

**Corino, Elisa; Marelo, Carla (2017).  
*Italiano di stranieri. I corpora VALICO e VINCA.*  
Perugia: Guerra, pp. 284**

Paola Capponi  
(Universidad Pablo de Olavide, Sevilla, España)

Da anni il gruppo torinese guidato da Carla Marelo lavora alla costruzione e all'accessibilità in rete di un corpus di testi scritti, prodotti da apprendenti di italiano lingua straniera. Si chiama VALICO (Varietà di Apprendimento della Lingua Italiana Corpus Online) ed è il primo corpus di apprendenti di italiano lingua straniera liberamente disponibile in rete (<http://www.valico.org>). Raccoglie testi prodotti da parte di studenti non italofofoni a partire da cinque stimoli iconici, ossia da cinque storie disegnate. A VALICO si affianca un altro corpus, VINCA (Varietà di Italiano di Nativi Corpus Appaiato), che riunisce testi stesi da italofofoni a partire dagli stessi stimoli iconici. È stato creato «a immagine di VALICO», cioè «annotato nello stesso modo e con una finestra di interrogazione simile» (p. 29). VALICO e VINCA sono strumenti estremamente duttili che si possono piegare alle esigenze di ricerca, insegnamento, apprendimento e formazione di quanti siano interessati alla lingua italiana come lingua straniera. Si possono esplorare attraverso chiare finestre d'interrogazione per campi, incrociando ricerche per parti del discorso (PoS) e basi di dati sociolinguistici oppure attraverso un'interrogazione georeferenziata che dà accesso a carte da cui a colpo d'occhio si può cogliere la composizione del corpus. La trascrizione dei testi, estremamente accurata, conserva traccia di autocorrezioni, inserzioni, variazioni e permette quindi di ricostruire il processo di stesura del testo. All'utente è poi data la possibilità di ritagliare sottocorpora omogenei in base alle proprie esigenze di ricerca.

In questo volume Elisa Corino e Carla Marelo, con la collaborazione, in due capitoli, di Simona Colombo, illustrano in modo chiaro e articolato come si sia arrivati a costruire tale flessibilità di ricerca e come se ne possa fare uso. Il libro riunisce quella molteplicità di letture che i corpora stessi stimolano e auspicano. La riflessione teorica sulle finalità e sui principi di costruzione dei due corpora (inserita nell'ambito più ampio dell'allestimento di corpora in genere, di apprendenti in specie, di italiano lingua straniera nello specifico) si combina con indicazioni puntuali in merito alle possibilità d'uso in classe di VALICO e VINCA e con l'analisi e l'esemplifi-

cazione dell'ampio spettro di ricerche realizzabili combinando i numerosi parametri dell'occhiuta maschera di interrogazione.

Nella prima parte del volume il lettore è accompagnato in un percorso che lo conduce a riflettere sulle scelte metodologiche alla base dell'odierna architettura di VALICO e VINCA. Si osserva come tali scelte si siano affinate nel tempo, con l'avanzare dei lavori, alla luce degli stimoli ricevuti da un gruppo di lavoro allargato nello spazio (si osservino le mappe dei punti di raccolta nella ricerca georeferenziata) e diversificato (docenti di lingua italiana all'estero, studenti, tirocinanti, tesisti, ricercatori). L'esigenza di garantire un corpus di testi omogeneo e comparabile ha indotto a preferire testi redatti in classe in un arco di tempo limitato e prodotti a partire da uno stesso input, anziché produzioni libere assegnate dai docenti e svolte in autonomia dagli studenti. Questioni operative volte a garantire la rappresentatività del corpus e ragioni di interesse scientifico hanno spinto a scegliere testi narrativi brevi anziché lunghi (*essays*) e a preferire un input grafico, disegnato, a un video: più facile da somministrare anche in contesti in cui la riproduzione audiovisiva potesse essere un ostacolo e stimolante per la ricerca sul 'non disegnato', ossia su ciò che l'immagine può lasciar supporre, immaginare (in sé e nel passaggio da una vignetta all'altra, con l'inserzione di informazioni di raccordo, «unità informative che fanno da cornice alla storia e introducono possibili 'mondi immaginari' diversi da quello suggerito dalle vignette», p. 123). La scelta dei dati sociolinguistici da rilevare è stata anch'essa oggetto di dibattito e revisione alla luce della diversità culturale e sociale dei punti di raccolta. Dal primo corpus nato nel 2003 (che passerà a chiamarsi GranVALICO), servito come *training corpus* su cui si è allenato *TreeTagger* per l'annotazione, nascerà così VALICO, nome «dal 2008 riservato alla porzione di corpus bilanciato, contenente solo testi scritti a partire dai cinque stimoli iconici disegnati da Leonardo Borazio» (p. 16). Il sito ha poi iniziato ad accogliere anche discussioni orali sugli esercizi registrate in classi di italiano L2 e le «Voci di VINCA», registrazioni di italofofoni che raccontano le storie disegnate.

