

## Gli scaffali digitali dell'Accademia della Crusca

Nicoletta Maraschio, Claudio Marazzini

Pubblicato: 15 dicembre 2021

### *Abstract*

The section of the Accademia della Crusca website called *Scaffali digitali* has been enriched over time, so much so that it has become one of the most appreciated and useful language resources on the Internet for scholars from all over the world. The report will illustrate the lines of action followed and tell the story of *Scaffali digitali*, in parallel with the growing importance of IT resources applied to the humanities sector in the Italian cultural framework. The first phase of the *Scaffali* repropose the 5 editions of the *Vocabulary of the Academy*, an operation being completed. The report will therefore highlight how, around this fundamental nucleus of Italian lexicography transferred from paper to digital, a valuable collection of new vocabularies and tools has been formed (e.g. *ArchiDATA* and the *Stazione lessicografica*) or sometimes obtained through a policy of free exchanges with publishing (Tommaseo-Bellini and GDLI). The report will illustrate the difficulties encountered in transposing from paper to site, taking into account that digital is an approximation that is gradually improving, and the result is continuous innovation.

La sezione del sito dell'Accademia della Crusca denominata *Scaffali digitali* si è andata arricchendo nel corso del tempo, fino a diventare una delle risorse linguistiche della Rete più frequentate e più utili per gli studiosi di tutto il mondo. La relazione illustrerà le linee di intervento seguite e darà conto della storia degli *Scaffali digitali*, parallela alla crescita d'importanza delle risorse informatiche applicate al settore umanistico nel quadro culturale italiano. La prima fase degli *Scaffali* ha riproposto le 5 edizioni del *Vocabolario dell'Accademia*, un'operazione che si va ora concludendo. La relazione metterà quindi in evidenza come, attorno a questo nucleo fondamentale della lessicografia italiana trasferito dalla carta a digitale, si sia costruita una raccolta preziosa di vocabolari e di strumenti, nuovi (ad es. *ArchiDATA* e la *Stazione lessicografica*) o talora ottenuti attraverso una politica di scambi non onerosi con l'industria editoriale (il Tommaseo-Bellini e il GDLI). La relazione illustrerà le difficoltà incontrate nella trasposizione dalla carta al sito, tenendo conto del fatto che il digitale è un'approssimazione che va migliorando via via, e l'esito sta nell'innovazione continua.

**Parole chiave:** Accademia della Crusca; archivi digitali; lingua italiana.

**Nicoletta Maraschio:** Università degli Studi di Firenze

✉ [nicoletta.maraschio@unifi.it](mailto:nicoletta.maraschio@unifi.it)

**Claudio Marazzini:** Università del Piemonte Orientale

✉ [claudio.marazzini@uniupo.it](mailto:claudio.marazzini@uniupo.it)

Copyright © 2021 Nicoletta Maraschio, Claudio Marazzini

The text in this work is licensed under Creative Commons BY-SA License.

<https://creativecommons.org/licenses/by-sa/4.0/>

Oltre dieci anni fa, durante la presidenza dell'Accademia della Crusca, Nicoletta Maraschio, insieme a Marco Biffi, ha tracciato un profilo dei nostri programmi 'digitali' e dei risultati ottenuti fino a quel momento.<sup>1</sup>

L'intenso programma informatico che l'Accademia della Crusca ha sviluppato, soprattutto nell'ultimo decennio, risponde ad alcuni scopi fondamentali tra loro strettamente collegati: tutelare e valorizzare il proprio patrimonio librario e documentario, facilitarne la consultabilità da parte di un pubblico il più largo possibile e creare strumenti di ricerca capaci di favorire in Italia e all'estero una migliore conoscenza dell'italiano di ieri e di oggi. La storia dell'Accademia, dei suoi libri, delle sue carte e soprattutto del suo *Vocabolario* (nelle cinque edizioni: 1612-1923) si intreccia inestricabilmente con la storia della nostra lingua. Proprio per questo l'Accademia della Crusca ha investito molte risorse e molte energie sia nella digitalizzazione di parti caratterizzanti della propria biblioteca e del proprio archivio, sia nella creazione di banche dati e strumenti informatici utili a ricostruire in modo sistematico il quadro evolutivo del lessico italiano e a permettere un'osservazione più ampia e documentata dell'italiano contemporaneo, in vari suoi aspetti.<sup>2</sup>

Non c'è molto da cambiare oggi per quanto riguarda le finalità. È stato Giovanni Nencioni ad aprire precocemente in Accademia la strada della digitalizzazione, pensando in primo luogo alla lessicografia, convinto che, grazie al «rovesciamento» della prima edizione (1612) del *Vocabolario degli accademici della Crusca* (ossia alla sua digitalizzazione in formato testo), sarebbe stato possibile fare emergere il moltissimo lessico nascosto che esso conteneva.<sup>3</sup> Tutti i presidenti che gli sono succeduti (Francesco Sabatini, Nicoletta Maraschio e Claudio Marazzini) hanno proseguito per questa strada. In particolare il «rovesciamento» ideato da Nencioni è stato esteso dalla prima fino alla quinta edizione del *Vocabolario* (1623, 1691, 1729-1738, 1863-1923) così da permetterne l'interrogazione in modo sistematico e rapido, con ricerche avanzate che consentono di selezionare sezioni specifiche (definizioni, esempi, parole greche e latine, forestierismi, locuzioni, proverbi, parole dell'uso vivo, fonti); ma anche di confrontare costantemente le diverse edizioni. È così possibile seguire con precisione le tappe evolutive del lavoro lessicografico degli accademici e insieme i cambiamenti dell'italiano da loro registrati nel corso dei secoli.<sup>4</sup>

<sup>1</sup> Marco Biffi da molti anni è il responsabile dell'attività informatica dell'Accademia.

