

AIO

I

PAROLE E NUMERI

Analisi quantitative
dei fatti di lingua

a cura di

Tullio De Mauro
Isabella Chiari



Copyright © MMV
ARACNE editrice S.r.l.

www.aracneeditrice.it
info@aracneeditrice.it

via Raffaele Garofalo, 133 A/B
00173 Roma
tel. (06) 93781065

ISBN 88-548-0040-6

*I diritti di traduzione, di memorizzazione elettronica,
di riproduzione e di adattamento anche parziale,
con qualsiasi mezzo, sono riservati per tutti i Paesi.*

*Non sono assolutamente consentite le fotocopie
senza il permesso scritto dell'Editore.*

I edizione: marzo 2005

INDICE

- 7 Premessa
- 9 Introduzione

LA FONETICA E SISTEMI FONOLOGICI

- 21 Gli studi statistici della forma e della sostanza dei suoni
- 27 *Numeri, fonemi e foni* di Federico Albano Leoni e Giuliana Clemente
- 45 *Coppie minime e misura del rendimento funzionale* di Isabella Chiari
- 67 *La fonotassi statistica dell'italiano: i nessi consonantici* di Isabella Chiari e Silvia Castagna
- 85 *Lunghezza e frequenza delle parole nel lessico e nel testo: il caso del tedesco* di Sabine Koesters Gensini

LA GRAMMATICA

- 103 Lo studio statistico della morfologia e della sintassi
- 107 *Aspetti quantitativi della produttività morfologica* di Livio Gaeta e Davide Ricca
- 125 *La misura delle categorie sintattiche* di Miriam Voghera
- 139 *Tendenze nella sintassi dell'italiano contemporaneo* di Gianna Policarpi e Maggi Rombi
- 157 *Lessico Grammatica dell'italiano* di Annibale Elia
- 171 *Le funzioni del che nell'italiano parlato* di Massimo Aureli

IL VOCABOLARIO DI BASE DELL'ITALIANO

- 187 La struttura statistica del lessico e del vocabolario
- 193 *La nozione di vocabolario di base alla luce della stratificazione diacronica del lessico dell'italiano* di Alessandro Giuliani, Claudio Iacobini e Anna M. Thornton
- 215 *Nuove rilevazioni sul vocabolario di base e di alta disponibilità* di Francesco De Renzo

- 233 *La rilevazione dei gradi di notorietà dei lemmi del vocabolario di alta disponibilità* di Domenico Russo
 247 *Il vocabolario di base dell'italiano degli stranieri* di Fiammetta Carloni e Massimo Vedovelli
 277 *Un lessico di frequenza della LIS* di Tommaso Russo

LA CONOSCENZA E GLI USI DELLE PAROLE

- 293 Le tendenze statistiche nella conoscenza di lessemi, significati e usi
 297 *Quantità dei lemmi nei dizionari* di Tullio De Mauro e Silvana Ferreri
 307 *L'estensione delle conoscenze lessicali individuali* di Silvana Ferreri
 335 *La reperibilità statistica di tendenze diacroniche nell'uso delle parole* di Sergio Bolasco
 355 *La legge di Zipf sul numero dei significati in italiano e inglese* di Fiammetta Carloni

I TESTI E GLI USI

- 373 La struttura statistica dei testi e la sua analisi
 377 *Misurazioni quantitative degli stili verbali e indici di leggibilità* di M. Emanuela Piemontese
 399 *Temperatura informazionale e leggibilità dei testi* di Rita Plantera
 417 *Strumenti automatici di analisi e gestione testuale: IntraText, UTM e Censor* di Nicola Mastidoro e Maurizio Amizzoni
 439 *L'analisi dei testi con INTEX* di Simonetta Vietri
- 449 Riferimenti bibliografici
 487 Gli Autori
- 495 Indice dei nomi
 503 Indice degli argomenti

Premessa

Abbiamo progettato questo volume, discutendone con collaboratrici e collaboratori, nell'autunno del 2003. Il lavoro di coordinamento si è rivelato più complesso di quanto immaginavamo. Ora che è terminato, ringraziamo gli autori e le autrici per la attenzione e pazienza con cui hanno rispettato le indicazioni redazionali. Nella raccolta figurano alcuni contributi di cui noi due curatori siamo autori o coautori. Inoltre la stesura dell'introduzione generale al volume è di Tullio De Mauro e le introduzioni alle sezioni sono di Isabella Chiari.

i.c. e t.d.m.

