



“Si è sempre qualitativi di qualcuno”: le ricerche quantitative e qualitative in semiotica

Piero Polidoro

Negli ultimi anni assistiamo a un crescente interesse della semiotica per i metodi quantitativi. In realtà questa tendenza si iscrive in un più generale sforzo, da parte di alcuni studiosi, di sondare la compatibilità della nostra disciplina con metodi e presupposti epistemologici simili a quelli adottati dalle cosiddette scienze “dure”, di cui la fisica rappresenta certamente l’esempio più citato (ma non necessariamente più conosciuto). Un percorso simile è stato intrapreso, già da molti decenni, da altre scienze umane e sociali, come la psicologia e l’economia. Questo fenomeno più generale non riguarda solo l’adozione di metodi quantitativi, ma anche quella di metodi sperimentali. Vale la pena ricordare che, anche non avendo sviluppato un approccio quantitativo e sperimentale, la semiotica è comunque una disciplina empirica, perché le sue teorie nascono e vengono continuamente messe alla prova nel confronto con i testi. D’altronde il principio empirico era uno dei fondamenti del pensiero di Hjelmslev¹ ed è stato notoriamente fatto proprio anche da Greimas (vedi, per esempio, la voce “empirisme” del *Dizionario*²). Volendo fornire una bibliografia non esaustiva, dovremmo cominciare da quello che è il testo più spesso citato (almeno in ambito italiano)

quando si parla di questi temi: si tratta di una breve appendice, di meno di due pagine, a *Lector in fabula* di Umberto Eco³. L'appendice (che non è presente nella versione sintetica inglese *The Role of the Reader*) è intitolata "Il lettore modello di 'Drame': un test empirico". Si tratta di un tentativo (nelle parole dello stesso Eco, «di formato modesto») di verificare attraverso i riassunti scritti da due gruppi di soggetti la comprensione di un brano ("Un drame bien parisien" di Alphonse Allais) e la sua corrispondenza con l'interpretazione critica data dal semiologo. Tentativi simili per molto tempo non ci sono stati (o almeno non hanno avuto la stessa notorietà).

Negli ultimi anni, per quanto riguarda il panorama italiano, il numero 122 della rivista *Versus*, curato da Guido Ferraro e Aldo Nemesio⁴, è stato dedicato all'uso dei metodi empirici nella ricerca semiotica. Nel numero successivo, il 123, è stato pubblicato un articolo di Giovanna Cosenza, Jennifer Colombari ed Elisa Gasparri⁵ in cui si propone un'analisi quali-quantitativa su un corpus molto vasto di campagne pubblicitarie composto da circa 7.000 testi (annunci stampa, spot, affissioni, radiocomunicati, ecc.): in una prima fase erano stati definiti alcuni stereotipi pubblicitari dell'uomo e della donna; nella fase successiva ogni pubblicità veniva assegnata a uno di questi stereotipi, in modo da poter calcolare il peso di ogni stereotipo nel corpus preso in considerazione. Un metodo simile (anche se mirato al rilevamento non di stereotipi, ma di stili pubblicitari) era stato applicato da Fabiana Savini⁶ in un lavoro sulla pubblicità per giovani che era stato anch'esso compiuto su un corpus molto vasto. Anche se esterni al campo semiotico, vanno ricordati anche i lavori sul cosiddetto *distant reading* di Franco Moretti⁷.

L'applicazione del metodo sperimentale in semiotica è stata poi affrontata in un volume curato da Bruno Galantucci e Simon Garrod⁸. Mentre scrivevo queste pagine ho avuto modo di vedere la raccolta dedicata a *Quantitative Semiotic Analysis*, curata da Dario Compagno⁹, alla cui introduzione rimando per una discussione più approfondita di molti dei temi che toccherò in queste poche pagine. È inoltre in preparazione, a cura di Dario Compagno e di Matteo Trealani, un numero speciale della rivista *Semiotica* sui metodi quantitativi in semiotica. Sempre in

quest'ambito, infine, avevo avanzato alcune proposte sui metodi di valutazione degli effetti semiotici in un articolo di qualche anno fa¹⁰.

