

A cura di Francesca Socrate, Carlotta Sorba

Tra linguistica e storia: incroci metodologici e percorsi di ricerca. Interventi di Jean-Philippe Genet, Michele A. Cortelazzo, Jacques Guilhaumou, Pietro Trifone, Marta Margotti, Rachele Raus, Francesca Socrate

(doi: 10.1409/73533)

Contemporanea (ISSN 1127-3070)

Fascicolo 2, aprile 2013

Ente di afferenza:

Universit Trieste (units)

Copyright © by Società editrice il Mulino, Bologna. Tutti i diritti sono riservati.

Per altre informazioni si veda <https://www.rivisteweb.it>

Licenza d'uso

L'articolo è messo a disposizione dell'utente in licenza per uso esclusivamente privato e personale, senza scopo di lucro e senza fini direttamente o indirettamente commerciali. Salvo quanto espressamente previsto dalla licenza d'uso Rivisteweb, è fatto divieto di riprodurre, trasmettere, distribuire o altrimenti utilizzare l'articolo, per qualsiasi scopo o fine. Tutti i diritti sono riservati.

Tra linguistica e storia: incroci metodologici e percorsi di ricerca

a cura di Francesca Socrate e Carlotta Sorba. Intervengono Jean-Philippe Genet, Michele A. Cortelazzo, Jacques Guilhaumou, Pietro Trifone, Marta Margotti, Rachele Raus, Francesca Socrate

Francesca Socrate e Carlotta Sorba

Due discipline a confronto

Perché porre la questione del rapporto tra storia e linguistica e dedicarle uno spazio a più voci? In Italia, diversamente che in Francia, tale incrocio ha avuto poche occasioni di sviluppo, nonostante le due discipline presentino punti di contatto del tutto evidenti: non solo la lingua è una delle più importanti istituzioni di una società, ma le fonti della storia sono prevalentemente, anche se non esclusivamente, costituite di parole. Le opportunità d'incontro e di colloquio sono state tuttavia poche e anche gli incroci più facilmente praticabili (a partire da quello preliminare con la filologia per continuare con la storia della lingua e con la sociolinguistica) hanno incontrato un interesse piuttosto scarso all'interno della corporazione degli storici. Oggi però le cose appaiono in via di mutamento e tale constatazione ci ha sollecitato a organizzare questo forum. Le enormi novità introdotte dai molti software di analisi testuale che permettono una verifica statistica del conte-

nuto linguistico di un *corpus* di testi aprono importanti opportunità di utilizzo anche per l'analisi storica dei linguaggi (politici, e non solo) e impongono un ripensamento d'insieme dei possibili incontri tra le due discipline. Perché accontentarsi di un approccio esclusivamente sociologico o politico ai fenomeni della lingua? Perché non avvicinarsi alle modalità di analisi del discorso che attingono anche alla linguistica? E soprattutto perché non utilizzare strumenti oggi in grado di sottoporre enormi insiemi di testi a un'analisi ravvicinata (ricerca di frequenze, calcolo del vocabolario specifico, distanze intertestuali, ecc.)? Che cosa questi strumenti possono dirci (e come) sulle lingue, le culture, le società del passato? E quali rapporti possono esserci tra analisi quantitative e qualitative degli atti linguistici?

C'è un presupposto teorico importante che accomuna le due discipline e va qui segnalato in apertura: quello di voler guardare, e cercare di vedere, al di là del testo, andando

alla ricerca non solo del suo significato esplicito, ma di quello più profondo. Entrambe si sforzano cioè, dal proprio punto di vista specifico, di portare alla luce l'implicito, e finanche l'inconsapevole. Ogni testo, come unità comunicativa dotata di senso, è visto in questa prospettiva comune come il precipitato delle pratiche sociali che lo producono; e oggetto dell'interesse di entrambe le discipline sono in fin dei conti quelle pratiche sociali. La linguistica può essere dunque uno strumento potenzialmente utile per aiutarci a interpretare il linguaggio nel suo rapporto con la realtà sociale anche in una prospettiva storica.

L'interesse per un dialogo tra le due discipline si è aperto originariamente in Francia, e non è un caso che il confronto che proponiamo si giochi essenzialmente tra studiosi italiani e francesi. Come ricostruiscono qui Jean-Philippe Genet e Jacques Guilhamou, è con gli studi pioneristici di Régine Robin e del gruppo che si coagula negli anni Settanta attorno all'École Normale Supérieure di Saint-Cloud che si avvia in Francia un percorso articolato e complesso. Da allora a oggi questo gruppo ha prodotto un notevole insieme di ricerche, di cui i due studiosi, nei loro interventi, propongono una panoramica che spazia tra epoche e linguaggi diversi. Piuttosto diversa appare la situazione nella

storiografia anglosassone, dove l'analisi del discorso «alla francese» e la lessicografia quantitativa hanno trovato sviluppi minori e comunque più tardivi. Come noto, la sociolinguistica nasce però negli Stati Uniti alla fine degli anni Sessanta e concentrandosi sul rapporto tra lingua e articolazione sociale (e sulla multidimensionalità della lingua) ha dimostrato subito connessioni evidenti con le domande degli storici sociali¹.

Che cosa è arrivato in Italia di tale dibattito? Nel 1981 Maurizio Gribaudo dedicava un bel saggio su «Quaderni storici» alle nuove prospettive di collaborazione interdisciplinare con la sociolinguistica, ma quell'intervento rimaneva sostanzialmente isolato². Più frequenti sono stati piuttosto i contatti tra storia e storia della lingua, come dimostrano ad esempio gli studi di Pietro Trifone³ o il convegno che l'Associazione per la storia della lingua italiana ha dedicato nel 1999 al rapporto con le discipline storiche⁴. Esperimenti di ricerca «sul campo», tutti incentrati sul linguaggio politico e istituzionale, non sono in realtà mancati anche in Italia, soprattutto in anni recenti, e tuttavia non sono stati condotti da storici. Si pensi al notissimo e ancora molto convincente studio di Erasmo Leso sulla nascita del linguaggio politico moderno nel triennio giacobino⁵; alle ricerche di

¹ W. Labov, *The Social Stratification of English in New York City*, Washington, Center for Applied Linguistics, 1966; D. Hymes, *Foundations in Sociolinguistics: An Ethnographic Approach*, London, Tavistock, 1974; J.K. Chambers, *Sociolinguistic Theory: Linguistic Variation and Its Social Significance*, Malden, Blackwell, 2010.

² M. Gribaudo, *A proposito di linguistica e storia*, «Quaderni Storici», 46, 1981.

³ Cfr. P. Trifone (a cura di), *Lingua e identità. Una storia sociale dell'italiano*, Roma, Carocci, 2009; Id., *Storia linguistica dell'Italia disunita*, Bologna, Il Mulino, 2010.

⁴ G. Alfieri (a cura di), *Storia delle lingue e storia: atti del 2° convegno ASLI (Catania, 26-28 ottobre 1999)*, Firenze, Cesati, 2003.

⁵ E. Leso, *Lingua e rivoluzione. Ricerche sul vocabolario politico italiano del triennio rivoluzionario 1796-1799*, Venezia, Istituto veneto di scienze, lettere e arti, 1991.

Michele Cortelazzo e Arjuna Tuzzi sulla lingua dei presidenti della repubblica⁶; alla storia dell'oratoria politica italiana proposta di recente da Gabriele Pedullà⁷; e ancora ai lavori sul discorso politico in televisione⁸, e ai risultati della ricerca sul discorso politico e mediatico relativo al conflitto in Iraq condotta da un gruppo di studiosi dell'Università di Bologna, dell'Università di Siena e della Luiss⁹.

Oggi le opportunità aperte dalla digitalizzazione dei testi e dalla loro accessibilità e decifrabilità attraverso applicativi specifici permettono di lavorare su *corpora* di testi anche colossali. La cosiddetta Linguistica dei *corpora*, ha scritto Paul Bayley, «mira a rendere esplicito l'uso del linguaggio, basandosi sull'analisi computerizzata dei *patterns* linguistici ricorrenti, osservabili in campioni rappresentativi di testo. Tra le informazioni linguistiche offerte da un *corpus* sono comprese: liste lessicali, dati sulle frequenze relative dei lessemi, sulla loro distribuzione all'interno del *corpus*, *patterns* di coselezione (coselezione di più elementi lessicali), *patterns* grammaticali (coselezione di elementi lessicali e particolari forme grammaticali) e associazioni semantiche (coselezione di aree semantiche)»¹⁰. Al di là dei tecnicismi, evidentemente inevitabili su questo terreno, signi-

fica poter lavorare sul fronte lessicale, su quello grammaticale e, seppure forse con maggiori difficoltà, su quello semantico per insiemi coerenti di testi, riscontrandone continuità, variazioni, dispositivi.

Ciò ha fatto sì che il cantiere delle interrelazioni tra storia e linguistica stia conoscendo una nuova vitalità, testimoniata a livello internazionale da convegni, seminari e pubblicazioni recenti¹¹. Il confronto di «Contemporanea» si svolge dunque tra storici e linguisti e vuole proporre un giro di orizzonte su diverse esperienze di incontro realizzate negli ultimi anni tra le due discipline. Abbiamo chiesto ad alcuni protagonisti di quelle esperienze di illustrare quali possibilità si aprano alla storiografia con l'avvicinamento ad approcci di tipo linguistico (sia qualitativi che quantitativi), e di proporre una riflessione sui percorsi fatti fin qui e sui loro possibili sviluppi futuri.

Jacques Guilhaumou, su un registro prevalentemente teorico, e Jean-Philippe Genet, con una particolare attenzione alla Linguistica dei *corpora* e alle sue possibili applicazioni in ambito storico, ripercorrono nei loro interventi gli sviluppi dell'analisi del discorso così come si è sviluppata nella storiografia francese, seguendone diverse traiettorie, sia teoriche che applicative. Il primo sviluppa l'idea di una «storia linguistica dei

⁶ M.A. Cortelazzo, A. Tuzzi (a cura di), *Messaggi dal colle: i discorsi di fine anno dei Presidenti della Repubblica*, Venezia, Marsilio, 2007.

⁷ G. Pedullà, *Parole al potere. Discorsi politici italiani*, Milano, Rizzoli, 2011.

⁸ R. Petrilli, D. Femia, *Il discorso politico in Tv: caratteristiche e variazioni in un trentennio*, in *1976-2006 L'italiano televisivo*, Firenze, Accademia della Crusca, 2010.

⁹ J. Morley, P. Bayley (eds.), *Corpus-Assisted Discourse Studies on the Iraq Conflict: Wording the War*, New York, Routledge, 2009.

¹⁰ P. Bayley, *Analysing Language and Politics*, www.mediazioni.sitlec.unibo.it/images/stories/PDF_folder/document-pdf/2005/articoli2005/4%20bayley.pdf (ultimo accesso il 24 febbraio 2013).

¹¹ J.P. Genet, *Langue et histoire: des rapports nouveaux*, in J.M. Bertrand, P. Boilley, J.P. Genet, P. Schmitt-Pantel (sous la direction de), *Langue et histoire*, Paris, Publications de la Sorbonne, 2011.

concetti» che si avvale del contributo decisivo della *Begriffsgeschichte* di Koselleck e della teoria degli atti linguistici di Pocock e Skinner, con la centralità da questi assegnata all'intenzionalità del parlante. Il secondo propone un articolato quadro di studi e di percorsi di ricerca che mostrano tutta la complessità di un uso in ambito storico dell'analisi del discorso e allo stesso tempo la sua imprescindibilità nel moltiplicare le prospettive di lettura dei testi del passato.

Concentrando il suo intervento su una metodologia specifica, l'analisi statistica del contenuto, Michele Cortelazzo mette a fuoco le possibilità insite negli approcci quantitativi agli atti linguistici, sottolineandone però il carattere aggiuntivo e non sostitutivo rispetto alle analisi di tipo qualitativo. E da linguista propone agli storici una serie di importanti considerazioni preliminari, necessarie nel momento in cui ci si pone di fronte ai testi, soprattutto istituzionali, utilizzando i metodi dell'analisi statistica. Pietro Trifone affronta invece un problema importante di storia della lingua italiana: quello di ricostruire una storia linguistica nazionale che dia peso e rilevanza

alle varietà cittadine, delineando così una geografia linguistica del paese più corretta e più articolata di quanto non sia successo fino a oggi.

Chiudono il confronto due interventi che ci forniscono esempi concreti di ricerca basati sull'avvicinamento tra storia e linguistica. Marta Margotti, storica, e Rachele Raus, linguista, hanno lavorato insieme, con le loro rispettive competenze disciplinari, su una fonte importante per la storia religiosa della Francia novecentesca e tentano qui un bilancio del loro percorso. Francesca Socrate illustra alcuni passaggi della ricerca compiuta su uno specifico *corpus* di fonti orali relativo ai movimenti del Sessantotto, cercando di trarre dal lavoro fatto indicazioni metodologiche più generali sull'utilità, ma anche la complessità, di un approccio statistico-quantitativo nel fare emergere elementi distintivi di quella generazione di militanti. Sono due lavori che potremmo definire sperimentali ma che fanno ben intravedere il senso e le modalità, i limiti e le opportunità, di percorsi interdisciplinari in cui si facciano interagire problemi storici e approcci linguistici.

Lo storico, il discorso e il linguaggio

Gli storici hanno da sempre fatto ricorso ai testi, ma una volta superate le prove della decifrazione e dell'edizione, si sono di solito dedicati semplicemente al loro contenuto, accontentandosi tutt'al più di note stilistiche o, nel migliore dei casi, sociolinguistiche. Al contrario, è molto più raro che abbiano affrontato i problemi linguistici che ponevano i testi che studiavano facendo ricorso a metodi sviluppati dalla linguistica; in questo modo hanno tuttavia abbandonato una parte considerevole delle informazioni che queste fonti – che pretendevano di rispettare – apportavano. Di fatto, se da una parte gli storici rispettano la filologia in nome dell'erudizione della quale si pregiano, dall'altra ignorano, in un certo senso altezzosamente, le implicazioni della linguistica saussuriana. Gli storici continuano cioè a «dirci» essi stessi ciò che la fonte già dice; questa sostituzione ci porta, tuttavia, come ha scritto Michel Foucault, «enunciando ciò che è stato detto», a «ridire ciò che non è mai stato pronunciato»: «la sintesi soggettiva» dello storico non è che un commento, testimone di «un eccesso del significato sul significante»¹, che viene letteralmente ab-

bandonato. Un simile atteggiamento non è scientificamente sostenibile: se si vuole rispettare la fonte, non si può negare una parte, a volte essenziale, di ciò che essa trasmette.

Trasformazioni disciplinari

Questa constatazione senza appello deriva dall'impatto sulle scienze sociali dell'opera di de Saussure, che ha rivoluzionato la linguistica abbandonando la prospettiva storica e diacronica, sino a quel momento dominante, per gettare le basi della linguistica strutturale con il suo *Corso di linguistica generale*, pubblicato postumo nel 1916². Qui Saussure distingue la lingua dalla parola, prodotta dall'individuo: la loro somma costituisce il linguaggio. Sposta così l'obiettivo dalla diacronia alla sincronia, dall'evoluzione del linguaggio alla sua struttura e si focalizza sul suo funzionamento: per lui, la lingua non svolge altro ruolo che quello di strumento di comunicazione tra parlante e ascoltatori. La lingua è «una», operativa in quanto sistema, benché ciascun elemento linguistico³ abbia un valore che corrisponde precisamente a una propria funzione e a un

¹ M. Foucault, *La nascita della clinica. Il ruolo della medicina nella costituzione delle scienze umane*, Torino, Einaudi, 1969 [Paris, 1963], pp. 10-11.

² F. de Saussure, *Corso di linguistica generale*, Roma-Bari, Laterza, 2009 [Paris, 1916]. Il testo in francese è scaricabile sul sito www.enonciation.com (ultimo accesso il 2 febbraio 2015). I manoscritti di de Saussure sono stati oggetto di un'edizione – in francese – più recente: *Écrits de linguistique générale*, Paris, Gallimard, 2002.

³ Va da sé che è compito del linguista suddividere queste unità delimitandole e individuandole. O. Ducrot, T. Todorov, *Dizionario enciclopedico delle scienze del linguaggio*, Milano, Isedi, 1972 [Paris, 1972], pp. 33-35.

proprio posizionamento nel sistema della lingua in un dato momento: non può, pertanto, avere pienamente senso che facendo riferimento a questo sistema. In questo modo, la lingua, prodotto delle forze sociali, è essa stessa un'istituzione sociale. Ma il legame tra sistema e sincronia comporta conseguenze pericolose per lo storico: la determinazione del valore di un qualunque elemento linguistico presuppone la comprensione dell'insieme del sistema. Ora, questo sistema è difficile da apprendere: se i metodi della linguistica permettono di studiare il sistema della lingua nel mondo contemporaneo, non è così – ovviamente – per i sistemi del passato, dei quali noi, spesso, non possediamo che delle tracce. Al di là dei problemi inerenti alla linguistica saussuriana, i linguisti sono stati influenzati in misura crescente dalla riscoperta di Peirce⁴, da cui è nato il loro interesse per la differenza tra l'enunciato e l'enunciazione, un'opposizione della quale Émile Benveniste⁵ aveva già mostrato la ricchezza: una parte dei loro lavori è ascrivibile al campo della pragmatica⁶, dove è l'enunciazione, ossia l'atto linguistico, a essere valorizzato. Per prendere un esempio lampante, si consideri come noi siamo privati degli apporti dell'oralità: senza risalire alle pionieristiche registrazioni delle attrici del XIX secolo, è sufficiente evocare i cinegiornali che mostrano i discorsi di Mussolini o di Léon Blum per constatare come il contesto retorico nel quale i nostri predecessori recepiamo la loro lingua fosse radicalmente differente

dal nostro. Ma qual era il contesto retorico ai tempi di Cavour o di Victor Hugo? Le parole, irrigidite dalla scrittura o dalla stampa, private delle sfumature, degli accenti e del ritmo della voce, difficilmente possono fornire una risposta a questo interrogativo: quando anche in qualche caso gli enunciati si siano conservati, l'enunciazione, per contro, non può che essere oggetto di ipotesi a partire dalle tracce che l'enunciato ha – in misura maggiore o minore – conservato. Ma si tratta di un campo – senza dubbio il più cruciale per lo storico – in cui ci si può sforzare di ricostruire almeno parzialmente questo sistema della lingua – che è quello della semantica – grazie ad alcune delle nozioni centrali della linguistica, tra cui quelle di variazione o di scarto rispetto a ciò che costituisce la norma di un sistema, e ai metodi di analisi del testo che permettono di constatare e di misurare questi scarti a partire dalla «connotazione», ossia ciò che è proprio del locutore e non appartiene all'esperienza comune degli altri utilizzatori del linguaggio, o dalla «variazione» (si veda Chomsky), che pone tuttavia altri problemi allo storico, che cerca di risolverli attraverso i metodi di analisi del discorso.

Il termine non definisce esattamente la stessa cosa in linguistica e nelle scienze sociali. La storia, come ogni scienza sociale, considera le condizioni di produzione degli enunciati e le loro relazioni con ciò che è extralinguistico. Ma, mentre l'analisi del discorso dei linguisti si concentra anzitutto sulla materialità linguistica del testo, gli

⁴ Si veda C. Peirce, *Écrits sur le signe*, Paris, Seuil, 1988.

⁵ E. Benveniste, *Problemi di linguistica generale*, Milano, Il Saggiatore, 2010 [Paris, 1966].

⁶ A. Reboul, J. Moeschler, *La pragmatica oggi: una nuova scienza della comunicazione*, Fasano, Schena, 2011 [Paris, 1998].

storici e i sociologi si concentrano, invece, sulle condizioni esterne di produzione del discorso. Molti sistemi concettuali – che non sono di competenza dei linguisti *stricto sensu* e ai quali questi s’ispirano in misura maggiore o minore a seconda dei casi – sono a loro disposizione per questo: se è importante precisare che le condizioni storiche del testo (scritto o discorsivo, modalità di diffusione, contesto sociale e politico della scrittura o dell’enunciazione) sono rilevanti (e generalmente ben approfondite dagli storici), vi sono anche elementi strutturali che informano che ciascun enunciato si integra in un *continuum* composto di migliaia di altri enunciati (ciò che Foucault, per esempio, chiama l’archivio), dei quali la conoscenza è più o meno diffusa e ai quali è collegato – o meno – per mezzo del suo lessico, del suo autore, o del suo genere, anche laddove avesse scelto di opporsi a questi con una rottura.

