



I corpora VALICO e VINCA:

Elisa CORINO, Carla MARELLO,  
*Italiano di stranieri.*  
*I corpora VALICO e VINCA,*  
Perugia, Guerra, 2017, 288 p.  
ISBN 9788855706063

Elisa CORINO,  
Cristina ONESTI (a c. di),  
*Italiano di apprendenti.*  
*Studi a partire da VALICO e VINCA,*  
Perugia, Guerra, 2017, 168 p.  
ISBN 9788855706094



---

Sandro CARUANA (L-Università ta' Malta)

Fino a una trentina di anni fa, insegnare e apprendere l'italiano, specialmente all'estero, era basato su pochi manuali e corsi di lingua perlopiù di metodo grammaticale-traduttivo e audio-orale. Si avvertiva la carenza di strumenti scientifici, che intanto stavano diventando maggiormente disponibili per altre lingue, in particolare per l'inglese, sui quali basare decisioni pedagogiche e soprattutto, effettuare ricerche scientificamente rigorose. Con il passare degli anni, i materiali in lingua italiana tramite i quali si studiavano le varietà d'apprendimento e le implicazioni glottodidattiche non sono solo aumentati quantitativamente ma hanno dato origine al campo della *Second Language Acquisition*, o linguistica acquisizionale per l'italiano. Questi studi, tra cui quelli citati da Chini (2016) e

in parte concepiti dal filone nato dal Progetto di Pavia, hanno identificato come si evolvono le sequenze d'acquisizione in italiano, specialmente per la morfologia verbale e nominale e per l'ordine sintattico, diventando punti di riferimento per la ricerca sull'italiano L2 di parlanti di diverse nazionalità. Sono serviti anche per indirizzare situazioni create in seguito ai cambi demografici avvenuti in Italia, per via dei quali diventava sempre più urgente imparare la L2 per scopi comunicativi e di immediata fruizione.

Tra gli strumenti che si sono sviluppati parallelamente ai lavori sull'interlingua dell'italiano vi sono i corpora linguistici dei dati di apprendenti. I lavori di Spina (2001), Andorno & Rastelli (2009), Palermo (2009) e Gallina (2015) tra gli altri, sono tappe rilevanti per i corpora di italiano L2 e hanno sfruttato l'evolversi degli strumenti tecnologici, fondamentali per renderli maggiormente accessibili.

I corpora VALICO (382.098 *token* / 20.497 *type*) e il corpus appaiato VINCA (167.727 *token* / 12.387 *type*) sono ideati per essere usufruiti non solo da linguisti, ma anche da docenti d'italiano. I due corpora, VALICO composto da testi di non italofofoni e VINCA da quelli di italofofoni, permettono una comparabilità immediata consentendo di indirizzare domande che fanno spesso linguisti e insegnanti: una parola, un'espressione, una frase ecc. che viene utilizzata da non nativi verrebbe anche adoperata, in contesti analoghi, dai nativi? E se sì, con la stessa frequenza o in un contesto paragonabile?

Nella parte iniziale volume di Elisa Corino & Carla Marello (2017) si spiega come sono nati i due corpora e come essi si inseriscono nel filone di studi, precedentemente citato, della *corpus linguistics*, campo ormai consolidato per varie lingue. È molto utile la disamina tramite la quale si spiega in che cosa VALICO, comparato con altri corpora di italiano L2, sia simile o diverso da questi. Il confronto con i corpora CAIL2, ADIL2 e MERLIN risulta particolarmente interessante. Sempre per quanto riguarda la comparabilità, si rileva che i lemmi che figurano più frequentemente nel corpus VALICO sono simili a quelli di altri corpora di italiano L2. Questa rilevazione, che sottolinea le analogie tra i corpora, ha delle implicazioni didattiche perché tra le domande degli insegnanti vi è non solo quali lemmi si devono insegnare ai propri apprendenti in base alla frequenza d'uso, ma anche quali sono quelli che gli stessi apprendenti adoperano regolarmente, avendoli imparati. Ovvero, ciò che dall'*input* krasheniano diventa *intake* e/o *ouput* ed è dunque adoperabile - e modificabile - anche per la produzione di testi scritti.

