

La Storia dell'Italiano nella Prospettiva della *Corpus Linguistics*

Francesco Sabatini
Accademia della Crusca

Abstract

Present day Italian is in great part the language used by Dante and other Tuscan writers at the end of the 13th century. 61% of contemporary Italian vocabulary has been used since the 14th century. If we include in our survey words used in Italian from the beginning to the 16th century, we see that the percentage of words still in use today rises and reaches 76,5%. Italy lacked a centre of political and cultural power which could establish an undisputed linguistic standard as had happened in France and Spain. Therefore the language of the great 14th century Florentines, namely Petrarca and Boccaccio, was adopted as a model. The diffusion of the printing process also contributed to the standardisation of the language.

Ever since 1590 the Accademia della Crusca had been working on a comprehensive *Dictionary* of the Italian language, based on the study of texts mostly written in 14th century Florentine, but also including several later authors, not all Tuscan. The dictionary was published in Venice in 1612 and was the first modern European lexicographical undertaking in terms of its content and methods. The works of the authors quoted formed a balanced *corpus*. Galileo and other scientists attended to the technical and scientific entries. For each meaning a large context was supplied and there were frequent links to other related words and definitions. The *Vocabolario* acted as a centre of standardisation and identity of the language in Italy for centuries.

1.

La storia della lingua italiana è il regno della continuità. Molti studiosi sanno, o lo pronono talora con qualche meraviglia, che un italiano mediamente colto di oggi può leggere con relativa facilità, e comprendere almeno in superficie, i nostri maggiori classici letterari anche di 700 e più anni fa. Un'espressione come *la bocca mi baciò tutto tremante* è un enunciato della nostra lingua quotidiana, con fonologia, morfologia, lessico e relativa semantica, ordine delle parole e reggenze sintattiche perfettamente rispondenti al nostro modo usuale di descrivere una simile vicenda personale, che il verbo al passato remoto ovviamente riferisce a un tempo alquanto trascorso. Notoriamente, si tratta di un celebre verso dantesco, nel quale si riscontra una sola piccola differenza, registrata dalle edizioni critiche più attendibili, la palatale sibilante in *basciò* (differenza che si annulla, per un concorso di circostanze, nell'odierna pronuncia italiana dell'area centrale). Trovandosi in uno dei primi canti (il V) dell'*Inferno*, questo verso, sovrapponibile all'enunciato in italiano di oggi, sta per compiere appunto i suoi 700 anni.

Non è un esempio tirato a forza, perché è ben dimostrato che la struttura linguistica – mettendo da parte aspetti stilistici specifici e difficoltà concettuali, che sono altra cosa – della

Commedia e di altri classici italiani di radice toscana prodotti dalla fine del '200 in poi è largamente coincidente con l'italiano di oggi. Per quanto riguarda il lessico, la parte più mutevole della lingua, calcoli finalmente condotti con mezzi informatici hanno mostrato che il 61% del vocabolario di alta disponibilità (circa 9000 lemmi) dell'italiano odierno resiste nell'uso fin dal secolo XIV e la percentuale sale al 76,5% se includiamo il secolo XVI.¹ Gli altri livelli del sistema linguistico si rivelano anche più stabili, e la coincidenza è maggiore se l'italiano di oggi viene messo a confronto proprio con la lingua dei trecentisti, anziché con quella degli scrittori più vicini a noi di un secolo o due. Questo andamento apparentemente bizzarro della nostra lingua non si spiega senza conoscere almeno nelle sue linee essenziali la nostra storia linguistica. Non basta, infatti, evocare le condizioni socio-politiche generali della comunità italiana, che ha tardato moltissimo a costituirsi in un organismo unitario dotato di un dinamismo complessivo, né è sufficiente riferirsi genericamente alla forte preminenza dell'uso scritto, al quale è stata affidata a lungo l'esistenza dell'italiano. Bisogna arrivare a considerare le circostanze particolari e i modi specifici in cui è stato conferito all'italiano quel carattere di stabilità che finora ha sfidato i secoli e, chissà, potrebbe raggiungere il millennio: le circostanze, cioè, nelle quali operatori dotati di grandissima autorità hanno proposto consapevolmente agli apprendenti della lingua quel carattere come un valore primario, riuscendo a imporlo. Si scopre così che più e più volte, nel corso storico della nostra lingua, l'accettabilità dell'uso linguistico, nonché il riconoscimento dell'esistenza della lingua stessa e ogni argomentazione anche teorica sui fatti linguistici sono stati derivati dalla costituzione preliminare di un concreto *corpus* di testi.