È qui necessario ricordare l’apporto dei concetti socio-culturali, come quello di «campo» di Pierre Bourdieu⁷, quello di «formazione discorsiva» di Michel Foucault⁸ e, nell’ambito della critica letteraria, quello di «orizzonte d’attesa» sviluppato da Hans-Robert Jauss⁹. È necessario citare anche «la storia linguistica dei concetti», di cui Jacques Guilhaumou in Francia si è

fatto sostenitore, sia che si tratti della sua declinazione tedesca (la *Begriffsgeschichte* di Rainer Koselleck) o, soprattutto, di quella anglosassone (John Pocock, Quentin Skinner)¹⁰. Lo sviluppo relativamente recente della storia culturale (in confronto a ciò che abbiamo per lungo tempo chiamato «la storia delle mentalità») ha inoltre spinto gli storici a prendere in considerazione questa dimensione. Ma, in termini di analisi del discorso, si tratta di elementi «esterni», che si devono combinare con un’analisi che si potrebbe definire «interna», prendendo spunto dalla materialità linguistica degli enunciati. È a questo punto – e a questo punto solamente – che si può sostenere che in Francia si sia sviluppata una tradizione specifica di analisi del discorso: forzando un po’ la mano, si può dire che i linguisti, seguiti in questo da qualche storico, si siano posti come obiettivo principale quello di mettere a punto un metodo scientifico – e in larga misura quantitativo – per lo studio delle ideologie, fondato sulla lessicologia e la semantica. Questo sviluppo non ha prodotto solamente dei risultati storici importanti, ha anche progressivamente definito un arsenale metodologico che è oggi provvisto di strumenti informatici efficaci, soprattutto nel campo della logometria¹¹.

⁷ P. Bourdieu, *La parola e il potere. L’economia degli scambi linguistici*, Napoli, Guida, 1988 [Paris, 1982] e Id., *Langage et pouvoir symbolique*, Paris, Seuil, 2001.

⁸ M. Foucault, *Le parole e le cose. Un’archeologia delle scienze umane*, Milano, BUR, 2009 [Paris, 1966] e Id., *L’archeologia del sapere: una metodologia per la storia della cultura*, Milano, BUR, 2011 [Paris, 1969]. Cfr. D. Maingueneau, *Archéologie et analyse du discours*, «Texte», 2005, 2, www.revue-texto.net/Reperes/Themes/Maingueneau_Archeologie.html (ultimo accesso il 12 gennaio 2013).

⁹ H.R. Jauss, *Apologia dell’esperienza estetica*, Torino, Einaudi, 1985 [Konstanz, 1972].

¹⁰ J. Guilhaumou, *Discorso ed evento: per una storia linguistica delle idee*, Roma, Aracne, 2010 [Besançon, 2006].

¹¹ J. Guilhaumou, *L’analyse du discours du côté de l’histoire: les historiens et le tournant linguistique*, «Langage et Sociétés», 1993, 3; cfr. J.P. Genet, *Langue et histoire: des rapports nouveaux*, in J.M. Bertrand, P.

Traiettorie nella storiografia francese

Un esempio ci permetterà di comprendere intuitivamente ciò che è stato per molti storici il punto di partenza di questo cammino, quello di Régine Robin. Confrontata, durante il suo lavoro sui *cahiers de doléances*¹², con i quaderni del baliato di Semur-en-Auxois, Robin segnalò la necessità di disporre di tecniche adatte per «leggere, ordinare, normalizzare e comparare» i dati di questa documentazione. L'enorme ricerca di Beatrice Hyslop sui *cahiers de doléances* aveva già spianato il terreno, ma, nel complesso, i suoi risultati erano stati ingannevoli. D'altronde era davvero necessario essere grandi esperti per immaginare ciò che i francesi andavano recriminando nel 1789? Il contenuto dei *cahiers* era inesorabilmente ripetitivo. Ma come erano state espresse queste lagnanze? Secondo lo *status* sociale (avvocato? curato? gentiluomo?) dei redattori che si esprimevano a nome della collettività? Secondo le sfumature, talvolta sottili, del vocabolario o dei costrutti utilizzati? Secondo le specificità politiche, sociali o religiose dell'una o dell'altra entità del baliato? Quando si giunse a considerare tutti questi aspetti, a un'uniformità arida si sostituì una diversità ridondante che era necessario padroneggiare. E del tutto natural-

mente ci si rivolse alla linguistica, che conosceva allora, in Francia come altrove – in quanto scienza faro dello strutturalismo –, uno sviluppo impressionante: l'inventario che Robin stabilì per lo stato delle relazioni fra storia e linguistica all'inizio degli anni Settanta rimane ancor oggi la miglior introduzione al tema; anche se i metodi si sono evoluti, i problemi teorici non sono molto cambiati, fatta eccezione per lo sviluppo della pragmatica¹⁵.

In effetti si riscontra come in Francia – per delle ragioni che accennerò brevemente – la linguistica abbia interessato relativamente presto alcuni storici sociali e della politica¹⁴, e come nuove metodologie abbiano conosciuto una certa fortuna, anche se sono ancora lontane dall'appartenere al bagaglio comune degli storici attuali. Lo sviluppo della linguistica ha conosciuto, in Francia come altrove, un picco durante gli anni Sessanta, quando è diventata la scienza faro di uno strutturalismo imperante. Ma l'interesse degli storici era stato in precedenza destato dagli sviluppi della semantica e della semiologia: i lavori di semiologi come Algirdas Julien Greimas¹⁵ e di lessicologi come Georges Matoré¹⁶, e soprattutto quelli di Jean Dubois, colui che introdusse in Francia il concetto di analisi del discorso di Zelig

Boilley, J.P. Genet, P. Schmitt-Pantel (sous la direction de), *Langue et histoire. Colloque de l'École Doctorale d'Histoire de Paris I*, Paris, Publications de la Sorbonne, 2011.

¹² R. Robin, *La société française en 1789: Semur-en-Auxois*, Paris, Plon, 1970.

¹³ R. Robin, *Histoire et linguistique*, Paris, A. Colin, 1973; Robin si è in seguito orientata in altre direzioni. Si veda anche F. Mazière, *L'analyse du discours: histoire et pratiques*, Paris, Puf, 2005.

¹⁴ Si veda, per esempio, A. Prost, *Vocabulaire et typologie des familles politiques*, «Cahiers de lexicologie», 1969, 2.

¹⁵ A.J. Greimas, *La mode en 1830: langage et société. Écrits de jeunesse*, Paris, Puf, 2000; il saggio rappresenta la ripubblicazione della sua tesi di dottorato del 1948. Cfr. anche il successivo *La semantica strutturale: ricerca di metodo*, Milano, Rizzoli, 1969 [Paris, 1966].

¹⁶ G. Matoré, *Le vocabulaire et la société sous Louis-Philippe*, Genève, Slatkine, 1951.

Harris, dal 1962 al 1969 chiarirono il cammino che conduce dalla «parola» – e dunque dalla semplice lessicologia – all'enunciato e, successivamente, al discorso¹⁷. Parallelamente, storici e studiosi di letteratura lavoravano fianco a fianco e svilupparono al Centre de lexicologie politique de l'École normale supérieure de Saint-Cloud (oggi trasferita all'École normale supérieure de lettres et sciences humaines di Lione) i metodi della lessicologia quantitativa, che riprendevano e ampliavano i lavori dei linguisti realizzati nei laboratori di Nancy e di Besançon, che sono stati in seguito riuniti con il laboratorio di Saint-Cloud e con altri per dar vita all'Institut de la langue française. Le ricerche condotte da questo laboratorio, che vertevano primariamente sulla lingua della politica dal XVII al XX secolo¹⁸, sono state all'origine di opere che hanno contrassegnato inde-

lebilmente la storiografia¹⁹. Queste hanno stabilito nuovi standard per lo studio delle fonti storiche che rispettino le regole di base della Linguistica dei *corpora*: la costruzione sistematica degli insiemi e il trattamento sistematico dei dati lessicali mediante programmi che si sono perfezionati con il tempo, hanno prodotto risultati eccezionali, specialmente nel campo della storia della Rivoluzione francese, della storia dei sindacati, dei partiti politici e dei movimenti sociali²⁰.

In conclusione, l'analisi del discorso attualmente si diversifica, anche se mi sembra che sia meno presente del passato nel lavoro degli storici²¹. Inoltre, l'analisi del discorso trascende le frontiere tra le discipline: è molto presente tra i sociolinguisti²², ma soprattutto tra gli studiosi di letteratura. Il solo campo nel quale questa occupa realmente uno spazio di primo piano è quello

¹⁷ J. Dubois, *Le vocabulaire politique et social en France de 1869 à 1872 à travers les œuvres des écrivains, les revues et les journaux*, Paris, Larousse, 1962. Dubois creò la rivista «Langages» nel 1966; il numero 13 di questa rivista è particolarmente importante.

¹⁸ École normale supérieure de Saint-Cloud. Centre de lexicologie politique, *Formations et aspects du vocabulaire politique français, XVIII^e-XX^e siècle*, Paris, Diderot-Larousse, 1969.

¹⁹ Mi limiterò a citarne qualcuno, secondo l'ordine cronologico dei periodi trattati: M. Launay, *Jean-Jacques Rousseau écrivain politique (1712-1762)*, Cannes, Cel, 1972; A. Geffroy, J. Guilhaumou, S. Moreno, *Dictionnaire des usages socio-politiques du français (1770-1815)*, 2 voll., Paris, Klincksieck, 1985; J. Guilhaumou, *La langue politique et la Révolution française: de l'événement à la raison linguistique*, Paris, Meridiens-Klincksieck, 1989; Id., *L'avènement des porte-parole de la république (1789-1792)*, Villeneuve-d'Ascq, Presses Universitaires du Septentrion, 1998; Id., *Sieyès et l'ordre de la langue. L'invention de la politique moderne*, Paris, Éditions Kimé, 2002; M. Tournier, *Un vocabulaire ouvrier en 1848: essai de lexicométrie*, 4 voll., Saint-Cloud, Centre de Documentation de l'Ens de Saint-Cloud, 1975; A. Prost, *Vocabulaire des proclamations électorales de 1881, 1885 et 1889*, Paris, Puf, 1974; P. Muller, *Jaurès, vocabulaire et rhétorique*, Paris, Klincksieck, 1994; D. Peschanski, *Et pourtant ils tournent: vocabulaire et stratégie du PCF (1934-1936)*, Paris, Klincksieck, 1988.

²⁰ Un'introduzione breve ed esplicativa è quella di Francine Mazière: F. Mazière, *L'analyse du discours*, cit.; si veda anche D. Maingueneau, *L'analyse du discours: introduction aux lectures de l'archive*, Paris, Hachette, 1991.

²¹ Jacques Guilhaumou è fortemente in disaccordo con questa prospettiva: per lui, l'analisi del discorso occupa oggi uno spazio importante in ambito storico. Ha senza dubbio ragione sul piano della ricchezza dei risultati scientifici acquisiti; mi sembra però che abbia torto su quello della pratica sviluppata dagli storici. Non esiste un insegnamento sistematico di questi metodi nelle università e sono troppo pochi i ricercatori che li praticano.

²² L.J. Calvet, *La sociolinguistique*, Paris, Puf, 1993.

della storia della Rivoluzione francese²⁵, dove tende tuttavia a confondersi con ciò che Jacques Guilhaumou chiama la «storia linguistica dei concetti» e che mi pare molto diverso dall'analisi del discorso «classica». Per i periodi a noi più vicini, i metodi di analisi del discorso sono da lungo tempo largamente utilizzati dai politologi²⁴, ed è soprattutto Damon Mayaffre che attualmente li ha usati per diversi *corpora* che possono essere considerati allo stesso modo come storici o politologici²⁵: molto informati sui metodi statistici e informatici, i suoi lavori rappresentano senza dubbio la miglior via d'accesso per ogni storico che voglia volgere la sua attenzione verso l'analisi del discorso.

Quest'ultima consiste oggi nella digitalizzazione del testo: questa tappa, che trovava un ostacolo nelle dimensioni del testo stesso, è resa più semplice dall'esistenza di un gran numero di *database*, raggiungibili su internet; e dalla possibilità di scannerizzare facilmente i testi a stampa, quanto meno a partire dalla nascita della stampa industriale nel XIX secolo. Il testo digitalizzato presenta già in se stesso un certo numero di prerogative che informano sul fatto che si tratta di un ipertesto; come ha affermato

Damon Mayaffre, «la structuration hyper-textuelle du *corpus* permet une navigation inespérée et facilite l'exploration documentaire de vastes ensembles»²⁶. Il *corpus* è allora il solo a permettere di ricostruire da una parte la grammatica soggiacente, dall'altra il vocabolario e le sue accezioni semantiche²⁷. In particolare, i *corpora* permettono di realizzare l'etichettatura (*tagging*), operazione preliminare all'elaborazione del linguaggio naturale, dal momento che ogni forma viene provvista di un *tag* che conferisce un certo numero d'informazioni (per esempio: il suo statuto grammaticale, il suo genere, se è al singolare o al plurale, ecc.); quello dell'etichettatura è un compito pesante e complesso, ma per le lingue contemporanee esistono strumenti di annotazione automatica. Queste informazioni consentono successivamente di lemmatizzare i testi, il che permette di passare dalla «forma grafica» alla parola e, a partire da questa nuova segmentazione del testo, di produrre dei dizionari di parole al posto delle liste di forme grafiche; e ancora, diviene così possibile interrogare il *corpus* a partire da queste informazioni, per esempio lavorando specificamente sui sostantivi, sui verbi, gli aggettivi o i pro-

²⁵ Si veda, tra gli altri, S. Wahnich, *L'impossible citoyen: l'étranger dans le discours de la Révolution française*, Paris, Albin Michel, 1997; J. Guilhaumou, R. Monnier, M.F. Pigué (sous la direction de), *Dictionnaire des usages socio-politiques du français (1770-1815)*, cit.

²⁴ Cfr. D. Labbé, *Le vocabulaire de François Mitterrand*, Paris, Fondation nationale des sciences politiques, 1990; L.J. Calvet, J. Véronis, *Les mots de Nicolas Sarkozy*, Paris, Seuil, 2008.

²⁵ D. Mayaffre, *Le poids des mots. Le discours de gauche et de droite dans l'entre-deux-guerres. Maurice Thorez, Léon Blum, Pierre-Étienne Flandin et André Tardieu (1928-1939)*, Paris, Honoré Champion, 2000; Id., *Paroles de président. Jacques Chirac (1995-2003) et le discours présidentiel sous la V^e République*, Paris, Honoré Champion, 2004; Id., *Le Discours présidentiel sous la V^e République. Chirac, Mitterrand, Giscard, Pompidou, De Gaulle*, Paris, Presses de Sciences Po, 2012.

²⁶ D. Mayaffre, *Histoire et linguistique: le redémarrage. Considérations méthodologiques sur le traitement des textes en histoire: la logométrie*, in J.M. Bertrand, P. Boilley, J.P. Genet, P. Schmitt-Pantel (sous la direction de), *Langue et histoire*, cit.

²⁷ Cfr. B. Habert, A. Nazarenko, A. Salem, *Les linguistiques de corpus*, Paris, Colin, 1997.

nomi o sui nomi propri. I *corpora* «taggati» accessibili agli storici sono, malgrado ciò, assai rari.

I *corpora* hanno un'altra utilità, nel senso che rendono la misurazione del testo assai efficace. Il testo è misurabile: se ne possono contare le parole; per prendere l'esempio di un conteggio banale, ma indispensabile senza *corpus* digitalizzato, nei dibattiti televisivi che opposero François Mitterrand e Valéry Giscard d'Estaing, l'utilizzazione da parte del primo di parole nettamente più lunghe di quelle del suo avversario è sembrato essere uno degli indici di un elitismo culturale che avrebbe allontanato da lui l'elettorato popolare²⁸. Tuttavia, al di là dei semplici conteggi, è soprattutto il ricorso alla statistica lessicale che permette lo sfruttamento sistematico degli indici e dei dizionari di forme grafiche, e che si trova così al cuore della logometria. I metodi della statistica lessicale sono stati applicati anzitutto nel campo degli studi letterari²⁹, ma questi sono stati in breve tempo adottati anche dalla Linguistica dei *corpora*³⁰. Essi pongono ancora molti problemi teorici, dal

momento che queste statistiche poggiano largamente sul modello delle statistiche probabilistiche, quando invece il linguaggio possiede proprie leggi matematiche, come dimostrato dall'esistenza della legge di Zipf³¹; cionondimeno, si crea un certo consenso intorno a questi metodi a partire dal momento in cui si trova un accordo su ciò che ci si può aspettare da questi³². Non si tratta in questa sede di darne un prospetto completo, ma di evocarne l'elemento essenziale, il calcolo della specificità.

Esempi di analisi

Un modo semplice di cogliere la specificità del vocabolario è quello di partire dall'idea che il *corpus* sia composto da sottoinsiemi. Se prendo il caso, studiato da Damon Mayaffre, del discorso della destra o della sinistra nella Francia del periodo tra le due guerre, i testi di ogni autore preso in esame rappresentano un sotto-*corpus*. Ma si possono anche classificare i testi per anno: si avrà così un sotto-*corpus* per il 1956, uno per il 1957 e così via. Si possono, allo stesso modo, classificare i *corpora* in funzione delle circo-

²⁸ J.M. Cotteret, C. Emeri, J. Gerstlé, R. Moreau, *Giscard D'Estaing-Mitterrand, 54.774 mots pour convaincre*, Paris, Puf, 1976.

²⁹ C. Muller, *Principes et méthodes de statistique lexicale*, Paris, Hachette, 1977; la tesi di Muller (*Etude de statistique lexicale: le vocabulaire du théâtre de Pierre Corneille*, Paris, Larousse, 1967) ha largamente contribuito a diffondere queste metodologie.

³⁰ P. Lafon, *Dépouillements et statistique en lexicométrie*, Genève-Paris, Slatkine-Champion, 1984; A. Salem, *Pratique des segments répétés: essai de statistique textuelle*, Paris, Klincksieck, 1987; L. Lebart, A. Salem, *Statistique textuelle*, Paris, Dunod, 1994.

³¹ Si vedano le critiche di Alain Guerreau, in *Statistiques pour historiens: il manuale* – in realtà si tratta di ben più che un manuale –, che è stato redatto nel 2004 per gli studenti dell'École nationale des Chartes, è disponibile online all'Url: elec.enc.sorbonne.fr/statistiques/stat.pdf (ultimo accesso il 6 febbraio 2013). Sulla legge di Zipf, cfr. G.T. Guilbaud, *Zipf et les fréquences*, «Mots», 1980, 1.

³² «La valeur des seuils exploités traditionnelles ne doit pas être fixée en fonction d'un risque encouru, comme c'est le cas dans les exploitations traditionnelles des probabilités, mais du volume des listes d'événements que nous souhaitons retenir dans nos résultats. Les modèles, ont ici une simple fonction heuristique, nous faire découvrir les faits les plus remarquables afin de mobiliser notre jugement autour de ceux-ci au moment de l'interprétation», cfr. P. Lafon, *Statistique et textométrie: position des problèmes*, in J.M. Bertrand, P. Boilley, J.P. Genet, P. Schmitt-Pantel (sous la direction de), *Langue et histoire*, cit.

stanze nelle quali i testi sono stati prodotti: i manifesti elettorali, i discorsi di fronte ai diversi partiti politici, i discorsi alla camera dei deputati, ecc. Il *corpus* può così essere suddiviso in molti modi, tutti significativi. Prendiamo uno di questi segmenti, per esempio la divisione per autore: in ciascuno dei testi, le parole hanno una frequenza osservata (che si può trasformare in frequenza relativa per tener conto della grandezza dei sotto-*corpus*). Questa può essere comparata alla frequenza che ci si sarebbe potuta attendere in funzione della grandezza del sotto-*corpus* e della frequenza della parola nell'insieme del *corpus*. Prenderò a prestito un esempio da Pierre Muller⁵⁵: ci sono ventiquattromilanovecentotrentotto occorrenze nel complesso dei *Fleurs du Mal* di Charles Baudelaire e ce ne sono novencentotantré in *Le vin*, la terza parte del poema. La parola «ni» ha venticinque occorrenze nei *Fleurs du Mal*: ci si potrebbe dunque attendere di trovarla una volta in *Le vin*; invece la si trova quattro volte. «Ni» è dunque una sovrarappresentazione, ma è per questo motivo specifica? Tutto dipenderà dal solco

tracciato dal ricercatore e dal metodo statistico impiegato per calcolare la specificità⁵⁴, che si tratti del calcolo della probabilità o di un calcolo fondato sulla distribuzione e il test del chi-quadrato (χ^2).

Tutte queste operazioni sono compiute con il supporto dei programmi di logometria, che sono numerosi ma di cui qui mi limito a segnalare l'esistenza⁵⁵. La prima conseguenza dei calcoli e delle misurazioni che questi permettono di effettuare è di distinguere tra il vocabolario specifico, quello che mostra le differenze fra i differenti sotto-*corpus*, e il vocabolario comune, ossia quello che è formato da parole aventi in ciascuno dei sotto-*corpus* una frequenza prossima a quella che si riscontra nell'insieme del *corpus*. Si tratta, spesso, delle parole più frequenti. Così, come si vede nella tabella 1, se si restringe ai sostantivi e agli aggettivi, le cinque parole più frequenti dei presidenti della Repubblica francese sono le stesse.

