

ITALIANO A STRANIERI

Rivista semestrale
per l'insegnamento dell'italiano
come lingua straniera/seconda

Anno 2022
Numero 32

Direttore responsabile
Telis Marin

Direttore scientifico
Matteo La Grassa

Comitato scientifico
Paolo Balboni
Università Ca' Foscari di Venezia
Ivana Fratter
Università degli Studi di Padova
Annarita Guidi
Università degli Studi Roma Tre
Elisabetta Jafrancesco
Università degli Studi di Firenze
Giuseppe Paternostro
Università degli Studi di Palermo
Daniel Stapek
Uniwersytet Wrocławski
Donatella Troncarelli
Università per Stranieri di Siena
Andrea Villarini
Università per Stranieri di Siena

Revisori
Luisa Amenta,
Università degli Studi di Palermo
Antonella Benucci,
Università per Stranieri di Siena
Mario Cardona,
Università degli Studi di Bari
Simone Casini, *University of Toronto*
Nevena Ceković, *Università di Belgrado*
Francesca Gallina,
Università degli Studi di Pisa
Giulia Grosso,
Università degli Studi di Cagliari
Maria Cecilia Luise,
Università degli Studi di Udine
Yahis Martari,
Università degli Studi di Bologna
Laura McLoughlin, *Galway University*
Anthony Mollica, *Brock University*
Francesca Nicora, *Galway University*
Graziano Serragiotto,
Università Ca' Foscari di Venezia

Responsabile di redazione
Antonio Bidetti

Tiratura
8.000 copie
ISSN 1790-5672

INDICE

Editoriale pag. 2
Matteo La Grassa

SEZIONE MONOGRAFICA

I corpora di apprendenti nella didattica dell'italiano LS/L2* pag. 3
Elisa Corino - Università di Torino

Risorse linguistiche *corpus based* per l'italiano dell'architettura e delle costruzioni* pag. 11
Mariapia D'Angelo - Università di Chieti-Pescara

Insidie dell'italiano. Strumenti e strategie per la gestione del testo didattico* pag. 16
Gianluca Biasci - Università per Stranieri di Siena

ARTICOLI

Didattica digitale. Strumenti, tecniche e formazione docente* pag. 20
Gerardo Fallani - Università per Stranieri di Siena

UNO SGUARDO IN CLASSE

Le OER per facilitare l'adattamento socioculturale e lo sviluppo di competenze socio-pragmatiche in apprendenti di italiano LS pag. 26
Andrea Civile - Università per Stranieri di Perugia

Appuntamenti pag. 31

Libri e novità pag. 32

Gli articoli segnalati nell'indice con * sono stati sottoposti a un processo di referaggio anonimo.

La rivista *Italiano a stranieri* è reperibile durante i nostri workshop e presso il nostro stand nelle fiere in cui Edilingua è presente.

Per maggiori informazioni e per consultare la versione sfogliabile visitate il nostro sito www.edilingua.it.

Editore
Edizioni Edilingua
Via Giuseppe Lazzati, 185 00166 Roma
tel. +39 06 96727307, +39 06 94443138
www.edilingua.it, info@edilingua.it

EDITORIALE

di Matteo La Grassa

Direttore scientifico Italiano a stranieri

Come è noto, la maggior parte dei testi (soprattutto se consideriamo i *testi input*) utilizzati nella didattica delle lingue è soggetta a un qualche tipo di “manipolazione”, o perché possano risultare più comprensibili agli apprendenti, o perché si intende farne notare particolari aspetti linguistici o stilistici. Si tratta di azioni che ogni docente mette quotidianamente in atto ma che, grazie all’uso di strumenti informatici, assumono contorni diversi rispetto alle sole attività svolte con carta e penna e, soprattutto, consentono di dare un ruolo maggiormente attivo agli apprendenti. All’uso di alcuni tra questi strumenti è dedicata la sezione monografica della rivista. Gli articoli non prendono in esame particolari *App* o *Software autore*, ma si concentrano su strumenti di interrogazione e analisi di testi che grazie alla diffusione della Rete e allo sviluppo di programmi sempre più *user-friendly* sono ormai utilizzabili anche in contesto didattico.

Il contributo di Corino ci introduce all’uso didattico e formativo dei *corpora*, ampie raccolte di testi in formato elettronico, uniformemente etichettati in modo da essere analizzati con vari scopi. In primo luogo, l’autrice prende in esame un tipo particolare di raccolte di testi, i *learner corpora*, prodotti da apprendenti di italiano L2/LS. Questi possono essere proficuamente utilizzati dai docenti di lingua al fine di osservare e analizzare produzioni libere o elicitate in vario modo di apprendenti con livelli di competenza diversi. Da tali osservazioni, i docenti potranno trarre indicazioni, per esempio per l’elaborazione di più adeguate prove di verifica. *Corpora* di parlanti madrelingua, d’altra parte, potranno essere utilizzati in classe per fare notare agli apprendenti, per esempio, particolari strutture verbali o regolari cooccorrenze lessicali.

L’articolo di D’Angelo presenta un interessante progetto multilingue che, con riferimento all’italiano, prende in esame un ampio corpus di testi della lingua dell’Architettura e delle Costruzioni. In particolare, viene presentata la struttura di unità incentrate su questi temi, che possono essere svolte in autoapprendimento o in classe. Le unità prevedono anche l’interrogazione del corpus da parte degli apprendenti, al fine di promuovere la riflessione su particolari caratteristiche dei testi. L’autrice, anche sulla base dei primi risultati di una indagine svolta con apprendenti universitari che hanno svolto le unità, auspica che le risorse del progetto possano essere utilizzate all’interno di corsi universitari, sottolinean-

do al contempo la possibilità di estenderne l’uso a contesti di insegnamento e gruppi di apprendenti diversi.

L’articolo di Biasci si focalizza sui testi, su alcune loro ‘insidie’ e sulle strategie e gli strumenti che si possono utilizzare per arginarle. L’autore si focalizza innanzi tutto sulla variabilità lessicale tipica dell’italiano e, dopo aver presentato numerosi esempi di geomonimi e geosinonimi non sempre immediatamente riconoscibili, segnala l’esigenza – per evitare ricadute negative sul piano didattico – che il docente sviluppi su questi aspetti una piena consapevolezza, anche servendosi di strumenti lessicografici. Nella seconda parte del contributo, allargando lo sguardo sulla testualità, l’autore prende in esame un testo giornalistico ed individua le difficoltà linguistiche che possono renderne ostica la comprensione. La riscrittura proposta mostra come il docente potrebbe puntualmente intervenire in chiave semplificatrice sul testo usando strumenti quali l’indice di leggibilità GULPEASE.