Testi scritti, naturalmente. E ciò fornisce la prova che il ricorso a quel principio era pur sempre la conseguenza (felice conseguenza?) della mancanza di un centro geografico-socio-politico dotato di potere di guida dell'uso linguistico vivo, quel potere che invece plasmò, com'è noto, il ben diverso corso del francese. Ma qual che ne fosse la causa oggettiva e profonda, è certo che in molti momenti cruciali per l'affermazione, il consolidamento e anche l'avanzamento dell'uso della nostra lingua fu risolutivo, nella coscienza e nella prassi degli attori principali di quei processi, il loro tenace appigliarsi a una tradizione di lingua letteraria consegnata in un canone di testi. Non resta, per misurare il peso di questo fenomeno, che ripercorrere velocemente le tappe di questo itinerario che dalla fine del secolo XIII ci conduce quasi all'inizio del XX, seguendo la prospettiva delle successive canonizzazioni testuali.

2.

Ben prima di dedicarsi alla *Commedia*, Dante compone il *De vulgari eloquentia* per dimostrare l'esistenza e le caratteristiche di una lingua letteraria italiana già formata e per lui

¹ I calcoli sono stati fatti sulla base delle datazioni segnate in F. Sabatini-V. Coletti, *Dizionario della lingua italiana*, Rizzoli-Larousse, 2006. In questo dizionario di ampiezza media (85074 lemmi, con esclusione delle varianti e dei derivati non posti a lemma) i vocaboli considerati di alta disponibilità sono 9059 (su un lemmario totale di 85074 unità): di essi 5513 sono documentati già nel secolo XIV e altri 1403 risultano in uso fin dal secolo XVI.

addirittura primeggiante sulle consorelle neolatine e fonda la sua dimostrazione sull'eredità lasciata dalla Scuola poetica siciliana, da lui conosciuta e fruita attraverso le sillogi costituite nei celebri Canzonieri trascritti qualche decennio prima in Toscana:² sono questi i pilastri della sua coscienza della realtà linguistica italiana e sono qui le radici anche delle sue teorie sull'uso illustre delle lingue.

Qualche decennio dopo, accresciuto enormemente il patrimonio di base della nostra lingua letteraria proprio dalla produzione di Dante e di una schiera di suoi contemporanei, si sottomettono allo stesso principio di autorità coloro che, in talune aree più attive del contesto italiano, desiderano conquistare la lingua di quella sfolgorante letteratura. Ha funzione di *corpus* paradigmatico il *Canzoniere* messo insieme dal trevisano Niccolò de' Rossi, e via via nel corso dello stesso secolo e nel successivo con varia calibratura rispondono a questo scopo le numerose sillogi di rime e anche prose toscane che circolano (insieme con le decine e decine di manoscritti del poema dantesco) per l'Italia: si utilizza naturalmente l'appiglio dei contenuti per giungere a possedere, in realtà, modelli di lingua e di stile. Episodio di prima grandezza, sotto questo profilo, è quello dell'invio, nel 1476, da Firenze a Napoli, da Lorenzo de' Medici a Federico d'Aragona, della cosiddetta *Raccolta aragonese*, una studiatissima antologia di poeti della più schietta tradizione toscana, pone in prima posizione Dante delle *Rime* ma intorno e di seguito a questo caposaldo dispone un ampio *corpus* che parte ancora dai Siciliani e prosegue col fiore della produzione trecentesca e quattrocentesca per giungere al nuovo vertice rappresentato da componimenti dello stesso Lorenzo: nell'epistola accompagnatoria, che si attribuisce al Poliziano, si esprime chiaramente il motivo dell'attesa e dell'offerta, attraverso i testi, di modelli linguistici.