Come ha affermato chiaramente Damon Mayaffre, «un certain nombre de termes apparaissent comme l'architecture immuable du discours présidentiel»⁵⁶ con le due

TAB. 1. *Le «parole piene» più frequenti del corpus presidenziale francese (Damon Mayaffre)*

| De Gaulle | Pompidou | Giscard | Mitterrand I | Mitterrand II | Chirac |
|-----------|------------|-----------|----------------|---------------|----------|
| France | France | France | France | France | France |
| Pays | Pays | Problème | Pays | Pays | Pays |
| Français | Europe | Pays | Grand | Europe | Problème |
| Peuple | Grand | Français | Gouvernement | Grand | Grand |
| Monde | Économique | Situation | Français [nom] | Homme | Europe |

⁵⁵ P. Muller, *Pistes. Pour une investigation systématique des textes*, Paris, Inrp, 1992, p. 58.

⁵⁴ P. Lafon, *Sur la variabilité de la fréquence des formes dans un corpus*, «Mots», 1980, 1.

⁵⁵ Sono numerosi i software presenti sul mercato. In Francia, gli storici hanno fatto uso soprattutto del programma Hyperbase, sviluppato a Nizza da Étienne Brunet e Lexico5, sviluppato all'Université Paris 3 da André Salem. Il progetto *Textométrie* ha permesso di sviluppare all'École normale supérieure di Lione Shs la piattaforma Txm, che riunisce pressoché tutte le funzionalità di questi software e ne implementa altre: si veda il sito www.textometrie.ens-lyon.fr (ultimo accesso il 6 febbraio 2015).

⁵⁶ D. Mayaffre, *Le Discours présidentiel sous la V^e République*, cit., p. 29.

parole, «France» e «pays» che si ritrovano per tutti i presidenti in prima o seconda posizione; altre parole («Europe», «français», «grand»), fanno parte di questa piattaforma comune. Si può allora procedere alla ricerca delle parole che possono essere considerate specifiche del singolo locutore o di un particolare periodo. Lo studio, se il *corpus* è taggato, può anche essere esteso alle strutture grammaticali dei testi. Il calcolo delle specificità presenta – lo si può vedere – dei risultati sorprendenti e, per giunta, facilmente interpretabili: per esempio, la sovrautilizzazione dell'avverbio «naturellement» da parte di Jacques Chirac è uno straordinario segnale rivelatore del conservatorismo inconscio del presidente; le sue scelte politiche «ne sont pas politiques mais réalistes ou pragmatiques, c'est-à-dire imposés par la nature des choses»⁵⁷. Va da sé che, in mancanza di un trattamento testometrico,

la funzione centrale nel discorso di questa parola-chiave – motore dell'artificio retorico che permette al presidente di dire tutto e il contrario di tutto – non avrebbe potuto essere individuata. Ugualmente non individuabili all'ascolto sono il «ce» o il «c'» specifici di Nicolas Sarkozy, che gli permettono di dimostrare con poco sforzo l'evidenza delle sue affermazioni⁵⁸.

Nel caso di un lessico che risulti a noi fondamentalmente estraneo, a differenza di quello del vocabolario presidenziale – che è al contempo in lingua francese e a noi contemporaneo – si può scoprire con certezza ciò che fa parte del lessico comune e le parole peculiari a partire dalle quali il contesto deve essere oggetto di uno studio specifico. La tabella 2 è relativa ad un *corpus* molto più esotico, quello della letteratura politica inglese del XV secolo⁵⁹. Se, da un lato, essa rivela la stessa architettura delle parole co-

TAB. 2. *I lemmi più frequenti nei testi politici inglesi del XV secolo*

| Gower | Hoccleve | Lydgate | Poesia lancasteriana | Poesia yorkista | Fortescue |
|---------|----------|---------|-------------------------|--------------------|-----------|
| King | King | King | King | England | King |
| Law | Lord | Peace | Lord | King | Council |
| Word | Heart | Lord | Peace | True | Realme |
| Fall | Love | Noble | Grace | Grace | Livelod |
| Right | Right | Worth | Christ | Pray | Land |
| Love | Give | Call | Right | Lord | Charge |
| People | Prince | Virtue | Love | Right | Lord |
| Pity | Peace | Grace | War | Gentle | Poor |
| Book | Poor | Give | Holy | Worth | Give |
| Worth | Virtue | Right | Faith | False | Office |
| Land | People | Glad | Crown | Prince | Reward |
| World | Rich | France | Heart | Might | People |
| Council | Know | Prince | Church | Name | Estate |
| Wise | Worth | Reign | Keep | Land | Might |
| Cause | Folk | Royal | Knight | Love | Royal |

⁵⁷ J. Guilhaumou, *Paroles de président. Jacques Chirac*, cit.

⁵⁸ D. Mayaffre, *Nicolas Sarkozy. Mesure et démesure du discours 2007-2012*, Paris, Presses de Sciences Po, 2012.

⁵⁹ J.P. Genet, *De Richard II à Richard III: le conseil*, in A. Marchandisse, J.L. Kupper (sous la direction de), *«L'ombre du pouvoir». Les entourages princiers au Moyen Âge*, Genève, Droz, 2003.

muni, dall'altra è necessario farne risaltare le differenze. Così «king» si ritrova nella stessa posizione in ciascun sotto-*corpus*, eccetto uno, quello della poesia politica yorkista, dove cede il posto a «England»; tutti gli altri *corpora* sono più o meno lancasteriani ed è significativo che gli yorkisti, che contestano la posizione del re di casata Lancaster insediato al trono, facciano riferimento non al potere di un monarca che essi combattono, ma all'Inghilterra e, attraverso questa, alla comunità politica nazionale. Tuttavia, le parole specifiche, che caratterizzano uno solo dei sotto-*corpus*, rivelano le particolarità di questi ultimi: così, il sotto-*corpus* dei testi di Sir John Fortescue (che arriverò a definire, *mutatis mutandis*, una sorta di antesignano inglese di Machiavelli), che sono gli unici a essere in prosa, è il solo sotto-*corpus* a prestare attenzione alle necessità materiali («livelod» definisce le risorse, «charge» è attribuito alle cariche fiscali, «office» all'amministrazione e «reward» ai salari e alle ricompense di cui beneficiano i detentori di questi incarichi). L'altro sotto-*corpus* che presenta il numero massimo di parole specifiche è quello della poesia lancasteriana: attaccati dagli yorkisti, i lancasteriani del sotto-*corpus* della «poesia lancasteriana» (brevi poemi polemicici della guerra delle Due Rose) introdussero nel vocabolario, diversamente dai grandi poeti moralisti (i laici John Gower e Thomas Hoccleve, e il benedettino John Lydgate, per quanto anch'essi lancasteriani) un

elemento religioso, grazie a termini come «Christ», «holy», «faith» e «church»; questi ultimi due si ritrovano appaiati.

Ma è ancora più importante conoscere l'apporto della specificità nell'esplorazione dei contesti di ciascuna delle parole, al cuore stesso dell'analisi semantica del testo, dal momento che in questo modo possiamo analizzare ciò che costituisce l'asse sintagmatico del senso di ciascuna parola. Abbiamo visto, poiché il *corpus* è stato diviso in sotto-*corpus*, come le specificità delle parole possano essere misurate e comparate in funzione di questi sotto-*corpus*: è così che viene esaminato l'asse paradigmatico del senso di ciascuna parola. Ora si può considerare che i contesti di una parola che si è scelto di studiare, e che sarà la parola-polo, possono costituire un sotto-*corpus* del testo (di cui si potrà fissare l'estensione⁴⁰), e si accosteranno le frequenze osservate in questo sotto-*corpus* a quelle osservate nell'insieme del testo.

Questa panoramica è evidentemente troppo rapida. Spero che ci sarà modo di mostrare i progressi che l'analisi del discorso permette di raggiungere nel campo dello studio dei testi, un metodo che è auspicabile che sia sistematicamente insegnato nel quadro della formazione degli storici. Bisogna pertanto ricordare due cose essenziali: i metodi statistici, per quanto sofisticati, non dispensano dal leggere il testo; il loro principale vantaggio consiste nel moltiplicare le prospettive di lettura del testo, di cui si sfrutta così tutto ciò che è stato possibile apportare

⁴⁰ Tutti i software di testometria presentati nell'indice Kwic (*key word in context*) o Kwoc (*key word out of context*) rappresentano in effetti la concordanza di una linea, ossia una decina di parole prima o dopo la parola-polo; la funzione «tema» di Hyperbase permette di determinare contesti più estesi, determinati dalla punteggiatura (la frase) o dal numero delle occorrenze (ad esempio cento forme prima, cento forme dopo).

dalla trasformazione in ipertesto per la digitalizzazione. Queste prospettive non dispensano da un lavoro di riflessione sui problemi più generali della semiologia, per cui è a maggior ragione essenziale che la storia culturale acquisisca un posto sempre più importante nella storiografia. Da questo punto di vista bisogna accogliere positivamente

l'interesse suscitato dal lavoro teorico sulla semantica dei *corpora* intrapreso da qualche anno da François Rastier⁴¹ e da altri, lavoro nei confronti del quale gli storici dovrebbero mostrarsi molto attenti, soprattutto quando fanno ricorso ai metodi della logometria.

[Traduzione di Jacopo Bassi]

Michele A. Cortelazzo

Metodi qualitativi e quantitativi di analisi dei testi

Gran parte della mia attività di ricerca è consistita nell'analisi di testi, singoli o organizzati in *corpora*¹. In particolare, mi sono occupato spesso di testi politici, che ho studiato dalla visuale della linguistica e della retorica, come documentano molti dei saggi ora raccolti nel volume *I sentieri della lingua*².

Negli ultimi anni ho affiancato ai tradizionali metodi qualitativi, dei metodi quantitativi. Si è sempre trattato di un affiancamento, mai di una sostituzione, in quanto dalla prospettiva disciplinare che coltivo, quella linguistica (e, in definitiva, stilistica), l'analisi qualitativa rimane lo strumento prioritario. Tuttavia, ho potuto sperimentare in

maniera positiva l'utilità di affiancare le prospettive quantitative a quelle qualitative. Questo affiancamento può avvenire in due modi diversi: da un'ottica esplorativa o, al contrario, da un'ottica confermativa.

Da un'ottica esplorativa, l'analisi quantitativa ci può indicare piste di ricerca che, a un approccio di tipo qualitativo, possono sfuggire, per esempio perché gli indizi che le giustificano risultano poco appariscenti nei testi, in quanto consistenti in corrispondenze dissimulate, diluite in testi di grande ampiezza, difficili da dominare, nella loro interezza, dalla memoria umana. Da un'ottica confermativa, l'analisi quantitativa ci può dare la certezza che le conclusioni alle

⁴¹ Si veda in particolare F. Rastier, *La mesure et le grain. Sémantique de corpus*, Paris, Honoré Champion, 2011.

¹ Questo contributo deve molto alle discussioni condotte nell'ambito del Giat – Gruppo interdisciplinare di analisi testuale (<http://www.giat.org/>).

² M.A. Cortelazzo, *I sentieri della lingua. Saggi sugli usi dell'italiano tra passato e presente*, a cura di C. Di Benedetto, S. Ondelli, A. Pezzin, S. Tonello, V. Ujcich, M. Viale, Padova, Esedra, 2012.

quali ci porta quella qualitativa sono davvero suffragate dai dati empirici; ci assicura che non sovrastimiamo fenomeni esistenti, magari molto evidenti, addirittura ostentati, ma poco significativi nel confronto, inevitabile, con un modello di riferimento; è, infine, un potente strumento di sintesi, soprattutto quando è in grado di dare una complessiva rappresentazione visiva dei risultati raggiunti.

Può essere utile anche allo storico confrontarsi con l'esperienza di un umanista che ha sperimentato l'ausilio di sistemi quantitativi di analisi dei testi, individuando, sul campo, le condizioni e i contesti nei quali il loro utilizzo può essere importante.

L'esperienza che ho maturato nella mia attività di ricerca più recente può essere utile allo storico contemporaneo anche perché si è svolta all'interno di un gruppo interdisciplinare, formato da linguisti, statistici, politologi, sociologi, economisti, che ha individuato principalmente nei testi istituzionali il luogo dove esercitare il comune spirito metodologico e, al tempo stesso, far interagire le diverse competenze che vengono dalle diverse origini disciplinari. Fino a ora ci siamo occupati dei messaggi di fine anno dei presidenti della Repubblica, dei loro discorsi d'insediamento, dei discorsi programmatici dei presidenti del Consiglio, dei discorsi annuali dei presidenti di Con-

findustria⁵, dei discorsi d'insediamento dei presidenti di Camera e Senato⁴.

Metodi quantitativi

Immagino che anche gli storici abbiano alle spalle ampie esperienze di analisi qualitative dei testi da diverse prospettive (ma certamente le hanno gli studiosi di letteratura e di lingua): c'è la «nostra» prospettiva stilistica, c'è la prospettiva dello storico che vede il testo come documento e quindi come fonte, e via dicendo. Per quanto si stia affermando, nel campo delle scienze «dure», della medicina, ma anche delle scienze sociali, l'idea che la ricerca o è quantitativa o non è, ritengo che, almeno nel campo delle scienze umane, un certo tipo di conoscenza si possa raggiungere solo con mezzi di tipo qualitativo. L'analisi dei testi è proprio uno di quei campi in cui la via primaria di ricerca è quella qualitativa. Ma ci sono casi nei quali la prospettiva quantitativa può aprire nuovi orizzonti alla ricerca, aumentando le capacità di conoscenza e d'interpretazione. Vedo due casi tipici di utilizzo dei metodi quantitativi: quando serve utilizzare statisticamente informazioni raccolte attraverso testi (per esempio interviste) o contenute in testi scritti, o quando serve estrarre dati da *corpora* testuali di grandi dimensioni (in questo secondo caso l'estrazione non può avvenire che in forma automatica

⁵ Rispettivamente in M.A. Cortelazzo, A. Tuzzi (a cura di), *Messaggi dal Colle. I discorsi di fine anno dei presidenti della Repubblica*, Venezia, Marsilio, 2007; M.A. Cortelazzo, A. Tuzzi, *Il discorso di insediamento del Presidente della Repubblica Giorgio Napolitano. Lessico e retorica*, «Lid'O. Lingua italiana d'oggi», 3, 2006; C. Di Benedetto, *Sessant'anni di discorsi programmatici governativi (1948-2008): tra ritualità e cambiamento*, «Lid'O. Lingua italiana d'oggi», 7, 2010; M.A. Cortelazzo, F. Gambarotto (a cura di), *Parole, economia, storia. I discorsi dei Presidenti di Confindustria dal 1945 al 2011*, Venezia, Marsilio, 2013.

⁴ Questo argomento finora è stato affrontato solo in una comunicazione orale, al convegno *Come parla il Parlamento. Giornata di studio sul linguaggio politico e parlamentare*, Roma, 12 febbraio 2013 organizzato dall'Associazione degli ex parlamentari della Repubblica.

o semiautomatica). Quella di avere un'ampia massa di testi, difficili da governare con la semplice lettura, è una situazione che capita sempre più spesso, grazie all'odierna facile disponibilità di ampi archivi digitali. Di fronte a *corpora* ampi, o accettiamo di ridurci all'analisi di campioni significativi (in senso generico, non nel senso statistico del termine) o ci gioviamo di sistemi di estrazione delle informazioni che si basino su tecniche diverse dall'onerosa lettura individuale dell'intero *corpus*.

Proprio per venir incontro alla necessità di analizzare ampi *corpora* di testi (nella ricerca, ma anche in ambito professionale e imprenditoriale), si stanno sviluppando strumenti di analisi automatica dei testi e correlate metodologie, in parte diverse tra di loro, che assumono diverse denominazioni (Analisi del testo, o dei testi o Analisi testuale; Analisi automatica dei testi; Text Mining; Natural language processing (Nlp); Information Retrieval; Analisi statistica dei dati testuali; Statistica testuale, Statistica linguistica, Statistica lessicale, Linguistica computazionale, Linguistica quantitativa, Linguistica dei *corpora*, Analisi del contenuto, Analisi statistica del contenuto). Vorrei porre l'accento sull'Analisi statistica del contenuto, che è un'evoluzione innovativa della classica Analisi del contenuto. «L'analisi del contenuto è un processo di acquisizione, sintesi e restituzione delle informazioni presenti in una comunicazione»⁵. Esistono due approcci: quello classico (iniziato negli anni Venti negli Stati

Uniti), basato sulla strutturazione dei dati testuali in categorie concettuali, quello più moderno (adottato inizialmente, negli anni Settanta, dall'analisi del contenuto di scuola francese e dalla statistica linguistica di matrice anglosassone), basato sulla codifica automatica delle forme lessicali presenti nei testi.

Il metodo classico consiste nell'attribuire alle porzioni del testo, che costituiscono l'unità statistica di analisi, una categoria concettuale. Il «contenuto» della porzione viene riconosciuto dal ricercatore come «coerente» con una o più categorie ed «etichettato» (in maniera manuale o parzialmente manuale), per permettere successive analisi qualitative o quantitative. L'etichettatura avviene in maniera manuale o parzialmente manuale, ora anche con l'aiuto di software specifici chiamati Caqdas (Computer Assisted Qualitative Data Analysis), come Atlas.ti, N6, Nvivo, Maxqda, Qda Miner, Tams. Il metodo moderno vede il testo come un insieme di stringhe ripetute di caratteri, le «unità testuali» o «forme» (sostanzialmente, quelle che nel linguaggio comune chiamiamo «parole»). L'unità statistica è costituita proprio dalle forme. L'analisi non mira a etichettare «porzioni» di *corpus*, ma proprio a individuare, secondo una prospettiva lessicale, le relazioni esistenti tra unità testuali e tra unità testuali e testi (o gruppi di testi). Ciò fa sì che la codifica possa avvenire solamente in modo automatico, con software come Taltac, Lexico, Alceste, Spad-T e, più di recente, R.⁶

⁵ A. Tuzzi, *L'analisi del contenuto. Introduzione ai metodi e alle tecniche di ricerca*, Roma, Carocci, 2005, p. 17.

⁶ Su tutto questo, cfr. A. Tuzzi, *L'analisi del contenuto*, cit., pp. 34-39.

Insomma, l'analisi statistica del contenuto lavora direttamente sul testo e considera il testo, con una prospettiva del tutto analoga a quella delle scienze del linguaggio, come una serie di stringhe, di frasi, di parole, a seconda del livello a cui si opera. Solitamente si lavora sulla parola così come si presenta nel testo, cioè la parola con la sua morfologia (quindi, la parola declinata o coniugata: è questo che si intende con il termine tecnico «forma»). Ma si può lavorare anche sui «lemmi», cioè sulle parole raggruppate neutralizzando la variazione morfologica (raggruppando, ad esempio, il singolare e il plurale di una stessa parola). Qui si può parlare di livello superiore rispetto alle forme (i lemmi, i segmenti, le categorie semantiche, altre caratteristiche come lunghezza, frequenza, ecc.) e di livello inferiore (le lettere, le sillabe, gli n-grams, i fonemi, i morfemi, ecc.).

Punti di forza e di debolezza dei metodi di analisi statistica

In sostanza, possiamo dire che i metodi di analisi statistica del testo riescono a restituirci il *corpus* analizzato in una forma diversa da quella lineare che ne costituisce la forma originaria. Vediamo, quindi, il testo trasformato in modo da darci tutta una serie d'informazioni per la successiva analisi qualitativa. Un esempio di trasformazione possibile con un'elaborazione di statistica linguistica è la rappresentazione mediante una lista di frequenza delle forme presenti in un testo. Nella tabella 1 possiamo vedere uno stralcio della lista di frequenza dei lemmi che costituiscono i testi dei messaggi di fine anno dei presidenti della Repubblica.

È anche possibile individuare automaticamente le specificità e le esclusività lessicali, intendendo per specificità quelle unità testuali che risultano significativamente sovrautilizzate (o sottoutilizzate) in un *sotto-corpus* rispetto al *corpus* complessivo e per esclusività le unità testuali presenti esclusivamente in un *sotto-corpus*. Nella tabella 2 sono presentate alcune specificità di tre presidenti della Repubblica (il segno + indica quelle che sono anche usate in maniera esclusiva).

Desidero insistere su un fatto: il dato quantitativo che esce dall'analisi statistica dei dati testuali non è, dal mio punto di vista, il risultato dell'analisi, ma una ristrutturazione del testo, tale da restituire con una relativa facilità le informazioni necessarie per la successiva analisi qualitativa.

