

La Crusca in rete

Marco Biffi

1. *Un po' di storia*

L'Accademia della Crusca si è sempre contraddistinta, fin dalla sua fondazione, per una spiccata attenzione alla tecnologia (lo testimonia la scelta di un simbolo come il *buratto*, o *frullone*, che alla fine del Cinquecento era una macchina d'avanguardia) e per una vocazione alla progettazione e realizzazione di strumenti fortemente innovativi (come di fatto è stato il *Vocabolario*). Non stupisce quindi che questa propensione dell'Accademia nel Novecento trovi precoce applicazione nel campo di intersezione tra informatica e linguistica, in particolare, tra informatica e lessicografia. La linguistica computazionale stava muovendo i suoi primi passi quando, agli inizi degli anni Sessanta, l'Accademia della Crusca si orientò con grande lungimiranza verso l'impiego degli strumenti che in questo ambito si stavano mettendo a punto. Sotto la presidenza di Giacomo Devoto e la direzione di Aldo Duro si cominciò infatti a realizzare un grande dizionario storico basato su un *corpus* di testi informatizzati. Questo avrebbe permesso ai lessicografi di effettuare con maggiore celerità e precisione lo spoglio delle fonti, vale a dire la tradizionale operazione di schedatura di ogni singola parola di un testo per raccoglierne sistematicamente i contesti d'uso al fine di compilare la voce di un dizionario che tenga conto dell'evoluzione dei significati nel corso del tempo. I computer erano all'epoca quelle apparecchiature di grandi dimensioni che si vedono ormai soltanto nei film di spionaggio degli anni Sessanta, i dati andavano inseriti in schede perforate e i risultati si radunavano in lunghissimi tabulati cartacei, ma l'Accademia già perseguiva tenacemente la strada dell'informatizzazione, con un metodo innovativo e anticipatore di quella che oggi può dirsi una prassi diffusa: l'impiego di *corpora* rappresentativi come base per la compilazione di strumenti lessicografici, storici e sincronici.

La progettazione di un dizionario storico era iniziata sotto la presidenza di Bruno Migliorini negli anni Cinquanta. Tra il 1953 e il 1954 Giovanni

Nencioni preparò una *Relazione dell'Accademia della Crusca sul Vocabolario della lingua italiana*, che fu presa in esame e discussa dal corpo accademico il 21 gennaio 1954 (e poi pubblicata negli «Studi di Filologia Italiana»); e nel 1957 Migliorini presentò il progetto al colloquio internazionale *Lessicologia e lessicografia francesi e romanze. Orientamenti ed esigenze attuali*. All'interno del convegno, organizzato dal CNR francese come punto di partenza per l'avvio ai lavori del proprio dizionario storico moderno, i partecipanti, provenienti dalle istituzioni accademiche europee di eccellenza, si scambiarono notizie sulle principali imprese lessicografiche in corso e condivisero l'idea di usare tecniche di spoglio e inventariazione meccanografici (che furono a breve seguiti da quelli elettronici). Così nel 1964 Devoto affidò la direzione dei lavori ad Aldo Duro, nel quadro di una fitta rete di contatti col *Trésor de la langue française* che consolidò e migliorò l'approccio informatico alle fasi di spoglio per la stesura delle voci del dizionario.

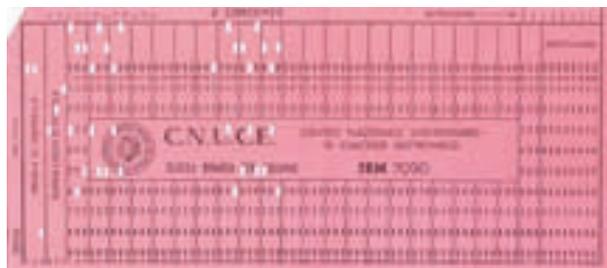
Per respirare quell'aria di novità, di entusiasmo e di fiducia in questo metodo, è di grande effetto "ascoltare in diretta" proprio la voce Aldo Duro, a lavori ormai iniziati e anzi nel momento della presentazione dei primi parziali frutti, nella *Prefazione alle Concordanze degli Inni Sacri di A. Manzoni*:

Questa che presentiamo è una "lista di concordanze" degli *Inni sacri* del Manzoni: un elenco alfabetico cioè di tutte le parole che compongono i cinque inni, seguite ciascuna dalla trascrizione di tutti i contesti nei quali ricorre. [...] È stata la nostra generazione che ha riscoperto la grande utilità che questi repertori possono recare agli studi, e che ne ha promosso la compilazione per altri autori e per altre opere. Si tratta però di compilazioni lentissime, faticose e ingrato, quasi anacronistiche rispetto all'epoca nella quale viviamo, incalzata dalla fretta di fare e di concludere. Poche perciò se ne vedrebbero pubblicate, se una macchina modernissima e divenuta quasi un simbolo del nostro tempo, il calcolatore elettronico, non si fosse rivelata velocissimo strumento dell'uomo nelle operazioni di

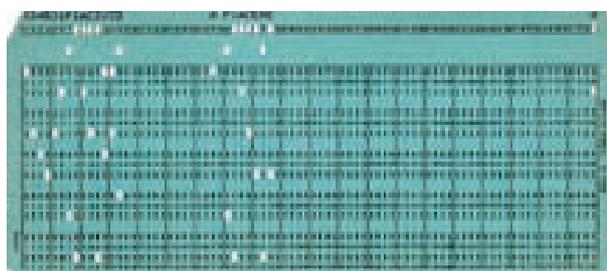
analisi, selezione, elaborazione che bisogna compiere sui dati linguistici per la formazione di liste di concordanze. A queste macchine ha fatto ricorso il Centro di lessicografia dell'Accademia della Crusca – così come molti altri centri e istituti di ricerca linguistica e lessicografica in quasi tutti i Paesi del mondo – per averne valido aiuto negli immensi spogli che esso ha predisposto per la redazione del grande Vocabolario storico della lingua italiana.

Nell'ambito dei lavori preparatori alla banca dati informatizzata che doveva costituire la base di partenza per la compilazione delle voci del dizionario, si contano alcuni prodotti collaterali parziali, come le appena citate *Concordanze degli Inni Sacri di A. Manzoni* (tra le prime concordanze per testi di lingua italiana), le concordanze della *Novella del grasso legnaiuolo*, le *Concordanze del «Decameron» di Giovanni Boccaccio* e le *Concordanze del Canzoniere di Francesco Petrarca*. Ma il progetto del dizionario storico non arrivò a compimento in seno all'Accademia: nel gennaio del 1983 il CNR assunse in proprio l'impresa (nel 1985 venne costituito il Centro di Studi Opera del Vocabolario Italiano del CNR in convenzione con l'Accademia della Crusca, diretto da Carlo Alberto Mastrelli fino al 1992, poi da Pietro G. Beltrami, al cui contributo in questo volume si rimanda per la storia del vocabolario da questo punto in poi).