Il corpus VALICO, oltre a fornire numerose informazioni in merito agli apprendenti che forniscono i dati (è fondamentale, per esempio, il dato che riguarda la loro permanenza in Italia, fattore determinante nello sviluppo delle competenze linguistiche), ha il merito notevole di basarsi su dati prodotti da apprendenti di livello intermedio e avanzato. In chiave di ricerca questo potrebbe contribuire maggiormente per delinearne le sequenze d'acquisizione special-

mente per gli stadi che superano il livello soglia, spesso sintatticamente complessi e in cui bisogna anche tenere conto dell'adeguatezza pragmatico-informativa.

Nel volume si rilevano i vari passaggi della costruzione del corpus, dalla sua architettura alle annotazioni. Come capita nella costruzione di corpora di L2, in questo caso il corpus VALICO, una parte dell'annotazione è avvenuta tramite una correzione manuale, con oltre 11.000 forme che sono state verificate individualmente. Questo è dovuto al fatto che l'interlingua è, per sua stessa natura, una varietà dinamica che contiene varie forme che non sono attestate nella varietà standard, talvolta create tramite calchi, traduzioni o commistioni che risultano dal transfer della L1 dell'apprendente. Si registrano, inoltre, 1.065 lemmi classificati con '*foreign word*' – in questa categoria, tuttavia, si includono anche lemmi usati spesso da italofofoni, per esempio *jeans* e *shopping*.

Una parte sostanziale del lessico presente in VALICO, è costituita da *hapax*, ossia parole che figurano in una sola occasione nel corpus stesso. Questa categoria rappresenta quasi la metà dei lemmi (a differenza di VINCA, dove gli *hapax* sono circa un terzo). Trattandosi di un corpus di L2 scritto, non è sorprendente che ci sia una quantità notevole di *token* con una sola occorrenza – infatti, in questa categoria si inseriscono parole “non giuste” (es. *croscio* e *platonosa*) e quelle ibride, create ricalcando forme della propria L1. Inoltre, sono da segnalare le parole “ricercate” (es. *celeberrimo*, *cibaria*, *boccheggiare*), anche se a volte non usate in modo appropriato a livello pragmatico. Risulta utile, in questi casi, il confronto con VINCA, specialmente tra gli apprendenti avanzati, le cui buone competenze linguistico-grammaticali non combaciano necessariamente con quelle pragmatiche, come è stato riscontrato anche in qualche studio svolto a Malta (Gauci et. al, 2017), ma non solo.

Una parte del volume, in linea con il proposito di rendere i corpora utili agli insegnanti, è dedicato agli usi didattici di VALICO e VINCA. L'avere a disposizione una ricchezza tale di dati può, infatti, essere sfruttato in classe

specialmente in considerazione del fatto che molti apprendenti avanzati hanno il potenziale di svolgere compiti di *editing*: identificare le forme errate, o quelle inadeguate, e formulare ipotesi per migliorarne lo stile e l'idoneità contestuale. Questi compiti, peraltro, rispecchiano anche il concetto di *process writing*, nel quale scrivere diventa un'attività collaborativa, sociale più che individuale.

La parte finale del volume è interamente dedicata a fornire esempi di testi, sia da VALICO sia da VINCA. Questi sono utili per avere un riscontro diretto, *hands on*, del lavoro svolto e del processo intrapreso, per esempio per il *PoS tagging* e per rendere disponibili i dati degli apprendenti. Il sito online (<http://www.valico.org>), consente di effettuare ricerche in maniera semplice e immediata in base a campi diversi tra cui il singolo lemma, l'uso di forme verbali, segni di interpunzione e la sintassi. Si auspica che questo strumento possa essere sviluppato ulteriormente in futuro, specialmente per raffinare le ricerche morfologiche, visto che il corpus presenta numerosi dati che si possono esplorare maggiormente in questo senso.

Al volume Corino & Marello (2017) si affianca il lavoro, pubblicato nello stesso anno e curato da Elisa Corino e Cristina Onesti, in cui la maggior parte dei contributi si basa su studi e riflessioni riguardanti i corpora. Come si evince dall'introduzione, vi sono lavori in cui gli strumenti per l'elicitazione dei dati di VALICO si usano per studiare apprendenti di madrelingua inglese (Bolognesi), per fornire contributi linguistici (Mulhall e Bettega & Russo) o contributi didattici (Capponi e Masla). *Le voci di VINCA* (Colombo e Romano & De Iacovo) invece illustra uno strumento recentemente aggiunto al sito del corpus online: un insieme di racconti orali di italofofoni di varie regioni italiane a partire dalle storie diseguate servite a stimolare i testi scritti di VALICO e VINCA.

