

## TENDENZE

Per 4 anni un'équipe di specialisti americani ha setacciato 500 miliardi di parole, tratte da 5,2 milioni

di libri (il 4% di quelli stampati in tutta la storia), a caccia del «genoma della letteratura globale»

È nata così la «culturomica», disciplina che studia le frequenze verbali e le loro variazioni



# Lingua

## Alla ricerca del «gene della cultura»

di Giuseppe O. Longo

**C**inquecento miliardi di parole, tratte da 5,2 milioni di libri (circa il 4% di tutti i libri mai stampati, che sono circa 129 milioni): è questo il corpus sterminato dal quale, mediante un'analisi quantitativa eseguita con l'ausilio del calcolatore, sono state ricavate moltissime informazioni di carattere linguistico e storico-culturale in genere. All'impresa si sono dedicati per 4 anni numerosi specialisti, capeggiati da due ricercatori dell'Università di Harvard, Jean-Baptiste Michel e Erez Lieberman Aiden, i quali hanno identificato quello che, alquanto pomposamente, è stato chiamato «il genoma della cultura universale» o *culturoma*. La banca dati è stata costruita sulla base di un terzo dei 15 milioni di libri scansionati finora da Google e finiti nella sua biblioteca digitale. I libri selezionati da Michel e colleghi appartengono all'Ottocento, al Novecento e ai primi anni del nostro secolo. L'analisi riguarda in sostanza le frequenze delle parole e la loro variazione nel tempo, e consente di studiare in termini quantitativi le tendenze linguistiche e culturali. L'avvento di questa nuova disciplina,



NOAM CHOMSKY

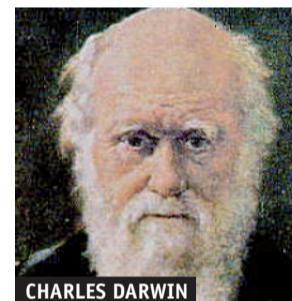
**Negli ultimi 50 anni il lessico è aumentato del 70%. Peccato però che l'analisi riguardi in misura preponderante il solito inglese; non è stata esaminata nessuna opera italiana (ma nemmeno giapponese o araba...)**

battezzata *culturomica*, che si colloca a metà tra la lessicografia, l'informatica e la storia, è stata accolta con entusiasmo durante l'ultimo congresso dell'Associazione Storica Americana, tenutosi a Boston l'8 maggio scorso. Soprattutto i linguisti potrebbero trarne utili indicazioni di carattere storico-evolutivo: quando è stata introdotta una certa parola, quando il suo uso è divenuto comune e magari quando è scomparsa. I risultati, comunque, riguardano in misura preponderante la lingua inglese: infatti dei 500 miliardi di parole, 361 sono inglesi, 45 francesi e altrettante spagnole, 37 tedesche e 37 russe, 13 cinesi e 2

ebraiche. Non è stato esaminato nessun libro italiano, nessun libro giapponese, nessun libro arabo.

**LA MIGRAZIONE DEI VERBI**  
Ad ogni modo, per quanto riguarda l'inglese si sono ottenuti alcuni risultati interessanti relativi al lessico e alla grammatica. Emblematico è il caso dei verbi. Come in molte altre lingue, anche in inglese esistono verbi "regolari" e verbi "irregolari": i primi formano il passato e il participio passato aggiungendo alla radice la desinenza *-ed* (per esempio *to love*, amare, fornisce *loved*, *loved*); invece i verbi irregolari si comportano in modo piuttosto indisciplinato (per esempio *to find*, trovare, fornisce *found*, *found*, oppure *to drive*, guidare, fornisce *drove*, *driven*). Tuttavia ogni verbo irregolare presenta, accanto alle forme devianti, anche le forme canoniche (*drived* e *drived* accanto a *drove* e *driven*) le quali esercitano una forte attrazione sulle prime, tanto che alcuni irregolari hanno progressivamente abbandonato

l'anomalia per rientrare a poco a poco nei ranghi. La rapidità della migrazione verso la forma regolare è stata più o meno accentuata a seconda dei casi: in genere i verbi più comuni sono i più refrattari al processo di regolarizzazione (per esempio *to come*, venire, *came*, *come*; oppure *to go*, andare, *went*, *gone*, non manifestano alcuna intenzione di conformarsi). In questo contesto la culturomica ha fornito informazioni quantitative importanti, per esempio ha indicato che se negli ultimi 200 anni gran parte dei verbi irregolari sono rimasti stabili, un 16% ha cominciato ad accogliere le forme regolari. L'omologazione è stata più rapida negli Stati Uniti che nel Regno Unito, e ciò non stupisce, visto il carattere conservatore del popolo britannico. Ma le forme regolari si fanno largo anche in Inghilterra: per esempio ogni anno circa 130.000 persone adottano la forma regolare *burned* (bruciato) invece di quella irregolare *burnt*.