<sup>2</sup> Cfr. N. Maraschio e M. Biffi, *Strumenti digitali dell'Accademia della Crusca*, in S. Magherini (a cura di), *Tradizione e modernità. Archivi digitali e strumenti di ricerca*, Firenze, Società Editrice Fiorentina, 2009, pp. 115-146 (il brano citato è a p. 115).

<sup>3</sup> Ivi, pp. 137-140. La prima versione è stata fatta in collaborazione con la Scuola Normale Superiore di Pisa ed è stata curata da Mirella Sessa, si può talitare negli *Scaffali digitali* dell'Accademia. Si veda anche M. Fanfani, *L'edizione elettronica del Vocabolario*, in *Una lingua, una civiltà, il Vocabolario* (volume di corredo al *Vocabolario degli Accademici della Crusca*. Riproduzione anastatica della prima edizione, Venezia 1612), Accademia della Crusca/ERA, Varese, 2008, pp. 63-78.

<sup>4</sup> Citiamo liberamente dalla presentazione del progetto, realizzato sotto la responsabilità di Marco Biffi e Massimo Fanfani: [www.lessicografia.it](http://www.lessicografia.it). Alla sua realizzazione (che ha incontrato non pochi ostacoli per la parte informatica), hanno lavorato molti collaboratori e collaboratrici (trascrizione, marcatura, revisione) che negli ultimi anni sono stati impegnati anche nella digitalizzazione in formato testo degli 11 volumi della V Crusca (finora consultabile solo per immagini). Tra le ricercatrici e i ricercatori più a lungo coinvolti: Silvia Dardi, Cecilia Palatresi, Mariella Canzani e più recentemente, insieme a loro e limitativamente per la V Crusca, Francesca Cialdini, Vera Gheno, Antonio Vinciguerra.

Crediamo occorra riconoscere che la forte continuità di questo importante settore di attività, legato alle nuove tecnologie e al loro rapido sviluppo, sia stata favorita dal carattere istituzionale dell'Accademia. È a tutti noto che la Crusca è molto antica (1583), ma nel tempo essa ha dovuto via via rinnovarsi, spinta in primo luogo dall'esigenza ineludibile di adeguare la propria grande impresa, il *Vocabolario*, ai mutamenti linguistici reali e ai nuovi modelli normativi che sono venuti imponendosi. Ma anche dopo il 1923, successivamente cioè alla fine dei lavori lessicografici imposta dal ministro Gentile, la Crusca non ha smesso di trasformarsi, in queste circostanze, per adeguarsi al decreto ministeriale che la impegnava a un compito diverso, preparare cioè «testi critici dei primi secoli».<sup>5</sup> E l'ha fatto potenziando la propria vocazione filologica, per altro originaria. Nel 1927 è stato creato infatti il Centro di studi di filologia italiana, animato da studiosi come Casella, Barbi, Pasquali, al cui interno si è voluto dare spazio a giovani e valenti filologi, attraverso un apposito comando.<sup>6</sup> Ne hanno usufruito per primi Gianfranco Contini e Vittore Branca. Per quanto riguarda poi il secondo dopoguerra, Nencioni ha parlato in molte occasioni di una «nuova Crusca», insistendo sulla svolta impressa da presidenti come Bruno Migliorini e Giacomo Devoto e, aggiungiamo noi, da lui stesso. Siamo convinti che nel momento storico attuale, caratterizzato da un'evidente accelerazione delle trasformazioni culturali e comunicative in atto, le grandi accademie (e la Crusca è tra queste) possano rappresentare solidi punti di riferimento grazie alla loro storia, alla loro autorevolezza e alla rete di rapporti nazionali e internazionali nella quale sono inserite spesso da secoli. Ci sembra che gli *Scaffali digitali* dell'Accademia della Crusca possano confermare questo punto di vista.

Se scorriamo le date delle molte realizzazioni, rileviamo che dal 2000 fino a oggi non c'è stata interruzione di sorta. Questa continuità deriva da molti fattori; innanzi tutto dal fatto che l'Accademia ha considerato la digitalizzazione un ambito privilegiato in cui investire ingenti risorse, e non solo per le finalità alle quali abbiamo prima fatto riferimento (tutela e valorizzazione del proprio patrimonio librario e archivistico), ma anche in ottemperanza ad altri due fondamentali scopi statutari: la ricerca e la formazione. Infatti nel settore informatico sono stati impegnati, accanto a specialisti informatici, molti giovani laureati o dottori di ricerca (soprattutto in storia della lingua italiana), che hanno imparato a sviluppare e realizzare progetti proprio in questo nuovo campo. A loro sono stati affiancati via via, a seconda dei casi, altri collaboratori, ad esempio esperti in archivistica, biblioteconomia e storia dell'arte. È stato inoltre creato un laboratorio fotografico (e sono state formate le persone anche in questo campo) per permettere all'Accademia di acquisire 'in casa' la maggior parte delle immagini delle sue banche dati. Si aggiunga che per circa venti anni Marco Biffi, storico della lingua ed esperto informatico, è stato un costante punto di riferimento (e continua ancora ad esserlo), avendo la responsabilità di coordinare l'intero settore, in stretta sinergia con il Consiglio direttivo dell'Accademia e con gli accademici che hanno diretto (spesso con lui) le diverse realizzazioni digitali.

<sup>5</sup> S.Parodi, *Quattro secoli di Crusca. (1583-1983)*, Firenze, Accademia della Crusca, pp. 168-174.