L'interruzione del progetto segnò una battuta d'arresto nell'attività informatica dell'Accademia, che fu ripresa successivamente nel quadro completamente diverso della valorizzazione del patrimonio, con la realizzazione della versione elettronica del *Vocabolario degli Accademici della Crusca*, nella prima edizione del 1612. Fu subito chiaro, infatti, all'allora presidente Giovanni Nencioni e a Severina Parodi (Accademica segretaria, archivista e storica dell'Accademia) che una versione elettronica avrebbe consentito un "rovesciamento" del testo, vale a dire un accesso alfabetico non solo al lemma ma anche al suo contenuto, così da far cadere nel setaccio della ricerca anche quelle parole (moltissime, lo si sapeva e percepiva, ma non si poteva numerarle ed elencarle) che, nascoste tra le pieghe delle definizioni e dei commenti degli Accademici, sfuggivano al lemmario legato al modello fiorentino trecentesco. In realtà l'idea di un'informatizzazione della prima edizione del *Vocabolario* aveva già cominciato a delinearsi negli anni Settanta, quando ancora erano in essere i lavori del dizionario storico. Sotto la guida e per impulso di Giovanni Nencioni,



Scheda perforata relativa al lemma *contento*.



Scheda perforata relativa al lemma *piacere*.



Le *Concordanze degli Inni Sacri di A. Manzoni*, 1967.

This image shows a page from the Concordanze section of the Vocabolario della Crusca. It features a list of words, with the word 'contento' highlighted as the lemma. The page is organized into columns, with the lemma in the first column and its occurrences in various texts in the subsequent columns. The text is dense and follows a traditional layout for concordances.

Una pagina delle *Concordanze* (con il lemma *contento*).

This image shows another page from the Concordanze section of the Vocabolario della Crusca. It displays a list of words and their occurrences in various texts, organized into columns. The page is dense with text, showing the frequency of each word across different sources. The layout is similar to the previous page, with a focus on the frequency of the words listed.

Una pagina delle *Concordanze* (indice di frequenza decrescente).

da un lato furono portate a compimento nuove ricerche sulla prima edizione del *Vocabolario* (*Gli atti del primo Vocabolario* a cura di Severina Parodi), accompagnate dalla ristampa anastatica del 1974; e dall'altro si formularono i primi approcci per una schedatura elettronica, resi pubblicamente noti fra il 1979 e il 1983 in vari convegni (in particolare il "rovesciamento" divenne uno dei punti di discussione di rilievo all'interno delle celebrazioni per il quarto centenario dell'Accademia nel 1983). Con il coordinamento di Mirella Sessa iniziò dunque, in collaborazione con la Scuola Normale Superiore di Pisa, un ventennale progetto di informatizzazione del dizionario, che, dopo fasi alterne (legate al rapido e continuo cambiamento delle procedure di indicizzazione e gestione dei dati), è approdato nel 2001 alla versione web del *Vocabolario del 1612*, consultabile dal sito dell'Accademia oltre che da quello della Scuola Normale.

Nel frattempo l'attività informatica era riemersa quasi carsicamente nel 1993 con la precoce adesione

dell'Accademia ai progetti di catalogazione in linea delle biblioteche. E l'informatica umanistica aveva avuto in Accademia un impulso molto forte sul finire degli anni Novanta, quando la rivoluzione digitale e il successo del web avevano favorito la realizzazione di numerosi strumenti, grazie anche ai vari finanziamenti pubblici e privati che sempre più frequentemente e insistentemente avevano come oggetto la costruzione di banche dati e più in generale la fruizione dell'informazione attraverso la rete. Nel 2001, dopo un primo esperimento minimale realizzato nel 1996, si stava già attivamente lavorando alla realizzazione del nuovo sito web dell'Accademia, che fu inaugurato il 30 settembre 2002 e che ora, dopo dieci anni di intensa frequentazione da parte degli utenti della rete, si avvia a essere ulteriormente rinnovato; e si stava portando a termine la *Fabbrica dell'Italiano*, all'interno del progetto "Biblioteche speciali". Infine, proprio mentre si stavano concludendo i lavori dell'edizione elettronica del *Vocabolario del 1612*, in seno all'Accademia era già stato

elaborato il progetto per il “rovesciamento” di tutte e cinque le edizioni, completato nel 2004 con la realizzazione della *Lessicografia della Crusca in rete*.

Nei primi anni del nuovo millennio l'attività informatica dell'Accademia è cresciuta in volume e in ambiti applicativi, spaziando dalla valorizzazione del patrimonio librario e archivistico, alla progettazione di strumenti linguistici. Da qui la decisione di puntare sulla formazione di un gruppo di lavoro informatico interno che sovrintendesse in modo organico alla realizzazione dei vari progetti. Nel 2004 è stato così costituito il Centro Informatico dell'Accademia della Crusca, all'interno del quale collaborano informatici e linguisti, con l'intento di offrire un supporto e un coordinamento a tutta l'attività informatica, ma anche di condividere con altre istituzioni ed enti di ricerca le esperienze e le competenze maturate. Un'importante linea strategica del Centro è anche quella della formazione di giovani laureati preparati ad affrontare i problemi legati all'applicazione di procedure informatiche allo studio della lingua italiana e alla linguistica in generale.

2. Un biglietto da visita digitale: il sito web dell'Accademia

Il sito web dell'Accademia si configura come un portale della lingua italiana e mette a disposizione degli utenti una grande quantità di informazione altamente qualificata, rivolgendosi tanto al largo pubblico quanto agli specialisti. È stato progettato e realizzato per essere una stazione di lavoro che divenisse un punto di riferimento per gli studi di lingua italiana, con un servizio di consulenza linguistica, un archivio bibliografico, una biblioteca virtuale di contributi scientifici e l'accesso alle banche dati e agli strumenti informatici realizzati, una centrale di smistamento dell'informazione sulla lingua italiana nel vasto oceano del web.

Particolare attenzione è stata posta alla modalità di accesso alle informazioni, che avviene secondo un duplice binario: da un lato il materiale è disponibile per essere consultato liberamente (come quando si cerca un libro fra gli scaffali di una biblioteca, lasciandosi tentare da legami impensati suggeriti da libri vicini che non avremmo mai preso in considerazione), dall'altro è indicizzato in modo da consentire una ricerca veloce a chi abbia qualcosa di specifico da trovare (e per questo sono previsti non soltanto un motore di ricerca globale, per rin-

tracciare parole all'interno di tutto il sito web, ma anche strumenti più raffinati studiati per le sezioni più specifiche).

