

ACCADEMIA DELLA CRUSCA

L'ITALIANO ELETTRONICO

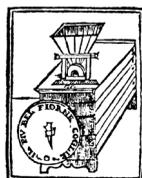
VOCABOLARI, CORPORA,
ARCHIVI TESTUALI E SONORI

Firenze, 6-8 novembre 2014

ATTI

a cura di

Claudio Marazzini e Ludovica Maconi



Firenze
2016

La *Piazza delle Lingue* 2014 è stata realizzata con i fondi dell'Accademia e con il contributo dell'Associazione degli Amici dell'Accademia della Crusca, utilizzato per la stampa degli Atti.



I contributi pubblicati in questo volume sono stati sottoposti a procedura di valutazione anonima affidata a specialisti scelti tra i membri dell'Accademia della Crusca.

Stampato in Italia
da Tipolitografia Contini S.r.l. - Sesto Fiorentino (Firenze)

ISBN 978-88-89369-65-4

LUDOVICA MACONI

RETRODATAZIONI LESSICALI CON GOOGLE LIBRI: OPPORTUNITÀ E INGANNI DELLA RETE

Dalla fine degli anni Novanta, l'informatica è stata applicata con successo alla ricerca di prime attestazioni di parole, permettendo di passare dallo spoglio manuale dei testi, ossia dalla lettura sistematica volta a individuare parole retrodatibili, alla ricerca semiautomatica per mezzo del computer. Nuove date sono state così ricavate dalla consultazione di corpora elettronici, in Cd-rom oppure *on-line*¹. Inestimabile risorsa della Rete, preziosa per la datazione del lessico, sebbene non ideata per questo scopo specifico, è Google Books (interfaccia in italiano "Google Libri"): biblioteca digitale in continua crescita che raccoglie ormai le scansioni di milioni di libri in tutte lingue.

Nato nel 2004 con l'intento di riversare e conservare in Rete la maggior quantità possibile di sapere, rendendolo disponibile a tutti in ogni momento e in ogni parte del mondo, Google Books è stato portato all'attenzione dei linguisti per il suo felice impiego nella ricerca di retrodatazioni da Yorick Gomez Gane, in un articolo del 2008 pubblicato negli «Studi linguistici italiani»². Da allora molti studiosi hanno messo a frutto il suggerimento,

¹ Menziono, a titolo d'esempio, solo la *LIZ* (ora *BIZ*), usata tra l'altro, come noto, per gli aggiornamenti del nuovo *DELI* (cfr. Manlio Cortelazzo, *Premessa alla seconda edizione*, in *Il nuovo Etimologico. DELI - Dizionario Etimologico della Lingua Italiana*, di M. Cortelazzo e P. Zolli, seconda edizione in volume unico, Bologna, Zanichelli, 1999, p. 3), e il catalogo internet del Servizio Bibliotecario Nazionale (SBN), ingegnosamente e utilmente sfruttato da Luigi Matt come banca-dati per retrodatazioni da titoli di pubblicazioni (cfr. Luigi Matt, *Retrodatazioni di tecnicismi da titoli di pubblicazioni*, in «Studi di lessicografia italiana», XXI, 2004, pp. 183-246). Rimando poi, almeno, al saggio di Stefano Telve, *Retrodatazioni di voci onomatopeiche e interietive. Un esempio di applicazione lessicografica degli archivi elettronici*, in «Studi di lessicografia italiana», XIX, 2002, pp. 229-77.

² Cfr. Yorick Gomez Gane, "Google Ricerca Libri" e la linguistica italiana: *vademecum per l'uso di un nuovo strumento di lavoro*, in «Studi linguistici italiani», XXXIV, f. II, 2008, pp. 260-78. Google Books, nato nel 2004 con diverso nome ("Google Print"), fu ribattezzato

usando Google Libri non solo per datare le parole, ma anche per ricostruirne la storia. Rinvio in particolare agli studi di D'Achille e Sgroi³. Ciò non esclude che altri preferiscano il metodo tradizionale dello spoglio del "testo miniera", col quale si ottengono retrodatazioni sicuramente valide, ma in molti casi migliorabili interrogando la Rete⁴. Ferma restando la consapevolezza che anche i risultati di Google potranno essere, presto o tardi, superati (facile supporlo, considerando che si tratta di un corpus aperto), mi sembra oggi più vantaggioso, e spesso di esito migliore, retrodatare appoggiandosi a Internet, soprattutto se la datazione riguarda il lessico moderno dal Seicento in avanti.

col nome attuale nel 2005, dotato di nuova maschera di ricerca e della possibilità di scaricare PDF nel 2006, di visualizzazione completa per i libri fuori dal diritto d'autore nel 2007. Obiettivo del progetto è anche la salvaguardia e valorizzazione, attraverso il digitale, dei testi rari e di pregio. Google Books conta oggi milioni di digitalizzazioni (nelle diverse lingue; i libri italiani, secondo una stima di Gomez Gane, presentata nell'articolo sopracitato, ammontavano a quarantamila nel 2008, e la cifra è andata ovviamente crescendo con l'adesione di nuovi enti e di tre grandi biblioteche italiane). Tra i partecipanti al progetto ci sono le principali biblioteche nazionali e universitarie di tutto il mondo: la Biblioteca di Harvard, la New York Public Library, la British Library di Londra, la Bodleian Library di Oxford, l'Österreichische Nationalbibliothek di Vienna. Per la storia di Google Libri: <http://books.google.com/googlebooks/about/history.html> (ultimo accesso: agosto 2015). L'Italia ha aderito al progetto nel 2010 con la Biblioteca Nazionale Centrale di Roma, la Biblioteca Nazionale Centrale di Firenze e la Biblioteca Nazionale Vittorio Emanuele III di Napoli (partecipa anche l'ICCU): l'obiettivo è la digitalizzazione di un minimo di 500.000 testi sino a un massimo di un milione di volumi (cfr. <http://www.bnrcrmlibrari.beniculturali.it/index.php?it/832/progetto-googlebooks>). Sull'adesione al progetto da parte di alcune biblioteche dell'Università di Roma "La Sapienza", cfr. Adriana Magarotto - Maura Quaquarelli - Mattia Vallania, *Il Progetto di digitalizzazione Google books presso le biblioteche della Sapienza, Università di Roma*, in «DigItalia. Rivista del digitale nei beni culturali», VIII, n. 2, 2013, pp. 117-25.

³ Cfr. Paolo D'Achille, *Parole nuove e datate. Studi su neologismi, forestierismi, dialettismi*, Firenze, Cesati Editore, 2012; Salvatore Claudio Sgroi, *La «glottologia» in «Google»*, in «Studi di lessicografia italiana», XXIX, 2012, pp. 227-30; Id., *Ramificazioni (e retrodatazioni) mafiose: la «mafia» in «Google»*, in «Studi di lessicografia italiana», XXX, 2013, pp. 273-89. Si veda anche la raccolta di retrodatazioni di Andrea Monaldi, *Mille retrodatazioni*, Roma, Bonacci editore, 2012.

⁴ Cfr., ad esempio, tra le pubblicazioni del 2012, sulle quali ho fatto una verifica dei risultati in Google Libri, ottenendo aggiornamenti per i già utili dati là raccolti: Gianluca Biasci, *Retrodatate con il RALIP. Mille retrodatazioni da opere narrative tra Otto e Novecento*, Roma, Aracne, 2012 (per la descrizione del metodo di raccolta con il RALIP, strumento informatico ideato da Biasci stesso per rendere rapido e automatico lo spoglio del "testo miniera", cfr. § 7); Id., *Nuove retrodatazioni da testi letterari otto-novecenteschi*, Roma, Aracne, 2012; Debora De Fazio, *Elenco delle retrodatazioni*, in appendice a Ead., *Cesare Lombroso e la lingua italiana. Psichiatria, etnologia, antropologia criminale nell'Italia di fine Ottocento*, Galatina, Congedo editore, 2012, pp. 339-52; Pierluigi Ortolano, *Retrodatazioni nel Vocabolario italiano-inglese di Giovanni Torriano (1659)*, in «Italianistica.it. Rivista internazionale di lingua e cultura italiana», I, 2012, pp. 51-64.

Ho dunque iniziato a testare la produttività di Google Libri nella ricerca di retrodatazioni, raggiungendo finora tremila nuove date. Come lemmario di riferimento ho preso quello dello *Zingarelli 2014*, con la cui redazione sto collaborando in vista del prossimo aggiornamento. Come ormai tutti i principali dizionari dell'uso in unico volume, anche lo *Zingarelli* indica, accanto all'etimologia, la data di prima attestazione delle parole⁵.

Per avere un'idea del rendimento di questa verifica e correzione, si consideri che, scegliendo come campione di prova, ad avvio del lavoro, i 100 lemmi che aprono il dizionario, con Google Libri ne ho retrodatati 41. Il dato, già di per sé significativo, va ulteriormente commentato: delle 100 parole prese come campione, quasi la metà, 45, erano parole antiche con datazione dal XII secolo al Quattrocento, e come tali non controllabili in Google Libri, un archivio che raccoglie testi a stampa⁶. Le 41 retrodatazioni sono dunque state ottenute sulle restanti 55 parole. E non si è trattato solo di retrodatazioni minute di qualche anno, bensì anche d'interi secoli.

