



Consonanze 29

METODI E PROSPETTIVE DELLA RICERCA LINGUISTICA

a cura di
Chiara Meluzzi & Nicholas Nese



Metodi e prospettive della ricerca linguistica

a cura di
Chiara Meluzzi & Nicholas Nese

LEDIZIONI

CONSONANZE

Collana del
Dipartimento di Studi Letterari, Filologici e Linguistici
dell'Università degli Studi di Milano

diretta da Alberto Cadioli

29

Comitato scientifico

Benjamin Acosta-Hughes (The Ohio State University), Giampiera Arrigoni (Università degli Studi di Milano), Johannes Bartuschat (Universität Zürich), Maria Patrizia Bologna (Università degli Studi di Milano), Alfonso D'Agostino (Università degli Studi di Milano), Maria Luisa Doglio (Università degli Studi di Torino), Bruno Falchetto (Università degli Studi di Milano), Alessandro Fo (Università degli Studi di Siena), Luigi Lehnus (Università degli Studi di Milano), Maria Luisa Meneghetti (Università degli Studi di Milano), Michael Metzeltin (Universität Wien), Silvia Morgana (Università degli Studi di Milano), Laurent Pernot (Université de Strasbourg), Simonetta Segenni (Università degli Studi di Milano), Luca Serianni (Sapienza Università di Roma), Francesco Spera (Università degli Studi di Milano), Renzo Tosi (Università degli Studi di Bologna)

Comitato di Redazione

Stefania Baragetti, Guglielmo Barucci, Virna Brigatti, Edoardo Buroni, Silvia Gazzoli, Maddalena Giovannelli, Cecilia Nobili, Marco Pelucchi, Stefano Resconi, Luca Sacchi, Francesco Sironi

Tutti i volumi della collana sono sottoposti a peer review

ISBN 978-88-5526-727-4

Metodi e prospettive della ricerca linguistica, a cura di Chiara Meluzzi, Nicholas Nese

© 2022

Ledizioni – LEDIpublishing

Via Boselli, 10 20136

Milano, Italia

www.ledizioni.it

È vietata la riproduzione, anche parziale, con qualsiasi mezzo effettuata, compresa la fotocopia, anche a uso interno o didattico, senza la regolare autorizzazione.

Volume stampato con il contributo dell'Associazione Giasoniani del Maino di Pavia e del Dipartimento di Studi Letterari, Filologici e Linguistici dell'Università degli Studi di Milano.

Indice

Introduzione	5
CHIARA MELUZZI & NICHOLAS NESE	
Prefazione	7
ILARIA FIORENTINI	
Studiare la lingua come fatto sociale	13
ROSALBA NODARI	
La raccolta dati sul campo (e in campo)	37
CHIARA MELUZZI	
La raccolta dati a distanza: metodologia per uno studio fonetico	53
NICHOLAS NESE	
<i>Corpus linguistics</i> nei nuovi media	65
GLORIA COMANDINI	
La linguistica delle lingue di attestazione frammentaria	83
LUGA RIGOBIANCO	
Linguistica storica e risorse linguistiche digitali	95
CHIARA ZANCHI	
Fare linguistica del testo	117
FILIPPO PECORARI	
I disturbi del linguaggio e la ricerca in linguistica clinica	135
GLORIA GAGLIARDI	

Linguistica computazionale. Fare i conti con quello che si dice GIULIA CAPPELLI	149
Come si imparano le lingue: la linguistica acquisizionale ELISA CORINO	169
Biografia degli autori (in ordine alfabetico)	185

Introduzione

Chiara Meluzzi* & Nicholas Nese**

*Università degli Studi di Milano; **Università degli Studi di Pavia

Per molti studenti di facoltà umanistiche l'ingresso all'università coincide anche con il "primo contatto" con una disciplina nuova e, fino ad allora, sconosciuta: la linguistica. Il rapporto iniziale con la materia può essere difficile, a tratti (a volte per sempre) anche burrascoso, ma spesso matura nello studente la consapevolezza che, attraverso la linguistica, si possa arrivare a comprendere meglio il mondo di parole all'interno del quale viviamo e nel quale siamo quotidianamente immersi. La linguistica diventa quindi uno strumento, ma anche materia di studio con molteplici sfaccettature, al cui interno lo studente scoprirà prospettive teoriche e metodologiche molto differenti, nonché oggetti di studio e di interesse tra i più vari tra le scienze umane.