Si sta propagando ormai, alla fine del Quattrocento, la produzione del libro a stampa e stanno maturando i tempi per una vera codificazione del volgare, imposta anche dal desiderio dei tipografi di raggiungere tirature destinate a più larga diffusione. È dalle loro officine, infatti, che si fa appello ai grammatici perché provvedano a tale bisogna, ed ecco che sotto l'insegna più celebre, quella di Aldo Manuzio, viene fissato un ormeaggio inamovibile: sono le edizioni modello (per l'epoca) dei due testi fondamentali del Parnaso italiano, il *Canzoniere* petrarchesco (1501) e la *Commedia* dantesca (1502), per cura di un giovane, ma agguerrito umanista ed estroso inventore di formati e segni tipografici, Pietro Bembo. A lui, veneziano, in concorrenza con un quasi conterraneo, il pordenonese Fortunio, va attribuito il merito (la colpa?) di aver incardinato definitivamente, mediante l'argomentazione implacabile delle sue *Prose della volgar lingua* (1525), l'intera struttura della lingua italiana nella realtà cristallizzata quasi di un solo testo, il *Canzoniere* petrarchesco, con forte penalizzazione di Dante, e comunque di aver circoscritto l'area di prelievo dei modelli al fiorentino letterario trecentesco, a scapito di quello contemporaneo. Uno sciame di suoi seguaci continua e consolida ra-

² Tre ne sono giunti fino a noi e sono i notissimi codici Vat. lat. 3793, Redi 9 della Biblioteca Laurenziana e B.R. 217 (già Palatino 418) della Biblioteca Nazionale Centrale di Firenze. Si possono ora consultare nella splendida riproduzione anastatica curata da Lino Leonardi, *I Canzonieri della lirica italiana delle Origini*, Sismel – Edizioni del Galluzzo, Tavarnuzze, Impruneta (Firenze), 2000, in tre voll., più un quarto di *Studi critici*.

pidamente la sua opera: già una “giuntina” di rime antiche, del 1518, forse ispirata proprio da lui, mette in circolazione un serrato *corpus* di testi canonici. Nei decenni successivi entrano in azione lessicografi e grammatici che estraggono a gara lessico e regole rigorosamente dalle “tre corone” (o “tre fontane”). Infatti, se esistono, come sappiamo bene, sostenitori di altri indirizzi, è la proposta del Bembo che, nell’assetto politico e socio-culturale generale d’Italia, ottiene una larghissima adesione: e così il fiorentino trecentesco, riesumato nei testi consacrati, risorge come *lingua italiana* per essere messo in cammino verso le età future, ricoperto però – fatto altrettanto decisivo – di una sontuosa veste stilistica rinascimentale. È il momento in cui la saldezza del canone di riferimento propria della lingua italiana genera invidia tra i dotti francesi.³

Il progetto di rifondazione dell’italiano (si potrebbe dire di sottofondazione, pensando alle opere che ridanno vigore alle basi di un edificio preesistente) è già ben delineato, certamente, nei decenni centrali del Cinquecento. Ma perché l’edificio sorga ampio e robusto occorrono allestimenti e materiali abbondanti, macchine edilizie adeguate: è questa l’immagine con cui si può raffigurare l’opera messa in cantiere nell’ultimo decennio del secolo dagli Accademici della Crusca, un ridotto manipolo di dotti e filologi fiorentini, capitanati per pochi anni da Lionardo Salviati, che riconducono l’iniziativa nella patria della nostra lingua. Salviati è il vero mediatore della tesi bembiana nella città che si sente ancora procreatrice di lingua ed è un convinto assertore della funzione stabilizzatrice di un canone autoriale ben evidente, nel quale egli, allargando i vincoli posti dal Bembo, integra decisamente Dante e accoglie autori anche di epoche posteriori, giungendo a includere il non toscano Ariosto, ma rifiutando, com’è noto, Tasso. I suoi compagni di lavoro, che si sono dedicati a un’impresa quasi temeraria, sono però più flessibili, ampliano la rosa degli autori moderni (ma non accettano ancora Tasso!), si preoccupano di immettere, sia pure in misura limitata, anche testi pratici e tecnici e perfino solo manoscritti. È evidentissima, e dichiarata nella presentazione dell’opera, il celebre *Vocabolario degli Accademici della Crusca* (questo il titolo scelto per evitare termini come *fiorentino*, *toscano*, *italiano*, giudicati variamente compromettenti) che appare nel 1612, l’intenzione di pervenire alla formazione di un vero *corpus* universale, capace di rappresentare tutta la lingua. Basta leggere questo brano della introduzione *A’ lettori*:

