

Data-driven Learning: tra lingue straniere e CLIL, tra ricerca e didattica

Elisa Corino

Università degli Studi di Torino, Italia

Sommario 1 Linguistica dei corpora e didattica: tra passato, presente e futuro. – 1.1 Il Data-driven Learning (DDL): tra teorie e pratiche. – 1.2 Il DDL tra punti di forza e debolezza. – 2 Il DDL tra L1, LS e CLIL. – 3 I contributi del volume.

1 Linguistica dei corpora e didattica: tra passato, presente e futuro

La linguistica dei corpora in Italia ha ormai una tradizione consolidata che va dalle prime schede perforate di Padre Busa alle più recenti esperienze di parsing (Bosco et al. 2019 tra gli altri) e corpora del parlato (Cerruti et al. 2019) passando per i numerosi *learner corpora* di italiano LS/L2 (Corino, Marello 2017; Abel 2014; Spina 2014).

Tuttavia, alla vivacità della ricerca corpus-based e al ruolo essenziale che i corpora hanno ormai assunto nell'analisi della lingua non corrisponde un'altrettanta diffusione in ambito glottodidattico: come afferma Boulton (2020, XV), «DDL [Data-driven Learning] work in applied linguistics remains relatively unknown in educational circles», sebbene gli usi e i benefici dell'uso dei corpora in ambito educativo siano da tempo ampiamente trattati dalla ricerca e discussi nella letteratura di riferimento (tra i molti studi si veda ad esempio O'Keeffe, McCarthy, Carter 2007; Boulton, Leńko-Szymańska 2015; Crosthwaite 2020).

Attualmente, i corpora sono risorse con le quali gli studenti entrano per lo più in contatto indiretto, attraverso la mediazione di opere

di riferimento e materiale didattico e, nonostante i corpora siano alla base della costruzione di tali prodotti, l'utente finale non deve nemmeno esserne consapevole: dai dizionari per apprendenti (un esempio tra i tanti, il *Macmillan English Dictionary for Advanced Learners*, Rundell 2007), alle grammatiche (cf. la *Longman Grammar of Spoken and Written English* - Biber et al. 1999; o la *Cambridge Grammar of English* - Carter, McCarthy 2006), dai manuali scolastici (ad es. la serie *Touchstone* - McCarthy, McCarten, Sandiford 2005) ai molti materiali integrativi (come la serie *English Vocabulary in Use series* - McCarthy, O'Dell 2005).

Gli usi diretti dei corpora nell'insegnamento delle lingue sono trattati piuttosto marginalmente nella letteratura; questo riflette tristemente la realtà della pratica in classe, in particolare al di fuori del livello di istruzione superiore, in cui tecniche esplicite basate sui dati sono raramente incorporate nelle normali procedure di insegnamento (Mukherjee 2004; Römer 2009; 2010; Tribble 2015).

Nella prefazione a uno degli ultimi volumi usciti sul *Data-driven Learning* (Crosthwaite 2020), Alex Boulton (2020) cita ben 378 pubblicazioni recenti che descrivono sperimentazioni e studi sul DDL e tutte giungono a conclusioni positive sull'impatto che l'uso dei corpora può avere sull'apprendimento linguistico. Eppure, al di fuori di un certo contesto anglofono, con riferimento particolare a EFL, ESL e EAP, gli strumenti e le tecniche della linguistica dei corpora in didattica sono del tutto accessori, se non completamente ignorati.

Le conclusioni di Boulton e Cobb (2017, 386), per i quali «DDL works pretty well in almost any context where it has been extensively tried», non sembrano di fatto trovare riscontro tra le buone pratiche glottodidattiche oggi in uso nella scuola; al contrario, la maggioranza delle sperimentazioni, anche a livello internazionale, sono limitate all'ambito universitario.

Tra le cause vi sono certamente l'assenza di una formazione sistematica che renda evidente come i risultati dell'analisi dei corpora possano essere resi pedagogicamente efficaci e rilevanti, la diffidenza del corpo docente verso delle pratiche che appaiono complesse sia per la gestione degli strumenti di ricerca sia per l'alta richiesta metalinguistica e - più in generale - cognitiva che esse richiedono, il tempo da dedicare alla pianificazione delle attività e dei materiali e all'introduzione degli studenti al DDL stesso (Tribble 2015; Corino, Onesti 2019).

La rivoluzione dei corpora predetta da Rundell e Stock (1992) dunque deve ancora raggiungere gli apprendenti.

Nel 2007, Braun affermava che ci sarebbe voluta una nuova generazione di insegnanti «for corpora to find their way into the language classroom» (Braun 2007, 308). A più di un decennio di distanza, tuttavia, il *Data-driven Learning* è ancora una pratica limitata a contesti specifici, sia per il grado scolastico in cui esso viene proposto sia - e soprattutto - per le lingue coinvolte.

A fronte di tali considerazioni si impone quindi la necessità di strutturare programmi di formazione che mirino a sviluppare competenze d'uso e pianificazione da parte di tutti i soggetti coinvolti nel processo di insegnamento/apprendimento, a partire dagli insegnanti per arrivare almeno fino alle classi della scuola secondaria.

