. (Duro, 1973 e 1975).

Di natura parzialmente diversa è il SEIOD (Spogli Elettronici dell'Italiano delle Origini e del Duecento), il ben noto progetto dell'Università di Utrecht, che si articola in molti livelli di analisi, ma che fino ad ora ha pubblicato essenzialmente l'Index Locorum e la lista di frequenza, entrambi non lemmatizzati, dei testi elaborati elettronicamente. Con criteri analoghi viene elaborato, ad Utrecht, un corpo di romanzi italiani del Novecento.

2.2. *Archivi lessicali per la redazione di Dizionari relativi a discipline particolari.* A titolo di esempio cito il progetto per il *Vocabolario Storico della Lingua Giuridica Italiana* (Ciampi, 1973) promosso dall'Istituto per la Documentazione Giuridica del CNR di Firenze.

2.3. *Lessici e spogli di autore.* Raggruppiamo sotto questa etichetta spo-

gli di opere letterarie di lingue diverse condotti per lo più da studiosi che operano nell'ambito delle ricerche letterarie, stilistiche e filologiche.

Alla elaborazione elettronica viene richiesta per lo più la produzione di concordanze, di rimari, di indici di frequenza. Le metodologie sono quelle ben note, lo spoglio è per lo più integrale: semmai, in sede di pubblicazione, si eliminano le parole grammaticali. È da osservare tuttavia che, essenzialmente per motivi economici, in particolare per i costi di stampa, la maggior parte degli spogli non viene pubblicata. Le pochissime eccezioni: Dante (Tagliavini 1965), Grasso Legnaiuolo, Manzoni e Petrarca (Accademia della Crusca 1967, 1968, 1971) Seneca (Busa-Zampolli 1975) La Bilancetta (Lessico Intellettuale Europeo 1975) La Costituzione (Bartoletti 1971) sono dovute in buona parte ad azioni di tipo «promozionale». Questa situazione rende urgente la messa in opera di nuovi strumenti per la diffusione e l'utilizzo dei dati memorizzati: microphiches prodotte su richiesta, consultazione interattiva, ecc.)

Può essere interessante segnalare qualche esperienza di concordanze parzialmente diverse dalle forme tradizionali.

Le concordanze che chiamiamo *contrastive* si sono rivelate utili per confrontare un testo di base con un insieme di testi satelliti (per esempio un originale con le sue traduzioni in altre lingue, o con le sue successive redazioni) (Penazzo 1968, su alcune traduzioni del *Faust*).

Le concordanze *sintattiche* hanno come esponente un classificatore o un insieme di classificatori dei tipi sintattici e come contesto una porzione di periodo delimitata con regole sintattiche (De Mauro-Policarpi 1971 operano un confronto tra italiano popolare e italiano letterario).

Le concordanze che chiamiamo *tematiche*, di solito richieste per l'analisi delle strutture di testi poetici di tradizione orale, hanno, come esponente, codici che designano i temi e i motivi della struttura compositiva.

Il progetto di indici e concordanze (1973 e 1976) sulla lingua poetica in Italia avanti la fine del XIII secolo, di D'Arco Silvio Avalle, è particolarmente interessante, perché presenta un ciclo completo di operazioni che, in interazione con il calcolatore, portano dalla trascrizione dei manoscritti alla pubblicazione, in fotocomposizione, del testo riprodotto diplomaticamente e delle sue diverse analisi (indici, concordanze, ecc.).

Un problema non completamente risolto è quello del *trattamento delle varianti* negli spogli. Si sono sperimentati metodi diversi. La soluzione più semplice prevede unicamente l'apposizione di un contrassegno alle

parole per avvertire che hanno delle varianti che non vengono specificate (Busa-Zampolli, 1975). Soluzioni più complesse prevedono che tutte le informazioni dell'apparato critico vengano inserite nelle elaborazioni. In questo modo, le singole varianti e i rapporti tra le singole varianti sono presentati esplicitamente e perspicuamente nelle concordanze, negli indici, nelle statistiche. I progetti più avanzati sono lo spoglio delle tre redazioni dell'*Orlando Furioso* di L. Ariosto diretto da C. Segre (Segre-Zampolli 1976) e lo spoglio delle diverse versioni dell'*Itinerarum Antonini Placentini*, a cura di C. Milani.

2.4. *Lessici di testi filosofici.* In questi progetti i risultati dello spoglio vengono utilizzati non solo per la compilazione di glossari, lessici d'autore, dizionari filosofici, ecc., ma anche come strumenti documentari che facilitano il reperimento degli argomenti trattati nei testi filosofici. L'*Index Thomisticus*, iniziato oltre 25 anni orsono, giunge ormai alla conclusione: sono stati pubblicati fino ad oggi 40 dei 60 volumi previsti. Si avvia alla conclusione anche la compilazione delle voci (6.000/7.000) del *Lessico Rosminiano*, curata dal «Centro di ricerca dei problemi filosofici e religiosi dell'800 italiano e francese del CNR» (Genova) sulla base di un archivio di circa 350.000 schede contestate (Ottonello 1976). Sono in pieno svolgimento gli spogli del «Centro di studio per il Lessico Intellettuale Europeo» (Roma), che cura una serie di lessici di autore. I materiali prodotti dagli spogli elettronici (le *Meditationes* e l'*Aesthetica* di Baumgarten, le opere italiane di Giordano Bruno e Galileo Galilei, le opere latine e italiane di Vico, e inoltre opere di Bacone, Descartes, Leibniz, Spinoza, Hume, Tatens, ecc.) assieme ai risultati di numerosi altri spogli manuali, verranno utilizzati anche per la compilazione di un lessico del linguaggio filosofico dei secoli XVII e XVIII (Gregory 1976).

2.5. *Testi e archivi per ricerche di linguistica storica.* Esempio a mio avviso è il programma tracciato da A. Prosdocimi (1975) relativo a *corpora* epigrafici etruschi e italici. Egli ha già condotto alcuni esperimenti sulle *Tavole Iguvine*, ottenendo concordanze, indici diretti e inversi, e dati statistici di diversa natura (Prosdocimi 1970).

I problemi di classificazione delle lingue su base quantitativa sono studiati con eccezionale competenza da P. Fronzaroli (1975, 1976). Egli applica raffinate tecniche statistiche per misurare la somiglianza tra

strutture linguistiche diverse, e per costruire modelli esplicativi e procedure euristiche, soprattutto nel campo delle lingue semitiche.

