

UNA RISORSA EDUCATIVA APERTA PER L'USO DEI CORPORA NELLA CLASSE DI ITALIANO: EXCURSUS STORICO-TEORICO E ANALISI DEI BISOGNI FORMATIVI

Luciana Forti¹, Alessio Amato², Davide Coccia³, Noemi Cuccu⁴, Claudia Latini⁵, Agnese Macocco⁶, Elena Michelini⁷, Sara Panetti⁸, Eleonora Zaffignani⁹

1. INTRODUZIONE

Da tempo, ricercatori e insegnanti di tutto il mondo sperimentano in maniera positiva l'uso dei corpora in classe mediante il *Data-driven learning* (DDL). I corpora sono definibili come raccolte di testi in formato elettronico, che vengono sottoposte ad una serie di procedimenti computazionali, atti a rendere i testi stessi interrogabili in modi diversi. Sfruttati in maniera estensiva, sono largamente impiegati nella ricerca linguistica (cfr. § 3 per i loro utilizzi in didattica). Un corpus elettronico è infatti in grado di rispondere a domande di ricerca disparate, attingendo ad una notevole mole di dati, ed utilizzando dei criteri di interrogazione specifici (*queries*), per incrociare gli elementi di maggiore interesse. In ambito glottodidattico, grazie ai diversi studi condotti sull'efficacia del DDL, si è giunti a una valutazione medio-alta di efficacia, pur riversandosi la sua diffusione spesso in ambiti accademici e ad alti livelli di competenza linguistica degli apprendenti.

La situazione poc'anzi descritta è simile nell'ambito italofono, in cui gli studi sul DDL sono aumentati, ma non in modo vertiginoso. Questi sono infatti più descrittivi, relativi alle potenzialità dell'uso dei corpora, mentre pochi sono quelli empirici (ad es. Forti, 2023, o Kennedy, Miceli, 2010). Sebbene, inoltre, gli ultimi anni siano stati caratterizzati dalla pubblicazione di risorse educative aperte, mirate alla formazione di studenti e docenti interessati ad utilizzare i corpora nella didattica, nessuna di queste è in lingua italiana, né si riferisce all'apprendimento della lingua italiana. Lo scopo di questo contributo è quello di presentare la prima fase di un progetto mirato a creare una risorsa educativa aperta (*Open Educational Resource* – OER) in lingua italiana, basata sulla collaborazione tra ricercatori e docenti, per l'insegnamento e l'apprendimento dell'italiano, sia esso come L2 che come L1, attraverso i corpora. Nello specifico, il

¹ Università per Stranieri di Perugia.

² Università di Parma.

³ I.I.S. "Angelo Frammartino" di Monterotondo, Roma.

⁴ Alma Mater Studiorum - Università di Bologna

⁵ Kids&Us, Roma.

⁶ La Scuola International School - San Francisco.

⁷ Università di Macerata.

⁸ I.C. Cocchi Aosta Todi, Perugia / CPIA, Terni.

⁹ Altas Language School, Dublino.

L'articolo è stato discusso e rivisto da tutti gli autori. In particolare, Luciana Forti è responsabile dei §§ 7, 8.1.1., 8.1.2 (insieme a Elena Michelini e Sara Panetti), 8.2.1., 8.2.2., 8.2.3 e 8.2.4. (insieme a Elena Michelini); Alessio Amato del § 2; Davide Coccia del § 3; Noemi Cuccu dei §§ 1 e 9; Claudia Latini del § 4; Agnese Macocco del § 5; Elena Michelini dei §§ 8.1.2. (insieme a Luciana Forti e Sara Panetti) e 8.2.4. (insieme a Luciana Forti); Sara Panetti dei §§ 8.1.2. (insieme a Elena Michelini e Luciana Forti) e 8.2.5.; Eleonora Zaffignani del § 5.

contributo illustra i risultati di una rilevazione dei bisogni formativi in relazione all'uso dei corpora in classe, effettuata tramite questionario. La rilevazione ha coinvolto 158 partecipanti attivi, a vario titolo, nella didattica della lingua italiana.

Il contributo è strutturato come segue. Passeremo dalla relazione dinamica tra ricerca e didattica con accenni alla letteratura sul tema nel secondo paragrafo, seguiti da un focus sugli usi che dei corpora si possono attuare, all'analisi più approfondita del DDL nel terzo, con focus della sua presenza nel panorama didattico italiano, nel sesto. Nel paragrafo quarto, invece, siamo andati ad analizzare diversi studi sulla conoscenza delle potenzialità pedagogiche dei corpora. Successivamente, dal quinto, è meglio precisato il concetto di OER, con una rassegna di quelle esistenti nell'ambito del DDL. Infine, settimo e ottavo paragrafo vanno ad esplicitare le varie componenti, e del progetto, e del questionario e ad analizzarne quantitativamente e qualitativamente le risposte. A conclusione dell'articolo vedremo una valutazione complessiva dei risultati ottenuti, e una parte finale comprendente una discussione a riguardo.

2. CORPORA E DIDATTICA DELLE LINGUE

2.1. Rivoluzioni e cambi di paradigma

La relazione che intercorre tra la *corpus linguistics* (di seguito LC) e la glottodidattica può essere efficacemente definita come dinamica, nella misura in cui le due prospettive di ricerca esercitano vicendevolmente una certa influenza. Infatti, le riflessioni glottodidattiche beneficiano dello strumento corpus e della metodologia che sottende la LC; al contempo, gli studi di carattere didattico supportano lo sviluppo di risorse e strumenti pedagogicamente rilevanti attraverso l'afflusso di dati e riflessioni significativi (Römer, 2008: 112-113). Questo scambio di risorse tra realtà scientifiche diverse è certamente coerente con l'approccio *enterprise* (McEnery *et al.*, 2006) e ramificato della LC (McEnery, Hardie, 2012: 266), che ha contribuito alla crescente considerazione di aspetti specifici della linguistica applicata quali il lessico e la fraseologia. La centralità di questi ha progressivamente sostituito l'attenzione prima riservata alla morfologia e alle indagini di stampo generativista sulla sintassi (Granger, 2009: 3-4; Flowerdew, 2019: 394; Hank, 2008: 219).