Molte parole sono state riportate dal Novecento all'Ottocento (ad esempio *abbacchiario*, *abbassalingua*, *abbattibile*)⁷; talvolta hanno migliorato la

⁵ Data che, in alcuni casi, grazie agli aggiornamenti annuali di questo fortunato dizionario, è migliore di quella fornita dal *GRADIT*, dal *DELI* e dagli altri repertori classici di riferimento (la fonte della retrodatazione è archiviata dalla redazione, non dichiarata per ovvie esigenze editoriali di spazio; certo al lettore resta in molti casi il problema della mancata specificazione dell'accezione a cui è riferita la data: cfr. le riserve di L. Serianni, contenute nell'intervento raccolto in questo stesso volume, sulla presenza della datazione nei dizionari dell'uso). Primo dizionario dell'uso in unico volume a introdurre la data di prima attestazione delle parole (segnalandola come una delle innovazioni più importanti ed essenziali in un vocabolario moderno) è stato il Palazzi-Folena del 1992, nella nuova edizione Loescher, con la collaborazione di Carla Marellò, Diego Marconi e Michele Cortelazzo, al quale ultimo si deve in particolare la cura delle datazioni. Il Sabatini-Coletti (Giunti 1997), con una nutrita redazione di storici della lingua, ha poi, a sua volta, accolto questa informazione, scegliendo di indicare solo il secolo di nascita per le parole attestate prima dell'Ottocento. Lo Zingarelli ha invece introdotto la data di prima attestazione dei lemmi dall'edizione 1999 (stampata nella primavera 1998): collaboratore principale per il nuovo aggiornamento è stato Riccardo Tesi.

⁶ Per queste parole più antiche ci si affidi a *TLIO*, a *LEI* e al Battaglia, dizionario storico che, pur essendo stato riferimento per la compilazione dei moderni dizionari etimologici, contiene talora informazioni sfuggite ai lessicografi, motivo per cui non è mai superfluo consultarlo (cfr. Luigi Matt, *Sulle prime attestazioni dei vocaboli nei dizionari*, in «Studi linguistici italiani», XXVIII, f. II, 2002, pp. 272-97). Cito solo un esempio di parola presente nel Battaglia con rimando a fonte non accolta nei dizionari etimologici: *legaccia* 'giarrettiera' (*DELI* av. 1655, ma Battaglia: 1542, Aretino, *Talanta*).

⁷ *Abbacchiario* dal 1923 al 1838 (in *Raccolta delle leggi e disposizioni di pubblica amministrazione nello Stato Pontificio emanate nell'anno 1837*, volume unico, Roma, nella stamperia della R.C.A., 1838, p. 238: «resta proibito ai macellaj, abbacchiari e tripparoli di manifatturare e vendere qualunque specie di carni e grassi di majale»; poi in «La frusta. Giornale politico morale», anno 1, n. 7, Roma, domenica 4 dicembre 1870, p. 28: «Ieri mattina verso le ore

datazione addirittura di due secoli: *abbeveraggio* è passato dal Novecento al Settecento, *abbigliatura* e *abbioccare* dall'Ottocento al Seicento⁸. La percentuale di retrodatazioni è certo molto alta per il lessico attualmente considerato otto-novecentesco, ma con Google la data si raffina anche per quello sei-settecentesco: lo si può verificare persino sul piccolo campione iniziale a cui ho appena fatto riferimento, che è stato aggiornato anche per le parole del XVII-XVIII secolo, con retrodatazione di cinque lemmi: *abbacinamento*, *abbarbagliamento*, *abbindolare*, *abbindolatura*, *abbindolamento*⁹.

Raggiunte tremila retrodatazioni, posso avanzare la stima che la data di oltre il 70% del lessico moderno è da correggere. Per attuare tale aggiornamento, Google è uno strumento da cui non si può prescindere per l'enorme quantità e varietà di libri italiani di tutte le epoche e di tutti i generi che

quattro due annessionisti sono penetrati in una bottega di abbacchiario in via del Leoncino»); *abbassalingua* dal 1915 al 1834 (in Francesco Bennati, *Studii fisiologici e patologici sugli organi della voce umana*, Milano, dalla tipografia di Sambrunico-Vismara, 1834, p. 55); *abbattibile* dal 1929 al 1864 (in Pietro Fanfani, *Il Borghini*, anno II, Firenze 1864, p. 196).

⁸ *Abbeveraggio* dal 1974 al 1775 (in *Editti, pregoni, ed altri provvedimenti emanati pel Regno di Sardegna*, t. II, nella reale stamperia di Cagliari, 1775, p. 139: «Le acque de' pubblici pozzi, o fonti destinate all'uso degli individui, ed anche all'abbeveraggio degli armenti, dovranno tenersi monde [...]»); *abbigliatura* dal 1819 al 1612, attestato nella Crusca, nel dizionario bilingue latino-italiano che fa da indice finale, s.v. *concinatio*; *abbioccare* dal 1846 al 1613, attestato nel Dizionario toscano di Adriano Politi (Roma, Mascardi, 1613, s.v.).

⁹ *Abbacinamento*: DELI e Zingarelli 1692, Google 1639 (in Girolamo Mautini, *Prediche fatte nel palazzo apostolico*, III edizione, Roma, Typis Vaticanis, 1639, p. 214: «Che abbacinamento d'occhi»); *abbarbagliamento*: Zingarelli 1690, Google 1562 (in Gaudenzio Merula, *Nuova selva di varia lettione*, Venezia, per G.A. Valvassori detto Guadagnino, 1562, p. 301: «dopo il morso della quale, segue abbarbagliamento de gli occhi»), *abbindolare*: DELI e Zingarelli 1691 (terza edizione della Crusca), Google 1683 (in Isidoro Lanfredini, *Nouvelle et facile methode pour apprendre la langue italienne*, parte seconda, Parigi, presso Laurent d'Houry, 1683, p. 125: «*Abbindolare, aggirar uno*. Amuser, trainer quelqu'un»); *abbindolatura*: Zingarelli av. 1742, Google 1736 (in Giovan Battista Fagioli, *Commedie*, t. VI, Firenze, nella stamperia di F. Moucke, 1736, p. 161: «e se conoscerò punta punta d'abbindolatura nel vostro operare, [...] vi farò camminar io più che di trotto»); *abbindolamento*: DELI e Zingarelli 1734, Google 1733 (in Dionisio Longino, *De sublimi libellus Graece conscriptus. Latino, italico et gallico sermone redditus*, Verona, dalla tipografia di Tumermani, 1733, p. 123: «in ischerno di se medesimo pigliando l'abbindolamento delle parole»). Pur non essendo lo strumento migliore per gli studi sul lessico cinquecentesco, per la cattiva lettura dei caratteri tipografici antichi e per la modestia di alcune riproduzioni fotografiche, Google restituisce talvolta datazioni utili anche per il XVI secolo. Qualche esempio: *adorabile* (1669, ma in Google 1546: lettera dell'Aretino), *affaccendare* (1659, ma in Google 1550: Luca Contile, *La comedia del Contile chiamata la Pescara*, Milano 1550; atto primo, p.s.n.: «Anch'egli è stato affaccendato, ma non fu giorno che [...]»), *disseminatore* (1735, ma in Google 1569: Alessandro Piccolomini, *Piena, et larga parafrase di m. Alessandro Piccolomini nel secondo libro della Retorica d'Aristotele*, Venezia 1569, p. 121: «disseminatori, et divulgatori di tutto quel che sanno de i fatti, et de i detti d'altri»), *furbacchione* (1840, ma in Google 1575: *Delle homelie del R.D. Onofrio Zarrabini da Cutignuola*, parte ultima, Venezia 1575, p. 217: «è un ladro, è un furbacchione»).

interroga automaticamente (giornali, almanacchi, editti, paraletteratura, manuali e riviste scientifiche, ecc.). I testi ottocenteschi, in particolare, qui digitalizzati in gran numero, sono fonti preziose per datare, tra l'altro, molto lessico tecnico finora erroneamente ritenuto di origine novecentesca.

1. ERRORI DI SCANSIONE DEL TESTO E CATTIVE LETTURE DELL'OCR

Non bisogna pensare che la ricerca di retrodatazioni in Google sia rapida e non problematica solo perché informatizzata. Per comprensibili difetti del sistema, serve il controllo vigile di un operatore che divida i risultati positivi da quelli ingannevoli.

Innanzitutto, la scansione digitale dei testi può presentare manchevolezze di vario tipo: libri monchi, frontespizi assenti o la cui riproduzione fotografica è tagliata male (così da far perdere porzioni di testo o da renderne difficile la lettura), riproduzioni che saltano pagine o coprono involontariamente parte del documento con la mano del digitalizzatore. Imperfezioni siffatte nuocciono prima di tutto al lavoro di chi usa la biblioteca di Google per la consultazione integrale di un libro specifico; incidono meno, invece, nella ricerca di retrodatazioni¹⁰: le sviste del personale nella preparazione dei file avranno sicuramente causato perdite di dati potenzialmente validi, ma ritengo che tali perdite siano bilanciate dall'enorme consistenza del corpus¹¹.

Si tenga presente, inoltre, che la perdita di informazioni non è da imputare solo agli errori materiali dovuti alla frettolosa scansione del testo, ma anche alle cattive letture dell'OCR (*Optical character recognition*, cioè lo strumento di riconoscimento ottico dei caratteri presenti in un'immagine acquisita tramite scanner), che non è certo infallibile. L'OCR può restituire letture sbagliate o non riconoscere parole in testi pur perfettamente scansionati. Nel primo caso, le cattive letture saranno pazientemente eliminate a una a una dallo studioso, aprendo il collegamento alla fotografia e controllando che la parola in esame figuri davvero nella fonte; nel secondo caso, in-

¹⁰ Non saranno ovviamente accolti i dati di pubblicazioni prive di frontespizio, considerando l'incuria di Google nella schedatura del materiale (cfr. § 2).