Roma, 7 febbraio 2005

Introduzione

1. I lavori di analisi statistica dei fenomeni linguistici che qui si raccolgono e l'analisi quantitativa stessa si muovono lungo un crinale reso sottile e impervio da qualche superficialità d'una parte della cultura linguistica e da opposizioni teoriche motivate anche se criticabili e, anzi, da respingere di un'altra parte della linguistica recente.

Il linguista di formazione umanistica rischia di presumere che si tratti di una pretesa matematizzazione della lingua e, se è sensibile alla storicità delle lingue, ritiene di potere e dovere respingere la pretesa. Quell'ampio versante della linguistica teorica contemporanea che invece effettivamente ha elaborato e segue o cerca di seguire una visione calcolistica, algebrica, del funzionamento delle lingue considera irrilevanti le analisi statistiche.

Al linguista di formazione umanistica si deve fare osservare che, se avrà pazienza di accostarsi a questi lavori, potrà constatare che essi vanno in direzione opposta a quella che egli paventa. Come in ogni altro campo delle scienze umane (e non solo) la quantificazione statistica serve a precisare e affinare percezioni intuitive su natura e ruolo, origini e relazioni di fatti accidentali, in definitiva serve, come torneremo più oltre a dire, a rendere evidenti peculiarità idiosincratiche variabili nel tempo delle singole lingue nella loro variabilità. È strumento di analisi qualitative, storiche, non le soffoca, le potenzia e spesso rivela tratti altrimenti nascosti di singole lingue e dello stesso linguaggio.

Accenno solo a due esempi. Dobbiamo alle analisi statistiche di estesi corpora l'emergere di una constatazione: nelle lingue flessive non solo i morfemi singoli, che troviamo in frasi e testi, ma anche i loro aggregati in parole hanno rivelato una inattesa presenza di forme omonime, plurietichettabili sotto il profilo morfologico e semantico. Nella frase precedente tali sono le parole *accenno*, verbo e sostantivo, *solo*, avverbio e aggettivo, *a*, preposizione e nome di una lettera, *analisi*, singolare e plurale, *di*, preposizione e sostantivo nome di una lettera, *estesi*, passato remoto e participio aggettivale, *l'*, articolo e pronomi ecc. L'esperienza dell'applicazione di programmi di lemmatizzazione automatica ci dice che nelle lingue europee la percentuale di omonimi

che troviamo nei testi, omonimi *testuali*, oscilla intorno al cinquanta per cento. In generale è rarissimo che un omonimo testuale turbi la comprensione della frase o del testo. Occorre costruire sequenze *ad hoc* perché un'intera frase risulti equivoca, come avviene nei giochi di parole bisensi o in esempi di scuola, come *Una vecchia porta la sbarra* o *Una giovane legge la regola*. Di solito la nostra comprensione non soffre equivoci perché il nostro leggere o ascoltare non procede atomisticamente, ma tenendo conto degli elementi precedenti e seguenti delle sequenze sintagmatiche. Che così stessero le cose faceva parte del comune sapere intuitivo. Ma come ogni sapere prescientifico non operava poi a tutto campo nella riflessione scientifica sui fatti linguistici.