Nella sezione *Articoli* ospitiamo un interessante contributo di Fallani che, riprendendo il concetto di tecnica didattica nell’accezione comunemente diffusa, suggerisce come nella didattica digitale tale concetto possa essere compreso in quello di Tecnologia Educativa Digitale: questa, infatti, si riferisce a un tempo sia agli strumenti in sé, sia alla perizia necessaria per gestirli e utilizzarli a scopo didattico. Dopo aver indicato le principali competenze e abilità che possono essere sviluppate con le tecnologie digitali, Fallani auspica che nella formazione del docente di lingua possa trovare adeguato spazio lo sviluppo della competenza strumentale, specie quando finalizzata alla realizzazione di materiali didattici.

Il numero della rivista si chiude con il contributo di Civile che riflette su aspetti inerenti allo sviluppo delle competenze sociopragmatiche in italiano LS. L’autore presenta e commenta una proposta didattica rivolta ad apprendenti universitari in mobilità che ha fatto uso di risorse digitali open source. Il percorso proposto ha inteso sollecitare la riflessione su alcuni aspetti sociopragmatici tipici della comunicazione in ambito accademico, con particolare riferimento all’appropriatezza di registro in base ai diversi interlocutori, elemento particolarmente difficile da gestire da parte di molti apprendenti di lingua italiana.

Buona lettura

SEZIONE MONOGRAFICA

I COPORA DI APPRENDENTI
NELLA DIDATTICA DELL'ITALIANO LS/L2

Elisa Corino - Università di Torino

1. Introduzione

Perché usare un corpus nella didattica della lingua? Quando e come usarlo? Sono certamente legittimi interrogativi che i lettori posti davanti al titolo di questo contributo si possono porre. L'introduzione dei corpora nelle buone pratiche didattiche non è da considerarsi una proposta innovativa e di recente formulazione: quello *corpus-based* è un approccio ormai da tempo consolidato in linguistica applicata – e, più generalmente, negli studi di linguistica *tout-court* – così come in contesti di insegnamento e apprendimento delle lingue.

È d'altra parte vero che l'uso di questi strumenti nelle classi della scuola secondaria rimane ancora confinato a esperienze legate a progetti di ricerca-azione che spesso non si traducono nel corrispettivo e auspicabile consolidamento di un approccio che a livello teorico è fortemente raccomandato.

La realtà della scuola impone dei vincoli logistici, infrastrutturali e temporali, ma anche una tradizione metodologica che fa guardare con sospetto delle attività che apparentemente sono complesse, più adatte alla ricerca scientifica che alla pratica di insegnamento/apprendimento, e che richiedono un importante cambio di prospettiva, con una rimodulazione dell'intervento didattico da parte del docente e l'acquisizione di nuove competenze trasversali e di *problem solving* linguistico da parte dei discenti.

Nei paragrafi che seguono vedremo quindi come la ricerca teorica che fa uso dei corpora si può applicare alle pratiche didattiche e come i corpora stessi possono venire utilizzati in modo attivo e costruttivo in classe, con particolare riferimento al contesto dell'insegnamento/apprendimento dell'italiano come LS e L2.

2. Corpora e didattica: tra teorie e pratiche

Dal punto di vista della ricerca linguistica i corpora sono oggi gli strumenti irrinunciabili per osservare l'uso della lingua in contesti diversi e per fare considerazioni sullo stato della *norma*. Le grandi quantità di dati organizzati e interrogabili¹ permettono di testare e validare

ipotesi in modo scientifico e fanno sì che si possano trarre conclusioni relativamente generalizzabili su fenomeni che pertengono a diversi livelli dell'analisi linguistica e sono precisi di contesti d'uso e varietà diverse. È attraverso i corpora e le liste di frequenza, ad esempio, che i lessicografi definiscono i lemmari dei dizionari e le entrate da inserire nelle nuove edizioni; l'accesso ai corpora ci permette di verificare l'andamento e lo sviluppo di alcune strutture morfosintattiche in chiave sincronica e in chiave diacronica; nei corpora paralleli possiamo poi dare risposte a questioni che riguardano la traduzione... Insomma, si tratta di strumenti raffinati che supportano le necessità di ricerca dei professionisti della lingua, in molte delle loro possibili declinazioni.

In questo particolare contesto ci concentriamo su una particolare tipologia di corpus: il corpus di apprendenti (o *learner corpus*). Come osservato in Corino (2022a), esso è il risultato dell'adattamento delle metodologie e delle tecnologie usate dalla linguistica computazionale tradizionale allo studio delle varietà di lingua scritte e parlate dagli apprendenti: i testi diventano interrogabili per parti del discorso, per errori, o per strutture sintattiche e permettono di dare conto delle caratteristiche delle interlingue. Addentrarsi nei contenuti di un corpus significa poter osservare da vicino le regole che governano i diversi stadi di acquisizione e apprendimento, identificare quei passaggi comuni che sono tappe obbligate del processo che va da una L1 a una L2/LS, definire con dati quantitativi, oltre che qualitativi, le proprietà lessicali, morfosintattiche e testuali di ciascun livello di competenza. Si ottengono così "istantanee" delle caratteristiche delle varietà di apprendimento in grado di condurre a descrizioni sem-

in numero finito in formato elettronico trattati in modo uniforme (ossia tokenizzati ed addizionati di markup adeguato) così da essere gestibili ed interrogabili informativamente; se (come spesso) le finalità sono linguistiche (descrizione di lingue naturali o loro varietà), i testi sono perlopiù scelti in modo da essere autentici e rappresentativi. Linguisticamente, inoltre, un corpus è una raccolta di atti di parole, e dai fatti di parole raccolti in un corpus si può risalire ai loro correlati stati di lingue, anche se certamente non tutti gli elementi di una lingua saranno contenuti in un corpus: è l'uso testimoniato dai corpora, anzi, che fonda la lingua, anche se i corpora, essendo per definizione finiti, ne rappresenteranno solo un sottoinsieme.

