

Ludovica Maconi

Archidata

Archivio elettronico di retrodatazioni lessicali

Storia del progetto e linee guida



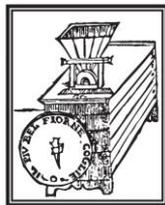
Accademia della Crusca

Ludovica Maconi

ArchiDATA

Archivio elettronico di retrodatazioni lessicali

Storia del progetto e linee guida



Accademia della Crusca

2017

© 2017 Accademia della Crusca, Firenze
ISBN 978-88-89369-84-5
Novembre 2017

Sommario

Progetto	5
Il sito di ArchiDATA	10
Accesso al materiale e maschera di ricerca	12
Struttura delle voci	19
Il caricamento dei dati	27
Prova di monitoraggio	29
Sviluppi	33

Progetto

Tra il 2014 e il 2016, in collaborazione con l'Accademia della Crusca e l'editore Zanichelli, mi sono occupata di retrodatazioni lessicali¹. Come noto, le date di prima attestazione fornite dai repertori etimologici di riferimento e dal GRADIT sono bisognose di una revisione, da attuare anche per mezzo di nuovi strumenti informatici. Frutto della ricerca sono state 4.000 nuove datazioni confluite nello *Zingarelli 2017 e 2018*².

Pensando agli sviluppi futuri della lessicografia *online*, e ritenendo inadatta (e superata), per le retrodatazioni lessicali, l'edizione su carta³, sto rendendo pubblico il materiale rac-

¹ Strumento principale di questa ricerca è stato Google Libri, ma mi sono servita anche di Internet Archive, degli archivi elettronici dei principali quotidiani nazionali, di corpora di testi non letterari allestiti da varie unità di ricerca per fini non solo lessicografici, e di programmi per la retrodatazione automatica di parole: *Ralip* (di Gianluca Biasci, cfr. <http://ralip.eu/> - ultimo accesso nel novembre 2017) e *TestRetro* (di Daniele Fusi).

² Di queste 4.000 retrodatazioni, 3.000 vengono da mie raccolte di prima mano, 1.000 sono state acquisite dagli studi elencati nella pagina di *Bibliografia* del sito ArchiDATA, ma solo dopo controllo in Google Libri, alla ricerca di eventuali miglioramenti. Nuove retrodatazioni sono in preparazione per lo *Zingarelli 2019*. Ringrazio Mario Cannella, curatore delle riedizioni annuali dello *Zingarelli*, per i preziosi consigli e per l'appoggio nelle ricerche lessicografiche.

³ Ovviamente, a meno che non ci si distacchi dal mero dato cronologico e si costruisca invece un discorso sulla storia della parola (questo sarà il secondo passo della ricerca qui avviata). Ben noti sono i limiti dell'edizione su carta: costi e tempi di stampa, invecchiamento del risultato, scarsa reperibilità, impossibilità di aggiornamenti. Ho ritenuto inadatta la pubblicazione in volume anche perché il prodotto finale sarebbe stato un dizionario, un libro di schede. Inoltre, gli studiosi avrebbero avuto accesso al materiale

colto in un archivio elettronico da me ideato, liberamente consultabile dal sito dell'Accademia della Crusca: nella sezione *Scaffali digitali*, accanto alle *Crusche in Rete* e al *Tommaseo online*, è disponibile da qualche mese *ArchiDATA - Archivio di (retro)datazioni lessicali*⁴. Le nuove datazioni sono accompagnate dalla riproduzione fotografica della fonte (scelta di sicura utilità, innovativa rispetto alla tradizione, e resa possibile grazie alle moderne tecnologie), oppure sono seguite dal rinvio alla bibliografia specifica (in base alle modalità spiegate a p. 19, cfr. anche fig. 10). Importante novità, che andrà ampliandosi nel corso del lavoro, è la datazione delle accezioni. Parole e accezioni sono marcate per ambito d'uso, quindi è possibile accedere all'archivio sezionando variamente il lemmario.

Sotto il controllo della Crusca, ArchiDATA vorrebbe assumere il ruolo di centro di raccolta di retrodatazioni, utilizzando i contributi che da lungo tempo escono in sedi disperate, con inevitabile dispersione di risultati che sarebbe invece utile

solo tramite ricerca alfabetica per lemma, mentre in un archivio elettronico è possibile arrivare ai dati (e valorizzarli) attraverso molteplici chiavi di accesso: datazione, autore, ambito d'uso, prefissi, suffissi, ecc. (cfr. *Accesso al materiale e maschera di ricerca*, pp. 12-18). Questi dunque i vantaggi della Rete, già messi in evidenza da grandi imprese come il TLIO: 1. il lavoro può essere pubblicato senza dover attendere la conclusione del progetto, mostrando l'avanzamento delle ricerche; 2. non si è vincolati al rispetto dell'ordine alfabetico; 3. è possibile tornare su voci già pubblicate, sia per correzioni sia per aggiunte (ess. datazioni di accezioni precedentemente assenti, rimandi bibliografici a nuovi saggi sulla storia della parola); 4. il materiale può essere ampliato a volontà, senza limitazioni di sorta e senza costi di stampa; 5. è possibile accedere ai dati attraverso più canali.

⁴ Cfr. <http://www.accaemiadellacrusca.it/it/scaffali-digitali/archidata-archivio-datazioni-lessicali> (ultimo accesso nel novembre 2017). Il sito è raggiungibile anche all'indirizzo www.archidata.info.

convogliare in un'unica sede, una sede accessibile, aggiornabile, riordinabile e ampliabile. L'esigenza di un centro unico di raccolta e aggiornamento a fini lessicografici è stata rilevata anche da Yorick Gomez Gane, che, su scala più ampia, ha dato avvio al cantiere dell'AVSI (*Archivio per il Vocabolario Storico Italiano*)⁵. Avvalendosi della collaborazione di studiosi che seguono la ricetta miglioriniana *nulla dies sine schedula*, ArchiDATA andrà dunque progressivamente arricchendosi di nuovi documenti per la storia delle nostre parole, con particolare attenzione per il lessico moderno e contemporaneo (che è quello più suscettibile di retrodatazione), e per i linguaggi settoriali⁶.

L'idea che ha portato alla realizzazione di ArchiDATA è maturata pensando innanzitutto al concetto di "dinamicità" che la Crusca vuole dare ai suoi prodotti lessicografici, in particolare al VoDIM (*Vocabolario dinamico dell'italiano moderno*). La *dinamicità* presuppone molteplicità di chiavi d'accesso, facilità di aggiornamento, collegamenti ipertestuali, pluralità di fonti: elementi tutti che la pubblicazione su carta non può offrire, ma che invece sono caratteristiche del sito di ArchiDATA. Stretto è stato inoltre il legame, almeno ideale, tra ArchiDATA e lo «schedario» di prime attestazioni di Migliorini, quella «cassettera fatta costruire nel 1941 a misura

⁵ Cfr. Y. Gomez Gane, *Una nuova rivista lessicografica: l' "Archivio per il vocabolario storico italiano" ("AVSI")*, in "Studi di lessicografia italiana", vol. XXXII, 2015, pp. 263-274. Il primo numero della rivista è previsto in uscita per il 2018.