Il punto di forza dell'analisi statistica dei dati testuali consiste nell'essere, come ho anticipato, uno strumento potente per analizzare *corpora* di grandi dimensioni, perché sostituisce la lettura umana con una lettura computerizzata dei dati. In altre parole, è utile quando ha senso sacrificare la ricchezza di un testo per ridurlo a rappresentazioni basate su grafici, indici, tabelle, percentuali, ecc.

L'analisi statistica dei testi, e le informazioni che ne derivano, rendono più solida l'argomentazione qualitativa; permettono ulteriori approfondimenti (per esempio generalizzazioni e classificazioni sicure, perché si possono fondare su dati sistematici); consentono la visualizzazione dei rapporti tra i testi (o insiemi di testi), tra unità testuali e tra unità testuali e testi.

Accanto a questi punti di forza, non mancano, ovviamente, i punti di debolezza, al-

TAB. 1.

| Lemma | Cat | Presidenti | | | | | | | | | Corpus |
|-----------------|-------|------------|---------|-------|---------|-------|---------|---------|----------|--------|--------|
| | | Einaudi | Gronchi | Segni | Saragat | Leone | Pertini | Cossiga | Scalfaro | Ciampi | |
| di | Prep. | 92 | 531 | 164 | 814 | 620 | 802 | 1.255 | 1.689 | 996 | 6.961 |
| : | : | : | : | : | : | : | : | : | : | : | : |
| essere | Verbo | 31 | 116 | 17 | 249 | 243 | 692 | 337 | 1.030 | 333 | 3.048 |
| : | : | : | : | : | : | : | : | : | : | : | : |
| avere | Verbo | 8 | 46 | 33 | 94 | 65 | 289 | 146 | 479 | 202 | 1.362 |
| : | : | : | : | : | : | : | : | : | : | : | : |
| anno | Sost. | 18 | 31 | 7 | 45 | 36 | 66 | 62 | 132 | 55 | 450 |
| : | : | : | : | : | : | : | : | : | : | : | : |
| popolo | Sost. | 8 | 26 | 14 | 20 | 13 | 120 | 42 | 54 | 31 | 328 |
| io | Pron. | 1 | 14 | 1 | 10 | 6 | 189 | 15 | 79 | 4 | 317 |
| italia | Sost. | 3 | 14 | 6 | 35 | 12 | 37 | 26 | 93 | 90 | 316 |
| : | : | : | : | : | : | : | : | : | : | : | : |
| paese | Sost. | 5 | 14 | 6 | 44 | 35 | 37 | 62 | 22 | 20 | 245 |
| pace | Sost. | 2 | 15 | 8 | 25 | 7 | 40 | 37 | 69 | 36 | 239 |
| : | : | : | : | : | : | : | : | : | : | : | : |
| italiano | Sost. | 8 | 15 | 6 | 16 | 15 | 72 | 12 | 27 | 48 | 219 |
| mondo | Sost. | 1 | 8 | 4 | 28 | 9 | 30 | 29 | 55 | 42 | 206 |
| : | : | : | : | : | : | : | : | : | : | : | : |
| italiano | Agg. | 3 | 9 | 5 | 15 | 7 | 72 | 11 | 41 | 26 | 189 |
| cittadino | Sost. | 3 | 11 | 3 | 20 | 15 | 7 | 59 | 47 | 24 | 189 |
| augurio | Sost. | 4 | 2 | 0 | 22 | 15 | 15 | 17 | 93 | 20 | 188 |
| libertà | Sost. | 5 | 16 | 6 | 23 | 8 | 20 | 57 | 23 | 21 | 179 |
| giovane | Sost. | 0 | 0 | 0 | 7 | 9 | 95 | 5 | 31 | 29 | 176 |
| problema | Sost. | 2 | 26 | 2 | 27 | 26 | 15 | 34 | 28 | 12 | 172 |
| vita | Sost. | 3 | 14 | 3 | 16 | 9 | 33 | 22 | 49 | 23 | 172 |
| : | : | : | : | : | : | : | : | : | : | : | : |
| stato | Sost. | 0 | 14 | 1 | 8 | 10 | 5 | 32 | 74 | 13 | 157 |
| : | : | : | : | : | : | : | : | : | : | : | : |
| forza | Sost. | 0 | 6 | 1 | 11 | 19 | 20 | 25 | 36 | 29 | 147 |
| Europa | Sost. | 0 | 2 | 0 | 8 | 3 | 15 | 31 | 44 | 45 | 146 |
| : | : | : | : | : | : | : | : | : | : | : | : |
| guerra | Sost. | 2 | 5 | 1 | 8 | 0 | 41 | 4 | 11 | 19 | 91 |
| pubblico | Agg. | 1 | 10 | 1 | 16 | 11 | 8 | 26 | 14 | 3 | 90 |
| democratico | Agg. | 0 | 7 | 2 | 19 | 8 | 4 | 26 | 16 | 8 | 90 |
| presidente | Sost. | 0 | 0 | 2 | 3 | 1 | 17 | 11 | 46 | 9 | 89 |
| parlare | Verbo | 0 | 1 | 0 | 3 | 1 | 23 | 4 | 49 | 8 | 89 |
| : | : | : | : | : | : | : | : | : | : | : | : |
| dignità | Sost. | 0 | 2 | 0 | 0 | 1 | 2 | 9 | 11 | 6 | 31 |
| antico | Agg. | 0 | 2 | 0 | 1 | 1 | 8 | 8 | 3 | 8 | 31 |
| cammino | Sost. | 3 | 7 | 1 | 1 | 1 | 2 | 7 | 5 | 4 | 31 |
| soluzione | Sost. | 1 | 9 | 1 | 5 | 3 | 1 | 7 | 3 | 1 | 31 |
| : | : | : | : | : | : | : | : | : | : | : | : |
| operaio | Agg. | 0 | 0 | 0 | 1 | 0 | 6 | 0 | 0 | 0 | 7 |
| contadino | Sost. | 0 | 0 | 0 | 5 | 0 | 2 | 0 | 0 | 0 | 7 |
| moglie | Sost. | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 5 | 7 |
| ringraziare | Verbo | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 1 | 7 |
| risorgimento | Sost. | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 6 | 7 |
| memoria | Sost. | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 3 | 7 |
| : | : | : | : | : | : | : | : | : | : | : | : |
| lager | Sost. | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| fascista | Sost. | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| nazifascista | Agg. | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| nazista | Sost. | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| allignare | Verbo | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| delegittimato | Agg. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| impunemente | Avv. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| inadempiente | Agg. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| processuale | Agg. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| condivisione. | Sost. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| identificazione | Sost. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| intermediario | Sost. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| pacatamente | Avv. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

TAB. 2.

| Saragat | Leone | Pertini |
|--------------------------------|--------------------------------|---------------------------------------|
| + Gran Bretagna | situazione economica | popolo italiano |
| + nostra cara patria | ordine pubblico | + italiane e italiani |
| + nostre libere istituzioni | + fino in fondo | + amici miei |
| + posti di lavoro | + gravi difficoltà | + giovani che mi ascoltate |
| senso di responsabilità | ciascuna famiglia | + nostra gioventù |
| + bilancia dei pagamenti | maggiore impegno | + per dannata ipotesi |
| + cari concittadini | momenti difficili | + popolo napoletano |
| + categorie dirigenti | popolo libero | + popolo siciliano |
| + Europa democratica | problemi economici | due superpotenze |
| + lotte di liberazione | + aumento dei prezzi | + barriera contro il terrorismo |
| + settori produttivi | + cassa integrazione | + fine dell'umanità |
| + sommo pontefice | + in questo difficile momento | in carcere |
| + ventesimo anniversario | + mondo produttivo | + movimento operaio |
| libere istituzioni | + risorse morali | + attentato contro il papa |
| giustizia sociale | + ristrutturazione industriale | + come fossimo antichi amici |
| livello di vita | + sicurezza individuale | + avvenire della patria |
| responsabilità politica | + sicurezza sociale | + creature umane |
| bene supremo | proprio dovere | + disarmo totale e controllato |
| problemi umani | altri paesi | + questa strage |
| + apertura dei mercati | ansia di giustizia | + questo terrorismo |
| + ceti popolari | assistenza sanitaria | + sarebbe la fine dell'umanità intera |
| + enti previdenziali | forze sociali | + sotto il fascismo |
| + enti pubblici | impegno di tutti | + camminiamo insieme |
| + minoranze violente e faziose | meno abbienti | + cari amici |
| + ordinato progresso | nostra capacità | + cataclisma sismico |
| + piccole minoranze | nostro avvenire | + codice morale |
| + potere di acquisto | senso il dovere | + fame nel mondo |
| + principi di libertà | | + illuminata dalla luce |
| + rapporto fiduciario | | + io credo nel popolo italiano |
| + riserve valutarie | | + io credo nella nostra gioventù |
| + società moderne | | + non armate la vostra mano |
| + tradizione di fine d'anno | | + ospedale da campo |
| classe dirigente | | + partigiani afgani |
| cosa pubblica | | + per costruire ordigni di morte |
| viva la repubblica | | + popolo calabrese |
| nostra costituzione | | + popolo polacco |
| | | + questi corrotti |
| | | + uomini liberi |
| | | noi anziani |
| | | al vostro fianco |
| | | mio animo |
| | | vostro animo |

cuni dei quali abbastanza pesanti soprattutto dal punto di vista della ricerca linguistica. Il limite maggiore è che i metodi di analisi statistica del contenuto si basano sul lessico e, quindi, servono solo indirettamente ad analizzare altri livelli linguistici, come la sintassi e la retorica, che spesso sono elementi fondamentali soprattutto nell'ambito dei testi politici, che in questa sede costitui-

scono i tipi di testi a cui facciamo continuamente implicito riferimento.

Ma, anche ammettendo che ci basti esaminare il lessico, gli strumenti di cui stiamo discutendo riescono a rappresentare l'aspetto formale del lessico, cioè il significante, e non l'aspetto semantico, il significato: non riconoscono sinonimi, non distinguono omografi (come «stato», parti-

cipio passato del verbo «essere», da «stato», sostantivo).

Può ovviare in parte a quest'ultimo problema la lemmatizzazione, cioè il raggruppamento delle forme in lemmi. È un'operazione molto onerosa, perché non completamente automatizzabile (anche se esistono software, come TreeTagger, che sono in grado di lemmatizzare una considerevole porzione dei testi) e pone problemi teorici-pratici non sempre facilmente risolvibili (per esempio la distinzione tra i participi, presenti e passati, con funzione verbale e quelli con funzione aggettivale; oppure il trattamento delle forme composte del verbo); ma permette di comparare più correttamente le scelte lessicali degli autori di testi diversi: solo con vocabolari lemmatizzati si può superare la variazione contingente dovuta al rispetto degli accordi richiesti dal contesto (una variazione che tocca il piano morfologico ma non quello lessicale, e coinvolge quindi solo marginalmente quello del contenuto).

Con *corpora* contenuti, il processo di lemmatizzazione è sostenibile, e la fatica dell'operazione si ripaga con il miglior affinamento dei risultati. La citata ricerca sui discorsi di fine anno dei presidenti della Repubblica si basa su un *corpus* lemmatizzato, che ha consentito di trarre conclusioni più sicure nei confronti tra gli stili discorsivi dei presidenti, basati sulle affinità e sulle distanze degli usi lessicali. Ma la lemmatizzazione è stata possibile prima di tutto perché si trattava di un *corpus* relativamente ridotto (circa novantamila parole).

Alla luce delle ricerche finora svolte, alcuni dei limiti indicati sono superabili o neutralizzabili. È vero, ad esempio, che i sistemi

di analisi statistica del contenuto non si occupano della semantica, ma comunque riescono a darci delle indicazioni quantomeno di semantica contestuale. I software per l'analisi statistica dei testi sono, infatti, in grado di offrirci dati relativi non solo alla frequenza delle singole parole, ma anche a quella delle sequenze di parole (con l'analisi dei cosiddetti segmenti ripetuti), aiutandoci, poi, con appropriati strumenti statistici, a individuare le sequenze più significative, quali sono quelle costituite da parole che occorrono prevalentemente o esclusivamente proprio nelle sequenze riscontrate.

È anche vero che l'analisi lessicale non può sostituire l'analisi retorica; ma raramente le due analisi vanno in direzione opposta. La riprova ci viene, ancora una volta, dalla ricerca sui presidenti della Repubblica: i risultati di chi ha centrato la propria analisi sul lessico e quelli di chi si è occupato di sintassi e retorica hanno finito per collimare.

La costruzione del corpus

Arriviamo alla questione cruciale: ogni analisi quantitativa, sia pure considerata come base per la successiva analisi qualitativa, è tanto migliore quanto migliore è il modo con cui è stato costruito il dato, cioè quanto migliore è il modo con cui viene costruito il *corpus*. Il *corpus* non è una semplice raccolta di testi; è una raccolta di testi che tiene conto, in maniera equilibrata, dell'omogeneità dei testi e della varietà della realtà che intende rappresentare, in relazione all'ipotesi di lavoro della ricerca specifica. Per esempio, chi volesse svolgere un'indagine sulla lingua dei giornali di

oggi, dovrebbe costruire un *corpus* di articoli nel quale siano rappresentate, in maniera equilibrata, sia le diverse testate, sia i diversi sottogeneri giornalistici (cronaca, politica, cultura, sport, ecc.).

Per la valutazione dell'omogeneità si devono considerare le principali dimensioni di variazione (diafasica, diamesica, diastratica, diacronica, diatopica)⁷, autore, lingua, stile e argomento⁸ e, naturalmente, la tipologia o genere testuale⁹.

Solo in alcuni casi il *corpus* non pone problemi di scelta, in quanto può essere esauritivo, poiché i testi trattati costituiscono un inventario chiuso: in tutte le ricerche che ho citato all'inizio, il *corpus* è costituito da tutti i discorsi di quel genere specifico dei presidenti della Repubblica, o dei presidenti del Consiglio, o dei presidenti di Confindustria. Un *corpus* deve essere facilmente articolabile in sotto-*corpus*, dal momento che l'analisi quantitativa è utile solo se ci permette di comparare tra di loro realtà diverse. Non è possibile trarre conclusioni sulle caratteristiche, ad esempio, di un presidente della Repubblica, se non si comparano i suoi discorsi con quelli degli altri presidenti, oppure se non si comparano i discorsi dei presidenti della Repubblica con quelli, di tenore analogo, di altre cariche istituzionali.

Alcuni esempi

Può essere utile presentare alcuni risultati dalle ricerche citate in apertura, per dare

degli esempi, sia pure in forma quasi aneddotica, delle informazioni che si possono trarre dall'elaborazione statistica di dati testuali.

Il primo esempio, che ho utilizzato in altre sedi, riguarda l'intreccio tra dimensione qualitativa e dimensione quantitativa. Studiando, con i tradizionalissimi strumenti dell'analisi stilistica e retorica, il discorso d'insediamento come Presidente di Giorgio Napolitano, è emerso chiaramente che la strutturazione portante del suo eloquio era di tipo binario, ricco di dittologie, con una forte tendenza alla *correctio*. Poteva, tuttavia, trattarsi di una caratteristica del genere testuale, o una caratteristica comune all'intero discorso politico; e quindi un tratto che non poteva caratterizzare una peculiarità di Napolitano. La comparazione, fondata sui dati quantitativi, con i discorsi d'insediamento degli altri presidenti conferma l'intuizione: la congiunzione «e» ha, in Napolitano, la frequenza relativamente più alta tra tutti i discorsi di insediamento (un dato che conferma la rilevanza specifica della strutturazione binaria); «pur(e)», «senza» e «non» anteposto ad aggettivi mostrano un'alta frequenza; i primi due, in particolare, presentano la frequenza più alta rispetto agli altri presidenti (a dimostrazione della significatività delle tattiche di attenuazione in Napolitano).

Un secondo esempio può venire dall'esame della distribuzione delle parti del discorso,

⁷ Su queste categorie cfr. E. Coseriu, *Lezioni di linguistica generale*, Torino, Boringhieri, 1975 e Id., *Einführung in die Allgemeine Sprachwissenschaft*, Tübingen, Francke, 1988; G. Berruto, *Sociolinguistica dell'italiano contemporaneo*, Roma, La Nuova Italia Scientifica, 1987.

⁸ C. Muller, E. Brunet, *La statistique résout-elle les problèmes d'attribution?*, «Strumenti critici», 1988, 3.

⁹ B. Mortara Garavelli, *Italienisch: Textsorten*, in G. von Holtus, M. Metzeltin, C. Schmitt (hrsgg.), *Lexikon der Romanistischen Linguistik, IV (Italienisch, Korsisch, Sardisch)*, Tübingen, Niemeyer, 1988.

TAB. 5.

| | Presidenti | | | | | | | | | Corpus |
|--------------|------------|---------|-------|---------|-------|---------|---------|----------|--------|--------|
| | Einaudi | Gronchi | Segni | Saragat | Leone | Pertini | Cossiga | Scalfaro | Ciampi | |
| Aggettivi | 15,8 | 14,4 | 16,2 | 14,8 | 14,6 | 12,0 | 16,1 | 12,2 | 14,0 | 13,7 |
| Avverbi | 4,5 | 5,1 | 4,1 | 4,5 | 5,2 | 5,5 | 5,2 | 5,9 | 4,5 | 5,2 |
| Sostantivi | 23,5 | 24,3 | 25,2 | 25,3 | 23,5 | 21,3 | 24,0 | 23,6 | 25,4 | 23,7 |
| Verbi | 14,5 | 12,7 | 11,9 | 13,2 | 15,6 | 19,7 | 12,4 | 17,9 | 15,2 | 15,9 |
| Articoli | 7,1 | 9,1 | 9,2 | 9,2 | 7,9 | 8,7 | 8,5 | 8,5 | 9,6 | 8,7 |
| Congiunzioni | 6,3 | 8,5 | 6,8 | 6,6 | 8,1 | 7,7 | 8,2 | 7,7 | 5,0 | 7,4 |
| Preposizioni | 20,3 | 19,6 | 20,4 | 19,6 | 18,0 | 14,4 | 19,2 | 15,8 | 18,7 | 17,4 |
| Pronomi | 7,8 | 6,0 | 4,9 | 6,1 | 6,6 | 8,6 | 5,3 | 7,0 | 6,1 | 6,7 |
| Altro* | 0,5 | 0,4 | 1,2 | 1,0 | 0,4 | 2,1 | 1,1 | 1,3 | 1,4 | 1,3 |
| | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 |

che nei discorsi di fine anno dei presidenti della Repubblica mostra i risultati indicati nella tabella 5.

Il dato parrebbe arido. Però, la distribuzione delle parti del discorso (in particolare di sostantivi e verbi) cambia notevolmente tra scritto e parlato. Può essere, quindi, un indice importante per individuare diversi stili discorsivi. Così, la percentuale di verbi più alta si trova in Pertini (a conferma del dato comunque evidente che il suo stile è fortemente colloquiale e orientato al parlato), mentre la più bassa si trova in Segni, che si dimostra, così, campione di un parlato che si caratterizza puramente come esecuzione orale di uno stile scritto.

Un terzo esempio ci viene dall'esame delle specificità e delle esclusività lessicali. Mi limito a un accenno. Nella sequenza dei presidenti, Saragat si caratterizza per aver usato il lessico politico e sociale con una frequenza maggiore di quello che ci potremmo aspettare se quello dei presidenti fosse un *corpus* unico e indifferenziato.

Un quarto e ultimo esempio mostra come la rappresentazione grafica delle similarità tra componenti di un *corpus* possa portare a ipotesi interpretative di carattere generale. Esiste la possibilità di rappresentare in un grafico il complesso delle somiglianze e delle differenze tra le parti di un *corpus* testuale. Quanto più due testi presentano le stesse parole con la stessa frequenza tanto più si collocano vicini nel grafico.

Bernardi e Tuzzi nel 2007¹⁰ hanno rappresentato, con questo sistema (che va sotto il nome di «analisi delle corrispondenze»), i discorsi dei presidenti della Repubblica (fino a Ciampi), ottenendo i grafici mostrati nella figura 1.

Di Benedetto¹¹, utilizzando la stessa tecnica, ha rappresentato nel modo illustrato nella figura 2 le affinità tra i discorsi programmatici dei presidenti del Consiglio.

È evidente che i due tipi di discorso presentano una configurazione ben diversa. I discorsi dei presidenti del Consiglio si dispongono sostanzialmente secondo una

¹⁰ L. Bernardi, A. Tuzzi, *Parole lette con misura (statistica)*, in M.A. Cortelazzo, A. Tuzzi (a cura di), *Messaggi dal Colle*, cit., pp. 109-134.

¹¹ C. Di Benedetto, *Sessant'anni di discorsi programmatici governativi (1948-2008)*, cit., p. 141.