1.1 Il Data-driven Learning (DDL): tra teorie e pratiche

Come notano Boulton e Leńko-Szymańska (2015), nella discussione sulle applicazioni dei corpora all'insegnamento delle lingue ci imbatiamo spesso nella nozione di *affordance* (cf. Hafner, Candlin 2007; Boulton 2012), ovvero la capacità di un oggetto - linguistico nel nostro caso - di suggerire all'utente come interagire con esso. E di fatto è questo il ruolo della didattica data-driven: offrire a studenti e docenti degli indizi sui fatti di lingua che portino induttivamente chi li indaga a trarre delle indicazioni su strutture, usi, funzioni, contesti che non sono prototipicamente contenuti nei libri di testo in uso nella scuola.

Gli usi diretti dei corpora coinvolgono in prima persona i discenti che sfruttano gli strumenti della linguistica computazionale per imparare ad eseguire analisi linguistiche con l'obiettivo di decodificare le strutture della lingua. Celebre è il suggerimento di Tim Johns, uno dei pionieri più influenti dell'approccio e padre della sua definizione come *Data-driven Learning*, di «cut[ting] out the middleman as far as possible and [...] giv[ing] the learner direct access to the data» (Johns 1991, 30) seguendo il motto «Every student a Sherlock Holmes» (1997, 101).

Attraverso il corpus, gli studenti applicano una metodologia di ricerca simile a quella che segue il linguista nelle sue indagini e mettono in atto processi cognitivi alti, che coinvolgono conoscenze linguistiche e metalinguistiche da una parte e competenze trasversali di classificazione, analisi e *problem solving* dall'altra. In primo luogo si tratta di definire i contenuti di una *query* in relazione alle informazioni alle quali gli studenti sono interessati e che risultano pertinenti per loro in un determinato momento. L'atto stesso di formulare un'interrogazione appropriata e dare senso ai suoi risultati contribuisce allo sviluppo non solo della consapevolezza linguistica, ma anche delle capacità di apprendimento generale e di autonomia che sono tra gli obiettivi primari della didattica contemporanea.

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 socio-culturale di Vigotsky, con un'attenzione particolare verso gli stili di apprendimento e i processi cognitivi che lo sottendono.

Da punto di vista glottodidattico i recenti sviluppi nell'ambito dell'approccio comunicativo, unitamente alle ricerche in ambito di

Second Language Acquisition (SLA), hanno dimostrato che il *focus on form* - soprattutto se la forma è contestualizzata e associabile a un significato e a un uso - è indispensabile nell'apprendimento delle lingue, in particolare quando gli studenti debbano raggiungere livelli alti di competenza (Doughty, Williams 1998; Norris, Ortega 2000). Certamente il processo di *noticing* non può sempre essere demandato alle capacità di osservazione dello studente ed è ormai un assunto condiviso che talvolta l'insegnamento esplicito di una struttura sia più efficace di un approccio induttivo, soprattutto laddove gli elementi potrebbero essere poco salienti per l'apprendente o poco evidenti (si pensi ad esempio alla riflessione sulla L1). Il ricorso ad un corpus, tuttavia, può rivelarsi utile per presentare un argomento in relazione alla sua frequenza, rendendone immediatamente evidenti le caratteristiche e le funzioni all'interno di un dato sistema linguistico. L'insegnante in questo caso 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.

Le alte domande cognitive del DDL possono essere risolte utilizzando i dati del corpus in modo più proattivo; ciò significa che essi possono essere utilizzati proficuamente anche in un approccio deduttivo in base al quale le occorrenze non vengono impiegate per elicitarne una regola o un modello ma per esemplificarlo. O ancora, secondo quanto proposto da Molés-Cases e Oster (2015), gli studenti possono lavorare su compiti corpus-based altamente strutturati in modo collaborativo, avendo così la possibilità di condividere ipotesi e inferenze e contemporaneamente di acquisire maggiore fiducia attraverso il lavoro di gruppo.

In questo contesto, il *noticing* dunque può essere un processo messo in atto spontaneamente dallo studente che si trova ad analizzare dei materiali, così come originariamente suggerito da Johns (1991) e dalla corrente sinclairiana della linguistica computazionale, che sostiene la ricerca corpus-driven, oppure può essere elicitato e diretto dal docente mediatore di conoscenza e mediatore tra lo studente e i dati. Si tratta in questo caso di un procedimento corpus-based di «induzione guidata» sul modello di quanto proposto da Johansson (2009) per cui l'approccio induttivo e quello deduttivo non sono visti come poli opposti, ma si integrano in base ai bisogni formativi degli studenti e al loro profilo.

Questo processo si inserisce appieno nelle specifiche di quella didattica di stampo costruttivista che vede l'acquisizione di conoscenza come un processo dinamico di cui il discente è protagonista: le attività DDL di estrazione, analisi, classificazione e attribuzione di senso ai risultati attivano processi cognitivi di livello alto (HOTS) che stimolano l'acquisizione. D'altra parte l'interazione con il docente e il suo ruolo da 'regista' del processo di apprendimento permettono quelle azioni di *scaffolding* e di co-costruzione di conoscenza attraverso il confronto promosse dalla teoria socioculturale vigotskiana.

