

L'agone per la Webgloria

GoogleFight, un metamotore contrastivo per l'antropologia della Rete

Massimo Manca

Immaginate questa scena – è così scolpita nell'immaginario che davvero non vi sarà difficile. Di fronte a un pubblico trepidante per l'incerto esito, ma in fondo già consapevole della decisione ultima del Fato, scendono in campo gli eroi Achille ed Ettore. Ettore si difende piuttosto bene, in modo da pagare il suo tributo alle ragioni del *kleos*, ma alla fine è costretto a piegarsi al volere della Moira, stabilita per lui fin dalla fondazione del tempo, e a soccombere. Dove si svolge questa scena? Facile, direte voi: geograficamente, siamo a Troia; letterariamente, siamo nel ventiduesimo libro dell'*Iliade*. Sbagliato: siamo su *Google*.

Come gli umanisti che vagavano per l'Europa a caccia di manoscritti, anche i neofilologi della Rete hanno i loro problemi di *recensio*. Si calcola che il *World Wide Web* indicizzabile contenga oggi circa 10-15 miliardi di pagine, in progressione, almeno per ora, geometrica. Ciò, come noto, pone il grande problema del reperimento delle informazioni e fa sì che chi detiene l'accesso alla loro selezione sia oggi più importante dello stesso autore: l'archivista conta più del libro (Borges direbbe che è sempre stato così); se oggi il Senato romano dovesse decretare la *damnatio memoriae* di un imperatore, invece di scalpellarne via il nome dagli archi di trionfo chiederebbe agli *staff* dei *search-engine* la rimozione dall'archivio, perché ciò che non esiste sui motori di ricerca non esiste *tout-court*. Questi strumenti più o meno automatizzati, concepiti per orientarsi nel *mare magnum* dell'informazione, nel corso degli anni sono stati tutti più o meno tutti soppiantati da *Googlez*. Non è questa la sede per parlarne diffusamente³, ma in effetti, il fatto che *Google* sia *cool* e ormai sempre più pervasivo ha fatto sì che la maggior parte dei motori di nuova concezione vi si ispiri direttamente (un esempio adatto agli umanisti: *Cruscle*, il motore di ricerca dell'Accademia della Crusca) o vi si appoggi come motore primario sfruttandone alcune caratteristiche o rifinendone la visualizzazione, fino ad arrivare a veri e propri *gadget* come *Berlusgoogle*, motore specializzato nelle ricerche su Silvio Berlusconi. Uno di questi motori 'scherzosi', *GoogleFight*, offre spazio per qualche riflessione di interesse per un filologo.



GoogleFight è un doppio metamotores di ricerca. Se si desidera controllare le occorrenze di due stringhe di testo, con i motori consueti occorre lanciare due ricerche successive. *GoogleFight* permette invece di lanciare due ricerche simultanee su *Google*, presentando la collazione in forma grafica. Il processo di ricerca è molto coreografico, accompagnato com'è da un'animazione in cui due omini combattono fra loro⁴.

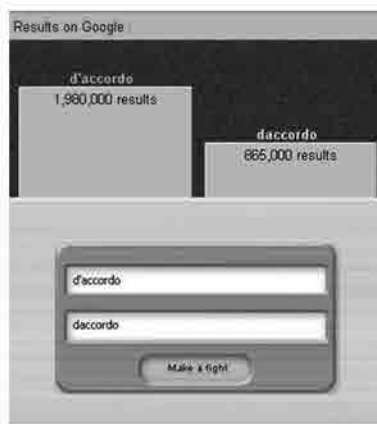
Al termine della ricerca, viene visualizzata la stringa "vincitrice", che presenta cioè il maggior numero di occorrenze. Per esempio, se si cerca "Luminar" contro "Internet e umanesimo" al termine si ottengono circa 400.000 risultati per la prima stringa contro circa 17.000 della seconda. Diamo qui una schermata d'insieme della pagina principale:

La presentazione dei risultati in forma non solo numerica, ma di istogramma, permette un'agevole comparazione degli ordini di grandezza. Benché *GoogleFight* sia nato come *variatio* ludica sul tema di *Google*, questo suo aspetto non secondario non deve impedire di scorgervi una certa utilità nel tentativo sempre approssimato di ricercare un *kosmos* nel *chaos* del Web. L'idea giocosa originale alla base del progetto è il confronto di



stringhe semanticamente differenti: Dio contro Satana, Robbie Williams contro Britney Spears, Pepsi contro Coca Cola, Cicerone contro Catilina, Bizantini contro Turchi, etc. I *GoogleFighter* sono sempre alla ricerca di match originali, che creino un “effetto di senso” fino a ottenere l’onore di essere inseriti nel *tab* “classics”, ed è inoltre possibile seguire *in fieri* le richieste che vengono proposte al metamotores, che tiene in memoria le ultime venti *query* e diversi confronti preselezionati.