Il consolidamento di quel sapere attraverso le analisi statistiche ci mette dinanzi a un fatto: il nostro comprendere linguistico non è lineare, non procede, come ancora da qualcuno si dice, dalla individuazione dei fonemi a quella dei morfî e parole, poi a quella dei sintagmi e, infine, alla individuazione del significato (il *Sinn* di Frege) e, nell'ambito di questo, all'attribuzione del senso (la *Bedeutung* di Frege) dell'enunciato che realizza una frase. Il comprendere linguistico procede invece con un continuo movimento di va e vieni, circolare, fondato su un intreccio di ipotesi di senso e significato e di individuazione di apigli, di segmenti di significante, continuamente sottoposti a verifica e spesso, nel procedere, abbandonati e sostituiti, e ciò anche in rapporto a fatti esterni alla forma della frase, dunque non formali, come la diversità del contesto situazionale o le caratteristiche di chi scrive o parla ecc. Il carattere *attivo e problematico*, nient'affatto descrivibile come un calcolo, proprio del comprendere linguistico ci si rivela con una evidenza nuova alla luce dei dati acquisiti attraverso l'analisi quantitativa. E ciò si riflette in una correzione sostanziale dell'idea che una lingua sia assimilabile a un'algebra.

Anche all'analisi quantitativa dobbiamo un'altra acquisizione, quella del grado di ridondanza che caratterizza le lingue. È un grado elevatissimo. A parità di differenziazione di frasi e di loro componenti, le lingue potrebbero prevedere meno fonemi oppure, anche, parole assai meno lunghe o, ancora, frasi assai meno sovrabbondanti nella marcatura di generi, casi, numeri, persone del verbo ecc. L'enorme carico di

ridondanza delle lingue e delle frasi è estraneo ai calcoli. E ciò di nuovo deve fare riflettere in una prospettiva teorica generale¹, che recupera un aspetto importante dell'insegnamento saussuriano: una lingua non è comprensibile e spiegabile in sé ma in quanto vive in funzione delle esigenze espressive e comunicative di una determinata, storica, *masse parlante*, in un tempo determinato.

Ci volgiamo con ciò a chi, invece, della lingua ha cercato di elaborare una visione calcolistica, come è avvenuto in parte dello strutturalismo, specie negli indirizzi generativistici.

È vero: se una lingua fosse un calcolo, come l'aritmetica elementare o un'algebra, le analisi statistiche dei fenomeni linguistici sarebbero una mera curiosità, faticosa come altre forme di curiosità, ma poco rilevante ai fini della comprensione della lingua stessa, così come senza rilevanza ai fini della comprensione dell'aritmetica sarebbe raccogliere un corpus di operazioni aritmetiche e stabilire quante volte vi appare questo o quel numero oppure il segno più ecc. Ma la lingua è assimilabile a un calcolo?

Abbiamo già accennato poc'anzi a aspetti che inducono a una risposta negativa. Non pare opportuno qui indugiare oltre nell'insistere su di essi² e nel ricordare, come ora pare farsi chiaro anche tra studiosi di origine generativista, come Ray Jackendoff³, le insufficienze della riduzione totalizzante e totale della lingua a calcolo.

Si può dunque procedere serenamente sul crinale cui accennavamo all'inizio, sulla via cioè delle analisi statistiche dei fenomeni linguistici.

¹ Lo ha fatto la curatrice di questa raccolta, Isabella Chiari, *Ridondanza e linguaggio. Un principio costitutivo delle lingue*, Carocci, Roma 2002.

² Rinvio a Franco Lo Piparo, *Linguaggio, macchine e formalizzazioni*, Il Mulino, Bologna 1974, e a De Mauro, *Minisemantica dei linguaggi non verbali e delle lingue*, Laterza, Bari-Roma 2000.

³ Cfr. per esempio, Ray Jackendoff, *Linguaggio e natura umana*, Il Mulino, Bologna 1998, p.17, sul carattere di "miscela" tra natura e cultura proprio della lingua, un tema ripreso poi, specie in riferimento al lessico, in scritti recenti di Chomsky: Marc D. Hauser, Noam Chomsky, W. Tecumseh Fitch, *The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?*, «Science» 298, 22 nov. 2002, pp. 1569-1579, in particolare p. 1574.