Racchiudere in un unico volume tutto ciò che l'aggettivo "linguistica" implica oggi nel campo delle scienze del linguaggio è impresa troppo ambiziosa e fuori dagli obiettivi del presente lavoro. Già altre raccolte esistono al riguardo, tra cui il volume curato da Francesca Masini e Nicola Grandi *Tutto ciò che hai sempre voluto sapere sul linguaggio e sulle lingue*, che ha in parte ispirato la preparazione di questo volume. Questa miscellanea nasce con l'esplicito intento di presentare agli studenti dei corsi triennali (ed eventualmente magistrali) la diversità degli approcci metodologici della linguistica, intesa come disciplina che studia il linguaggio come qualcosa in cui *tout se tient*. Proprio in considerazione del target primario del nostro volume, al termine di ogni capitolo è stato chiesto agli autori di inserire alcune letture di riferimento rispetto agli argomenti trattati, nella speranza che questo alimenti l'interesse verso gli studi linguistici.

Lo scopo primario di questo lavoro è quello di mostrare, tramite una lunga sequenza di esempi e lavori precedenti, come avviene la ricerca in ambito linguistico e, in particolare, in alcuni dei filoni di studio, classici e moderni, che rientrano in questa disciplina. L'interesse principale del volume, così come dei vari capitoli che lo compongono, è rivolto dunque ai dati e alle metodologie di ricerca che caratterizzano diversi campi dell'indagine linguistica. È importante però specificare che questo approccio strettamente empiristico non rappresenta assolutamente l'unico modo (e, men che meno, il miglior modo) di "fare lingui-

stica”, ma riflette solamente gli interessi di ricerca dei curatori e degli autori dei singoli capitoli.

Nel rispetto dell’interesse primario sulle metodologie di ricerca, il volume è organizzato in due sezioni tematiche, la prima dedicata ai metodi di raccolta e trattamento dei dati in linguistica, con una prospettiva preferenzialmente, come detto, di stampo empirico. Rientrano in questa sezione i contributi di Rosalba Nodari, Chiara Meluzzi, Nicholas Nese e Gloria Comandini. La seconda sezione del volume presenta invece alcuni approcci all’analisi delle lingue e del linguaggio, evidenziando come ciascuna branca dell’indagine linguistica sia guidata da proprie domande di ricerca e come tali domande determinino le metodologie di raccolta, organizzazione e analisi dei dati. In questa sezione, i contributi spaziano dalla linguistica storica (lavori di Luca Rigobianco e Chiara Zanchi), alla linguistica del testo e l’acquisizione delle lingue (contributi di Filippo Pecorari ed Elisa Corino), fino alle più recenti applicazioni dell’indagine linguistica nell’ambito clinico e computazionale (lavori di Gloria Gagliardi e di Giulia Cappelli). Completa il volume la prefazione redatta da Ilaria Fiorentini, che ben rispecchia lo spirito e le intenzioni di questa miscellanea.

Come curatori, siamo doppiamente grati agli autori che hanno voluto prendere parte a questo volume con un proprio prezioso contributo, dal momento che gli stessi autori sono anche stati relatori delle serate di divulgazione scientifica “Linguistica e cinema”, tenutesi al Collegio “Giasone del Maino” di Pavia dal 2018 al 2020; l’iniziativa è stata realizzata con i fondi ACERSAT dell’Università di Pavia e con il contributo dell’EDiSU di Pavia. Questi momenti di incontro hanno permesso di sviluppare le riflessioni scientifiche che vedete poi espresse in questo volume.

Nel congedare quest’opera, i curatori vorrebbero ringraziare, oltre agli autori dei singoli contributi che hanno preso parte e creduto fortemente in questa impresa, anche l’Associazione Giasoniani del Maino, che ha permesso l’organizzazione di “Linguistica e Cinema” e che ha contribuito economicamente alla stampa del volume. Si ringrazia il rettore del Collegio, prof. Stefano Colloca, e tutto lo staff del “Giasone del Maino” per avere ospitato negli anni l’iniziativa delle serate linguistiche. Un doveroso ringraziamento va inoltre a Erica Maria Rinaldi, per il supporto ricevuto nella fase finale di questo progetto, nonché a Francesca Cotugno e Francesca Nicora per la puntuale e dettagliata revisione del volume nella sua prima stesura. Un ringraziamento sentito va alla casa editrice “Ledizioni”, per la disponibilità con cui ha accolto l’idea originaria e ha aiutato i curatori a portarla a termine, nonché al prof. Alberto Cadioli per aver incluso il volume nella collana “Consonanze”.