<sup>6</sup> Il Centro è stato diretto dal 1971 al 2000 da Domenico De Robertis che dal 1966 al 1980 ha diretto anche l'Ufficio filologico dell'*Opera del Vocabolario italiano* (OVI) per la preparazione della tavola dei citati del *Tesoro della lingua italiana delle origini* (TLIO), quindi in stretta connessione con la ripresa dell'attività lessicografica.

Ma guardando fuori dallo spazio accademico possiamo fare qualche ulteriore osservazione oltre il tema della continuità: molti progetti sono stati ideati e sviluppati insieme ad altre istituzioni, in primo luogo ad alcune università, ma anche a biblioteche, centri di ricerca e istituti culturali. Insomma la digitalizzazione ha stimolato l'Accademia ad aprirsi a nuove collaborazioni. In particolare una delle più antiche banche dati realizzate dall'Accademia *La fabbrica dell'italiano* (2001),<sup>7</sup> nasce da un progetto di grande interesse *Biblioteche centrali e biblioteche speciali*, coordinato dalla BNCF (direttrice Antonia Ida Fontana),<sup>8</sup> al quale hanno partecipato, oltre all'Accademia della Crusca (Giovanni Nencioni), la Fondazione Franceschini (Claudio Leonardi), l'Istituto nazionale di studi sul Rinascimento (Michele Ciliberto), l'Istituto e il Museo della scienza (Paolo Galluzzi). Si è trattato di un'esperienza di alto valore culturale. Se guardiamo specificamente all'Accademia, le ha dato la possibilità di collaborare con enti già da tempo impegnati nelle nuove tecnologie e di costruire per la prima volta un archivio digitale integrato. *La fabbrica dell'italiano* si fonda, infatti, sul ricco patrimonio bibliotecario e archivistico dell'accademia: sono stati inseriti oltre duemila dizionari, quasi quattrocento grammatiche; sono stati catalogati i centosettanta manoscritti giunti in Crusca per i concorsi letterari banditi nell'Ottocento e sono stati lemmatizzati i circa novemila termini tecnici raccolti nella seconda metà del '600 dal cardinale Leopoldo de' Medici. Questo progetto ha spinto l'Accademia a valorizzare sempre più, proprio attraverso la digitalizzazione, la propria storia, i propri libri e manoscritti e le proprie carte. Se ne può avere un'idea dalle molte banche dati che sono state create successivamente alla *Fabbrica*, grazie al lavoro dei dipendenti, dei collaboratori della Biblioteca e dell'Archivio e da un gruppo di informatici guidato da Giovanni Salucci, da molti anni legato all'Accademia:<sup>9</sup> *Fondo dei citati* (2003-2004); *Cinquecentine* (2010); *Catalogo degli accademici* (2011-2018); *Quinta Crusca virtuale* (2011-2018);<sup>10</sup> *Incunabuli* (2015); *Biblioteca: infografica* (2018); *Manoscritti (anche Fondo Castellani)* (2018).

Un altro ente storicamente legato all'Accademia è stato l'Istituto di Linguistica Computazionale di Pisa (all'interno del CNUCE, Centro Nazionale Universitario di Calcolo Elettronico). È noto che il rapporto con Antonio Zampolli, direttore dell'Istituto, è stato fondamentale per la rifondazione del *Vocabolario*, voluta da una commissione di accademici (Barbi, Pasquali e Nencioni) fin dagli anni Cinquanta, ma i cui lavori furono avviati, di fatto, nel decennio successivo. Come spiega molto bene Aldo Duro, che divenne direttore dell'impresa nel 1964, è stato merito di Zampolli avergli: «infuso fiducia nell'impiego dell'elaboratore elettronico per

<sup>7</sup> La banca dati è stata originariamente pubblicata su DVD.

<sup>8</sup> Una sintesi in A. I. Fontana, *Le biblioteche speciali nel sistema nazionale*, «Biblioteche oggi», gennaio-febbraio 1999, p. 79.

<sup>9</sup> La maggior parte delle banche dati di questo settore è stata realizzata appunto da Giovanni Salucci e dalla società Progettirete da lui amministrata. Ricordiamo alcuni dei responsabili e collaboratori più coinvolti qualitativamente e quantitativamente nei diversi progetti: Delia Ragionieri, Giuseppe Abbatisa, Marta Ciuffi, Elisabetta Benucci, Fiammetta Fiorelli, Raffaella Setti, Silvia Dardi, Cecilia Palatresi, Mariella Canzani, Francesca Carletti, Maria Cristina Torchia, Alessio Misuri, Barbara Fanini.

<sup>10</sup> Come è noto, gli Accademici nel 1923 non hanno interrotto i lavori di schedatura. Nell'Archivio accademico sono conservate le schede relative alle ultime lettere dell'alfabeto (dalla O alla Zeta). Nella *V Crusca virtuale* sono per ora consultabili i 9 faldoni (pezzi 715-723) contenenti le schede di preparazione, in fase avanzata, dei lemmi della lettera P, da *p* a *p* *perlustrazione* per la continuazione della quinta impressione del *Vocabolario* (1863-1923), databili tra il 1918 e il 1931. Inoltre la stessa banca dati contiene anche i neologismi schedati dallo scrittore e giornalista Cesare Donati (1826-1913), tratti da giornali e periodici. Nel 1913 sua figlia ha regalato alla Crusca il prezioso schedario del padre.

lavori di lessicografia» (p. 433). La Crusca ha preso contatto allora con le principali imprese lessicografiche europee e il modello del *Trésor de la langue française* (diretto da Paul Imbs) è stato decisivo per la scelta dell'applicazione informatica al nuovo *Vocabolario* in fabbricazione.<sup>11</sup>