¹ La definizione di corpus data da Barbera (2013: 18-19) è: raccolta di testi (scritti, orali o multimediali) o parti di essi

pre più dettagliate e utili ai linguisti, ai glottodidatti, ma anche agli apprendenti stessi, che potranno così acquisire maggior consapevolezza dei propri mezzi.

Se ripercorriamo per un breve momento il percorso che ha visto la cruciale intersezione tra la ricerca linguistica e l'uso dei corpora in Italia, non possiamo non soffermarci sul ruolo fondamentale ricoperto dal Progetto di Pavia (Giacalone Ramat 2003) nella definizione della scale di acquisizione e nella descrizione dell'interlingua: l'osservazione sistematica di dati longitudinali organizzati ha portato a stabilire una progressione implicazionale degli elementi² che vengono acquisiti nell'ambito della morfologia nominale e verbale, del sistema pronominale, della negazione etc. E la linguistica acquisizionale con i suoi risultati è stata ed è tutt'ora una fonte preziosa di informazioni per chiunque si occupi di didattica delle lingue seconde e delle lingue straniere: conoscere i meccanismi che governano l'acquisizione spontanea può aiutare a valutare l'opportunità di alcuni interventi in contesti di apprendimento guidato, considerando, nell'ottica della teoria dell'insegnabilità (Pienemann 1998; Bettoni, Di Biase 2005), quali elementi estranei all'ordine naturale di acquisizione inserire in modo "forzato" nel sillabo rivolto ad una certa varietà di apprendimento³.

2 Un esempio piuttosto conosciuto di scala acquisizionale è quella relativa ai tempi verbali che vede la seguente progressione: presente (e infinito) → (ausiliare) + participio passato → imperfetto → futuro → condizionale → congiuntivo.

3 Secondo la teoria dell'insegnabilità (Pienemann 1998; Bettoni, Di Biase 2005), l'insegnamento può favorire l'acquisizione linguistica solo se l'interlingua è vicina al punto in cui la struttura da insegnare viene acquisita nella situazione naturale. Se dunque la sequenza di acquisizione del sistema verbale è presente (e infinito) → (ausiliare) + participio passato → imperfetto → futuro → condizionale → congiuntivo l'ipotesi dell'insegnabilità ci suggerisce che il condizionale

I corpora di apprendenti di italiano L2/LS di cui disponiamo oggi sono strumenti frutto di un'esperienza di ricerca nel campo della linguistica computazionale ormai quasi trentennale in Italia, ma di ben più lunga tradizione in ambiente anglosassone: le origini della ricerca sui *learner corpora* risalgono alla fine degli anni Ottanta, quando accademici ed editori, contemporaneamente ma indipendentemente, iniziarono a raccogliere dati linguistici prodotti da studenti di lingue straniere/secondo: gli uni con lo scopo di far progredire la comprensione dei meccanismi di acquisizione di una L2, gli altri per sviluppare strumenti e metodi pedagogici accuratamente tarati sui bisogni degli studenti. Così, con il passare del tempo e con l'investimento in un campo della ricerca estremamente promettente in termini di ricadute pratiche, i corpora si sono trasformati da strumento fatto da linguisti per altri linguisti in dispositivi *user friendly* rivolti a un pubblico molto più ampio, tra il quale gli insegnanti sono i destinatari privilegiati.

Tra i corpora di italiano che oggi sono a disposizione della comunità scientifica e glottodidattica abbiamo raccolte di italiano parlato, di certificazioni, ma anche corpora di varietà particolari di italiano di apprendenti legate al luogo di produzione dei testi, alla L1 degli informanti, alle tipologie testuali etc. Alcuni di essi hanno generato liste di frequenza lessicale, altri sono stati – e sono tutt'ora – alla base di studi *corpus-based* su vari aspetti delle varietà di apprendimento. Tra i *learner corpora* di italiano LS/L2 ad oggi liberamente accessibili citiamo VALICO (www.valico.org), MERLIN (<https://merlin-platform.eu/>), i corpora di Perugia (LOCCLI e CAIL2 <https://www.unistrapg.it/cqpwebnew/>) e riman-

dovrebbe essere insegnato solo dopo che l'apprendente ha il saldo controllo di tutte le proprietà di imperfetto e futuro indicativo (oltre che di presente e passato composto).



Progetto
ITALIANO
nel tempo libero



- ✓ Un ripasso divertente, con attività motivanti e ludiche
- ✓ Da usare in classe o in autoapprendimento
- ✓ Durante l'anno scolastico o le vacanze estive o invernali

Perché giocando s'impara!

diamo a Corino (2017 e 2022b) per una discussione più approfondita delle loro caratteristiche. Ci sono poi i corpora raccolti a partire dai compiti delle certificazioni sui quali sono stati basati studi sull'ampiezza del lessico e sulle competenze grammaticali (cfr. Spinelli, Parizzi 2010) e corpora di parlato (uno su tutti il LIPS) che hanno fornito importanti liste di frequenza per lo studio delle varietà diafasiche.

Gli usi indiretti dei corpora nell'apprendimento di una L2/LS si ritrovano in approcci quali l'Analisi Contrastiva delle Interlingue (CIA, Granger 1996), che prevede il confronto e l'analisi delle varietà d'uso della lingua nativa vs non nativa, ma ha anche ricadute sulla formulazione di test linguistici, laddove gli errori trovati in un corpus di studenti vengono usati come distrattori nello sviluppo di domande (Marello 2009). Inoltre, usi indiretti si trovano nelle pratiche lessicografiche che si rivolgono agli studenti (Grange, Paquot 2015; Paquot 2012; Spina 2010) e anche per informare la progettazione di libri di testo (McCarten, 2010).

Learner corpora per la ricerca dunque, ma anche corpora per formare e corpora per insegnare (Corino 2014; Forti, Spina 2019), così come anticipato fin dagli anni Sessanta da Randolphe Quirk che auspicava il ricorso ad essi per supplire all'inadeguatezza di materiali didattici basati su una lingua artificiale e artificiosa e che non tenevano conto dell'effettivo livello degli studenti. Le ricadute pratiche e dirette della ricerca sui corpora di apprendenti, inoltre, sono evidenti nella stretta relazione con le buone pratiche di linguistica educativa promosse dalle tecniche di *data-driven learning* (DDL), ancora poco impiegato nella didattica dell'italiano LS (si veda ad esempio Forti 2019), ma in costante crescita e affermazione in contesto anglofono.