⁶ Benché il progetto sia nato pensando al lessico moderno, si sta ora avviando una collaborazione con il TLIO per convogliare in ArchiDATA le retrodatazioni di lessico antico raccolte dalla redazione dell'OVI. Ringrazio Lino Leonardi per l'interessamento e per la disponibilità a lavorare insieme in questa direzione.

delle sue personalissime schede di 10 x 11 cm», come ha descritto il figlio Paolo in un ricordo, raccontando che il padre, in un'epoca in cui ancora non esistevano i computer, «di ogni parola sapeva dire in pochi istanti, consultando il suo schedario, la data di nascita e il nome di chi quella parola aveva usata per primo»⁷: ArchiDATA si prefigge questo stesso compito. Vorrei infine legare l'ideazione anche alla suggestione di immagini familiari a chi ha interesse per studi di storia della lessicografia: nel periodo in cui più intense si facevano le mie schedature per lo *Zingarelli*, mi sono corsi alla memoria i noti versi di Riccardi di Lantosca che ritraggono Tommaseo intento a lavorare al grande Dizionario Pomba, mentre «ordinava certi quadrellini / di carta, esaminandoli col piglio / della civetta prima che avvicini / al becco l'uccellin che ha nell'artiglio»⁸. I «quadrellini di carta», avidamente trattiene e consultati da Tommaseo, così come le schede cartacee meticolosamente compilate nel secolo scorso da Migliorini, si tramutano oggi nelle voci elettroniche di ArchiDATA.

L'acronimo ArchiDATA è coniato mediante composizione delle parole *archivio* e *datazioni* (da intendersi come iperonimo di *retrodatazioni*, e da sottintendersi *lessicali*)⁹. C'è poi

⁷ P. Migliorini, *Un ricordo di mio padre*, in *Bruno Migliorini, l'uomo e il linguista*, Atti del convegno di studi (Rovigo, 11-12 aprile 2008), a cura di M. Santipolo e M. Viale, Rovigo, Accademia dei Concordi Editore, 2009, p. 4. Lo schedario di Migliorini, che una volta si trovava nella casa del Maestro in via La Marmora, è oggi custodito in Accademia della Crusca, accanto alla celebre sala delle pale. L'avvio dei lavori di ArchiDATA potrebbe anche essere una buona occasione per riesaminare questo prezioso materiale lasciato da Migliorini.

⁸ V. Erdiel [Vincenzo Riccardi di Lantosca], *Pape Satan Aleppe. Macchietta*, Assisi, Tipografia Froebel del Collegio Principe di Napoli, 27 settembre 1882, p. 17.

⁹ Per la scelta del nome, ringrazio i professori Massimo Fanfani e Luca Serianni, con i quali ho potuto utilmente confrontarmi, traendone preziosi

una doppia valenza, perché *archi-* coincide con il prefissoide che in parole composte dotte significa ‘primo’ o ‘capo’, indicando quindi un *primato*: nel nostro caso, un primato di origine e antichità; *-DATA*, oltre a riferirsi alle *datazioni*, indirettamente richiama i *data*, la dimensione elettronica dell’archivio, le informazioni elaborate in digitale e consultabili tramite motore di ricerca¹⁰.

consigli, durante un colloquio per la presentazione di ArchiDATA in Accademia della Crusca.

¹⁰ Modello per coniare il nome ArchiDATA è stato anche *Archilet* (*Archivio delle corrispondenze letterarie di età moderna*; cfr. <http://www.archilet.it/HomePage.aspx>, ultimo accesso nel novembre 2017; progetto coordinato da C. Carminati, P. Procaccioli ed E. Russo): in anni in cui proliferano lavori umanistici volti a fornire banche dati e strumenti di ricerca pubblicati in Rete, sarebbe funzionale, a mio avviso, una maggiore uniformità nella nomenclatura di tali nuovi prodotti (raggruppabili per tipologie: archivi, corpora di testi, dizionari ecc.) e, soprattutto, sarebbe importante un sito che raccogliesse l’elenco, gli indirizzi e una breve descrizione di queste preziose nuove risorse, talvolta poco conosciute e solo parzialmente sfruttate.

Il sito di ArchiDATA

Nelle pagine che seguono descrivo il funzionamento dell'archivio elettronico di retrodatazioni da me progettato. La parte informatica, eseguita sempre secondo mio disegno, è stata realizzata dall'ingegnere Michele Lavezzi.

Dalla *homepage* di ArchiDATA (www.archidata.info) si accede al materiale dell'archivio tramite la maschera di ricerca, oppure scorrendo il lemmario generale. In prima pagina è dichiarato il numero totale di parole raccolte, aggiornato automaticamente ad ogni nuova immissione¹¹. Altre pagine del sito sono dedicate alla descrizione del progetto, alla presentazione di autori e collaboratori¹², alla bibliografia e ai contatti. La pagina di *Bibliografia*, che come l'archivio è in continuo aggiornamento, è divisa in due sezioni, le quali registrano le progressive acquisizioni di ArchiDATA: studi e fonti. Nella sezione "Studi" sono elencati libri e articoli che raccolgono retrodatazioni utili non confluite nel DELIN e nel GRADIT, ma ora accolte (o in fase di inserimento) in ArchiDATA¹³. Quando opportuno, queste pubblicazioni sono richiamate

¹¹ Al momento è stato avviato un primo caricamento; gli inserimenti continueranno e verranno quantitativamente monitorati ogni sei mesi. Si tenga presente che il numero indicato non include le datazioni delle accezioni, archiviate sotto lo stesso lemma.

¹² La direzione del sito è di Claudio Marazzini. Il comitato scientifico è composto da Vittorio Coletti, Paolo D'Achille, Ludovica Maconi e Claudio Marazzini.

¹³ Per rendere questa sezione quanto più possibile completa, si procederà a una sistematica raccolta bibliografica sul tema delle retrodatazioni. Per segnalare eventuali lacune, i collaboratori e gli studiosi interessati al progetto possono scrivere ad archidata.crusca@gmail.com.

nelle voci dell'archivio. La sezione “Fonti” elenca invece le “fonti d'autore”, o i “testi miniera”, cioè quei testi selezionati dai responsabili scientifici (membri delle unità di ricerca del PRIN 2015) per essere spogliati elettronicamente con il *Ralip* di Gianluca Biasci o con *TestRetro* di Daniele Fusi (programmi che estraggono automaticamente retrodatazioni da testi in formato RTF immessi per la ricerca)¹⁴.



Fig. 1. Pagine del sito ArchiDATA (www.archidata.info)

¹⁴ Su questi strumenti cfr. L. Maconi, *Retrodatazioni lessicali con Google Libri: opportunità e inganni della Rete*, in *L'italiano elettronico. Vocabolari, corpora, archivi testuali e sonori*, Atti della Piazza delle lingue 2014 (Firenze, 6-8 novembre), a cura di C. Marazzini e L. Maconi, Firenze, Accademia della Crusca, 2016, pp. 73-93 (in particolare pp. 88-90). Primi testi ad essere spogliati elettronicamente saranno quelli digitalizzati nel PRIN 2012 e che costituiscono il corpus del VoDIM. Si tratta di testi settoriali di scienza, oratoria parlamentare, stampa periodica, cucina, paraletteratura: settori tutti poco indagati finora dai lessicografi, e dai quali potranno dunque venire aggiornamenti utili. Parallelamente, si procederà allo spoglio elettronico di testi più antichi, che forniranno retrodatazioni meno suscettibili di successivo aggiornamento nel corpus di Google Libri. Prove per queste due vie (su testi scientifici del corpus del VoDIM e su testi cinquecenteschi) sono state da me già avviate durante la fase di raccolta di retrodatazioni, ottenendo buoni risultati: cfr. *ibidem*.