Un'obiezione certamente legittima avanzata da molti (cf. ad esempio Kirschner, Sweller, Clark 2006) è che questo tipo di approccio non è adatto a tutti i tipi di apprendente. La relazione tra DDL, costruttivismo e stili di apprendimento è ancora in fase esplorativa e la ricerca presenta considerazioni controverse e risultati che andranno approfonditi nel prossimo futuro con indagini su ampia scala. In relazione agli stili cognitivi, ad esempio, Flowerdew (2008) sostiene che apprendenti dipendenti dal campo traggono maggiori benefici dalla pedagogia corpus-based, sebbene uno studio di Turnbull e Burston (1998, in Boulton 2009) provi come lo stesso tipo di apprendenti fossero estremamente critici verso l'approccio e i risultati di Chan e Liou (2005) riportino solo minime differenze tra le reazioni di studenti con una propensione verso la didattica induttiva e i loro pari più inclini all'approccio deduttivo.

Certamente il DDL, grazie anche all'uso di strumenti con interfacce *user friendly* e di una dimensione grafica sempre più curata e capace di far risaltare gli elementi salienti (Corino, Onesti 2019), presenta una stretta correlazione con lo stile di apprendimento visivo, e la manipolazione dei dati da parte dell'apprendente in prima persona favorisce sia chi ha inclinazioni di tipo procedurale e sequenziale sia i cinesztetici che sono chiamati a ordinare e classificare concretamente i dati.

1.2 Il DDL tra punti di forza e debolezza

Se le ricadute positive del DDL in termini linguistico-pedagogici sono ormai dimostrate da numerosi studi empirici e nuovi progetti che testimoniano la vitalità dell'approccio vedono la luce sempre più frequentemente, è d'altra parte vero che raramente la ricerca e la pratica accademica incontrano il mondo della scuola. Il recente volume curato da Crosthwaite (2020) dedicato ai «young learners» è certamente un tentativo di dimostrare come l'integrazione del DDL all'interno delle buone pratiche più comuni sia possibile anche a livelli che prescindono dal contesto accademico. Ci sembra però legittimo chiederci se le riflessioni degli autori che hanno contribuito all'opera raggiungeranno mai gli insegnanti impegnati sul campo e, nel fortuito caso che lo facessero, se questi ultimi integreranno effettivamente i suggerimenti alle modalità di insegnamento tradizionali.

Il primo convegno italiano sul DDL tenutosi a Torino nel 2018 (*DDL: un metodo a supporto delle lingue straniere e delle classi CLIL*) si proponeva di diffondere buone pratiche didattiche corpus-based tra gli insegnanti di lingue straniere e i disciplinari impegnati nel CLIL; nonostante l'interesse dimostrato dal folto pubblico presente, le proposte presentate non hanno avuto seguito in progetti concreti di didattica in aula.

Tra le criticità del DDL individuate dagli insegnanti c'è sicuramente l'aspetto tecnologico, come già osservato da molti studi nel campo (Mukherjee 2004; Römer 2009; Tribble 2015, Leńko-Szymańska 2017): si tratta da una parte di una obiezione legittima ma meramente logistica, legata alle attrezzature disponibili nelle scuole, dall'altra - questione decisamente più spinosa - di una mancanza di formazione specifica dei docenti che non solo non hanno familiarità con la linguistica dei corpora e le ICT (Internet and Communication Technologies) ad essa legate, ma, soprattutto, mancano delle competenze specifiche su come didattizzare le risorse a loro disposizione (si vedano ad esempio alcuni dei risultati riportati in Scheffer-Lacroix 2020).

Come nota Chambers (2019), la maggioranza degli studi empirici coinvolgono insegnanti-ricercatori con un livello elevato di abilità nella creazione e nella consultazione dei corpora, figure pronte a dedicare molto tempo e sforzi a un'attività che produrrà esperienze di apprendimento per gli studenti e allo stesso tempo costituirà una valida attività di ricerca per loro. E, data la specificità del contesto, la replicabilità dei percorsi sarà limitata a situazioni analoghe, poiché difficilmente esportabili nella stessa forma ad altri livelli di scuola.

Un'adeguata formazione dei docenti sugli strumenti e su come usarli concretamente in classe è quindi cruciale per colmare il divario tra ricerca e pratica. Römer (2006) descrive l'alfabetizzazione in DDL con la metafora religiosa della «missione», un'opera di «diffusione della parola» che deve partire dalla formazione dei futuri insegnanti fin dai banchi dell'università. In questo modo i (futuri) docenti acquisiranno competenze nelle tre aree indispensabili all'applicazione delle strategie corpus-based (Leńko-Szymańska 2017): tecnologia, metodo di indagine legato alla linguistica dei corpora, didattica e pedagogia del DDL. Alla fine del percorso conosceranno quindi gli strumenti disponibili, saranno in grado di selezionarli a seconda delle necessità e usarli per estrarre informazioni, analizzarle e interpretarle, e avranno sviluppato le competenze necessarie a implementare sessioni DDL in un quadro metodologico più ampio.

Un'ulteriore riserva alla diffusione del DDL nell'insegnamento è legata alla disponibilità di corpora pedagogici che soddisfino i bisogni dei discenti, più che rispondere alle esigenze dei ricercatori.