2. Anche per quanto si è detto, siamo ben consapevoli che l'area intersettiva che unisce il dominio delle comuni parole e il dominio dei numeri (e i numeri del resto sono anche parole), il dominio del linguaggio verbale e quello dei linguaggi formali matematici⁴, è assai vasta. L'analisi quantitativa si occupa solo di un aspetto particolare, anche se a nostro avviso essenziale per la linguistica descrittiva, teorica e educativa. È l'aspetto dell'utilizzazione di valutazioni quantitative nello studio delle lingue, delle loro parti, dei testi e usi cui danno luogo. È su tale utilizzazione che verteranno i contributi di autori e autrici che hanno collaborato al presente volume.

Questa introduzione intende offrire un succinto profilo del progressivo emergere delle analisi quantitative di fatti linguistici e del maturare recente della consapevolezza di quanto la dimensione quantitativa sia rilevante ai fini della determinazione degli aspetti storici e qualitativi dei fenomeni.

Nella Grecia classica già prima di Aristotele, probabilmente anzitutto tra rapsodi e maestri di scuola, l'attenzione si appuntò sulla presenza in Omero e in altri testi canonici di parole come ad esempio *bôtor* "pastore" usate una sola volta nell'*Iliade* (XII 302) e una sola volta nell'*Odissea* (XIV 102), gli *onómata hápax legómena* "parole dette una sola volta" o, come per tempo dovette dirsi, le *glóssai (glóttai* in attico). Di Democrito, pensatore dai marcati interessi linguistici, sappiamo che scrisse sistematicamente *Perì Homérou (...) glosséon*. Da trattazioni del genere in età alessandrina, tra i filologi, partì una lunga

⁴ Sui sistemi di numerazione, di verbalizzazione dei numeri e di notazione in cifre (tre concetti e ordini di fenomeni interrelati ma da ben distinguere, e talora confusi) Carl P. Boyer, *Storia della matematica*, prefazione di Lucio Lombardo Radice, Mondadori, Milano 1980, pp.1 sgg., 10 sgg., 19 sgg., 68 sgg., Giovanni Buffa, *Tra numeri e dita*, Zanichelli, Bologna 1986 e, fondamentale, Georges Ifrah, *Histoire universelle des chiffres. L'intelligence des hommes raccontée par les nombres et le calcul*, 2 voll., Robert Laffont, Parigi 1994². Sulle parole assunte a termini per i numeri naturali, specie in fasi arcaiche e in lingue c.d. primitive, molti contributi in Domenico Silvestri (a cura di), *Numeri e istanze di numerazione tra preistoria e protostoria linguistica del mondo antico*, Atti del Convegno dell'Istituto Orientale, Napoli 1-2 dicembre 1995, «Annali del Dipartimento di Studi del Mondo Classico e del Mediterraneo Antico. Sezione linguistica», vol. XVII; da un punto di vista semiotico generale sui rapporti di analogia e differenza tra le lingue storico-naturali e sistemi di numerazione, sistemi di cifrazione, calcoli, rinvio a De Mauro, *Minisemantica dei linguaggi non verbali e delle lingue*, Laterza, Bari-Roma 2000, *ad indices*.

serie di inventari sistematici di parole rare perché obsolete o perché di scarsissima frequenza, confluiti poi intorno al V sec. d.C. nel lessico di Esichio⁵. Nella tradizione latina, già da un passo di Varrone (*'tesca' aiunt sancta esse qui glossas scripserunt: Ling. Lat.VII 19*), pare di capire che le 'glosse' fossero non tanto e più le parole rare, ma, metonimicamente, le annotazioni esplicative di tali parole, ciò che prelude all'uso moderno di *glossa* nel senso di commento esplicativo epilinguistico o metalinguistico reso da chi sta parlando o scrivendo. Come in molti casi, dobbiamo al genio di Aristotele avere trasformato fatti altrimenti accertati in un dato di rilevanza teorica: la presenza delle *glôttai* nei testi attrae ripetutamente la sua attenzione nella *Poetica* in quanto esse sono la componente principale di quello *xenikón*, di quel "peregrino" che a suo avviso caratterizza lo stile letterario, soprattutto l'eroico. Venti secoli dopo Leopardi, Antoine Meillet o Leo Spitzer hanno ripreso questa idea aristotelica.