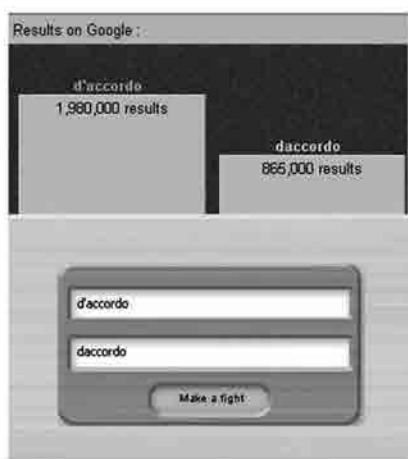
Fin qui, come nello spirito di *GoogleFight*, abbiamo giocato. Al filologo, tuttavia, potrebbe interessare, accantonata la “battaglia semantica”, l’esame delle varianti grafiche. Il *Web*, essendo una forma scritta estremamente eterogenea, che contiene le lingue e i registri linguistici più disparati, è un serbatoio preziosissimo per chi voglia rendersi conto dello *status* linguistico attuale o voglia seguirne gli sviluppi. Nel momento in cui scrivo, è il nove dicembre 2006. Ho appena lanciato su *GoogleFight* la ricerca contrastiva di “d’accordo”, con l’apostrofo, contro “daccordo”, tutto attaccato. La seconda forma è errata, ma una certa esperienza di insegnante liceale (e le cose non cambiano troppo all’università) mi dice che sta soppiantando la prima. Ecco il responso di *GoogleFight*: vince “d’accordo” con un po’ più del doppio delle occorrenze; guardando però il bicchiere mezzo vuoto, possiamo dire che, in Rete, ogni due-tre volte, “d’accordo” segue la *scriptio continua*; se consideriamo che nel *Web* non esistono solo siti di carattere amatoriale, ma anche di carattere accademico o comunque professionale, come quello che ospita questo lavoro, in cui esistono *referee* ed *editor* e dunque l’ortografia è ben controllata, possiamo immaginare che, al di fuori di queste aree protette, l’italiano dell’uso abbia ormai preso l’abbrivio verso la crasi. Invito chi mi legge – nel frattempo sarà passato del tempo – a provare a rifare l’esperimento (o a provare d’altronde contro



daltronde, “mah” contro “mha”, “beh” contro “bhe”, “aut aut” contro “out out”. Di qui a dieci anni è facile ipotizzare che, almeno per il caso che ho esaminato, la colonna di destra avrà superato quella di sinistra (a dire il vero “out out” stravinca già, non per semplice insipienza degli scriventi, che pure esiste ed è ben attestata nei quotidiani, ma perché è il nome di un gruppo *techno*).

Non necessariamente la variabilità coinvolge evidenti (per ora) errori di ortografia. Si può per esempio cercare “pronuncia” vs. “pronunzia”. Il grafico conferma chiaramente la sensazione istintiva che la seconda opzione sia ormai *lectio difficilior* riservata a una scrittura ‘intenzionale’, affettata o comunque percepita come arcaizzante. Ancora più fine, e un buon *divertissement* da docente universitario, è cercare la “*variatio* accademica”, cioè allografi che individuano spesso diverse “scuole”; Teodorico vs. Teoderico, Ulfila vs. Vulfila (e Wulfila; purtroppo *GoogleFight* non permette confronti a tre), Aristotele vs. Aristotile, etc. È un peccato che la lingua italiana non segni gli accenti, perché sarebbe davvero interessante iscrivere al *GoogleFight Club* Èdipo vs. Edipo, Giàsone vs. Giasòne, Capanèo vs. Capàneo vs. Capàneo...

La filologia, come è noto, presenta sempre dei pericoli, il computer anche, e prima di fare convolare a nozze l'uno e l'altra è opportuno stipulare patti prematrimoniali ben precisi. Il problema di strumenti come *GoogleFight* è che è necessaria una accurata pianificazione delle ricerche per evitare l'anfibolia sempre in agguato e ogni sorta di interferenza nei risultati. Per esempio, una ricerca classica su Internet, scherzosa, ma anche antropologicamente piuttosto interessante, è “sex” vs “religion”, che in questo