La varietà di strumenti disponibile oggi, tuttavia, permette di poter selezionare ciò che più si presta agli scopi prefissati e molti sono i progetti che si rivolgono ad un ampio pubblico di non-esperti e alcuni sono realizzati appositamente per gli apprendenti. Tra questi: Lextutor, la suite di corpora corredata di attività didattiche elaborate da Tom Cobb; SkELL, la piattaforma con corpora multilingui estratti da Sketch Engine e creata proprio per il «Language Learning»; SACODEYL, i corpora di interviste di adolescenti descritto come un «pedagogical mediator in the language learning process»; FLAX, che con-

tiene preziosi riferimenti anche per le lingue specialistiche (LSP) e il CLIL; o ancora software di concordanze come AntConc...

Il fatto che molte delle risorse citate siano anche multilingui è indice di una certa determinazione a superare l'egemonia che la lingua inglese ha avuto fino ad ora in questo campo. Certamente la preponderanza dell'inglese è dovuta alla sua influenza come lingua di comunicazione globale e, di conseguenza, al suo status di lingua più insegnata/imparata, alla fortuna della *corpus linguistics* - e dei *learner corpora* - in contesti legati a tale lingua, al lavoro di ricerca e definizione teorica da parte di figure di riferimento che si inseriscono nel panorama accademico anglosassone. Si sente però il bisogno di superare il monolinguisimo di strumenti e letteratura, per estendere il DDL a una effettiva dimensione metodologica che può essere definita tale solo se abbraccia la glottodidattica a tutto tondo, indipendentemente da quale lingua sia oggetto dell'apprendimento/insegnamento.

2 Il DDL tra L1, LS e CLIL

Quando si discute di *Data-driven Language Learning* è quasi un assunto dato per scontato che ci si riferisca alla lingua straniera. Se è vero che il DDL è nato essenzialmente per essere inserito nell'ambito della classe di lingue, è altrettanto vero che l'esplorazione di corpora può supportare efficacemente la riflessione sulle lingue specialistiche (LSP), anche quando si tratta di codici e sottocodici della L1.

Gli studenti delle scuole secondarie sono spesso costretti a confrontarsi con i linguaggi delle discipline senza però avere adeguati supporti linguistici che consentano loro di operare le necessarie differenziazioni e categorizzazioni della LSP rispetto alle varietà più generali.

Solo nel CLIL questo aspetto è tenuto in considerazione e si tende a dimenticare che l'educazione linguistica è compito che spetta al *curriculum* formativo nella sua globalità, per cui gli stessi insegnanti di fisica, matematica, storia dell'arte, diritto... sono di fatto chiamati ad agire anche sul piano linguistico sulla L1 dei loro studenti.

La dimensione disciplinare, quella morfosintattica e quella lessicale non possono essere considerate separatamente, perché sviluppare competenze di LSP - siano esse in L1 o in LS - significa essere in grado di padroneggiare diversi aspetti linguistici, tra cui caratteristiche lessicogrammaticali, modelli di testualizzazione e caratteristiche di strutturazione del genere, oltre l'adeguatezza e l'accettabilità di alcuni elementi in relazione al contesto.

L'uso dei corpora e del DDL può supportare gli insegnanti non linguisti nell'individuare modelli di fraseologia specializzata, lessico e parole chiave la cui specificità può sfuggire all'occhio del disciplinarista e che non necessariamente sono menzionate nei dizionari bililingui e monolingui usati dai loro studenti (Corino 2014).

L'argomento principale per l'utilizzo di DDL nei contesti disciplinari e CLIL è che la lingua è la chiave di accesso al contenuto, lo codifica e lo rende accessibile. Ma, come sottolinea Snow (2010), il linguaggio della scienza – o più in generale il linguaggio disciplinare – è «alienante», sfocato, se non addirittura oscuro, tanto da trasformare spesso anche la stessa L1 in una lingua straniera. E la comprensione diventa ancora più difficile quando si tratta di una LSP 'alienante' che pertiene a un codice linguistico diverso e che implica necessariamente lo sviluppo di un'educazione specifica che abbia come obiettivo lo sviluppo della consapevolezza della relazione tra contenuto e lingua.

In questo senso, un approccio induttivo corpus-based e *bottom up* che permette il focus su porzioni di lessico e grammatica in contesto può favorire la competenza LSP sia degli insegnanti di disciplina che degli studenti, offrendo la possibilità di rilevare dati sull'uso della lingua che potrebbero altrimenti passare inosservati: la scelta tipica delle parole (ordinandole per frequenza), che significa sfumature di significato e uso appropriato delle collocazioni, la definizione di preferenze semantiche nella selezione di modificatori, l'individuazione di pattern morfosintattici.

E, come già osservato, lo sforzo cognitivo di decodifica e *problem solving* legato alla relazione tra la dimensione paradigmatica verticale e quella sintagmatica orizzontale attiva processi HOTS che probabilmente si tradurranno in una conoscenza più duratura e migliori competenze linguistiche.

E questo è tanto vero per i contesti CLIL in cui la lingua di comunicazione è la LS, quanto per l'educazione linguistica e disciplinare il L1.

Negli ultimi anni abbiamo assistito alla pubblicazione di alcuni testi importanti in contesto anglofono che fungono da guida all'uso delle pratiche DDL e che propongono attività, esercizi di osservazione e definiscono percorsi di esplorazione di corpora accessibili (si vedano ad esempio Thomas 2015; Poole 2018; Friginal 2018). Tuttavia si tratta di suggerimenti che riguardano usi e funzioni di una lingua comune, mentre poca attenzione è data alle buone pratiche in quei contesti specialistici per i quali gli insegnanti faticano a trovare strumenti di supporto linguistico pronti all'uso.