Si tratta di anni lontani, ma la collaborazione con l'ILC (Istituto di Linguistica Computazionale) è continuata fino a un periodo molto più vicino a noi e, come vedremo meglio in seguito, continua ancora oggi. Dal 1993 al 2003, l'Accademia della Crusca (con alcune università, in particolare con il CLIEO, Centro d'eccellenza dell'Ateneo fiorentino) ha realizzato, insieme a Eugenio Picchi, allora ricercatore dell'ILC, un progetto innovativo basato sul suo DBT e dedicato al parlato radiofonico: il *Lessico dell'italiano radiofonico* (LIR), che è stato pubblicato prima in DVD ed è ora consultabile nel VIVIT (*Vivi italiano, il portale dell'italiano nel mondo*).<sup>12</sup> Il VIVIT raccoglie altri archivi digitali relativi all'italiano novecentesco (il LIT, *Lessico dell'italiano televisivo* e il LIS, *Lessico dell'italiano scritto*). La piattaforma è rivolta in primo luogo agli italiani all'estero che vogliono «stabilire un solido contatto culturale a distanza con il nostro Paese, conoscere meglio la lingua, (le sue strutture grammaticali e la storia). È stato inserito anche il *Dizionario elettronico di italianismi*, corrispondente al cartaceo *Dizionario di Italianismi in Francese, Inglese e Tedesco*, diretto da Harro Stammerjohann (DIFIT).<sup>13</sup> Il VIVIT prevede sia percorsi didattici (coordinati da Maria Pia Lo Duca) sia percorsi descrittivi sulla storia della lingua italiana (coordinati da Francesco Sabatini), in collegamento ai più significativi fenomeni storici, artistici e di costume (ai quali hanno collaborato diversi studiosi). La sezione sulle comunicazioni di massa è stata curata da un grande esperto come Peppino Ortoleva.<sup>14</sup>

Già da questa breve sintesi risulta la grande varietà di ambiti che sono stati interessati dalla digitalizzazione. Sempre per quanto riguarda la lessicografia aggiungiamo altri due progetti, uno avviato qualche anno fa, l'altro più recente: il *Vocabolario del fiorentino contemporaneo* (2008) e il *Vocabolario dantesco* (2017). Quest'ultimo nasce in collaborazione con l'OVI (responsabili Paola Manni e Lino Leonardi) e si pone l'obiettivo di raccogliere l'intero patrimo-

<sup>11</sup> È stato ripubblicato recentemente il prezioso volumetto M. Barbi, G. Pasquali, G. Nencioni, *Per un grande vocabolario storico della lingua italiana*, Firenze, Le Lettere, 2012. Per la prima fase dei lavori, iniziati negli anni Sessanta, è di grande interesse A. Duro, *L'impianto del nuovo vocabolario: profilo storico*, in *La Crusca nella tradizione letteraria e linguistica italiana*, Atti del Congresso internazionale per il IV Centenario dell'Accademia della Crusca (29 settembre 2 ottobre 1983), Firenze, Accademia della Crusca, 1985, pp. 431-442. Duro ricorda anche l'incontro (gennaio 1962) al Centro di elaborazione dati di Gallarate con il direttore padre Busa. Come è noto dall'idea di un *Vocabolario* cronologicamente esteso si passò a quella di un *Tesoro delle Origini* e i lavori passarono all'OVI, prima Centro e poi Istituto del CNR che continua a condividere con la Crusca la sede e la biblioteca. La collaborazione tra le due istituzioni continua a essere molto stretta, come dimostra, tra l'altro, il progetto del *Vocabolario dantesco* su cui torneremo.

<sup>12</sup> Il LIR (*Lessico dell'Italiano Radiofonico*) è una banca dati testuale e audio costituita da due *corpora* di parlato radiofonico raccolti in periodi distinti, nel maggio del 1995 e nello stesso mese del 2003: il primo contiene 108 ore complessive di registrazione (di 9 emittenti a livello nazionale, oltre alle tre reti Rai, Radio DJ, Radio 102,5, Rete 105, Italia Radio, Radio Radicale e Radio Vaticana) con la trascrizione delle parti parlate per un totale di circa 60 ore; il secondo, limitato alle sole reti Rai (Radio1, Radio2, Radio3), raccoglie 30 ore di parlato su 36 complessive di registrazione, scelte in base ad una griglia statistica analoga a quella applicata al corpus del 1995 (stesso periodo dell'anno, stesse fasce orarie, ecc.). Il parlato è stato trascritto, opportunamente marcato e allineato al sonoro. Tuttavia nel VIVIT non tutte le funzioni presenti nel DVD sono attive. Si presenta insomma per il LIR, ma anche per il LIT, il grave problema, sul quale ritorneremo, dell'obsolescenza tecnologica che richiede grandi sforzi per recuperare, materiali, strumenti e banche dati non più utilizzabili.

<sup>13</sup> H. Stammerjohann *et al.* (a cura di), *Dizionario di italianismi in francese, inglese e tedesco*, Firenze, Accademia della Crusca, 2008.