Prefazione

Ilaria Fiorentini
Università degli Studi di Pavia

Per sua stessa natura, la linguistica è composita e sfaccettata: negli ultimi anni, hanno cominciato a emergere delle “guide” per permettere di orientarsi sia agli specialisti, sia a chiunque nutra interesse per il tema. Limitandoci al contesto italiano, possiamo testimoniare un discreto fervore scientifico ed editoriale che ha interessato la disciplina da diversi punti di vista. Da un lato, si è cominciato a riflettere più diffusamente sulla necessità di un’opera di divulgazione della linguistica; ciò ha portato per esempio all’organizzazione di convegni sul tema (tra cui il IV Convegno interannuale della Società di Linguistica Italiana, dal titolo “La linguistica della divulgazione, la divulgazione della linguistica”, tenutosi a Bologna il 14-15 giugno 2018),¹ alla redazione di volumi (per esempio, *Tutto ciò che hai sempre voluto sapere sul linguaggio e sulle lingue*, a cura di Nicola Grandi e Francesca Masini, vincitore tra l’altro del Premio Nazionale di Divulgazione Scientifica 2017), e alla creazione di blog e risorse online dedicate (come *Linguisticamente*,² sito di divulgazione scientifica nato nel 2020 e dedicato agli ambiti di interesse e alle principali direzioni della ricerca linguistica contemporanea).³

Dall’altro lato, si è reso sempre più evidente il bisogno di una riflessione metodologica trasversale,⁴ specialmente a fronte delle recenti applicazioni linguistiche anche in settori tangenti (di cui si discuterà poco più avanti). In questo filone in particolare si inserisce il presente volume, che tra i suoi diversi meriti ha quello di mettere insieme due aspetti fondamentali: in prima battuta il dato, e le

1. Gli atti del convegno sono stati pubblicati in Grandi, Masini (2020).

2. <https://www.linguisticamente.org/>

3. Non va dimenticato a questo proposito il magazine online *Lingua italiana* di Treccani (https://www.treccani.it/magazine/lingua_italiana), dedicato ad articoli di approfondimento (sempre con taglio divulgativo) su temi linguistici.

4. Di taglio diverso, ma altrettanto importante da menzionare, è la riflessione sulla didattica delle discipline linguistiche; si ricorda a questo proposito il LIII Congresso SLI (svoltosi presso l’Università dell’Insubria, a Como, nel settembre 2019), che aveva come titolo (e tema) “Insegnare Linguistica: basi epistemologiche, metodi, applicazioni”. Gli atti del congresso sono stati pubblicati in Sansò (2020).

questioni di metodo a esso legate; in seconda, le prospettive della ricerca linguistica, anche in senso applicativo.

Innanzitutto, dunque, viene ribadita la centralità del dato in linguistica, in tutte le sue realizzazioni. D'altronde, come ricordava Löfstedt (1942: IX, citato in Lehmann 2004: 176), «Die Theorien vergehen, aber das Material bleibt bestehen»: le teorie passano, ma il materiale resta. (Ri)partiamo dunque dai dati: quali proprietà devono possedere per potersi configurare come dati linguistici? Qual è il loro ruolo nella ricerca e nel processo scientifico? Come si ottengono? Come si analizzano?