¹¹ Si tenga presente che molti testi hanno due o più digitalizzazioni, perché provenienti da biblioteche diverse, ognuna delle quali ha immesso la copia del volume in proprio possesso (sui doppioni digitalizzati, cfr. Maurizio Messina, *Aspettando Google Book Search*, in «Bollettino AIB. Rivista italiana di biblioteconomia e scienze dell'informazione», 50, n. 4, dicembre 2010, pp. 384-85). Se un esemplare è lacunoso per incuria dell'operatore che ha scansionato il volume, non è dunque impossibile trovarne un secondo buono nella biblioteca di Google, recuperando così anche il materiale lessicale ipoteticamente perso dalla digitalizzazione cattiva.

vece, le parole non individuate dall'OCR saranno irrimediabilmente perse¹².

Gli errori di lettura sono i più vari e non riguardano, come si potrebbe supporre, soltanto il riconoscimento dei caratteri tipografici antichi, dove facile è confondere *f* e *s*, *u* e *v*. Google può sbagliare persino nell'interrogazione di testi dell'Ottocento. Mi sono ad esempio imbattuta nei seguenti errori del sistema: cercando attestazioni ottocentesche di *cinematico*, il motore di ricerca ha restituito anche attestazioni di *matematico*, *sistematico* e *grammatico*; Google ha letto *idioletto* anche laddove c'era scritto *Micheletto* e *idoletto*; *canalizzare* laddove c'era *e analizzare* o *nazionalizzare*; *colemia* dove invece c'era scritto *colonia* e *cotenna*.

Non essendo programmato per riconoscere la punteggiatura, e non tenendo conto talvolta degli spazi bianchi (in alcuni casi per cattiva qualità della stampa, con spaziatura tra parole troppo stretta), Google Libri restituisce insieme a tutti gli *abbassalingua* anche "*s'abbassa; lingua*", spezzetta *ergodico* in "*ergo dico*", *adesione* in "*adesso ne*" o "*ad esso ne*" (scambiando inoltre *-ss-* per *-si-*), *parruccaio* in "*parrucca. Io*", *contempo* in "*con tempo*" (restituendo anche "*alcun tempo*"), e così via. Questo difetto può d'altro canto essere un vantaggio nella ricerca di parole composte, le cui occorrenze sono restituite sia che la grafia sia unverbata, con trattino, o con i due elementi lessicali in successione. Le parole che vanno a capo a cavallo di rigo sono in alcuni casi riconosciute e restituite, in altri perse.

Come noto, il corpus di Google Books è interrogabile dalla maschera di ricerca "Google Libri Ricerca avanzata", che permette di cercare singole parole, o stringhe di testo, all'interno di un arco cronologico delimitato. Ad ogni invio di ricerca, il computer restituisce tante schede (e possono essere migliaia) quanti i libri in cui ha trovato il rimando desiderato. Le schede sono fornite di titolo, autore e data del libro di riferimento, con la parola cercata marcata in grassetto (evidenziata in giallo nella riproduzione

¹² Di queste parole non abbiamo notizia diretta, ma giungiamo a ipotizzarne la perdita osservando che, in alcuni file, la parola cercata è presente un numero di volte maggiore rispetto a quelle rilevate ed evidenziate dall'OCR. Ci sono casi in cui l'OCR estrae la parola desiderata dall'indice del libro, ma inspiegabilmente non la recupera a testo. E anche in altre occasioni mi è capitato di constatare il mancato riconoscimento di parole: solo alcune delle retrodatazioni che ho estratto dalla *Vita di Alberto Pisani* dello scapigliato Carlo Dossi, disponibile in Google in libera lettura, sono riconosciute e restituite dal motore di ricerca; altre sono recuperate da Google solo affinando le richieste nella maschera di ricerca avanzata e chiedendo di cercare le attestazioni specificamente nei libri dell'autore Dossi; altre non sono trovate nemmeno in questo modo (per recuperarle servono nuovi programmi informatici: cfr. § 7). Cfr. Ludovica Maconi, *Retrodatazioni lessicali da Carlo Dossi e Giovanni Faldella*, in F. Mereta - A. Sisti (a cura di), *Sette studi e un croquis. Omaggio a un maestro*, introduzione di G. Tesio, Vercelli, Edizioni Mercurio, 2014, pp. 17-53.

fotografica ad apertura del collegamento). Queste schede, però, contengono talvolta errori e non sono riordinate cronologicamente dal computer (né è possibile impostare il riordinamento ascendente, quello che serve per avere d'un colpo la data più antica)¹³.

Il nucleo di attestazioni più antiche, che a noi interessa, andrà quindi recuperato manualmente, riducendo progressivamente l'arco cronologico della ricerca agli anni più lontani e scorrendo pazientemente tutte le pagine dei risultati di Google. È consigliabile ripetere la ricerca su segmenti cronologici diversi, di un secolo, cinquanta e trent'anni, perché a probabili banchi del sistema si deve la perdita di informazioni settando in modo differente l'intervallo di date¹⁴.

2. MANCANZA DI RIGORE FILOLOGICO NELLA SCHEDATURA DEL MATERIALE

Difetto maggiore di Google Libri è la mancanza di rigore filologico nella schedatura del materiale, talvolta immesso con dati bibliografici scorretti. Facile accorgersi dell'errore se si tratta di opere famose, come nel caso del libro *Cuore* di De Amicis (1886), che compare, tra l'altro, come riferimento per la retrodatazione della parola *sobbalzo*, ma con la data sbagliata del 1822, anno in cui De Amicis non era nemmeno nato. L'errore marchiano è dovuto

¹³ Utile strumento per indagini statistiche sulla frequenza delle parole nel corso dei secoli, da affiancare a Google Books nelle indagini sul lessico, è Google Ngram Viewer (cfr. http://en.wikipedia.org/wiki/Google_Ngram_Viewer, ultimo accesso: agosto 2015). Ringrazio Tullio De Mauro per avermi segnalato le potenzialità di questo strumento.

¹⁴ Può capitare che un libro pubblicato nel primo decennio dell'Ottocento, e così schedato in Google, compaia tra i risultati della ricerca cercandolo nella prima metà del XIX secolo, ma scompaia restringendo l'indagine ai primi vent'anni. Inspiegabile (o spiegabile solo ricorrendo alla consulenza di un informatico che conosca bene il sistema adottato da Google per l'indicizzazione) anche il seguente caso in cui mi sono imbattuta: cercando l'attestazione più antica della parola *pollicoltore*, ponendo il limite cronologico del XIX secolo, Faldella non compariva tra le fonti restituite; raffinando però l'impostazione della maschera di ricerca avanzata, e restringendo il campo a "Faldella" come autore, ecco segnalatami l'attestazione di *pollicoltore* nelle *Figurine*. Certo, se non avessi già saputo per altra via della presenza di quella parola in Faldella, il dato sarebbe andato perso. D'altro canto i risultati non sarebbero cambiati significativamente, perché, anziché *Figurine* 1875, Google avrebbe restituito per *pollicoltore* una data e una fonte di poco posteriore. La parola *pollicoltore* è attualmente datata 1884 nel *GRADIT* e nel *DELI*, 1875 nello *Zingarelli 2014* (che ha aggiornato la data usando la *LIZ* e recuperando così l'attestazione presente nelle *Figurine* di Faldella). Google recupera per *pollicoltore* attestazioni da fonti degli anni Ottanta dell'Ottocento, in visualizzazione protetta: *L'uomo delinquente* di Cesare Lombroso (in Google presente nell'edizione 1889, ma la prima edizione del saggio è del 1876), *Giornale di agricoltura, industria e commercio del Regno d'Italia* (vol. 33, 1880), *Lessico dell'infima e corrotta italianità* di P. Fanfani e C. Arlia (ed. 1890).

al fatto che sul frontespizio c'è l'indicazione 1822° migliaia, riferito alla tiratura della fortunata opera (la parola *sobbalzo*, *DELI* 1873, è retrodata-bile con Google al 1828, scartando i riferimenti ingannevoli più antichi)¹⁵.

Facile accorgersi dello sbaglio anche quando, cercando retrodatazioni di *bon ton* (*DELI* 1813, retrodata-bile con Google al 1784)¹⁶, Google dà come attestazione più antica una fonte erroneamente schedata come secentesca: le *Opere* di Giuseppe Pitre, noto etnologo ottocentesco. In questo caso l'errore è dovuto a un numero (presumibilmente catalografico) presente sul retro del frontespizio, in caratteri piccoli sopra i dati di stampa, ignorati da chi ha immesso il file (ed. Firenze, 1950). Essendo qui l'indicazione vistosamente sbagliata, anche solo aprendo il collegamento alla fotografia, il ricercatore si accorge che la veste editoriale non può certo essere del Seicento. In altri casi, però, l'errore potrebbe farsi più insidioso, e si tenga presente che, usando il corpus di Google, solo raramente si retrodata con autori celebri¹⁷.

La fretteolosità con cui è catalogato il materiale è palese in refusi lasciati nei titoli e nei nomi degli autori: subito ce ne accorgiamo per "Targioni Torzzetti" anziché "Tozzetti", forse è necessario aprire il collegamento e controllare il frontespizio in altri casi, in base alle conoscenze di chi utilizza questa risorsa della Rete. Personalmente, ad esempio, ho fatto la verifica aprendo il collegamento e confrontando i dati bibliografici nel catalogo del Servizio Bibliotecario Nazionale (SBN) per "Ettore De Ruggieno", così schedato da Google anziché "De Ruggiero", scoprendo che per questa fonte Google sbaglia anche l'indicazione dell'anno di pubblicazione (nel febbraio 2016, mentre correggo le bozze, l'errore in Rete è stato eliminato).