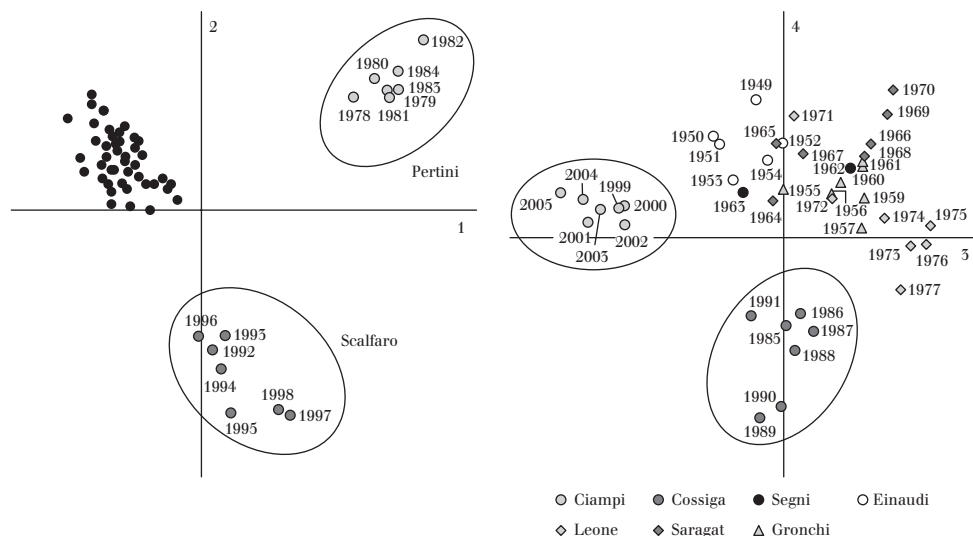


FIG. 1.

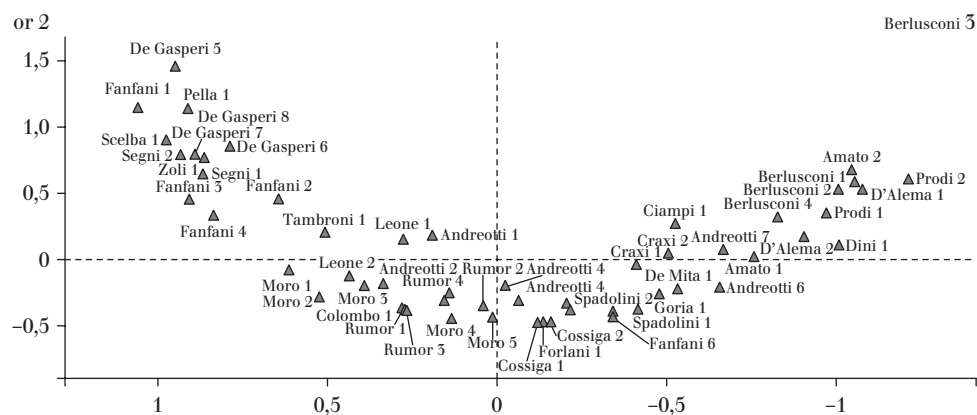


FIG. 2.

sequenza storica, al punto che presidenti del Consiglio ideologicamente antagonisti, ma cronologicamente vicini si collocano in posizioni contigue (è il caso di Prodi e Berlusconi), mentre persone che hanno assunto più volte l'incarico di presidente del Consiglio si collocano in parti diverse del grafico, in relazione al periodo dei singoli

mandati (emblematico il caso di Andreotti che si colloca, con i suoi sette discorsi, in tutti e quattro i quadranti del grafico). Diverso è il caso dei presidenti della Repubblica: tutti i discorsi dello stesso presidente si addensano nella stessa zona del grafico, mostrando quindi la sostanziale coerenza discorsiva di ogni presidente; ma le affinità

che si evidenziano tra i discorsi di diversi presidenti prescindono da legami di tipo cronologico o ideologico.

Si possono ipotizzare così, almeno a livello di ipotesi di lavoro, due tipologie di discorsi istituzionali: quelli, come i discorsi dei presidenti della Repubblica, nei quali la personalità dell'oratore mostra una forte indipendenza, tale da travalicare steccati ideologici e barriere cronologiche; altri, come i discorsi dei presidenti del Consiglio, nei quali la sequenza cronologica è l'asse principale attraverso la quale si dispiegano affinità e lontananze discorsive.

Una nuova frontiera: la distanza intertestuale

Concludo con un riferimento agli studi che il gruppo interdisciplinare che si occupa di analisi testuale sta sviluppando in questi ultimi tempi. Il tema è, ancora, quello della similarità dei testi, nel tentativo, ora, di riuscire a dare un contributo alle tematiche dell'attribuzione d'autore. Stiamo perfezionando metodologicamente e applicando ai testi italiani il concetto e la formula della distanza intertestuale di Labbé¹². La distanza intertestuale di due testi si basa sulle differenze in termini di frequenza di tutte le forme utilizzate. Se due testi sono identici, tutte le parole compaiono nei due testi con la stessa frequenza e la distanza risulta pari a zero. Il massimo teorico è pari a uno

e viene raggiunto quando due testi non hanno parole comuni (distanza massima). Abbiamo sperimentato la formula su un ampio *corpus* di testi letterari italiani, comprendente anche testi scritti dallo stesso autore. La formula riesce ad abbinare, con un'alta percentuale di successo, le opere dello stesso autore¹⁵.

Applicando la formula ai discorsi dei presidenti della Repubblica, si giunge al dendrogramma della figura 3, che rappresenta graficamente la vicinanza tra i testi.

Come si vede, in generale i testi degli stessi oratori risultano vicini tra di loro.

Un perfezionamento del modello della distanza intertestuale può portare a importanti risultati nell'individuazione di ipotesi di lavoro per le classiche indagini qualitative. Per esempio, osservando il dendrogramma, viene da chiedersi perché i discorsi di Gronchi del 1955 e di Cossiga del 1991 risultino così vicini (ma in realtà il discorso di Cossiga del 1991 è un non-discorso, in quanto quell'anno il Presidente si limitò, dopo una breve premessa, a fare gli auguri agli Italiani) o perché vi è un addensamento così eterogeneo, che incunea i discorsi di Segni 1963 e Ciampi 2005 tra i discorsi di Saragat del 1965 e 1966.

Ma la misura della distanza intertestuale può aiutare a smascherare i plagii e può dare elementi oggettivi nei processi di attribuzione di opere di autore incerto. Da

¹² C. Labbé, D. Labbé, *Inter-Textual Distance and Authorship Attribution. Corneille and Molière*, «Journal of Quantitative Linguistics», 2001, 4.

¹⁵ M.A. Cortelazzo, P. Nadalutti, A. Tuzzi, *Una versione iterativa della distanza intertestuale applicata a un corpus di opere della letteratura italiana contemporanea*, in A. Dister, D. Longrée, G. Purnell (sous la direction de), *Jadt 2012. Actes des 11^{es} Journées internationales d'Analyse statistique des Données Textuelles*, Liège, 15-15 June 2012, Liège-Bruxelles, Lasla-Sesla, 2012, pp. 295-307 e lid., *Improving Labbé's Inter-textual Distance: Testing a Revised Version on a Large Corpus of Italian Literature*, «Journal of Quantitative Linguistics», 2013, 2.

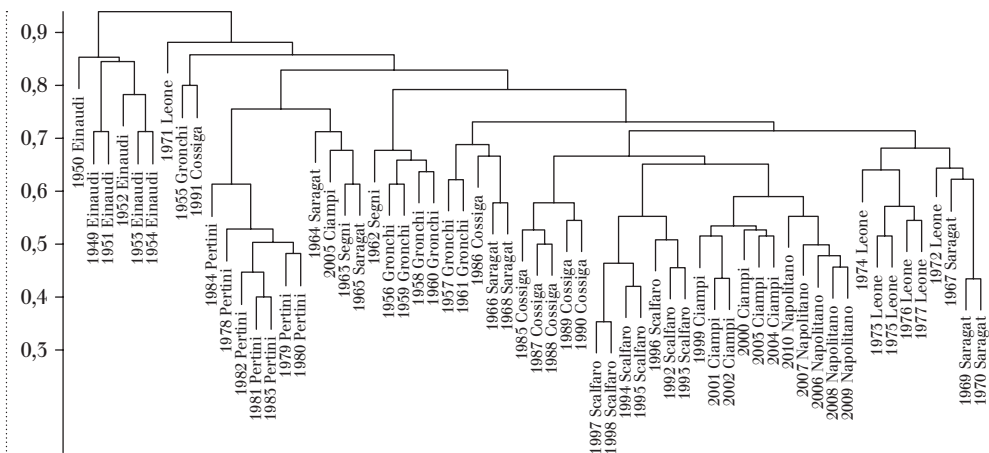


FIG. 3.

quest'ultimo punto di vista potrà essere interessante applicare la formula di Labbé a testi di incerta o discussa attribuzione ad Antonio Gramsci, già affrontati, con una diversa tecnica d'indagine dal gruppo guidato da Mirko degli Esposti¹⁴. Per concludere, il perfezionamento di queste formule potrà costituire un importante

ausilio agli studi umanistici, compresi quelli storici: pur non potendo offrire argomenti decisivi, i metodi che misurano la distanza tra testi possono offrire un importante supporto quando si debbano attribuire documenti a un autore o a un gruppo, o quando si vogliono smascherare dei falsi.

¹⁴ C. Basile, D. Benedetto, E. Caglioti, M. Degli Esposti, *An example of mathematical authorship attribution*, «Journal of mathematical Physics», 2008; D. Benedetto, E. Caglioti, M. Degli Esposti, *The Unreasonable Effectiveness of Mathematics in Human Sciences: The attribution of Texts by Antonio Gramsci*, in M. Emmer (ed.), *Imagine Math. Between Culture and Mathematics*, Milano, Springer, pp. 143-154.

La storia linguistica dei concetti

Nell'eccellente lavoro d'introduzione alla traduzione italiana del nostro volume *Discorso ed evento: per una storia linguistica delle idee*¹, Rachele Raus insiste, a giusto titolo, sull'apporto dei recenti contributi dell'analisi del discorso alla storia delle idee e dei concetti. In tale prospettiva, facendo convergere la tradizione ormai sedimentata dell'analisi storico-linguistica del discorso² e i primi frutti della tendenza attuale all'ontologia storica del soggetto parlante, proponiamo un approccio sintetico alla storia linguistica dei concetti, limitandoci ai lavori dei ricercatori francesi, seppure all'interno degli attuali dibattiti di ricerca internazionali.

Un percorso all'interno dell'analisi del discorso

I metodi strutturalisti di ricerca linguistica, lessicometrica e semantica vengono per la prima volta introdotti in campo storico su iniziativa di Régine Robin³, e in seguito con l'appoggio di ricercatori che hanno partecipato alla stesura della miscellanea *Langage et idéologies* (1974), il cui sottotitolo *Le discours comme objectif de l'histoire*⁴ è rivelatore. Questi metodi sono stati utilizzati

specificatamente per l'analisi del discorso politico dal Laboratorio di lessicologia politica della Scuola normale superiore (Ens) di Saint-Cloud, diretto da Maurice Tournier, fondatore della rivista «Mots. Les langages du politique», e hanno inoltre ispirato la rivista «Langage & Société» fondata più tardi da Pierre Achard.

Tale metodologia prevede una fase iniziale di selezione di parole ed enunciati attestati all'interno del *corpus*, più precisamente nell'«universo del discorso» intendendo per discorso quello politico, sulla base di criteri lessicali, sintattici e/o semantici.

Basandosi sulla conoscenza delle condizioni di produzione del discorso, l'analista si sforza di rendere visibili le strategie discorsive degli autori/attori della storia. Si tratta di approcciare le pratiche discorsive nella loro storicità e specificità piuttosto che di costituire, in modo sistematico, una teoria del discorso a partire da una teoria delle ideologie. Lo storico del discorso si dota quindi di un insieme di strumenti che restano sempre attuali: tale approccio, infatti, collegando la storia con la linguistica, inaugura un rapporto stabile con quegli strumenti lessicali, sintattici e semantici di

¹ Guilhaumou utilizza l'espressione «storia dei concetti», disciplina che ha una tradizione tedesca e anglofona. Nel titolo del volume *Discorso ed evento: per una storia linguistica delle idee*, Roma, Aracne, 2010 [Besançon, 2006], abbiamo tradotto con il più generico «storia delle idee» per inquadrare l'opera nel complesso più ampio della storia delle idee, ivi compresa quella dei concetti, in accordo con l'autore [N.d.T.].

² L'«analisi storico-linguistica del discorso» traduce l'espressione «analyse du discours du côté de l'histoire», per la quale rimandiamo al volume *Discorso ed evento* cit., p. XXII [N.d.T.].

³ Cfr. Il suo volume *Histoire et linguistique*, Paris, Colin, 1973.

⁴ «Il discorso come obiettivo della storia» [N.d.T.].

analisi che rispondono ai bisogni permanenti di una metodologia volta a descrivere l'uso delle parole in modo sistematico. D'altronde, l'archivio⁵ non è più meramente considerato come un documento, stampato o manoscritto, dal quale attingere dei referenti, ma diventa partecipe di un «gesto di lettura» in cui si attualizzano configurazioni significanti e dispositivi particolari.

Questo primo aspetto della problematica dell'enunciato raggiunge la prospettiva etnometodologica sulla riflessività delle descrizioni sociali⁶: gli individui-membri di una data società, considerati nei loro rapporti intersoggettivi, utilizzano il linguaggio naturale come contesto, risorsa e tema interpretativi, dando così un senso al proprio agire tramite la mediazione di un discorso autoriflessivo.

A essere privilegiato è il rapporto con l'evento nella misura in cui la portata riflessiva dell'enunciato scaturisce dal suo inserirsi specificatamente nell'azione situata. La messa in narrazione, che si dispiega lungo tutto un percorso tematico, acquisisce una portata globale e realizza il suo significato più profondo nel momento in cui emergono una o più espressioni capaci di riassumere l'intelligibilità del processo discorsivo che opera nell'evento.

Di fatto, lo spazio rarefatto, e perciò configurante, dell'enunciato riflessivo assume piena significazione all'interno della di-

spersione di enunciati eterogenei, che permette una gran varietà di norme che regolano il passaggio da un enunciato all'altro.

Negli anni Novanta del secolo scorso, attorno alla nozione di «formazione discorsiva», introdotta da Michel Foucault e ripensata poi da Michel Pêcheux nell'ambito dell'analisi del discorso⁷, si affina il passaggio dalla configurazione metodologica iniziale dell'analisi storico-linguistica del discorso alla problematica dell'enunciato, a seguito della quale la comprensione del concetto assume una nuova dimensione.

Evidenziando il carattere «transvalutatore» – da un momento all'altro della storia dell'analisi del discorso – della nozione-concetto di «formazione discorsiva», questa categoria centrale nella conoscenza delle pratiche discorsive si rivela cofondatrice della storia linguistica dei concetti.

Un primo avvicinamento è reso possibile dalla questione del «tempo storico», formulata dallo storico tedesco Reinhart Koselleck tramite le categorie di «spazio di esperienza» e «orizzonte di attesa»⁸, a seguito della considerazione che la storia è fatta di esperienze vissute e delle attese di un'umanità che agisce e che soffre.

Lo storico del discorso ha quindi progressivamente cambiato il proprio terreno di ricerca. Non si limita più a indagare esclusivamente le strategie discorsive che permet-

⁵ La nozione di «archivio» è ripresa da M. Foucault, *L'archeologia del sapere. Una metodologia per la storia della cultura*, Milano, Rizzoli, 1971 [Paris, 1969] [N.d.T.].

⁶ J. Widmer, *Langage et action sociale. Aspects philosophiques et sémiotiques du langage dans la perspective de l'ethnométhodologie*, Fribourg, Éditions Universitaires, 1986.

⁷ M. Pêcheux, *L'inquietude du discours*, testi scelti e presentati da Denise Maldidier, Paris, Éditions des Cendres, 1990.

⁸ R. Koselleck, *Futuro passato: per una semantica dei tempi storici*, Genova, Marietti, 1986 [Frankfurt am Main, 1979].

tono di rintracciare il «senso nascosto» delle varie attestazioni delle forme linguistiche, ma si sforza piuttosto di ridare la possibilità di osservare praticamente gli attori, gli autori e gli spettatori della storia, attraverso gli argomenti e i concetti. Opta, perciò, per un approccio «genealogico» del discorso a partire dall'interrogativo ontologico.

L'ontologia storica della lingua (politica)

Per comprendere quale sia il rapporto attuale tra la storia e la linguistica, occorre tornare sul modo in cui gli storici, soprattutto Jean-Marie Bertrand, Jean-Philippe Genet e altri⁹, non hanno colto sino a che punto i lavori nella storia linguistica dei concetti, in particolare quelli di Reinhard Koselleck e di Quentin Skinner¹⁰, accordino un posto centrale alla questione dell'intenzionalità, partecipando perciò significativamente al dibattito che ha caratterizzato la linguistica pragmatica e la filosofia del linguaggio sul terreno della storia.

Questi storici concepiscono la storia dei concetti nei termini esclusivi dell'approccio focalizzato sulle intenzioni più o meno coscienti degli attori e sul modo in cui esse si dispiegano nello spazio-tempo, senza legame alcuno con l'azione storica e quindi con la realtà.

Vari linguisti e filosofi hanno rifiutato la concezione dell'intenzione «svincolata» dall'azione¹¹.

Adottare l'approccio intenzionale vuol dire considerare anzitutto gli oggetti nel loro modo di esistenza, e poi ciò che pertiene più propriamente al lavoro mentale che dà loro consistenza. L'insistere su tale approccio presuppone quindi la conoscenza dei vissuti individuali, attribuendo agli individui delle credenze, che certamente sono colte nel loro significato mentale, ma soprattutto in merito al loro senso contestuale, alla loro eternalità.

Più precisamente, l'intenzionalità è una delle proprietà linguistiche principali che possiedono certi fenomeni mentali, e *in primis* il pensiero e le credenze, per rappresentare qualcosa al di fuori di sé, per rinviare a una cosa altra, ovvero a ciò che lo storico chiama contesto. Mettere l'accento sull'intenzionalità vuol dire allora prendere in considerazione l'attività linguistica dell'io/ agente nella storia.

In tal senso, la prospettiva genealogica e ontologica nella storia dei concetti si arricchisce tanto dal punto di vista storico che linguistico grazie all'apporto dei diversi approcci discorsivi. Anzitutto, ci riferiamo ai lavori innovatori di Reinhart Koselleck e di Quentin Skinner, che abbiamo presentato ampiamente nel nostro testo *Discours et événement*. Ma ci riferiamo anche ai lavori di filologia politica nell'ambito del Laboratorio italiano dell'Ens di Lione (Umr «Triangle») attorno a Jean-Louis Fournel e a Jean-Claude Zancarini, nonché alle ricerche di

⁹ Cfr. l'introduzione al volume J.M. Bertrand, P. Boilley, J.P. Genet e P. Schmitt (sous la direction de), *Lan-gue et histoire*, Paris, Publications de la Sorbonne, 2012.

¹⁰ Q. Skinner, *Visions of Politics*, 3 voll. (dedicati rispettivamente a *Regarding Method*, *Renaissance Virtues*, *Hobbes and Civil Science*), Cambridge, Cambridge University Press, 2002; trad. it. del II vol., Id., *Virtù rinascimentali*, Bologna, Il Mulino, 2006.

¹¹ G.E. Anscombe, *L'intention*, Paris, Gallimard, 2001.

linguistica testuale condotte a Losanna da Jean-Michel Adam e Ute Heidmann¹².

Nell'associare la filologia alla politica, quest'ultima si interroga sulla vita delle parole, restituendocene gli usi passati all'interno dei singoli contesti di credenze legittime. L'obiettivo del Laboratorio italiano si configura nel modo in cui la correttezza e la legittimazione degli usi testuali della lingua sono considerate indissociabili, e con esse la pertinenza dei suoi effetti nella città, dal momento che la lingua diviene il centro del potere. In questa prospettiva, un dato testo s'inserisce all'interno di una tradizione e di un movimento di rottura che gli attribuisce un'individualità e una storicità proprie.

Il principale «cantiere» del suddetto Laboratorio si interessa alla genealogia del repubblicanesimo. Infatti, Jean-Louis Fournel e Jean-Claude Zancarini lavorano sulla scrittura della politica e della storia nell'Italia del XVI secolo. Entrare nel vivo degli usi linguistici consiste nel mostrare in cosa «il tempo degli attori» venga considerato come «qualità dei tempi». Si tratta, in altri termini, di evidenziare qual è il peso della congiuntura degli specifici rapporti di forza che viene tradotta nei testi. Così avviene per la filologia politica, vero e proprio spazio testuale in cui si creano legami tra gli autori – Machiavelli e Guicciardini, Savonarola, Campanella e altri – tramite gli usi osservati nella trasformazione stessa delle situazioni politiche. Sull'esempio della libertà, nella quale si mescolano le idee

d'indipendenza della città e della forma di governo di cui i cittadini si fanno carico, l'intero spaccato della storia della lingua politica repubblicana viene riportato alla luce.

