

# Recensione: Italiano di stranieri. I corpora VALICO e VINCA

Elisa Corino, Carla Marellò.

Perugia: Guerra Edizioni, 2017

Katarina Zavisin

Università di Belgrado

Nel presente contributo, prendiamo in esame il volume di Elisa Corino e Carla Marellò, in cui sono descritti e analizzati i due corpora VALICO (*Varietà Apprendimento Lingua Italiana Corpus Online*) e VINCA (*Varietà di Italiano di Nativi Corpus Appaiato*), frutto del lavoro del gruppo di ricerca dell'Università di Torino guidato dalla professoressa Carla Marellò. Il libro si propone come guida per docenti di italiano L2 e linguisti, fruitori principali del volume. Essendo i due corpora destinati alla fruizione in rete, le autrici consigliano ai lettori del volume di consultare il sito contestualmente alla lettura del testo<sup>1</sup>.

Il libro comprende sei principali capitoli ed è corredato da un indice dei concetti principali, da una ricca bibliografia, la quale permette ulteriori ricerche e approfondimenti del corpus, nonché dagli esempi degli stimoli iconici e tematici usati nella ricerca.

Nel primo capitolo, intitolato *VALICO e VINCA*, le autrici spiegano le ragioni della ricerca, descrivono le fasi del lavoro della formazione dei due corpora, chiarendone il funzionamento e la struttura anche attraverso un confronto con altri corpora come ICLE (*International Corpus of Learner English*), il “cugino” tedesco FALCO (*Fehler Annotiertes Lerner Korpus des Deutschen als Fremdsprache*), e i corpora italiani ADIL2 (*Archivio digitale di Italiano L2*), MERLIN (*Multilingual Platform for European Reference Levels*) e LISSICS (*Lessico dell'Italiano Scritto della Svizzera Italiana in Contesto Scolastico*). Il corpus VINCA, con i suoi 167.727 token, è una risorsa di piccole dimensioni, la

---

<sup>1</sup> [www.valico.org](http://www.valico.org)

cui principale funzione è quella di fungere da corpus di controllo di VALICO, a sua volta di discreta grandezza (382.098 token).

I due corpora, descritti come comparabili, raccolgono testi prodotti da apprendenti non italofofoni (VALICO) e scritti di italofofoni (VINCA). VINCA è annotato in base a metadati simili a quelli di VALICO, etichettato per parte del discorso (*part-of-speech*, PoS) e composto di testi elicitati con l'input dalle stesse storie illustrate usate per VALICO. Esso funge da corpus di controllo di VALICO: è indispensabile, infatti, che tale risorsa possa servire da strumento di verifica della trama, della ricchezza del lessico e dei connettivi, dell'uso dei tempi, della lunghezza e della composizione delle catene anaforiche e altro ancora.

Inizialmente nel corpus sono stati inseriti testi molto vari per lunghezza e contenuti, testi che rappresentano sia la raccolta dei materiali messi a disposizione dagli insegnanti, sia i materiali elicitati dai ricercatori torinesi. Questo corpus eterogeneo prende il nome di GranVALICO; il nome VALICO, a partire dal 2008, si riferisce invece al corpus "bilanciato", contenente i testi scritti a partire dai cinque stimoli iconici disegnati da Leonardo Borazio: *Amore, Equivoco, Scontro, Sogno e Stazione*.

I corpora non raccolgono soltanto testi scritti. Nella sezione *Le voci di VINCA* troviamo, infatti, racconti orali stimolati dalle storie disegnate. Le registrazioni orali rappresentano una ricca fonte di varietà diatopiche dell'italiano orale, permettendo ai docenti di italiano all'estero di avere a disposizione registrazioni che possono fungere da esempio di diverse varietà regionali. Tali varietà sono facilmente consultabili sul sito dove, cliccando su una città a scelta, è possibile ascoltare la descrizione delle vignette da parte dei parlanti di una determinata città e regione. A nostro avviso, questa possibilità rappresenta un contributo particolare nel campo didattico dell'insegnamento di italiano LS/L2. Essa è da considerarsi, innanzitutto, una fonte rilevante per i docenti di italiano all'estero che, molto spesso, negli esercizi di ascolto proposti dai manuali di italiano, non possono accedere a quelle che sono le varietà regionali, le quali invece rappresentano una realtà significativa della società e cultura italiana, da sottoporre all'attenzione degli studenti stranieri. Non avendo la possibilità di ascoltare i parlanti nativi o, per lo meno, i parlanti di diverse provenienze regionali, gli studenti stranieri possono essere in tal modo sensibilizzati alle caratteristiche delle varietà diatopiche dell'italiano orale.