Gli esempi riportati in Corino (2014) sul CLIL in fisica e in Corino, Onesti (2019) sulla LSP nell'educazione professionale dei parrucchieri mettono in luce come attività DDL possono essere usate in modo efficace in contesti diversi legati ai linguaggi specialistici, ma si tratta pur sempre di esperienze di ricerca-azione legate a collaborazioni e progetti limitati nel tempo e legati unicamente alla LS.

Alcune recenti esperienze (Corino c.d.s.; Corino et al. c.d.s.) hanno provato che le pratiche di *noticing* corpus-based possono efficacemente essere estese anche all'educazione linguistica in L1. Gli studenti spesso non hanno una percezione definita di come funziona la propria lingua al di là delle nozioni di grammatica descrittiva e pre-

scrittiva apprese nelle ore di italiano, sono spesso inconsapevoli delle relazioni tra le parole e della loro dipendenza dal co-testo, soprattutto in contesti specialistici. D'altra parte i docenti di discipline non linguistiche si concentrano sul contenuto, talvolta ignari delle proprietà della lingua che parlano ai loro discenti, dando per scontato che questi ultimi siano in grado di comprendere e decodificare il sottocodice e che la LSP 'fluisca' senza bisogno di spiegazioni e approfondimenti linguistici.

Anche in testi disciplinari apparentemente semplici, in cui la maggior parte del lessico fa parte del vocabolario di base, che presenta una struttura sintattica affatto marcata e un'articolazione testuale-informativa lineare, permangono degli elementi linguistici che celano delle difficoltà legate ad aspetti collocazionali, alla polisemia, all'uso di tecnicismi collaterali e rideterminazioni semantiche, ecc.

È ad esempio il caso della Costituzione italiana, testo oggi al centro del dibattito educativo, comunemente descritta come un documento pienamente intellegibile (cf. De Mauro 2008; Korzen 2010; La Fauci 2011), ma che a ben guardare cela insidie legate alla lingua specifica del diritto: una *norma* (lemma che compare 41 volte nel testo) può essere *dettata, applicata, osservata, stabilita, emanata* a seconda del contesto e se si considera la reggenza preposizionale abbiamo sintagmi quali *norme di procedura, norme sull'ordinamento giudiziario* o *norme dell'ordinamento vigente*. O ancora *Amministrazione* (art 97) ha contemporaneamente l'accezione specialistica di *ufficio* e quella comune e generica, che pertiene invece al lessico fondamentale, di *gestione*.

Ci si chiede dunque quanto gli insegnanti di diritto siano consapevoli delle possibilità combinatorie offerte dal sistema e quanto i loro studenti siano in grado di sfruttarle in modo attivo e produttivo.

Nell'ambito di un Percorso per le competenze trasversali e per l'orientamento (PCTO) è stato dunque proposto ad alcune classi di istituti tecnici superiori un'esperienza di analisi e ricerca corpus-based per 'investigare' le caratteristiche dei testi specialistici, a partire da quelli di diritto per poi, una volta acquisita la metodologia di ricerca, estendere le buone pratiche anche ad altri ambiti (cf. Corino et al. c.d.s.).

Con l'aiuto del *concordancer* AntConc, gli studenti poco abituati all'osservazione e all'analisi linguistica hanno così descritto la lingua dei bugiardini, scoprendo le frequenze relative di *medicinale* (867), *farmaco* (76) e *medicina* (0) in un corpus da loro creato, e, riflettendo sul significato di tale distribuzione, hanno identificato i verbi che occorrono con più frequenza con i termini chiave (come prevedibile il *prendere un medicinale*, e il meno comune *assumere un medicinale*) e alcune strutture tipiche del tipo testuale [fig. 1].

o al fegato non deve assumere questo medicinale (vedere "Non prenda BRUFEN"). Anziani S	Brufen_Italia
uò verificare brevemente dopo aver assunto questo medicinale. I medicinali come FROBEN DOLORE E FE	Froben_Italia
se il suo bambino ha assunto questo medicinale per errore, contatti sempre un medico o	Actibu_Italia
se il suo bambino ha assunto questo medicinale per errore, contatti sempre un medico o	Actibu_Italia

Figura 1 Ricerca su AntConc_bugiardini: medicinale

Inoltre hanno riflettuto sulla polisemia dei termini e sulle possibili sfumature di significato, riempiendo di nuovi tratti semantici significanti fino ad ora usati con una sola accezione o in un solo contesto: non solo *assumere*, ma anche *manifestare (sintomi)* o *orale (per via orale, uso orale, sospensione orale)*. Hanno insomma definito le caratteristiche della LSP tracciando in modo chiaro e finalmente consapevole il confine tra tecnicismo e lingua comune.

Le ricadute sulla competenza linguistica degli studenti sono state evidenti a tutti gli insegnati, tanto che la modalità corpus-based (e in parte corpus-driven) è stata estesa anche nella classe di lingua straniera, attraverso un lavoro su testi paralleli di confronto e traduzione.

Non vanno infatti dimenticate le applicazioni del DDL nella didattica della traduzione dove i corpora vengono efficacemente usati come strumento complementare ai dizionari e le concordanze parallele aiutano gli studenti a definire i contesti d'uso dei traduttori.