Ogni disciplina scientifica ha i suoi dati: si pensi ad esempio alla demografia, per cui il dato può essere la distribuzione in fasce d'età della popolazione; o alla semiotica, la scienza del segno, per cui un dato può essere qualsiasi forma di testualità. In linguistica, in generale, può considerarsi dato «qualunque fenomeno o fatto che abbia caratteri che lo rendono pertinente per una certa teoria e una certa analisi: il riconoscimento di un dato presuppone il filtro di una teoria, che escluda i fatti che non hanno rilevanza per il problema da studiare» (Berruto, Cerruti 2015: 211). La definizione ricorda e sottolinea il peso che ha la teoria che soggiace all'analisi, e il filtro che essa sempre mette in atto: oltre a filtrare le infinite informazioni che ci arrivano continuamente dal mondo esterno, la teoria decide quali fra queste «riceveranno lo statuto stesso di *dato* e diventeranno pertanto elementi delle nostre successive costruzioni teoriche. Informazioni sensoriali, fenomeni che sono considerati *dati* in certe scienze possono non esserlo in altre, o - più pericolosamente - persino in altre scuole di quella stessa scienza» (Iannàccaro 2000: 23).

Il dato presuppone tre entità fondamentali (tante quante sono le valenze del verbo *dare*): chi dà (il produttore, la fonte del dato); chi riceve (in questo caso, è chi scopre o utilizza il dato); infine, l'oggetto che viene trasferito (l'entità che costituisce effettivamente il dato). Se a prima vista le prime due entità potrebbero sembrare meno importanti, almeno per una concezione scientifica del dato, nella ricerca linguistica (così come in altre discipline) così non è: al contrario, è questione assolutamente cruciale quella relativa a «who produces the data and who receives them» (Lehmann 2004: 177). A questo proposito, va sottolineato come un dato sia tale solo per una determinata persona (o per una comunità scientifica) e in una determinata prospettiva. Si consideri la seguente situazione, ipotizzata dallo stesso Lehmann (2004: 180-1):

Linguist A has a tape which records a story in Yucatec Maya. The recording is A's data. He produces an orthographic representation of the story and publishes it as the result of his research. Linguist B uses A's orthographic representation as data for his grammar of Yucatec Maya, which he publishes as the output of his research. Linguist C is a typologist whose sources of information are grammars. He uses B's descriptive statements on Yucatec Maya clause

structure as data, puts them into a database and arrives at a couple of cross-linguistic generalizations which he publishes as a typology of clause structure.

L'esempio mostra che l'analisi di una persona (una rappresentazione ortografica, una grammatica e così via) può costituire il dato di un'altra; allo stesso tempo, dallo stesso dato possono scaturire più analisi, diverse tra loro. Di conseguenza, un dato non è tale in sé e per sé, quanto piuttosto in relazione a una particolare ricerca empirica: a diverse prospettive corrispondono diversi tipi di dati, e teorie diverse permetteranno di ottenere da una stessa base reale dati differenti (anche all'interno della stessa disciplina). È la differente prospettiva che distingue, ad esempio, un fonetista da un fonologo: la [r], la [R], la [ʁ] rappresentano «per il fonologo che studi il sistema dell'italiano la stessa lettera (...) e per il fonetista quattro 'cose' assolutamente diverse» (Iannàccaro 2000: 28). Questo aspetto emerge particolarmente bene dai contributi che compongono il presente volume: vediamo per esempio come per il linguista clinico l'oggetto di studio siano i disturbi del linguaggio; in questo caso, saranno fondamentali anche ulteriori distinzioni, per esempio relative all'insorgenza del disturbo (evolutivo o acquisito), o il livello a cui questo si manifesta (Gagliardi, questo volume). L'analisi si baserà su testi orali o scritti dei pazienti, che verranno analizzati secondo le finalità specifiche della ricerca (a livello fonetico, fonologico, morfologico, sintattico, e così via), per determinare la competenza del parlante e il disturbo di cui soffre. Questi oggetti, queste distinzioni, questi testi non avranno lo stesso interesse (in altre parole, non costituiranno lo stesso dato) per, poniamo, il sociolinguista, che si occupa dei rapporti tra lingua e fatti sociali, e che cercherà oggetti diversi, osservando "cose" diverse: per esempio, potrà costituire un dato la presenza di determinate lingue nello spazio pubblico (su cartelloni, segnaletica, insegne...), indicando un maggiore peso, sociale e/o politico, della lingua in questione nello spazio in cui viene osservata (Nodari, questo volume). Ancora, la linguistica acquisizionale (Corino, questo volume) osserva le produzioni di persone che stanno apprendendo una lingua: in questo caso, frasi come *Io arabo*, *Io cinese* o *Io rumeno*, prodotte da parlanti diversi, potranno essere interpretate non tanto (o non solo) come "errori", quanto come testimonianze di un processo di semplificazione (ovvero l'uso di una forma semplificata, come l'omissione della copula), tipico delle prime fasi di apprendimento linguistico.

