

INTRODUZIONE*

Tradizionalmente, la *Piazza delle Lingue* accoppia il tema principe della lingua italiana al tema della varietà delle lingue viventi: dopo la prima edizione del 2007, dedicata a *Le Lingue d'Europa patrimonio comune dei cittadini europei*, quando il piazzale antistante la Villa medicea di Castello, la sede dell'Accademia della Crusca, è stato denominato "Piazza delle lingue d'Europa", sono seguite le *Piazze* dedicate al *Multilinguismo in Europa* (2008), a *Esperienze di multilinguismo in atto* (nel 2009, con ospite d'onore la Confederazione Svizzera, in quanto esempio di nazione con una realtà multilingue, in grado di esportare maturi modelli di coesistenza), a *L'italiano degli altri* (2010), a *L'italiano in Europa nel XXI secolo* (2011), quest'ultima in onore di Gianfranco Folena, grande studioso dell'italiano fuori d'Italia nel Settecento. In questa prima fase, come si vede, l'Europa è ritornata più volte, ed è stata posta in vario modo al centro dell'attenzione, in anni in cui era vivo il sentimento di costruzione dell'Europa medesima come casa comune anche linguistica, nella speranza di farne una realtà vantaggiosamente plurilingue e non il terreno uniforme del monolinguisma anglofono.

Nel 2012 ricorreva il IV centenario del *Vocabolario* della Crusca, e dunque quella volta la *Piazza* cambiò strada: fu dedicata appunto all'*Italiano dei vocabolari*, cioè al tema della lessicografia, che per la Crusca resta un impegno fondamentale. L'edizione del 2013 tornò al multilinguismo, con una novità, però, cioè la lettura dei problemi in chiave giuridica e normativa: ecco *Lingue e Diritti*. Giungiamo così all'edizione 2014, di cui qui si presentano gli Atti. Come nel 2012, nel 2014 è stato trattato il tema della lessicografia, ma in una forma nuova e assai moderna, legata all'informatica e al destino futuro dei vocabolari e dei corpora nell'era di Internet, in una stagione che sembra segnare inesorabilmente la crisi del vocabolario a stampa come lo abbiamo conosciuto per secoli.

* Il progetto della presente *Introduzione* è comune ai due autori; la stesura si deve per le pp. 5 e 9-11 a Claudio Marazzini, per le restanti a Ludovica Maconi.

Il titolo della *Piazza delle Lingue* 2014 è stato *L'italiano elettronico. Vocabolari, corpora, archivi testuali e sonori*. Con *italiano elettronico* non deve in questo caso intendersi l'italiano impiegato nella scrittura telematica, oggi usuale nella comunicazione rapida attraverso i nuovi *media* (per questo italiano digitale, si rimanda ai recenti lavori di Elena Pistolesi e Massimo Prada¹). Come circoscrive e aiuta a intendere il sottotitolo, l'italiano elettronico oggetto di questo volume è quello raccolto in vocabolari, già realizzati o ancora da farsi (elettronici e *online*), l'italiano raccolto in banche dati e in corpora testuali: risorse tutte fruibili in Rete e utilmente applicabili a fini linguistici e lessicografici.

Gli Atti qui pubblicati, in cui sono purtroppo venuti a mancare per ragioni contingenti tre contributi che hanno destato interesse e ammirazione nelle esposizioni orali, si aprono con le relazioni di due maestri, Raffaele Simone e Luca Serianni, entrambi attenti alla "lessicografia del futuro": Simone illustra le caratteristiche del dizionario del nuovo millennio; Serianni si interroga sul futuro del dizionario dell'uso. La lessicografia di domani appare inevitabilmente legata all'informatica, come era prevedibile. Simone propone il superamento del modello su carta a lista, fatto di voci isolate, e immagina un modello a "rete" che permetta la navigazione in tutte le direzioni, e che tessa in una "rete di reti" il vocabolario monolingue, quello dei sinonimi e l'analogico. Non a caso Raffaele Simone è autore di un magnifico dizionario analogico dell'italiano, un libro che fa invidia a tutta l'Europa per la sua qualità e ricchezza, e che purtroppo le non fortunate vicende della casa editrice hanno reso meno popolare anche tra gli specialisti. La lezione di questo libro si riverbera ora nelle considerazioni di Raffaele Simone, che ci sollecita a considerare il vocabolario come un luogo di incontri e intrecci tra parole. I lessicografi del futuro dovranno, anche più che in passato, riservare attenzione alle strutture, alle combinazioni, ai costrutti, alle relazioni semantiche di varia natura; dovrà essere rinnovato e ripensato il metalinguaggio, affinché sia più vicino al linguaggio naturale degli utenti, dovranno essere rivisti i formati delle entrate.