3 I contributi del volume

Questo volume monografico si propone quindi di descrivere, per la prima volta nel panorama italiano, gli usi e le funzioni che i corpora e il DDL possono fattivamente avere nella pratica didattica. Gli autori sono sia accademici attivi nella ricerca sulla didattica corpus-based sia insegnanti di scuola secondaria impegnati in esperienze di DDL nell'ambito delle LS o del CLIL. Obiettivo della pubblicazione è stimolare e spingere i docenti di lingue e di discipline non linguistiche a sperimentare questo approccio, integrandolo tra le buone pratiche già messe in atto nelle loro classi.

Ecco perché il contributo di apertura è proprio a cura di un'insegnante della scuola secondaria che si occupa di CLIL: **Emma Abbate** presenta una serie di attività DDL sperimentate nelle proprie classi di geografia con studenti di livello B1 e fornisce preziosi suggerimenti su come usare i corpora per lo *scaffolding* nella comprensione e produzione di testi, con particolare riferimento agli strumenti offerti dalla piattaforma FLAX e dal suo software di concordanze.

Di sperimentazione di metodologia e strumenti tratta anche **Sonia Di Vito**, che dimostra come la ricerca e la didattica corpus-based non siano unicamente legate alla lingua inglese, ma possano es-

sere estese anche a contesti di *Français Langue Étrangère* (FLE) e a livelli di competenza basica e postbasica quale quello degli studenti della scuola secondaria di primo grado. Attraverso l'osservazione dei dati di lingua autentica e l'individuazione delle sue regolarità grammaticali e morfologiche, le classi coinvolte nel progetto *Apprentissage sur corpus* hanno acquisito maggiore consapevolezza metalinguistica e autonomia nello studio. Di Vito rileva inoltre come il DDL possa essere considerato come un approccio inclusivo in classe che permette a tutti di sviluppare un atteggiamento di sicurezza nelle proprie capacità e uno stile di apprendimento consapevole.

Il contributo dei corpora nell'insegnamento/apprendimento delle LSP è esemplificato da **Carolina Flinz** e **Peggy Katelhoe**n che si occupano di lingua del turismo per *il Deutsch als Fremdsprache* (DaF). Le autrici riflettono sull'importanza della scelta del corpus di riferimento e offrono una panoramica sui corpora di tedesco che possono essere usati nell'insegnamento DaF in generale e della LSP del turismo in particolare, corredando le proprie considerazioni di esempi e attività pratiche da usare in classe.

Non manca nel volume l'attenzione verso l'italiano L2/L2. In particolare, **Luciana Forti** e **Stefania Spina** esplorano lo stato dell'arte del DDL e dei corpora *di* e *per* apprendenti in Italia, definendo la distanza che ancora esiste tra i corpora nati per ricerca accademica e quelli creati con intento pedagogico. Il divario tra le due prospettive può certamente essere colmato integrando i risultati della ricerca sui *learner corpora* con l'approccio DDL, di cui vengono dati alcuni esempi.

Nel suo contributo, poi, **Luciana Forti** presenta gli effetti del Data-driven Learning sullo sviluppo della competenza fraseologica nell'italiano L2 di studenti sinofoni, adottando una duplice chiave di lettura: la misurazione empirica dei dati e della progressione nelle competenze degli apprendenti e la percezione circa l'efficacia delle attività proposte. Si ha così una visione a tutto tondo della metodologia, che include anche aspetti motivazionali ed emotivi che contribuiscono in modo significativo all'efficacia delle pratiche didattiche.

Cristina Caimotto affronta l'uso dei corpora nella didattica della traduzione: osservare e confrontare varie traduzioni dello stesso testo sorgente è un'attività che può migliorare le abilità linguistiche degli studenti e la loro consapevolezza della pratica di trasferire un testo da un codice ad un altro. Focus della proposta operativa sono i discorsi inaugurali tenuti dai Presidenti degli Stati Uniti d'America (raccolti nel corpus POTUS), spesso tradotti da più agenzie di stampa e testate giornalistiche, qui usati con lo scopo di rendere gli studenti consapevoli della pratica in gran parte invisibile e ancora sottostimata della traduzione, delle sue strategie e delle sue caratteristiche, attraverso un'analisi contrastiva che abbraccia anche il tema della distribuzione globale delle notizie.

L'uso dei dati linguistici in didattica si estende anche settori che esulano dalla competenza sul significante e sul significato, ma che investono tratti soprasegmentali. **Valentina De Iacovo** e **Antonio Romano** mostrano come il Computer-Assisted Language Learning (CALL) e l'approccio data-driven possa essere integrato con procedimenti tipici della ricerca fonetica: gli spettrogrammi che descrivono curve intonative di nativi diventano quindi strumenti per definire le competenze prosodico-intonative degli studenti di LS e possono venire usati per migliorare le loro abilità orali.

Infine **Milica Vitaz** e **Milica Poletanović** confrontano lo stato dell'arte e la diffusione del DDL con la situazione serba, dove l'uso dei corpora nella didattica non è solo limitato ma addirittura inesistente. Sebbene circoscritto alla situazione locale, lo studio di caso apre riflessioni generalizzabili sul punto di vista degli insegnanti e sulla diffusione dell'uso dei corpora in didattica al di fuori del contesto universitario.