Accesso al materiale e maschera di ricerca

ArchiDATA è consultabile attraverso motore di ricerca che permette di accedere ai dati tramite:

- Ricerca alfabetica: si può scorrere il lemmario di ogni singola lettera (cliccando sulla sezione desiderata), o aprire il lemmario generale.
- Ricerca lessicale: anche con caratteri jolly, ess. *logo, capo*, contro*¹⁵.



Fig. 2. Maschera di ricerca alfabetica e ricerca per lemma

¹⁵ Da ricerche di questo tipo si può ad esempio osservare che le professioni mediche con suffisso *-logo*, *biologo*, *ematologo*, *embriologo*, *ginecologo*, *oftalmologo*, quelle attualmente schedate, trovano prima attestazione attorno agli anni Trenta dell'Ottocento (alcune di queste parole nel DELIN hanno invece datazione molto più tarda, novecentesca). Osservazioni generali, di più ampio respiro e attendibilità, potranno farsi solo quando il materiale schedato raggiungerà numeri più consistenti. Anche per i composti di *capo-* (*capofabbrica*, *capofficina*, *caporedattore*, *capostazione*, *capotribù* - attestato in una traduzione di W. Scott) la datazione risale alla prima metà dell'Ottocento; inoltre, si può osservare che in queste fonti antiche le grafie sono perlopiù con trattino, solo *caporedattore* è univertato (l'attestazione viene da un giornale del 1838, "La voce della verità. Gazzetta della Italia Centrale"; nel DELIN *caporedattore* è datato 1955, nel GRADIT 1962).

Selezionando l'opzione "ricerca avanzata":

- Ricerca cronologica.
- Ricerca per autore.
- Ricerca per ambito d'uso.
- Estrazione di forestierismi.

Questi campi di ricerca sono ovviamente interrogabili singolarmente o in abbinamento. Sarà ad esempio possibile chiedere al motore di ricerca di estrarre le parole della medicina (ambito d'uso) dell'Ottocento (ricerca cronologica), e, volendo circoscrivere ulteriormente l'insieme dei risultati, si potrà verificare se alcune di queste prime attestazioni si trovano in testi di Cesare Lombroso o di Giulio Bizzozero, inserendo il nome dello scienziato nel campo "autore" (cfr. fig. 3). In questa prima fase di caricamento, ArchiDATA contiene 66 parole della medicina, delle quali 52 ottocentesche, e una di queste, *algometria* 'misurazione dell'intensità della sensazione dolorosa', con attestazione più antica in un saggio di Lombroso del 1867. Nel GRADIT *algometria* è datata 1942.

Ricerca: [chiudi](#) 🔍

da: a: [?](#) autore: ambito d'uso: ▾

Filtri:

Lemmi 1

[algometria](#) (1867)

Fig. 3. Maschera di ricerca avanzata

Si osservi che togliendo dalle impostazioni di questa ricerca il filtro “ambito d’uso - medicina”, ArchiDATA restituisce un’altra parola con prima attestazione in Lombroso: *ebefrenia* ‘forma di schizofrenia’, parola marcata con “ambito d’uso - psicologia” e per questo non restituita nella prima ricerca. Di Bizzozero non sono attualmente schedate prime attestazioni di termini della medicina, ma dal suo saggio *Contro la tubercolosi* (1899) viene la retrodatazione di una parola di uso comune: *portasciugamani* (nel GRADIT datato sec. XX come sostantivo maschile invariabile; DELIN e GRADIT indicano invece 1970 per la forma *portasciugamano*).

Propongo ancora un esempio di possibile ricerca. Selezionando l’opzione *francesismi* dal menù a tendina dei forestierismi, si estraggono i termini francesi finora archiviati, con la data che attesta il loro ingresso in italiano (cfr. fig. 4).

The screenshot shows a search interface with the following elements:

- A dropdown menu labeled "forestierismi:" with "Francesismi" selected. The menu options are: nessuno, Anglismi, Arabismi, **Francesismi**, Germanismi, Ispanismi, Latinismi e grecismi, and Altre lingue.
- A "Cerca" button.
- A "Filtri:" section with a red 'x' icon and the text "forestierismo".
- A "Lemmi 4" section listing the following terms with their first attestation years:
 - bon ton (1784)
 - brie (1878)
 - chiné (1846)
 - grisaille (1857)

Fig. 4. Estrazione di forestierismi

Tra questi francesismi, *brie*, noto formaggio a pasta molle, ha prima attestazione in Carlo Dossi. Potrà allora sorgere la curiosità di verificare se altre parole della cucina sono datate con testi dello scrittore scapigliato. A questa ricerca, ArchiDATA

restituirà, accanto a *brie*, altri tre lemmi: *mentino* ‘caramella alla menta’, *pirotta* ‘mortaiolo di legno a forma di scodella’ (lombardismo), *secchetto* ‘pezzo di pane secco’¹⁶. Gli esempi qui illustrati sono stati proposti con l’intento, spero raggiunto, di suggerire le potenzialità di ArchiDATA come strumento nello studio del lessico moderno. È ovvio che tali potenzialità saranno pienamente esprimibili quando l’archivio conterrà diverse migliaia di voci.

Ricerca per ambito d’uso

La ricerca per ambito d’uso è disponibile con menù a tendina. Ho circoscritto il numero delle categorie a una ventina di ambiti tecnici, accorpendo discipline affini, per evitare il proliferare di sottoinsiemi che avrebbero segmentato in maniera inutilmente analitica il lemmario. Mi sono per ora limitata a categorie che in parte riflettono gli argomenti dei nove corpora allestiti per il *Vocabolario dinamico dell’italiano moderno* (VODIM)¹⁷.

¹⁶ Attestati tutti e tre nella *Vita di Alberto Pisani* (1870). Nel GRADIT queste parole hanno prima attestazione novecentesca.

¹⁷ I nove corpora raccolgono: testi scientifici (C. Marazzini, Università del Piemonte Orientale), paraletteratura, galatei e libri per ragazzi (G. Alfieri, Università di Catania), testi d’arte e cucina (M. Biffi, Università di Firenze, e U. Vignuzzi, “Sapienza” – Università di Roma), giornali e fumetti (I. Bonomi, Università di Milano), melodramma e canzone (L. Coveri, Università di Genova), discorsi parlamentari e documentari industriali del dopoguerra (R. Gualdo, Università della Tuscia), scritti dei fondatori del cattolicesimo sociale (R. Librandi, Università di Napoli “L’Orientale”), testi giuridici (E. Marinai, ITTIG di Firenze). Cfr. *Il progetto “Corpus di riferimento per un nuovo Vocabolario dell’italiano moderno e contemporaneo”*, in *L’italiano elettronico. Vocabolari, corpora, archivi testuali e sonori*, Atti della Piazza delle Lingue 2014, a cura di C. Marazzini e L.

In questa prima fase di pubblicazione, non tutti i campi di ricerca a disposizione raccolgono materiale (alcuni daranno ricerca nulla), ma sto ovviamente lavorando in prospettiva futura, e sto predisponendo ArchiDATA pensando sia al materiale, già in mio possesso, che dovrà essere inserito nei prossimi mesi¹⁸, sia alle future acquisizioni.