Riconosciuta l'importanza (e la natura) del dato, restano da illustrare i diversi paradigmi di acquisizione dei dati stessi. Da questo punto di vista, la linguistica «è stata meno ingenua di altre discipline, dacché ci si è resi presto conto che il modo stesso di acquisizione del dato è ovviamente in grado di cambiare l'essenza e l'aspetto del dato stesso, prima ancora che la sua interpretazione» (Iannàccaro 2000: 27). La letteratura scientifica in materia ha sempre dedicato grande attenzione e ha strutturato con accuratezza la questione delle tecniche di elicitazione (questio-

nari, conversazione libera, scelta dei punti d'inchiesta o dei testimoni, tipo di registratore più adatto), proprio perché tradizionalmente esse sono «esplicitamente inserite nell'ambito della riflessione teorico-metodologica» (*ibidem*).

Guardando ai normali processi di acquisizione dei dati in linguistica, emergono due modalità fondamentali, distinte tra loro (Iannàccaro 2000: 25):

- il *raccogliere* dati (una volta che siano stati riconosciuti come tali) per interpretarli e sistematizzarli, e con essi a) costruire una nuova teoria o tentare di falsificarla, oppure b) semplicemente fornire esempi reali per teorie già esistenti;
- l'*inventare* dati, ossia produrre in prima persona dati virtuali, testandoli poi su sé stessi o su altri, al fine di corroborare o falsificare la propria teoria.

Ogni branca della linguistica avrà dunque i suoi dati e i suoi specifici metodi per raccogliarli; per esempio, in sociolinguistica (disciplina fortemente empirica, che si occupa di indagare i rapporti tra lingua e società; si veda anche Nodari, questo volume) il problema della raccolta dati è centrale: il dato non potrà in nessun caso essere fornito dall'introspezione del linguista stesso (come invece è possibile per altre sottodiscipline), ma dovrà essere il più possibile autentico, ovvero raccolto in una situazione effettiva di produzione linguistica da parte dei parlanti.

Alla raccolta dati, l'unica modalità di acquisizione possibile per le sottodiscipline empiriche della linguistica, il volume dedica dunque la sua prima parte, descrivendo, in maniera metodologicamente fondata, le diverse modalità di elicitazione dei dati. Oltre ai metodi più classici di raccolta (tra cui, per esempio, quella sul campo; si veda Meluzzi, questo volume), ne vengono introdotti e descritti di nuovi, ovvero quelli in un certo senso imposti dalle restrizioni dovute alla pandemia, che non hanno fermato la ricerca linguistica; si veda a questo proposito il capitolo di Nicholas Nese, che descrive, tra le altre cose, i software che permettono di portare avanti anche a distanza esperimenti sociofonetici percettivi. Diverso è il caso della linguistica testuale, di cui si parla nel capitolo a cura di Filippo Pecorari, per la quale i testi reali possono essere affiancati, se necessario, da testi manipolati; oppure, si può fare ricorso a brevi sequenze create *ad hoc* al fine di mettere in luce il fenomeno oggetto di analisi (analisi che si baserà sull'intuizione del ricercatore, a sua volta poggiata su una competenza testuale costruita nel tempo).

Se numerosi sono i modi di guardare ai dati (e, di conseguenza, di acquisirli), altrettanto vari saranno gli approcci e le metodologie attraverso le quali i dati saranno analizzati. Oltre alla riflessione sul dato, dunque, è altrettanto importante e necessaria una riflessione sulle prospettive della linguistica, la cui portata va ampliandosi con l'ampliamento degli ambiti di impiego della disciplina. La seconda parte del volume va in questa direzione: qui, l'importanza del dato si intreccia con gli aspetti più "pratici", più applicativi. A questo proposito, vengono introdotti e descritti approfonditamente sia alcuni nuovi settori della linguistica (come la linguistica clinica nel capitolo di Gloria Gagliardi), sia i più recenti sviluppi e

applicazioni delle discipline tradizionali (come i rapporti tra linguistica storica e *digital humanities* nel capitolo di Chiara Zanchi).