Non rare sono le cattive letture delle date di stampa, che scambiano 3 per 5, 8 per 6, 9 per 0, oppure interpretano male i numeri romani. L'errore è talvolta evidente fin dall'incoerenza della scheda di Google: non potrà certo essere del 1808 un libro che ha per titolo *La rivoluzione del 1848 e la*

¹⁵ Fonte della retrodatazione: *Biografia universale antica e moderna*, vol. XLIV, Venezia, presso G.B. Missiaglia, 1828, s.v. *Plinio il Giovane*, p. 440: «lo stile n'è a sobbalzi e saltellante, e con profusione vi è usata l'antitesi».

¹⁶ Fonte della retrodatazione: *Opere di Francesco Albergati Capacelli*, tomo V, Venezia, nella stamperia di C. Palese, 1784, p. 4: «il bel mondo allora e il *bon ton* si scaricherebbe sopra di me».

¹⁷ Alcune retrodatazioni tra le non numerose ricavate da opere di autori celebri: F. Soave, *appiattimento* (retrodatato dal 1906 al 1822, ma av. 1806, anno di morte dell'autore: in *Istituzioni di metafisica*); V. Malacarne, noto scienziato piemontese, usa l'aggettivo *anteroposteriore* (retrodatato dal 1970 al 1803, in *Memorie di matematica e di fisica della Società Italiana delle Scienze*, tomo X, parte I, Modena 1803); V. Gioberti, *incomunicabilità* (dal 1869 al 1844, in *Degli errori filosofici di Antonio Rosmini*, ed. II, tomo III, Bruxelles, 1844), P. Petrocchi, *pensucchiare* (Zingarelli av. 1908, Google 1891, nel *Novo dizionario della lingua italiana*, s.v.).

provincia di Caltanissetta, né del 1615 il *Bollettino delle leggi del Regno d'Italia*, e pure entriamo (giustamente) in sospetto se un opuscolo che Google dice pubblicato nel 1744 s'intitola *Vedute di Firenze nel '700*.

Per tutti i testi presenti in visualizzazione completa, il problema dell'errore di schedatura si risolve facilmente controllando il frontespizio, passaggio che è assolutamente d'obbligo. Per le pubblicazioni che invece hanno visualizzazione parziale, e per tutti i casi incerti, è bene controllare i dati bibliografici nel catalogo SBN e decidere cosa accogliere come certo, cosa controllare in biblioteca e cosa eliminare.

3. PRIME ATTESTAZIONI IN RIVISTE E GIORNALI

Il problema delle schedature sbagliate o imprecise riguarda in particolare le riviste, per le quali spesso è indicato un intervallo di date, oppure una sola data associata a più volumi¹⁸. Capita inoltre che una stessa rivista sia immessa più volte in Google, con digitalizzazioni doppie provenienti da biblioteche diverse, che diversamente hanno schedato il materiale, creando incoerenze.

Come spiegare queste fuorvianti indicazioni? Nelle biblioteche, annate consecutive di una stessa rivista possono essere rilegate in unico volume. Da accordi con Google, le biblioteche forniscono il materiale per la digitalizzazione, ma nessuno si cura della descrizione del volume, né tantomeno della marcatura del testo¹⁹. Per questo motivo un unico *link* può aprire annate

¹⁸ Le schede sbagliate o imprecise creano ovviamente problemi per la retrodatazione quando le riviste sono presenti in visualizzazione limitata (*snippet* 'ritaglio'), con il solo zoom di una piccola sezione del testo (zoom che oltretutto non sempre riproduce il passo con la parola desiderata, ma può essere centrato su una diversa porzione della pagina).

¹⁹ Gli accordi con Google stabiliscono la quantità di materiale da scansionare, non importa cosa si sceglie per la digitalizzazione, conta la quantità: «L'accordo Google books/MiBAC prevede fino ad un massimo di 1 milione di volumi digitalizzati. I volumi comunque dovranno essere almeno 500.000. Google intende per volume il "nominal book", cioè un volume nominale standard di 330 pagine. La reale unità di misura è rappresentata dal carrello *S-cart* Google, con capacità di 260 volumi nominali o "n" volumi reali, per un'equivalenza di ca. 86.000 pagine digitalizzate a carrello» (<http://www.bnrcrmlibrari.beniculturali.it/index.php?it/832/progetto-googlebooks>, ultimo accesso: agosto 2015). Per questo prevalere del dato puramente quantitativo, Osvaldo Avallone, già direttore della Biblioteca Nazionale Centrale di Roma, ha criticato il procedimento di selezione delle opere destinate alla digitalizzazione, la mancanza di coordinamento e l'esclusione di molti libri del Sei e Settecento per problemi di formato delle tavole in essi contenute (cfr. Osvaldo Avallone, *Il Progetto Google books: la prima grande esperienza di accesso diretto al patrimonio bibliografico nazionale*, in «DigItalia. Rivista del digitale nei beni culturali», VIII, n. 1, 2013, pp. 9-13).

diverse (non sempre segnalate nella scheda di Google); così si spiegano più volumi associati a un solo anno (quello del primo fascicolo nell'ordine di rilegatura), o a un intervallo di date (la data del fascicolo in apertura e quella del fascicolo in chiusura del volume rilegato). I controlli in biblioteca sulla fonte cartacea saranno sempre d'obbligo se la digitalizzazione è in visualizzazione protetta.

Le retrodatazioni recuperate da Google in riviste andranno dunque accolte con cautela. Per avere una prima indicazione sull'attendibilità del rimando di Google, può essere utile il supporto del Servizio Bibliotecario Nazionale. Come noto, la scheda del catalogo SBN porta indicazione dell'anno di prima pubblicazione, periodicità ed eventuali interruzioni della rivista. Da questi elementi si può valutare quale catalogazione di Google regge e merita di essere approfondita e quale invece è da rifiutare²⁰.

Per alcune riviste, la fonte protetta in Google è disponibile in libera consultazione in Internet Archive, sito tramite il quale ho recuperato le fonti in visualizzazione completa per confermare la retrodatazione delle parole *aerosbarco* nella «Nuova Antologia» (1941, vol. 415; *DELI* 1942) e *alettone* nella rivista «Emporium» (1912, vol. 35; *DELI* 1925)²¹. Negli Stati Uniti le riviste, come i libri, se non più sotto diritto d'autore, sono generalmente disponibili in visualizzazione completa, quindi la catalogazione sbagliata o approssimativa, pur rallentando l'esame, crea minor rischio di errore, perché la fonte è direttamente accessibile²².

²⁰ Porto l'esempio del termine *coassiale*, oggi datato 1948 in *DELI* e *GRADIT*, retrodatabile al 1878 mediante Google, che recupera il passo «è coassiale colla superficie» in una rivista di matematica, schedata in tre modi diversi, in tre *link* a visualizzazione protetta: «Giornale di matematiche ad uso degli studenti delle università italiane» (vol. 16, 1878), «Giornale di matematiche di Battaglini» (voll. 15-16, 1877), «Giornale di matematiche» (1878, senza indicazione del volume). Dalla scheda del catalogo SBN leggiamo che la rivista «Giornale di matematiche ad uso degli studenti delle università italiane» inizia le pubblicazioni nel 1863; non sono dichiarate interruzioni, quindi, se il primo volume è del 1863, il sedicesimo sarà proprio dell'anno 1878. A queste valutazioni sull'attendibilità della fonte fornita da Google, seguirà il controllo in biblioteca. In America, la data 1878 per *coassiale* è confermata dal *link* in Rete alla pagina della rivista, disponibile in visualizzazione completa.

²¹ Cfr. sito <https://archive.org> (ultimo accesso: agosto 2015). Sul fruttuoso impiego di Internet Archive per aggirare alcuni divieti che bloccano la visualizzazione delle digitalizzazioni in Google Libri, cfr. Alberto Petrucciani, *Con Robert Darnton nella giungla digitale*, in «Bollettino ALB. Rivista italiana di biblioteconomia e scienze dell'informazione», 51, n. 1/2, 2011, pp. 97-105 (in particolare, cfr. pp. 100-1). Per la datazione di *aerosbarco*, cfr. «Nuova Antologia», 415, 1941, p. 404: «quegli strumenti generatori di nebbia che hanno in tanti casi protetto l'aerosbarco, e impedito ogni efficace reazione a terra»; per *alettone*, cfr. «Emporium», 35, 1912, p. 69, articolo di C. Canovetti, *Storia dell'aviazione*: «equilibratore ad alettone portato dietro il timone posteriore».

²² Per le numerose informazioni e verifiche sul funzionamento di Google Books dagli

Le schede di Google possono ricomparire a distanza di tempo modificate. Così mi è capitato di osservare per la schedatura di «Lingua nostra», fonte della retrodatazione di *aeroreattore* (DELI 1955): inizialmente la scheda portava indicazione dell'anno 1951 associato a tre volumi (XII, XIII e XIV, rispettivamente degli anni 1951, 1952 e 1953), con rimando alla pagina 120 per la parola cercata; oggi, invece, la rivista è schedata con l'anno 1952, associato a due volumi (XIII-XIV). Ma in quale volume, e quindi in quale anno, è effettivamente attestata la parola *aeroreattore*? Nel vol. XII del 1951, nel vol. XIII del 1952 o nel vol. XIV, che è dell'anno seguente? Solo andando in biblioteca a verificare possiamo averne la conferma, perché la rivista in Rete è protetta, in visualizzazione *snippet*: «l'aereo era chiamato *aeroreattore*» è in «Lingua nostra», vol. XIII, 1952, p. 120. Se accolta senza ulteriori verifiche, la prima schedatura avrebbe fornito una data scorretta (per questa fonte anche negli USA la visualizzazione è protetta, nel rispetto dei diritti d'autore).