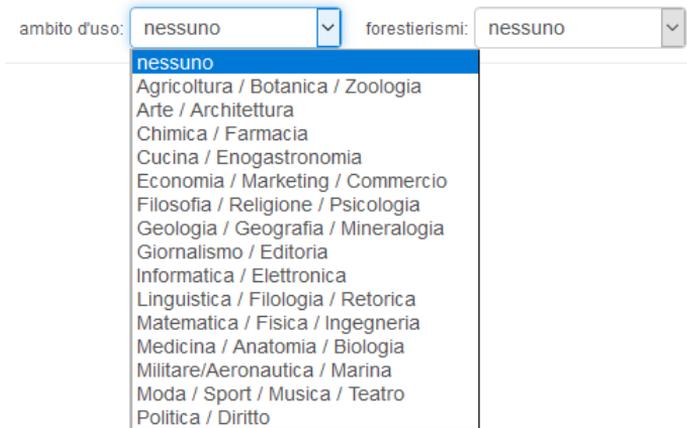


Fig. 5. Menù di ricerca per ambito d'uso

Maconi, Firenze, Accademia della Crusca, 2016, pp. 159-280 (con contributi dei responsabili delle nove unità di ricerca).

¹⁸ Bisogna considerare che il progetto e la raccolta di retrodatazioni sono stati finora portati avanti con le mie sole forze, mentre lavoravo, parallelamente, all'allestimento e marcatura di testi scientifici per il corpus del VoDIM. Prima di procedere alla seconda fase di caricamento, ritengo inoltre importante studiare (come sto facendo insieme al Comitato scientifico di ArchiDATA) le modifiche tecniche da apportare al programma e alle marcature del materiale inserito, per predisporre al meglio questo nuovo strumento.

Ricerca per autore

Per gli autori non è previsto un menù a tendina perché le fonti delle retrodatazioni, com'è facile intuire, sono migliaia. Lo studioso interessato al lessico di un determinato autore, eseguirà la ricerca immettendo direttamente il cognome dell'autore da esaminare¹⁹. Non molte, sul lemmario generale di un comune dizionario dell'uso monovolume (quello da cui è partita questa raccolta), sono le retrodatazioni di età moderna che trovano prima attestazione in scrittori celebri. È più facile che ciò accada per parole di secoli antichi, o per parole di sapore espressivo, presenti ad esempio nei testi letterari della Scapigliatura, o di un autore come l'ingegner Gadda, scrittore ricettivo verso i neologismi. Hanno ad esempio prima attestazione in Dossi *rompitorta* 'rompiscatole' e *margnuccione* 'stupido', in Faldella *faccia da schiaffi*²⁰; sono parole gaddiane, individuate da Luigi Matt con prima attestazione in articoli giornalistici dello scrittore, *antologismo*, *astronautico* e *minutaggio*²¹. Tra le voci retrodate con testi della Scapigliatura, si

¹⁹ Non sono stati marcati i nomi di battesimo. Si dovrà quindi inserire solo il cognome. Per gli autori con lo stesso cognome si procederà manualmente alla verifica.

²⁰ Per le fonti e i passi di riferimento rimando al sito di ArchiDATA. Rilevo che l'attestazione faldelliana di *faccia da schiaffi* (presente in *Figure* 1875) migliora di quasi un secolo la datazione di DELIN e GRADIT, che danno come prima attestazione della locuzione il 1965. Una raccolta di un centinaio di retrodatazioni da autori della Scapigliatura, ottenute con spoglio elettronico dei testi e successivo controllo in Google Libri e in altre risorse elettroniche, si trova in L. Maconi, *Retrodatazioni lessicali da Carlo Dossi e Giovanni Faldella*, in *Sette studi e un croquis*, a cura di F. Mereta e A. Sisti, Vercelli, Edizioni Mercurio, pp. 17-53 (questa raccolta è stata inserita in ArchiDATA).

²¹ Cfr. L. Matt, *Sulle prime attestazioni dei vocaboli nei dizionari*, in "Studi linguistici italiani", vol. XXVIII, f. 2, 2002, pp. 272-297: per le retrodatazioni gaddiane, pp. 289-297; segnalo che alcune retrodatazioni qui

conta un buon numero di parole composte e di parole formate col prefisso *in-*, tra le quali *inzitellonire*, *inciullire* e *immusonito*. Ricordo che nella pagina di *Bibliografia*, alla sezione “Fonti”, è dato l’elenco degli autori e dei testi spogliati in modo sistematico per mezzo di strumenti informatici, e inclusi nell’archivio²².

Per le prime attestazioni ricavate da traduzioni, il nome del traduttore italiano (quando dichiarato nel frontespizio) viene promosso nel campo autore. Compare tra parentesi quadre accanto al nome dell’autore straniero, e nella maschera di caricamento (non visibile all’utente, ma per la quale cfr. fig. 11) è marcato in modo che sia possibile estrarlo tramite ricerca avanzata. Il medico Annibale Omodei, che visse tra Sette e Ottocento e diresse gli “Annali universali di medicina”, è ad esempio presente in ArchiDATA tra i traduttori del *Dizionario dei termini di medicina, chirurgia, veterinaria, farmacia ecc.*, opera trasposta dal francese e stampata a Milano nel 1829. A lemma in questo vocabolario si trova la più antica attestazione di *splenectomia*, termine tecnico che indica l’asportazione della milza (DELIN e GRADIT datano la parola *splenectomia* 1888).

raccolte sono oggi migliorabili consultando l’Archivio storico del quotidiano “La Stampa” (ma non le tre a cui ho fatto riferimento a testo): ad esempio, *catacombale* (DELIN e GRADIT 1965, Gadda 1942) è ulteriormente retrodatabile al 1932 con G. Piazza, *L’altro Faust*, in “La Stampa”, 15 gennaio 1932, p. 3; *intersiderale* (GRADIT 1961, Gadda 1952) è nell’articolo *La pioggia delle stelle filanti*, in “La Stampa”, 11 ottobre 1933, p. 2.

²² Per gli autori schedati in questa pagina, lo studioso che consulta ArchiDATA può essere certo di trovare retrodatazioni in archivio. Sono state al momento raccolte prime attestazioni da Aretino, Castiglione, Dossi e Faldella tra i letterati; tra gli scienziati sono stati esaminati Bizzozzero, Golgi e Lombroso, limitatamente ai testi digitalizzati per il corpus del VoDIM. Per alcuni di questi autori le voci saranno caricate nei prossimi mesi, cfr. n. 18.

Struttura delle voci

I lemmi compaiono incolonnati a sinistra dello schermo, affiancati dalla più antica datazione recuperata²³. Selezionando il lemma desiderato, sulla destra dello schermo si apre la voce, strutturata in quattro campi fissi (e un quinto campo di note, non sempre presente):

1. Definizione.
2. Nuova (retro)datazione: in grassetto, seguita, tra parentesi quadre, dalla (superata) data del DELIN e del GRADIT; talvolta compaiono anche il Battaglia e *L'Etimologico* di Nocentini, quando l'indicazione fornita in questi repertori è migliore o colma una lacuna.
3. Fonte della retrodatazione.
4. Firma del collaboratore e data di inserimento.