Il sito web si articola in otto sezioni: *L'Accademia*, *Storia*, *Attività*, *Pubblicazioni*, *Notiziario*, *Biblioteca virtuale*, *Lingua in rete*, *Crusca in gioco*. La prima sezione unisce la necessità di presentare l'istituzione a quella di raccogliere in maniera ordinata varie informazioni di carattere pratico legate all'Accademia e alle sue strutture (tra cui Biblioteca e Archivio, le cui sottosezioni sono predisposte in modo da costituire di fatto delle pagine web autonome). Nella sezione *Storia* si ripercorrono le tappe fondamentali del lungo cammino dell'Accademia, dal 1583 (anno della sua stabile costituzione) fino a oggi. Le vicende contemporanee invece sono state meglio delineate nella sezione *Attività*, in cui vengono illustrate le varie collaborazioni che l'Accademia ha con università e istituti di cultura italiani ed esteri, e in cui ampio spazio è dedicato a illustrare i progetti attualmente in via di realizzazione. Fra le varie attività dell'Accademia va inserita anche quella editoriale, che ha un accesso specifico nella sezione delle *Pubblicazioni*, con un quadro sistematico sulle riviste (alla «Crusca per Voi», rivolta al largo pubblico, si affiancano tre riviste specialistiche «Studi di Filologia Italiana», «Studi di Grammatica Italiana» e «Studi di Lessicografia Italiana», per le quali si fornisce uno spoglio dettagliato degli articoli) e sui volumi (il *Catalogo* costituisce una guida al mondo editoriale dell'Accademia, con le informazioni necessarie all'acquisto dei libri, un catalogo interrogabile per autore, titolo, anno, collana). La sezione *Crusca in gioco* propone una raccolta di curiosità relative all'Accademia: dai suoi simboli alle imprese editoriali dal Seicento a oggi, dalle sue figure più illustri alle più recenti presenze televisive.

È nelle sezioni del *Notiziario*, della *Biblioteca virtuale* e della *Lingua in rete* che il sito web diventa portale della lingua italiana: nel *Notiziario* sono raccolte le indicazioni sui convegni e congressi, così come su qualunque iniziativa di interesse linguistico; dalla *Biblioteca virtuale* si entra nello spazio dedicato ai progetti informatici dell'Accademia; mentre la sezione *Lingua in rete*, senza dubbio quella più dinamica del sito, è proiettata sul fronte di una consulenza diversificata, con risposte ai quesiti linguistici, articoli, schede lessicografiche sui neologismi, informazioni sulla lingua italiana nel web.

Il primo e principale livello di servizio è offerto dalla *Consulenza linguistica* (sulla cui attività si rimanda per maggiori dettagli al contributo di Raf-



La "copertina" del sito web dell'Accademia.



La Biblioteca virtuale.



La pagina d'entrata del nuovo sito web dell'Accademia.

faella Setti in questo volume). La sezione relativa del sito è dedicata alle risposte sui dubbi linguistici proposti dai lettori e diventa quindi una stanza di incontro virtuale tra l'Accademia e un pubblico vasto ed eterogeneo. In essa è possibile porre un quesito linguistico, attraverso un modulo elettronico; e consultare le risposte già date su temi più generali (su cui si appunta l'attenzione del maggior numero di richiedenti), ma anche su argomenti più circoscritti o specifici (che soddisfano la curiosità o la necessità di un pubblico più limitato). Fin dai primi giorni d'apertura del sito web le pagine di consulenza sono quelle di gran lunga maggiormente consultate: i quesiti che hanno raggiunto la redazione dal settembre 2002 al 2011 sono oltre 36.000: una media di quasi 4.000 quesiti l'anno (considerando che nel 2002 il sito è stato attivo solo due mesi), in media 11 quesiti al giorno. E va notato che il loro numero è cresciuto enormemente negli ultimi anni: 2.353 nel 2006, 2.463 nel 2007, 3.333 nel 2008, 5.415 nel 2009, 7.693 nel 2010, 6.631 nel 2011; e la media giornaliera supera ormai da tempo i 30 quesiti, con punte di 40-50. Tra il 2002 e il novembre 2011 è stato possibile rispondere a oltre 6.000 persone, un numero considerevole tenuto conto delle esigue risorse disponibili per la consulenza: attraverso la pubblicazione sul sito di 337 schede sono state soddisfatte richieste su dubbi sollevati da più di 2.000 persone (negli altri casi si è provveduto con una risposta via mail).

Il sito web dall'ottobre 2002 al dicembre 2011 ha avuto un totale di circa 7.500.000 visite con oltre 200 milioni di contatti. L'attuale media mensile è di circa 222.000 visite (oltre 7.000 al giorno) con 6 milioni di contatti. Visto il volume di consultazione, all'interno di un mondo informatico e di una comunicazione web in continua e rapida evoluzione, nel 2011 l'Accademia, come si accennava più sopra, ha avviato la progettazione e la realizzazione di una nuova versione del suo sito web, che entrerà in funzione nel 2012, nel quarto centenario dell'uscita della prima edizione del *Vocabolario*.

3. Scaffali digitali: manoscritti e libri in rete

Molti dei progetti informatici sono dedicati alla valorizzazione del patrimonio archivistico e librario dell'Accademia. A questo filone vanno ricondotte tutte le banche dati che possono essere riunite sotto l'etichetta di "Biblioteca Digitale" e di "Archivio Digitale", vale a dire quelle banche dati bibliografiche

e archivistiche che forniscono l'accesso diretto alla riproduzione digitale dei materiali schedati.