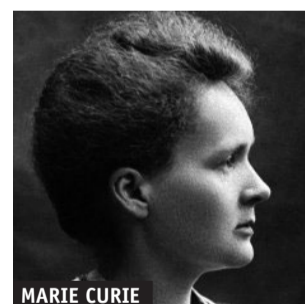


CHARLES DARWIN

**Di grande interesse la tendenza degli idiomi a conformarsi sulle forme regolari, cancellando le eccezioni. Ma oltre la metà dei vocaboli è una sorta di «materia oscura» che non compare nemmeno nei dizionari**

**LA RICCHEZZA DEL LESSICO**  
Certo, per il grande pubblico si tratta forse di curiosità, ma per gli storici della lingua queste indagini sull'evoluzione delle forme verbali sono importanti. Altrettanto importanti sono i risultati forniti dalla culturomica sull'ampiezza del lessico, cioè sul numero di parole appartenenti alla lingua (inglese, come al solito). Si sono ottenute queste stime: secondo il catalogo di Michel e Aiden le parole inglesi erano 544.000 nel 1900 e 597.000 nel 1950, mentre nel 2000 erano aumentate fino a 1.022.000. La ricchezza del lessico è cresciuta enormemente, a un tasso di circa

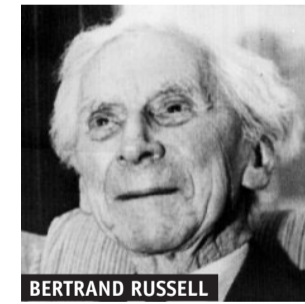
8500 parole all'anno, tanto che negli ultimi cinquant'anni l'incremento è stato di oltre il 70%. Per contro il numero delle parole registrate nei dizionari è molto minore: nel Webster (edizione del 2000), che è uno dei più ampi e diffusi, se ne trovano soltanto 348.000. Questa discrepanza si spiega in parte con l'assenza nei dizionari dei nomi propri e di molti nomi composti, ma dipende soprattutto dalla circostanza che un vocabolario deve raggiungere un equilibrio ragionevole tra la ricchezza e la concisione: dev'essere abbastanza esteso da essere utile all'utente medio e abbastanza stringato da consentirne la stampa e l'uso maneggevole. Perciò molte parole rare sono omesse. I metodi statistici della culturomica hanno consentito di stimare che il 52% delle parole che si trovano nei libri inglesi non si trovano nei dizionari. Insomma il repertorio inglese consiste per oltre metà in una sorta di "materia oscura" lessicale, costituita da quelle parole che entrano ed escono rapidamente dall'uso per ragioni di moda o per la volatilità creativa dei letterati o degli specialisti.



MARIE CURIE



EREZ LIEBERMAN E JEAN-BAPTISTE MICHEL (IN PRIMO PIANO)



BERTRAND RUSSELL

### LE APPLICAZIONI

#### Ma poi migliora la conoscenza?

**S**e la linguistica può trarre beneficio dalla culturomica, lasciano un pochino perplessi le sue possibili applicazioni ai campi della storia, della filosofia e della cultura in genere. Sapere di quanto è aumentata la frequenza del termine «nazista» durante la seconda guerra mondiale ci dice qualcosa che non sapessimo già? Sapere di quanto è cresciuto l'uso della parola «donna» dopo l'avvento del femminismo incrementa davvero la nostra comprensione del fenomeno? È vero che prima ne avevamo una nozione imprecisa, impressionistica, qualitativa, istintiva e ora ne possediamo una conoscenza quantitativa. Ma che cosa aggiunge davvero la quantità alla conoscenza? In certi casi aggiunge molto, in altri niente. Non tutto ciò che si può misurare è importante e non tutto ciò che è importante si può misurare. La culturomica rappresenta un esempio notevole di «delega tecnologica»: per decenni se non per secoli le analisi e le sintesi culturali sono state compiute dagli esseri umani. Oggi l'analisi è affidata alle macchine, ma la sintesi dev'essere fatta ancora dagli uomini, perché siamo noi a possedere i criteri di giudizio e gli strumenti di valutazione per interpretare e inquadrare i dati e per trasformare la loro massa caotica in cultura organica e sistemata. Inoltre l'affermazione di Michel e Aiden (e dei loro numerosi collaboratori) secondo la quale «la culturomica estende i confini di una rigorosa indagine quantitativa fino a comprendere la vasta gamma dei nuovi fenomeni che permeano le scienze sociali e umanistiche» deve fare i conti con l'assoluta preponderanza della cultura inglese nel campione considerato. Il quadro rischia di essere un tantino parziale. (G.O.L.)

### GLI SCIENZIATI FAMOSI

#### E Russell «vale» 1500 milliDarwin

**I**metodi quantitativi della culturomica sono stati applicati anche al calcolo della fama degli scienziati degli ultimi due secoli, sulla base non del parere (pur sempre soggettivo) degli esperti bensì del numero di ricorrenze del loro nome. Si è così scoperto che lo scienziato di gran lunga più celebrato è Charles Darwin, il cui nome compare quasi 150 mila volte tra il 1839 e il 2000. Ciò ha consentito di creare un'unità, il «darwin», con cui misurare la notorietà degli altri scienziati. In pratica, siccome il darwin è molto grande, si usa il millidarwin, mD. Si scopre così che Marie Curie ha 189 mD, Noam Chomsky 507, Bertrand Russell 1500, Albert Einstein 878, Isaac Asimov 183, Richard Dawkins 90. L'analisi riserva molte sorprese: per esempio vincere un premio Nobel non comporta necessariamente una valutazione elevata; Chomsky ritiene che il proprio quoziente sia dovuto più all'impegno politico e alle denunce subite che alla fama scientifica; per Asimov vale un discorso analogo: la sua fama di scrittore di fantascienza prevale di gran lunga sui meriti di biochimico. Per salire nella scala dei mD è molto meglio essere sociologo o psicologo che chimico, fisico o matematico (questi ultimi sono quasi ignorati). L'unico matematico con molti mD è Russell, che secondo alcuni maligni è famoso per non aver fatto matematica... Un libro divulgativo di successo alza molto il punteggio (come dimostrano i casi di Dawkins, Carl Sagan e Rachel Carson). Ma più di tutto conviene buttarsi nelle polemiche, specie se violente e dissacranti, urlare in Tv, e in genere comportarsi in modo stravagante. Se ne può concludere che la culturomica non sa distinguere tra fama e infamia. (G.O.L.)