Per le retrodatazioni attinte in Google Libri, un *link* rimanda alla riproduzione fotografica della fonte. Per le retrodatazioni accolte da altri è invece trascritto il passo di riferimento accanto alla fonte, senza fotografia, e in un campo aggiuntivo di note e commenti è indicato il rimando bibliografico alla pubblicazione da cui è stata ricavata l'informazione. Queste retrodatazioni acquisite di seconda mano sono inserite in Archi-DATA specificando che si tratta di "Scheda di redazione".

²³ Per i lemmi con più accezioni, in questo campo figura solo la datazione del significato più antico, le altre si recuperano aprendo la voce.

Per le parole con più accezioni, la struttura è sempre la medesima, ma la voce è formata da più schede, una per ogni significato. Le accezioni compaiono, numerate, in ordine cronologico, dalla più antica alla più recente. È possibile aggiungere al lemma tante schede quanti i significati che si vogliono datare (queste aggiunte possono essere fatte in momenti diversi; il programma le riordina automaticamente). Porto qui l'esempio della voce *androceo*, formata da due schede: la prima per il significato botanico ('insieme degli stami di un fiore'), che attualmente ha prima attestazione nel 1829; la seconda per il più conosciuto significato architettonico, indicante la 'parte della casa greca riservata agli uomini', retrodatato agli anni Settanta dell'Ottocento. Per questo secondo significato il DELIN indica come prima attestazione il Vocabolario della Reale Accademia d'Italia (1941). ArchiDATA migliora quindi la datazione di oltre mezzo secolo. Si osservi che le due schede sono state da me inserite in momenti diversi, e che la seconda è frutto di una ricerca in Google seguita da controllo in biblioteca sulla prima edizione della fonte, al fine di affinare il risultato: Google restituisce, infatti, come attestazione più antica, la seconda edizione del saggio di Marco Monnier *Pompei e i pompeiani* (1875, p. 79); verificata la presenza dello stesso passo nella prima edizione, del 1873 (a p. 74), la datazione è stata ulteriormente migliorata. Come mostrato nella figura 6, cliccando sulla fonte della retrodatazione si apre la foto di frontespizio e passo di riferimento (la parola in esame è evidenziata in giallo). Un altro esempio di voce con più accezioni è riportato in figura 10.

ANDROCEO

1 s. m. (*bot.*) insieme degli stami di un fiore.

1829 [DELIN e GRADIT 1865]

- G. Moretti e C. Chiolini, *Istruzione teorico-pratica nell'arte de' giardini di piace* Stella e figli, 1829, p. 159.

[Ludovica Maconi - 13/07/2017]

2 s. m. parte della casa greca riservata agli uomini.

1873 [DELIN 1941]

- M. Monnier, *Pompei e i pompeiani*, Milano, Treves, 1873, p. 74 [II edizione, Milano, Fratelli Treves, 1875, p. 79].

[Ludovica Maconi - 07/09/2017]



Fig. 6. Voce androceo in ArchiDATA

Talvolta sono state archiviate più fonti sotto uno stesso significato per dar conto di successivi miglioramenti nella datazione del lemma (si vedano, nel sito, esempi alle voci *bubbonico*, *benzoile*, *assordimento*), oppure per documentare diverse grafie della parola: *capofficina*, parola che nel GRADIT e nel DELIN ha datazione novecentesca (1941), è retrodatato al 1846 nella grafia *capo-officina*, al 1867 con grafia univervata. L'indicazione della grafia è in grassetto tra parentesi quadre, subito accanto al rimando alla fonte (cfr. fig. 7).

- *Nuovo dizionario universale tecnologico o di arti e mestieri*, prima traduzione italiana, tomo XL, Venezia, G. Antonelli, 1846, p. 261, s.v. *morsa*. [**capo-officina**]

- *L'Esposizione universale del 1867 illustrata*, redattore in capo Francesco Ducuing, vol. I, Milano - Firenze - Venezia, Sonzogno, 1867, p. 296. [**capofficina**]

Fig. 7. Ritaglio della voce capofficina con indicazione di due diverse grafie

Campo di note

La sezione facoltativa di note e commenti è usata principalmente per i rimandi bibliografici agli studi raccolti in *Bibliografia* (nei casi in cui la retrodatazione è acquisita di seconda mano). Il “campo note”, però, può anche contenere informazioni più dettagliate sulla storia della parola, o avvertimenti di altro tipo. Sotto alcune voci si trovano note filologiche: così, s.v. *parlottio*, si avverte che la retrodatazione viene dalla seconda edizione della *Desinenza in A* di Dossi, e non dalla prima, nella quale l’autore aveva invece usato *cicioramento*²⁴.

Sono inoltre possibili note più estese e discorsive. È l’estensore della voce che sceglie se limitarsi al dato cronologico o se approfondire la trattazione. Porto l’esempio della voce *ubiquo*, retrodatata al 1871, aggiornando le datazioni novecentesche di DELIN e GRADIT (cfr. fig. 8). Il campo di note approfondisce spiegando che le più antiche occorrenze di *ubiquo* risalgono al XIX secolo, ma che il termine ha avuto maggior diffusione a partire dal Novecento: documentato inizialmente in testi filosofici (Z. Zini, P. Orano, L. Giusso), è stato poi impiegato da Gadda (anche con inedita valenza di sostantivo), e nella seconda metà del secolo scorso si è diffuso nella prosa giornalistica. La nota prosegue informando che più antichi di *ubiquo*, e di uso settoriale storico-religioso e teologico, sono i tecnicismi *ubiquitario* e *ubiquista*, con i quali si indicano gli appartenenti a una corrente protestante luterana

²⁴ Il riscontro è stato fatto sul testo curato da Dante Isella: C. Dossi, *Opere*, a cura di D. Isella, Milano, Adelphi, 1995, con apparato critico delle varianti (riferimenti per *parlottio* alle pp. 871 e 1504). Altre parole che hanno prima attestazione nella *Desinenza in A* sono state invece schedate in ArchiDATA con la data di pubblicazione della prima edizione dell’opera (1878): *ess. intrasparire, fracassosità e socchiave*. Cfr. L. Maconi, *Retrodatazioni lessicali da Carlo Dossi e Giovanni Faldella*, cit.

(le prime attestazioni dei due tecnicismi si trovano in traduzioni settecentesche dal francese). Infine, in coda a queste brevi notizie sulla storia della parola, si rimanda alla bibliografia di riferimento.

UBIQUO

agg. che ha la facoltà o il dono dell'ubiquità.

1871 [GRADIT 1946; DELIN 1963]

- Henry Yule [trad. it. di Guglielmo Berchet], *Marco Polo e il suo libro*, in "Archivio veneto", vol. I, t. II, parte II, 1871, p. 350.

Sebbene la più antica attestazione di *ubiquo* risalga al XIX secolo, il termine ha
documentato in testi filosofici (Z. Zini, P. Orano, L. Giusso), è stato poi impiegato
si è diffuso anche nella prosa giornalistica. Più antichi di *ubiquo*, e di uso settor
ubiquitario e *ubiquista* (con prime attestazioni in traduzioni francesi del Settec
"Lingua e Stile", LI, fasc. 2, 2016, pp. 307-313. Attestazione di *ubiquo* come s
Invenzioni lessicali gaddiane. Glossarietto di "Eros e Priapo", in "I quaderni d
97-182 (*ubiquo* a p. 178).