3. I corpora per formarsi

L'uso dei corpora nella formazione degli insegnanti è una pratica che spesso viene citata come buona prassi nella letteratura glottodidattica di riferimento, ma che di fatto non si traduce nella sistematica introduzione di questa attività nei sillabi dei corsi. Eppure il valore dei *learner corpora* in termini di linguistica educativa

rivolta ai docenti è ovvio: essi offrono la possibilità di venire in contatto con le interlingue dei propri studenti prima ancora di averli incontrati e di costruire *a priori* dei profili linguistici specifici basati su dati reali e quantificabili (in base ad esempio alla L1 di provenienza, al livello di competenza, al grado di scolarizzazione, etc.) che possono poi guidare l'insegnante nella costruzione del proprio intervento didattico. Utilizzare i risultati della ricerca *corpus-based* permette di decidere su cosa concentrarsi e su come mettere in sequenza e modulare i materiali all'interno della progettazione dei contenuti. Se un insegnante è informato sull'evidenza empirica legata a uno specifico tipo di errore, sarà in grado di decidere con cognizione di causa come trattarlo nei casi in cui lo incontrerà: se dai risultati emerge che un particolare errore non è molto comune, l'insegnante potrà decidere di valutarlo come qualcosa di temporaneo; al contrario, se un risultato dice che un errore è abbastanza comune, soprattutto con il passare del tempo, in questo caso l'insegnante potrà decidere di dedicare un'attenzione speciale alla deviazione dalla norma, al fine di contrastare il modello di apprendimento legato a una determinata forma.

3.1. Formarsi su VALICO

VALICO (Varietà di Lingua Italiana Corpus Online) è ad oggi il più grande corpus di Italiano LS liberamente consultabile online e si compone di 2502 testi nella sua versione bilanciata per L1 di provenienza⁴⁴ (sono 4778 invece i testi raccolti nella sua forma estesa GranVALICO). È una risorsa etichettata per parti del discorso che offre la possibilità di fare ricerche testuali e linguistiche precise, lavorando anche su sottocorpora creati di volta in volta in base alle esigenze dell'utente e a partire dalle variabili sociolinguistiche presenti nei metadati (tra le più rilevanti troviamo l'età degli apprendenti, il loro livello di competenza, le altre LS conosciute, ma anche il gruppo classe in cui si è inseriti o l'eventuale tempo di permanenza in Italia, cfr. fig.1)

⁴⁴ La carta di identità del corpus recita per la precisione 382.098 token, 20497 type, 6934 lemmi. Per una descrizione esaustiva del corpus si veda Corino, Marello (2017).



fig. 1. Finestra di interrogazione di VALICO: le variabili sociolinguistiche.

Possiamo quindi richiedere tutti i testi prodotti da apprendenti germanofoni di età compresa tra i 19 e i 25 anni ed elicitati da un particolare stimolo iconico, suddividerli per annualità di studio della lingua italiana e analizzarli tenendo conto delle informazioni sul profilo linguistico degli informanti per avanzare ipotesi

su possibili interferenze da altre LS. Se volessimo, ad esempio, studiare come all'interno del nostro sottocorpus si sviluppa l'uso dei tempi passati, potremmo fare una query per PoS (*parts of speech*) ed estrarre tutte le occorrenze di passato prossimo, imperfetto o passato remoto (come in fig. 2).

Id	Count	Text
9	900	molto felice , ma purtroppo Carla comincia a sgridare . Non se che cosa succede e ad un tratto Carla sparita . # La donna sgrida sempre e Luigi realizza che non ha vacanze , che non guida la barca a
10	1530	bellestima . Tutto era bene , ma subito la ragazza è di stata arrabita . Ind " Incredibile " , ha detto , " quando finirà questi test ? Li gie aspetto aspetto . . . a aspettati aspetto ho qualche giorno
11	1573	accompagna . Il sole brilla , il mare è tranquillo e tutto sembra perfetto . Ma improvvisamente questo momento incantato viene interrotto brutalmente . La voce dolce della Sua ragazza diventa arrabbiata , acula e lui è totalmente perplesso . Cosa è
12	1802	bella faccia di prima . Questo sorriso sottile ? Dov ' era ? ... Quando Paolo alzava aiza gli occhi ha visto quando vedeva xxx Maria . Lei sembrava veramente arrabbiata . " non mai ti vedo un lavoraro lavorando . "

fig. 2. Finestra di interrogazione di VALICO: l'interrogazione linguistica.

Verifichiamo allora, come diffusamente argomentato in Corino (2012), che gli apprendenti germanofoni presentano significative difficoltà legate alla differenziazione aspettuale tra passato prossimo e imperfetto, al legame tra aspetto e *Aktionsart*⁵, e ai condizionamenti legati alla semantica avverbiale, ma contemporaneamente divergono dal solco predittivo tracciato dalla linguistica acquisizionale con dati che evidenziano usi dell'imperfetto imprevisti nella scale di acquisizione, ossia valori modali che non dipendono esclusivamente da parametri temporali o aspettuati, come l'imperfetto onirico e l'imperfetto narrativo.

Certamente per l'insegnante che conosca il tedesco, tali considerazioni appaiono piuttosto prevedibili, data la distanza del sistema verbale della L1 da quello della lingua di arrivo, ma non tutti i docenti conoscono il tedesco, così come non tutti conoscono l'arabo, il romeno, il cinese, o – perché no – i pur familiari francese e spagnolo. Con questa consapevolezza, da alcuni anni ormai, l'interrogazione dei sottocorpus di VALICO è stata inserita e si è consolidata come prassi nei corsi magistrali di Didattica dell'italiano come lingua straniera e nel master MITAL2 dell'Università di Torino: agli studenti viene chiesto di usare un approccio scientifico all'analisi del *learner corpus*, formulando ipotesi di ricerca relative alle caratteristiche delle varietà di

apprendimento rappresentate dal gruppo linguistico selezionato – possibilmente scegliendone uno la cui L1 è relativamente sconosciuta agli studenti stessi – secondo la seguente consegna⁶:

1. Selezionate un sottocorpus di VALICO (www.valico.org)
2. Indicate quali sono stati i criteri di selezione e quali sono i vostri obiettivi/domande di ricerca
3. Analizzate i testi: quali sono gli errori più frequenti? A che cosa sono dovuti? Qual è il livello degli studenti e quali sono gli elementi che lo rivelano?