La raccolta si chiude con il contributo di **Fanny Meunier** che idealmente conclude con riflessioni sul quadro teorico di riferimento, sui bisogni ai quali va ancora data una risposta e sulle possibili azioni da intraprendere per diffondere il DDL e rendere le attività corpora-based più accessibili a insegnanti e studenti.

Bibliografia

- Abel, A. (2014). «A Trilingual Learner Corpus Illustrating European Reference Levels». *RiCognizioni*, 1(2), 111-26. <http://dx.doi.org/10.13135/2384-8987/702>.
- Biber, D.; Johansson, S.; Leech, G.; Conrad, S.; Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Bosco, C.; Di Nuovo, E.; Mazzei, A.; Sanguinetti, M. (2019). «Towards an Italian Learner Treebank in Universal Dependencies». Bernardi, R.; Navigli, R.; Semeraro, G. (eds), *Proceedings of the Sixth Italian Conference on Computational Linguistics* (Bari, November 13-15, 2019). CEUR-WS. <http://ceur-ws.org/Vol-2481/paper28.pdf>.
- Boulton, A. (2009). «Corpora for All? Learning Styles and Data-Driven Learning». Mahlberg, M.; González-Díaz, V.; Smith, C. (eds), *Proceedings of 5th Corpus Linguistics Conference* (University of Liverpool, July 20-23, 2009). Lancaster: UCREL. <http://ucrel.lancs.ac.uk/publications/cl2009/>.
- Boulton, A. (2012). «Beyond Concordancing: Multiple Affordances of Corpora in University Language Degrees». *Procedia: Social and Behavioral Sciences*, 34, 33-8. <https://doi.org/10.1016/j.sbspro.2012.02.008>.
- Boulton, A. (2020). «Data-Driven Learning for Younger Learners: Obstacles and Optimism». Crosthwaite 2020, XIV-XX.
- Boulton, A.; Cobb, T. (2017). «Corpus Use in Language Learning: A Meta-Analysis». *Language Learning*, 67(2), 348-93. <https://doi.org/10.1111/lang.12224>.

- Boulton, A.; Leńko-Szymańska, A. (eds) (2015). *Multiple Affordances of Language Corpora for Data-Driven Learning*. Amsterdam; Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/sc1.69>.
- Braun, S. (2007). «Integrating Corpus Work into Secondary Education: from Data-Driven Learning to Needs-Driven Corpora». *ReCALL*, 19(3), 307-28. <https://doi.org/10.1017/S0958344007000535>.
- Carter, R.; McCarthy, M. (2006). *Cambridge Grammar of English*. Cambridge: Cambridge University Press.
- Cerruti, M.; Ballarè, S.; Mauri, C.; Gorla, E. (2019). «Il corpus KIParla. Tra linguistica dei corpora e sociolinguistica dell'italiano». *RiCognizioni*, 6(12), 275-8. <http://dx.doi.org/10.13135/2384-8987/3989>.
- Chambers, A. (2019). «Towards the Corpus Revolution? Bridging the Research-Practice Gap». *Language Teaching*, 52(4), 460-75. <https://doi.org/10.1017/S0261444819000089>.
- Chan, T.-p.; Liou, H.-C. (2005). «Effects of Web-Based Concordancing Instruction on EFL Students' Learning of Verb-Noun Collocations». *Computer Assisted Language Learning*, 18(3), 231-51. <https://doi.org/10.1080/09588220500185769>.
- Corino, E. (2014). «Bottom-Up Specialized Phraseology in CLIL Teaching Classes». *Workshop Proceedings of the 12th Edition of the KONVENS Conference*. Hildesheim: Universitätsverlag Hildesheim, 68-76.
- Corino, E. (c.d.s.). «Lingua e diritto: la Costituzione a scuola». Visconti, J. (a cura di), *Linguaggi settoriali e specialistici: sincronia, diacronia, traduzione, variazione = Atti del XV Congresso della Società internazionale di Linguistica e Filologia italiana*. Firenze: Cesati.
- Corino, E.; Longo, F.; Onesti, C.; Pallante, F.; Sobrino, G. (c.d.s.). *Leggere la Costituzione a scuola. Tra lingua e diritto*. Roma: Carocci.
- Corino, E.; Marello, C. (2017). *Italiano di stranieri. I corpora VALICO e VINCA*. Perugia: Guerra.
- Corino, E.; Onesti, C. (2019). «Data-Driven Learning: A Scaffolding Methodology for CLIL and LSP Teaching and Learning». *Frontiers in Education*. <https://doi.org/10.3389/educ.2019.00007>.
- Crosthwaite, P. (ed.) (2020). *Data-Driven Learning for the Next Generation*. New York: Routledge.
- De Mauro, T. (2008). *Il linguaggio della Costituzione*. Resoconto del convegno tenutosi a Palazzo Minerva il 16 giugno 2008, Senato della Repubblica. Convegni e seminari 18. https://www.senato.it/application/xmanager/projects/leg18/file/convegni_seminari_n18.pdf.
- Doughty, C.; Williams, J. (eds) (1998). *Focus on Form in Classroom Second Language Acquisition*. Cambridge: Cambridge University Press.
- Flowerdew, L. (2008). «Corpus Linguistics for Academic Literacies Mediated Through Discussion Activities». Belcher, D.; Hirvela, A. (eds), *The Oral/Literate Connection: Perspectives on L2 Speaking, Writing and Other Media Interactions*. Ann Arbor (MI): University of Michigan Press, 268-87.
- Flowerdew, L. (2015). «Data-Driven Learning and Language Learning Theories: Whither the Twain Shall Meet». Boulton, Leńko-Szymańska 2015, 15-36.
- Friginal, E. (2018). *Corpus Linguistics for English Teachers*. New York; London: Routledge.
- Hafner, C.; Candlin, C. (2007). «Corpus Tools as an Affordance to Learning in Professional Legal Education». *Journal of English for Academic Purposes*, 6(4), 303-18. <https://doi.org/10.1016/j.jeap.2007.09.005>.