I fatti di lingua, soprattutto lessicali, ma anche, come si osserva nella *Poetica*, sintattici, si collocano dunque su due piani, quanto a frequenza: da un lato i fatti rari, di bassa frequenza, le "cose che nessuno direbbe nel parlare" (*hà oudeis àn eîpoi en têi dialéktoi*, Ar., *Poet.*, 1458 b 32), e dall'altra l'uso comune. Oggi abbiamo raggiunto una visione più complessa e lo dobbiamo a un cammino secolare verso la comprensione precisa dell'assai più articolato diversificarsi quantitativo dei fatti linguistici, che non paiono più distinguibili solo nelle due categorie del raro da un lato e del comune dall'altro.

Ma il punto di vista antico ha resistito a lungo. I linguisti e gli altri studiosi che a vario titolo si occupano di linguaggio nascono, lo abbiamo già detto, prevalentemente umanisti, appartengono al versante "letterato" della cultura più che al "numerato". Non stupisce che l'utilizzazione di dati quantitativi sia in genere difficile da promuovere. E non sorprende quindi che, per esempio, nella lessicografia europea moderna, anche in quella più avvertita e attenta, l'ormai ricco patrimo-

⁵ Peter Barr Reid Forbes, Robert Browning, *Glosse, glossari (greco)*, James Frederic Mountford, *Glosse, glossari (latini)*, in Nicholas Geoffrey Lemprière Hammond, Howard Hayes Scullard, Mario Carpitella (a cura di), *Dizionario di antichità classiche Oxford*, Edizioni Paoline, Roma 1981, s.vv., Valerio Casadio, *Lexicographers, Ancient Greek*, in Harro Stammerjohann, *Lexicon Grammaticorum. Who's Who in the History of World Linguistics*, Max Niemeyer Verlag, Tübingen 1996, s.v.

nio di analisi lessicostatistiche continui a essere sottoutilizzato e si trovino usate solo in modo non sistematico soltanto alcune generiche qualificazioni di natura solo intuitivamente statistica. Così, ad esempio, nei dizionari inglesi di Oxford, esemplari per altri aspetti, troviamo *generally* opposto a *especial, especially; frequent, commonly, more commonly, usually, rare, more rare, occasional, occasionally*. La bipartizione antica tra raro e comune tende dunque a persistere, anche se di recente un nuovo punto di vista emerge in una parte almeno delle opere lessicografiche.

A questo nuovo punto di vista si è giunti attraverso vie inizialmente distanti e disparate, che già altrove ho avuto occasione di evocare sommariamente⁶.

Una prima via, percorsa inizialmente per motivi religiosi e, anche, umanistico-filologici, è stata quella degli *indices verborum* (raccolte di lessemi con sola indicazione numerica dei loro luoghi) e delle concordanze (lessemi con luoghi e anche contesti) di testi religiosi, Confucio, Corano, Bibbia, e di testi dell'antichità greco-latina (Cicerone)⁷. In modo preterintenzionale indici e concordanze misero e mettono sotto gli occhi che i dislivelli di frequenza dei lessemi sono numerosi. Alcuni lessemi (congiunzioni, articoli, pronomi) hanno tale frequenza e almeno apparente ovvietà semantica da sconsigliarne spesso in *indices* e concordanze la registrazione che riempirebbe pagine e pagine. Ma anche osservando solo i lessemi registrati le due categorie "raro" e "comune" appaiono chiaramente inadeguate. Per citare tra le recenti una delle ultime concordanze frutto di pazienti spogli manuali durati lunghi anni⁸, nel *Decamerone guanto, abito e vestimento* sono tutte parole "comuni", ma *guanto* occorre 5 volte, *vestimento* 26, *abito* 31,

⁶ T. De Mauro, *Statistica linguistica*, in *Enciclopedia italiana*, App. III, Istituto dell'Enciclopedia Italiana, Roma 1961, vol. II, pp. 820-1, s.v.; rinvio anche a *Quantità-qualità: un binomio indispensabile per comprendere il linguaggio*, in S. Bolasco, R. Cipriani (a cura di), *Ricerca qualitativa e computer. Teorie, metodi e applicazioni*, Franco Angeli, Milano, pp.21-30.