Allo stesso modo, è possibile mettere a frutto la notevole esperienza che la filologia ha acquisito sul terreno dell'alterità dell'atto ermeneutico, cosa da cui scaturirebbe, inoltre, il rinnovato interesse per le problematiche cognitive concernenti le «scienze del testo» e per i vari dibattiti che essa suscita.

Se un testo è un fatto di discorso grazie alla sua relazione con l'interdiscorso, soprattutto in quanto formazione socio-discorsiva intesa come luogo di circolazione dei testi e del loro raggrupparsi in generi testuali, cionondimeno questo tipo di categorizzazione discorsiva si fa incerta nel raffronto con l'interrogativo ontologico che si pone in due occasioni diverse: 1) Quali entità (naturali, individuali, collettive, sociali, eventemenziali, ecc.) sono sollecitate quando qualcuno parla di qualcosa che esiste? 2) Quali sono le operazioni cognitive, che implicano delle risorse specifiche degli attori e dei ricercatori, che permettono di dar corpo alle categorizzazioni discorsive?

Queste domande portano a porsi un interrogativo di tipo semiotico. Tale approccio, che articola il sostrato cognitivo con procedure enunciative e con costituenti discorsivi vari storicamente situati, conduce ad ammettere una dimensione variabile, stratificata e gerarchizzata del concetto. Ad esempio,

¹² J.L. Fournel, J.C. Zancarini, *La Politique de l'expérience. Savonarole, Guicciardini et le républicanisme florentin*, Alessandria, Edizioni dell'Orso, 2005; J.L. Fournel, *La cité du soleil et les territoires des hommes. Le savoir du monde chez Campanella*, Paris, Albin Michel, 2012; J.M. Adam, U. Heidmann (sous la direction de), *Science du texte et analyse de discours. Enjeux d'une interdisciplinarité*, Genève, Slaktine Erudition, 2005.

citiamo il caso specifico dei legami semiotici tra il buon senso e il senso comune¹⁵ o quello dell'enunciazione storica del racconto utopico funzionale all'invenzione della lingua politica e basato sulla volontà della condivisione (Jean-Louis Fournel).

In altre parole, ci sembra che l'analisi del discorso non possa non considerare le operazioni cognitive che presiedono alla delimitazione delle categorie che essa stessa sollecita nel percorso obbligato che va dal testo al discorso. Essa deve, quindi, interrogarsi sull'ontologia che occorre adottare in questa o quella descrizione discorsiva iscritta in una data storicità.

Ci siamo posti questo interrogativo assieme con Laurence Kaufmann nella miscellanea *L'invention de la société*¹⁴, e in modo autonomo nelle nostre ricerche recenti su Sieyès. In questa prospettiva, ricorrere all'ontologia equivale a rintracciare il percorso di nozioni-concetti, in particolare della nozione di ordine sociale e dell'io, durante tutto il processo genealogico d'invenzione della lingua politica nei tempi moderni. Occorre considerare il soggetto della lingua politica dal punto di vista dell'«ontologia storica di se stessi»¹⁵, interrogandosi sul modo in cui qualcosa di esistenziale, il tutto dell'ordine sociale confrontato con l'espressione dell'io, può divenire l'obiettivo dell'azione per il soggetto parlante.

Questo approccio può essere eventualmente esteso alla contemporaneità, nella

misura in cui il riferirsi a una tale ontologia della lingua politica perdura nel XIX e XX secolo, tramite la ripresa costante del modello politico-linguistico della Rivoluzione francese, per il fatto stesso di prendere in considerazione il paradigma della metafora della vita nelle descrizioni degli usi del linguaggio¹⁶.

All'inizio del XXI secolo, ci si stupisce allora che il concetto di «ideologia» perda la sua preminenza a causa dell'uso comune che ne viene fatto e che è contrario alla metafisica politica sottostante al movimento storico dell'individuo e del tutto che è alla base della nuova lingua politica.

Se si considera che la nozione d'ideologia viene elaborata in contrasto con la metafisica politica dell'io e in contrasto con l'espressione del sostrato sociologico del nuovo ordine sociale, il fatto che il modello linguistico della Rivoluzione francese perduri incessantemente ne preannuncia gli esiti negativi successivi. È infatti un'altra nozione, più euristica, che attualmente è divenuta predominante: quella della credenza associata al senso comune, allo stare assieme, al convivere, all'agire condiviso nell'ambito dell'universo di persone e oggetti comprensibili con l'ausilio della descrizione discorsiva situata.

Più nello specifico, ci sembra che nel dibattito all'interno della storia linguistica dei concetti seguito alle proposte dell'analista politico Mark Bevir, l'apporto della nozione

¹⁵ G.E. Sarfati (sous la direction de), *Discours et sens commun*, «Langages», 170, 2008.

¹⁴ *L'invention de la société. Nominalisme politique et science sociale au XVIIIème siècle*, «Raisons pratiques», 14, 2003.

¹⁵ M. Foucault, *L'archéologie del sapere*, cit.

¹⁶ S. Stadius, *Langage de l'enfant, langage du peuple. Qu'est-ce que la «vie du langage»?* Dijon, Les Presses du réel, 2012.

di credenza si sia ormai precisato¹⁷. È perciò possibile affermare che tutte le nostre esperienze linguistiche, osservate nel percorso che va dall'esprimere le credenze al produrre i concetti, sono portatrici di teorie legate a un particolare stato di cose del mondo in uno stadio preciso della formazione della lingua politica. A questo titolo, tali esperienze dipendono da stati intenzionali, e quindi da credenze individuali piuttosto che da rappresentazioni, che costituiscono altrettanti referenti dei discorsi sotto l'aspetto di forme significanti situate in classi di oggetti (di tipo lessicale, sintattico, retorico, ecc.) a partire dalle quali vengono prodotti i significati specifici.

D'altronde, questo punto di vista nominalista non contraddice l'approccio realista degli eventi né il costituirsi dei collettivi al loro interno. Il volume recente *Qu'est-ce qu'un collectif?*, curato dai sociologi Laurence Kaufmann e Dany Trom¹⁸, sottolinea l'importanza dell'ancoraggio preriflessivo delle credenze reciproche. L'aggregarsi dei punti di vista individuali non consiste nel sommare gli «io» tramite la mera interazione delle credenze reciproche nelle rappresentazioni collettive, ma ci confronta con un ancoraggio preriflessivo che interroga la natura stessa del fatto sociale. Al di là del rapporto autoriflessivo rispetto al sé, occorre circoscrivere l'ontologia sociale, la quale lega naturalmente, ovvero necessariamente, le credenze ai fatti sociali all'interno della situazione di comunicazione condivisa.

Contrariamente a ogni pantestualismo e diversamente da qualsiasi approccio analitico aprioristico, è necessario ammettere che i significati contenuti nei testi non esprimono solamente le credenze, e che quindi l'esistenza di queste ultime è situata in relazione alla loro esistenza significante, materiale.

Gli storici dei concetti non si limitano ad attribuire le credenze agli attori sulla base del mero significato del discorso, ma mettono anzitutto queste credenze in relazione a oggetti storicamente e intenzionalmente significanti, ovvero moltiplicando, ovviamente sulla base della consistenza stessa del materiale testuale, le inferenze su questo o quello stato intenzionale che ci orienta poi verso la miglior comprensione possibile degli argomenti utilizzati nel discorso in atto.

Nell'universo delle credenze messe in luce dall'analisi del discorso, la radicalità storica acquisisce un ruolo di primo piano, non soltanto sulla base del suddetto interrogativo ontologico (qualcuno parla e qualcosa esiste in vista dell'emancipazione politica), ma anche dell'approccio nominalista, a beneficio della spiegazione storica delle azioni e delle pratiche politiche del tutto indissociabili degli individui e dei collettivi nell'ambito dell'ordine sociale in divenire. L'analisi storico-linguistica del discorso si dispiega perciò nel lungo corso della storia moderna e contemporanea all'interno di spazi di sapere diversificati per porne le fondamenta nella prospettiva dell'ontologia storica dell'io. In questo modo, essa integra

¹⁷ M. Bevir, *The Logic of the History of Ideas*, Cambridge, Cambridge University Press, 1999.

¹⁸ L. Kaufmann, D. Trom (sous la direction de), *Qu'est-ce qu'un collectif? Du commun à la politique*, «Raisons pratiques», 20, 2010.

la storia linguistica della formazione della lingua politica alle conoscenze sull'ordine sociale. Evitando accuratamente ogni reificazione disciplinare, la storia linguistica dei concetti, attraverso l'approccio nominalista, rimette al centro l'invenzione congiunta della storia, della sociologia, della lingu-

stica, dell'antropologia e della scienza politica. L'esempio probabilmente più classico di tale commistione disciplinare è quello dell'invenzione del discorso rivoluzionario dei giacobini francesi¹⁹ e la ripresa del loro messaggio nel discorso operaio²⁰.

[Traduzione di Rachele Raus]

Pietro Trifone

Comunità linguistica e modelli storiografici

Alberto Varvaro, nel saggio *Storia della lingua: passato e prospettive di una categoria controversa*, notava: «Al centro della problematica della storia della lingua si pone il concetto della coesistenza di sistemi linguistici molteplici nell'ambito di una stessa comunità». Ne consegue la necessità di «coordinare, nell'ambito di una storia della lingua italiana, all'asse evolutivo principale (latino volgare tardo → dialetto fiorentino → italiano colto [per lo più scritto] dei secoli scorsi → italiano parlato attuale) le diverse linee evolutive valide per Milano o Venezia o Napoli o Palermo, centri in cui la trasmissione linguistica da una generazione all'altra ha subito ad un certo momento una qualche frattura, perché altrimenti [...]

non si sarebbe giunti all'italiano parlato attuale, relativamente omogeneo»¹. Si tratta di considerazioni giustissime, e tuttavia occorre anche riconoscere che in molti casi, purtroppo, noi non riusciamo a stabilire con precisione il momento in cui si sono prodotti nei dialetti i cambiamenti linguistici connessi alla diffusione dell'italiano parlato. La supremazia dei dialetti nella lingua parlata da una parte, dell'italiano nella lingua scritta dall'altra, ha provocato una forte disparità nella documentazione, che privilegia nettamente l'italiano, e in particolare l'italiano letterario, rispetto alle varietà locali.

Questo squilibrio delle fonti ha avuto inevitabili riflessi sulla stessa elaborazione sto-

¹⁹ C. Vetter (a cura di), *La felicità è un'idea nuova in Europa. Contributo al lessico della rivoluzione francese*, tomi I e II, Trieste, Edizioni Università di Trieste, 2005, 2015.

²⁰ M. Margotti, R. Raus, *Du mot à l'action. Histoire et analyse linguistique de «La France, pays de mission?»*, Roma, Aracne, 2008.

¹ A. Varvaro, *Storia della lingua: passato e prospettive di una categoria controversa*, in Id., *La parola nel tempo. Lingua, società e storia*, Bologna, Il Mulino, 1984, pp. 51-52 e 45.

riografica: mentre è stato possibile scrivere un'organica storia della lingua italiana, impresa che Migliorini attuò in modo mirabile oltre mezzo secolo fa, nessuno ha mai osato concepire una compiuta storia linguistica d'Italia o degli italiani, che includesse sia le vicende della lingua (letteraria e non letteraria) sia quelle più eterogenee e sfuggenti dei dialetti. I due studiosi che si sono avvicinati di più all'obiettivo di un'integrale storia linguistica italiana, e non di una storia mirata alla sola lingua comune, sono Tullio De Mauro, cui si deve la magistrale sintesi della *Storia linguistica dell'Italia unita*, e Francesco Bruni, che ha curato *L'italiano nelle regioni*, una magnifica compagine di profili storico-linguistici regionali, che estendono lo sguardo a tutta la fase preunitaria. Entrambe queste opere si propongono essenzialmente di ricostruire le fasi dell'italianizzazione, ovvero le circostanze, gli agenti e le modalità del processo di diffusione della lingua comune in un contesto di generalizzata dialettologia. In virtù di intelligenti e fruttuosi tagli trasversali, la vicenda dei dialetti viene tratteggiata non tanto in sé, quanto nei principali rapporti dinamici con l'italiano e negli importanti effetti linguistici che ne sono conseguiti.

Città e varietà linguistiche

In effetti la realizzazione di una compiuta storia linguistica del paese, capace di abbracciare in una visione d'insieme le linee di sviluppo della lingua comune, dei diversi dialetti e delle varietà intermedie, sarebbe un compito arduo anche per la più

efficiente delle *équipe* scientifiche, tenendo conto del fatto che la *Carta dei dialetti d'Italia* di Giovan Battista Pellegrini distingue all'interno del sistema italo-romanzo circa ottanta varietà², molte delle quali prive di testimonianze che consentano di ricostruirne la vicenda evolutiva. L'impegno diviene accettabile se invece si sceglie di concentrare l'analisi su alcuni ambienti e centri altamente rappresentativi: penso in particolare a città come Milano, Torino, Venezia, Firenze, Roma, Napoli, Palermo, dove la documentazione disponibile permette di avanzare ipotesi storicamente fondate riguardo all'intero repertorio linguistico della comunità urbana. Nasce di qui la mia idea di dedicare una collana alla lingua delle principali città italiane, intitolando le rispettive monografie *Storia linguistica* della città di volta in volta considerata, e non solo *Storia dell'italiano* o *Storia del dialetto* in quella città.

Nella premessa alla recente *Storia linguistica di Torino*, uno dei volumi che compongono la collana, Claudio Marazzini osserva:

Mi affascina l'idea di mostrare una volta di più che la storia linguistica è capace di reggere da sola un percorso diacronicamente esteso. [...] Gli storici obietteranno forse che questo non esaurisce l'orizzonte degli eventi: ovviamente non lo esaurisce, ma la storia linguistica mi sembra regga assai bene alla prova della globalità. Essa, infatti, è quintessenza di storia della cultura, ma allo stesso tempo è storia sociale, perché nella lingua si manifesta l'identità di tutti, dai ceti elevati al popolo più umile³.

² G.B. Pellegrini, *Carta dei dialetti d'Italia*, Pisa, Pacini, 1977.

³ C. Marazzini, *Storia linguistica di Torino*, Roma, Carocci, 2011, p. 10.

Come dicevo, la sfida di una storia linguistica integrale non può essere affrontata con speranze di successo senza disporre di una documentazione idonea, ovvero dotata di tre prerogative fondamentali: l'ampiezza, la varietà, la continuità. Le grandi comunità urbane sono i contesti in cui si ritrovano più facilmente questi ingredienti propizi alla ricerca storica in generale e storico-linguistica in particolare. Come scriveva Carlo Cattaneo in una pagina famosa, le città «sono il centro antico di tutte le comunicazioni di una larga e popolosa provincia; vi fanno capo tutte le strade, vi fanno capo tutti i mercati del contado, sono come il cuore del sistema delle vene; sono termini a cui si dirigono i consumi, e da cui si diramano le industrie ed i capitali; sono un punto d'intersezione o piuttosto un centro di gravità, che non si può far cadere su di un altro punto preso ad arbitrio. Gli uomini vi si congregano per diversi interessi, perché vi trovano i tribunali, le intendenze, le commissioni di leva, gli archivi, i libri delle ipoteche, le amministrazioni militari e sacerdotali, le grosse guarnigioni, gli ospitali»⁴.

Si capisce, quindi, che le città (con i rispettivi domini territoriali) costituiscano uno straordinario banco di prova per una storiografia linguistica modernamente sensibile all'ottica variazionale, che non può prescindere dagli aspetti geografici, sociali, economici richiamati da Cattaneo; né può fare a meno di interrogare i cospicui depositi di

scritture tipologicamente eterogenee, e non riferibili soltanto all'ambito letterario, che i centri maggiori mettono a disposizione dei ricercatori⁵.

Un esempio di analisi: i verbali processuali

Collegandomi al passo appena citato di Cattaneo, che include «i tribunali» tra le istituzioni peculiari dei nuclei urbani, tornerò a soffermarmi su un genere testuale che ha suscitato spesso l'interesse dei linguisti, in particolare per le sue potenziali relazioni con l'uso parlato. Mi riferisco ai verbali dei processi giudiziari, che riproducono, sia pure con la mediazione di un cancelliere, il dialogo tra un giudice colto e un imputato o un testimone spesso di estrazione subalterna. Abbiamo quindi, almeno all'origine, un parlato formale o semiformale, con le due parti che tendono a venirsi incontro, a favorire le occasioni di solidarietà linguistica.

Si rivelano di particolare interesse, in questo ambito, i processi per stregoneria, la cui carica emozionale e visionaria, accentuata dal frequente ricorso alla tortura, si riflette in una notevole ricchezza e intensità espressive. Dal confronto tra alcuni brevissimi estratti dei verbali di quattro processi per stregoneria del secondo Cinquecento e del primo Seicento, tenuti rispettivamente a Venezia, Roma, Milano e Palermo, risalta a prima vista un diverso grado di marcatezza dialettale, assai più alto nei testi di

⁴ C. Cattaneo, *Le più belle pagine scelte da Gaetano Salvemini*, postfazione di L. Cafagna, Roma, Donzelli, 1995, p. 100.

⁵ Sulla «linguistica urbana» di impronta variazionale mi permetto di rinviare a P. Trifone, *Le città, capitali della varietà linguistica*, in C. Marazzini (a cura di), *Italia dei territori e Italia del futuro. Varietà e mutamento nello spazio linguistico italiano*, Firenze, Le Lettere, 2012.

Venezia e Palermo che in quelli di Roma e Milano⁶.

1. Venezia, 1554. Processo contro Giovanna detta l'Astrologa.

Et interrogatus: Cognosséu donna Zuana milanese? Respondit: Monsignor sì, la è nostra vesina, et sta in la calle medesma per mezzo de mi. La se domanda la strolega. Ghe va purrassai persone là. Interrogatus: Èstu mai stà là, tu? Respondit: Monsignor sì. Interrogatus: Che cosa fala? Respondit: La indivina. Interrogatus: A che modo? Respondit: La varda in un gotto de aqua con una candela⁷.

2. Roma, 1557. Processo contro Caterina sicula.

Dopo ho cominciato andarci segretamente da dui mesi in qua, et l'altra sera essendo andato in casa de detta Catherina per negoziare la detta Francesca et anchora perché voleva gli facesse imprestare dodici scudi sopra certi svinatori, et in questo venne il Bargello, che doveva essere un'ora de notte, et trovandomi lì in casa in una camera me menò in prigione⁸.

3. Milano, 1607. Processo contro la Monaca di Monza.

E finalmente stando io assentata sopra il bassello della prima porta esso Osio mi violentò

gettandomi per terra, e nonostante ch'io cri-dassi e dicessi ah traditore ah traditore hebbe comertio contro di me, dicendoli ah l'onor mio, dicendolo la mia virginità, racordatevi chi io sono, et insoma lui hebbe comertio carnale meco una volta sola, perché subito ch'io poti rihavermi e levar su corsi via e lo piantai lì⁹.

4. Monreale (Palermo), 1623. Processo contro Orazio di Adamo.

Et esso testimonio mandao a chiamare a Norella Bruno sua matre, alla quale ci domandao la detta copia dello detto batisimo; la quale la sira dello stesso giorno ci portao detta copia et esso testimonio la detti allo detto Oratio. Quale copia di batisimo nella predetta stantia lo detto Horatio leggìo, et esso testimonio mentre la liggia ci dissi allo detto Oratio: «Come ijamo, havi li sittanta anni?»¹⁰.

Tratti come le forme interrogative veneziane *cognosséu* «conoscete» e *èstu* «sei» (usate, si badi, dallo stesso giudice) rinviano a una tradizione linguistica in cui il dialetto trovava libero campo non solo nel parlato familiare, ma anche nel discorso pubblico. Ricorda a questo proposito Lorenzo Tomasini che «al secolo XVI risalgono in effetti le prime testimonianze di un'usanza che a lungo colpì gli osservatori e i visitatori stranieri: la consuetudine, cioè, di discorrere in

⁶ Recupero in parte, rielaborandoli in una chiave diversa, alcuni spunti di un mio contributo precedente: P. Trifone, *Italiano letterario, dialetti, varietà regionali*, in P. Mazzamuto (a cura di), *La letteratura dialettale preunitaria*, Palermo, Università degli Studi, 1994, vol. 1, pp. 161-179.

⁷ M. Milani (a cura di), *Streghe e diavoli nei processi del S. Uffizio (1554-1592)*, postfazione di L. Cafagna, Padova, Centro Stampa Palazzo Maldura, 1989, p. 5.