Tale stato di cose è stato anticipato e notato da alcuni studiosi, primo fra tutti John Sinclair, professore dell'Università di Birmingham, il quale predice i risultati delle analisi delle concordanze e delle possibili collocazioni attraverso il concetto di *idiom principle* ovvero il fatto che «a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments» (Sinclair, 1991: 110). L'idea che da ogni dinamica interpretazione di una collocazione derivi un significato flessibile e probabilistico al di là del significato “da dizionario” e la convinzione che la grammatica e il lessico siano intrinsecamente collegati per la tendenza della lingua ad essere *highly patterned* (Römer, 2008: 114; Hanks, 2008: 219) si concretizza nel pionieristico progetto COBUILD. Le intuizioni di Sinclair concorrono alla costruzione del primo dizionario per apprendenti *corpus-based*, primo di una serie di supporti didattici; questi, mostrando all'utente le regolarità in cui occorrono i lemmi e stratificando a posteriori, grazie alle analisi sulla frequenza d'uso degli item lessicali di interesse, le sfumature semantiche potenzialmente più utili e significative (per una trattazione esaustiva si rimanda al sito ufficiale dell'editore Collins¹⁰) (Sinclair, 1987), rispondono efficacemente alla critica avanzata da Quirk (1960)

¹⁰ <https://blog.collinsdictionary.com/the-history-of-cobuild/>.

sull'inadeguatezza dei materiali pedagogici a lui coevi per la lingua inglese e seguono l'impostazione delle grammatiche di Quirk (Quirk *et al.*, 1985), nutrite di dati linguistici reali. Proprio l'analisi empirica sui fenomeni linguistici allontana questi studi dai “*myths and distortions*” caratterizzanti l’ “*armchair linguistics*” (Johns, 1991: 30), inaugurando così gli usi indiretti dei corpora in prospettiva glottodidattica. Nonostante il vivace clima accademico internazionale che circonda tale ambito di ricerca, si ritiene opportuno ricordare che la maggior parte di grammatiche come *Cambridge Grammar of English* (Carter, McCarthy, 2006), *Exploring Grammar in Context* (Carter *et al.*, 2000), dizionari per apprendenti *Cambridge Advanced Learner's Dictionary* (2013), *Longman Dictionary of Contemporary English* (2016) e libri di testo come *Touchstone* (McCarthy *et al.*, 2004, 2015), *Evolve* (Clandfield *et al.*, 2018, 2023), *corpus-informed* così come manuali destinati ai docenti al fine di evidenziare e mettere in pratica le potenzialità applicative dei corpora nel contesto didattico, sono stati realizzati specificamente per l'insegnamento e l'apprendimento della lingua inglese come lingua non materna (Poole, 2018; Friginal, 2018; Viana, 2022).

Nella stessa università in cui il progetto COBUILD prende forma, le analisi sulle informazioni linguistiche autentiche si fondono all'esperienza di Tim Johns come docente di lingua inglese a studenti internazionali, generando le fondamenta di un approccio a base induttiva basato sull'osservazione delle linee di concordanze da parte degli apprendenti (Johns, 1986; 1988; 1990; 1991). Tale metodologia sarà approfondita dagli altri autori, ma si ritiene utile anticipare che le prime applicazioni del DDL si traducono nella possibilità da parte degli apprendenti di notare regolarità e tracciare generalizzazioni a partire dai dati che possono essere sfuggite anche ai docenti (Johns, 1991: 4).

Infine, tra i mutamenti degni di nota che hanno rafforzato il collegamento tra glottodidattica, LC e Second Language Acquisition (in seguito SLA), si cita sì l'avvento del computer e del World Wide Web, ma anche la presenza massiccia di software CALL (*Computer-assisted language learning*), il cui fulcro attuale ruota intorno ai bisogni specifici dell'apprendente e alla volontà di incoraggiarlo ad esercitare maggior controllo sul processo di apprendimento (Friginal, 2018: 33).

2.2. *Ramificazioni pedagogiche dirette e indirette della LC*

Una prima distinzione tra tipologie di applicazioni didattiche dei corpora, in grado di interagire e modificare le pratiche di insegnamento nel contesto di apprendimento, identifica usi diretti e indiretti (Leech, 1997; Römer, 2008). Questi ultimi, originatisi dalle indagini *corpus-based* dei ricercatori, forniscono input rilevanti per la costruzione di una nuova generazione di materiale didattico, dizionari, sillabi e, più generalmente, per la definizione del contenuto dei libri di testo (Römer, 2008: 114-116). Un esempio sono le caselle all'interno di corsi di lingua il cui scopo è illustrare la frequenza d'uso comparativa di specifici item lessicali o grammaticali (Curry, Mark, 2024: 188). Al contrario, le ramificazioni dirette sono rappresentate dal DDL, metodo che in questa sede ci si limiterà a definire rivoluzionario per il modo in cui, rendendo immediata la relazione tra apprendente e dati linguistici, stimola lo studente ad essere parte attiva del processo di apprendimento linguistico (Johns, 1991) e si aggancia ad una serie di concetti chiave come il *lifelong learning*, *critical thinking*, *authenticity autonomy*, *cognitive depth* (Boulton, 2016: 6).