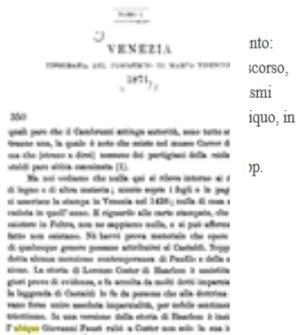


Fig. 8. Esempio di compilazione del campo note: voce *ubiquo*

Data di inserimento e firma dei collaboratori

La data di caricamento si aggiorna ad ogni intervento sulla voce, anche solo per piccole aggiunte o modifiche. Tornando sulla stessa voce a distanza di tempo, quindi, potrebbero esserci informazioni nuove, come del resto accade in altri dizionari progettati per la Rete. L'introduzione della firma del compilatore non stupisce pensando sia al TLIO, sia al più antico modello del Tommaseo-Bellini. I contributi dei collaboratori sono dichiarati nella pagina *Autori* (sezione "Contributi"). Al momento c'è solo un nome e qualche scheda di redazione (cfr. fig. 10), ma dal prossimo anno ci saranno altri collaboratori, ordinati alfabeticamente, con indicazione del numero di voci retrodate. Cliccando su questo numero si accederà al lemario di voci firmate. Ogni collaboratore potrà dunque facilmente attestare il lavoro svolto per il sito. Pensando a questa finalità, ArchiDATA è stato dotato di ISSN, come un periodico. I nuovi inserimenti verranno monitorati quantitativamente ogni sei mesi e registrati nella finestra "Periodicità" presente nella pagina *Progetto*: si darà qui indicazione del numero totale di voci inserite e l'elenco dei collaboratori nel semestre.

Dopo il convegno tenutosi in Accademia della Crusca nel settembre 2017, quello durante il quale è stato firmato l'importante accordo con la UTET per la digitalizzazione del Battaglia, il TLIO e la Consulenza linguistica di Crusca si sono mostrati interessati a far confluire in ArchiDATA le retrodatazioni in loro possesso. Auspicando quindi una prossima collaborazione di questi due gruppi di ricerca, che avranno un loro accesso al sito e completa autonomia, ho pensato alla possibilità di inserire un *link* in calce alle voci di ArchiDATA, per consentire eventuali collegamenti esterni alle voci del TLIO e

alle pagine del sito di Crusca, curate dalla Consulenza linguistica, nelle quali è spiegata la storia della parola.

In attesa di approfondire e concordare con Lino Leonardi, il direttore del TLIO, la forma e le modalità di caricamento del materiale antico, ho sperimentato un primo inserimento dei lemmi *genio* e *grumo* (cfr. fig. 9). Per riconoscere, anche a colpo d'occhio, la specificità di questo materiale antico, ho voluto riprendere graficamente, nei colori, la veste tradizionale del *Tesoro della lingua italiana delle Origini*, riportando il collegamento al TLIO con un fondino grigio e *link* bordò. L'acquisizione di prova dei due lemmi antichi ha subito posto un problema: il programma ArchiDATA, progettato per lessico moderno e fonti a stampa, non prevedeva infatti la possibilità di indicare genericamente un secolo o un intervallo di anni, situazione che per le parole del TLIO, come ovvio, si presenta frequentemente. Dopo qualche modifica tecnica, è ora consentita anche questa modalità di datazione. Resta tuttavia da mettere a punto l'estrazione tramite ricerca cronologica: in una prima fase di caricamento le voci e accezioni antiche saranno probabilmente estraibili in blocco, marcate sotto la categoria "Retrodatazioni del TLIO – lessico antico". Nessuna differenza invece per quanto riguarda le marcature di autore e ambito d'uso, che possono fin da ora essere applicate anche al materiale del TLIO, permettendone l'estrazione tramite ricerca avanzata: la voce *genio* è stata ad esempio marcata con "ambito d'uso – filosofia, religione, psicologia", in riferimento al significato antico di 'divinità tutelare' (saranno in seguito da datare e marcare in modo opportuno le altre accezioni della parola), *grumo* è stato marcato con "ambito d'uso – medicina".

GENIO

s. m. nell'antichità classica, divinità tutelare.

av. 1327 [DELIN e GRADIT av. 1484 (L. Pulci)]

- Cecco d'Ascoli, *L'Acerba*, "Quel genio della morte sarà giunto".

Vai alla voce del TLIO 

Fig. 9. Esempio di retrodatazione acquisita dal TLIO

<p>2 s. m. <i>inv.</i> (<i>bot.</i>) nome comune di pianta insettivora.</p> <p>1819 [GRADIT 1950]</p> <p>- <i>Nuovo corso completo d'agricoltura teorica e pratica</i>, prima edizione italiana, t 1819, p. 13. [<i>acchiappa-mosche</i>]</p> <p>[Ludovica Maconi - 13/07/2017]</p>	
<p>3 s. m. <i>inv.</i> paletta per uccidere mosche e insetti.</p> <p>1874 [DELIN e GRADIT 1950]</p> <p>- Vittorio Imbriani, <i>Mastr'Impicca</i>, 1874: "per ogni stanza si tenevan tre o quattro piattelli con carta moschicida, cinque o sei <i>acchiappamosche</i> prussiani".</p> <p>Cfr. G. Biasci, <i>Nuove retrodatazioni da testi letterari otto-novecenteschi</i>, Roma, Aracne, 2012, s.v.</p> <p>[Scheda di redazione - 07/09/2017]</p>	

Fig. 10. Ritaglio di due accezioni della voce *acchiappamosche*: esempio di scheda firmata dal collaboratore (fonte raccolta in Google con immagine) e scheda di redazione (fonte importata di seconda mano)

Il caricamento dei dati

L'accesso alla maschera di caricamento è consentito solo ai collaboratori di ArchiDATA, e solo dopo registrazione. Per inserire retrodatazioni nell'archivio elettronico bisogna compilare i campi raffigurati nella figura 11. La voce comparirà in archivio nella veste già illustrata.

Durante il lavoro di caricamento, è importante prestare attenzione a marcare manualmente tutti gli elementi che dovranno poi essere estratti tramite opzioni di ricerca avanzata, altrimenti i dati non saranno analizzabili in modo automatico dal motore di ricerca, e non potranno essere messi a disposizione degli studiosi (mi riferisco alle marcature di autore, ambito d'uso e forestierismi). In questa operazione, le marcature riguardanti l'ambito d'uso e i prestiti sono selezionabili da un menù a tendina (cfr. fig. 11, colonna "Significati" della maschera di caricamento); l'autore della fonte della retrodatazione, invece, è marcato scrivendo il cognome tra simboli di trattino basso (cfr. fig. 11, colonna "Fonti").

Tra gli accorgimenti tecnici che ho intenzione di introdurre, segnalo la marcatura dei lemmi polirematici, che permetterà la loro estrazione automatica anche se la ricerca è impostata solo per un elemento della parola: es. *cintura di castità* sarà restituito anche cercando solo *cintura* o solo *castità*. Inoltre, il motore di ricerca sarà adattato per ignorare grafie con accenti e trattini interni al lemma²⁵.

²⁵ Attualmente, lemmi come *branco* e *parlò* non sono estratti se cercati con grafia priva di accento; *accendisigaro* non è estratto se cercato con grafia *accendi-sigaro*.

Lemmi

Salva lemma **Significati**

Significati

Salva significato **Fonti**

Fonti

Salva fonte **Immagine**

Immagini

Salva immagine

Lemmi

Lemma: gastroepatico

Stato Marca come "verificato"

Significati salvati

agg (med.) relativo allo...