Al passaggio in digitale del patrimonio librario è stata posta particolare attenzione negli ultimi anni sia a livello pubblico (ad esempio con varie iniziative del Ministero dei Beni e le Attività Culturali e della Regione Toscana che hanno coinvolto direttamente l'Accademia) sia a livello privato (si pensi a *Google Libri*). L'Accademia ha iniziato molto precocemente il suo rapporto con l'idea di biblioteca digitale, ed esattamente durante la progettazione (nel 1998) della *Fabbrica dell'Italiano*, che ha rappresentato un'interessante novità anche dal punto di vista metodologico. Due delle banche dati della *Fabbrica* sono infatti dedicate ai dizionari e alle grammatiche possedute dalla Biblioteca dell'Accademia: alle schede del catalogo in linea sono associate immagini ritenute rappresentative, perché funzionali a una più precisa descrizione del volume. L'idea sottesa è che, volendo conoscere le caratteristiche e gli scopi di un dizionario o di una grammatica senza prendere i volumi dallo scaffale, serve in particolar modo considerare alcune pagine specifiche: il frontespizio, la prefazione e/o l'introduzione (dove l'autore espone i criteri adottati), alcuni campioni della trattazione e, nel caso delle grammatiche, gli indici. Per i dizionari sono state scelte alcune pagine del lemmario all'inizio della lettera *M*, in modo da rendere disponibili informazioni sulla struttura della voce e sulla "densità" del repertorio (per capire quanto il dizionario vada in profondità e quali settori del lessico tratti). Per entrambe le sezioni è stato realizzato un "approfondimento" sui dizionari anteriori al 1738 (anno in cui si conclude la quarta edizione del Vocabolario degli Accademici) e sulle grammatiche edite tra XVI e XVII secolo. Questa parte prevede un arricchimento dei dati: si tracciano agili biografie degli autori, si ricostruisce la storia editoriale di ciascuna opera, allargando il campo delle informazioni agli esemplari non posseduti dalla Biblioteca. Per quanto riguarda poi gli esemplari posseduti si infittisce l'apparato delle immagini (nella sezione grammatiche sono incluse le pagine del testo dedicato ai pronomi personali soggetto di terza persona) e le si accompagnano con trascrizioni messe in parallelo. Su questa grande massa di dati (2.374 opere lessicografiche e 376 opere grammaticali per un totale di circa 35.000 immagini; con approfondimento su 43 dizionari, per un totale di 67 edizioni, e su 16 grammatiche in 34 edizioni) sono state messe a punto strategie di ricerca informatica differenziate: da un lato, con maschere puntuali e dettagliate, si offre la possibilità di effettuare indagini molto precise;



La *Fabbrica dell'italiano*.



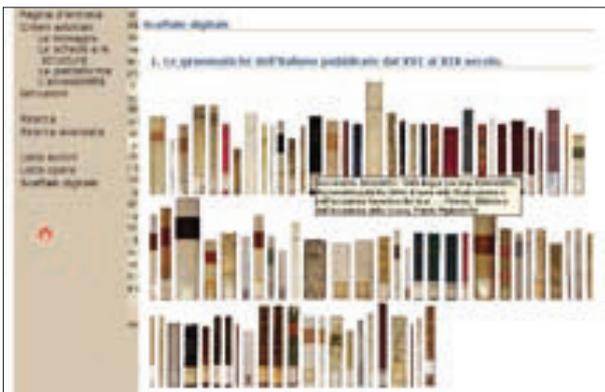
Scheda "opera" dei dizionari della *Fabbrica dell'Italiano*.



Scheda di approfondimento delle grammatiche della *Fabbrica dell'Italiano*.



La Biblioteca Digitale dell'Accademia della Crusca.



Lo scaffale digitale della Biblioteca.



Schermata dei risultati di una ricerca nella Biblioteca Digitale.

dall'altro, chi sfoglia gli scaffali virtuali può recuperare il ruolo della biblioteca come "suggeritrice" di informazioni e legami insospettiti.

La scelta della riproduzione di alcune pagine rappresentative del libro comportava la realizzazione di una biblioteca digitale "parziale" che, a fronte dei costi della riproduzione digitale assai elevati in quegli anni, costituiva una buona soluzione di compromesso e gettava le fondamenta per la futura realizzazione di una vera e propria biblioteca digitale integrale, in cui i libri sono consultabili per intero in rete; una possibilità che di lì a poco il progresso tecnologico avrebbe consentito grazie a nuovi strumenti web e all'abbassamento del costo della riproduzione digitale.

La *Fabbrica* contiene anche due sezioni archivistiche: una è dedicata all'*Archivio storico* e in essa sono consultabili la schedatura informatizzata dei *Diari e Verbali* effettuata da Severina Parodi, quella dei manoscritti relativi ai concorsi letterari promossi dall'Accademia fra il 1810 e il 1835, e infine lo schedario degli Accademici; l'altra, intitolata *Lessico tecnico* (sottotitolo *Banca dati del lessico tecnico enciclopedico nelle carte del Cardinale Leopoldo de' Medici*), è una schedatura informatizzata del materiale lessicale raccolto a partire dal 1650 dal principe cadetto medico e ora pubblicato dall'Accademia per cura di Raffaella Setti (*Le parole del mestiere*). Le due sezioni dedicate all'Archivio costituiscono la preistoria di un archivio digitale, in cui la catalogazione fornisce l'accesso ai materiali in riproduzione digitale.

Alla *Fabbrica* è seguita la realizzazione di un'altra biblioteca digitale parziale, *Il Fondo dei citati*, una banca dati dedicata al fondo librario usato per la compilazione del *Vocabolario* e che quindi, per quanto riguarda le immagini, punta l'attenzione sui frontespizi e sui fogli di guardia (che registrano la storia del libro e delle sue collocazioni consentendo quindi di tracciarne la storia) e sulle pagine postillate dagli Accademici (per ricostruire la delicata attività preparatoria delle schede di spoglio poi alla base delle voci lessicali).

Nel 2007 l'Accademia è riuscita a mettere a disposizione del largo pubblico anche i primi scaffali di una sua biblioteca digitale di volumi riprodotti integralmente. La banca dati, che prende il nome di *Biblioteca Digitale dell'Accademia della Crusca*, contenente circa 60.000 immagini, è stata successivamente ampliata nel 2008 e comprende così 121 edizioni per un totale di circa 111.000 immagini. Il suo primo "fondo digitale", dal titolo *Fonti descrittive e normative dell'italiano: corpus digitale di testi dal XVI*

al XIX secolo, si snoda su 4 nuclei fondamentali: 1) le grammatiche dell'italiano pubblicate dal XVI al XIX secolo; 2) i testi di discussione linguistica intorno alla polemica anti-Crusca; 3) le edizioni non ufficiali del *Vocabolario degli Accademici della Crusca*; 4) la lessicografia ottocentesca.

I volumi digitali possono essere sfogliati pagina dopo pagina, e possono essere “navigati”, parzialmente, attraverso un indice che propone una struttura molto articolata. Questo è forse uno degli aspetti più degni di nota della biblioteca digitale dell'Accademia, perché si è posta particolare attenzione alla descrizione delle strutture, che consentono un'ampia fruizione delle opere: da un lato sono messi in evidenza i luoghi di interesse librario (coperta e tagli, carte di guardia, frontespizio, note tipografiche ecc.), dall'altro la dettagliata scansione dei capitoli e dei paragrafi pone a esponente gli argomenti trattati e guida il lettore nella consultazione. L'accesso ai volumi segue la doppia impostazione consueta per gli strumenti dell'Accademia: quella per il visitatore e quella per il ricercatore. Quest'ultimo può accedere a un motore di ricerca libera o avanzata (con possibilità di filtro sui campi descrittivi previsti e sulle strutture). Nel caso di una “visita” è possibile scorrere una lista di autori e una di opere; ma l'accesso più bello è sicuramente quello dello *Scaffale digitale*, che proietta il lettore in uno scaffale virtuale su cui tutti i volumi sono appoggiati e visibili di costola: passando il puntatore del mouse sopra il dorso di ciascun libro, un'etichetta mostra il titolo dell'opera e il consultatore può con un semplice clic estrarre il volume dalla libreria e iniziare la lettura.