Inoltre, si ritiene particolarmente interessante citare una terza categoria approfondita da Römer (2008: 114) e accennata da Leech (1997: 5-6), ovvero i *learner corpora*. Questi realizzano la convergenza tra LC, teoria linguistica, glottodidattica e SLA, configurandosi come risorse preziose (Granger, 2009: 2-3) per identificare le tipologie di errori e comprendere a quali elementi dare la priorità nella pratica didattica. Nonostante ciò, l'uso

dei *learner corpora* solleva interrogativi ai quali la ricerca e la classe docente possono porvi rimedio attraverso la cooperazione, come la definizione dell'errore, quali feedback dovrebbero seguire alla loro identificazione e riconoscimento. Pertanto, al fine di promuovere un uso efficace delle risorse *corpus-based* da parte dell'insegnante, sarebbe utile che questo assumesse il ruolo di ricercatore e che fosse coinvolto attivamente nel contesto didattico nel quale è inserito (McCarthy, 2008: 564-570). Appare infatti necessario che lo sviluppo di corpora e di software user-friendly destinati all'insegnamento e all'apprendimento di collocazioni, registri, item lessicali e grammaticali attraverso i corpora si affianchi ad una formazione dei docenti in grado di declinare l'interesse verso il framework della linguistica applicata *corpus-based/corpus-driven* e alle sue diramazioni applicative (McCarthy, 2008: 570; Corino, 2019: 11).

3. IL DATA-DRIVEN LEARNING (DDL)

3.1. *Caratteristiche fondamentali*

Se i dati estratti da un corpus possono essere utili alla didattica anche quando non sono visibili agli apprendenti, come ad esempio nella fase di progettazione di un syllabo, nella costruzione di un test o nella definizione degli obiettivi di apprendimento, questi possono costituire anche un importante strumento didattico con cui gli apprendenti hanno la possibilità di interfacciarsi direttamente. Ogni metodologia che preveda tale rapporto diretto rientra nel dominio del DDL (cfr. Forti, 2021: 102), secondo una locuzione formulata da Tim Johns (1990; 1991) nei primi anni Novanta e che indica un processo di apprendimento guidato, appunto, dai dati estratti da un corpus e, quindi, di natura tipicamente induttiva. In questo approccio, infatti, gli studenti sono chiamati a scoprire autonomamente regolarità grammaticali, a inferire delle regolarità o delle consuetudini d'uso, avvicinandosi alle regole per approssimazioni progressive (Aston, 1998: 13).

Si tratta, quindi, di una metodologia didattica in linea con le più recenti teorie linguistiche *usage-based* (cfr. Tomasello, 2003), con i principi del costruttivismo (Cobb, 1999) e con le ricerche della SLA che mostrano l'importanza di valorizzare i processi *bottom-up* e *form-focused*. Si allinea, inoltre, ai concetti fondanti della teoria socioculturale (Vygotskij, 2012 [1934]). L'approccio DDL riflette la naturale tendenza del cervello umano a identificare dei pattern nel linguaggio, che è la base dell'interiorizzazione di una lingua secondo la *noticing hypothesis* (Schmidt, 2001).

La crescente autonomia dell'apprendente presuppone un ruolo diverso per il docente, il quale deve saper guidare i suoi studenti nel loro percorso di scoperta e di interpretazione dei dati linguistici, supportandoli e aiutandoli ad orientarsi nei loro stessi processi di scoperta e di apprendimento. I docenti possono servirsi dei dati estratti dai corpora in una duplice modalità: da un lato, possono fornire un input «condensato di regolarità» (Gabrielatos, 2005: 8), attraverso la presentazione di concordanze, collocazioni o liste di frequenza stampate e consegnate in fotocopia, senza che gli studenti si interfaccino direttamente con il corpus attraverso il computer; secondo un altro approccio, invece, possono prevedere attività ideate dai docenti – come avviene con i *kibbitzers*¹¹ di Tim Johns – o dagli stessi apprendenti nelle quali è previsto che questi si servano in modo autonomo o guidato del corpus¹².

¹¹ Alcuni esempi di *kibbitzers* sono disponibili online: <https://lexically.net/TimJohns/#gsc.tab=0>.

¹² A tal proposito si veda l'approccio *hard* (Gabrielatos, 2005: 13) descritto in Aston (1996), Bernardini (2002) e Tognini Bonelli (2001).

I corpora sono una risorsa utilizzabile in un'ampia varietà di soluzioni didattiche e, proprio per via della loro versatilità, sono stati sperimentati in vari contesti di insegnamento per lo sviluppo di diversi *task* specifici. Infatti, il campo di applicazione del DDL, che inizialmente era concentrato sulla sintassi e sulle abilità linguistiche orali, ben presto si è allargato alla pragmatica e agli aspetti discorsivi, lessicali e prosodici della lingua¹³ con un'attenzione particolare alle finalità di arricchimento del vocabolario e alla lessico-grammatica. Anche la platea di destinatari con cui sono state sperimentate attività DDL sembra decisamente ampia: nonostante il contesto di applicazione più frequente sia quello universitario, non mancano esempi di sperimentazioni condotte a livelli bassi e intermedi di competenze, con giovani e addirittura bambini (cfr. Crosthwaite, 2020).