Il passaggio necessario alla digitalizzazione (alla versione elettronica) dei maggiori dizionari dell'uso è stato solo il primo passo nel campo della lessicografia dell'era nuova, passo che comunque si è rivelato insufficiente a frenare il declino del dizionario generalista, fino a qualche tempo fa presente

¹ Elena Pistolesi, *Scritture digitali*, in G. Antonelli - M. Motolese - L. Tomasin (a cura di), *Storia dell'italiano scritto*, III. *Italiano dell'uso*, Roma, Carocci, 2014, pp. 349-75; Massimo Prada, *L'italiano in rete. Usi e generi della comunicazione mediata tecnicamente*, Milano, Franco Angeli, 2015.

in tutte le case italiane, ma oggi in seria difficoltà a resistere nel panorama editoriale e persino nell'uso scolastico, dove talvolta è relegato a strumento d'appoggio per la stesura del tema. I vantaggi del sistema informatico e multimediale, che pure provoca una crisi della produzione tradizionale dei vocabolari, restano indubbi. Serianni osserva come, in alcuni casi, animazioni, riproduzioni e filmati possano assolvere meglio al fine della definizione e comprensione del lemma rispetto a quanto faccia un comune vocabolario dell'uso monovolume.

Non è detto però che tutto il futuro della lessicografia sia racchiuso nella dimensione elettronica. Ancora Luca Serianni ritiene infatti che il dizionario cartaceo dell'uso possa sopravvivere se saprà trasformarsi in libro di lettura e dizionario d'autore. Si tratterebbe in tal caso di operare scelte coraggiose: sfoltire il lemmario e scrivere voci a mo' di brevi saggi sulla «storia della parola e del suo uso nella letteratura del passato», dando spazio anche all'uso degli «scrittori esperti, dai grandi giornalisti, agli autori letterari della seconda metà del secolo scorso» (così suggerisce lo studioso); storia linguistica e culturale, dunque, con indicazioni grammaticali e ortografiche (norma e uso d'oggi e del passato), implicazioni semantiche e collocazioni. Serianni propone qui tre voci campione, appositamente compilate: *cane*, *fare*, *flebile*. Il modello dichiarato per lo stile discorsivo di questo nuovo vocabolario è il *DIR - Dizionario italiano ragionato* diretto da Angelo Gianni e Luciano Satta, pubblicato a Firenze dall'editore D'Anna nel 1988. Il futuro del dizionario cartaceo, quindi, potrebbe inaspettatamente richiamare un modello del passato, seppure non troppo lontano, ma potrebbe ancor più inaspettatamente trovare riscontro e ispirazione in almeno due modelli del passato remoto, se pensiamo alla fortuna che ebbero un tempo i dizionari dei sinonimi, largamente usati nella scuola italiana dell'Ottocento: si faccia riferimento, in particolare, al *Saggio intorno ai sinonimi* di Giuseppe Grassi e ai celebri *Sinonimi* di Nicolò Tommaseo. Com'è noto, anche queste opere lessicografiche di mole ridotta (ridottissima quella del Grassi) nacquero come libri disponibili alla lettura continuata: non proponevano semplici elenchi di parole accostate per sinonimia, ma ogni voce spiegava in modo discorsivo le sottili differenze tra un sinonimo e l'altro, per chiarire e giustificare l'uso più corretto; Grassi inserì anche informazioni etimologiche, ritenute utili per la storia della parola.

Alla storia della parola si lega la data di prima attestazione dei lemmi, oggi presente in tutti i principali dizionari dell'uso monovolume. Tale data, però, come osserva Serianni, è «insufficiente e addirittura fuorviante in presenza di accezioni molto distanti semanticamente e storicamente le une dalle altre», quindi da eliminare come impropria in uno strumento destinato al largo pubblico non specialistico. A metà del secolo scorso, Bruno Migliorini,

dopo essersi dedicato all'attività lessicografica come revisore del Cappuccini e del Panzini, lamentava il fatto che nessuno dei nostri vocabolari fornisse informazioni sulla datazione delle parole, e spiegava questa mancanza con la maggior importanza attribuita in Italia al compito "normativo" del vocabolario rispetto a quello "documentario" che pure gli spetta². Allora, però, ancora mancavano i dizionari etimologici, che hanno in parte presto colmato questa lacuna. Negli anni Novanta (solo poco più di vent'anni fa) la data di prima attestazione è passata anche ai dizionari dell'uso, andando così ad ampliare la documentazione per alcuni lemmi, non registrati nei più ridotti repertori etimologici, ma creando anche le ambiguità e grossolanità a cui Serianni fa giustamente riferimento. Si ricordi, però, che quando per primo il vocabolario dell'uso Palazzi-Folena introdusse la datazione, presentandola come una delle innovazioni più significative, lo fece dichiarando l'accezione a cui la data andava riferita (se non coincidente con la prima accezione nell'ordine), evitando in questo modo errori e fraintendimenti dell'utente.