Un ulteriore scaffale della Biblioteca Digitale di Crusca è costituito dalla banca dati dedicata all'«Antologia» di Gian Pietro Vieusseux (*Antologia Vieusseux 1821-1832*), con i suoi 48 volumi resi disponibili in versione digitale (per un totale di 26.000 pagine). Gli articoli della rivista sono stati schedati e questo permette la consultazione di uno spoglio digitale con l'accesso diretto al testo ricercato. Un motore di ricerca consente di selezionare gli articoli in base all'autore, al titolo, alla data di pubblicazione, all'argomento trattato (un filtro di particolare utilità nell'ampio spettro di varietà di lingua, anche tecnica, presente nell'«Antologia»).

All'interno di specifiche esigenze progettuali ha continuato a essere impiegata anche la strategia della biblioteca digitale “parziale”, come avviene in due recenti progetti realizzati dall'Accademia. La prima, *Le Cinquecentine della Crusca*, è una banca dati che mette a disposizione degli studiosi la raccolta di ope-



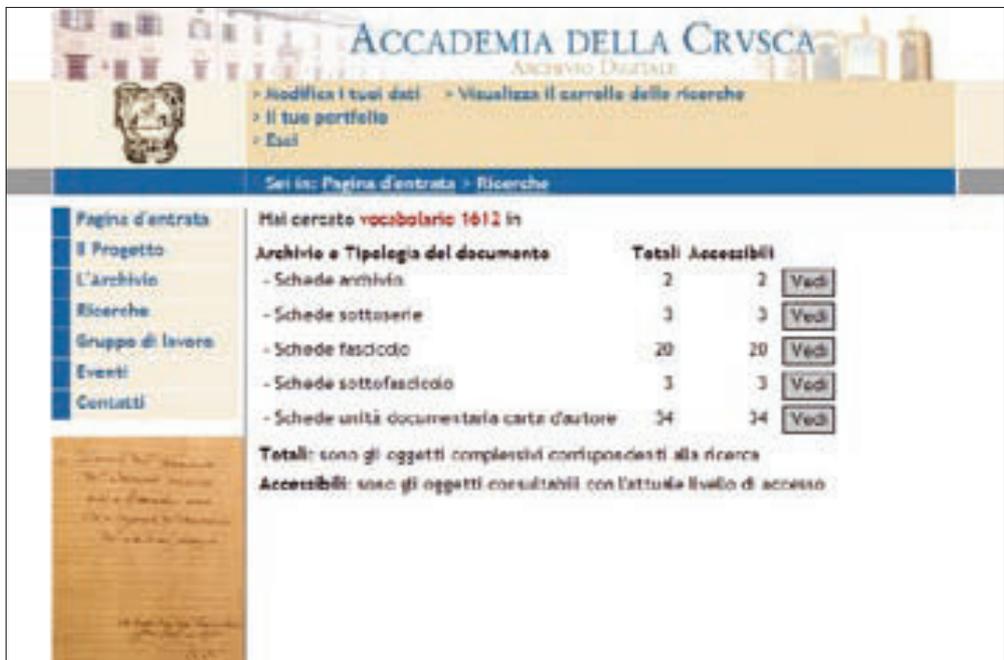
Scheda della *Biblioteca Digitale* con riproduzione del facsimile del documento.



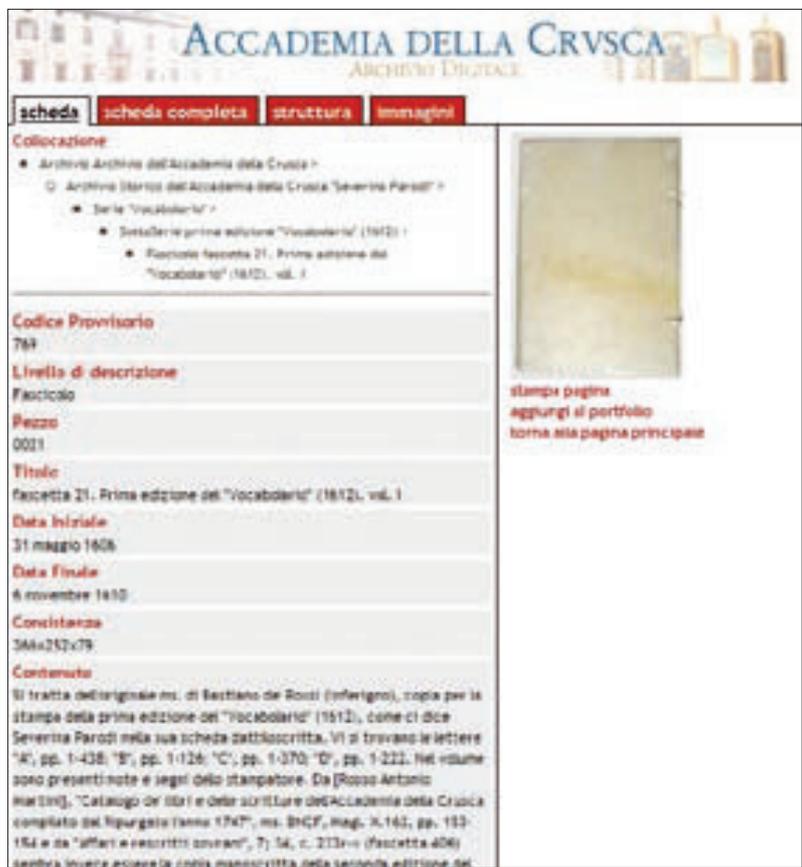
Antologia Vieusseux 1821-1832.



Le Cinquecentine della Crusca.



Schermata dei risultati di una ricerca nell'Archivio Digitale.



Scheda dell'Archivio Digitale con riproduzione del facsimile del documento.



Riproduzione del facsimile del documento d'archivio e relativa struttura.



La *Lessicografia della Crusca in rete*.