In generale, gli studi condotti sull'uso dei corpora nelle lezioni di lingua paiono confermarne la polivalenza e l'efficacia: infatti, nonostante sembri dare i suoi frutti migliori nello sviluppo della competenza lessicale e con apprendenti di livello almeno intermedio – specialmente quando possono interrogare direttamente il corpus (Cfr. Boulton, Cobb, 2017; Lee *et al.*, 2019) – «il DDL funziona pressoché in ogni contesto in cui è stato ampiamente sperimentato» (Boulton, Cobb, 2017: 386)¹⁴. Anche per questo, forse, è lecito attendersi, in un prossimo futuro, un incremento di sperimentazioni didattiche *DDL-based* che possano contribuire ad instaurare un dialogo «permanente e orizzontale tra docenti e ricercatori», mirato alla realizzazione di materiali «funzionali ai reali bisogni degli apprendenti» (Forti, 2021: 104) che possano colmare la distanza tra la ricerca e la pratica didattica quotidiana. Non è certo questa, però, l'unica distanza che il DDL dovrà cercare di colmare nel prossimo futuro: O'Keeffe (2021: 259)¹⁵ si augura, ad esempio, un «broader research gaze» che possa spingere la ricerca sul DDL oltre il costruttivismo e che possa legarlo in modo più stabile alla SLA.

3.2. Un esempio di attività DDL per l'italiano

L'attività che segue è immaginata per un contesto con studenti L2 di livello A2/B1. L'obiettivo è favorire l'apprendimento della distinzione tra *gli/le* con funzione di articolo determinativo e di pronomi personale. L'attività si basa su dati estratti dal *Corpus La Repubblica*¹⁶. In una prima fase, agli studenti viene consegnato un handout che contiene una serie di concordanze di *gli* tratte dal *corpus La Repubblica* e si chiede loro di evidenziare le frasi in cui *gli* ha funzione di articolo. In un secondo momento, si propongono le medesime concordanze con l'esplicitazione dei tag grammaticali (Tabella 1), invitando gli studenti a correggere eventuali errori e spiegando che il tag “CLI” indica un pronome clitico.

¹³ Traduzione mia. Originale: «DDL works pretty well in almost any context where it has been extensively tried». Per un prospetto diacronico delle abilità linguistiche si veda Boulton, Vyatkina (2021: 79-80) e le *Period-Specific Keywords* della loro indagine raccolte nell'Appendice C.

¹⁴ A conclusioni simili giungono anche le meta-analisi di Mizumoto, Chujo (2015) e di Lee *et al.* (2018).

¹⁵ Sullo stesso argomento si vedano anche Chambers (2019) e Boulton, Vyatkina (2021: 67).

¹⁶ Il corpus è consultabile al sito <https://corpora.dipintra.it>.

Tabella 1. *Selezione di concordanze di gli estratte dal Corpus La Repubblica, con etichetta grammaticale*

1	#152794522	sacrifici . </s><s>	Quelli di terra e gli assistenti di volo – ha aggiunto – h
		fi NOUN SENT	PRO:demo PRE NOUN CON ART NOUN PRE NOUN PUN AUX:fin VER:ppast PUN AI
2	#7897866	dal Papa , processato dagli psicologi per gli incidenti , dopo quello accaduto a Roma	
		PRE NOUN PUN VER:ppast ARTPRE NOUN PRE	ART NOUN PUN PRE PRO:demo VER:ppast PRE NPR
3	#209534628	i di routine . </s><s>	Alla truccatrice che gli passa la cipria sul volto chiede con fa
		PRE NOUN SENT	ARTPRE NOUN CHE CLI VER:fin ART NOUN ARTPRE NOUN VER:fin PRE VER
4	#249385819	ombrellone è un Alesi che rifulge , che gli dà dentro , che non ha paura e	
		NOUN VER:fin ART NPR CHE VER:fin PUN CHE	CLI VER:fin ADV PUN CHE NEG VER:fin NOUN CON AF
5	#33679481	. </s><s>	In realtà , chi " legge " gli ultimi due anni di presidenza della Repu
		SENT	PRE NOUN PUN WH PUN NOUN PUN ART ADJ DET:num NOUN PRE NOUN ARTPRE NO
6	#44744615	me comunista , dopo che nel gennaio 1990 gli venne asportato un tumore a un rene ,	
		JN ADJ PUN ADV CHE ARTPRE NOUN NUM	CLI AUX:fin VER:ppast ART NOUN PRE ART NOUN PUN
7	#42747869	nell' elenco dei difensori d' ufficio solo gli avvocati iscritti al proprio albo ma s	
		ARTPRE NOUN ARTPRE NOUN PRE NOUN ADV	ART NOUN VER:ppast ARTPRE DET:poss NOUN CON CA
8	#352716370	sciiti iracheni . </s><s>	Il turbante nero che gli avvolge la testa ricorda a tutti che è
		NOUN ADJ SENT	ART NOUN ADJ CHE CLI VER:fin ART NOUN VER:fin PRE PRO:indef CHE VER
9	#111903681	sempre il massimo . </s><s>	Differenti sono gli irlandesi : loro vivono di realtà , e
		ADV ART NOUN SENT	ADJ VER:fin ART NOUN PUN PRO:pers VER:fin PRE NOUN PUN COI
10	#136213720	una " autorità indipendente " di tutti gli appalti e sospensione delle procedure già a	
		ART PUN NOUN ADJ PUN PRE DET:indef	ART NOUN CON NOUN ARTPRE NOUN ADV VE
11	#102658901	temporanea al Mit di Boston . </s><s>	Gli storici della Guerra nel Pacifico ci dicono
		ADJ ARTPRE NPR PRE NPR SENT	ART NOUN ARTPRE NPR ARTPRE NPR CLI VER:fin
12	#345473632	la cena . </s><s>	Dice che la solitudine gli fa venire le idee . </s><s>
		RT NOUN SENT	VER:fin CHE ART NOUN CLI VER2:fin VER:infi ART NOUN SENT ART NOU
13	#373283433	e " . </s><s>	Fatto il sondaggio , gli psicologi della Fip hanno chiesto al c
		PUN SENT	VER:ppast ART NOUN PUN ART NOUN ARTPRE NPR AUX:fin VER:ppast ARTPRE
14	#124596754	ppure di palle in porta giocando a tennis gli psicologi ne hanno sempre tirate molte	
		ADV PRE NOUN PRE NOUN VER:geru PRE NOUN	ART NOUN CLI AUX:fin ADV VER:ppast DET:indef AR
15	#152795257	a possibile , clamorosa esclusione . </s><s>	' Non gli prometto nulla ' . </s><s>
		PUN NEG	CLI VER:fin PRO:indef PUN SENT ADJ NP