«A me pare che *computer sciences* e linguistica oggi si muovano sempre più in direzioni convergenti», scriveva Tullio de Mauro nella sua prefazione all'*Introduzione alla linguistica computazionale* di Isabella Chiari (2007: VI). E se la nascita ufficiale della disciplina risale ormai a cinquant'anni fa (1962, con la fondazione dell'*Association for Computational Linguistics* - ACL), gli aspetti applicativi a essa legati sono oggi sempre più diffusi in moltissimi ambiti (non solo della linguistica). Oltre a quelli citati nel capitolo di Giulia Cappelli (tra gli altri, la sintesi vocale, o la traduzione automatica), ne troviamo alcuni legati ai *corpora*: la linguistica dei *corpora* (Comandini, questo volume) nasce proprio come una branca della linguistica computazionale, basandosi su raccolte di testi dai quali vengono estratti dati grazie a strumenti automatici e semiautomatici; e, come ben sottolinea Zanchi (questo volume), oggi «chi studia lingue antiche è “condannato” a usare i *corpora*», con l'aiuto che arriva da strumenti computazionali sempre più raffinati. Se nel caso di lingue frammentarie (approfondite da Luca Rigobianco) è ancora necessario un minuzioso lavoro sostanzialmente manuale da parte del linguista (ma i recenti sviluppi permettono anche qui di avere a disposizione una quantità sempre maggiore di strumenti), in quello della già citata linguistica testuale la situazione cambia: la disciplina, che originariamente preferiva basarsi su testi fittizi, di piccole dimensioni, analizzati sulla base della competenza testuale del linguista, può ora contare «su corpora ampi (dell'ordine di milioni o miliardi di parole) e bilanciati, rappresentativi di numerosi tipi testuali e varietà linguistiche, che facilitano notevolmente i compiti di descrizione e analisi testuale» (Pecorari, questo volume).

In conclusione, nuovi metodi, nuove fonti, nuove tecniche hanno negli ultimi anni trovato spazio nella ricerca linguistica, permettendo la nascita di nuovi approcci e l'ampliamento di prospettive di quelli già esistenti: tuttavia, rimane costante e viene ribadita la centralità del dato, nelle diverse prospettive a cui abbiamo accennato, anche nell'ottica di una riflessione sul metodo di acquisizione e di analisi. Oggi, la linguistica, con tutto il suo apparato di conoscenze teoriche e applicative, si è fatta strada in molti settori, non solo strettamente scientifici, ed è presente in innumerevoli forme nella vita quotidiana di molti; e se questa strada è ancora lunga, non mancano più gli strumenti – tra cui questo volume - che aiutano a percorrerla.

Bibliografia

- G. Berruto, M. Cerruti, *Manuale di sociolinguistica*, Torino, UTET, 2015.
- I. Chiari, *Introduzione alla linguistica computazionale*, Roma-Bari, Laterza, 2007.
- N. Grandi, F. Masini (a cura di), *Tutto ciò che hai sempre voluto sapere sul linguaggio e sulle lingue*, Bologna, Caissa, 2017.
- N. Grandi, F. Masini (a cura di), *La linguistica della divulgazione, la divulgazione della linguistica. Atti del IV Convegno interannuale della Società di Linguistica Italiana (Bologna, 14-15 giugno 2018)*, Milano, Officinaventuno, 2020.
- G. Iannàccaro, «Per una semantica più puntuale del concetto di ‘dato linguistico’: un tentativo di sistematizzazione epistemologica», *Quaderni di Semantica*, 41 (1), 2000, pp. 21–49.
- C. Lehmann, «Data in linguistics», *The Linguistic Review*, 21 (3-4), 2004, pp. 175-210.
- E. Löfstedt, *Syntactica. Studien und Beiträge zur historischen Syntax des Lateins. Erster Teil: Über einige Grundfragen der lateinischen Nominalsyntax*. (Acta Reg. Societatis Humaniorum Litterarum Lundensis, X:1) (2., erw. Auflage), Lund, Gleerup, 1942.
- A. Sansò, *Insegnare Linguistica: basi epistemologiche, metodi, applicazioni. Atti del LIII Congresso della Società di Linguistica Italiana (Università dell’Insubria, 19-21 settembre 2019)*, Milano, Officinaventuno, 2020.