La *Lessicografia* in numeri

- 20.000 pagine in facsimile
- 209.547 forme
- 11.605.463 occorrenze
- 142.169 voci complessive
- 400.715 esempi d'autore
- 6.443 espressioni dell'uso vivo
- 2.658 proverbi
- 35.779 locuzioni
- 32.614 forme latine
- 26.076 forme greche

allarga in funzione dello specifico contenuto (così, ad esempio, la scheda di un'unità documentaria carteggio presenta il campo "mittente" e quello "destinatario"). Le schede descrittive sono raggiungibili sia visitando virtualmente gli armadi dell'Archivio, sia compiendo ricerche mirate con strumenti linguisticamente raffinati. Il sistema di gestione prevede per altro accessi controllati e personalizzati a seconda dei materiali e dei permessi degli utenti. L'*Archivio Digitale* rende a oggi consultabili oltre 3.670 schede di unità documentarie (manoscritti, lettere, dattiloscritte ecc.), a cui si aggiungono le schede dei livelli di articolazione superiore (fondo, serie, sottoserie, fascicolo, sottofascicolo) per un totale complessivo di 7.644 schede, con oltre 19.000 immagini e varie trascrizioni a esse associate (sui materiali si veda più dettagliatamente il contributo di Elisabetta Benucci in questo stesso volume). L'*Archivio Digitale* conta circa 500 visite mensili, con oltre 2.000 visualizzazioni di pagina e con una percentuale di nuove visite pari al 91%.

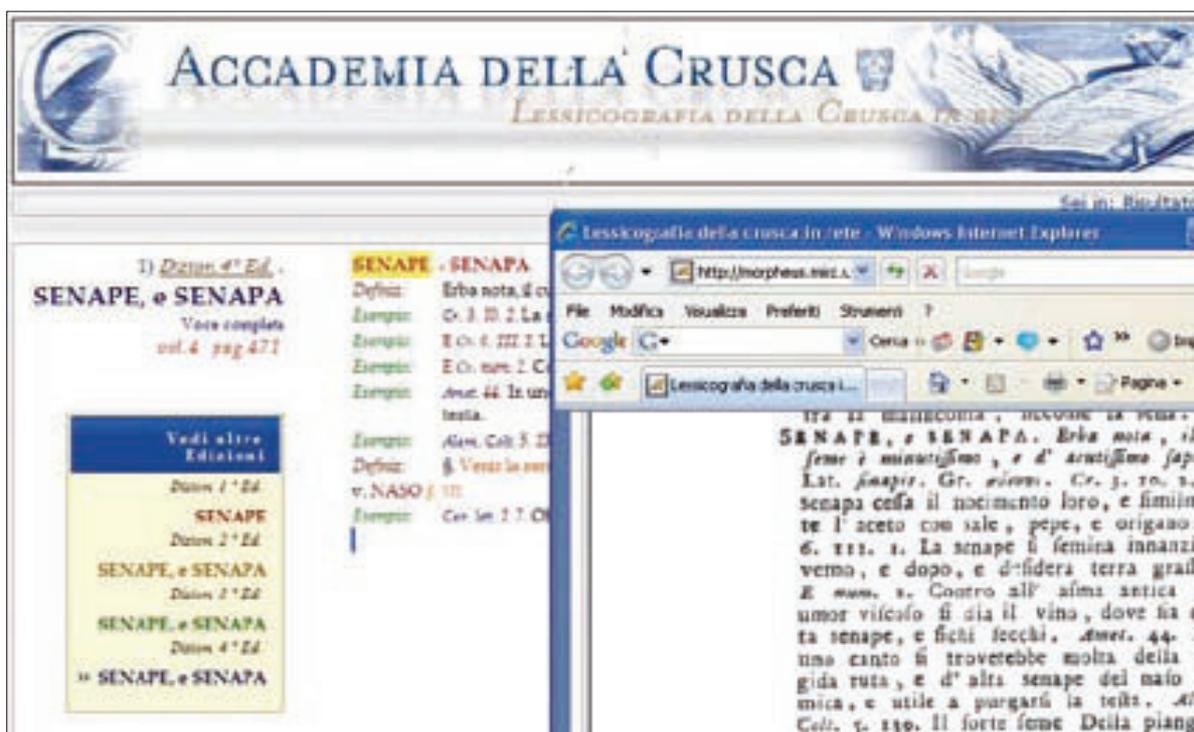
4. I vocabolari elettronici

Tra le banche dati realizzate, di particolare importanza è, senza dubbio, la *Lessicografia della Crusca in rete*, con cui l'Accademia ha reso disponibile sul web, per immagini, tutte le edizioni del suo *Vocabolario* (1612, 1623, 1691, 1729-1738, 1863-1923) e ha realizzato la versione elettronica delle prime quattro edizioni, con marcatura di macrocontesti e microcontesti. I numeri riassunti nella pagina di presentazione del sito sono già di per sé indicativi dell'importanza dello strumento: 20.000 pagine in fac-simile, 11 milioni di occorrenze, più di 200.000 forme, 142.000 voci complessive, più di 400.000 esempi d'autore, circa 6.443 espressioni dell'uso vivo, 2.658 proverbi, 35.779 locuzioni, oltre 32.000 forme latine e oltre 26.000 forme greche.

La trascrizione e la marcatura degli undici volumi sei-settecenteschi (in base agli standard XML/TEI) consentono di consultare le prime quattro edizioni in modo sistematico e rapido, sia con una ricerca semplice, sia con modalità di interrogazione più avanzate, così da rispondere alle diverse esigenze di un pubblico variegato. È sufficiente digitare una qualunque parola per ottenere istantaneamente i contesti in cui questa è usata, non soltanto a lemma, ma anche all'interno delle voci; risultato, questo, di non secondaria importanza per i *Vocabolari*, che si sono così rivelati ricche miniere di lessico nascosto fra le pieghe delle definizioni. Il consultatore



Schermata dei risultati di una ricerca nella *Lessicografia*.



Lessicografia: la voce *senape*.



La pagina d'entrata del *Vocabolario del Fiorentino Contemporaneo*.



Vocabolario del Fiorentino Contemporaneo: la voce *acquaio*.

nell'interrogazione: l'ambito semantico, la categoria grammaticale, la localizzazione fiorentina, le fonti, la vitalità d'uso e il registro all'interno dell'italiano, per citare i principali. Il *Vocabolario* è uno strumento aperto: attualmente conta 1.064 schede lessicali su un totale previsto di circa 7.000 (a partire da circa 20.000 entrate di partenza, ricavate con un'indagine sul campo con registrazioni per oltre 250 ore di parlato raccolto nei quartieri di Santa Croce, Rifredi e San Frediano).

5. Sotto le onde del web

Sotto le onde del web, in quel mare in cui gli internauti della rete amano *surfare*, numerosi sono gli strumenti informatici che l'Accademia ha realizzato per un uso interno o che sta tuttora mettendo a punto in vista di una pubblicazione in linea.