- Johansson, S. (2009). «Some Thoughts on Corpora and Second-Language Acquisition». Aijmer, K. (ed.), *Corpora and Language Teaching*. Amsterdam: John Benjamins, 33-44.
- Johns, T. (1991). «From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-Driven Learning». Johns, T.; King, P. (eds), *Classroom Concordancing, English Language Research Journal*, 4, 27-45.
- Johns, T. (1997). «Contexts: The Background, Development and Trialling of a Concordance-Based CALL Program». Wichmann, A.; Fligelstone, S.; McEnery, T.; Knowles, G. (eds), *Teaching and Language Corpora*. Harlow: Addison Wesley Longman, 100-15.
- Kirschner, P.; Sweller, J.; Clark, R. (2006). «Why Minimal Guidance During Instruction Does not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching». *Educational Psychologist*, 41(2), 75-86. https://doi.org/10.1207/s15326985ep4102_1.
- Korzen, I. (2010). «Lingua, cognizione e due Costituzioni». Visconti, J. (a cura di), *Lingua e diritto. Livelli di analisi*. Milano: LED, 163-201.
- La Fauci, N. (2011). «La Costituzione della Repubblica italiana. Spunti per una riflessione linguistica». Nesi, A.; Morgana, S.; Maraschio, N. (a cura di), *Storia della lingua italiana e storia dell'Italia unita. L'italiano e lo stato nazionale = Atti del IX Convegno ASLI* (Firenze, 2-4 dicembre 2010). Firenze: Franco Cesati Editore, 383-94.
- Leńko-Szymańska, A. (2017). «Training Teachers in Data-Driven Learning: Tackling the Challenge». *Language Learning & Technology*, 21(3), 217-41.
- McCarthy, M.; McCarten, J.; Sandiford, H. (2005). *Touchstone*. Cambridge: Cambridge University Press.
- McCarthy, M.; O'Dell, F. (2005). *English Vocabulary in Use*. Cambridge: Cambridge University Press.
- Molés-Cases, T.; Oster, U. (2015). «Webquest in Translator Training: Introducing Corpus-Based Tasks». Boulton, Leńko-Szymańska 2015, 201-23.
- Mukherjee, J. (2004). «Bridging the Gap Between Applied Corpus Linguistics and the Reality of English Language Teaching in Germany». Connor, U.; Upton, T.A. (eds), *Applied Corpus Linguistics: A Multi-Dimensional Perspective*. Amsterdam: Rodopi, 239-50. https://doi.org/10.1163/9789004333772_014.
- Norris, J.M.; Ortega, L. (2000). «Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-Analysis». *Language Learning*, 50(3), 417-528. <https://doi.org/10.1111/0023-8333.00136>.
- O'Keeffe, A.; McCarthy, M.; Carter, R. (2007). *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge (UK): Cambridge University Press.
- Poole, R. (2018). *A Guide to Using Corpora for English Language Learners*. Edinburgh: Edinburgh University Press.
- Römer, U. (2006). «Pedagogical Applications of Corpora: Some Reflections on the Current Scope and a Wish List for Future Developments». *ZAA*, 54(2), 121-34.
- Römer, U. (2009). «Corpus Research and Practice: What Help Do Teachers Need and What can We Offer?». Aijmer, K. (ed.), *Corpora and Language Teaching*. Amsterdam: John Benjamins, 83-98. *Studies in Corpus Linguistics* 33. <https://doi.org/10.1075/scL.33.09rom>.
- Römer, U. (2010). «Using General and Specialized Corpora in English Language Teaching: Past, Present and Future». Campoy-Cubillo, M.C.; Bellés-Fortu-

- ño, B.; Gea-Valor, L. (eds), *Corpus-Based Approaches to English Language Teaching*. London: Continuum, 18-35.
- Rundell, M. (2007). *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan.
- Rundell, M.; Stock, P. (1992). «The Corpus Revolution». *English Today*, 8(4), 45-51.
- Scheffer-Lacroix, E. (2020). «Barriers to Trainee Teachers' Corpus Use». *Crosthwaite 2020*, 47-64.
- Snow, C.E. (2010). «Academic Language and the Challenge of Reading for Learning About Science». *Science*, 328, 450-2.
- Spina, S. (2014). «Il *Perugia Corpus*: una risorsa di riferimento per l'italiano. Composizione, annotazione e valutazione». Basili, R.; Lenci, A.; Magnini, B. (eds), *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014*, vol. 1. Pisa: Pisa University Press, 354-9.
- Thomas, J. (2015). *Discovering English with Sketch Engine*. Versatile.
- Tribble, C. (2015). «Teaching and Language Corpora: Perspectives from a Personal Journey». Boulton, Leńko-Szymańska 2015, 37-62.