<sup>14</sup> La progettazione generale si deve a Sabatini, Biffi, De Martino e Maraschio.

nio lessicale contenuto nelle opere di Dante, a cominciare dalla *Commedia*. Il *Vocabolario* affianca a un'impostazione lessicografica tradizionale (definizione e articolazione semanticamente puntuale dei lemmi) un dialogo continuo (attraverso opportuni collegamenti ipertestuali) con altre importanti risorse informatiche: dalla *Lessicografia della Crusca in rete*, al *Tesoro della Lingua Italiana delle origini* (TLIO), alla versione informatizzata dell'*Enciclopedia dantesca Treccani* (ED). Il *Vocabolario del fiorentino contemporaneo*, invece, risale a un'idea di Nencioni degli anni Novanta (1994), è diretto da Teresa Poggi Salani e vuole documentare «tutta quell'area del repertorio in cui si verifica al giorno d'oggi una mancata o parziale sovrapposizione tra lessico fiorentino e lessico italiano». Sono state finora pubblicate oltre 3670 schede, ed altre 2800 circa sono in corso di pubblicazione.<sup>15</sup> Si tratta di uno strumento fondamentale per capire meglio il diverso processo di italianizzazione che si è verificato a Firenze, nel corso del Novecento, rispetto a quello di altre grandi città italiane. All'ambito lessicografico, in particolare alle voci tecnico artistiche, appartengono inoltre un portale dedicato alle *Parole dell'arte* (2018) e una banca dati sulle *Antiche guide della città* (2019). Entrambi i progetti sono stati realizzati dell'Accademia della Crusca insieme alla Fondazione Memofonte e hanno coinvolto storici dell'arte e storici della lingua.<sup>16</sup>

Scorrendo ancora gli *Scaffali digitali* dell'Accademia possiamo consultare altre banche dati, alcune delle quali presentano problemi di manutenzione e di aggiornamento. Non è questo il caso della banca dati *Dalla Scuola senese all'Accademia della Crusca (2010-2011)*: un articolato percorso virtuale che consente di navigare fra i testi dell'importante corrente linguistica senese (1500-1700: da Claudio Tolomei a Girolamo Gigli), con il supporto di strumenti diversi: la bibliografia delle fonti, le biografie dei diversi esponenti, le schede critico-informative sulle principali opere e una bibliografia critica.<sup>17</sup> Per quanto riguarda *l'Antologia Vieusseux* (2011-2018), nel sito sono consultabili per immagini tutti i 48 volumi (per un totale di 26.600 pagine) della rivista, il motore di ricerca lavora sugli indici, consentendo ricerche per autori, titoli, date, argomenti, ma non è disponibile la digitalizzazione in formato testo che pure era stata avviata. Il progetto affonda le sue radici in una collaborazione con il Dipartimento di Italianistica dell'Università di Firenze (Prof. Andrea Dardi) e rientra tra gli obiettivi dell'Accademia, che negli ultimi anni è impegnata ad approfondire il periodo ottocentesco della storia linguistica italiana, anche in vista della realizzazione del progetto VoDIM (Vocabolario Dinamico dell'Italiano) In questo quadro rientra quindi anche il recupero di un'opera fondamentale, *La stampa periodica milanese della prima metà dell'Ottocento: testi e concordanze*, curata da Stefania

<sup>15</sup> Ne è stata pubblicata anche una breve versione cartacea: T. Poggi Salani, N. Binazzi, M. Paoli, C. Torchia (a cura di), *Parole di Firenze dal Vocabolario del fiorentino contemporaneo*, Firenze, Accademia della Crusca, 2017.

<sup>16</sup> Il portale delle *Parole dell'arte* (2018) comprende diverse banche dati costruite nel tempo: *Trattati d'arte cinquecenteschi* (2015), *Per un lessico artistico: testi dal XVIII al XX secolo* (2016), *La lingua nella storia dell'arte nel XX secolo: Roberto Longhi* (2017), *Manifesti futuristi* (2017-2018). Le antiche Guide raccolgono testi fiorentini e napoletani. Coordinatori: Donata Levi, Francesco Caglioti, Nicoletta Maraschio. Predisposizioni dei testi, marcature, formari e lemmari, per la parte artistica di entrambe le realizzazioni: Martina Nastasi, Claudio Brunetti, Martina Visentin, Giovanni Rubino, Michela Tarallo, per la parte linguistica Barbara Fanini, Valentina Paggini, Massimiliano Bertelli, Antonio Vinciguerra.

<sup>17</sup> I testi sono stati inseriti da Margherita Quaglino che ha curato la realizzazione del progetto insieme ad altre collaboratrici (S. Dardi, B. Fanini, C. Torchia) che si sono occupate della marcatura. È interessante rilevare che la digitalizzazione completa della maggior parte delle opere è stata fatta successivamente grazie a progetti diversi della Biblioteca.

De Stefanis Ciccone, Ilaria Bonomi e Andrea Masini.<sup>18</sup> La realizzazione informatica è stata predisposta anni fa dall'Istituto di Linguistica Computazionale del Consiglio Nazionale delle Ricerche (ILC-CNR), da Eugenio Picchi e Remo Bindi. L'archivio testuale è stato arricchito con annotazioni linguistiche di varia natura (in particolare, morfo-sintattiche, lessicali e semantiche). Il recupero attuale è stato possibile grazie a una collaborazione con l'ILC «all'interno di una linea strategica di attività dedicata alla salvaguardia, conservazione e valorizzazione dei testi dell'Archivio Storico».<sup>19</sup> Il fatto che la piattaforma di gestione si appoggi a Adobe Flash Player, ormai da anni non più supportato dalla casa produttrice, implica l'aggiornamento degli strumenti di interrogazione.

Analoghi problemi presentano le banche dati radiofoniche e televisive (LIR, LIT) nate con il DBT e il Portale dell'italiano televisivo (PIT 2013) che ha visto il coinvolgimento delle Università di Firenze, Catania, Genova, Milano e della Toscana, e ha portato a un arricchimento del LIT (168 ore di parlato televisivo trascritto e allineato alle immagini e al sonoro, in VI-VII) di altre circa 40 ore di parlato televisivo in diacronia.