VALICO e VINCA sono interrogabili attraverso una base di dati sociolinguistici raccolti per lo scrivente di ciascun testo: dati come età, durata dello studio dell'italiano ed eventuale permanenza in Italia, lingua madre e altre lingue studiate, scuola e contesto in cui viene affrontata la prova sono disponibili per ogni testo e costituiscono possibili filtri di interrogazione dei corpora.

Dall'analisi di tali dati risulta che l'autore tipico delle produzioni scritte è una donna fra i 19 e i 25 anni. I dati relativi allo status indicano che l'italiano è studiato da varie tipologie di studenti. La domanda sulle conoscenze delle altre lingue da parte degli scriventi rivela che lo studio dell'italiano mediamente segue l'apprendimento di altre lingue indoeuropee.

Nel quinto paragrafo del primo capitolo, intitolato *Dal testo scritto all'interrogazione online: trascrizione e annotazione*, Elisa Corino spiega il meccanismo della trascrizione e annotazione dei testi, offrendo un quadro dettagliato sul funzionamento dei corpora e sulle possibilità di consultazione. L'autrice mette in rilievo alcuni possibili utilizzi dei dati raccolti, come, ad esempio, il confronto tra le strutture, gli errori e il lessico di un apprendente particolare con le produzioni scritte di apprendenti a lui simili, con i quali condivide lo stesso contesto di insegnamento e lo stesso libro di testo. Ciò ha lo scopo di individuare eventuali regolarità o fenomeni isolati o singoli. Questo paragrafo è una preziosa fonte di idee dell'utilizzo e della consultazione dei corpora, sia per i docenti di italiano L2/LS sia per i ricercatori linguisti.

Il secondo capitolo, *Lessico e sintassi in VALICO e VINCA*, si apre con il contributo di Simona Colombo, in cui si offre un quadro completo di liste e indicatori che consentono un'analisi accurata della struttura e del contenuto di VALICO, VINCA e GranVALICO. L'autrice riporta i dati relativi a diverse categorie, quali le lingue madri degli scriventi, la distribuzione per parti del discorso nei testi di scriventi con diversa lingua madre, nonché gli indicatori per storia e le parole utilizzate in VALICO, storia per storia, lingua per lingua.

Il lessico di VALICO viene accuratamente analizzato da Carla Marelli tenendo in considerazione le varietà lessicali e la frequenza dei lemmi divisi per categorie (verbi, nomi e aggettivi). Inoltre, particolarmente interessante è il risultato dell'analisi dell'uso delle parole straniere nei testi raccolti in VALICO. I dati riportati rivelano infatti che, in linea generale, gli apprendenti di italiano L2/LS sono molto attenti a non usare le parole straniere, ma anche che i parlanti di lingue romanze tendono ad usare maggiormente le parole di un'altra lingua, di solito la propria lingua madre.

Elisa Corino analizza la sintassi in VALICO, concentrandosi su alcuni aspetti specifici, e individua alcune linee generali che possano definire prospettive di ricerca del corpus. Il focus del suo lavoro è l'analisi contrastiva dei testi degli apprendenti di italiano L2/LS e dei nativi, sia del corpus dei testi scritti, sia di quelli orali. L'aspetto della sintassi della frase semplice maggiormente studiato è quello relativo all'ordine dei costituenti. In VALICO, ad esempio, l'ordine prevalente è quello lineare SVO. La gran parte delle imprecisioni sull'ordine dei costituenti riguarda la posizione degli avverbi, il che è dovuto all'interferenza

dalla lingua madre. Inoltre, Corino mette in rilievo che il verbo è uno degli oggetti più studiati nel campo delle varietà di apprendimento: in particolare, l'opposizione tempo-aspetto e l'alternanza tra passato prossimo e imperfetto sono i temi affrontati dalla stessa autrice, il che rappresenta una proficua fonte, sia per i linguisti, sia per i docenti di italiano.