«Non è stata nostra intenzione di fare scelta di vocaboli dispersé [‘di per sé’, ‘singoli, particolari’], ma di raccorre, e dichiarare universalmente, le voci e maniere di questa lingua: però [‘perciò’] non abbiamo sfuggito di metterci le parole, e modi bassi e plebei, giudicandogli noi necessari alla perfezione di essa, per comodità di chiunque volesse usargli nelle scritture, che gli comportano».

Chi non credesse alle promesse dei compilatori può, *more adolescentium*, andare a cercare tra i lemmi del *Vocabolario* anche le ‘parolacce’ canoniche e le troverà, scientificamente trattate e, specie nelle edizioni successive, filologicamente documentate. Non occorre un lun-

³ Notizie in proposito nella *Storia della lingua italiana* di B. Migliorini, p. 304 (cito dalla decima edizione, a c. di G. Ghinassi, Sansoni, Firenze, 1991).

go commento, insomma, per capire che nell'ambiente fiorentino, agitato dalla provocazione bembiana e pur sempre depositario di un senso vivo della lingua, l'assidua riflessione di almeno due generazioni di dotti aveva fatto compiere un lungo passo avanti nella rappresentazione e quindi nella pianificazione della nostra lingua. L'idea di un canone ristretto di autori modello da imitare per produrre altri testi modello si era trasformata nel concetto di un vero *corpus* integrato e bilanciato, dal quale si cercava di estrarre tutta la lingua per dare agli utenti la possibilità di usarla in tutte le maniere e per tutte le occorrenze possibili. Si rifletta almeno sulla presenza, nel brano citato, dei termini *raccorre* (che si oppone a *fare scelta*), *universalmente*, *voci e maniere*, *modi* (e sono spesso 'modi di dire' tratti dal parlato contemporaneo), *perfezione* cioè 'completezza', e non si trascuri quel *chiunque* con cui si allude a un pubblico vasto e indifferenziato. Se poi si bada al fatto che è precisa intenzione dei compilatori del *Vocabolario* fornire con dovuta ampiezza i contesti delle voci, osservabili così anche nelle connessioni sintattiche, e che tutto il materiale lessicale viene organizzato in un'ampia rete attraverso rinvii, ci si renderà conto che già i primi ideatori dell'opera avevano chiara l'idea del *sistema* globalmente inteso – strutturale, funzionale, socio-stilistico, storico-culturale – della lingua.

Non si trascuri, d'altra parte, il fatto che gli Accademici avevano fatto professione di modernità anche attraverso i loro simboli: come emblema principale dell'istituzione avevano scelto il *frullone*, una macchina di recente invenzione che sostituiva l'antico setaccio a mano e che permetteva di ottenere farina più raffinata e di produrne, con il lavoro di un solo operaio, una quantità maggiore. Era metafora del loro lavoro di vaglio filologico dei testi e di analisi più approfondita della lingua.

3.

Tra le intenzioni e la prassi corrono certamente delle distanze. Dai testi "citati" non viene estratto tutto il lessico, gli autori "moderni" sono, nelle prime due edizioni, ancora limitati (Tasso verrà accolto solo nella terza edizione, la più innovativa), il settore meno rappresentato resta quello della terminologia tecnica e scientifica. Ma i conti che oggi possiamo fare con notevole precisione grazie alla versione elettronica dell'intera serie delle edizioni del *Vocabolario*⁴ permettono di pronunciare giudizi più equilibrati.