Converrà aggiungere che gli *Scaffali digitali* della Crusca sono cresciuti negli anni in maniera davvero considerevole, fino a porre il problema di un ordinamento del materiale, quando non era più possibile sommare via via nuove risorse con aggiunta meccanica, collocandole in una pagina elettronica la quale tecnicamente aveva la potenzialità di srotolarsi all'infinito, ma ciò avrebbe violato le regole universalmente suggerite per un web accessibile e ben ordinato, facilmente consultabile, amichevole. Per questo gli *Scaffali*, nel corso della presidenza di Claudio Marazzini, sono stati non solo arricchiti, ma anche riordinati secondo una gerarchia logica, per facilitare il lettore. Al primo posto, in apertura, in testa all'elenco, sono state collocate *Le Crusche in rete*, le cinque *Crusche* di cui si è già ampiamente discusso. La collocazione si giustifica considerando che il *Vocabolario* è il risultato della vita dell'Accademia nei secoli, e inoltre questa è stata la prima palestra della nostra intensa digitalizzazione. La sezione che segue i cinque vocabolari è dedicata alla *Storia e patrimonio librario dell'Accademia della Crusca*. Comprende, come abbiamo accennato, una serie di banche dati specifiche: prima di tutto il *Catalogo degli Accademici*, poi l'*Infografica* della biblioteca dell'Accademia, quindi il *Catalogo degli incunaboli* con digitalizzazioni fotografiche, il catalogo delle cinquecentine con digitalizzazioni parziali, il catalogo dei 'citati' con digitalizzazioni selettive, la banca dati dei manoscritti della Crusca, la *Fabbrica dell'italiano*, la *Quinta Crusca virtuale* (una banca dati che raccoglie i numerosi materiali preparatori alla quinta edizione del *Vocabolario* conservati presso l'Archivio), e infine un DB del *Vocabolario* del 1612, nella forma raggiunta dal vecchio progetto pionieristico di trattamento informatico, successivamente superata dall'edizione completa delle cinque edizioni.

Come si è detto, tutta questa ricca messe di strumenti, fondamentali per conoscere la storia e il patrimonio dell'Accademia, è accessibile da una sotto-sezione apposita. Ma torniamo invece alle partizioni fondamentali. Si incontrano, nell'ordine, vari strumenti lessicografici, anche estranei alla storia dell'Accademia, ma comunque, nella forma attuale, tutti 'nostri', in quanto

<sup>18</sup> Pisa, Giardini Editore, 1984.

<sup>19</sup> Recupero coordinato da Manuela Sassi in collaborazione con l'Accademia della Crusca e, in particolare, con Marco Biffi.

sono il frutto della nostra specializzazione informatica sviluppata in maniera esemplare proprio nel campo della lessicografia. Per prima cosa, abbiamo a disposizione una versione elettronica del Tommaseo-Bellini, quello che resta pur sempre il maggior vocabolario completo dell'Ottocento, per di più caratterizzato dalla fisionomia di affascinante 'dizionario d'autore'. Questa versione, interrogabile attraverso un motore di nuova concezione realizzato da Daniele Fusi, è frutto della politica di collaborazione messa in atto dall'Accademia attraverso contatti con il mondo imprenditoriale. Infatti il testo elettronico del dizionario Tommaseo-Bellini è stato offerto senza oneri dalla casa editrice Zanichelli di Bologna, nel quadro di un accordo di scambio e collaborazione formalizzato nel gennaio 2014. A seguire, troviamo l'accesso al più grande vocabolario del Novecento italiano, il GDLI (*Grande dizionario della lingua italiana*) della UTET, comunemente noto come 'Battaglia', dal nome del primo direttore. Sono attualmente presenti in libera consultazione tutti i volumi del GDLI (tranne il *Supplemento 2009*) in una versione provvisoria e sperimentale, certamente perfettibile, a cui stiamo dedicando un lavoro di revisione che proseguirà nei prossimi anni. La realizzazione è stata prima di tutto una sfida di fattibilità e di velocità, visto che l'opera a stampa si presentava di enormi dimensioni, non era disponibile digitalizzata (anche rozzamente tale) dall'editore dei volumi a stampa (non si sono ritrovati nemmeno i *file* di tipografia dei volumi più recenti); i caratteri di stampa erano di dimensione molto piccola e (nonostante le apparenze) non erano regolari nel corso della lunga realizzazione, durata quarant'anni. Purtroppo questa irregolarità tipografica ha reso impossibile il riconoscimento automatico dei lemmi. Si è comunque deciso di mettere immediatamente lo strumento (realizzato a tempo di record) nelle mani degli studiosi, benché presentasse difetti. Anche in questo caso, vale la pena insistere su questo punto, siamo di fronte al risultato di una collaborazione esemplare con una casa editrice, perché i diritti per l'uso del GDLI sono stati concessi dall'editore alla Crusca in forma non onerosa, secondo un accordo delineato nel settembre del 2014, poi formalizzato in maniera piena il 12 settembre 2017, durante un apposito evento svoltosi nella Villa medicea di Castello. Lo storico accordo ha avuto come esito la consultazione gratuita via web di un'opera lessicografica monumentale e di fondamentale importanza per lo studio e la conoscenza della lingua italiana. UTET Grandi Opere ha rinunciato a qualsiasi genere di compenso, mostrandosi sensibile alle esigenze della cultura, al servizio di un grande progetto che ha unito le forze di un'azienda editoriale privata (parte del gruppo Cose Belle d'Italia) e quelle della nostra istituzione statale. L'Accademia si è fatta carico, oltre che della digitalizzazione, del motore di interrogazione, realizzato da Giovanni Salucci. Come si è detto, la digitalizzazione attuale non è certo perfetta. Non intendiamo nascondere. Carente è la marcatura: per le ragioni già esposte, non è stato possibile marcare in maniera automatica i lemmi, e non era praticamente pensabile, con le risorse disponibili e con i tempi brevi che ci eravamo imposti, realizzare una marcatura manuale. La ricerca da parte dell'utente avviene dunque per forme, non per lemmi, anche se il lemma può essere individuato intuitivamente, badando all'ordinamento alfabetico e alla quantità di occorrenze negli esempi. Tuttavia, pur con questo limite, lo strumento è stato apprezzato dagli studiosi di tutto il mondo, perché permette non solo l'utilizzazione, in qualunque luogo ove esista un collegamento Internet, di uno strumento che in carta era ben poco maneggevole (tutti sanno che l'opera consiste in 21 volumi di grande formato e peso, senza contare supplementi e indice degli autori). Inoltre la ricerca per forme, per quanto rozza, consente il reperimento di ma-