Nel caso specifico dei testi riuniti in VINCA centrale è stata la raccolta per varietà diatopiche: «la linguistica dei corpora italiana è costituzionalmente una linguistica della varietà, perché eredita la consapevolezza degli studi dialettologici, sociolinguistici e acquisizionali che l'hanno preceduta e accompagnata. Il linguista che allestisce corpora di italiano non può rinunciare a confrontare racconti/descrizioni raccolte in regioni diverse, quando ne ha la possibilità» (p. 26). E si ricordi poi che VALICO/VINCA nasce come corpus di una varietà (italiano di apprendenti) tra le varietà: si inserisce infatti all'interno di tutta una serie di corpora già allestiti dal gruppo torinese (<http://www.corpora.unito.it>) su italiano delle origini (*Corpus Taurinense*, italiano antico, fiorentino del XIII secolo), italiano di testi universitari (*Athenaeum*, corpus di italiano scritto accademico) e italiano di gruppi di discussione in rete (NUNC, *Newsgroups UseNet Cor-*

*pora*, in italiano, francese, spagnolo, inglese e tedesco che accoglie più di 600 milioni di parole per ogni lingua). Da rilevare poi che VINCA raccoglie testi autentici, ma non spontanei: si tratta di studenti universitari e in minor numero studenti delle secondarie inferiori e superiori non usi alla traduzione intersemiotica, che producono testi per lo più marcatamente stereotipati e da cui a volte si può ricavare l'impressione «che la lingua italiana sia straniera in patria» (p. 12).

La seconda parte del volume («Lessico e sintassi in VALICO e VINCA») indaga il lessico e la sintassi dei due corpora. Si sviluppano analisi quantitative e qualitative che, tenendo in considerazione le peculiarità dei testi raccolti (in primo luogo i condizionamenti dovuti all'input visivo, ossia alla grammatica visiva, e alla consegna), tracciano una visione d'insieme degli aspetti salienti, in un interessante confronto tra produzioni di italofoeni e non italofoeni, dando conto degli studi già svolti in merito e segnalando ambiti di interesse per ulteriori ricerche. Del lessico si analizzano varietà e rappresentatività; la frequenza di nomi, aggettivi e verbi; la presenza di unità lessicali formate da più unità grafiche e si sviluppa un'analisi differenziata per ciascuna storia disegnata. Si dà poi spazio all'analisi delle parole non italiane (che gli studenti di italiano lingua straniera tendono ovviamente a non usare): si rileva la presenza accanto a prestiti consolidati (i più frequenti: *yacht*, *boss*, *bikini*, p. 104) di parole 'tappabuchi' inglesi o della propria lingua madre usate per coprire vuoti, salti di conoscenza in L2 («Finalmente Carnel-Sanders è stato reached», p. 105) e di creazioni ibride «composte da una parte radicale straniera e una morfologia ispirata all'italiano» (*Super marcheto* - sudtirolese -; *levanta*, *levantava*, *levantato* - spagnolo -; *blasfemare*, *mustacche*, *monstro* - francese -, p. 107), segnale di una morfologia *in fieri*, ancora incerta (p. 107). Infine, attenzione specifica è volta all'elevato numero di hapax di «parole ricercate e non frequenti usate dai non italofoeni» (p. 110).

La sintassi è sensibile, come il lessico, ai condizionamenti dell'input visivo, delle consegne e dal fatto che la quasi totalità degli autori dei testi siano apprendenti di italiano LS (e non L2). L'analisi per corpora appaiati fa emergere come i testi di VALICO divergano rispetto a quelli di VINCA in caratteristiche che rientrano «nella sfera della costruzione del testo e della relazione tra contenuto informativo e struttura sintattica» (p. 112): alla maggiore complessità sintattica di VINCA costellata da numerose Unità di appendice, corrisponde la maggiore linearità e esplicitezza sintattica di VALICO (pp. 112-13). Dall'analisi per livelli di interlingua diversi si osserva un miglioramento complessivo della correttezza formale e dell'articolazione sintattica del testo, mentre lunghezza e articolazione in unità informative si mantengono pressoché omogenee, scandite dalle vignette e ad esse strettamente vincolate. L'analisi conduce poi a riflettere, oltre che sul ruolo guida uniformante delle immagini, sul fatto che a L1 diverse corrispondono diversi *thinking for writing*. Speciale attenzione è riservata agli studi di

sintassi e morfosintassi realizzati a partire da materiali VALICO in merito all'ordine dei costituenti della frase (all'intersezione tra sintassi e pragmatica) e all'opposizione tempo-aspettuale. Le considerazioni finali sono invece riservate alle possibilità di parsing sintattico applicato a VALICO.