momento mi dà il seguente esito: circa 400.000 attestazioni per “sex” e la metà per “religion”. Provo a lanciare la stessa ricerca, ma in italiano, “sesso” vs. “religione”, ottenendo dati differenti (“sesso” circa 2.000.000 e “religione” ben undici milioni; lascio eventuali deduzioni maliziose al lettore). Provo poi a lanciare la stessa ricerca in francese, “sexe” vs. “religion”, ottenendo il risultato un po’ controintuitivo che la laica e sensuale Francia pare essere sul Web piuttosto ascetica e pia. In realtà, *GoogleFight* non è sensibile all’accento, e dunque ingloba in “religion” i risultati di tutte le lingue in cui “religione” si scrive con questa sequenza grafica, compreso l’inglese, che in Internet pesa moltissimo; ma anche la prima ricerca, quella in inglese, è in realtà falsata dal medesimo problema, per cui, in realtà, “sex” sopravanza “religion” ben di più di quanto non appaia. *GoogleFight*, che non è in grado di disambiguare, non è lo strumento adatto per questa ricerca, e può essere usato solo quando i significanti non siano soggetti a confusione. Se ciò non è possibile, occorre tornare all’uso di *Google*, che permette di filtrare i risultati limitandoli a determinate lingue. E così, in francese, *sexe* compare all’incirca 5.000.000 di volte, *religion* circa 1.500.000, *et voilà, l’honneur de la Patrie est sauvé*.

Si noti ancora che una semplice comparazione lessicale interlinguistica può portare assai facilmente a conclusioni errate, poiché termini analoghi in lingue diverse, anche se sono effettivamente corrispondenti e non ‘falsi amici’ come “morbido” vs. *morbid* (“malsano” < *morbus*), difficilmente coprono la medesima estensione semantica. Per stare agli esempi precedenti, “sesso” in italiano ha una connotazione più neutra, laddove l’inglese *sex* tende a restringersi alla sfera strettamente erotica e lasciare il passo a *gender* in campo sociologico; “religione” corrisponde certo a *religion*, che però è affiancato a ‘quasi sinonimi connotativi’ come *cult* o *church* in modo piuttosto differente rispetto alla relazione che in italiano passa fra “religione”, “culto”, e “chiesa”, e il latino *religio* sarebbe ancora un’altra cosa. Occorre dire che tutte queste cautele ‘a priori’, e molte altre si possono apprendere facilmente ‘a posteriori’ facendo un po’ di esperimenti con il metamatore. Qualche “battaglia contrastiva” è sufficiente a dare un’idea dei problemi che si presentano a chi voglia seriamente affrontare



ricerche di affinità o opposizione semantica o lessicale, e per serendipità, come spesso accade cercando in Rete, ci si imbatte spesso in scoperte interessanti, secondo modalità che prima di Internet sarebbero state inconcepibili.

NOTE

1. La medaglia ha il suo rovescio e, come dice *Spiderman*, che di reti se ne intende, da grandi poteri derivano grandi responsabilità, tanto che *Google* è stato in effetti denunciato per aver permesso la pubblicazione di filmati che documentavano atti di bullismo (e che proprio in conseguenza della pubblicazione su *Google* sono divenuti di dominio pubblico e poi perseguiti); è un po' come dare al termometro la colpa della febbre o, come si usava talora in età prescientifica, curare l'arma al posto del ferito.
2. Difficilmente un navigatore dell'ultima ora avrà sentito nominare *Altavista* (dal 2013 è stato sostituito dal motore di ricerca *Yahoo*), *Lycos*, *HotBot*. I navigatori di più vecchia data si sorprenderanno invece forse di sapere che ancora esistono (*Lycos* si è tristemente ridotto a motore di ricerca per anime gemelle; *Altavista* ha clonato l'interfaccia minimalista di *Google* con poche speranze di poter rosicchiare un po' di share dal rivale, ma mantenendo una certa popolarità fra il pubblico degli *indocti* grazie al suo traduttore automatico *Babelfish*, reso celebre dagli esercizi oulipiani dedicatigli da Eco (*Babelfish* nel 2012 venne assorbito da *Yahoo*). Ha inoltre ispirato onomasticamente gli autori di *Astalavista* (di cui non diamo il *link* diretto perché è opportuno connettersi opportunamente catafratti dal punto di vista della sicurezza informatica), motore di ricerca più o meno legale al servizio di hacker, cracker e phreaker. *Hotbot*, che pareva il più promettente, è stato inglobato da *Lycos*, con cui condivide la *balba senectus*.
3. Sui motori di ricerca, e *Google* in particolare, si veda Elisa Bastianello, *Google it! La ricerca on-line*, "engramma" 46, marzo 2006.
4. Sullo stesso principio opera <http://www.googlebattle.com/>, ma con grafica molto inferiore.