I risultati sono molto positivi, sia in termini di evidenti ricadute pedagogiche dell'attività, sia per quanto riguarda il feedback degli studenti stessi sulla sua utilità. Dal punto di vista educativo, gli studenti imparano a riflettere sulle interlingue degli studenti in modo attivo, scompongono i testi in segmenti minimi e, come scienziati con il microscopio, ne ingrandiscono le parti più minuscole e nascoste, portando in superficie elementi che visti da lontano non emergerebbero. Si tratta, per dirla in termini didattici, di un'attività HOTS (*Higher Order Thinking Skill*) che obbliga chi la svolge a classificare, mettere in relazione, trovare le cause e dedurre le conseguenze pensando a possibili soluzioni. Comprendere un errore e attribuirlo ad una categoria, capire se e come correggerlo in relazione al livello dello studente e come spiegare il perché lo si

5 Le ricerche hanno sottolineato la tendenza degli apprendenti a produrre la marca morfologica più affine al valore azionale del predicato, ovvero la marca di aspetto perfetto per i predicati telici e la marca di aspetto imperfettivo per i predicati atelici (Bernini 1990; Andersen, Shirai 1994; Bardovi Harlig 2000).

6 A questi punti se ne aggiunge poi un altro relativo alla produzione di esercizi, di cui tratteremo nel prossimo paragrafo.

ritiene una deviazione dalla norma, non è sempre così scontato: in primo luogo perché una singola occorrenza può racchiudere in sé più di un errore, come in [1] in cui “Qualchi minuti” può essere corretto con “alcuni minuti” (pl.) o “qualche minuto” (sing.) e la preposizione “fra” è usata in luogo dell’avverbio “dopo”, o in [2] in cui si ravvisa un’errata scelta lessicale che in realtà “copre” un errore aspettuale (vederla = guardarla = stare(la) a guardare).

[1] **Qualchi minuti fra** il ragazzo si è reso conto che il brutto sognava.

[2] Matteo non poteva **vederla** senza fare niente,

In secondo luogo perché alcuni elementi dell’italiano sono poco prevedibili ed è difficile dar conto di possibili regole laddove non c’è regolarità strutturale (si pensi, ad esempio alla morfologia derivativa: perché “spazzino” e non “spazzatore”?); e non in ultimo perché alcuni elementi della lingua italiana oggettivamente complessi e difficili da gestire anche per il docente (consideriamo ad esempio la polifunzionalità della particella *si*: riflessiva, impersonale, reciproca, benefattiva, etc.).

Nell’analisi del sottocorpus francofono di prima annualità, una studentessa di master⁷ si è proposta, ad esempio, di osservare alcuni elementi della varietà di apprendimento rappresentata, tra i quali la concordanza interna al sintagma nominale, l’utilizzo di preposizioni nelle costruzioni polirematiche e nelle collocazioni, la *consecutio temporum* dei passati, ricavando indicazioni preziose per un eventuale intervento didattico mirato a potenziare alcune aree che emergono come più problematiche. Fra gli errori presenti si evidenziano a livello ortografico nella trascrizione delle consonanti geminate (*indosava, belisima, immaginare*), meno frequente il fenomeno inverso (*ragazza*); a livello lessicale l’interferenza della L1 (*ho turnato la testa, fatigante, giusto* per “appena/juste”). Per quanto riguarda l’accordo all’interno del sintagma nominale si notano errori nell’accordo articolo-nome, soprattutto in presenza di declinazione irregolare (*la mia capo*), la cui portata investe anche l’aggettivo possessivo (*il suo vacanze, il suo direttrice, nella mio capo*), e nell’accordo soggetto pronominale-participio passato ([*lei se è innervosato*]). I francofoni poi tendono ad omettere l’articolo davanti agli aggettivi possessivi, ma anche a sovraestenderne l’uso anche in contesti in cui l’italiano non lo richiede (*un uomo in su ufficio sogna di su vacanze*), inoltre tendono all’ipercorrettismo in relazione all’espressione del pronome personale e all’esplicitazione del soggetto, tratto peraltro comune a molti parlanti di lingue non *pro-drop* (*Georgio sogna di essere in barca con una bel-*

la donna che se chiama Alyssa. Georgio è contento di essere con Alyssa. Lui guida la barca con un bel tempo. Ma al improvvisamente Alyssa è nervosa e Georgio non capisce perchè Alyssa è nervosa).

Curiosa è la presenza in alcuni testi di elementi propri dello spagnolo: gli apprendenti francesi che utilizzano espressioni di questo tipo possiedono una conoscenza dello spagnolo, oltre che dell’italiano, come L2; nel momento in cui sono in difficoltà nella produzione probabilmente sentono come più vicino all’italiano l’utilizzo di termini e costruzioni morfologiche provenienti dallo spagnolo:

L’altro giorno al lavaro , Antonio ha sognato il suo vacanze . # Si ha è immaginato con il suo direttrice en una barca al mare . # è x muy più bella que ordinare con il capello longui , belli , blondi . # La ragazza se ha è innervosato , Antonio no comprende . il uomo no ha comprendito perchè se si nervosa innervosisce . Il um uomo rendersi si rende conto che è la realidad perchè sua amica la fache si innervosisce

Così come potrebbe essere imputata alla conoscenza dello spagnolo (più che del francese) anche la presenza della prima persona singolare dell’imperfetto in -a: (*io mi annoiava così tanto; io portava soltanto una piccola camicia a fiori; io andava troppo veloce*).

Simili relazioni su gruppi specifici da apprendenti sono state negli anni presentate da numerosissimi studenti e futuri insegnanti di italiano LS/L2, che hanno potuto così scoprire in prima persona che gli ispanofoni alle prime annualità utilizzano il verbo “stare” al posto di “essere” e viceversa (*Io sono pensando*), trasferiscono impropriamente alcune costruzioni spagnole VERBO + PREP (*pensando en le mie vacanze; sogna con il mare*), hanno difficoltà nella scelta dell’ausiliare (*la moglie ha diventato molto furiosa*); oppure che i russofoni, che difettano della categoria “articolo”, trovano grandi difficoltà nell’uso di determinanti; o ancora che tra gli errori più frequenti dei sinfoni, a dispetto di un lessico relativamente avanzato, si riscontrano frequenti errori nella formazione degli avverbi (es: *di repente*), con un uso sovraesteso di avverbi in -mente (*contentamente, repentinamente, comodamente, affezionatoamente, attentamente, etc.*), errori morfosintattici che riguardano la concordanza persona-tempo verbale (anche con il presente indicativo) e l’accordo genere e numero, l’uso dei verbi riflessivi e, come per molti altri gruppi linguistici, la scelta dell’ausiliare essere/avere.