Anche E. Campanile (Campanile-Zampolli 1976) è interessato al significato euristico delle misure statistiche in uno studio sul profilo etimologico del cornico antico. Egli ha iniziato poi, in collaborazione con M. Ott, la registrazione in *machine readable form* del Dizionario dell'Antico Irlandese.

Alcune operazioni in questo settore possono risultare tecnicamente molto complesse. Per esempio, per la produzione dell'archivio lessicale dei testi storici in ittito cuneiforme curato da P. Meriggi (Meriggi 1973), si sono dovuti risolvere i problemi posti dalla complessità, a livello grafico, dei documenti.

Il progetto di lessicografia sanscrita appena iniziato da M. Ott con lo spoglio del *R̥gveda* ha dovuto risolvere anzitutto il problema di ottenere, a partire da una unica codificazione del testo in *machine readable form*, due forme di trascrizione, una in *pada patha* e una in *samdhi*, sulle quali si possa lavorare contemporaneamente, assumendo la seconda nelle operazioni a livello sintagmatico e contestuale, la prima nelle operazioni a livello di forma lessicale.

Il lavoro condotto in collaborazione tra il Department of Eastern Languages della University of California at Los Angeles, l'Istituto per gli studi Micenei del CNR e la DLC, ha sviluppato, oltre a un accurato sistema di codificazione, un sistema per la riproduzione grafica dei testi mediante calcolatore, e un algoritmo per la analisi morfologica automatica (Buccellati-Saporetti 1976).

2.6. *Banche di dati dialettali.* Sono state già pubblicate alcune analisi del *Questionario* dell'ALI (Atlante Linguistico Italiano, 1971, 1973).

La memorizzazione delle risposte dialettali, che si sta ora iniziando, dovrebbe consentire di comporre automaticamente indici e carte e di tentare l'impiego di procedure euristiche tendenti a suggerire analogie e contrasti tra i fenomeni registrati (Grassi 1973).

2.7. *Elaborazione di archivi testuali a scopi documentari.* Raduniamo sotto questo titolo progetti condotti nell'ambito di discipline diverse, i quali hanno in comune il fatto di elaborare i testi non solo (o non princi-

palmente) per lo studio della loro lingua, ma anche (o soprattutto) per accedere più facilmente alle informazioni in essi contenute.

L'interesse linguistico è ancora nettamente prevalente negli spogli di testi grammaticali, di commenti, di scolia, etc. (per es., nei progetti da poco iniziati sull'edizione dei *Grammatici Latini* del Keil presso gli Istituti di Filologia Classica di Torino e Milano e sull'edizione teubneriana degli *Scholia Pyndari* presso l'Istituto di Filologia Classica di Pisa). Prevale invece nettamente l'interesse documentario nella elaborazione di fonti storiche di tipo letterario, soprattutto medioevale e di fonti epigrafiche (cfr. Baldacci et alii 1976).

I progetti nel campo dell'*information retrieval* propriamente detto, anche se non si propongono alcun intento linguistico, in alcuni casi, e cioè quando operano la ricerca delle informazioni richieste sull'intero corpus dei documenti che costituiscono la base informativa (il cosiddetto sistema del *full-text*), impiegano procedure e producono conoscenze ed esperienze che possono interessare i linguisti (Perschke 1975).

Con il titolo «Corpus sistematico delle tradizioni italiane» viene designato un progetto di sistematizzazione documentaria delle informazioni attualmente disponibili sul folklore italiano (Cirese 1973). La costruzione di numerosi inventari (incipit, rimari, formulari, ecc., in parte già realizzati) costituisce l'aspetto documentario che ci ha indotti a collocare il progetto in questo paragrafo. Esso però si propone anche alcune ricerche estremamente interessanti le quali, riunendo componenti presenti in molti altri settori del LDP, gli meriterebbero una collocazione a parte.

2.8. *Banca Internazionale di Dati Linguistici*. La registrazione di testi in *machine readable form*, la loro analisi linguistica (anche a livelli molto semplici, come la lemmatizzazione), gli strumenti linguistico-informatici per le elaborazioni (programmi per il calcolatore, dizionari di macchina per la lemmatizzazione semiautomatica, ecc.) presentano un elevato costo di tempo e denaro. La DLC ha promosso un progetto per una *banca internazionale dei dati*, che ha ricevuto in seguito il patrocinio di importanti organizzazioni internazionali. I rapporti già stabiliti con centri operanti in diversi paesi europei e nordamericani permettono fin d'ora ai ricercatori italiani di accedere senza difficoltà tecniche e organizzative agli archivi di testi elaborati nelle diverse lingue presso tali Centri, tra-

mite i canali tecnico-organizzativi della DLC (Cfr. Zampolli 1976D pp. 174-179).

2.9. *Dizionari di Macchina*. I dizionari di macchina sono di solito concepiti come strumenti che, proponendo automaticamente per ogni occorrenza del testo le analisi lessicali e morfosintattiche possibili secondo il sistema linguistico di cui il testo è una attuazione, alleggeriscono radicalmente il compito e i costi della fase di lemmatizzazione e favoriscono la coerenza dei criteri tra ricercatori diversi (Duro-Zampolli 1968; Durante 1965 p. 143-4; Zampolli 1975). Con questi scopi la DLC ha avviato la preparazione di dizionari di macchina per il latino e per lo spagnolo (Zampolli *et alii*) e ha acquisito dizionari preparati in altri paesi (per es. quello greco di Packard, 1976 e quello latino di Liegi). Il progetto del *Dizionario Italiano di Macchina* (DMI) si propone anche altri scopi. Ognuna delle sue voci (circa 100.000 lemmi desunti dai normali dizionari di consultazione, in particolare lo Zingarelli, 10^a ed.) è corredata da informazioni linguistiche a diverso livello: fonologico, morfosintattico, semantico, ecc.

Queste informazioni vengono utilizzate per ricerche che perseguono l'intento di raccogliere il più gran numero possibile di dati ai diversi livelli del sistema linguistico italiano, e di strutturarli mediante modelli costruiti elaborando euristicamente i dati stessi (cfr. Zampolli 1968C, Ferrari-Prodanof 1976, Zampolli *et alii*).