Successivamente, si stimolano gli studenti a riflettere intorno alla posizione nella frase delle forme che hanno sottolineato e al loro contesto, attraverso un'attività di scoperta guidata come quella mostrata nella Tabella 2.

Tabella 2. *Attività DDL paper-based di scoperta guidata 1 (gli)*

Osservando le frasi dell'esercizio precedente, riempi la seguente tabella:

	<i>Gli CLI</i>	<i>Gli ART</i>
<i>Da quale parte del discorso è seguito più di frequente?</i>		
<i>Quale posizione può assumere nella frase (iniziale, intermedia, finale)?</i>		
<i>Con quali lettere iniziano le parole che lo seguono?</i>		
<i>Può essere sostituito con "a lui"?</i>		

Dopo la fase di identificazione delle regolarità si passa a quella di generalizzazione. Considerando il contesto destro di *gli*, si cerca di individuare una regola nella scelta dell'articolo determinativo tra *i* e *gli*, attraverso l'attività riportata nel Tabella 3.

Tabella 3. *Attività DDL paper-based di scoperta guidata 2 (gli)*

Osservando le frasi del primo esercizio, rispondi alle domande seguenti:

- Quali nomi sono preceduti dall'articolo "gli"?

- Con quali lettere/suoni iniziano questi nomi?

- Quali articoli useresti nei corrispondenti nomi al singolare?

A questo punto, eventualmente, si può ripetere la stessa procedura della prima fase presentando delle concordanze di *le*, per far emergere, attraverso un'altra attività di scoperta guidata la preponderanza di nomi nel contesto destro di *le* articolo e di verbi in quello di *le* clitico, come quella mostrata nella Tabella 4.

Tabella 4. *Attività DDL paper-based di scoperta guidata 1 (le)*

Osservando le frasi dell'esercizio precedente, riempi la seguente tabella:

	<i>Le</i> CLI	<i>Le</i> ART
<i>Da quale parte del discorso è seguito più di frequente?</i>		
<i>Quale posizione può assumere nella frase (iniziale, intermedia, finale)?</i>		
<i>Può essere sostituito con "a lei"?</i>		

In questa operazione, ci si può servire anche di tabelle e schemi vuoti, per facilitare la «formulazione delle regolarità presenti»¹⁷ nelle concordanze fornite. Dopo eventuali attività di sintesi e di formulazione della regola, si può passare ad esercizi di reimpiego, che possono essere condotti anche su concordanze "bucate", nelle quali gli studenti sono chiamati ad inserire il tag grammaticale all'interno di una lista di concordanze tratte dallo stesso corpus Repubblica.

4. DIFFUSIONE E CONOSCENZA DEL DDL IN AMBIENTI DI ISTRUZIONE LINGUISTICA

Qual è il grado di conoscenza delle potenzialità pedagogiche dei corpora in ambienti educativi extra-universitari? Per rispondere a tale domanda, senza pretesa di esaustività, menzioneremo alcuni tra gli studi più importanti che hanno cercato di fornire una risposta. Mukherjee (2004) indaga l'uso e la conoscenza dei corpora da parte di insegnanti di inglese delle scuole secondarie di North Rhine-Westphalia (Germania) tra il 2001 e il 2002, attraverso la diffusione di un questionario. La somministrazione è avvenuta

¹⁷ Come indica Lo Duca (2004 [1997]: 177), sulla scorta di Cesarini (1995).

attraverso formati *ad hoc* a seconda del contesto istituzionale di riferimento, ma con una struttura comune formata da:

- una lezione introduttiva sulla linguistica dei corpora;
- un seminario di confronto tra i dati ottenuti da ricerche su corpora (es. la frequenza di una stringa lessico-grammaticale) e quelli forniti da grammatiche pedagogiche e libri di testo;
- una parte pratica di applicazione dei dati del corpus sperimentata all'interno di una classe, in riferimento all'uso di una forma linguistica da parte di parlanti L1: *example for/example of*.

L'obiettivo principale dello studio era quello di individuare il grado di conoscenza pregressa in materia di corpora da parte dei 248 insegnanti intervistati. Dallo studio è emerso che quasi l'80% degli intervistati non aveva mai sentito parlare di corpora, dato che non può non tenere conto dell'età degli insegnanti (la maggior parte tra i 50 e i 65 anni). Una percentuale così alta mette in luce la necessità di introdurre alla linguistica dei corpora anche gli insegnanti della scuola secondaria, che non hanno ricevuto una formazione al riguardo, permettendogli così di familiarizzare con i metodi *corpus-based*.