Il problema della rilegatura in unico volume di materiale pubblicato in anni diversi riguarda inoltre estratti di articoli (anche articoli di varia materia e provenienza eterogenea), raccolti e schedati tutti sotto titolo e data del primo nell'ordine²³, e i giornali, molti ottocenteschi, disponibili però, quasi sempre, in visualizzazione completa²⁴; in alcuni casi mi è anche capitato

Stati Uniti, ringrazio Eugenio Refini, che mi ha dato un grande aiuto in questo lavoro. Sulla questione del rispetto del *copyright* in Google, cfr. M. Messina, *Aspettando Google Book Search*, cit., pp. 381-84; Adrian Johns, *Pirateria. Storia della proprietà intellettuale da Gutenberg a Google*, Torino, Bollati Boringhieri, 2011, pp. 667-71.

²³ In generale, quando i titoli rimandano a raccolte di fascicoli eterogenei, il ricercatore deve aspettarsi la possibilità d'errore. Talvolta queste fonti hanno addirittura titolo inglese, perché il materiale appartiene a biblioteche britanniche o americane, che l'hanno rilegato e schedato sotto titolo generico (es. *Collection of works...* o simili formule). Cfr. la retrodatazione di *peristalsi* (s.f. 'movimento di contrazione e rilassamento dei visceri muscolari cavi con effetto di avanzamento progressivo del contenuto': DELI e Zingarelli 1895, Google 1859), parola che in Google sembra avere attestazione più antica in un volume della Bodleian Library di Oxford, schedato sotto il titolo *Collection of works, mainly in Italian* (1850): in realtà l'anno indicato nella scheda di Google è quello dell'estratto di L. Amabile sulla flora irpina che apre il volume; la parola *peristalsi* è attestata invece in un saggio di medicina che segue, sempre di L. Amabile, ma dell'anno 1859 (L. Amabile - T. Virnicchi, *Sulle soluzioni di continuo dell'intestino e sul loro governo. Commentario*, Napoli, Stamperia De Angelis, 1859, p. 45).

²⁴ Segnalo, tra le tante date sbagliate assegnate alle digitalizzazioni immesse nella Rete, quella del «Corriere delle Dame», presente in Google con una riproduzione datata 1815 sebbene i fascicoli siano del 1828. Tra le prime attestazioni attinte tramite Google al «Corriere delle Dame», elenco: *chiné* (agg. 'detto di tessuto screziato', DELI 1885), presente in un numero del 1846 (28 marzo, p. 137): «Le stoffe per gli abiti saranno: [...] il *chiné* a colonne, disegno largo e grandioso»; *grisaille* (s.f. 'qualsiasi tessuto, spec. di lana di filati grigi, o bianchi e neri con effetto di grigio', DELI 1886), parola attestata nel fascicolo del 12 maggio 1857, p. 145: «Le stoffe di moda sono - la *grisaille*, tessuto composto di bianco, nero e grigio»; *toque* (s.f.

di trovare accorpati due libri (senza segnalazione alcuna, come comporta, ormai è chiaro, la meccanicità e frettolosità del lavoro di archiviazione in Google)²⁵.

4. ASSENZA DI MARCATURE

L'assenza di marcatura nelle digitalizzazioni può creare errori di vario tipo nella ricerca di prime attestazioni. Google offre la riproduzione fotografica e la digitalizzazione del testo, ma nulla dice sulla struttura del libro; inoltre, come più volte ribadito, la scheda catalografica è rudimentale, talvolta addirittura portatrice di errori. È quindi necessario appurare la "sincerità" della fonte prima di dichiarare la parola retrodatata.

Attenzione deve prestare il ricercatore, per non lasciarsi ingannare, nei casi in cui la parola si trova nelle note di commento o nell'introduzione del curatore moderno a un testo più antico. Cercando *idioletto*, ad esempio, Google dà come fonte per la retrodatazione il Dizionario di Tommaseo (1865): apro il *link* e non limitandomi alla lettura del passo evidenziato, ma avendo la pazienza di scorrere il testo poco più oltre, si verifica che il tecnicismo compare nelle pagine prefatorie al Dizionario firmate da Gianfranco Folena nel 1977 per la ristampa anastatica Rizzoli. Sarebbe dunque un errore grossolano attribuire la parola *idioletto* al linguista dalmata, assumendo meccanicamente il risultato offerto da Google (il file non è più reperibile in Google nel febbraio 2016, mentre correggo le bozze).

Possibile errore di simil fatta anche per la retrodatazione della parola *aprioristico*, per la quale Google rimanda al *Vocabolario toscano dell'arte del disegno* di Filippo Baldinucci (1681), nell'edizione anastatica con nota

'cappello da donna senza tesa', *DELI* 1905) nel fascicolo del 30 agosto 1823, p. 280: «Fra gli altri adornamenti del capo vuoi citare un *toque*». Per altre retrodatazioni di parole della moda dell'Ottocento, cfr. Giuseppe Sergio, *Parole di moda: il "Corriere delle Dame" e il lessico della moda nell'Ottocento*, Milano, Franco Angeli, 2010 (da verificare in Google).

²⁵ Cfr. la digitalizzazione del saggio del padre gesuita Giampietro Secchi, *La cattedra alexandrina di S. Marco evangelista e martire conservata in Venezia entro il tesoro marciano delle reliquie* (Venezia, Tipografia di P. Naratovich, 1853), che in coda porta le 82 pagine di un testo di qualche anno prima, *La Pala d'oro dell'I.R. patriarcale Basilica di San Marco, considerata sotto i risguardi storici, archeologici ed artistici dal Mons. Giovanni Bellomo* (Venezia, Co' tipi di P. Naratovich, 1847). Per evitare errori nella ricerca di retrodatazioni in Google, è sempre opportuno far scorrere il cursore laterale all'inizio e alla fine del file per verificare, oltre al frontespizio, l'indice e il numero totale di pagine digitalizzate, alla ricerca di possibili incoerenze che compromettano l'attendibilità della fonte (la scheda SBN dà indicazione del numero di pagine: se l'ultima pagina del file non corrisponde alla numerazione dichiarata dal catalogo SBN, bisogna entrare in sospetto e approfondire).

critica di Severina Parodi: il testo è del Seicento, ma ovviamente non lo è la nota critica della Parodi, nella quale è usata la parola *aprioristico*. In questo secondo caso, a smascherare l'errore, aiuta decisamente la veste tipografica, che differenzia a colpo d'occhio la parte moderna novecentesca dal testo secentesco riprodotto in anastatica²⁶.

L'assenza di marcature comporta inoltre il mancato riconoscimento di lingue diverse all'interno di un libro archiviato come "italiano". Può così capitare che il rimando a dizionari bilingui sia scorretto: cercando *corrugare*, ad esempio, Google restituisce un dizionario volgare-latino (Venezia, 1568), in cui però il verbo è nella colonna del latino: «corrugare vel complicare manicas ad cubitum» ('tirare su le maniche'), quindi in una lingua e in un significato diverso da quello desiderato.

Sbagliato, ma interessante per avere informazioni sulla storia della parola prima del suo ingresso ufficiale nella lingua nazionale, è il riferimento recuperato nei dizionari dialettali, quando la parola individuata sta nella colonna del dialetto (da non accogliere dunque come retrodatazione). Il termine *paciugo*, 'pasticcio', solo per fare un esempio, è segnalato da Google nel *Dizionario domestico genovese-italiano* dell'Olivieri (1841), rinvio non valido perché qui la voce è dialettale (di essa è fornito l'equivalente italiano). Grazie a questa fonte, però, abbiamo conferma dell'origine settentrionale, e più precisamente ligure, di *paciugo*, come dichiarato nei prontuari etimologici.

5. RICERCHE COMBinate

Nella ricerca di parole omografe, che il computer ovviamente non distingue per significati, Google restituisce un numero altissimo di schede, difficile da scremare isolando i risultati che interessano. In questi casi, può essere utile tentare qualche ricerca combinata: si possono cioè cercare delle stringhe e non una singola parola, oppure la singola parola in abbinamento a un'altra che ne delimiti l'ambito semantico²⁷.

Ho usato questo espediente in due casi: per isolare i sostantivi che hanno

²⁶ L'aggettivo *aprioristico* (*DELI* e *GRADIT* 1905, *Zingarelli* 1894) trova in Google nuova attestazione nel 1846 (*Annali delle scienze religiose*, compilati dal Prof. Giacomo Arrighi, serie seconda, volume II, Roma, presso Pietro Capobianchi, 1846, p. 56: «Non già ch'è sieno in ciò proceduti a quel modo puramente astratto ed aprioristico che testè per noi si riprovava»).

²⁷ Questa seconda parola andrà inserita nel campo «Trova risultati che contengano una qualunque delle seguenti parole», presente nella maschera di ricerca avanzata di Google Libri; il lemma da retrodatare andrà invece nel campo «Trova risultati che contengano la seguente frase», come pure l'eventuale stringa lessicale.

la stessa forma del participio dalle forme verbali omografe, e per i nomi propri passati a nomi comuni. Aniché cercare *abbeverata*, ho quindi tentato di distinguere empiricamente con *una/l'abbeverata*, *bella abbeverata*, perdendo sicuramente qualche dato utile, ma evitando la restituzione di centinaia di schede in cui la parola compariva come verbo²⁸. Infinitamente lunga sarebbe stata anche la scrematura di *Fahrenheit* (*DELI* e *Zingarelli* 1892; Google 1765) se non avessi cercato le combinazioni *gradi Fahrenheit*, parallelamente a *termometro Fahrenheit* e *scala Fahrenheit*²⁹.