⁷ Franz Josef Hausmann, *Die gesellschaftlichen Aufgaben der Lexikographie in Geschichte und Gegenwart*, in Hausmann, Franz Josef; Oskar Reichmann; Ernst Wiegand y Ladislav Zgusta (eds.), *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie*, 3 voll., Walter de Gruyter, Berlino-New York 1989, vol. I, pp. 1-19.

⁸ Alfredo Barbina, *Concordanze del "Decamerone"*, sotto la direzione di Umberto Bosco, 2 voll., Accademia della Crusca-Giunti, Firenze 1969.

e tutte e tre sono poca cosa di fronte alle circa 1.400 occorrenze di *donna*, *donare* è comune come *vedere* e *dire*, ma occorre un centinaio di volte, mentre sono oltre 1.200 le occorrenze di *vedere* o circa 3.000 quelle di *dire*. Una concordanza recente della *Scienza nuova* di Vico nel testo del 1744, interamente informatizzata e attenta a fornire tutte le necessarie indicazioni di frequenza dei lessemi⁹, ci permette osservazioni più precise: le 154.593 occorrenze del testo si riconducono a 13.110 forme (flesse se variabili), ricondotte a loro volta a 6.142 lessemi (lemmi nella terminologia del lavoro), che occorrono quindi con una frequenza media di circa 25. Dei lessemi 660 hanno frequenza superiore alla media, 5.482 inferiore: già questo dato indica una sproporzione. Questa si fa più evidente se osserviamo che:

- il 50% delle occorrenze è coperto dai primi 44 lessemi;
- i primi 10 lessemi coprono il 24,7% delle occorrenze;
- i primi 50 lessemi coprono il 51,5% delle occorrenze;
- i primi 500 coprono l'81,6%;
- i primi 1.000 coprono l'88,7%;
- i primi 2.000 coprono il 94,5%;
- i residui 4.142 lessemi coprono il residuo 5,5% del testo.

Cifre del genere rendono evidente che nella distribuzione delle frequenze in un qualunque testo o nel campionamento di testi d'una lingua non ci troviamo dinanzi all'opposizione binaria di un generico "raro", "poco comune", "basso" e di un generico "comune", "frequente", "alto". Nell'ultimo caso citato si va da centinaia di lessemi che occorrono una sola volta, con una frequenza di $1/6.142 = 0,000163\%$, o dall'insieme dei 4.142 lessemi dopo i primi 2.000, con una frequenza media di 0,00133, fino ai primi dieci lessemi che hanno frequenze comprese tra 1,72% e 6,93%, cioè superiori tra diecimila e quarantamila volte alla frequenza media dei precedenti. Non c'è solo un basso e un alto. Il paesaggio è assai più movimentato: bassure, pianure, ondulazioni, altopiani, colline, montagne e picchi altissimi, svettanti assai

⁹ Giambattista Vico, *Principj di Scienza Nuova d'intorno alla comune natura delle nazioni. Concordanze e indici di frequenza dell'edizione Napoli 1744*, a cura di Marcello Veneziani, Olschki, Firenze 1997.

più di quanto le più alte vette della Terra si levano sul livello dei mari e delle terre.