2.10. *Morfofonologia*. È stato costruito un algoritmo per la flessione automatica il quale rappresenta una descrizione esplicita completa e formalizzata del sistema morfofonologico dell'italiano, interamente verificata mediante la applicazione ai 106.000 lemmi del DMI (cfr. Ferrari 1973). Questa descrizione sta per essere completata con un modello della derivazione alterativa e mediante affissi in italiano (cfr. i contributi di E. Picchi, N. Ruimy, G. Turrini e di J. Prodanof in Zampolli *et alii*, 1976), costituito da regole di derivazione espresse sotto forma di tratti selettivi classificatori, associati ad ogni suffisso del sistema italiano ed ad ogni singola base del DMI. Procedure euristiche applicate al DMI dovrebbero facilitare la identificazione di questi tratti, presumibilmente di natura fonetica sintattica e semantica. L'algoritmo per la trascrizione fonemica automatica delle forme del DMI presenta l'interessante caratteristica di

poter operare, a richiesta, la trascrizione secondo uno qualsiasi dei diversi inventari fonematici dell'italiano fin qui proposti. In questo senso l'algoritmo contiene una descrizione formalizzata e operativamente applicabile delle equivalenze e delle divergenze tra questi inventari (cfr. Zampolli 1973B).

E. Menucci e M. Monreale (1976) hanno affrontato invece il problema della analisi morfologica automatica, partendo dal presupposto che il DM per la lingua su cui si opera non esista ancora, e debba essere costruito in base alle forme trovate nei testi da analizzare.

2.11. *Sintassi*. Non mancano modelli relativamente efficienti e ben noti per il trattamento automatico della sintassi (*parser* e generazione di frasi): come per altre lingue, anche per l'italiano gli sforzi devono essere rivolti soprattutto alla raccolta di fatti e tipi sintattici dell'italiano, dei quali mancano a tutt'oggi inventari soddisfacenti e alla interazione tra componente sintattico e componente semantico. È agli inizi un tentativo di formalizzare la sintassi relativa a un sottoinsieme abbastanza esteso dell'italiano scritto contemporaneo, in modo da poterla descrivere con il formalismo adottato dal GETA di Grenoble e da poterla applicare a un *parser* automatico ispirato all'*Augmented Transition Network* di Woods (1976). Scopo non secondario di questi studi è quello di contribuire a un sistema per automatizzare almeno in parte la distinzione degli omografi (Zampolli 1972). Ci si chiede però se tale scopo non sia più realisticamente raggiungibile con *parser* limitati, i quali non si propongano di ricostruire l'intera struttura di un enunciato, ma solo di identificare «localmente» la struttura dei sintagmi che la compongono. Questo è anche lo scopo del semplice sistema di *parser* descritto da B.D.R. Smith (1976), il quale sviluppa in dettaglio la strategia proposta da M. Alinei per un *parser* dell'italiano contemporaneo.

A un *parser* sintattico dell'italiano ha lavorato anche Von Stechow (1973), con l'intento di mostrare come sia possibile aumentare l'efficienza di una grammatica *context-free* in «forma normale», introducendovi alcune convenzioni per il trattamento dell'accordo grammaticale. Bartoli, Dell'Orco (1974) descrivono un progetto per un *parser* sintattico-semantico allo studio presso il Centro Scientifico IBM di Bari, da impiegare nell'interazione uomo-macchina nella *computer assisted instruction*.

2.12. *Semantica*. Non occorre, credo, insistere sulla totale impossibilità di applicare «in extenso» al lessico di una lingua una qualsiasi delle embrionali teorie formulate nel corso della «tragic history of semantics» (Katz 1972 p. 2). Per rispondere alle esigenze di alcuni utenti, i quali richiedono al DMI l'indicazione non solo delle omografie ma anche la segnalazione delle polisemie, al fine di distinguerle nella lemmatizzazione, si è corredato il DMI con la definizione di tutte le accezioni riportate, per ciascun lemma, dai dizionari di riferimento (in primo luogo lo Zingarelli). Per ciascuna accezione è stata riportata, assieme alla indicazione degli eventuali limiti stilistici o relativi a linguaggi speciali, la definizione lessicale fornita dal dizionario o, più spesso, una sua versione «semiformalizzata». Il progetto di N. Calzolari, L. Moretti, L. Pecchia (Calzolari, Moretti 1976) si propone di mettere a profitto questo corpus di definizioni pressoché unico per le sue dimensioni, con il duplice scopo di recuperare, grazie alla gestione elettronica, la somma di conoscenze, dati, e assunzioni teoriche e metodologiche più o meno coscienti prodotte negli anni dalla tradizione lessicografica e implicite nei dizionari, e di dare una prima strutturazione esplicita al materiale così raccolto. Tale strutturazione dovrebbe essere realizzata per mezzo di un *network* o grafo orientato, nel quale i nodi sono occupati da *markers* semantici e le frecce da relazioni semantiche di varia natura, interpretabili almeno in parte nella logica del calcolo dei predicati. Questa organizzazione dovrebbe permettere una valutazione anche quantitativa del cosiddetto indice di circolarità del lessico (Calzolari 1977).

Cappelli (1976), e Cappelli-Picchi (1976) operano sul corpus dei sinonimi riportati come tali nel DMI. Per individuare ed esplicitare i processi classificatori operanti nella competenza «semantica» del lessicografo.

Anche il lavoro di M. Alinei (1974) ha preso le mosse da una elaborazione elettronica delle definizioni contenute in un dizionario di consultazione: il piccolo *Dizionario Garzanti* della lingua italiana, Firenze 1963, che contiene circa 42.000 lemmi. Per una sua descrizione si veda l'articolo di Berruto in questo volume.

3. LINGUISTICA QUANTITATIVA

Sembra opportuno mantenere la distinzione tra statistiche sui testi e statistiche sul sistema, già introdotta e giustificata teoricamente dalla scuola praghese.

3.1. *Statistica linguistica sui testi.* Gli spogli di testi moltiplicatisi grazie alla diffusione dei calcolatori, hanno mostrato la necessità di modificare i modelli che negli anni '50, con la celebre proporzione «*langue: parole* = universo statistico: campione», avevano tentato di inserire gli studi sugli aspetti quantitativi del linguaggio direttamente al centro della teoria linguistica strutturalista di tipo europeo. Questi modelli, proposti da D. W. Reed (1949), P. Guiraud (1954) e G. Herdan (1960) e lucidamente esposti da L. Heilmann (1963), si sono dimostrati inadeguati quando, crescendo la quantità dei dati a disposizione, è risultato chiaro che le probabilità di occorrenza che essi postulavano e attribuivano alle unità linguistiche, non esistono indipendentemente da fattori tematici e stilistici. Non esistono ancora dei modelli sostitutivi i quali diano alla statistica linguistica una base teorica adeguata, anche se alcuni ricercatori hanno portato recentemente alcuni contributi relativi alle caratteristiche che dovrebbero avere, alle esigenze cui dovrebbero rispondere, ai fatti che dovrebbero spiegare. Questi modelli (Cfr. Zampolli 1969 e 1975), pur non rifacendosi necessariamente alle stesse teorie linguistiche, propongono degli schemi operativi di ricerca molto simili, e concordano sull'urgenza di attuarli concretamente per fare uscire la statistica linguistica dalla crisi teorica attuale.