Questa breve, e incompleta, rassegna bibliografica mostra bene come ci sia un crescente interesse verso i metodi quantitativi e sperimentali. I motivi di questa tendenza sono molteplici. Sicuramente il modello delle scienze dure è quello che, nella nostra società, incarna maggiormente l'idea diffusa di scienza. Nel dibattito pubblico, quando si parla di scienza si pensa appunto a discipline che funzionano o dovrebbero funzionare come la fisica (dove "funzionare" è usato – con una voluta semplificazione – come sinonimo di "adottare metodi e presupposti epistemologici"). I metodi sperimentali (fra quelli empirici) e quantitativi offrirebbero una maggiore verificabilità e, in generale, una maggiore rigore scientifico. In tal senso si delineerebbe una contrapposizione, forse troppo brutale, fra una semiotica tradizionale, troppo legata a metodi quali l'introspezione e a un'impostazione sostanzialmente filosofica, e una semiotica ancora da costruire, che deve guardare al quantitativo e allo sperimentale per potersi reinventare. François Rastier, per esempio, scrive:

The role of sciences is to strip philosophy of its objects; in short, to abandon the speculative world and enter the empirical one. As natural philosophies have been stripped by physics and the other natural sciences, philosophy of language has been rejected by linguistics, with which it does not share objects or aims.

As long as semiotics does not clarify these points, it will remain a philosophy of signs, certainly stimulating, but not a science capable of producing new observables and of rejecting hypotheses, as appealing as this may appear¹¹.

Da una parte si potrebbe pensare che questo approccio trascuri le evidenti differenze fra i vari oggetti di studio e quelle fra le premesse epistemologiche che, di conseguenza, ogni disciplina deve avere. Inoltre, metodi quantitativi applicati alla semiotica sembrano essere adatti alla verifica di ipotesi interpretative (come nel caso del test svolto da Eco) o utili a mappare il peso, la presenza e la diffusione di determinate linee interpretative o di certe unità culturali. Sembrano però meno adatti a scorgere le possibilità ancora non realizzate, ma presenti – almeno vir-

tualmente – nel sistema; non dimentichiamo che per Hjelmslev la *norma* comprendeva non solo le concatenazioni del piano dell'espressione o del contenuto effettivamente attestate (che componevano invece l'*uso*), ma anche quelle non esistenti che potevano essere comunque “calcolate” a partire dalle regole del sistema¹²: unità virtuali che sarebbe difficile rilevare con un'indagine quantitativa, ma che rappresentavano, per il linguista danese, un possibile oggetto d'indagine. Potremmo anche richiamare il caso del quadrato semiotico di Greimas, dove il fatto che la lessicalizzazione in una certa lingua possa non coprire tutti e quattro i vertici non è un problema, ma indica anzi possibilità semantiche non ancora esplorate da quella cultura.

Dall'altra parte i grandi successi dei metodi quantitativi esercitano oggi un potere d'attrazione innegabile, come – più di un secolo fa – fu per il metodo comparativo che, a partire dalla linguistica e in particolare dalla fonologia, venne progressivamente applicato a vari campi del sapere umanistico.

La discussione di questa contrapposizione non è oggetto di questo articolo, che avrà come suo punto di partenza la convinzione che, qualunque posizione si finisca per avere, vale comunque la pena di analizzare l'opportunità di applicare metodi quantitativi alla semiotica. Dato lo spazio limitato a disposizione, l'unico scopo delle pagine che seguono sarà quello di chiarire cosa effettivamente si intenda quando si parla di “quantitativo” e di “qualitativo”, nella convinzione che spesso ci sia ambiguità nell'uso di questi termini e che questo comprometta ogni possibilità di una discussione epistemologica e metodologica realmente utile. La prima domanda che dobbiamo porci è quale sia l'essenza del metodo quantitativo. Riprendendo la tradizione aristotelica, potremmo identificarlo nella possibilità di una misurazione. La misurazione può essere, secondo Aristotele (*Metafisica*, V 13), di una grandezza (misurazione propriamente detta) oppure di una molteplicità (numerazione). Ai nostri fini, è evidente che questa misurazione debba sempre essere espressa numericamente, altrimenti verrebbe meno la possibilità da parte dell'analisi quantitativa di sottoporre a elaborazione matematica il risultato di questa misurazione.