teriale che altrimenti sfuggirebbe alla ricerca alfabetica per lemmi. Ne deriva quella che potremmo definire una lettura 'nuova' di un GDLI rinato dalle proprie ceneri.

Con il GDLI digitalizzato si chiude per ora la serie dei vocabolari 'storici' dell'italiano resi nuovamente vivi e vitali dall'Accademia. Lo strumento che segue, negli *Scaffali digitali*, è il *Vocabolario dantesco* (VD), il contributo più importante che l'Accademia offre alle celebrazioni dei 700 anni dalla morte dell'Alighieri. La raccolta degli *Scaffali digitali* prosegue con una serie di realizzazioni progettuali specifiche, alcune delle quali sono state già descritte in precedenza. Dunque ci soffermeremo ora sulle nuove realizzazioni, su cui finora non abbiamo ancora portato il discorso, cioè *ArchiDATA* e la *Stazione lessicografica*.

L'origine di *ArchiDATA* congiunge nuovamente, come il GDLI, l'attività del nostro ente pubblico statale e il contributo dell'imprenditoria privata. Si collega a un progetto PRIN 2012, proseguito con un PRIN 2015. Ancora prima, tuttavia, ci fu il collegamento con il vocabolario *Zingarelli* della casa editrice Zanichelli: un primo nucleo del materiale *on line* è infatti costituito da retrodatazioni che furono consegnate all'editore Zanichelli per lo *Zingarelli 2017*. *ArchiDATA*, *Archivio di (retro)datazioni lessicali*, è in sostanza un sito dedicato all'aggiornamento delle date di prima attestazione delle parole, con particolare attenzione per il lessico moderno. Caratteristica peculiare e unica di questa banca dati sta nel fatto che la prova delle nuove datazioni non è una trascrizione (con i possibili errori e refusi), ma una riproduzione fotografica della fonte. Non soltanto viene mostrata la fotografia della pagina in cui ricorre la parola cercata, ma anche compare la documentazione fotografica del frontespizio dell'opera di riferimento. La fotografia è assente solo nel caso in cui il riferimento stesso sia ricavato da bibliografia secondaria, cioè nei casi in cui *ArchiDATA* riprende di seconda mano un risultato ottenuto da qualche studioso nel corso delle sue ricerche pubblicate altrove. Va precisato dunque che questo DB non si limita a proporre i suoi risultati originali, cioè le sue retrodatazioni, anche se questo è l'obiettivo primario, ma mira ad affermarsi come punto di riferimento nazionale per la datazione del lessico. Per questo raccoglie, debitamente citandoli, i risultati raggiunti da altri studiosi estranei alla Crusca. Inoltre *ArchiDATA* fornisce la datazione di singole accezioni. Parole e accezioni sono marcate per ambito d'uso, quindi è anche possibile accedere all'archivio sezionando variamente il lemmario. L'ideatrice della piattaforma e attuale Direttrice tecnica, Ludovica Maconi dell'Università del Piemonte Orientale, ama ricordare che questo portale si lega al fascino che ancora oggi esercita su chi pratica studi lessicografici il «prezioso schedario» di Migliorini, quella «cassettiera fatta costruire nel 1941 a misura delle sue personalissime schede di 10 x 11 cm», descritta dal figlio Paolo.<sup>20</sup> La cassetta, che ancora si conserva in Crusca, e che è stata riprodotta fotograficamente sulla copertina del primo dei *Quaderni di ArchiDATA* uscito alla fine del 2020,<sup>21</sup> era stata progettata in un'epoca in cui ancora non esistevano i computer. Grazie a questo schedario, Bruno Migliorini «di ogni

<sup>20</sup> P. Migliorini, *Un ricordo di mio padre*, in M. Santipolo e Matteo Viale (a cura di), *Bruno Migliorini, l'uomo e il linguista (Rovigo 1896 – Firenze 1975)*, Atti del Convegno di studi (Rovigo, Accademia dei Concordi, 11-12 aprile 2008), Rovigo, Accademia dei Concordi, 2009, p. 4.

<sup>21</sup> Cfr. L. Maconi (a cura di), *Laboratorio di ArchiDATA 2020. Retrodatazioni lessicali: storia di cose e di parole*, Firenze, Accademia della Crusca, 2020 (si tratta del primo *e-book* realizzato integralmente dall'Accademia della Crusca, senza l'appoggio di editori esterni).