Un altro dato rilevante evidenziato dall'indagine di Mukherjee è la risposta data dall'83,9% degli intervistati dopo la partecipazione al primo workshop, secondo la quale solamente gli insegnanti beneficerebbero dall'uso dei dati del corpus. Considerando che anche dopo il secondo workshop solamente l'11,7% concepisce un uso diretto del corpus da parte degli apprendenti, si può dedurre che tra gli insegnanti prevale decisamente una visione del processo di apprendimento *teacher-centred* piuttosto che *learner-centred*, visione che contrasta con la natura stessa del DDL, che mette in discussione il ruolo dell'insegnante come figura di riferimento per la trasmissione della conoscenza linguistica. Il docente potrebbe quindi sperimentare una perdita di controllo sui contenuti della lezione, trovandosi allo stesso livello dell'apprendente, per osservare insieme le strutture linguistiche, porsi domande e fare inferenze sul loro uso effettivo da parte dei parlanti L1.

In ultima analisi, i test somministrati durante il workshop hanno evidenziato la curiosità della maggior parte dei partecipanti nei confronti dell'uso dei corpora come parte integrante della lezione di L2, dovendosi inevitabilmente ricredere di fronte all'evidenza dei dati su “norme” grammaticali comunemente presenti nelle grammatiche scolastiche. Come spiegato dallo studioso, infatti, «There are many other examples of corpus-based findings that call into question the way English language teachers go about correctness in learner language. Specifically, the discussion of such examples makes it clear to all teachers that their own intuition is often at odds with linguistic reality» (Murherjee, 2004: 246).

Nel 2015, Christopher Tribble pubblica gli esiti di una indagine simile a quella condotta da Murkherjee (Tribble, 2015). Lo scopo è quello di rispondere alla seguente domanda: chi usa attualmente i corpora per l'insegnamento della L2 e con quale scopo? Il questionario è stato compilato da 560 tra insegnanti di lingua inglese e coloro che si occupano della formazione degli insegnanti, parlanti di più di 50 lingue diverse e provenienti da 63 paesi: un campione decisamente più ampio rispetto alle indagini precedenti. Analizzando le risposte, lo studioso ha rilevato che il 75% degli intervistati – prevalentemente insegnanti universitarie tra i 30 e i 49 anni – utilizza i corpora nella propria lezione di lingua. Conseguentemente, le lezioni sono rivolte a studenti universitari con un alto livello di competenza in inglese; dato rilevante se paragonato all'indagine del 2008, dove buona parte degli intervistati insegnava a studenti di scuole secondarie (statali o private) con un livello B1-B2 di competenza linguistica. Con la popolazione di riferimento cambia anche lo scopo di insegnamento: se nel 2008 “English for general purposes” superava “English for academic purposes”, nel 2012 troviamo una situazione inversa.

Dopo una serie di domande sugli strumenti e sui programmi utilizzati dagli insegnanti per l'analisi del corpus, il questionario si concentra sui benefici riscontrati dagli intervistati. L'impatto positivo sul proprio modo di insegnare è percepito dagli insegnanti intervistati da Tribble come più significativo rispetto all'impatto dell'approccio sull'apprendimento degli studenti, sebbene il 37% abbia riscontrato una differenza importante anche dal punto di vista dell'apprendimento dei propri studenti. Tra i principali motivi per cui i corpora non sono considerati utili nell'insegnamento della L2 si segnala la limitata conoscenza di questi strumenti e metodi didattici tra gli insegnanti e il timore di "perdere tempo" per la preparazione dei materiali e per la spiegazione sul funzionamento del corpus ai propri studenti.

A distanza di anni da questi due studi, la situazione non sembra migliorata (Leńko-Szymańska, 2024). Farr e Karlsen (2023) annoverano tra i principali ostacoli alla diffusione del DDL proprio la scarsa conoscenza dei concetti base di linguistica dei corpora e, di conseguenza, nessuna familiarità con il corpus. Infatti, come sostenuto in Corino (2019: 276): «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».

5. RISORSE EDUCATIVE APERTE PER FAVORIRE L'USO DEI CORPORA IN CLASSE

5.1. Cosa sono le risorse educative aperte

Le risorse educative aperte (OER, *Open Educational Resources*) rappresentano uno strumento fondamentale per rendere i corpora accessibili a docenti e studenti, offrendo materiali didattici gratuiti, adattabili e basati su dati autentici. Oltre a supportare l'insegnamento, le OER contribuiscono a rendere la ricerca più accessibile, promuovendo una conoscenza scientifica rigorosa, trasparente, collaborativa e inclusiva, in linea con i principi dell'*Open Science* e dell'*Open Education* (Le Foll, 2023). Per essere considerate OER, le risorse devono soddisfare una serie di requisiti chiave, che si basano sulla loro accessibilità, licenza e utilizzo. Innanzitutto, è necessario che sia garantita un'accessibilità gratuita: devono essere liberamente accessibili online, senza costi di acquisto o restrizioni economiche, e possono essere materiali digitali (come e-book, video, piattaforme interattive) o stampabili. Inoltre, devono essere rilasciate con una licenza che permetta il riutilizzo, l'adattamento, la modifica e la condivisione. Devono, poi, prevedere la possibilità di personalizzazione, ovvero devono consentire agli utenti di modificare i contenuti per adattarli a contesti specifici di insegnamento o apprendimento. Infine, devono essere trasparenti, ovvero devono includere una descrizione chiara del contenuto e delle modalità d'uso. Questi requisiti rendono le OER strumenti che garantiscono l'accesso gratuito ai materiali didattici, permettendo a chiunque di utilizzarli, modificarli e condividerli. In questo modo, le OER favoriscono la collaborazione tra educatori, studenti e ricercatori a livello globale, promuovendo una diffusione aperta e inclusiva della conoscenza.