Riassumendo molto sommariamente, essi propongono di misurare e descrivere accuratamente le frequenze e le variazioni di frequenza del più grande numero possibile di «caratteristiche» linguistiche (unità, strutture, costrutti ai diversi livelli linguistic) nel più grande numero possibile di testi di vario tipo (autore, genere, epoca, argomento, ecc.). Metodi statistici appropriati dovrebbero permettere di suddividere i testi in classi omogenee rispetto ad una certa «caratteristica» (o fascio di «caratteristiche»), e cioè di raggruppare assieme testi nei quali si può ritenere che tale «caratteristica» (o fascio) presenti lo stesso comportamento dal punto di vista quantitativo. Si dovrebbe quindi cercare di individuare se e quali fattori

pragmatici di produzione i testi di un gruppo abbiano in comune: fattori legati all'interlocutore e all'effetto che si vuol produrre, al tema e al contenuto, al genere letterario, ecc. Ne risulterebbe un procedimento euristico ricorsivo mirante a determinare la « natura statistica specifica » delle diverse « caratteristiche »: quali di esse non siano descrivibili statisticamente, quali mostrino una certa correlazione con alcuni fattori di produzione.

Alcuni ricercatori italiani sono stati tra i primi ad avviare concretamente procedure di questo tipo, e sembra che questo orientamento cominci a prevalere anche in altri paesi. Molti dei lavori di statistica linguistica condotti in Italia possono essere ricondotti, che lo dichiarino o no esplicitamente, a una problematica di questo tipo.

Per es., a livello fonemico, la analisi statistica del corpus di italiano « parlato » eseguita da B. Marioni (1974), ha mostrato che anche le frequenze dei fonemi variano a seconda del contesto di produzione (tipo di conversazione: familiare, colta: monologo, dialogo; ecc.).

A livello sintattico, lo studio di De Mauro-Policarpi ha fornito i materiali per mettere a confronto l'inventario dei tipi sintattici e le rispettive frequenze nell'italiano letterario e nell'italiano popolare (De Mauro 1974, Policarpi 1974), ed è stato completato da uno studio sui tipi di proposizione e di periodo nella versificazione italiana tra otto e novecento (Rombi 1974).

A livello lessicale, sono stati elaborati i materiali forniti dal *Frequency Dictionary of Italian Words* (FDIW) di A. Juilland e A. Traversa (1973), dal *Lessico di Frequenza della Lingua Italiana Contemporanea* (LIF) di U. Bortolini, C. Tagliavini, A. Zampolli (1971), dal corpus di quotidiani spogliati da M. Crocetti (1975). Questo confronto ha mostrato alcune straordinarie regolarità nelle variazioni delle frequenze di diversi parametri (lessicali, grammaticali, ecc.) tra i sottoinsiemi che costituiscono i 3 corpora citati (Zampolli 1975). Analoghe variazioni si sono riscontrate tra campioni di linguaggio di diverse categorie di malati mentali (Castrogiovanni 1973). O. Brattö ha a sua volta annunciato (1974) un dizionario di frequenza basato su un corpus di romanzi moderni.

I risultati ottenuti debbono però essere completati con ulteriori ricerche condotte su corpora più vasti. La Banca di testi *in machine readable form* esistente presso la DLC costituisce un materiale prezioso sul quale operare induttivamente. È però indispensabile che la definizione delle

unità linguistiche da «conteggiare» sia coerente nell'intera banca: in particolare è necessario che la lemmatizzazione dei testi di una stessa lingua sia condotta con criteri uniformi o almeno compatibili tra loro.

A questo programma di ricerca portano un importante, anche se indiretto, contributo anche quegli studi che rivolgono la propria attenzione non alla lingua o ai sottoinsiemi che la compongono, ma a un autore o un'opera specifici, e utilizzano le metodologie statistiche per mettere in evidenza fenomeni linguistico-stilistici particolari, spesso studiandoli in relazione all'evoluzione dell'autore nel tempo. Queste ricerche sono caratterizzate spesso da un apparato statistico particolarmente complesso. A titolo di esempio si cita la ricerca, guidata da A. Mioni, sui *patterns* fonostilistici in alcuni sonetti di Shakespeare, di P. M. Bertinetto (1973) sui modelli ritmici dell'endecasillabo dantesco, di G. Sartoni (1974) sulle strutture lessicali della *Divina Commedia*, di A. Finzi su alcune strutture morfosintattiche e semantiche di poeti spagnoli del Novecento, di F. Albano Leoni sulla distribuzione lessicale, in particolare degli omografi, nell'opera del Belli (1970). In genere, possiamo affermare che le procedure di spoglio elettronico forniscono immediatamente, e senza alcun lavoro aggiuntivo, se non altro le liste di frequenze lessicali che costituiscono il dato di partenza per qualsiasi studio stilostatistico. Solo una piccola parte dei dati disponibili grazie allo straordinario sviluppo del *text processing* in Italia è stata sfruttata in questa direzione: il resto è a disposizione della comunità degli studiosi.