Immaginiamo, per esempio, che io voglia studiare l'andamento delle temperature in un dato punto di una città. Dovrò posizionare in quel punto uno strumento di misurazione, per esempio un termometro digitale. Questo strumento traduce un certo stato di cose (che a livello microscopico consiste nel moto di particelle e a livello macroscopico emerge sotto forma di quel fenomeno che indichiamo con il termine "temperatura") in un numero, anzi in una coppia di numeri, che indicano la temperatura T maiuscolo e il tempo t nella quale è stata rilevata. Queste coppie di numeri potrebbero essere trasmesse a intervalli regolari a un computer, dove verrebbero immagazzinate come una serie di dati. A sua volta questa serie sarebbe suscettibile di elaborazioni matematiche (probabilmente statistiche: media, mediana, ecc.), più o meno complesse, più o meno modellizzate. È da notare che, una volta scelto il punto da monitorare e installato il termometro, non è più richiesto, almeno in linea di principio, l'intervento del ricercatore. Il rilevamento della temperatura avverrà in modo automatico, sulla base di una serie di processi fisici e chimici (reazione dei sensori, funzionamento dei circuiti). Se questo è un chiaro esempio di metodo quantitativo, quale potrebbe essere un esempio di metodo qualitativo? A mio modo di vedere (e riprendendo anche in questo l'impostazione aristotelica) è il riconoscimento di differenze, di soglie che ci permettono distinzioni, cioè quell'attività molto semiotica che solitamente indichiamo con il termine *categorizzazione*. Il che non esclude, come è fin troppo ovvio, il rigore o la formalizzazione. Per restare in ambito semiotico, consideriamo ad esempio l'analisi componenziale fatta da Eco della parola "invece"¹³. A partire da alcuni esempi di uso del termine, Eco costruisce un modello, un insieme di regole formali che spiegano l'uso e l'interpretazione di questa parola nella lingua italiana. Non ci sono misurazioni o numeri, ma un modello formale e falsificabile che permette di compiere delle scelte sulla base del riconoscimento di situazioni diverse (vale a dire le selezioni contestuali, la cui descrizione è inglobata nella definizione semantica del semema).

Il problema, però, è che non è possibile tracciare una linea così netta fra metodi quantitativi e metodi qualitativi. La misurazione non è l'unico

fattore in gioco. Bisogna prendere in considerazione anche l'origine di questa misurazione, cioè il modo in cui viene costruito il dato numerico che verrà elaborato matematicamente. Nell'esempio della temperatura, come abbiamo visto, il dato deriva dal funzionamento automatico di un dispositivo di misurazione. La teoria alla base di questo dispositivo potrebbe essere contestata e superata, oppure potrebbero essere inventati nuovi modi di implementarla; si potrebbero contestare margini di errore troppo alti o altri limiti della procedura. Ma è chiaro che, una volta impostato e accettato il sistema, questo comincerà a produrre dati senza la necessità di altri interventi (e dunque senza che questi possano influenzare il dato stesso). Quelle che, soprattutto in ambito sociale e umanistico, chiamiamo ricerche quantitative non funzionano sempre così. Prendiamo l'analisi del contenuto. In alcuni casi può assomigliare al caso del rilevamento della temperatura. È quanto accade quando in un testo rileviamo la frequenza di occorrenza di certi termini, o le co-occorrenze. Un software analizza il testo, conta occorrenze e co-occorrenze, e restituisce dati che possono servire da stimolo o da verifica di certe ipotesi interpretative. Ma consideriamo cosa accade nel caso di un'analisi tematica. Qui è necessario un lavoro preliminare di codifica del testo, che consiste nel taggarlo, cioè nell'assegnare un certo tag, indicativo di un tema, a una porzione del testo. Procedure simili possono essere seguite anche per rilevare altre caratteristiche del testo. Il tagging, in realtà, non è altro che un'operazione di categorizzazione. Infatti è una categorizzazione quella compiuta da Cosenza *et al.* nell'articolo che abbiamo già citato: hanno creato un sistema di codifica o – se vogliamo – di categorizzazione (l'identificazione e descrizione degli stereotipi) e lo hanno applicato a un corpus di pubblicità, assegnando a ognuna di esse il tag corrispondente allo stereotipo che era stato identificato. E un approccio simile è quello alla base delle ricerche sul *distant reading*. Se infatti un ricercatore dice “Il romanzo X appartiene al genere epistolare” sta certamente esprimendo un giudizio qualitativo, risultato finale di una serie di considerazioni. Non cambia molto, dal punto di vista metodologico, se il ricercatore dice invece “Il romanzo X_1 appartiene al genere epistolare, il romanzo X_2 appartiene al genere epistolare, il

romanzo X_3 appartiene al genere gotico...” e alla fine conta quanti romanzi epistolari, gotici e storici ci sono nel suo corpus per tracciare un grafico dell’andamento dei vari generi in un certo arco di tempo. Anche in questo caso ha formato qualitativamente un dato, che ha poi elaborato quantitativamente.