⁴ Cfr. *Lessicografia della Crusca in rete*, banca dati consultabile nel sito dell'Accademia della Crusca: www.accademidellacrusca.it/biblioteca_virtuale.shtml. Sull'intero progetto informa, in questi Atti, la relazione dei suoi coordinatori e responsabili, M. Fanfani e M. Biffi. Le mie considerazioni sono limitate sostanzialmente alla prima edizione del *Vocabolario*. Questa viene ora riproposta anche in carta: *Vocabolario degli Accademici della Crusca, 1612*, ristampa anastatica, accompagnata da un volume di presentazione, *Una lingua, una civiltà*, il *Vocabolario*, e dalla versione elettronica in CD-rom, a cura di Domenico De Martino e Simone Casini, ERA Edizioni, Carbonate – Como, 2006. Un filo di ragionamento complessivo sulla tradizione lessicografica italiana, con esame più attento all'opera degli Accademici, è nel saggio di Pietro G. Beltrami e Simone Fornara, *Italian Historical Dictionaries: from the Accademia della Crusca to the Web*, in "International Journal of Lexicography", vol. 17, num. 4, dicembre 2004, *Special Issue. Lexicography in Italy: specific themes and trends*, a cura di Carla Marengo, pp. 357-384.

Nella sua prima edizione (1612) il *Vocabolario* constava di 1092 pagine *in folio*, presentava 25.056 lemmi di entrata, articolati in varie accezioni, e raccoglieva 52.862 citazioni (che contengono 1.152.999 occorrenze) ricavate da 208 autori per un totale di 309 opere (27 sono gli autori “moderni”, con 52 opere). La presenza di lingua non “d’autore”, e quindi suggerita dal parlato, è molto ridotta, ma va cercata in alcuni sviluppi che talora si presentano in coda alle voci e soprattutto nella lingua dei compilatori: un’operazione non certo inconsapevole, paragonabile, in lessicografia, a quella che nei trattatisti di lingua (come Bembo) è stata qualificata come «grammatica silenziosa» (Patota). Ebbene, si tratta di una “quantità” di lingua mai prima di allora estratta da qualsiasi sorgente, scritta o parlata, commentata e messa a disposizione in bell’ordine per i parlanti di una lingua moderna. Le edizioni successive dell’opera (II, 1623, ancora in un volume; III, 1691, in tre volumi, arricchiti da molto lessico scientifico e tecnico; IV, 1729-1738, in sei volumi; lasciando da parte la V, 1863-1923, in 11 volumi ma incompiuta) portarono alla fine a triplicare tale quantità. Dimensioni e impostazione dell’opera concorrono a spiegarne il successo non solo in Italia ma nell’intera Europa, che risuonò a lungo di lodi all’istituzione fiorentina, espresse da personaggi, per citarne solo alcuni più “vicini” a noi nel tempo, come Voltaire, Samuel Johnson, i fratelli Grimm, Arthur Schopenhauer.

Com’è noto, fin dal suo apparire il *Vocabolario* fu sottoposto, in Italia, anche a vivaci critiche, alcune certo dettate da malumori o differenti gusti personali, altre nascenti da rilievi fin troppo facili di mancanze e squilibri. Ciò non ostante, per almeno due secoli non vi fu scrittore (fino a Manzoni, quando intraprese la stesura del suo romanzo) che non lo consultasse assiduamente. Facendo un bilancio tra meriti e difetti dell’opera i primi risultano sicuramente superiori ai secondi. Pensando soprattutto all’epoca più tarda della circolazione dell’opera, bisogna anche concludere che non poteva un vocabolario, comunque concepito, sopperire all’assenza di un centro unificatore della vita culturale del Paese e promotore di un vivo e largo flusso di nuova lingua all’interno della società. Perché questo avvenisse occorreva che cambiasse «le condizioni d’Italia», come ebbe occasione di dire Ugo Foscolo nel 1826.