Talvolta l’interrogazione del corpus è stimolo per l’insegnante per approfondire degli aspetti delle L1 dei suoi apprendenti che poi potranno tornargli utili nella gestione di una didattica tarata sui bisogni specifici, come riporta una studentessa del MITAL2:

⁷ Si tratta della relazione finale del modulo di Editoria di settore presentata dalla dott.ssa Giulia Guzzo.

Incuriosita dalla presenza di questo errore, sono andata a ricercare le possibili cause nella letteratura presente, scoprendo che la parola cinese è invariabile e che questa lingua non contiene quindi le caratteristiche del genere e del numero. [...]

Andando a ricercare le possibili cause di questi errori, ho potuto apprendere che nella lingua cinese il verbo non è una categoria grammaticale statica, ma a seconda del contesto assume il significato di verbo, aggettivo e nome; inoltre è una forma invariabile [...].

Ecco allora che l'analisi *corpus-based*, inizialmente elicitata e guidata dal formatore, diventa pretesto e spunto per ampliare ed arricchire le proprie conoscenze; si tratta di un risultato che risponde pienamente alla nozione di *affordance* (cfr. Hafner, Candlin 2007; Boulton 2012), ovvero la capacità di un oggetto – linguistico nel nostro caso – di suggerire all'utente come interagire con esso: il corpus diventa strumento di autoformazione che sviluppa l'autonomia dello studente e promuove processi di formazione di nuove competenze.

4. I corpora per formare

Una volta avvicinati gli insegnanti all'uso dei *learner corpora* come strumenti di conoscenza e autoformazione sulle varietà di apprendimento, è necessario mostrare loro anche come usarli attivamente della didattica. Due sono le possibili strade: il corpus come fonte di esercizi basati proprio sugli errori degli apprendenti stessi, oppure il corpus come strumento di educazione linguistica attiva da far scoprire e manipolare dagli studenti stessi.

Tra le esperienze di creazione di esercizi *corpus-based* possiamo certamente citare la ricerca-azione intorno a VALICO, come ad esempio la creazione di test a scelta multipla basati su distrattori estratti dal corpus (Marello 2009; Marello, Corino 2011, <<http://www.valico.org/esercizi.html>>) o l'uso di VINCA, corpus appaiato a VALICO, per esercizi di riordino di testo e vignette (Marello 2012).

L'insegnante che debba produrre esercizi mirati ad un gruppo linguistico particolare, potrà, come suggerito fin da Corino e Marello (2009), selezionare nel corpus gli enunciati che contengono gli errori più frequenti e, a partire da quelli, creare esercizi grammaticali e lessicali a scelta multipla con "distrattori efficacemente distraenti" perché non creati a tavolino, ma estratti dalle produzioni reali degli apprendenti e quindi maggiormente rappresentativi di una certa varietà di apprendimento. Tra le indicazioni ritroviamo sicuramente il fatto di enucleare il tipo di errore su cui si vuole focalizzare l'attenzione all'interno di un enunciato prodotto da apprendenti l'italiano L2 che, mol-

to probabilmente, ne contiene altri: bisogna lasciarne uno solo di tipo sintattico, al massimo insieme ad uno di tipo lessicale oppure ad uno di accordo morfologico. Utilizzare enunciati che sono contemporaneamente errati sul piano ortografico, lessicale o sintattico confonde ancora di più l'apprendente, che può perdere di vista l'errore più importante "bersaglio" del test. Come suggerivano Corino e Marello (2009: 282): «non è una procedura banale costruire scelte multiple plausibili; è un percorso a ritroso nell'applicazione delle regole».

A partire dall'occorrenza

Quando c'è il tempo, bisogna a cercare di approfondire gli argomenti che ci interessano

l'esercizio da proporre potrebbe presentarsi come segue (Livello A2 o B1):

- Quando c'è tempo, bisogna cercare di approfondire gli argomenti che ci interessano.
- Quando c'è tempo, bisogna cercare ad approfondire gli argomenti che ci interessano.
- Quando c'è tempo, bisogna cercare per approfondire gli argomenti che ci interessano.
- Quando c'è tempo, bisogna cercare da approfondire gli argomenti che ci interessano.

Si può poi pensare di attribuire pesi diversi ai vari distrattori in relazione al livello degli studenti: un errore di accordo soggetto-verbo, non riconosciuto da studenti di livello intermedio, è sicuramente più preoccupante della scelta di un tempo sbagliato. Allo stesso modo, la scelta dell'ausiliare errato in verbi importanti utilizzati fin dalle prime lezioni del livello A1 va sanzionato in modo più pesante rispetto a una forma riflessiva incompleta. L'esercizio proposto sopra, ad esempio, potrebbe essere corredato dei seguenti valori:

- 4 punti. Risposta esatta *cercare* regge *di* + infinito.
- 0 punti. Al livello B1 si deve sapere che *cercare* regge *di* + infinito.
- 0 punti. Al livello B1 si deve sapere che *cercare* regge *di* + infinito. Interferenza di *to look for*, grave se si considera che *cercare* in questo contesto è *to try*.
- 0 punti. Al livello B1 si deve sapere che *cercare* regge *di* + infinito.

4.1. Il data-driven learning (DDL)

Una volta formato se stesso, usato quanto appreso per creare esercizi mirati da somministrare ai propri studenti, l'insegnante può compiere un terzo passo nella scalata all'uso dei corpora e passare alla fase di introduzione in classe e al coinvolgimento attivo dello studente nel processo di esplorazione della lingua attraverso strumenti e metodi della linguistica applicata e computazionale: il DDL.