L’operazione di codifica, di tagging, è però manuale. Il ricercatore (o meglio, di solito una coppia di ricercatori che operano autonomamente, in modo da non influenzarsi a vicenda) legge il testo, lo interpreta e, quando è il caso, assegna uno dei tag previsti dalla sua ipotesi a una certa porzione di testo. È quindi il giudizio del ricercatore a costruire il dato (l’assegnazione del tag a una porzione di testo), che sarà poi contato, operazione preliminare all’elaborazione matematica. Questo tipo di analisi viene solitamente considerata quali-quantitativa, perché è evidente che non è quantitativa nello stesso senso in cui lo è il rilevamento della temperatura. Se l’elaborazione e l’output finali sono quantitativi, la costruzione originaria del dato deriva da un giudizio singolare, certo formalizzabile, ma solo in parte e quindi qualitativo. La differenza fra costruzione del dato nell’esempio del termometro e in una ricerca qualitativa è quindi la differenza fra una procedura totalmente automatizzabile e una in cui si deve lasciare almeno una parte del compito a una decisione non riducibile a un algoritmo. È questa, vale la pena ricordarlo, l’opposizione introdotta tempo fa da Eco fra relazioni cieche A-B (fra cui, per esempio, il meccanismo comportamentista stimolo-risposta) e il cosiddetto spazio C, cioè lo spazio della scelta, che non è altro che lo spazio dell’interpretazione¹⁴. La ricerca quali-quantitativa, dunque, si riconosce per la natura interpretativa della costruzione del dato, mentre in quella puramente quantitativa il dato dovrebbe emergere da una procedura completamente automatizzata (indifferentemente dal fatto che venga condotta da uno strumento di misurazione, da un’intelligenza artificiale oppure da un essere umano).

L’asse categoriale qualitativo/quantitativo, dunque, diventa lo spazio di un continuum. E per collocare una ricerca su questo continuum dobbiamo innanzitutto prendere in considerazione due fattori: 1) la presenza o meno di una misurazione che consente la formazione di dati

numerici e la loro successiva elaborazione, 2) la natura deterministica (cieca e automatica) oppure interpretativa della costruzione del dato. La semiotica sembra maggiormente (ma non esclusivamente) rivolta a una costruzione interpretativa del dato, che non implica automaticamente una sua aleatorietà. Il lavoro che va fatto, quindi, se si vuole rendere realmente credibile un approccio quali-quantitativo della semiotica è quello di rendere il più rigoroso e controllato possibile questo processo interpretativo dell'analista. In che modo? Per esempio attraverso procedure di segmentazione e analisi definite in modo sistematico e dettagliato, la definizione precisa e operativa dei concetti impiegati nell'analisi, l'adozione di tecniche di analisi incrociata (comuni nell'analisi del contenuto, in cui di solito un singolo testo viene trattato da due o più giudici che confrontano poi i risultati del loro lavoro).

Note

158

- 1 L. HJELMSLEV, *Omkring sprogteoriens grundlæggelse*, København, Munksgaard kowski, 1943 (trad. it. *I fondamenti della teoria del linguaggio*, Torino, Einaudi, 1968).
- 2 A.J. GREIMAS e J. COURTÈS, *Sémiotique. Dictionnaire raisonné de la théorie du langage*, Paris, Hachette, 1979.
- 3 U. ECO, *Lector in fabula*, Milano, Bompiani, 1979.
- 4 G. FERRARO e A. NEMESIO, *Empirical Research in Semiotics*, numero monografico di «Versus», n. 122, 2016.
- 5 G. COSENZA, J. COLOMBARI e E. GASPARRI, *Come la pubblicità italiana rappresenta le donne e gli uomini*, in «Versus», n. 123, pp. 323-361.
- 6 F. SAVINI, *Giovani nelle pubblicità per giovani*, Padova, Libreria Universitaria, 2014.
- 7 F. MORETTI, *Graphs Maps Trees. Abstract Models for a Literary Theory*, London-New York, Verso, 2005; F. MORETTI, *Distant Reading*, London-New York, Verso, 2013.
- 8 B. GALANTUCCI e S. GARROD (a cura di), *Experimental Semiotics*, Amsterdam-Philadelphia, John Benjamins, 2012.
- 9 D. COMPAGNO (a cura di), *Quantitative Semiotic Analysis*, Berlin, Springer, 2018.
- 10 P. POLIDORO, *Alcune riflessioni sui metodi di verifica degli effetti semiotici*, in «Rivista italiana di Filosofia del linguaggio», Special Issue – Italian Society of Philosophy of Language, pp. 206-218.

- 11 F. RASTIER, *Computer-Assisted Interpretation of Semiotic Corpora*, in D. Compagno (a cura di), *Quantitative Semiotic Analysis*, Berlin, Springer, 2018, p. 125.
- 12 L. HJELMSLEV, *Saggi linguistici I*, Milano, Unicopli, 1968. Su questi argomenti vedi anche P. POLIDORO, *Il concetto di sostanza e di significato in Hjelmslev*, in A. Gałkowski e T. Roszak (a cura di), *Semiotica generale-semiotica specifica*, Wydawnictwo Uniwersytetu Łódzkiego, 2018, pp. 49-74.
- 13 U. ECO, *Lector in fabula*, op. cit.
- 14 U. ECO, *I limiti dell'interpretazione*, Milano, Bompiani, 1990.