Segnalo qui che i libri latini sono classificati e compaiono in Google assieme a quelli italiani, creando rallentamenti in caso di ricerca di latinismi, per i quali il computer restituirà tutti i testi latini insieme a quelli italiani, senza possibilità di selezione automatica. Anche per i latinismi, dunque, sarà utile la ricerca combinata “articolo + sostantivo”: anziché *otium* (*DELI* 1855, Google 1852), sarà più pratico cercare *l'otium*, per eliminare i rimandi a opere e citazioni in latino³⁰.

Più ricerche servono per i verbi, perché, non essendo possibile la ricerca con caratteri jolly, bisogna coniugarli in vari modi; sono poi necessarie interrogazioni parallele per parole che hanno varianti grafiche e fonetiche (es. *abbonamentol abbuonamento*; *froebelianolfröbeliano*; *shakespearianol scespiriano*; *bancarella/bancherella*). Infine, ricerche combinate, nelle quali si associa alla parola da retrodatare una seconda parola che ne circoscrive l'ambito semantico, sono utili per la datazione di singole accezioni, non difficili da recuperare nel caso dei tecnicismi³¹.

²⁸ Retrodatazione di *abbeverata* (*DELI* e *GRADIT* 1853, *Zingarelli* 1928): Google 1830, con attestazione in Vincenzo Mazza, *Il veterinario militare*, Napoli, dalla Reale Tipografia della Guerra, 1830, p. 27: «Filtrata così l'acqua per l'abbeverata della mattina».

²⁹ Ho scelto sintagmi che escludessero la restituzione del nome proprio preceduto dalla preposizione *di*, nei quali il nome non ha ancora funzione aggettivale o appositiva (es. *gradi di Fahrenheit*). Fonte della retrodatazione: *Supplemento di Giorgio Lewis al Dizionario universale delle arti e scienze di Efraimo Chambers*, tomo VI, Venezia, G. Pasquali, 1765, s.v. *termometro*: «gradi del Termometro Fahrenheit». Altri esempi per la ricerca di nomi propri passati a comuni: anziché Maddalena, si cercherà “fare la maddalena”; anziché Cagliostro, “un Cagliostro”, o “quel cagliostro”; anziché Lloyd, “il Lloyd” o “del Lloyd”. Retrodatazione di Lloyd: 1836 (in *Annali universali di statistica*, vol. 49, Milano 1836, p. 323: «Nuova organizzazione del Lloyd austriaco a Trieste»).

³⁰ Fonte della retrodatazione: Giuseppe Micali, *L'Italia avanti il dominio dei Romani*, vol. II, Torino, Pomba e Compagnia, 1852, p. 161, n. 3: «Si celebra l'*otium* di Napoli dai poeti».

³¹ Esempio 1: *acchiappamosche* può essere retrodatato al 1794 come termine zoologico cercando la parola in co-occorrenza con *animale*; al 1819 nell'accezione botanica, cercando *acchiappamosche* in co-occorrenza con *pianta*. Non migliora invece con Google la datazione del termine nel significato di ‘strumento per uccidere le mosche’ (*DELI* 1950), retrodatabile però con un passo di V. Imbriani, *Mastr'Impicca* (1874), segnalato da G. Biasci in *Nuove retrodata-*

6. COME MIGLIORARE I RISULTATI DI GOOGLE

Senza aspettare nuove immissioni di testi in Google Libri, alcune date di prima attestazione possono essere migliorate nei seguenti modi:

1. Recuperando le prime edizioni, quando Google restituisce come fonte più antica della parola una riedizione. Porto un esempio. Per *gallomane* (*DELI* av. 1853) Google non dà attestazioni settecentesche ma solo ottocentesche, la più antica delle quali è la seconda edizione del trattato *Dell'uso e dei pregi della lingua italiana* di Galeani Napione. Recuperando però la prima edizione dell'opera (assente nel corpus di Google), e verificando la presenza del passo in esame, ho spostato al 1791 la data di prima attestazione del termine.

2. Metodo analogo vale per il miglioramento delle date restituite da Google in edizioni *post mortem*. Così mi è capitato per *ladruncolo* (Aretino, av. 1556; Google 1829), *appiattimento* (Francesco Soave, av. 1806; Google 1822) e *allombato* (Antonio Bresciani, av. 1862; Google 1868)³².

3. Quando fonte della prima attestazione è l'edizione di un epistolario, si può accogliere la data della lettera che documenta la parola, anziché quella, posteriore, dell'edizione a stampa (facendo attenzione ai falsi e ai falsi narrativi)³³.

4. Per le prime attestazioni ricavate da testi di cui esistono redazioni diverse ed edizioni critiche moderne, sarà utile la consultazione degli apparati critici, per verificare a quale redazione assegnare la parola. Così ho fatto per le prime attestazioni recuperate dall'*Altrieri* e dalla *Desinenza in A* di Carlo Dossi³⁴.

zioni, cit., p. 16. Esempio 2: in Google *glomerulo* è retrodatato al 1843 come termine botanico (*DELI* 1906) e 1859 come termine anatomico (*DELI* 1939-40); *scaglionare* (*DELI* 1872) passa al 1836, ma 1848 nell'accezione militare; *accollatura* è retrodatato al 1768 in riferimento al giogo dei buoi (*DELI* 1860), ma resta ottocentesco come termine della moda (*DELI* 1855): nel primo caso l'ho cercato in co-occorrenza con *giogo*, nell'altro con *veste e donna*.

³² Cfr. *Teatro classico italiano antico e moderno*, Lipsia, presso Ernesto Fleischer, 1829 (alla p. 180 l'attestazione di *ladruncolo*, forma *ladrunculo*, nel *Marescalco* di Aretino, atto II, scena XI: «un presuntuoso, uno inetto ladrunculo»); Francesco Soave, *Istituzioni di metafisica*, Venezia, Molinari, 1822, p. 277: «Ma siccome il moto di rotazione, che ha la terra presentemente, non sarebbe stato bastante a produrre l'appiattimento, che in lei si osserva [...]»; Antonio Bresciani, *Opere*, vol. XIV, Roma-Torino, Ufficio della civiltà cattolica-Marietti, 1868, p. 403: «camminando però bene allombati perché ben nutriti dalla pesca e dalla caccia».

³³ Esempio: retrodatazione di *chef* (*DELI* 1905, *Zingarelli* 1894), parola per la quale Google recupera l'edizione delle *Lettere del conte Gio. Battista Sommariva, a suo figlio Luigi*, edite a Parigi nel 1842. La parola è attestata nella lettera del conte datata 2 settembre 1824 (p. 301 dell'edizione parigina): «licenziava lo *chef* nostro di cucina».

³⁴ In questo caso, a parole che hanno come fonte lo stesso testo ho assegnato diversa data

5. In qualche caso, in particolare per tecnicismi, il miglioramento della data recuperata da Google può venire dal catalogo SBN³⁵: *freniatria* (1875, laddove Google dà 1868, per errore di schedatura, visto che il frontespizio porta data 1888; *GRADIT* e *Zingarelli* 1881), *aerotecnica* (*GRADIT* e *Zingarelli* 1925, SBN 1915), *grafologia* (*DELI* 1878, SBN 1876), *mancinismo* (*DELI* 1900, SBN 1884), *mineralogia* (*DELI* av. 1770, *Zingarelli* 1748, SBN 1663)³⁶.

7. PROGRAMMI INFORMATICI CHE UTILMENTE POSSONO AFFIANCARE GOOGLE NELLA RICERCA DI RETRODATAZIONI

Dal 2012 è liberamente disponibile in Internet il *RALIP*, *Retrodatatore automatico del lessico italiano postunitario*, ideato e realizzato da Gianluca Biasci (Università per Stranieri di Siena)³⁷. Questo programma estrae automaticamente retrodatazioni da testi immessi per la ricerca, ma funziona, come il suo stesso nome avverte, soltanto per testi otto-novecenteschi posteriori al 1860. Non di libero utilizzo è *TestRetro*, analogo programma realizzato da Daniele Fusi (Università di Roma “La Sapienza”) per gli aggiornamenti del dizionario *Zingarelli*³⁸. A differenza del precedente, questo programma lavora su testi di ogni epoca.

in base alla loro presenza fin dalla prima edizione dell’opera o soltanto nell’ultima. Cfr. L. Macconi, *Retrodatazioni lessicali*, cit., pp. 34-53.

³⁵ Sull’impiego del catalogo SBN come banca-dati per retrodatazioni di tecnicismi, cfr. L. Matt, *Retrodatazioni di tecnicismi da titoli di pubblicazioni*, cit.

³⁶ Per il termine *aerotecnica* Google restituisce due schede del 1915, ma l’una in visualizzazione *snippet*, in cui si legge solo *aerotecnico* aggettivo (trattasi quindi di cattiva lettura), l’altra in visualizzazione totalmente coperta (quindi non affidabile). Il termine *grafologia* compare in Google con rimando al manuale di Paolo Colombetti (Torino, 1876), schedato con dati bibliografici sbagliati e senza possibilità di visualizzazione. Per la maggiore affidabilità di schedatura di SBN, ho ritenuto indispensabile la sua consultazione parallela all’interrogazione di Google. Altri corpora elettronici possono occasionalmente migliorare le retrodatazioni di Google: interrogando *AITER* (Archivio Italiano della Tradizione Epistolare in Rete: <http://aiter.unipv.it/ricerca/>, ultimo accesso: agosto 2015), ad esempio, ho retrodatato *illustrissimo* come titolo reverenziale della scrittura epistolare (*DELI* 1541), recuperando l’attestazione in una lettera di Baldassarre Castiglione del 1520 (indirizzata a Federico Gonzaga). Per il lessico otto-novecentesco, utili anche gli archivi dei quotidiani nazionali, ora disponibili in Rete.