⁸ I. Fosi, *Un processo per «streghe e furfanterie» nella Roma di Paolo IV (1557)*, «Ricerche per la storia religiosa di Roma», 1980, 4, p. 225; testo controllato sull'originale manoscritto, in Archivio di stato di Roma, Tribunale del governatore, *Processi criminali del XVI sec.*, b. 58, f. 629 v.

⁹ U. Colombo *et al.* (a cura di), *Vita e processo di Suor Virginia de Leyva Monaca di Monza*, Milano, Garzanti, 1985, p. 517.

¹⁰ G. Millunzi, S. Salomone-Marino, *Un processo di stregoneria nel 1623 in Sicilia*, «Archivio Storico Siciliano», 25, 1900, p. 297.

veneziano nei consigli e nei tribunali della Repubblica – costume che rimaneva ben solido anche dopo che nella maggior parte degli usi scritti il modello dell'italiano letterario si era ormai affermato»¹¹. Anche la lingua del processo palermitano presenta una spiccata coloritura locale, nella vocale tonica di *sira* «sera», in quella finale di *dissi* «disse», nel morfema del perfetto *mandao* «mandò», nell'accusativo preposizionale *chiamare a Norella*; ma al tempo stesso mostra una certa articolazione diafasica, come risulta dall'alternanza tra *iddu* nel discorso diretto di un testimone (fuori del passo citato) ed *esso* nell'uso corrente del verbalizzatore.

La sensibilità per la componente diafasica si accentua nel testo di Roma e in quello di Milano¹², in particolare nel primo, dove l'area d'origine è riconoscibile solo a stento, e in modo approssimativo, da una forma come *svinatori* per «svinato». Qualche elemento in più tradisce invece la provenienza del verbale milanese: si vedano ad esempio la sorda ipercorretta di *cridassi*, lo scempiamento delle geminate in *insoma* e in *comertio*, i dialettismi lessicali *assentata* «seduta» e *basello* «gradino». Quest'ultima forma è un adattamento della parola dialettale *basèll*, attestata nei principali vocabolari del milanese, a cominciare dal *Vocabolario milanese-italiano* di Francesco Cherubini; mentre per *assentata* si può allegare un riscontro del

secondo Cinquecento nei *Rabisch*, gli «arabeschi» in versi mistilingui della bizzarra Accademia della Valle di Blenio capeggiata da Giovan Paolo Lomazzo: «assentat sora un sass», «seduto sopra un sasso»¹³.

Anche da un microsondaggio come questo si possono trarre indicazioni significative sugli atteggiamenti tutt'altro che omogenei assunti nelle diverse comunità, dal Cinquecento in poi, riguardo all'opzione lingua-dialetto: mentre a Venezia la varietà locale affianca e talvolta sorpassa l'italiano persino in alcuni ambiti dell'uso colto e ufficiale, a Roma non riesce quasi mai a forzare lo sbarramento della scrittura, a parte naturalmente la letteratura dialettale e la produzione dei semicolti. Superata – con il contributo di Tullio De Mauro e della sua *Storia linguistica dell'Italia unita* – la riduttiva unilateralità dell'ottica toscanocentrica, occorre prendere atto che la storia linguistica italiana, in quanto espressione di un costitutivo, ramificato e persistente policentrismo, è caratterizzata da un numero di linee di sviluppo almeno pari a quello delle celebrate «cento città» del paese. Ma va anche sottolineato che queste molteplici piste prevedono spesso collegamenti e incroci tra individui e gruppi appartenenti a luoghi e tempi diversi, dando vita a un articolato reticolo di pratiche linguistiche e tradizioni testuali all'interno del più vasto «universo discorsivo» dell'area italo-romanza¹⁴.

¹¹ L. Tomasin, *Storia linguistica di Venezia*, Roma, Carocci, 2010, p. 72.

¹² Sulle vicende delle due città cfr., rispettivamente, P. Trifone, *Storia linguistica di Roma*, Roma, Carocci, 2008; S. Morgana, *Storia linguistica di Milano*, Roma, Carocci, 2012.

¹³ G.P. Lomazzo, *Rabisch*, a cura di D. Isella, Torino, Einaudi, 1993, p. 260. In area settentrionale, anche G. Boerio, *Dizionario del dialetto veneziano*, Venezia, Santini, 1829, attesta la diffusione dell'italiano regionale *assentato* «seduto» (s.v. *sentà*).

¹⁴ Interessanti, in tale direzione, le prospettive di ricerca aperte alla storia linguistica dall'integrazione del concetto di «comunità linguistica» con quello di «comunità discorsiva», che pone l'accento sull'identità

La scarsità di ricerche che in Italia hanno inteso incrociare storiografia e linguistica spinge a interrogarsi su questa mancanza, soprattutto perché in altri contesti culturali (in particolare in Francia) simili studi hanno prodotto risultati di notevole interesse. Se maggiore attenzione è stata dedicata in Italia alla storia della lingua, minore seguito hanno avuto i lavori che si sono avvalsi dell'analisi del discorso come strumento per ricostruire vicende storiche particolari o fenomeni di più lunga durata. Certamente l'alta specializzazione e l'estrema formalizzazione raggiunte da molte delle tecniche usate dai linguisti possono aver contribuito a scoraggiare gli storici dal ricorso a tali strumenti; allo stesso tempo, la percezione che i risultati degli studi storici non potessero offrire un arricchimento sostanziale alle ricerche linguistiche può aver spinto i ricercatori a rivolgersi ad altri metodi di analisi. Del resto, non tutti i testi si prestano

a un'analisi interdisciplinare: alcuni possono essere considerati dallo storico non particolarmente rilevanti per la propria ricerca; altri, invece, possono risultare per il linguista di scarso interesse analitico. L'interdisciplinarietà appare, dunque, spesso una «utopia metodologica» difficile da esercitare per storici e linguisti, in quanto richiede di superare sia limiti oggettivi (dovuti alle caratteristiche dei documenti che si hanno a disposizione), sia limiti soggettivi (la volontà degli studiosi di scambiarsi il punto di osservazione e di mettere in discussione approcci più usuali).

L'analisi del libro pubblicato nel settembre 1943 da Henri Godin e Yvan Daniel, *La France pays de mission?*, mi ha permesso di verificare, insieme a Rachele Raus, le potenzialità degli scambi tra discipline scientifiche diverse, ma anche le molte difficoltà che non appaiono sempre superabili¹. La duplice analisi – storica e linguistica – cui

multipla del parlante. In particolare, Raymund Wilhelm osserva che «l'individuo di solito fa parte contemporaneamente di più comunità linguistiche», come quella più ristretta dei parlanti di un dialetto e quella più estesa dei parlanti dell'italiano. «In più l'individuo fa parte di un certo numero, e di solito è un numero molto più grande, di comunità discorsive: oltre a quelle che tramandano le forme dell'interazione quotidiana possiamo pensare alle comunità professionali, alle comunità religiose», anche a gruppi di persone «lontane nello spazio o nel tempo l'una dall'altra». Si veda R. Wilhelm, *Che cos'è una comunità discorsiva? Le molteplici identità del parlante e i modelli della linguistica storica*, in S. Dessi Schmid et al. (hrsgg.), *Rahmen des Sprechens. Beiträge zu Valenztheorie, Varietätenlinguistik, Kreolistik, Kognitiver und Historischer Linguistik. Peter Koch zum 60. Geburtstag*, Tübingen, Narr, 2011, pp. 164 e 168.

¹ M. Margotti, R. Raus, *Du mot à l'action. Histoire et analyse linguistique de «La France pays de mission?»*, prefazione di J. Guilhaumou e postfazione di E. Poulat, Roma, Aracne, 2008. La prima edizione del libro di Godin e Daniel fu edita dalle Éditions du Cerf di Parigi e dalle Éditions de l'Abeille di Lione, per permetterne la diffusione sia nella Francia occupata, sia nel territorio controllato dal governo del maresciallo Pétain; il volume fu ristampato nel 1946 e nel 1950 (Éditions du Cerf) e riedito nel 1962 (Union générale d'édition).

è stato sottoposto questo libro fondamentale nella storia religiosa della Francia contemporanea ha consentito di valutare la complessità del testo, di ricostruire con maggiore esattezza le fasi della sua redazione e di meglio comprendere le ragioni della sua rapida diffusione e delle forti reazioni di segno opposto suscitate anche fuori dell'Esagono.

Dalla parola all'azione

Nell'agosto del 1942, durante l'occupazione tedesca della Francia, gli *abbés* Godin e Daniel, due assistenti della Jeunesse ouvrière chrétienne, furono incaricati dall'arcivescovo di Parigi, il cardinal Emmanuel Suhard, di preparare un rapporto sulla situazione religiosa degli ambienti operai della capitale e di presentare alcune soluzioni per «abbattere il muro» che separava la Chiesa cattolica dal proletariato². Dopo una prima bozza dattiloscritta terminata nel marzo 1943, intitolata *Mémoire sur la conquête chrétienne dans les milieux prolétaires*, i due preti presentarono il mese successivo una seconda versione rivista a Suhard il quale, profondamente colpito dai dati che confermavano le sue preoccupazioni circa lo stato religioso della diocesi parigina, chiese che il testo fosse distribuito ad alcuni ecclesiastici influenti della capitale. L'obiettivo era duplice: il testo doveva essere rapidamente pubblicato, allargando l'osservazione da Parigi all'insieme della Francia, e una nuova istituzione doveva sorgere per sperimentare forme originali

di azione dei cattolici per la «conquista cristiana» della classe operaia. Il libro, stampato nell'estate del 1943 dopo aver superato il controllo delle autorità ecclesiastiche e di quelle politiche, circolò nonostante le difficoltà dovute allo stato di guerra, mentre parallelamente Godin e Daniel si impegnarono per fondare un nuovo gruppo, denominato Mission de Paris, attraverso cui preti e laici si sarebbero dedicati a tempo pieno all'«apostolato operaio».

Questo sovrapporsi di «parole» e di «azione» presente nella genesi del libro del 1943 ha reso di estremo interesse il lavoro di «smontaggio» e «rimontaggio» analitico dello scritto dei due preti parigini: proprio per la sua carica innovatrice, resa possibile anche dalla scelta degli autori di riclassificazione e ricategorizzazione dei termini usati, il volume scritto nel pieno del conflitto mondiale può essere considerato uno dei testi «fondatori» dell'attività del cattolicesimo sociale francese nel secondo dopoguerra, in particolare dell'esperienza dei preti-operai (tra i quali molti si sarebbero formati nella Mission de Paris), oggetto di una dura condanna da parte del Vaticano nel 1954.

La particolarità del documento analizzato, che si presenta come il frutto di una meditata elaborazione teorica dei due autori e della pressante necessità d'intervento nella realtà sociale diffusa in una parte della Chiesa francese nella prima metà del Novecento, risulta ancora più evidente dall'esame delle diverse fasi redazionali del testo. La ricostruzione delle complesse vicende di

² Cfr. E. Poulat, *I preti operai (1943-1947)*, Brescia, Morcelliana, 1967 [Paris-Tournai, 1965]. Il libro è stato riedito in Id., *Les prêtres-ouvriers. Naissance et fin*, Paris, Éditions du Cerf, 1999. Cfr. anche M. Margotti, *Preti e operai. La Mission de Paris dal 1943 al 1954*, Torino, Paravia Scriptorium, 2000.

edizione del volume ha richiesto di mettere alla prova le competenze della storica e le competenze della linguista e, ancor prima, di confrontare i diversi metodi di lavoro, gli studi su temi simili condotti all'interno delle rispettive discipline e le personali ipotesi di ricerca. Il contributo della storica ha reso possibile il reperimento delle fonti archivistiche conservate in Francia, la rassegna degli studi sulla società e sulla Chiesa francese durante il conflitto mondiale e la conoscenza delle numerose iniziative e riflessioni dei cattolici intorno alla «questione operaia». Da parte sua, l'apporto della linguista ha permesso, attraverso l'analisi del discorso e la comparazione delle diverse versioni (prima rivolte a un pubblico più ristretto, poi a una platea indistinta di lettori), di attribuire ai due autori la paternità delle diverse parti del libro e, soprattutto, di considerare come un messaggio fortemente originale e potenzialmente sovvertitore potesse essere trasmesso attraverso la mediazione di un linguaggio ritenuto accettabile dalle gerarchie ecclesiastiche parigine e, più in generale, all'interno del cattolicesimo.

L'evento storico (le vicende che resero possibile la preparazione e la pubblicazione del volume) e l'evento discorsivo e linguistico (la forma e i contenuti assunti dal testo) si sono illuminati a vicenda, offrendo la possibilità di osservare nella sua com-

plexità la formazione di un testo per molti versi innovatore, se non propriamente «rivoluzionario»: la forza sovvertitrice del libro di Godin e Daniel risulta non soltanto dall'uso che in seguito ne fu fatto dai «missionari operai» (costantemente osservati con sospetto per le loro posizioni teologiche e politiche da una parte notevole dell'istituzione ecclesiastica), ma dall'originario intento riformatore dei due autori che segnando la gravità dell'allontanamento dalla Chiesa degli ambienti operai e proponendo nuovi metodi di azione – non più legati alle parrocchie e alle tradizionali associazioni cattoliche – intendevano sollecitare dall'interno il cambiamento dell'intera istituzione ecclesiastica.

Proprio i risultati ottenuti dal sondaggio compiuto sul libro di Godin e Daniel sollecitano a considerare con maggiore attenzione le potenzialità dell'innesto tra metodi della storiografia e metodi della linguistica: soprattutto la possibilità di comparare testi scritti a distanza di tempo dai medesimi autori o negli stessi ambienti culturali può permettere di osservare le trasformazioni intervenute nel tempo sia nella forma, sia nei contenuti del discorso, come anche avvalorare ipotesi di ricerca altrimenti difficilmente documentabili e proporre nuove domande alle numerose fonti scritte (e pure a quelle orali) che affollano la storia contemporanea.

L'osservatorio dell'analista del discorso

Dal punto di vista dell'analista del discorso¹, lavorare con lo storico permette di ancorare maggiormente l'oggetto d'analisi al suo contesto reale di produzione. In tal senso, studi francesi hanno proposto una fusione transdisciplinare delle metodologie storiche e linguistiche nella figura dello «storico-linguista», che affronta in modo nuovo le questioni discorsive per contribuire con un approccio originale a entrambe le discipline e, più in generale, alla storia delle idee².

Attraverso l'analisi della genesi e la filologia di *La France pays de mission*³ di Henri Godin e Yvan Daniel, a partire dai due dattiloscritti alla base della redazione del volume finale pubblicato nel 1943, abbiamo inteso capire i motivi per i quali il testo dei due preti sia risultato così «rivoluzionario» rispetto ad altri documenti coevi o precedenti che pure si erano già interessati ad argomenti simili; allo stesso tempo, attraverso la metodologia dell'analisi del discorso abbiamo inteso proporre risposte alla questione se il libro avesse posto le basi della successiva frattura tra il gruppo dei

preti-operai – largamente ispirati da questo scritto – e l'istituzione ecclesiastica, sfociata nella condanna vaticana del 1954.

Antiretorica del testo e ridefinizione del ruolo del prete missionario

Le modalità discorsive usate da Godin e Daniel attestano la presenza di un'antiretorica, posizionata polemicamente nei confronti delle istituzioni ecclesiastiche, che intendeva agire sul reale partendo da riclassificazioni e ricategorizzazioni delle parole usate dalla Chiesa, ridefinendole in modo nuovo.

I prassemi⁵ così riconfigurati permettono, infatti, di attivare dei programmi linguistici tali da garantire al testo una portata realmente rivoluzionaria, perché i suoi contenuti sono trasmessi in modo innovativo e appaiono capaci di coinvolgere un gran numero di persone. Si tratta di un messaggio costantemente co-costruito con il lettore ideale, in un percorso in cui ne viene richiesta la partecipazione e la connivenza, tramite meccanismi ironici e impliciti di-

¹ Per una presentazione degli studi linguistici denominati «analisi del discorso», sviluppatasi in Francia dagli anni Settanta del Novecento, cfr. F. Mazière, *L'analyse du discours. Histoire et pratiques*, Paris, Puf, 2005, cfr. anche D. Maingueneau, *Analyse de discours. Introduction aux lectures d'archive*, Paris, Hachette, 1991.

² Cfr. i lavori di J. Guilhaumou, e in particolare il volume *Discorso ed evento. Per una storia linguistica delle idee*, Roma, Aracne, 2010.

³ Cfr. C. Détrie, P. Siblot, B. Vérine, *Termes et concepts pour l'analyse du discours: une approche pragmatique*, Honoré Champion, Paris, 2001; B. Gardin, *Langage et lutte sociale*, Limoges, Lambert-Lucas, 2005. La nozione di «prassema» in luogo di «lessema, segno, parola», normalmente utilizzati in linguistica, rinvia all'approccio «prassematico», per il quale il fatto che le parole siano presenti nel discorso in un certo modo (all'interno di determinati contesti, con determinate denominazioni, ecc.) dipende dal modo in cui i soggetti parlanti/scriventi intendono rappresentare la propria azione tramite il linguaggio.

scorsivi che la riscrittura del testo aveva necessariamente comportato al fine di ottenere l'*imprimatur*.

Nell'opera in questione, inoltre, i missionari divengono i nuovi «portavoce» degli operai e quindi i reali mediatori, attraverso la loro azione rivolta all'esterno, nei confronti di un'istituzione cattolica che s'intendeva rinnovare nelle forme e nei contenuti. Dal punto di vista storico-linguistico, è interessante notare come il ruolo di mediazione svolto dal «missionario operaio» richiami quello svolto dal «missionario patriota» della Rivoluzione francese, come precisato da Jacques Guilhaumou nella prefazione al volume *Du mot à l'action*⁴: si può ipotizzare che in particolari momenti storici di crisi, coincidenti rispettivamente con il crollo della Monarchia francese nel XVIII secolo e con la Seconda guerra mondiale, alcune figure specifiche polarizzino elementi di criticità e si pongano come personaggi chiave del rinnovamento. In altre parole, tali mediatori assumono un ruolo centrale nella circolazione dei saperi in una fase di crisi, contribuendo a diffondere nuove idee attraverso forme discorsive altrettanto innovative.

Quando si parla di «innovazione», come quella prodotta dal libro fondatore della Mission de Paris, occorre però relativizzare l'evento storico-discorsivo, nel senso che esso è frutto di un contesto storico-linguistico che ha contribuito a renderlo tale, nell'ottica quindi di un'innovazione *ex novo*, non *ex nihilo*. Nell'opera di Daniel e

Godin confluiscono varie tendenze, tra le quali una complessa tradizione di contestazione dei privilegi ecclesiastici. Parimenti, la formula⁵ «prete-operaio» era stata utilizzata in Francia nel 1906 da alcuni sacerdoti che avevano fondato un'*alliance* di preti che si dedicavano al lavoro manuale e poi da Henri Perrin nel 1945, quando questo prete pubblica il suo diario di prete-operaio in Germania. La formula ha all'inizio un valore puramente designativo, indicando il prete che lavora come operaio, ma è solo a seguito della ridefinizione del ruolo missionario presente nel testo di Daniel e Godin che tale espressione potrà poi assumere nuovi significati per i preti della Mission de Paris, additando al prete missionario una soluzione di «rottura nella continuità».

Certamente, il volume *Du mot à l'action* vuole contribuire a fornire degli strumenti ulteriori per indagare tali meccanismi, mostrando soprattutto come le parole creino aspettative di mondi realizzabili, che assumono una portata rivoluzionaria nella misura in cui tali parole circolano e legittimano nuove memorie. L'analisi dell'azione dei cattolici francesi negli ambienti popolari, come proposta da Marta Margotti, si pone sul confine mobile che unisce storia sociale e storia delle istituzioni, per mostrare quanto la diffusione di idee potenzialmente sovvertitrici della realtà possa essere favorita da strutture gerarchicamente organizzate (in questo caso, la diocesi di Parigi attraverso il cardinal Suhard) che tentano di reagire in modo innovativo

⁴ Cfr. J. Guilhaumou, *Préface*, in M. Margotti, R. Raus, *Du mot à l'action*, cit., p. 9.

⁵ Per «formula» deve intendersi una sequenza linguistica dal carattere problematico che circola in modo polemico in un determinato spazio-tempo; cfr. A. Krieg-Planque, *«Purification ethnique». Une formule et son histoire*, Paris, Éditions du Cnrs, 2003, pp. 305-306.

alla crescente tensione provocata dai mutamenti in atto nella società. Attraverso l'indagine dell'analista del discorso si può trovare conferma di tale dinamica, dato che sul piano linguistico è possibile ricostruire l'intersezione tra discorso sociale e discorso istituzionale, considerando i reciproci condizionamenti e le divergenze di prospettive che emergono proprio nel momento in cui si intende tradurre le parole in azioni. Non soltanto, infatti, il discorso istituzionale della Chiesa è messo in discussione in quanto discorso dominante, ma un discorso a carattere sociale crea, per il tramite dei propri portavoce (in questo caso, Godin e Daniel), una memoria alternativa, le cui parole non possono essere silenziate⁶ e che

minaccia perciò l'autorità del primo, generando una situazione storico-linguistica di tipo «polemico»⁷.