La seconda metà del volume (parti tre, quattro e cinque) ha un taglio maggiormente applicativo rispetto alla prima. «*Oltre il VALICO*» guida il lettore all'uso del corpus in classe: viene indicato come consultarlo, quali strumenti metta a disposizione, quali tipi di attività didattiche siano già state costruite a partire dal corpus e testate in classe, quali siano stati i risultati. La tipologia di esercizi è strettamente legata alla pluridimensionalità del corpus. Alla dimensione intersemiotica, al passaggio da figura a parola e alla teoria del testo sono legate riflessioni e esercizi su presenza e peso dell'immaginario e del bagaglio interpretativo degli studenti nella produzione dei testi. La distribuzione spaziale, essenziale in testi che partono da uno stimolo iconico, è oggetto specifico di un'attività basata su scomposizione e ricomposizione dell'immagine per elicitarne espressioni riguardanti i rapporti nello spazio. Alla dimensione multimodale e contrastiva si legano invece attività volte ad attingere alle voci e agli scritti di nativi e non nativi, stimolando *noticing* positivo e negativo. All'autenticità e rappresentatività del corpus è legata l'elaborazione di esercizi a scelta multipla (disponibili sul sito e consultabili non in rete), costruiti con distrattori basati su errori da interferenza estremamente accurati, accompagnati da commenti, da feedback metalinguistici che spiegano sia la risposta corretta sia i motivi di non correttezza delle altre risposte. Le prospettive di studio si intrecciano tra loro, com'è naturale, e, ad esempio, attività su questioni di tecnica narrativa non escludono, anzi sollecitano, il *focus on form* (p. 151). Le nuvole di parole (accessibili dalla sezione «Statistiche» del sito) consentono tutta una fioritura di attività tra sintassi, semantica e lessico: possono essere usate ad esempio come testi disidratati o per esercitazioni sulla morfologia (p. 155). Si segnalano insomma le potenzialità di uno strumento che permette di ritagliare sottocorpora calibrabili dal docente in base alla finalità dell'attività e gruppo classe a cui si rivolge, grazie all'innesto della base di dati sociolinguistici sul corpus annotato. Esistono poi strumenti specifici a disposizione in rete, sul portale: MorfoWeb, «un glossario di prefissi e suffissi derivativi e compositivi italiani corredati di definizioni ed esempi» (p. 141), utilissimo se consideriamo che «il lessico dei testi italiani è formato di una grande quantità di parole derivate» e So.F.I.A., un modulo sperimentale creato da Luca Procopio che offre all'utente (apprendente o docente) la possibilità di scegliere le domande in base alle quali sarà valutato, strumento di particolare interesse per il futuro docente che intenda allenarsi alla valutazione dell'errore (p. 144).

«VALICO Testi 'esemplari' e risultati di interrogazioni» e poi «VINCA Testi del corpus appaiato» esemplificano come possano essere svolte ricerche su alcuni aspetti specifici (accordo nome-aggettivo; passato prossi-

mo; perifrasi gerundiviale; proposizione completiva) e quali testi vengano restituiti dalla ricerca. Al lettore sono proposte vere e proprie antologie di testi «costruite intorno a criteri (morfo)sintattici» che «si propongono di offrire una finestra sulle possibilità offerte da VALICO e VINCA per lo studio della struttura del periodo e dei costituenti che lo compongono» (p. 111). Infine, nell'ultima parte, «Corredo iconico, liste, bibliografia» si pongono a disposizione del lettore le vignette da cui la raccolta prende avvio, le nuvole di parole, le 400 parole più frequenti in VALICO e in VINCA, la lista delle tesi realizzate sui due corpora e un fitto elenco di collaboratori a ulteriore testimonianza della profonda e allargata condivisione della ricerca che è nel fondo punto di partenza e approdo del progetto, il cui risultato, il corpus, è liberamente disponibile in rete.

*Italiano di stranieri. I corpora VALICO e VINCA* si presta dunque a una lettura orizzontale per le indicazioni d'uso che offre ed anche a una lettura verticale, legata allo sviluppo nel tempo del progetto, una stratificazione che resterebbe invece per lo più invisibile a una piana consultazione del repertorio messo a disposizione in rete. Nel suo insieme, il libro di Elisa Corino e Carla Marellò racchiude dunque passato, presente e futuro di un corpus di apprendenti plurimodale e intersemiotico che occupa uno spazio di primo piano nell'ambito degli studi sull'italiano lingua straniera.