³⁷ Cfr. il sito internet del *RALIP*: <http://www.ralip.eu/index.php> (ultimo accesso: agosto 2015). Cfr. anche G. Biasci, *Retrodatate con il RALIP. Mille retrodatazioni*, cit.; Id., *Nuove retrodatazioni da testi letterari otto-novecenteschi*, cit.

³⁸ Il programma mi è stato gentilmente messo a disposizione da Enrico Lanfranchi (già direttore editoriale Zanichelli), ai fini della mia ricerca per l’aggiornamento delle date di prima attestazione dello *Zingarelli*.

RALIP e *TestRetro* possono utilmente affiancare Google nello studio delle retrodatazioni, fornendo lemmari di partenza già retrodatati da verificare nella biblioteca digitale di Google Libri (controllo che in molti casi raffinerà la data della parola)³⁹. Entrambi gli strumenti confrontano automaticamente le parole del testo immesso per la ricerca (in formato RTF e con indicazione della data di stampa), con il lemmario datato che hanno in memoria (preso, nel caso del *RALIP*, principalmente dal *GRADIT*, nel caso di *TestRetro* dallo *Zingarelli*). Da questo confronto restituiscono elenchi di possibili retrodatazioni. Dico “possibili” perché anche questi programmi incorrono, come Google, in errori di significati e categorie grammaticali.

Per ottimizzare la resa e ottenere retrodatazioni che abbiano poi maggiori possibilità di essere confermate dal successivo controllo in Google, è consigliabile scegliere testi più antichi, o fonti non letterarie. Utilizzando il programma di Fusi, ho così provato a esaminare testi cinquecenteschi: tre opere religiose dell’Aretino, lettere, dediche e avvisi ai lettori del Ruscelli, ottenendo una trentina di retrodatazioni utili, non migliorate poi da Google Libri⁴⁰. Cito a campione, dal Ruscelli: *elogio* (1561), *monastico* (1554), *rassomiglianza* (1554); dall’Aretino: *germinale* (1552), *idolatrico* (1552), *indeterminabile* (1540), *inviscerare* (1552), *ostetrica* (1552), *presciente* (1552), *sconquassamento* (1540)⁴¹.

³⁹ Si prenda, ad esempio, l’analisi compiuta sul testo della *Vita di Alberto Pisani* di Carlo Dossi (1870) e sulle *Figurine* di Faldella (1875). Per questi due testi il *RALIP* fornisce, fatte le opportune screature di doppiini e di errori di significato, 126 retrodatazioni, che messe poi alla prova in Google trovano data ancora più antica in 91 casi su 126, quindi solo 35 risultati del *RALIP* sono confermati, ben meno di un terzo del campione di partenza (cfr. L. Maconi, *Retrodatazioni lessicali*, cit.).

⁴⁰ Ringrazio Chiara Gizzi e Paolo Marini per aver messo a mia disposizione materiale utile per l’esame informatico dei testi.

⁴¹ Commenterò qui solo l’ultima parola in elenco, *sconquassamento*, presente nella *Vita di Santa Caterina* dell’Aretino (forma con *i* prostetica: *isconquassamento*). Google Libri retrodata *sconquassamento* (*GRADIT* av. 1694) con un testo religioso di inizio Seicento, che da controllo effettuato nel febbraio 2015 è però scomparso (sulla sparizione di materiali in Google Libri, cfr. l’esempio riportato al § 4; cfr. inoltre A. Petrucciani, *Con Robert Darnton nella giungla digitale*, cit., p. 101, con esempio di alcuni volumi non più consultabili del «Library journal»). Sembrano non esserci attestazioni cinquecentesche del termine, sia cercando la parola scritta con e senza *i* prostetica, sia cercandola con le *f* al posto delle *s*, ipotizzando una cattiva lettura dell’OCR per confusione dei caratteri tipografici antichi. Eppure nella biblioteca di Google Libri ci sono ben tre digitalizzazioni in visualizzazione completa della *Vita di Santa Caterina*: l’edizione del 1552 (immessa da due biblioteche) e quella del 1540. L’OCR, però, non legge la parola presente in questi testi, né con ricerca generale né con ricerca nel singolo volume. In questo caso, senza il programma informatico di Daniele Fusi e senza la digitalizzazione del testo messa a disposizione dall’editore moderno, avremmo perso una retrodatazione di oltre un secolo.

Il controllo in Google dei dati estratti con *RALIP* e *TestRetro* può inoltre servire a segnalare rapporti tra scrittori, testimoniati dalla ripresa di lessico. Retrodatazioni individuate nelle *Figurine* di Faldella (1875), per esempio, hanno trovato in Google documentazione precedente in note fonti dello scrittore scapigliato: Pietro Fanfani, il poeta Giusti e Giambattista Giuliani. Ne elenco alcune: il toscanismo *tipizzarsi* ‘contendere con parole ingiuriose, bisticciare’, *brugnoccolo* ‘grumo o bernoccolo’, *riacciuffare*, *motriglio* ‘fanghiglia’, *cibaglia*⁴².

8. HATHITRUST E CONTROLLI ALL'ESTERO

Esiste una biblioteca digitale che, oltre ai libri di Google, include digitalizzazioni di biblioteche americane che non partecipano al progetto “Google Books”. Si tratta di HathiTrust Digital Library, progetto partito nel 2008 con gli stessi scopi di Google ma con più ampia collaborazione di Università statunitensi⁴³. Si potrebbe dunque pensare che la ricerca di retrodatazioni abbia miglior esito in questa apparentemente più ampia banca-dati, ma così non è. Ho controllato le retrodatazioni ottenute con Google in HathiTrust, pensando di poter raffinare i risultati. Mi sono fermata dopo i primi settanta controlli, perché non solo la data recuperata da Google non è mai stata migliorata, ma in moltissimi casi, HathiTrust non ha nemmeno recuperato la fonte segnalata da Google come prima attestazione (forse per diversa indicizzazione).

Pur ricordando che Google e HathiTrust sono biblioteche virtuali, e che quindi, cercando prime attestazioni di parole, chiediamo a questi strumenti di aiutarci in un compito che non è quello previsto da chi li ha ideati, il funzionamento del motore di ricerca e la presentazione dei dati offerti da HathiTrust sono meno funzionali rispetto a quelli di Google.

Per accedere al materiale di HathiTrust bisogna, infatti, passare attraverso troppi passaggi che rallentano il lavoro; i tempi di caricamento del file sono molto lunghi; sono richiesti più invii per arrivare alla fonte fotografica, che, una volta raggiunta, diversamente da Google, non porta evidenziata la parola cercata, rallentando ulteriormente il lavoro del ricercatore, costretto a scorrere l'intera pagina per trovare il riferimento (con possibilità, oltretutto, che la parola cercata non sia in realtà presente in quella pagina, per errore

⁴² Cfr. L. Maconi, *Retrodatazioni lessicali*, cit., p. 23.

⁴³ Cfr. <http://www.hathitrust.org/> (ultimo accesso: agosto 2015).

di lettura dell'OCR)⁴⁴.

Ho dunque interrotto la verifica in HathiTrust e ho ritenuto più produttivo effettuare controlli all'estero nel corpus di Google Books, dall'Europa e dall'America, avviando l'interrogazione dalle diverse maschere di ricerca nazionali, per capire se e quanto i risultati ottenuti cambiassero in base al luogo di interrogazione⁴⁵. Ebbene: le retrodatazioni acquisite con Google in Italia sono state confermate anche all'estero. Segnalo tuttavia, come già anticipato (cfr. § 3), che negli Stati Uniti si ha visualizzazione completa per molte fonti che, benché ormai fuori diritto d'autore, in Europa sono ancora protette⁴⁶.

Faccio infine presente che, riverificando in Google alcune retrodatazioni, a nove mesi di distanza dalla raccolta, ho avuto aggiornamenti per circa il 20% del materiale (si è trattato di un miglioramento di qualche anno o decennio: più importante era stato il risultato raggiunto dalla prima ispezione nella Biblioteca digitale di Google Libri).

9. PER UN AGGIORNAMENTO SISTEMATICO DELLE DATE DI PRIMA ATTESTAZIONE DELLE PAROLE NEL VOCABOLARIO ITALIANO

Un progetto di revisione completa delle date di prima attestazione delle parole non può prescindere da Google e, a mio avviso, potrebbe prendere le mosse dall'aggiornamento del lemmario di un dizionario italiano dell'uso in unico volume: non certo perché questa sia la sede idonea per la raccolta delle datazioni, ma perché mi sembra utile partire da un lemmario-base rappresentativo della lingua italiana, idealmente racchiuso in quelle 100.000 parole (circa) del dizionario monovolume. Tale lemmario è grossomodo così composto: 10.000 parole del X-XIII secolo, 13.000 parole trecentesche,

⁴⁴ Critico nei confronti di HathiTrust è anche A. Petrucciani, *Con Robert Darnton nella giungla digitale*, cit., pp. 100-1. Come pregio di HathiTrust segnalo la possibilità di fare ricerche con caratteri jolly (sembra inoltre che qui si possano dividere i volumi italiani da quelli latini, ma, in realtà, i libri latini schedati da Google sotto la categoria "italiano" rimangono tali anche per HathiTrust).