I dati inclusi in VALICO e VINCA servono per approfondire maggiormente temi specifici. Mulhall, per esempio, studia come *High Frequency Verbs* (es. *andare, avere, fare*, ecc.) vengono usati da soggetti di madrelingua in-

glese per formare collocazioni poi adoperate per descrivere le vignette VALICO. Bettega & Russo propongono una disamina dell'italiano L2 di arabofoni, partendo da testi di egiziani e tunisini presenti in VALICO e sottolineando che alcune difficoltà di apprendimento della morfologia verbale sono da attribuire al sistema morfologico – molto più analitico rispetto a quello dell'italiano – della loro L1.

Sia Capponi sia Masla usano il corpus VALICO per fare, rispettivamente, considerazioni traduttologiche (spagnolo-italiano) e interlinguistiche (russo-italiano). Ambedue i lavori portano a riflessioni didattiche, anche in chiave metalinguistica. Colombo, che in uno scritto del 2009 aveva studiato la morfologia degli alterati nei testi di francofoni in VALICO, qui indaga gli alterati presenti nei testi di parlanti nativi d'italiano raccolti in VINCA e ne ricava suggerimenti per la didattica di italiano L2.

Bolognesi, dal canto suo, avanza alcune ipotesi di natura semiotica sulle vignette di VALICO, partendo dalla constatazione che la comunicazione visiva viene interpretata diversamente da parte di apprendenti diversi: si torna qui a sottolineare il merito di VALICO nel fornire informazioni biografiche dettagliate sui soggetti che hanno fornito i lavori scritti raccolti nel corpus.

I sette contributi inclusi nel volume di Corino e Onesti (2017) rappresentano un esempio di quanto sia estesa l'applicazione della metodologia, degli strumenti e dei dati in VALICO e VINCA: non a caso le curatrici le definiscono sette 'sfumature'. Pur affrontando tematiche diverse, che non si collegano necessariamente direttamente l'una all'altra, il loro filo conduttore è rappresentato da un approccio *bottom-up*, in cui si parte dai dati per fare constatazioni linguistiche.

I siti di corpora, elencati in Corino & Marello (2017), nonché l'indice analitico presente in ambedue i lavori, arricchiscono i volumi e rendono più facile l'identificazione dei temi portanti, consentendo ai lettori di ampliare le proprie ricerche o di identificare facilmente ciò che vorrebbe approfondire.

I vari cambiamenti demografici, per via dei quali l'insegnamento dell'italiano L2 è diventato fondamentale anche in Italia, nonché la disponibilità e la duttilità di strumenti online, hanno portato a degli sviluppi che rendono lo studio dell'interlingua più rigoroso e scientificamente valido. In quest'ottica VALICO e VINCA forniscono un contributo di tutto rispetto, e i corpora si prestano a vari usi per utenti diversi, in Italia e all'estero.

### BIBLIOGRAFIA

- Andorno, C. / Rastelli, S. (a cura di) (2009). *Corpora di italiano L2: tecnologie, metodi, spunti tecnici*. Guerra: Perugia.
- Chini, M. (2016). "Elementi utili per una didattica dell'italiano L2 alla luce della ricerca acquisizionale", in *Italiano LinguaDue*, 2: 1-18  
<https://riviste.unimi.it/index.php/promoitals/article/view/8172/7806>, data d'accesso 30 aprile 2020.
- Colombo, O. (2009). "L'apprendimento dell'alterazione in italiano L2 in produzioni scritte di francofoni", in Corino, E.; Marelli, M. (a cura di), *VALICO. Studi di linguistica e didattica*, Perugia, Guerra Edizioni, pp. 111-135.
- Gallina, F. (2015). *Le parole degli stranieri. Il lessico dell'italiano parlato da stranieri*. Guerra: Perugia.
- Gauci, P. / Ghia, E. / Caruana, S. 2017. *The pragmatic competence of student-teachers of Italian L2 in Current Issues in Intercultural Pragmatics*, eds. I. Kecskes & S. Assimakopoulos. Amsterdam: John Benjamins, 323-346.
- Palermo, M. (a cura di) (2009). *Percorsi e strategie di apprendimento dell'italiano lingua seconda. Sondaggi su ADIL2*. Guerra: Perugia.
- Spina, S. (2001). *Fare i conti con le parole. Introduzione alla linguistica dei corpora*. Guerra: Perugia.