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».<sup>22</sup> *ArchiDATA* si prefigge questo stesso compito, ovviamente aggiornando con coerenza le ambizioni, adeguandole allo stile del nostro tempo, all'era dell'elettronica. Il sito è curato per la parte informatica (per la maschera di consultazione, per la maschera di immissione dei dati e per i meccanismi di ricerca); dall'ingegnere pavese Michele Lavezzi, ovviamente sotto le indicazioni di L. Maconi e del comitato scientifico del progetto (formato da V. Coletti, M.A. Cortelazzo, P. D'Achille e C. Marazzini). L'archivio va progressivamente arricchendosi di nuovi documenti per la storia del lessico italiano, e inoltre, come già si è detto, *ArchiDATA* vorrebbe assumere il ruolo di centro di raccolta di dati su retrodatazioni lessicali, utilizzando anche i contributi che escono in sedi disparate, con inevitabile dispersione di risultati che sarebbe invece utile convogliare in un'unica sede, accessibile, aggiornabile, riordinabile e ampliabile. Al dicembre 2020, *ArchiDATA* è arrivato a 7500 retrodatazioni lessicali e 1200 retrodatazioni di locuzioni (nel giugno del 2017, all'inizio dei lavori, le retrodatazioni erano 515). Sicuramente oggi indicare la datazione di una parola facendo solo riferimento al GRADIT o allo *Zingarelli*, o al DELI, comporta il rischio di dimenticare qualche cosa: la necessità di una rapida e comoda verifica dei dati presenti in *ArchiDATA*, dunque, si impone, e si imporrà sempre di più, ci auguriamo, con il procedere dei lavori. Fra l'altro, benché *ArchiDATA* sia uno strumento ideato soprattutto per la retrodatazione di parole moderne, sono state inserite retrodatazioni di parole antiche mediante il TLIO, con apposito rinvio all'OVI. Inoltre è stata attivata una collaborazione tra *ArchiDATA* e AVSI, l'*Archivio per il vocabolario storico italiano*, mediante la quale vengono acquisite le retrodatazioni pubblicate negli articoli della nuova rivista lessicografica (per ora quasi quattrocento).

Concluderemo con una breve panoramica sulla *Stazione lessicografica*, nata da un'idea di Marco Biffi, che si presenta come uno strumento per specialisti lessicografi e che dovrebbe facilitare il confronto tra una grande quantità di dati. Si tratta di un sistema integrato di consultazione di fondamentali dizionari in rete (storici, sincronici, di ortografia e pronuncia), banche dati di retrodatazioni, corpora dell'italiano, archivi di giornali. Si tratta di risorse che sono raggiungibili anche per altra via, non tutte di proprietà dell'Accademia della Crusca, ma che qui si trovano riunite per comodità dell'utilizzatore, in modo che sia possibile avere in una sola schermata la risposta a un'interrogazione unica, con notevole risparmio di tempo e di fatica. La *Stazione* è nata come piattaforma di lavoro per il VoDIM, il Vocabolario Dinamico dell'Italiano Moderno messo in cantiere dalla Crusca tra i propri programmi strategici, in connessione con un PRIN 2015, ma ora a disposizione degli studiosi e dei cultori della nostra lingua, senza limitazione alcuna per l'accesso e l'impiego. La stazione lessicografica si avvia cercando un lemma o una forma, e restituisce da sé il risultato, se possibile, del lemma, altrimenti della forma di cui il sistema non è stato in grado di trovare il lemma. L'esito della ricerca si presenta con un ordine variabile: se è stato individuato un lemma, viene mostrata prima di tutto la definizione in uno dei vocabolari post-unitari scelti tra quelli preselezionati dal sistema, a partire dalla V Crusca e dal Tommaseo-Bellini. Se la ricerca in questi vocabolari non ha avuto esito, viene proposta la definizione tratta da un dizionario contemporaneo. Il lemma

<sup>22</sup> P. Migliorini, *Un ricordo di mio padre*, cit., p. 4.

proposto in apertura della ricerca, il primo in alto tra i risultati, in forma di voce lessicografica, è seguito da una serie molto ampia di altre risorse, alcune elaborate e poste in Rete dall'Accademia, altre frutto esclusivo del PRIN 2015 – Bando 2015 (Prot. 2015PTXT4Y). Accanto ai dizionari generali compaiono banche dati (quella del VoDIM, il già citato *ArchiDATA* ecc.), repertori speciali (dizionari di ortografia e pronuncia), archivi di giornali. Un apposito indicatore colorato specifica il tipo di strumento: rosso per i Dizionari generali, granata per i Dizionari di ortografia e pronuncia, verde per *ArchiDATA*, azzurro per la Banca dati (tra le quali la Banca dati VoDIM e CoLIWeb, sviluppate nel citato PRIN 2015), viola per gli archivi di giornali. In caso di mancanza di risultato nella ricerca in una di queste risorse, compare l'indicazione numerica '0'. Lo strumento si presenta dunque come un approccio complessivo al lavoro lessicografico, con particolare evidenza assegnata a strumenti nati nel periodo post-unitario, quindi in grado di documentare il lavoro lessicografico svolto nella fase dell'italiano moderno e contemporaneo. Da ognuna delle risorse si può transitare immediatamente al sito che presenta la risorsa specifica, e ciò permette di utilizzare lo strumento prescelto in maniera più completa, concentrandosi su di esso e staccandolo dal complesso articolato e coordinato della *Stazione lessicografica*. Proprio per questo carattere comprensivo e sintetico, abbiamo descritto per ultima la *Piattaforma*. Essa rappresenta, più che un punto d'arrivo, un punto di partenza, l'avvio di altre ricerche lessicografiche, per le quali, ci pare di poter concludere, l'Accademia della Crusca, grazie al suo impegno costante, mostra di aver raggiunto risultati ragguardevoli, collocandosi in una posizione avanzata, come punto di riferimento indiscutibile e solidamente costituito, e inoltre con la garanzia di durata nel tempo. Molte volte abbiamo visto siti interessanti cadere in oblio e in abbandono per ragioni contingenti. Un'accademia con alle spalle cinquecento anni di storia e di lavori lessicografici, in questo senso, può fornire le migliori garanzie.