5.2. Principali risorse educative aperte per imparare a usare i corpora in classe

Gli ultimi anni hanno visto la pubblicazione in rete di una serie di risorse educative aperte mirate alla formazione di studenti e docenti interessati a usare i corpora in classe, per favorire l'apprendimento delle lingue. Questa rassegna esplora alcune delle principali risorse educative aperte disponibili, senza pretesa di esaustività, analizzandone le

caratteristiche, i contenuti e le modalità di impiego, con l'obiettivo di evidenziare i punti di forza e le aree di miglioramento.

Incorporating Corpora di Nina Vyatkina (2020) è progettato per l'insegnamento del tedesco a studenti anglofoni. Include una guida all'uso di corpora tedeschi e attività basate su corpora, divise per livelli di competenza e con feedback automatico. L'obiettivo è quello di migliorare l'apprendimento del tedesco attraverso l'analisi di corpora, focalizzandosi su aspetti come le desinenze degli aggettivi, le collocazioni e la negazione. Per questa risorsa, viene usato il *Das Digitale Wörterbuch der Deutschen Sprache* (DWDS), organizzato in diversi subcorpora, che include un'ampia gamma di testi storici e contemporanei della lingua tedesca. Il corpus principale è diviso sia per periodo che per tipo di testo e include quattro categorie principali: narrativa, saggistica, scienza e giornalismo. I punti di forza consistono nel fatto che include esercizi mirati a diversi aspetti grammaticali della lingua tedesca, che possono essere svolti online dagli apprendenti stessi. La risorsa si limita all'apprendimento del tedesco e richiede competenze linguistiche sia in inglese che in tedesco.

Corpora in English Language Teaching. Classroom Activities for Teachers New to Corpus Linguistics, di Larissa Goulart e Ingrid Veloso (2023) introduce gli strumenti e le tecniche della linguistica dei corpora a insegnanti in formazione. I capitoli, scritti da studenti universitari, presentano attività didattiche basate sui corpora per l'insegnamento dell'inglese. La risorsa è divisa in due sezioni: una guida sull'uso degli strumenti di corpus online e attività pratiche. L'obiettivo generale della risorsa è quello di migliorare la comprensione e l'uso dell'inglese attraverso l'analisi di corpora, nonché incoraggiare gli insegnanti in formazione a creare e adattare materiali didattici utilizzando i corpora. Tra i corpora utilizzati figurano il *Corpus of Contemporary American English* (COCA) e il *British National Corpus* (BNC) tramite Sketch Engine. Tra i punti di forza della risorsa possiamo menzionare il fatto che offre tutorial dettagliati per l'uso di strumenti corpus, include attività pratiche e diversificate che rendono l'apprendimento coinvolgente, e che, più in generale, si adatta a insegnanti nuovi alla linguistica dei corpora. La risorsa si limita alla lingua inglese, richiede una certa familiarità con l'uso di strumenti digitali, e potrebbe porre dei limiti a chi non ha accesso a strumenti online.

Using Language Data to Learn About Language: A Teachers' Guide to Classroom Corpus Use di Paula Tavares Pinto, Peter Crosthwaite, Carolina Tavares de Carvalho, Franciele Spinelli, Talita Serpa, William Garcia e Adriane Orenha Ottaiano (Tavares Pinto *et al.*, 2023) contiene attività didattiche per l'insegnamento dell'inglese, portoghese e spagnolo. Ogni piano di lezione include una descrizione degli strumenti usati, un breve video in cui l'autore (o gli autori) della lezione spiega come utilizzarli e attività pratiche. Si propone di facilitare l'integrazione dei corpora nell'insegnamento delle lingue, migliorando competenze come vocabolario, grammatica e comprensione culturale. Contiene attività basate su corpora sia di L1 che di L2, in relazione a più lingue. Ogni lezione contiene una descrizione dello strumento usato e un video esplicativo. Inoltre, le attività si possono svolgere con o senza accesso a Internet (*hands-on* o *hands-off*). Tra i limiti menzioniamo il fatto che può richiedere tempo per adattare le lezioni a specifici contesti educativi, che la varietà di lingue presenti all'interno della risorsa potrebbe confondere alcuni insegnanti, e che alcune delle risorse usate sono a pagamento.

Creating Corpus-Informed Materials for the English as a Foreign Language Classroom di Elen Le Foll (Le Foll, 2021) si presenta come una guida per insegnanti sull'uso di risorse online per sviluppare materiali didattici basati su corpora. È organizzata in sedici capitoli suddivisi in quattro sezioni: scuola primaria e secondaria di primo grado, scuola secondaria di secondo grado, CLIL (*Content and Language Integrated Learning*) nella scuola secondaria, e *English for Specific Purposes* (ESP). L'obiettivo della risorsa è quello di supportare gli

insegnanti nella creazione di materiali didattici per l'insegnamento dell'inglese come lingua straniera, migliorando competenze come vocabolario, grammatica e uso della lingua. La risorsa si basa sull'uso di corpora di inglese L1, accessibili tramite strumenti come *English-corpora.org*, *Sketch Engine* e *SKELL*. Tra i punti di forza, menzioniamo la presenza di una guida dettagliata per l'uso dei corpora nell'insegnamento, l'inclusione di spunti e idee per lezioni adattabili a diversi livelli scolastici e contesti educativi, e la presenza di una lista di risorse aggiuntive utili per insegnanti e studenti. La risorsa si limita all'apprendimento della lingua inglese e richiede una certa familiarità con strumenti digitali. La Tabella 5 fornisce una sintesi delle risorse descritte.