Il ruolo della didattica data-driven è quello di portare induttivamente l'utente che esplora i dati a trarre delle indicazioni su strutture, usi, funzioni, contesti in un processo di indagine e problem solving linguistico (dalla definizione della *query*, alla selezione dei risultati, alla loro classificazione e decodifica) che stimola la competenza linguistica – ma soprattutto quella metalinguistica – attraverso il *focus on form* contestualizzato. Flowerdew (2015) descrive il DDL come una metodologia che si pone alla confluenza di approcci diversi: la teoria del *noticing*, l'apprendimento di stampo costruttivista, le teorie di interazione socioculturale di Vigotsky, con un'attenzione particolare verso gli stili di apprendimento e i processi cognitivi che lo sottendono.

Attraverso l'approccio DDL si concretizza quel ruolo, spesso teorizzato ma di fatto poco realizzato, di insegnante come mediatore e facilitatore, che usa il corpus per esemplificare e sostenere con dati concreti la definizione di pattern, costruzioni, significati e usi, mettendo in atto l'auspicata transizione da un insegnamento basato su regole a un approccio basato su prove e fatti di lingua.

Gli apprendenti possono, ad esempio, imparare ad usare corpora di nativi per verificare le proprie ipotesi o scoprire nuovi elementi che arricchiscono il proprio bagaglio linguistico.

VINCA (Varietà di Italiano di Nativi Corpus Appaiato, cfr. Corino, Marello 2012), il corpus appaiato a VALICO elicitato a partire dagli stessi stimoli iconici (e basato sulla stessa architettura computazionale), potrà essere usato per verificare l'aggettivazione relativa ad alcuni elementi della traduzione intersemiotica, per osservare quali tempi e modi i nativi usano per narrare un determinato evento o per segnalare un cambiamento nel punto di vista, quali sono le collocazioni con un certo nome o un verbo, etc. Si veda ad esempio la ricerca in fig. 3, in cui si chiede quale verbo preceda il lemma "terra", che restituisce collocazioni V-N e costruzioni quali "cadere per terra", "stendere a terra" o "finire per terra"⁸.

⁸ Il riferimento è la vignetta *Amore*, <<http://www.valico.org/vignette/amore.pdf>>

fig. 3. Finestra di interrogazione linguistica di VINCA: VERBO + Terra.

Moltissimi sono ormai i corpora di italiano di nativi che possono utilmente essere utilizzati per attività DDL, soprattutto per studenti avanzati. Per apprendenti intermedi, piuttosto, ci sono strumenti espressamente creati per la didattica della lingua come itSkELL⁹, che consentono agli studenti di digitare una parola o una stringa di parole e di esplorarla in termini di concordanze o di relazioni (*work sketch*) che mostrano i *type* e i *cotesti* che più frequentemente co-occorrono con la parola cercata. Si veda banalmente il *word sketch* del lessema "patata" (cfr. fig. 4), in cui sono mostrati

tutti i verbi che possono co-occorrere con il nome: la schermata restituisce immediatamente e in una sola esaustiva "carta di identità della parola" tutte le possibili azioni che possiamo fare quando cuciniamo una patata, così come gli aggettivi ad essa riferiti, le preposizioni che seguono il sostantivo, etc.

Questo corrisponde al tipo più semplice di attività DDL: guida l'apprendente a osservare se e come una certa forma è effettivamente usata dai nativi oppure no; altri tipi di attività possono comportare una serie di passi diversi che costituiscono un processo di scoperta guidata e più complessa, a seconda del livello dei discenti.

⁹ <<https://skell.sketchengine.eu/#home?lang=it>>



fig.4. Word sketch da itSkELL: "patata".

5. Conclusioni

Abbiamo visto come i corpora di apprendenti possono essere usati per formarsi e per formare e abbiamo accennato a come questi strumenti possono utilmente essere consegnati agli stessi studenti per sviluppare competenze linguistiche e metalinguistiche attraverso percorsi di ricerca scientifica.

Siamo ben consapevoli che né la formazione né le pratiche didattiche che riguardano le lingue – e nel nostro caso l'italiano LS/L2 – possono essere basate unicamente su questi strumenti e l'approccio data-driven. Oltre alle ovvie ragioni metodologiche e glottodidattiche bisogna tenere conto dei vincoli logistici e della necessità di attrezzature, così come del tempo e dell'impegno intenso che le proposte qui avanzate richiedono. La letteratura di riferimento è ormai ampia e ricca di spunti e possiamo però affermare con una certa sicurezza, in base all'esperienza diretta in progetti di ricerca-azione e nella pluriennale formazione degli insegnanti, che l'introduzione dei corpora in questo sistema a tre poli fatto di *scoperta* (l'analisi del corpus di apprendenti da parte degli insegnanti) – *applicazione della scoperta* (la creazione di esercizi) – *guida alla scoperta* (il DDL) costituisce un elemento importante di costruzione e consolidamento di competenze che difficilmente si ottengono in altro modo. L'auspicio è quindi quello che gli insegnanti comincino ad esplorare e a sperimentare in un campo che è di fatto uno degli anelli di congiunzione tra ricerca e didattica.

Riferimenti bibliografici

Boulton A., 2012, *Beyond Concordancing: Multiple Affordances of Corpora in University Language Degrees*. *Procedia: Social and Behavioral Sciences*, 34, pp. 33-8.

Bettoni C., Di Biase B., 2005, *Sviluppo obbligato e progresso morfosintattico: un caso di processabilità in italiano L2*. In *ITALS III*, pp. 27-48.

Corino E., 2022a, *Come si imparano le lingue: la linguistica acquisizionale*. In Meluzzi C., Nese N. (a cura di), 2022, *Metodi e prospettive della ricerca linguistica*. Milano, Ledizioni, pp. 169-187.

Corino E., 2022b, *Corpora di italiano L2: tra ricerca e didattica*. In Borghetti C., Tamburini F., 2022, *CLUB Working Papers in Linguistics*, Vol. 6, pp. 7-26.

Corino E., 2017, *Learner corpora e corpora di apprendenti di italiano. Confronti*. In Corino E., Marellò C., 2017, *Italiano di stranieri. I corpora VALICO e VINCA*, Perugia, Guerra, pp. 31-64.

Corino E., 2014, *Didattica delle lingue corpus-based*, «EL. LE Educazione Linguistica – Language Education», 3 (2), pp. 231-257.

Corino E., 2012, *Italiano di tedeschi. Una ricerca corpus-based*, Guerra, Perugia.