Un'ultima annotazione riguarda la scelta di scrivere il volume *Du mot à l'action* in lingua francese. Il testo, infatti, dal punto di vista dell'analista del discorso, ha inteso inserirsi nel filone di ricerche che ha innovato gli studi linguistici a partire dalla Francia, oltre a voler contribuire alla diffusione di questi stessi studi in Italia per il tramite di un prodotto editoriale italiano. In questo sta la nostra «mediazione» nei confronti del dibattito della storia delle idee in Italia, che come tale diventa anche una sfida metodologica rispetto a quanto nel nostro paese si è fatto nelle discipline sia storiche, sia linguistiche.

Francesca Socrate

Le fonti orali alla prova della Linguistica dei *corpora*

Nel corso di una ricerca sul Sessantotto italiano ho raccolto, come una delle molti fonti su cui lavorare, una trentina d'interviste. Le persone intervistate erano in quel periodo studentesse e studenti per lo più universitari, che parteciparono al movimento tra il 1967 e il 1969, e che oggi si identificano con la generazione del Sessantotto. Sono quattro le principali domande con cui mi sono avvicinata a queste interviste,

domande fortemente legate alla lunga riflessione e alla ricca esperienza della storia orale:

1. La memoria del Sessantotto: ricostruirne la trama, individuare il nesso che la lega, nella sue articolazioni e variazioni, ai tratti socio-culturali degli intervistati e indagare modi e ragioni delle differenze (le caratteristiche socio-culturali che prendo in considerazione si riferiscono alla situazione

⁶ Per le forme di silenziamento della parola, cfr. E. Orlandi, *Les formes du silence*, Paris, Éditions des Centres, 1996.

⁷ Per le categorie di «discorso polemico» e «autoritario», cfr. E. Orlandi, *Typologie du discours et règles de la conversation*, «Langage et Société», 29, 1984.

personale di allora, dei tempi di cui raccontano, ma anche all'oggi, al presente dell'intervista). Elemento rilevante in questo quadro è la tensione tra memoria pubblica, memoria culturale, memoria collettiva, e, appunto, memoria individuale¹.

2. La parte giocata dal Sessantotto nella biografia di ciascuna delle persone intervistate, il ruolo che queste gli attribuiscono nel loro racconto: una svolta esistenziale o il compimento di un processo di formazione giovanile? l'inizio di una nuova fase o invece una parentesi? un errore, o piuttosto un colpo di fortuna?

3. Il rapporto con quell'esperienza: ci si riconoscono? è ormai troppo lontana o è qualcosa di cui vale la pena parlare ancora e pensarci su? è un segno forte della loro identità o uno dei tanti passaggi esistenziali, meno rilevante rispetto a qualcos'altro accaduto in contemporanea o pressappoco a quei tempi? quale impronta emotiva ha lasciato?

4. Per finire, sotteso a queste prime tre questioni e questione centrale in sé, l'impatto delle esperienze politiche e personali sulla soggettività o, meglio, il rapporto tra soggettività e esperienze politiche e personali.

Il principale oggetto della mia analisi non è ciò che è stato detto volontariamente, né tanto meno i giudizi pronunciati esplicita-

mente. Mi interessa l'implicito, se non l'inconsapevole; la narrazione, non la cronaca. Così come, dal punto di vista narrativo, m'interessa rintracciare, sulla scorta di quello che la storia orale pratica ormai da tempo, i fili del racconto e della memoria attraverso un'analisi della costruzione del discorso (montaggio, sottolineature e omissioni, *de-tournement*, ritmo, tono, retoriche, ecc.)².

Il software

Con queste domande ho tentato un approccio per me nuovo, e in genere direi abbastanza inusuale non solo per la storia orale ma, almeno in Italia³, per la storia in generale, ovvero la Linguistica dei *corpora*.

L'attenzione alla lingua è ovviamente parte importante del patrimonio della storia orale italiana, ma la sua utilizzazione è stata finora occasionale e episodica, quando non interna all'impianto teorico e metodologico della critica letteraria. Adottare la Ldc e i suoi strumenti informatici implica invece innanzitutto un approccio quantitativo, e quindi un'analisi sistematica delle fonti raccolte, e poi una messa a fuoco di fatti linguistici letti con strumenti presi a prestito dalla stessa linguistica.

Il software per il trattamento e l'analisi dei testi che ho utilizzato è Taltac⁴.

¹ Sulla negoziazione fra questi livelli, v. il recente R. Gildea, *Utopia and Conflict in the Oral Testimonies of French 1968 Activists*, «Memory studies», 2013, 1.

² Per una attenta e aggiornata ricognizione della storia orale, dei suoi metodi e dei suoi fondamenti epistemologici, v. B. Bonomo, *Voci della memoria. L'uso delle fonti orali nella ricerca storica*, Roma, Carocci, 2013.

³ Sarebbe interessante un censimento tra gli storici e le storiche italiane in questo senso. Recentemente è nato ad esempio un piccolo gruppo – formato dallo storico Alessandro Pes e dalla sociologa e statistica Francesca Atzeni – che applica la Ldc all'analisi dei discorsi di Mussolini. In Francia invece, come risulta evidente anche dagli interventi pubblicati in questa rubrica, la storiografia utilizza da decenni la Ldc e l'approccio quantitativo, soprattutto per ricerche incentrate sul linguaggio politico.

⁴ Taltac (Trattamento automatico lessicale e testuale per l'analisi del contenuto di un corpus) è un software sviluppato da S. Bolasco, F. Baiocchi e A. Morrone (www.taltac.it/it/index.shtml) che lavora nella duplice

Come sostengono Jean-Philippe Genet e Michele Cortelazzo in questo stesso forum, le capacità e i risultati dei software di analisi dei *corpora* possono essere molteplici. Vorrei qui limitare il mio discorso all'esperienza diretta di ricerca che ho condotto finora.

Con le trascrizioni delle interviste ho costruito il *corpus*, secondo le procedure di trattamento del testo previste da Taltac e in genere dai programmi informatici analoghi, associando a ogni intervistato/a i dati necessari alla mia analisi: l'età, il genere, il percorso scolastico, la sede universitaria del suo Sessantotto, le origini familiari, la formazione politica e l'appartenenza ideologica pre-Sessantotto, il rapporto con la politica dopo il Sessantotto, la collocazione sociale e l'area politico-culturale cui appartiene oggi⁵.

Il mio *corpus* (ventiquattromila parole circa) è composto da trentanove interviste: ventinove raccolte da me e un'altra decina già pubblicate⁶, sostanzialmente omogenee alle mie per il modo in cui sono state costruite e per il criterio di scelta delle persone.

Taltac mi ha dato in primo luogo il vocabolario complessivo delle interviste, ovvero la lista delle parole usate, con il numero delle occorrenze per ogni singola parola, le forme grammaticali, i lemmi.

Sempre sulla base del numero delle occorrenze, ma ponderate rispetto al peso delle interviste volta a volta prese in esame, Taltac ha costruito poi, per ogni singola categoria di intervistati definita rispetto alla variabile scelta, i rispettivi vocabolari specifici⁷. Ed è su questi vocabolari specifici che ho lavorato prevalentemente finora.

Approccio statistico e linguistico: i risultati

Ritorno alle domande da cui sono partita e ai miei primi due «esperimenti» con Taltac.

Nel primo esperimento ho usato la variabile dell'età, individuando nel movimento studentesco, a partire dalla formazione socio-culturale (famiglia, scuola, politica, consumi, ecc.), due diversi gruppi generazionali: quelli nati durante la guerra e quelli nati nel dopoguerra⁸. L'obiettivo dell'analisi

logica di *Text Analysis* (Ta) e di *Text Mining* (Tm): la prima è l'analisi automatizzata dei testi, la seconda è un'applicazione specifica della prima, che si riferisce in particolare a tutte le tecnologie informatiche necessarie a gestire grandi quantità di dati e di informazioni. La versione di riferimento di questo lavoro è la versione 2.

⁵ Una esposizione accurata e sistematica della Ldc e delle tecnologie informatiche in I. Chiari, *Introduzione alla linguistica computazionale*, Roma-Bari, Laterza, 2007.

⁶ Dieci interviste sono riprese da Circolo Gianni Bosio, *Un anno durato decenni. Vite di persone comuni prima, durante e dopo il '68*, Roma, Odradek, 2006; una, invece, è stata raccolta da F. Costantino e ora in F. Socrate (a cura di), *Un altro Sessantotto. La protesta nella memoria dei docenti dell'Università di Roma «La Sapienza»*, Roma, Bilibin, 2008.

⁷ Le parole specifiche sono quelle sovra-rappresentate (specificità positive) o sotto-rappresentate (specificità negative) nel gruppo preso in considerazione. Le specificità si basano su un valore di probabilità (p-value) fissato di default $\alpha \leq 0,025$ per ciascuna parola, dove *alfa* è la soglia di probabilità al di sotto della quale possiamo ritenere che le forme siano specifiche: il p-value indica infatti se la presenza di quella parola in quella determinata partizione è sovra- o sotto-dimensionata rispetto alle attese, senza che questo risultato sia dovuto al puro effetto del caso. Le liste che ho preso in considerazione in entrambi i casi sono composte dalle prime cento parole specifiche positive per ognuno dei due gruppi.

⁸ Un primo abbozzo dei due profili generazionali sta in F. Socrate, *Le differenti gioventù del '68*, «Annale Irsifar», 2013, in corso di stampa.

era di focalizzare, nel vocabolario specifico di queste due «generazioni», la differenza che le separa nel rapporto con il passato di cui raccontano o, meglio, nel rapporto tra l'io che narra e l'io di cui si narra. L'ottica che avevo adottato nell'analisi dei due vocabolari specifici era centrata prevalentemente su alcune funzioni grammaticali delle parole (nomi, pronomi, avverbi, forme verbali, ecc.)⁹.

Nel secondo esperimento, invece, ho cambiato i termini del confronto – dalla generazione al genere, mettendo a confronto le interviste degli uomini con quelle delle donne –, e ho cambiato anche l'approccio – da grammaticale a semantico¹⁰. Per avvicinarmi alle differenze tra uomini e donne nel modo di raccontare e nel contenuto del racconto ho analizzato i campi semantici in cui si articolano le narrazioni, ovvero gli insiemi di parole che si riferiscono a una stessa area di significati, e che possono essere in relazione tra loro secondo rapporti di vario tipo (antonimia, iponimia, sinonimia, ecc.)¹¹.

E i risultati? Schematicamente direi che i due casi mi hanno offerto, ai fini delle domande che come storica mi sono posta, una dimostrazione del valore e della capacità interpretativa, rispettivamente, dell'approc-

cio linguistico il primo, e di quello quantitativo, il secondo.

Nel primo caso la valenza euristica delle forme grammaticali mi ha permesso di raggiungere l'inconsapevole di cui dicevo, di andare cioè al di là dell'intenzione degli intervistati.

Un esempio per tutti: i tempi e i pronomi verbali dicono del loro rapporto con il passato molto più di quello che le persone intervistate mi avrebbero – e in alcuni casi mi hanno – dichiarato nel corso del nostro incontro. La prevalenza, nella «prima» generazione, del presente commentativo e del passato remoto alla terza persona singolare o nelle forme impersonali, tipici della narrazione storica «che esclude ogni forma linguistica autobiografica»¹² e invece quella, nella seconda generazione, del passato prossimo e dell'imperfetto alla prima persona singolare, «forma autobiografica per eccellenza»¹³, riflettono un diverso rapporto con quel loro passato: un occhio esterno e oggettivo quello della prima generazione, lo sguardo in soggettiva, specchio di un legame vivente tra l'io che narra e l'io narrato, nelle interviste della seconda.

L'apporto della linguistica, insomma, permette di arrivare all'analisi del discorso sottostante, di individuare, «tra le righe»,

⁹ Ead., *Classici e romantici. Le generazioni del '68 nel racconto di sé: un'analisi linguistica*, in P. Capuzzo, C. Giorgi, M. Martini, C. Sorba (a cura di), *Pensare la contemporaneità. Studi di storia per Mariuccia Salvati*, Roma, Viella, 2011.

¹⁰ Ead., *Maschile e femminile: memorie del '68*, in B. Bonomo, F. Bartolini, F. Socrate (a cura di), *Lo spazio della storia. Studio per Vittorio Vidotto*, Roma-Bari, Laterza, 2015.

¹¹ Una proposta metodologica interessante su come costruire i campi semantici nella Ldc, e sui risultati che questa può ottenere, in R. Heuser, L. Le-Khac, *A Quantitative Literary History of 2.958 Nineteenth-Century British Novels: The Semantic Cohort Method*, litlab.stanford.edu/?page_id=255 (ultimo accesso il 6 febbraio 2015).

¹² É. Benveniste, *Le relazioni di tempo nel verbo francese*, in Id., *Problemi di linguistica generale*, Milano, il Saggiatore, 2010, p. 285.

¹³ *Ibidem*, p. 289.

quello che il testo «ci lascia intendere, senza averlo voluto dire in maniera esplicita»¹⁴. Al di là del contenuto manifesto, la linguistica applicata all'analisi del discorso ricerca quindi il senso attraverso oggetti linguistici che non sono limitati al piano semantico. E qui la lingua nella sua forma grammaticale si dimostra una spia straordinaria della costruzione di senso del racconto.

Nel secondo caso, mi sono invece concentrata sul valore semantico del lessico specifico delle due diverse partizioni, uomini e donne, attraverso l'analisi delle singole parole – soprattutto sostantivi – e dei campi semantici che quelle delineano.

Questa volta il fuoco dell'indagine era la soggettività di genere: cosa separa e distingue uomini e donne nel loro vissuto e nel racconto di esso?

Il taglio di genere attraversa il *corpus* delineando due vocabolari distanti tra loro, ma in linea di massima prevedibili: secondo infatti quello che è stato sedimentato in questi ultimi decenni nella memoria culturale e in quella collettiva a proposito del nesso specifico che lega le donne alla sfera del privato, la memoria individuale delle intervistate privilegia fortemente parole legate agli aspetti relazionali, emotivi e quotidiani, contrariamente alle prime cento parole maschili che appartengono invece tutte al lessico politico. Un'analisi ravvicinata, grazie soprattutto alla definizione dei campi semantici, ha articolato

ulteriormente la lettura di questi dati dando risultati più interessanti. Non entro qui approfonditamente nel merito, ma mi preme sottolineare il fatto che l'approccio quantitativo ha permesso un rilevante avanzamento nell'analisi.

Anche qui un esempio per tutti: quell'elenco forse appunto prevedibile di parole tutte legate al privato che sono le parole specifiche delle donne è inaugurato da una parola che invece sorprende: un participio passato femminile singolare, «andata». Anche la coorte maschile mostra il termine «andato» tra le prime specificità, ma non è la prima parola: si trova al nono posto della lista.

«Sono andata», «ero andata»: movimento, spostamento da un luogo all'altro, da una condizione a un'altra: la storia al femminile, il modo in cui le donne la presentano, il loro racconto così fortemente autobiografico e individuale, comincia con una svolta.

Dopo un confronto, attraverso le concordanze, dei contesti in cui quella parola viene pronunciata nelle interviste, emerge che per le donne «ero andata» e «sono andata» ha un significato molto chiaro: racconta passaggi sofferti, spostamenti dallo spazio privato a quello pubblico, o anche viceversa. Nessuna casualità: piuttosto degli strappi. Nella memoria maschile, invece, l'equivalente autobiografico «andato» riflette scarti più contenuti, prevalentemente all'interno dello spazio pubblico, da un partito a un gruppo, da una facoltà a un'altra.

¹⁴ M. Bloch, *Apologia della storia*, Torino, Einaudi, 1969, p. 69: la citazione per intero («Ciò che il testo ci dice espressamente non costituisce più l'oggetto preferito della nostra attenzione. A noi di solito interessa maggiormente quello che ci lascia intendere, senza averlo voluto dire in maniera esplicita») era l'*incipit*, appositamente, di un articolo con cui Maurizio Gribaudi introduceva nel dibattito storiografico italiano la questione del rapporto tra storia e linguistica individuando ormai più di trent'anni fa alcuni nodi ancora oggi da riprendere e su cui riflettere (M. Gribaudi, *A proposito di linguistica e storia*, cit.).

Questo elemento inaspettato – l'importanza e la centralità della rottura e della svolta nel racconto delle donne, su cui evidentemente ha senso riflettere – è frutto della logica statistica con cui procede il software.

L'approccio quantitativo sottrae così l'analisi delle interviste al limite e ai condizionamenti imposti dai nostri pre-giudizi: gli elementi a prima vista incongrui che quell'approccio introduce sono segnali che indicano dove andare a cercare, cosa provare a capire, ponendo interrogativi da prendere sul serio, tanto più là dove non confermano ciò che già ci aspettavamo.

Quantitativo e qualitativo: le difficoltà

Vorrei concludere con alcune osservazioni sui terreni della storia orale che la Ldc mi sembra non possa coprire. C'è innanzitutto un'altra storia che volta per volta, intervista per intervista, dicono i testi. Anzi, le persone intervistate. Parlo dell'architettura generale del racconto: da dove chi narra comincia a

narrare, cosa sottolinea, la trama, le associazioni, il registro retorico, il ritmo, ecc.

Insomma, la costruzione narrativa, che mi sembra almeno finora sfuggire alle possibilità di quantificazione della Ldc¹⁵. E che, per citare solo uno dei più noti e felici esempi, Alessandro Portelli ha utilizzato invece in tante occasioni come suo principale strumento di analisi.

C'è poi un altro aspetto, specifico delle fonti orali, che opporrebbe sicuramente delle serie difficoltà alla quantificazione. Penso in primo luogo a tutto ciò che discende dalla caratteristica dialogica dell'intervista. Intanto le componenti pragmatiche e interattive della comunicazione (l'alternarsi di intervistatore e intervistato, tempi e modi¹⁶), ma soprattutto le dinamiche intersoggettive quali emergono attraverso i registri emotivi e nei condizionamenti reciproci che incidono sulla costruzione del dialogo¹⁷.

Un'ultima annotazione. In questa prima fase sono andata avanti in modo artigianale e da autodidatta: con molte scoperte

¹⁵ È in questo senso significativa la riflessione sui risultati raggiunti in una prima fase della loro ricerca dal gruppo dello Stanford Literary Lab che, diretto da Franco Moretti, lavora nel campo della storia letteraria attraverso un approccio di tipo digitale e quantitativo. In gioco, in questo intervento di Moretti, la questione dello stile: «When we have been doing quantitative stylistics, we have ended up quantifying units that cannot really be defined as stylistic. Stylistic analysis has, on the other hand, defined certain units that are certainly stylistic – free indirect discourse, stream of consciousness, analytic description, and a few others – but the trouble is, it's very hard to find ways of quantifying them. So we find ourselves caught between the quantification of something that is not stylistic – and the incapacity to quantify what instead is stylistic», (da un documento di discussione interna). Il sito del laboratorio è litlab.stanford.edu/ (ultimo accesso il 9 febbraio 2013).

¹⁶ In questo ambito è importante, e per me ancora tutto da esplorare, il contributo della pragmatica computazionale, ovvero di quella branca della linguistica computazionale che studia le relazioni fra le espressioni linguistiche ed il contesto.

¹⁷ A partire dai lavori di Alessandro Portelli e Luisa Passerini, l'attenzione all'intersoggettività è sempre più presente nella storia orale, soprattutto nelle generazioni di storici e storiche più giovani: a titolo di esempio, E. Asquer, *Storia intima dei ceti medi. Una capitale e una periferia nell'Italia del miracolo economico*, Roma-Bari, Laterza, 2011; C. Hughes, *Negotiating Ungovernable Spaces between the Personal and the Political: Oral History and the Left in Post-War Britain*, «Memory studies», 2013, 1; fino allo sguardo psicoanalitico di M. Roper, *Analysing the Analysed: Transference and Counter-Transference in the Oral History Encounter*, «Oral History», 31, 2003.

e anche alcune soddisfazioni, ma sono sicura che un simile lavoro, come succede ad esempio nel laboratorio Triangle di Lyon¹⁸ o nella ricerca integrata di Marta Margotti

e Rachele Raus (di cui si parla in questa rubrica), darebbe – e darà – maggiori risultati quando le diverse competenze saranno messe al lavoro insieme.

¹⁸ www.cnrs.fr/inshs/Lettres-information-INSHS/lettre_infoINSHS_16.pdf (ultimo accesso il 9 febbraio 2015).