Alcuni strumenti sono strettamente legati all'attività di consulenza dell'Accademia e sono stati pensati per una fruizione interna. Tra questi grande importanza ha la *Banca dati della consulenza linguistica*, una piattaforma informatica che raccoglie e convoglia i vari quesiti provenienti dal sito, organizza lo smistamento ai redattori e raccoglie tutte le risposte inviate. Le risposte sono archiviate e affiancate da alcuni marcatori di classificazione (grafia, morfologia, sintassi, lessico, neoformazioni ecc.) e da parole chiave. Nel tempo si è in questo modo formata una ricca raccolta di oltre 6.000 schede che i redattori della consulenza possono interrogare in modo sistematico, grazie a un motore che le rende consultabili con un accesso per parole chiave o con una ricerca libera (di una o più forme), con la possibilità di applicare un filtro per classificazione, data, provenienza del quesito, redattore, stato di completamento della risposta. In questo modo è a volte possibile ridurre i tempi di risposta ad alcuni quesiti, recuperando un testo già pronto e limitandosi soltanto al suo eventuale aggiornamento. Sempre di grande importanza per il lavoro di consulenza è poi la *Banca dati del «Corriere della Sera» per gli anni 1992-2005*, realizzata grazie a una collaborazione con il Gruppo Editoriale Rizzoli-Corriere della Sera: i materiali messi a disposizione direttamente dall'editore sono stati organizzati in un corpus elettronico che riunisce 3.338.757 articoli di giornale opportunamente schedati, da cui è stata estratta una banca dati di potenziali neologismi composta da 4.790 schede interrogabili in base al lemma, all'anno di ingresso, all'anno di eventuale scomparsa, all'ambito d'uso.

Il Vocabolario del Fiorentino Contemporaneo in numeri

- circa 20.000 entrate nel corpus di partenza
- 1.064 schede realizzate
- 7.000 schede stimate
- 300 ore di parlato registrato nelle indagini di verifica

Fra gli strumenti a uso interno va ricordato anche l'*Archivio Iconografico Digitale* gestito da *Photocrusca*, uno strumento realizzato dal Centro informatico che consente la catalogazione delle immagini sulla base di ontologie descrittive affiancate anche da una *folksonomy* (cioè una classificazione in base a parole chiave assegnate dagli utenti), con un sistema di ricerca che prevede una ricca gamma di filtri. Attualmente la banca dati, che intende raccogliere tutte le immagini presenti in Accademia in modo da renderle disponibili ai vari servizi in modo centralizzato, contiene 6.490 schede e 7.709 immagini e funziona da raccordo anche per le immagini di alcune banche dati pubblicate, per un totale di circa 40.000 immagini catalogate e interrogabili globalmente.

Nascosta invece solo in via provvisoria, in vista del suo completamento e della messa a punto delle procedure informatiche di interrogazione, è la raccolta dei *Proverbi italiani*. Il progetto globale è teso a costituire un vasto corpus, comprendente testi manoscritti e opere a stampa, in collegamento con altre banche dati esistenti: si spazia dalle prime raccolte cinquecentesche sistematiche di Francesco Serdonati e Lionardo Salviati a quelle ottocentesche, in particolare di Giusti e di Capponi. Si prevede l'arricchimento con i proverbi contenuti nelle cinque edizioni del *Vocabolario degli Accademici della Crusca* (raggiungibili in modo automatico grazie alla *Lessicografia della Crusca in rete*) e con le componenti proverbiali emergenti dal *Vocabolario del Fiorentino Contemporaneo*. La banca dati offrirà un facile accesso a una porzione significativa del ricco patrimonio paremiologico italiano, permettendo vari tipi di interrogazione (tanto per argomenti quanto per parole) e consentirà di visualizzare immediatamente, in parallelo, lo stesso proverbio nelle sue varianti minime o sostanziali. Per ora è stata realizzata la piattaforma di gestione e interrogazione, mentre

Le domande alla Consulenza linguistica

Anno	Quesiti
2002	1.333
2003	2.307
2004	2.085
2005	2.680
2006	2.353
2007	2.463
2008	3.333
2009	5.415
2010	7.693
2011	6.631
	36.293

la banca dati è stata popolata con alcune raccolte: i *Proverbi toscani* di Giuseppe Giusti (sia quelli riportati alla forma originale secondo gli autografi, sia quelli riuniti nelle edizioni di Gino Capponi: risultano così consultabili 2.630 proverbi del *manoscritto A*, 2.888 del *manoscritto B*, circa 9.000 dell'edizione Capponi 1853, circa 11.500 dell'edizione Capponi del 1871); i proverbi di Lionardo Salviati (tenendo conto delle varianti emergenti dai manoscritti, per un totale di 3.136 proverbi); i proverbi di Serdonati (10.764 proverbi, per le lettere A-E).

Proprio nell'ambito della celebrazione dei 150 anni dell'Unità d'Italia è stata avviata nel 2011 una campagna di acquisizione digitale dei materiali preparatori alla quinta edizione del *Vocabolario* e dei neologismi raccolti da Cesare Donati nella seconda metà dell'Ottocento e conservati nell'Archivio, in vista della realizzazione di uno scaffale digitale che sarà pubblicato entro l'anno (per un totale di 35.000 immagini relative alle schede di circa 30.000 lemmi).

6. Oltre le colonne d'Ercole

L'obiettivo principale per il prossimo futuro è quello di omogeneizzare gli strumenti esistenti, in modo da creare una banca dati centralizzata dei materiali e degli strumenti informatici esistenti e di quelli in via di completamento.

Un primo elemento di aggregazione è stata la centralizzazione della procedura di registrazione necessaria per consultare le banche dati dell'Accademia: tutte infatti, a eccezione della *Lessicografia della Crusca in rete* e della *Fabbrica dell'Italiano*, prevedono un ingresso controllato attraverso un nome identificativo e una password, con vari livelli di accesso. Da tempo ormai la registrazione consente l'accesso base a tutte le banche dati dell'Accademia, mentre per una consultazione più articolata (ad esempio per accedere alle riproduzioni delle immagini di qualità maggiore e senza marcatura) sono richiesti specifici permessi a tempo che l'amministratore può attribuire in modo differenziato.

Ma la vera nuova frontiera è la creazione di un "metamotore" di ricerca che consenta di interrogare globalmente tutto l'insieme delle banche dati realizzate, rendendo comunicanti le diverse piattaforme, le diverse strutture dei dati, i diversi e specifici motori di ricerca.

Il primo passo verso questa nuova modalità integrata è stata la creazione di *AFLI* (*un Authority File per la Linguistica Italiana*), che unisce l'archivio di



La pagina di smistamento dei quesiti nella Banca dati della consulenza linguistica.



AFLI, un Authority File per la Linguistica Italiana.



Il DVD *La Fabbrica dell'Italiano*, 2011.