Non mancano anche ricerche statistiche basate su spogli manuali. Salvo poche eccezioni (per esempio gli ottimi lavori di M. Cortelazzo (1973) sul vocabolario del primo capitolo dell'*Orlando Furioso* e di O. Andreani (1971) sulle caratteristiche morfosintattiche e lessicali del linguaggio in età scolare) si tratta per lo più di spogli non integrali, volti a rilevare, nel corpus esaminato, le frequenze di pochi elementi specifici. Si tratta di tesi di laurea su opere letterarie, delle quali non è facile avere notizia: di ricerche di tipo psicolinguistico e sociolinguistico, in genere su «protocolli» di tests, registrazioni di interviste, componimenti scolastici, ecc., che si propongono di correlare l'acquisizione e l'uso di alcune *features* linguistiche (lessicali, morfologiche, sintattiche) con variabili quali il coefficiente d'intelligenza, il grado di scolarità, la provenienza regionale, la posizione sociale, ecc. (si vedano per es. i lavori di Andreani (1971) e De Negri Trentin (1971)); di rilevamenti statistici intesi a misurare la

preferenza per una certa unità o costrutto, per lo più in un certo sottoinsieme di italiano: dei giornali, dei romanzi, della pubblicità, ecc. (si vedano per es. i lavori di Van Molle-Marechal (1974) Vagni (1974), Rocher-Tanugi (1974), e quelli preannunciati dal gruppo guidato da M. Dardano (Dardano 1973 p. 203).

Per quanto riguarda contributi e discussioni generali sui fondamenti teorici della linguistica quantitativa si vedano, tra gli altri, i lavori di M. Cortelazzo (1973) T. De Mauro (1962 e 1965) P. Fronzaroli (1975) L. Heilmann (1963, 1971) A. Prodocimi (1975) L. Rosiello (1965) A. Zampolli (1969 e 1975).

3.2. *Statistiche sul sistema linguistico.* Per quanto suggerite e raccomandate da illustri linguisti di diverse scuole, le statistiche sul sistema sono di gran lunga meno note e diffuse di quelle sui testi. Anche se non mancano difficoltà di ordine teorico, le difficoltà che fundamentalmente hanno impedito la realizzazione dei programmi più volte tracciati sembrano essenzialmente operative. In primo luogo, mentre i testi sono, per così dire, «oggetti» dati naturalmente, gli inventari di unità, funzioni, regole devono essere costruiti dal linguista e questo lavoro spesso è lungo e faticoso.

È in corso una serie di ricerche, sui lemmi e sulle forme registrate nel DMI, che vanno da quelle, già concluse, o in via di conclusione, a livello fonemico (frequenza dei fonemi, delle sillabe, di diversi tipi di strutture fonemiche, delle opposizioni fonologiche, ecc.) a quelle appena iniziate sul rendimento dei diversi meccanismi di produzione lessicale per derivazione mediante affissi, allo studio del rendimento dei diversi patterns sintattici dei verbi, alla composizione etimologica, etc. (cfr. Zampolli 1973).

Molto resta ancora da fare, ma la presenza del DMI consentirà di compiere per l'italiano ricerche spesso programmate ma mai eseguite neppure per altre lingue.

4. CONVEGNI E SCUOLE SPECIALI ORGANIZZATI IN ITALIA

Sono contrassegnati con * i Convegni nei quali solo una sezione è stata esplicitamente riservata al settore cui la presente rassegna è dedicata. Di tutti sono stati pubblicati o sono in corso di stampa gli Atti, che sono citati nella bibliografia.

* 1967 L'uomo e la Macchina, XXI Congresso Nazionale di Filosofia, Pisa.

* 1968 L'automazione elettronica e le sue implicazioni scientifiche tecniche e sociali (Accademia dei Lincei) Roma (Accademia Nazionale dei Lincei, 1968).

1968 Seminario Internazionale sui Lessici latini di macchina (Divisione Linguistica CNUCE) Pisa (Busa 1968).

1970 Colloque Internationale sur l'Elaboration Electronique en Lexicologie et Lexicographie (Divisione Linguistica CNUCE) Pisa (Zampolli 1973A)

1970 I^a Scuola Estiva Internazionale di Linguistica Computazionale e Matematica (Divisione Linguistica CNUCE) Pisa (Zampolli 1973)

* 1971 1er Table Ronde Internationale des Directeurs d'entreprises lexicographiques (Accademia della Crusca) Firenze (Accademia della Crusca, 1973)

1972 Le tecniche di classificazione in Linguistica (Accademia dei Lincei) Firenze (Accademia Nazionale dei Lincei, 1975)

1972 II International Summer School «Computational and Mathematical Linguistics» (Divisione Linguistica CNUCE) Pisa

1973 COLING 73, Vth International Conference on Computational Linguistics (Divisione Linguistica CNUCE) Pisa (Zampolli, Calzolari 1973)

1974 III International Summer School «Computational and Mathematical Linguistics» (Divisione Linguistica Cnuce) Pisa (Zampolli 1977B)

* 1975 I Colloquio Internazionale del Lessico Intellettuale Europeo (Lessico Intellettuale Europeo) Roma (Fattori, Bianchi 1976)

* 1976 Logiche, Calcoli, Formalizzazioni e Lingue Storiconaturali (Società di Linguistica Italiana) Catania.

Nel 1977 si terrà la IV International Summer School «Computational and Mathematical Linguistics». Riteniamo vada particolarmente se-

gnalato il contributo dato da questa alla diffusione, in Italia e in Europa, non solo della linguistica matematica e computazionale, ma in genere delle correnti più recenti della linguistica teorica.

5. CONCLUSIONI

Lo sviluppo del settore *elaborazione di testi* sembra, anche in confronto con la situazione di altri paesi, soddisfacente, soprattutto a motivo dei progetti e delle realizzazioni relative alla Banca Nazionale e alla Banca Internazionale di Dati Linguistici. Particolare attenzione sembra dovrà essere data nel prossimo futuro alle seguenti direzioni di lavoro:

- completare l'attuale archivio di testi italiani con corpora provenienti dal XIX e XX secolo, opportunamente scelti e stratificati;
- generalizzare l'uso del DMI come strumento di lemmatizzazione dei testi, in modo da perseguire una comparabilità dei risultati degli spogli che affianchi la standardizzazione già raggiunta nella fase di registrazione;
- mettere a punto metodologie e strumenti informatico-linguistici che permettano di rendere interattiva e semiautomatica la fase di consultazione e utilizzo dei materiali linguistici già memorizzati e ne permettano una diffusione più razionale ed economica (*microfiches*, terminali, ecc.), estendendone così l'utilizzazione a nuove classi di utenti (terminologi, sociolinguisti, ecc.).

Il DMI si presenta come strumento utilissimo per la fase di raccolta e di organizzazione di inventari, anche esaustivi, di conoscenze sul sistema linguistico italiano e sul suo funzionamento. Sembra perciò venuto il tempo per avviare e sviluppare le ricerche nel settore dei modelli per la rappresentazione di strutture e conoscenze pragmatiche, semantiche, sintattiche, e di strategie e algoritmi per il loro « calcolo » a partire da enunciati italiani, e viceversa. A questo proposito, sembra indispensabile una più stretta cooperazione tra linguisti computazionali propriamente detti, linguisti generali e studiosi di linguistica italiana.