Il moltiplicarsi degli indici e delle concordanze ha reso sempre più netta la percezione di questo paesaggio statistico¹⁰. Essa è si è rafforzata con il moltiplicarsi di liste lessicali di frequenza estratte da testi delle varie lingue e nate per le esigenze più varie: individuazione di stenogrammi per i nessi grafici e i vocaboli più frequenti; individuazione del vocabolario "di base", del *Grundwortschatz*, da offrire agli apprendenti della lingua, sia della prima sia della L2; decrittazione di testi cifrati di una lingua; pura ricerca statistica, per il carattere stocastico che gli statistici sempre di più scorgevano nel configurarsi dei testi rispetto alle unità fonologiche, grafiche, morfolessicali potenzialmente esistenti in una lingua¹¹.

Dopo alcuni lavori pionieristici¹², negli anni quaranta del Novecento ha inizio lo sfruttamento teorico linguistico e psicolinguistico della massa di dati già disponibili nelle varie liste di frequenza che ordinano i fenomeni per rango decrescente, da quelli più occorrenti, di *rango* 1, a quelli di rango minimo, cioè di frequenza 1, reperiti in singoli testi o gruppi di testi, più o meno rappresentativi di uno stato di lingua¹³. Anche mettendo a frutto l'ormai affermata teoria matematica della comunicazione¹⁴ emergono una serie di costanti nella distribuzione dei fenomeni e se ne ricavano indicazioni importanti per qualificare sia le unità e strutture delle lingue, sia i caratteri idiosincratici delle lingue e del loro divenire, sia infine dei testi e delle modalità diafasiche di utilizzazione delle lingue. Un primo rendiconto ancora prezioso fu dato

¹⁰ Suzanne Hanon, *La concordance* e Paul Sappler, *Der Index/Das Belegstellenwörterbücher*, in *Wörterbücher. Dictionaries*, cit., vol.II, pp. 1562-67, 1567-73.

¹¹ *Statistica linguistica*, cit., Willy Martin, *The Frequency Dictionaries*, Peter Kühn, *Das Grundwortschatzwörterbücher*, in *Wörterbücher. Dictionaries*, cit., vol.II, pp. 1314-22, 1353-64.

¹² Ricordati e discussi, per es., in un'altra opera pionieristica di George Kingsley Zipf, *The Psychobiology of Language*, Houghton Mifflin, Boston 1935, e poi nella sua trattazione più canonica, *Human Behavior and the Principle of Least Effort*, Hafner, New York-Londra 1949, in particolare pp. 546-47.

¹³ Pierre Guiraud, Joshua Whatmough, *Bibliographie de la statistique linguistique*, Mouton, Utrecht-Anversa 1954.

¹⁴ Claude E. Shannon, Warren Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana 1949.

da George A. Miller¹⁵. Delle particolari interpretazioni e discussioni che si sono sviluppate, nonostante il pesante ostracismo antistatistico di alcune posizioni teoriche, daranno conto i singoli contributi di questo volume.

È stato ripetutamente osservato che, per usare le parole di Giulio Lepschy, «gli eventi linguistici si rivelano particolarmente adatti al trattamento statistico»¹⁶. Proprio perché una lingua non è un'aritmetica o un'algebra, un calcolo in senso formale, ma un aggregato di forme e regole fluttuanti nel *temps* e nella *masse parlante* come voleva giustamente Saussure, i fatti di esecuzione (perfino quelli che possono apparire aberranti) incidono sulle unità del sistema e sulle regole, le modificano, le sopprimono, ne introducono nuove, rivelano un diverso grado di possesso degli elementi che l'intero repertorio linguistico può mettere a disposizione di una comunità a un momento dato. Gli accertamenti statistici così gettano luce da un lato sulla lingua, sulla consistenza delle sue unità e strutture, dall'altra su chi la lingua utilizza, sui locutori e sulle diverse modalità con cui essi di volta in volta, da un tipo all'altro di situazione, mettono a frutto le risorse loro offerte dalla lingua.

¹⁵ *Language and Communication*, McGraw-Hill Book Comp., New York 1951.

¹⁶ Giulio C. Lepschy, *La linguistica del Novecento*, Il Mulino, Bologna 2000³, p.101.