Il terzo capitolo, *Oltre il VALICO*, offre una guida pratica ed efficace al portale con le istruzioni d'uso di diversi menu e le ricerche che possono interessare diversi tipi di utenti (sia i linguisti sia i docenti). Le istruzioni sono chiare e corredate dalle foto tratte dal sito. Inoltre, Marellò si occupa dell'uso didattico dei corpora VALICO e VINCA ponendo l'enfasi sulla loro multimodalità: sono strumenti validi per preparare esercizi sull'uso della lingua italiana e riflessione sulla stessa. Ad esempio, VALICO ben si presta come un ambiente di apprendimento per futuri docenti di italiano nei corsi universitari di glottodidattica, sia per capire quali sono gli errori più frequenti nei parlanti di una certa lingua, sia per individuare le strategie che gli apprendenti adottano per ovviare alle carenze lessicali. D'altronde, è rilevante l'analisi delle motivazioni che muovono uno studente a ritenere un errore meno grave di quanto non lo consideri l'insegnante.

Il quarto capitolo riporta campioni significativi delle produzioni degli apprendenti del corpus VALICO come esempi per le interrogazioni (*query*) che gli utenti del corpus potrebbero effettuare.

L'analisi, nonché i criteri di selezione dei testi VINCA, sono i temi del quinto capitolo. Corino spiega che sono stati inseriti campioni rappresentativi sia per contenuti linguistici dei testi, sia per profili sociolinguistici degli autori. Ai testi di VINCA si riconosce non solo valore scientifico, ma anche didattico in quanto base per la costruzione di esercizi.

L'ultimo capitolo offre il materiale usato come stimolo per le produzioni scritte e orali insieme alla raccolta delle 400 parole più frequenti in VALICO e in VINCA. Nel capitolo vengono inoltre elencati sia i collaboratori esteri che hanno messo a disposizione i materiali dei loro studenti, sia le tesi di laurea discusse all'Università di Torino a partire da VALICO e VINCA. La ricca bibliografia e la sitografia di corpora possono servire da linee guida ai ricercatori, agli studenti e ai docenti di italiano per lavori di approfondimento.

Condividiamo infine l'opinione delle autrici su alcuni punti di forza di VALICO. *In primis*, vorremmo porre l'enfasi sul carattere versatile dei due corpora, in quanto pensati per categorie eterogenee di utenti (linguisti, glottodidatti, insegnanti, studenti), la cui fruizione è resa particolarmente agevole dalla loro disponibilità in rete. Un altro punto di forza è il carattere dinamico dei due corpora: entrambi possiedono una struttura pensata per l'incremento. A ciò si aggiunge il fatto che il carattere omogeneo dei testi all'interno di VALICO dà la possi-

bilità al ricercatore o all'insegnante di individuare sottocorpora omogenei, mentre l'accuratezza della trascrizione dei testi manoscritti in cui si sono registrate le autocorrezioni e le variazioni degli scriventi offrono all'utente dei corpora la possibilità di analizzare accuratamente, per esempio, gli errori e gli sbagli di studenti di diversa lingua madre. Pertanto, i docenti di italiano LS/L2 possono confrontare le produzioni di nativi italofoni e di non nativi su più livelli (lessico, sintassi, organizzazione testuale e altro) per confermare eventuali strategie di evitamento negli apprendenti stranieri.

Un altro punto di forza è l'accessibilità dei due corpora, i quali sono liberamente interrogabili in rete insieme ai numerosi altri creati dal gruppo di ricerca torinese<sup>2</sup>. I tre corpora VALICO, Gran VALICO e VINCA, offrendo un valido quadro dei risvolti applicativi della linguistica dei corpora di apprendenti, rappresentano uno strumento significativo nell'auto-formazione dei docenti di italiano LS/L2, ormai parte integrante dell'aggiornamento glottodidattico.

---

<sup>2</sup> <http://www.corpora.unito.it>