Per quanto riguarda la statistica linguistica, la crisi teorica in cui questa disciplina si dibatte per la assenza di modelli adeguati, accomuna

la situazione italiana a quella di altri paesi. Tuttavia la presenza della Banca di Dati e quella del DMI forniscono validi strumenti e incentivi per tentare la costruzione di un nuovo modello esplicativo.

BIBLIOGRAFIA

- Accademia della Crusca (1967), *Concordanze degli Inni Sacri di A. Manzoni*, Firenze.
- Accademia della Crusca (1968), *Novella del Grasso Legnaiuolo. Testo, frequenze, concordanze*, Firenze.
- Accademia della Crusca (1971), *Concordanze del Canzoniere di Francesco Petrarca*, Firenze.
- Accademia della Crusca (1973), *Tavola rotonda sui grandi lessici storici*, Firenze.
- + Accademia Nazionale dei Lincei (1968), *Atti del Convegno sul tema: L'Automazione elettronica e le sue implicazioni scientifiche, tecniche e sociali* (Accademia dei Lincei, Roma, 1967), Roma.
- Accademia Nazionale dei Lincei (1975), *Colloquio sul tema: Tecniche di classificazione e loro applicazione linguistica* (Firenze 1972), Roma.
- F. Albano Leoni (1971), *Concordanze Belliane*, Göteborg.
- M. Alinei (1968), *Spogli elettronici dell'italiano delle origini e del duecento*, The Hague.
- M. Alinei (1974), *La struttura del Lessico*, Bologna.
- Almanacco Letterario Bompiani 1961 (1962), Milano.
- O. Andreani (1971), *Classe sociale, intelligenza e abilità linguistica. Ricerca su bambini di 9-10 anni*, in *L'insegnamento dell'Italiano in Italia e all'estero*, Atti SLI 4, Roma, Bulzoni, 1971, pp. 411-468.
- Atlante linguistico italiano (1971), *Questionario, I, Testo*, Torino.
- Atlante linguistico italiano (1973), *Questionario, II, Indici*, Torino.
- D'Arco Silvio Avalle (1973), *Projet pour une liste des concordances de la langue poétique en Italie avant la fin du XIII.e siècle*, in A. Zampolli (1973A), pp. 19-27.
- D'Arco Silvio Avalle (1976), *La formalisation des graphies dans le domaine de la rime*, in Zampolli, Calzolari (1977).
- A. M. Bartoletti Colombo (a c. di) (1971), *La Costituzione della Repubblica Italiana, Testo, Indici, Concordanze*, Firenze.
- G. Bartoli, P. Dell'Orco (1971), *Un approccio all'analisi automatica della risposta*, in *L'insegnamento dell'Italiano in Italia e all'estero*. Atti SLI 4, Roma, Bulzoni, 1971, pp. 635-638.
- P. M. Bertinetto (1973), *Ritmo e modelli ritmici*, Torino.
- + U. Bortolini, C. Tagliavini, A. Zampolli (1971), *Lessico di frequenza della lingua italiana contemporanea*, IBM Italia.
- O. Brattö, *Un esperimento di linguistica statistica, Un vocabolario italiano di frequenza. Romanzi moderni (1961-1963)*, in *Fenomeni morfologici e sintattici nell'Italiano contemporaneo*. Atti SLI 7, Roma, Bulzoni, 1974, pp. 390-408.
- G. Buccellati, F. Saporetti (1974), *The Old Babylonian Linguistic Analysis Project: Goals, Procedures and First Results*, in A. Zampolli N. Calzolari (1977).

- R. Busa (1951), *Sancti Thomae Aquinatis Hymnorum Ritualium. Varia Specimina Concordantiarum. Primo saggio di Indici di parole automaticamente composti e stampati da macchine IBM a schede perforate*, Milano.
- R. Busa, C. Croatto-Martinolli, L. Croatto, C. Tagliavini, A. Zampolli (1962), *Una ricerca statistica sulla composizione fonologica della lingua italiana parlata eseguita con un sistema IBM a schede perforate*, in International association of logopedics and phoniatrics (1962), *Proceedings of the XIIth International Speech and Voice Therapy Conference*, Padova, pp. 542-562.
- R. Busa (ed.) (1968), *De lexico electronico latino*, Pisa.
- R. Busa, A. Zampolli (1968), *Centre pour l'Automation de l'Analyse Linguistique (C.A.A.L.)*, Gallarate, in *Les machines dans la linguistique*, Prague, pp. 25-34.
- R. Busa, A. Zampolli (1975), *Concordantiae Senecanae*, Hildesheim-New York.
- N. Calzolari, L. Moretti (1976), *A Method for a Normalization and a possible Algorithmic Treatment of Definitions in the Italian Dictionary*, in *Preprints del COLING 76*.
- E. Campanile, A. Zampolli (1976), *Some Experiments in Historical Computational Linguistics*, in Zampolli, Calzolari (1977).
- A. Cappelli (1976), *La transitività del lessico nell'analisi meccanica dei sinonimi*, in «Linguistica e Letteratura», 2.
- A. Cappelli, E. Picchi (1976), *Synonymie Lexicale: un Tentative d'analyse*, in *Preprintis del COLING 76*.
- P. Castrogiovanni, A. Telara (1973), *Primi risultati di un'analisi statistica morfologica e lessicale delle risposte al test di Rorschach nella prospettiva di uno studio dei rapporti tra psicopatologia e linguaggio*, in A. Zampolli (1973) pp. 307-323.
- S. Ceccato (1962), *La storia di un modello meccanico dell'uomo che traduce*, in *Almanacco Letterario Bompiani* (1962), pp. 122-134.
- C. Ciampi (1973), *Les projets de recherche automatique des informations juridiques dans l'Institut pour la documentation juridique du Conseil Nationale des Recherches*, in Zampolli (1973), pp. 249-268.
- A. M. Cirese, *Inventaires et répertoires lexicaux, formulaires et métriques des chants populaires italiens*, in Zampolli (1973A), pp. 209-239.
- M. Cortelazzo (1973), *Il Vocabolario del primo canto dell'Orlando Furioso*, in *Miscellanea II*, Udine, pp. 5-67.
- C. Crocetti (1976), *Criteri di analisi del testo per una statistica del lessico dei quotidiani*, Tesi di Laurea dell'Università di Pisa.
- M. Dardano (1973), *Il linguaggio dei giornali italiani*, Bari.
- T. De Mauro (1962), *Statistica Linguistica*, in *Enciclopedia Italiana, Appendice III*, vol. II, Roma.
- T. De Mauro (1965), *Il Linguaggio nella critica d'arte*, Roma.
- T. De Mauro, G. Policarpi (1971), *Ricerche sulla struttura del periodo italiano*, in *L'insegnamento dell'Italiano in Italia e all'estero. Atti SLI 4*, Roma, Bulzoni, 1971, pp. 683-694.
- T. De Mauro (1974), *Premessa a una raccolta di tipi sintattici*, in *Fenomeni morfologici e sintattici nell'Italiano contemporaneo. Atti SLI 7*, Roma, Bulzoni, 1974, pp. 551-574.