La seconda parte degli Atti che qui presentiamo raccoglie, sotto il titolo *L'italiano negli archivi della Rete*, i contributi su risorse disponibili in Internet e utili per gli studi linguistici italiani. Daniele Fusi presenta l'edizione digitale del Tommaseo-Bellini, realizzata per il sito dell'Accademia della Crusca e ora disponibile *online* in forma moderna, "adattiva" o "responsiva". Ludovica Maconi illustra la propria individuazione di oltre 3000 retrodatazioni lessicali, reperite utilizzando la banca dati di Google Libri. Il suo contributo vuole spiegare le potenzialità e anche i difetti di Google applicato alla ricerca di retrodatazioni, suggerendo come ovviare agli errori del sistema. Illustra inoltre l'utilità di un programma informatico realizzato da Daniele Fusi per estrarre automaticamente retrodatazioni da testi in formato RTF. Questo strumento si è rivelato utile in particolare per l'esame dei testi antichi, perché i dati così ricavati sono stati raramente migliorati nel successivo controllo in Google Libri, mentre nel caso di testi ottocenteschi le retrodatazioni estratte in modo automatico dal programma sono state poi nuovamente retrodate con Google nella maggioranza dei casi.

La Rete è luogo di apparenti certezze, spesso di inganni, e la tecnologia ci dà non di rado una sensazione di durata e di robustezza dei risultati, e anche questa sensazione può essere illusoria. Pier Marco Bertinetto e Silvia Calamai ci raccontano nel loro intervento i rischi di obsolescenza e di

² Bruno Migliorini, *Che cos'è un vocabolario?*, Firenze, Le Monnier, 1951, p. 61 (I ed. 1946).

perdita corsi da un'importante raccolta di materiali linguistici di interesse dialettologico. I due autori si soffermano sul legame tra supporto materiale e documento, descrivono gli archivi orali del progetto *Gra.fo* e affrontano il tema dell'identificazione dell'unità documentale "di base" in un archivio orale digitale. Bisogna ricordare che i dati informatici necessitano sempre di aggiornamenti per rimanere adeguati agli standard in uso. Questo costringe spesso a ritornare sul già fatto, a spendere energie per riconvertire e ricodificare, come è accaduto in passato per molti testi digitalizzati con il sistema DBT. Dobbiamo quindi essere attenti all'evoluzione tecnologica e cauti verso sistemi proprietari che non siano in grado di garantire la durata temporale. Il segreto dell'informatica, quanto a stabilità dei dati, sta nella moltiplicazione e differenziazione dei supporti di memoria, e, quanto a durata, sta nella prontezza con cui si procede alla riconversione, quando essa diventa necessaria. Carla Marello chiude la parte sugli archivi dell'italiano in Rete presentandoci l'evoluzione dell'interfaccia di interrogazione di VALICO, un corpus di testi di apprendenti italiano lingua seconda. Numerosi sono ormai gli archivi di testi in italiano, di tutti i generi e le epoche, disponibili in Internet in libera consultazione, ma considerando che talvolta restano poco conosciuti e solo parzialmente sfruttati dagli studiosi, non sarebbe forse inutile un sito che raccogliesse informazioni e indirizzi di queste preziose risorse.

Ospite d'onore della *Piazza 2014*, un convegno proiettato verso il futuro e la modernità, è stata una lingua antica morta, il latino. Nel dialogo sui corpora intessuto con i colleghi classicisti, abbiamo scelto un gruppo di ricerca specializzato nel tardoantico, il periodo storico più vicino alla formazione della lingua italiana, per presentare un progetto in corso da qualche anno presso l'Università del Piemonte Orientale e già ampiamente avviato, ideato e diretto da Raffaella Tabacco: la biblioteca digitale *digilibLT* (*Digital Library of late-antique Latin texts*), che mette in libera consultazione in Rete un corpus di testi in prosa della tarda latinità (II d.C.-VII d.C.), permettendo sia l'interrogazione generale sia lo scaricamento dei singoli testi in vari formati. Ovviamente la lingua latina, conclusa la sua stagione produttiva, può ambire con ottimi risultati a un completo trasferimento delle proprie intere risorse nella Rete, cosa che per una lingua viva risulta francamente impossibile.