⁴⁵ Ringrazio Benedict Buono e Eugenio Refini, che gentilmente mi hanno aiutato in questa verifica da Spagna e Stati Uniti; io stessa ho compiuto il controllo in Francia e in Svizzera (la verifica è stata effettuata su un piccolo campione di parole, che però non ha mai dato esiti discordanti).

⁴⁶ Alberto Petrucciani (*Ancora su Google e la giungla digitale: altri mestieri, novità e tendenze*, in «Bollettino AIB. Rivista italiana di biblioteconomia e scienze dell'informazione», 52, n. 2, 2012, p. 199), per aggirare le limitazioni di visualizzazione, suggerisce di effettuare le ricerche da un servizio *proxy* considerato da Google americano.

4.000 parole del Quattrocento, 8.000 del Cinquecento, 11.000 parole del XVII-XVIII secolo, oltre 16.000 parole ottocentesche, oltre 28.000 novecentesche, qualche centinaio del XXI secolo (questi i dati dello *Zingarelli 2014*, recuperati dalla versione elettronica tramite l'opzione "datazione", selezionabile nella maschera di ricerca avanzata).

Data tale composizione, è palese che produttivo sarà l'aggiornamento delle 45.000 parole attualmente considerate otto-novecentesche, parole che costituiscono circa il 45% del dizionario, e che in buona parte possono trovare nuova data, cambiando talora anche secolo di appartenenza. Per lo studio e controllo di questo lessico, Google è lo strumento migliore. Considerata la mole e difficoltà del lavoro, nella ricerca sarà bene partire dalle parole invariabili (che creano ovviamente meno problemi nelle modalità di interrogazione della banca-dati) e dai tecnicismi, con i quali raramente si incontrano errori di significato nelle fonti recuperate da Google⁴⁷.

Meno produttivo, sia per quantità sia per entità e facilità della retrodatazione, sarà invece l'aggiornamento del lessico antico: quello due-trecentesco è da affidare al *TLIO*, al *LEI* e al Battaglia; quello cinque-settecentesco (circa 19.000 parole del dizionario monovolume) è verificabile in Google, ma i dati raccolti dovranno essere utilmente integrati con quelli, sicuramente più numerosi, recuperabili mediante una parallela ricerca con *TestRetro* su file di testo disponibili (cfr. § 7)⁴⁸. Limite in questo lavoro sarà la disponibilità (e affidabilità) di file in formato "solo testo" da immettere per l'interrogazione. Tutto il materiale fornito da *Liber Liber* è sfruttabile⁴⁹, ma sarebbe importante soprattutto la collaborazione di quanti hanno allestito edizioni di testi, o digitalizzato materiale cinque-settecentesco, e anche primo-ottocentesco,

⁴⁷ Delle parole attualmente considerate otto-novecentesche, ho per ora esaminato forestierismi, parole della medicina, dell'agricoltura, della botanica, della zoologia, ecc., estraendole dal lemmario dello *Zingarelli* in versione elettronica, attraverso le opzioni previste dalla maschera di ricerca avanzata.

⁴⁸ Per il lessico antico, dal X al XV secolo, saranno da schedare e accogliere anche tutti i lavori pubblicati in varie sedi su retrodatazioni da fonti antiche non migliorabili con Google: es. Davide Puccini, *Retrodatazioni quattrocentesche*, in «Lingua nostra», LXVII, n. 1-2, 2006, pp. 57-58; Id., *Nuove retrodatazioni quattrocentesche*, in «Lingua nostra», LXVII, n. 3-4, 2006, pp. 80-83 (retrodatazioni ricavate in particolare da Pulci e Poliziano).

⁴⁹ Cfr. www.liberliber.it (ultimo accesso: agosto 2015). In questo lavoro, si presti attenzione a dividere il materiale della fonte antica dalle prefazioni e note dell'editore moderno (spesso presenti nei file di *Liber Liber*), le quali potrebbero creare errori nella lettura dei dati (scambiando parole moderne per lessico antico), e si ricordi che i file di *Liber Liber* hanno bassa affidabilità (quindi ogni lezione restituita andrà controllata sull'edizione di riferimento). Per la prima potenziale causa d'errore, cfr. gli esempi da me forniti in L. Maconi, *Retrodatazioni lessicali*, cit., p. 19, nota 5.

per biblioteche elettroniche o per lavori filologici⁵⁰. Mi riferisco ad esempio ai testi a suo tempo preparati per il Ci-Bit (*Centro interuniversitario - Biblioteca italiana telematica*), per Biblioteca italiana, e più recentemente per il CEOD (*Corpus epistolare ottocentesco digitale*) e per AITER (*Archivio Italiano Tradizione Epistolare in Rete*)⁵¹. Fonti privilegiate dovrebbero essere quelle meno spogliate dai lessicografi (es. epistolari, testi tecnici e giornali).

Utile può essere inoltre l'esame informatico di pubblicazioni specialistiche del Novecento, per gli anni ancora coperti da diritto d'autore (in Google presenti in visualizzazione protetta, parziale o addirittura del tutto coperta). Giornali, paraletteratura, testi politici, giuridici, divulgativi e scientifici di questi anni più recenti saranno presto interrogabili nel "Corpus di riferimento per un Nuovo Vocabolario dell'Italiano moderno e contemporaneo", in allestimento sotto la direzione di Claudio Marazzini⁵². Seguendo queste vie di ricerca, ritengo che possano procedere con successo gli studi sulla storia e datazione del nostro lessico.

⁵⁰ Un invito a raccogliere a livello nazionale quanto già disponibile in digitale è stato fatto, seppur non per questo mio specifico fine di ricerca, da A. Petrucciani, *Con Robert Darnton nella giungla digitale*, cit., pp. 103-4.

⁵¹ Cfr. <http://www.bibliotecaitaliana.it/>; <http://ceod.unistrasi.it/>; <http://aiter.unipv.it/> (ultimo accesso: agosto 2015).

⁵² Si tratta del PRIN 2012 *Corpus di riferimento per un Nuovo Vocabolario dell'Italiano moderno e contemporaneo. Fonti documentarie, retrodatazioni, innovazioni* (coordinatore nazionale: Claudio Marazzini, Università del Piemonte Orientale). Cfr. la presentazione del progetto pubblicata in questo stesso volume di Atti. Ringrazio infine Matteo Motolese e Lorenzo Tomasin per i preziosi consigli.

INDICE

<i>Introduzione</i> di Claudio Marazzini e Ludovica Maconi	pag.	5
<i>Programma della Piazza delle Lingue 2014</i>	»	13

LA LESSICOGRAFIA DEL FUTURO

Raffaele Simone, <i>Il dizionario del futuro</i>	»	17
Luca Serianni, <i>Ha un futuro il dizionario dell'uso?</i>	»	33

L'ITALIANO NEGLI ARCHIVI DELLA RETE

Daniele Fusi, <i>Edizione digitale del Tommaseo-Bellini</i>	»	49
Ludovica Maconi, <i>Retrodatazioni lessicali con Google Libri: opportunità e inganni della Rete</i>	»	73
Silvia Calamai - Pier Marco Bertinetto, <i>L'unità documentale negli archivi sonori</i>	»	95
Carla Marellò, <i>Alla ricerca della finestra d'interrogazione amichevole ma completa</i>	»	113

CONFRONTO DI METODI: IL LATINO TARDOANTICO IN RETE

Raffaella Tabacco, <i>La prosa latina pagana tardoantica e la Biblioteca digitale digilibLT</i>	»	125
Laura Mosca - Simona Musso, <i>Costruire una biblioteca digitale di testi latini tardoantichi: problemi e scelte</i>	»	143

IL PROGETTO
 “CORPUS DI RIFERIMENTO PER UN NUOVO VOCABOLARIO
 DELL’ITALIANO MODERNO E CONTEMPORANEO.
 FONTI DOCUMENTARIE, RETRODATAZIONI, INNOVAZIONI”
 (PRIN 2012)

Claudio Marazzini, <i>Presentazione del progetto PRIN 2012 e compiti dell’unità dell’UPO nel settore del linguaggio scientifico</i>	pag. 161
Gabriella Alfieri - Giovanna Alfonzetti - Rosaria Sardo, <i>Il corpus dell’unità di Catania: paraletteratura e divulgazione, galatei e libri per ragazzi</i>	» 171
Ilaria Bonomi, <i>Il corpus dell’unità di Milano: giornali, fumetti e libretti d’opera</i>	» 193
Lorenzo Coveri, <i>Unità di Genova: l’italiano in versi. Per un corpus dell’italiano del melodramma, della canzone e della poesia del nuovo secolo</i>	» 199
Maria Vittoria Dell’Anna - Elisabetta Marinai - Francesco Romano - Jacqueline Visconti, <i>Un corpus di testi giuridici per il Nuovo Vocabolario dell’Italiano moderno e contemporaneo: il patrimonio dell’unità ITTIG di Firenze e altre risorse digitali</i>	» 223
Riccardo Gualdo - Maria Vittoria Dell’Anna - Stefano Telve, <i>Il corpus dell’unità di Viterbo: scartafacci parlamentari. Filologia del parlato trascritto e stenografato</i>	» 239
Rita Librandi, <i>Il corpus dell’unità di Napoli “L’Orientale”: scritti dei fondatori del cattolicesimo sociale. Un vocabolario documenta la storia</i>	» 251
Marco Biffi, <i>Unità di Firenze: progettare il corpus per il vocabolario postunitario</i>	» 259
<i>Indice dei nomi</i>	» 281