- R. De Negri Trentin (1971), *L'uso di modelli sintattici diversi per esprimere rapporti causali finali in bambini dai 7 ai 10 anni*, in *L'insegnamento dell'Italiano in Italia e all'estero*, Atti SLI 4, 1971, pp. 499-524.
- M. Durante (1975), *La linguistica sincronica*, Torino.
- A. Duro, (1973), *Élaborations électroniques de texts effectuées par l'Accademia della Crusca, pour la préparation du Dictionnaire Historique de la langue italienne*, in Zampolli (1973A), pp. 53-75.
- A. Duro (1975), *L'elaborazione elettronica dei dati linguistici: problemi filologici*, in Accademia nazionale dei Lincei (1975), pp. 11-21.
- A. Duro, A. Zampolli (1968), *Analisi lessicali mediante elaboratori elettronici*, in Accademia Nazionale dei Lincei (1968).
- A. Faedo (1973), *Discorso di Apertura*, in Zampolli (1973A), VII-IX.
- M. Fattori, L. Bianchi (a c. di) (1976), *I Colloquio Internazionale del Lessico Europeo*, Roma.
- G. Ferrari (1973), *Procedés et méthodes pour la création d'un algorithme de flexion de la langue italienne*, in Zampolli (1973) pp. 97-100.
- G. Ferrari, I. Prodanof (1976), *Machine Dictionary and Lexicon*, in *Preprints del COLING 76*.
- Ch. Fillmore (1976), *Scene-and-Frames Semantics*, in Zampolli (1977B).
- P. Fronzaroli (1975), *Problemi di classificazione delle lingue su base quantitativa*, in Accademia Nazionale dei Lincei (1975), pp. 123-141.
- C. Grassi (1973). *Perspectives de l'emploi de l'élaborateur électronique en géographie linguistique et en dialectologie*, in Zampolli (1973A), pp. 233-239.
- T. Gregory (1976), *Rapport sur les activités du « Lessico Intellettuale Europeo »*, in Fattori, Bianchi (1976), pp. 21-43.
- P. Guiraud (1954), *Les caractères statistiques du vocabulaire*, Paris.
- L. Heilmann (1962-1963), *Considerazioni statistico-matematiche e contenuto semantico*, in «Quaderni dell'Istituto di Glottologia» (Università di Bologna), VII, pp. 34-45.
- L. Heilmann (1971), *Corso di Linguistica Teorica*, Milano.
- G. Herdan (1960), *Type-token mathematics. A textbook of mathematical linguistics*, The Hague.
- A. Juilland, V. Traversa (1973), *Frequency Dictionary of Italian Words*, The Hague.
- J. J. Katz, (1972), *Semantic Theory*, Harper and Row, Publishers.
- L'Uomo e la Macchina*, Atti del XXI Congresso Nazionale di Filosofia (Pisa, 22-25 Aprile 1967) (1967) Torino.
- G. Lepschy (1966), *La linguistica strutturale*, Torino.
- Lessico Intellettuale Europeo (1973), *Galileo Galilei, La Bilancetta, testo concordanze indici*, Roma.
- B. M. Marioni (1974), *Studi di statistica fonematica su un corpus di italiano parlato*, Tesi di laurea dell'Università di Pisa.
- M. Menucci, E. Monreale (1976), *An interactive System for Stem-suffix Discrimination in Italian words*, in Zampolli, Calzolari (1976).
- P. Meriggi (1973), *Un lexique de l'hittite cunéiforme*, in Zampolli (1973A), pp. 111-113.

- Ottonello (1976), *Problemi e struttura del Lessico Rosminiano*, in Fattori, Bianchi (1976), pp. 113-127.
- D. Penazzo (1972), *Analisi comparativa di cinque traduzioni del Faust di Goethe condotta con l'ausilio degli elaboratori elettronici*, Tesi di laurea dell'Università di Parma, (non pubblicata).
- S. Perschke, *Automatic Thesaurus Construction for Information retrieval*, in Accademia Nazionale dei Lincei (1975), pp. 109-122.
- G. Policarpi (1974), *Tipi di proposizione e periodo nell'italiano popolare contemporaneo e in Croce*, in *Fenomeni morfologici e sintattici nell'Italiano contemporaneo*. Atti SLI 7, Roma, Bulzoni, 1974, pp. 651.
- A. Prosdocini (1970) *Kutef Persimu āreps arves: analisi interna e problemi redazionali nelle tavole iguivine*, in *Mille. I Dibattiti del Circolo Linguistico Fiorentino (1945-1970)* Firenze, pp. 185-207.
- A. L. Prosdocimi (1975), *Sull'applicazione di metodi statistici e computazionali a corpora epigrafici*, in Accademia Nazionale dei Lincei (1975), pp. 143-157.
- D. W. Reed (1949), *A statistical Approach to quantitative Linguistic Analysis*, in «Word», pp. 235-247.
- J. Rocher-Tanugi (1974), *Modèles syntaxiques dans les titres de journaux et les slogans*, in *Fenomeni morfologici e sintattici nell'Italiano contemporaneo*. Atti SLI 7, Roma, Bulzoni, 1974, pp. 474-490.
- M. Rombi (1974), *Tipi di proposizione e periodo nella versificazione tra Otto e Novecento*, in *Fenomeni morfologici e sintattici nell'Italiano contemporaneo*. Atti SLI 7, Roma, Bulzoni, 1974, pp. 575-650.
- A. Roncaglia, M. Pacifico (1961), *Un esperimento di filologia elettronica: l'omogeneizzazione degli Indici del Franck e dello Spanke*, in *Almanacco Letterario Bompiani*, pp. 135-142.
- L. Rosiello (1965), *Struttura, uso e funzioni della lingua*, Firenze.
- M. Sartoni (1974), *Il Lessico della Divina Commedia. Saggio statistico*, Tesi di Laurea dell'Università di Bologna.
- C. Segre, A. Zampolli (1976), *Le Concordanze diacroniche dell'Orlando Furioso*, in *Ludovico Ariosto: lingua, stile, tradizione*, Feltrinelli, Milano, (in stampa).
- C. Tagliavini (1965), *Concordanze della «Divina Commedia»*, Pisa (Ci si riferisce all'introduzione dell'edizione IBM Italia).
- C. Tagliavini (1968), *Applicazioni dei calcolatori elettronici all'analisi e alla statistica linguistica*, in Accademia Nazionale dei Lincei (1968), pp. 111-118.
- F. Vagni (1974), *La proposizione finale nell'Italiano contemporaneo*, in SLI (1974), pp. 329-337.
- P. Van Molle-Marechal, «Andare» e «venire» ausiliari del passivo, in SLI (1974), pp. 357-372.
- A. Von Stechow (1973), *Syntactic analysis of Italian*, in Zampolli (1973), pp. 541-562.
- T. Winograd (1976), *Five Lectures on Artificial Intelligence*, in Zampolli (1976B).
- W. Woods (1976), *Lunar Rocks in Natural English: Explorations in Natural Language Question Answering*, in Zampolli (1976B).
- A. Zampolli (1960), *Studi di statistica linguistica eseguiti con impianti IBM*, (Tesi di laurea dattiloscritta), Padova.