Corino E., Marellò C., 2009, *Elicitare scritti a partire da storie diseguate: il corpus di apprendenti VALICO*. In Andorno C., Rastelli S. (a cura di), 2009, *Corpora di italiano L2: tecnologie, metodi, spunti teorici*, Perugia, Guerra, pp.113-138.

Flowerdew L., 2015, *Data-Driven Learning and Language Learning Theories: Whither the Twain Shall Meet*. In Boulton A., Leńko-Szymańska A., 2015, *Multiple Affordances of Language Corpora for Data-Driven Learning*, Amsterdam/Philadelphia, John Benjamins Publishing Company, pp. 15-36.

Forti L., 2019, *Evaluating Data-driven Learning Effects in the Italian L2 Classroom: Elic and Elic perspectives combined*, «EL.LE. Educazione linguistica – Language education», 8 (2), Venezia: Ca' Foscari Digital Publishing, pp. 363-378.

Forti L., Spina S., 2019, *Corpora for linguists vs. corpora for learners. Bridging the gap in Italian L2 learning and teaching*, EL.LE. Educazione linguistica – Language education, 8(2), pp. 349-361.

Giacalone Ramat A., 2003, *Verso l'italiano*, Roma, Carocci.

Granger S., 1996, *From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora*. In Aijmer K., Altenberg B, Johansson M. (eds.), 1996, *Languages in Contrast. Text-based cross-linguistic studies*, Lund, Lund University Press, pp. 37-51.

Granger S., Paquot M., 2015, *Electronic lexicography goes local: Design and structures of a needs-driven online academic writing aid*, «Lexicographica - International Annual for Lexicography», 31 (1), pp. 118-141.

Hafner C., Candlin C., 2007, *Corpus Tools as an Affordance to Learning in Professional Legal Education*, «Journal of English for Academic Purposes», 6 (4), pp. 303-18.

Marello C., 2012, *Corpora di apprendenti. Come usarli nella didattica dell'italiano in Svizzera*. In Di Pretoro P.A., Unfer Lukoschik R. (a cura di), *Lingua e letteratura italiana 150 anno dopo l'Unità*, Monaco, Meidenbauer, pp. 299-315.

Marello C., 2009, *Distrattori tratti da corpora di apprendenti di italiano LS/L2*. In Corino E., Marello C. (a cura di), 2009, *VALICO. Studi di linguistica e didattica*, Perugia, Guerra, pp.179-193.

Marello C., Corino E., 2011, *Der Fehler als Distraktor. Multiple-choice Verfahren und metalinguistische Überle-*

ungen im DaF-Unterricht. In Bosco Coletos S., Costa M., Eichinger L. (a cura di), *Atti del convegno Tedesco-Italiano: confronti linguistici/Deutsch Italienisch: Sprachvergleiche*, Heidelberg, Winter, pp. 189-207.

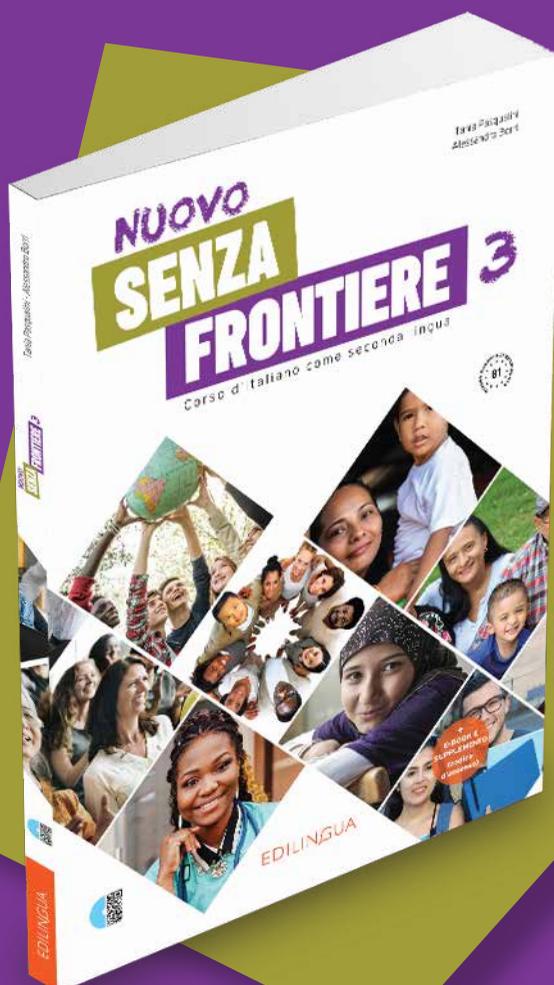
McCarten J., 2010, *Corpus-informed course book design*. In O'Keeffe A., McCarthy M. (eds.), *The Routledge handbook of corpus linguistics*, London, Routledge, pp. 413-427.

Paquot M., 2012, *The LEAD dictionary-cum-writing aid: An integrated dictionary and corpus tool*. In Granger S., Paquot M. (eds.), *Electronic lexicography*, Oxford, Oxford University Press, pp. 163-185.

Pienemann M., 1998, *Language Processing and Second Language Development. Processability Theory*, Amsterdam, Benjamin.

Spina S., 2010, *The dictionary of Italian collocations: Design and integration in an online learning environment*. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 10)*, pp. 3202-3208.

Spinelli B., Parizzi F., 2010, *Profilo della lingua italiana. Livelli di riferimento del QCER A1, A2, B1, B2*, Firenze, La Nuova Italia.



Nuovo Senza frontiere 3 si rivolge ad apprendenti immigrati e immigrate che vogliono raggiungere il livello B1 del Quadro Comune Europeo e sostenere il Test di lingua italiana B1 per cittadinanza.

- ✓ Libro e Quaderno in un unico volume
- ✓ 3 moduli di 4 Unità didattiche ciascuno, a loro volta suddivise in Lezione A e Lezione B
- ✓ Sviluppo dei domini: personale, professionale, educativo e pubblico
- ✓ QR code per accedere alle tracce audio
- ✓ autovalutazione e test di verifica finale
- ✓ video e attività create ad hoc
- ✓ Sviluppo della competenza sociolinguistica
- ✓ e-book sulla piattaforma i-d-e-e.it
- ✓ Guida per l'insegnante