Significati

Descrizione: <>agg </> (<>med </>) relativo allo stomaco e al fegato.

Data precedente: DELINI e GRADIT

Data attuale: 1794

Ambito di uso: Medicina / Anatomia / Biologia

Fonsterismo: nessuno

Inserito da: Ludovica Maconi

Fonti

Autore/i: Lorenzo Nannoni

Titolo: Trattato delle materie chirurgiche, e delle operazioni loro rispettive

Dati di stampa: seconda edizione, tomo II, nella stamperia di Francesco Pieracchi, 1794, p. 114.

Immagine

Cerca immagine

Scegli... Nessun file selezionato.

Immagini salvate

Source TypeID: Libro | Position: 1

Immagine:

Trattato delle materie chirurgiche, e delle operazioni loro rispettive. Per opera di Lorenzo Nannoni, medico, e di Francesco Pieracchi, stampatore. Firenze, 1794.

Fonti salvate

[1] Trattato delle mater...

Fig. 11. Maschera di caricamento

Prova di monitoraggio

Dopo due anni di lavoro a questo progetto lessicografico, mi è sembrato opportuno effettuare una prova di monitoraggio nel corpus di Google Libri, per valutare l'entità di cambiamenti e miglioramenti nella ricerca di prime attestazioni lessicali a distanza di tempo. Ho preso come campione dell'indagine mille retrodatazioni raccolte in Google Libri nel 2011 (ma pubblicate nel 2012) da Andrea Monaldi²⁶ (Università di Urbino), e le ho controllate in Google nel novembre 2015, quindi a quattro anni di distanza dal lavoro di Monaldi; poi ancora nel gennaio 2017, per avere informazioni anche su eventuali aggiornamenti nel più breve periodo di un anno (o poco più). Nel primo controllo del 2015, le datazioni sono state aggiornate in circa il 40% dei casi: su 1090 retrodatazioni, ne ho migliorate 408²⁷.

La percentuale di aggiornamento del 40% in quattro anni può sembrare alta, ma bisogna considerare che i cambiamenti significativi, quelli che hanno portato un miglioramento di almeno mezzo secolo nella retrodatazione, hanno riguardato solo il 10% del campione (un centinaio di parole)²⁸. Per le altre parole si è trattato invece di un miglioramento di qualche anno, che non ha cambiato, quindi, il risultato generale fornito

²⁶ Cfr. A. Monaldi, *Mille retrodatazioni*, Roma, Bonacci, 2012.

²⁷ L'elenco è fornito in L. Maconi, *Fonti di Google per la lessicografia: prova di monitoraggio*, in "Italiano LinguaDue", vol. 9, 2017/1, pp. 185-198.

²⁸ Solo 71 lemmi hanno cambiato secolo di appartenenza, passando in gran parte al Settecento, e qualcuno al Sei e Cinquecento; non hanno cambiato secolo, ma sono state aggiornate di circa cinquant'anni una trentina di parole. Per questi risultati, cfr. *ibidem*.

da Monaldi nella sua prima ispezione in Google (ispezione che resta fondamentale anche oggi, a distanza di oltre cinque anni). Possiamo dire che, il più delle volte, è la prima ricerca nella biblioteca di Google ad aggiornare in modo considerevole le date di prima attestazione indicate nei repertori etimologici di riferimento e nel GRADIT.

Bisogna inoltre tener presente che nel tempo trascorso tra la raccolta di Monaldi e la mia verifica c'è stata un'immissione importante di digitalizzazioni nel corpus di Google Libri: tre grandi biblioteche italiane sono entrate a partecipare al progetto, inserendo tra il 2013 e il 2015 un consistente numero di libri del loro patrimonio²⁹. Molti dei miei aggiornamenti sono stati ricavati proprio da libri che sul frontespizio hanno il timbro della Nazionale Vittorio Emanuele III di Napoli (non disponibili quando Monaldi ha svolto la sua ricerca), o anche della Nazionale Centrale di Firenze e di Roma³⁰. Segnalo infine che controlli a distanza di tempo possono essere utili per annotare datazioni di accezioni precedentemente ignorate.

Nel gennaio 2017, dopo poco più di un anno dal secondo controllo, i miglioramenti sono stati minimi³¹. Diversamente

²⁹ Sulla storia di Google Libri e sulla partecipazione di biblioteche italiane, cfr. Y. Gomez Gane, "Google Ricerca Libri" e la linguistica italiana: *vademecum per l'uso di un nuovo strumento di lavoro*, in "Studi linguistici italiani", vol. XXXIV, f. 2, 2008, pp. 260-278; L. Maconi, *Retrodatazioni lessicali con Google Libri: opportunità e inganni della Rete*, cit., pp. 73-74 n. 2.

³⁰ Possiamo supporre che al termine dell'inserimento dei volumi promessi dalle biblioteche italiane a Google, i dati saranno forse meno soggetti a variazione, e che quindi la percentuale di aggiornamento sarà ulteriormente abbattuta.

³¹ Hanno riguardato meno del 5% del campione, e solo tre parole hanno cambiato secolo di prima attestazione: *assordimento* e *illustrativo*, passati

da quanto accaduto nel 2015, però, molte digitalizzazioni non sono più recuperabili (o sono ben nascoste) nel *mare magnum* di Google, e non vengono estratte dal motore di ricerca³². Per questo motivo è sempre bene procedere salvando con immagini il materiale messo a disposizione da Google Libri.

Alcuni ritengono forse che raccolte sistematiche di prime attestazioni dalla Rete siano superflue, perché lo studioso può farsele da sé nel momento in cui necessita dell'informazione. Ma possiamo essere sicuri che queste preziose risorse restino disponibili anche in futuro? Google ha già introdotto alcuni filtri, e libri che anni fa erano scaricabili sono oggi sotto protezione: talvolta ancora interrogabili con motore di ricerca, ma non più consultabili e quindi filologicamente inattendibili ai nostri fini, poiché conosciamo le imperfezioni del sistema di archiviazione e digitalizzazione di Google Libri (cfr. fig. 12)³³.

dal XVIII al XVII secolo, e *augite* dal XVII al XVI (cfr. s.vv. nel sito ArchiDATA); per le altre parole si è trattato nuovamente di un miglioramento di pochi anni.

³² I risultati di questa verifica sono illustrati nel dettaglio in L. Maconi, *Fonti di Google per la lessicografia: prova di monitoraggio*, cit., pp. 195-197. Il motore di ricerca restituisce oggi altre fonti, che, benché in parte peggiorino le datazioni ottenute qualche anno fa, spostano di pochi anni la data di prima attestazione. Solo per un piccolo gruppo di parole il peggioramento è stato pesante: perse le precedenti attestazioni settecentesche, risultano oggi attestate in Google a partire dall'Ottocento *bromatologia*, *circumnavigare*, *circumnavigatore*, *circumnavigazione*, *deteriorabile*, *dicotomico* (ultimo controllo nel giugno 2017). In ArchiDATA queste parole sono tuttavia archiviate con datazione settecentesca, perché avevo a suo tempo salvato le fonti.