Dall’età illuministica in poi la fede nel repertorio di modelli, e quindi nel canone di autori, era stata scossa, ma «le condizioni d’Italia», per riprendere subito l’espressione foscoliana, non favorivano la formazione di un “uso” linguistico generale ben conoscibile che potesse sostituire ampiamente il principio di autorità. Nel campo della lessicografia, il più sensibile a questo problema, i primi tentativi di attingere più liberamente agli autori recenti, ai settori tecnici e scientifici e al parlato (depurato di forma dialettale) furono quelli del nizzardo Francesco D’Alberti di Villanova (1797-1805) e dell’editore napoletano Tramater (1829-1840). Due riedizioni private del *Vocabolario della Crusca* (quella curata dal veronese Antonio Cesari, 1806-1811, e quella del cesenate Vincenzo Manuzzi, 1833-1840) riproponevano invece strettamente il principio del canone di autori e quella del Cesari in particolare ridusse fortemente questo canone al secolo XIV, escludendo comunque il linguaggio tecnico e scientifico. Negli ultimi quarant’anni del secolo si giocò la partita decisiva della lessicografia dell’Italia unita. Due opere cercarono di conciliare il principio del canone d’autori con il libero prelievo dall’ “uso”, un uso di cui si poteva cominciare a captare l’esistenza. Nel 1863 vide la luce il primo volume della V edizione del *Vocabolario della Crusca* (che fu fermato all’XI

volume, con la lettera *O*, nel 1923, con decreto del ministro fascista Gentile), totalmente basato su un *corpus* di autori, anche se esteso via via ai moderni e contemporanei (solitamente presi in considerazione dopo la morte) e con aperture alla lingua d'uso e tecnico-scientifica anche senza sostegno di citazioni. Un'impostazione non dissimile, ma con maggior propensione per gli autori moderni, un'accoglienza più larga delle voci liberamente captate dall'uso e una sensibilità più aggiornata per le voci tecnico-scientifiche, dette Nicolò Tommaseo al suo *Dizionario della lingua italiana* (1861-1879), rimasto per oltre un secolo il caposaldo della lessicografia italiana. Una terza opera tentò l'innovazione radicale: il *Novo vocabolario della lingua italiana secondo l'uso di Firenze* (1870-1897), mirava programmaticamente a cancellare la tradizione del canone degli autori, ispirandosi alla tesi manzoniana della necessità di una totale rifondazione dell'italiano nel parlato dei Fiorentini. La scarsissima accoglienza ottenuta dall'opera dimostrò l'inaccettabilità storico-culturale e l'inattuabilità pratica di un simile proposito.

Se si volge rapidamente lo sguardo a un campo vicino a quello della lessicografia, quello dell'educazione allo scrivere, si può cogliere un interessante parallelismo. Durante tutto l'Ottocento ebbero ancora molta fortuna le raccolte di autori esemplari per l'uso della lingua. Letterati di fama vennero incaricati, dalle case editrici, di allestire simili raccolte e tra questi incontriamo, negli stessi anni, il purista napoletano Basilio Puoti (*Antologia di prose italiane*, Napoli, 1828) e Giacomo Leopardi (*Crestomazia italiana*, I, *La prosa*, II, *La poesia*, Milano, 1827 e 1828). Ma il caso più significativo è forse quello degli *Esempi del bello scrivere in prosa e in poesia proposti agli studenti di retorica*, raccolti in due volumi dal toscano Luigi Fornaciari: pubblicati la prima volta nel 1829 e 1830, ripubblicati nel 1838, furono ancora riediti, anche se accresciuti e aggiornati, a cura del figlio Raffaello (autore di una ben "moderata" tra le nostre grammatiche di quel secolo, fronteggiata dai manuali dei manzoniani) fino all'inizio del Novecento (1907).

Il grande traguardo dell'unificazione politica d'Italia era stato raggiunto, ma solo l'avanzare dei fondamentali processi conseguenti, quali l'innalzamento culturale dell'intera popolazione, un avvicinamento delle classi sociali e uno stemperarsi delle concentrazioni regionali, avrebbe avviato la formazione di un "uso comune" riconoscibile e captabile dai lessicografi, e praticabile, si aggiunga, dai nuovi scrittori (e saggisti e giornalisti): un uso distinguibile, ma non separabile dal fiume di lingua scritta che attraversando i secoli aveva permesso che esistesse, giungesse fino a noi e perdurasse la civiltà italiana.