- A. Zampolli (1968A), *Recherche statistique sur la composition phonologique de la langue italienne exécutée avec un système IBM* in *Les Machines dans la Linguistique*, Prague, pp. 25-34.
- A. Zampolli (1968B), *L'elaboratore elettronico negli studi linguistici* in «Rivista IBM», n. 2, pp. 14-19.
- A. Zampolli (1968C), *Projet d'un dictionnaire italien de machine, Intervention*, in Busa (1968) pp. 109-126.
- A. Zampolli (a c. di) (1973A), *Linguistica Matematica e Calcolatori. Atti del Convegno e della prima Scuola Internazionale*, Firenze.
- A. Zampolli (1973B), *La Section Linguistique du CNUCE*, in Zampolli (1973A), pp. 133-199.
- A. Zampolli (1973C), *L'automatisation de la recherche lexicologique: état actuel et tendances nouvelles*, in «META», XVIII, 1-2, pp. 101-136.
- A. Zampolli (1973E), *Humanities Computing in Italy*, in «Computers and the Humanities», 7, 6, pp. 343-360.
- A. Zampolli (1975), *L'elaborazione elettronica dei dati linguistici: stato delle ricerche e prospettiva*, in Accademia Nazionale dei Lincei (1975), pp. 23-107.
- A. Zampolli, N. Calzolari (eds.) (1977A), *Computational and Mathematical Linguistics*, Olschki, Firenze.
- A. Zampolli (1977B), (ed.), *Linguistic Structures Processing*, Amsterdam, North-Holland, (in stampa).
- A. Zampolli (1977C), *Introduction* in Zampolli (1976A).
- A. Zampolli (1976), *Les dépouillements électroniques: quelques problèmes de méthode et d'organisation*, in Fattori, Bianchi (1976), pp. 173-197.
- A. Zampolli, N. Calzolari, A. Cappelli, G. Ferrari, L. Moretti, L. Pecchia, E. Picchi, I. Prodanof, N. Ruimy, G. Turrini, *Il Dizionario di Macchina dell'Italiano*, in Atti del Convegno: «Logiche, Calcoli, Formalizzazioni e Lingue Storiconaturali, Catania 17-19 Settembre 1976, (in stampa).

INDICE

VII PAOLO RAMAT DANIELE GAMBARARA Premessa

DIECI ANNI DI LINGUISTICA ITALIANA (1965-1975)

Parte prima : STORIA E SITUAZIONE ATTUALE DELLA LINGUISTICA IN ITALIA

- 5 ALFREDO STUSSI Storia della linguistica italiana
15 DANIELE GAMBARARA Tradizione e rinnovamento della linguistica
in Italia
31 LUIGI ROSIELLO Il periodo delle traduzioni

Parte seconda : STORIA E SITUAZIONE ATTUALE DELLE LINGUE D'ITALIA

- 51 FRANCESCO SABATINI Storia della lingua italiana
107 MANLIO CORTELAZZO Dialettologia italiana e italiano popolare

Parte terza : SETTORI DELLA LINGUISTICA

- 127 ALBERTO M. MIONI Fonetica e fonologia
157 RICCARDO AMBROSINI Morfologia
173 GUGLIELMO CINQUE Sintassi
189 GAETANO BERRUTO Semantica
209 ALDO DURO Lessicologia
221 MARIO MEDICI Analisi dell'espressione pubblicitaria

Parte quarta : APPROCCI TEORICI

- 229 PAOLO RAMAT Linguistica storica
247 LORENZO COVERI Sociolinguistica e pragmatica
273 PAOLO LEGRENZI Psicolinguistica
284 LORENZO RENZI Tipologia
291 MARIA-ELISABETH CONTE Linguistica testuale
303 VINCENZO LO CASCIO Linguistica contrastiva
327 RENZO TITONE Didattica delle lingue straniere
339 FRANCESCO LO PIPARO Formalizzazioni linguistiche
349 ANTONIO ZAMPOLLI Trattamento automatico di dati linguistici e
linguistica quantitativa

Parte quinta : LINGUISTICA E ALTRE SCIENZE DEL LINGUAGGIO

- 373 CESARE SEGRE Semiotica
385 MARIA CORTI Dalla stilistica alla semiologia letteraria
395 ALBERTO VARVARO Critica del testo e linguistica

Appendici

- 405 CARICHE SOCIALI DELLA SOCIETÀ DI LINGUISTICA ITALIANA NEL
1966-1977
409 SOCI DELLA SOCIETÀ DI LINGUISTICA ITALIANA NEL 1966-1976
419 PUBBLICAZIONI DELLA SOCIETÀ DI LINGUISTICA ITALIANA
- 437 INDICI I degli autori e delle opere
II dei centri e istituti di ricerca