³³ Su questi difetti, che in ogni caso non compromettono l'utilità del corpus di Google Libri ai nostri fini lessicografici, perché si tratta comunque di un corpus impareggiabile per quantità e per varietà del materiale raccolto, cfr. L. Maconi, *Retrodatazioni lessicali con Google Libri: opportunità e inganni della Rete*, cit., pp. 77-85.

Google libri

Ricerca Libri avanzata

Opere complete di Giuseppe Pitrè..

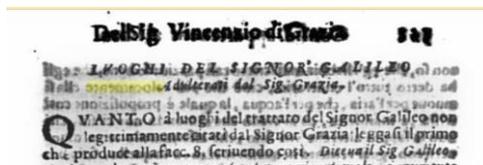
books.google.it/books?id=-QgOQAQAAI
Giuseppe Pitrè - 1614 - Leggi - Altre ec

OPERE
COMPLETE DI

Cuore: libro per i ragazzi - Pagina

books.google.it/books?id=Y-1WAAI
Edmondo De Amicis - 1822 - Leggi
Non c'è!; - si risvegliò di sobbalzo,

CUORE
LIBRO PER I
RAGAZZI



Corrugare nel complicare mani-
cas ad cubitum.

NAPOLI MDCCCIX.
Presso GENARO DE TONIS Strada S. Gregorio Armeno
N. 41.

Fig. 12. Esempi di imperfezioni di Google Libri. 1. Libri schedati con data sbagliata: G. Pitrè, noto antropologo ottocentesco, non può certo aver pubblicato libri nel Seicento, come erroneamente indicato nella scheda di Google; E. De Amicis, che pubblicò *Cuore* nel 1886, nel 1822 non era ancora nato. 2. Digitalizzazioni deformate che compromettono la lettura del testo tramite OCR. 3. Errato riconoscimento di parole per cattiva qualità della stampa (in alcuni casi anche per difficile lettura dei caratteri tipografici antichi). 4. Impossibilità di isolare lingue diverse all'interno di uno stesso libro (nella foto, parola latina restituita come attestazione italiana di *corrugare*). 5. Errori nella lettura della data di stampa per cattiva digitalizzazione.

Sviluppi

Per ampliare in modo significativo ArchiDATA, quattro sono a mio avviso le vie di ricerca produttive (e queste vie potrebbero corrispondere ad altrettanti gruppi operativi):

1. Raccogliere in modo sistematico la recente bibliografia avente per oggetto retrodatazioni, e acquisirne i dati, dopo controllo nei corpora disponibili in Rete³⁴.
2. Spogliare con programmi informatici testi allestiti per il corpus del VoDIM, ma, se possibile, anche testi più antichi digitalizzati per progetti che hanno avuto corso negli ultimi anni: penso soprattutto ai testi epistolari di AITER, *Archivio italiano tradizione epistolare in Rete*, e CEOD, *Corpus epistolare ottocentesco digitale*³⁵.
3. Continuare l'interrogazione di Google Libri per retrodatare parole che attualmente hanno datazione compresa tra il XVIII e il XX secolo (cioè quelle che me-

³⁴ Se il controllo non è già stato fatto e dichiarato dagli autori, o se è stato compiuto parecchio tempo fa.

³⁵ Si tratta di corpora di testi epistolari digitalizzati e liberamente disponibili in Rete, con consultazione tramite motore di ricerca: cfr. *Aiter*, <http://aiter.unipv.it/>, progetto che ha coinvolto, tra le altre, l'Università di Pavia, curatrice del sito Internet; cfr. *CEOD*, <http://193.204.192.242/index.htm>, realizzato dalle Università di Siena per Stranieri, Cassino, Roma "La Sapienza" e Milano Statale, le quali hanno partecipato anche ad *Aiter* (ultimi accessi nel novembre 2017).

glio si prestano all'aggiornamento usando una collezione digitale di libri a stampa), con particolare attenzione per il lessico tecnico. La ricerca andrà poi estesa negli archivi dei quotidiani nazionali³⁶, che talvolta, a sorpresa, si possono rivelare utili per datare anche lessico specialistico³⁷: “La Stampa” mette in libera consultazione l'archivio più completo.

³⁶ L'interrogazione di Internet Archive (<https://archive.org/>) è meno fruttuosa perché non permette la ricerca lessicale nell'intero corpus, ma solo in un singolo testo selezionato. Può quindi servire soltanto per ricerche mirate, come quelle descritte nei seguenti esempi: 1. la fonte di una retrodatazione estratta da Google è la terza edizione di un libro di cui Archive mette a disposizione la prima edizione; 2. Google restituisce come fonte di una retrodatazione un saggio scientifico di un autore presente in Archive con digitalizzazioni di altri testi più antichi, che potrebbero a loro volta contenere la parola in esame. Si tratta quindi di un lavoro di cesello di cui non è sicuro l'esito positivo.

³⁷ Su retrodatazioni ricavate dall'archivio della “Stampa”, cfr. n. 21. Menziono inoltre due esempi in cui, consultando l'archivio elettronico del quotidiano nazionale, è stata affinata la datazione di lessico medico, abbandonando la generica indicazione di “secolo XX” fornita dai repertori lessicografici di riferimento: il *cardiotocografo* (non a lemma nel GRADIT, nello *Zingarelli* datato sec. XX), strumento per monitorare il battito cardiaco fetale, è attestato nella “Stampa” del 9 maggio 1975, p. 8, nell'articolo di E. Robecchi, *I rischi del parto*; l'aggettivo *carcinogeno* (GRADIT sec. XX) ha attestazione nella “Stampa” del 25 marzo 1955, p. 3, nell'articolo *Esperienze che accerchiano il tumore maligno. Isolato un virus del cancro?*. I risultati ottenuti interrogando l'archivio della “Stampa” concordano con quelli di Google Libri, che per *cardiotocografo* restituisce fonti degli anni Settanta e per *carcinogeno* fonti di metà Novecento: si tratta però di attestazioni in riviste scientifiche in visualizzazione bloccata, motivo per cui, anche se si può considerare sicura la presenza della parola (individuata in modalità *snippet* ‘ritaglio’), senza la consultazione diretta, non si può essere altrettanto certi della corretta indicazione dell'annata della rivista, che potrebbe essere falsata di qualche anno per cattiva catalogazione (cfr. gli esempi illustrati in L. Maconi, *Retrodatazioni lessicali con Google Libri: opportunità e inganni della Rete*, cit., pp. 81-84).

4. Sviluppare la datazione delle accezioni, settore ancora troppo carente di studi.

Già ho menzionato le collaborazioni che potranno venire dal TLIO e dalla Consulenza linguistica di Crusca. Segnalo, in chiusura, gli importanti materiali che stanno per uscire dalla redazione dell'AVSI (*Archivio per il Vocabolario Storico Italiano*), con il quale sarà fondamentale instaurare uno stretto dialogo, anche per valorizzare il più possibile risultati raggiunti da ricercatori che si occupano di studi affini. Dal canto mio, procederò con il caricamento del materiale fornito negli scorsi anni allo *Zingarelli*, e aggiungendone di nuovo, consapevole che per non restare esercizio strettamente tecnico, le retrodatazioni avranno valore culturale quando si potrà ragionare su un grande archivio di parole, frutto di lavoro collettivo. Allora ArchiDATA diventerà, mi auguro, uno strumento utile per nuove ricerche, e allora sarà possibile costruire un discorso storico, argomentato e ben documentato sull'evoluzione del nostro patrimonio lessicale.

ISBN 978-88-89369-84-5