La *Piazza delle Lingue 2014* è stata anche l'occasione per un primo grande raduno dei gruppi di ricerca aderenti al progetto del "Corpus di riferimento per un Nuovo Vocabolario dell'Italiano moderno e contemporaneo. Fonti documentarie, retrodatazioni, innovazioni", a cui è stato assegnato

un finanziamento consistente nei PRIN 2012 (distribuiti nel 2014). Il convegno si è chiuso con una tavola rotonda collocata appositamente alla fine dei lavori, allo scopo di raccogliere le esperienze diverse presentate durante la *Piazza* da parte di lessicografi, linguisti, informatici, studiosi del mondo moderno e del mondo classico. Il gruppo è costituito da otto unità, che qui hanno avuto spazio per dare conto, per la prima volta, delle proprie ricerche, dopo pochi mesi dalla partenza ufficiale del progetto nel marzo 2014.

Si tratta delle seguenti unità, qui elencate in ordine geografico: 1. unità dell'Università del Piemonte Orientale (capofila Claudio Marazzini, anche P.I.), che si occupa del linguaggio scientifico e divulgativo nei settori specifici dell'astronomia, della geologia e della letteratura di montagna, oltre che delle retrodatazioni e delle datazioni di accezioni (tema affidato alla dott.ssa Ludovica Maconi); 2. unità dell'Università di Milano (capofila Ilaria Bonomi), che si occupa del linguaggio di giornali e fumetti; 3. unità dell'Università di Genova (capofila Lorenzo Coveri), che costituisce un corpus capace di integrare le zone lasciate scoperte dalle banche dati attuali relativamente alla lingua poetica del secondo Novecento, alle canzoni e ai libretti d'opera; 4. unità dell'Università di Firenze (capofila Marco Biffi), che sta fornendo il supporto linguistico-informatico per le diverse fasi del progetto e provvede all'aggregazione delle varie banche dati fin qui realizzate in un'unica piattaforma, compito importantissimo per evitare che il lavoro già fatto si frantumi e disperda in mille rivoli; inoltre il gruppo fiorentino predisporrà un sottocorpus sull'italiano delle arti dall'Ottocento a oggi e un altro sottocorpus relativo ai libri di cucina; 5. unità dell'ITTIG (Istituto di Teoria e Tecniche dell'Informazione Giuridica) di Firenze, istituto del CNR specializzato nel linguaggio giuridico; 6. unità dell'Università di Viterbo (capofila Riccardo Gualdo), che sta costruendo un corpus dell'italiano istituzionale e parlamentare in età postunitaria, con un'appendice preziosa dedicata al cinema industriale; 7. unità dell'"Orientale" di Napoli (capofila Rita Librandi), che sta allestendo un corpus dell'italiano politico attingendo a testi di area cattolica (autori come Romolo Murri, Luigi Sturzo, Alcide De Gasperi); 8. unità dell'Università di Catania (capofila Gabriella Alfieri), che sta allestendo un corpus di paraletteratura e divulgazione, di galatei, di libri per ragazzi.

La tavola rotonda permette dunque di gettare un primo sguardo su tutte queste ricerche, le quali dovranno concludersi nella costruzione di un grande "corpus dei corpora" strettamente legato alle attività della Crusca, destinato alla sperimentazione di metodi e forme atte a costruire il nuovo edificio di un grande vocabolario dell'italiano moderno e contemporaneo, o grande vocabolario dell'italiano postunitario che dir si voglia (il titolo è ancora oscillante nei vari contributi, non essendo ancora formalizzato in

maniera definitiva), un'opera la cui realizzazione emerge in maniera sempre più chiara tra gli obiettivi strategici dell'Accademia della Crusca: e un ritorno della Crusca alla lessicografia, lasciatecelo dire, sarebbe evento di grande significato storico. Ci auguriamo che questo evento possa davvero realizzarsi, che i tempi di attesa siano brevi, non troppo lunghi i tempi di esecuzione. In ogni modo gli Atti che qui presentiamo sono segno che il percorso è iniziato, e ci auguriamo sia fruttuoso.

CLAUDIO MARAZZINI
LUDOVICA MACONI