Il DVD *Insegnare italiano nelle classi multilingui*, 2010.

autorità della Biblioteca e dell'Archivio dell'Accademia della Crusca in un organico strumento di controllo sia della forma del nome degli autori che dell'accesso ai dati. L'architettura delle schede di registrazione è basata su un tracciato comune che tiene conto degli standard internazionali di archivi e biblioteche. La piattaforma informatica consente un allineamento sincrono con i record di autorità provenienti dai programmi di catalogazione dell'Archivio e dall'OPAC della Biblioteca e una loro corretta e costante integrazione nel record di autorità comune. La base di dati è per ora popolata con un primo campione di 2.200 schede, limitato a un gruppo di autori significativo nell'ambito della storia della lingua italiana e di quella dell'Accademia (941 registrazioni d'autorità sono state create a partire dagli autori dei lessici presenti nella sezione DIZ della Biblioteca; 1.265 sono relative agli Accademici e ai Soci corrispondenti dell'Accademia). Di ognuno dei nominativi è stata indicata la corretta intestazione (secondo le norme stabilite e la verifica di 10 repertori di riferimento internazionale), le date di esistenza comprensive dei luoghi, una breve biografia e le "qualifiche" che individuano l'attività.

AFLI costituirà la base di accesso normalizzato a tutti i nomi presenti nella galassia digitale dell'Accademia della Crusca, ma si sono anche sviluppati i primi stadi del "metamotore" generale. La versione *beta* finora realizzata consentirà di testare le criticità dello strumento in vista di un suo perfezionamento e della sua estensione a tutti i materiali presenti nelle banche dati dell'Accademia.

7. Le pubblicazioni digitali

Tra gli strumenti informatici realizzati dell'Accademia vanno ricordate anche numerose pubblicazioni su supporti ottici.

All'interno della collana "Testi e strumenti digitali" sono raccolti alcuni CD e DVD, fra cui numerosi titoli sono legati ai corsi di aggiornamento per insegnanti svolti all'interno dell'iniziativa "Insegnare italiano":

Insegnare italiano. Modelli per lo studio della lingua, 2006.

Francesco di Giorgio Martini, *Il Vitruvio magliabechiano in versione elettronica*, 2006.

Insegnare italiano. Il lessico tra grammatica e dizionari, 2008.

Insegnare italiano come lingua materna e come L2, 2009.

Insegnare italiano nelle classi multilingui, 2010.

La fabbrica dell'italiano. Dizionari e grammatiche, 2011.

Nel 2008 è uscita la versione elettronica del *Vocabolario* del 1612, realizzata per la parte informatica insieme all'Istituto dell'Opera del Vocabolario, di corredo alla ristampa anastatica del volume: *Vocabolario degli Accademici della Crusca. Riproduzione anastatica della Prima Edizione, Venezia 1612*, promossa dall'Accademia della Crusca in collaborazione con ERA Edizioni, 2008, Firenze-Varese, Accademia della Crusca-ERA, 2008.

8. La galassia della Crusca esplorabile in rete

Sito web: <<http://www.accademiadellacrusca.it>>

Dizionari elettronici

- *La Lessicografia della Crusca in rete*: <<http://www.lessicografia.it>>

- *Il Vocabolario del Fiorentino Contemporaneo*: <<http://www.vocabolariofiorentino.it>>

Biblioteche e archivi digitali

- *La Fabbrica dell'Italiano*: <<http://www.fabbricadellitaliano.it>>

- *Il Fondo dei Citati*: <<http://www.citatinellacrusca.it>>

- *Biblioteca Digitale dell'Accademia della Crusca*: <<http://www.bdcrusca.it>>

- *Archivio Digitale dell'Accademia della Crusca*: <<http://www.adcrusca.it>>

- *Le Cinquecentine della Crusca*: <<http://www.cinquecentine-crusca.org>>

- *Dalla Scuola senese all'Accademia della Crusca*: <<http://www.opere-senesi.org>>

- *Antologia Vieusseux 1821-1832*: <<http://www.antologia-vieusseux.org>>

La Crusca digitale del 2011 in numeri

Banche dati	Immagini	Schede	Opere	Lemmi/Voci	Spogli
<i>Archivio Digitale</i>	19.150	7.644	0	0	0
<i>Fabbrica dell'italiano</i>	33.104	4.000	2.750	0	0
<i>Fondo dei citati</i>	12.908	3.163	1.684	0	0
<i>Cinquecentine di Crusca</i>	52.885	1.132	1.132	0	0
<i>Biblioteca Digitale</i>	110.086	119	121	0	0
<i>Scuola senese</i>	10.522	75	47	0	0
<i>Neologismi di secondo Ottocento</i>	35.000	0	0	32.000	0
<i>Archivio Iconografico</i>	7.709	6.490	0	0	0
<i>Antologia Vieusseux</i>	26.585	0	0	0	6.040
<i>Lessicografia</i>	20.490	0	0	142.169	0
<i>Vocabolario Fiorentino Contemporaneo</i>	0	0	0	1.064	0
<i>AFLI Authority File</i>	0	2200	0	0	0
Totale	328.439	24.823	5.734	175.233	6.040

Nota bibliografica

Nel testo si è fatto riferimento ad alcune concordanze uscite a stampa: *Concordanze degli Inni Sacri di A. Manzoni*, Firenze, Accademia della Crusca, 1967 (alle pagine 7-15 è contenuta la *Prefazione* di Aldo Duro, citata per le pp. 7-8); *Novella del grasso legnaiuolo nella redazione del codice Palatino 200*, con testo, frequenze e concordanze, Firenze, Accademia della Crusca, 1967; *Concordanze del «Decameron» di Giovanni Boccaccio*, a cura di Alfredo Barbina, sotto la direzione di Umberto Bosco, Firenze, Giunti-Barbèra, 1969; *Concordanze del Canzoniere di Francesco Petrarca*, Firenze, Accademia della Crusca-Opera del Vocabolario, 1971.

Altri contributi e volumi citati sono: Giovanni Nencioni, *Relazione dell'Accademia della Crusca sul Vocabolario della lingua italiana*, in «Studi di Filologia Italiana», XIII, 1955, pp. 395-420; *Gli atti del primo Vocabolario* a cura di Severina Parodi, Firenze, Sansoni, 1974 (ristampa con aggiunta di indici: Firenze, Accademia della Crusca, 1993); *Vocabolario degli Accademici della Crusca*, Firenze, Le Lettere, 1974 (con un'Avvertenza scritta da Giovanni Nencioni); Raffaella Setti, *Le parole del mestiere. Testi di artigiani fiorentini della seconda metà del Seicento tra le carte di Leopoldo de' Medici*, Firenze, Accademia della Crusca, 2010.