Tabella 5. *Sintesi delle principali risorse educative aperte per imparare a usare i corpora in classe*

Risorsa	Tipo di corpus	Obiettivi linguistici	Punti di forza	Limiti
<i>Incorporating Corpora</i> (Vyatkina, 2020)	Tedesco L1	Migliorare l'apprendimento del tedesco	Esercizi specifici sulla grammatica	Limitato al tedesco; richiede competenze in inglese e tedesco
<i>Corpora in English Language Teaching. Classroom Activities for Teachers New to Corpus Linguistics</i> (Goulart e Veloso, 2023)	Inglese L1	Migliorare la comprensione e l'uso dell'inglese	Tutorial dettagliati; attività diversificate	Focalizzato solo sull'inglese; richiede familiarità con strumenti digitali
<i>Using Language Data to Learn About Language: A Teachers' Guide to Classroom Corpus Use</i> (Tavares Pinto et al., 2023)	L1 e L2	Integrare l'uso dei corpora nell'insegnamento di inglese, portoghese e spagnolo	Risorse per più lingue; lezioni adattabili; video esplicativi	Richiede tempo per adattamento; la varietà di lingue può risultare dispersiva; alcune risorse a pagamento
<i>Creating Corpus-Informed Materials for the English as a Foreign Language Classroom</i> (Le Foll, 2021)	Inglese L1	Supportare la creazione di materiali didattici per l'insegnamento dell'inglese	Guida dettagliata; idee per lezioni per vari livelli scolastici; risorse aggiuntive gratuite	Focalizzato sull'inglese; richiede familiarità con strumenti digitali

Le risorse educative aperte sono strumenti innovativi e flessibili per l'insegnamento delle lingue, ma presentano alcune sfide. Molte risorse non sono adattate a contesti specifici o a lingue meno comuni rispetto a quelle predominanti, come l'inglese e lo spagnolo. Inoltre, alcune piattaforme, pur essendo aperte, richiedono competenze

tecniche per essere sfruttate al meglio. Infine, come già messo in luce, manca del tutto una risorsa di questo tipo per la didattica della lingua italiana.

6. DDL E INSEGNAMENTO DELL'ITALIANO L1/L2

Dopo aver preso in esame il contributo apportato dall'utilizzo di corpora linguistici in termini di conoscenza e di comprensione della lingua target da parte di apprendenti di una lingua straniera, e il valore delle risorse educative aperte a diffondere tale opportunità, si andrà ora a porre l'attenzione sul ruolo che il DDL occupa nell'insegnamento dell'italiano L1 e L2. Constatato, infatti, come il panorama scientifico abbia messo abbondantemente in luce gli innumerevoli punti di forza dell'apprendimento linguistico induttivo e autodiretto mediato dai corpora, ci si concentrerà qui, da un lato, ad illustrare lo stato attuale della ricerca con un focus specifico nell'ambito dell'insegnamento-apprendimento in italiano L1 e L2 e, dall'altro, si prenderanno in esame le sfide legate all'implementazione di successo del DDL con studenti di italiano evidenziando il profondo iato che ancora intercorre tra l'auspicato incremento dell'uso di quest'ultimo e la reale applicazione nella prassi didattica. Questi due aspetti appaiono, infatti, strettamente correlati se si esplora e analizza la quantità e la tipologia di studi pubblicati nel tempo su DDL e italiano L2.

Si può mettere in luce, facendo riferimento a Forti (2024), come seppur ci sia fervore scientifico attorno al tema dal 1993 (Polezzi, 1993) ci sia, tuttora, dopo un trentennio, una quantità molto bassa di studi, equamente scritti in italiano e in inglese, per lo più in atti di convegno e riviste, e prevalentemente a carattere descrittivo. L'aspetto maggiormente interessante, dopo l'aspetto quantitativo, è la fondamentale carenza di sistematicità e consistenza per quanto riguarda gli studi empirici in contesto italiano, il che spiegherebbe l'ancora persistente diffidenza tra insegnanti di lingua italiana riguardo a interventi didattici DDL: data la scarsità di studi empirici risulta infatti difficoltosa sia la generalizzabilità degli interventi sia il provarne l'efficacia. Sono aspetti, questi, che si influenzano reciprocamente in un mutuo circolo, dove ciascun elemento alimenta l'altro, rendendo difficoltosa una reale implementazione e diffusione del DDL nel panorama didattico italiano. Questo perpetua da un lato la scarsa conoscenza di questo approccio tra gli insegnanti e dall'altro la mancanza di risorse e conoscenze adeguate da parte degli insegnanti, impedendo un progresso significativo nell'adozione di approcci più innovativi nell'insegnamento dell'italiano.

Nell'analizzare alcune esperienze di DDL nell'insegnamento dell'italiano occorre fare un distinguo riguardo ad italiano L1 e italiano L2. Gli studenti di italiano L1 e L2 presentano esigenze, infatti, molto diverse, a cui il DDL deve essere in grado di rispondere. Concentrandosi sull'insegnamento dell'italiano L2 e prendendo in considerazione i principali corpora italiani L1, è possibile evidenziare come, in termini di adattabilità pedagogica, questi ultimi presentino caratteristiche poco user-friendly e che, allo stato attuale, SkELL rappresenti la risorsa più adatta agli studenti per un approccio pratico al DDL (Forti, 2023). Se ci si focalizza, poi, su corpora L2 per apprendenti di italiano si evidenzia come essi debbano ancora essere ampliati in termini di dimensioni e collezionare testi orali o scritti in un arco temporale più ampio, con strumenti più avanzati per l'analisi dei dati e interfacce maggiormente fruibili dall'utente, auspicando il "bridging the gap" tra la ricerca e le applicazioni pedagogiche del DDL, in sintesi tra teoria e pratica (Forti, 2024). Ne viene tuttora sottostimato il potenziale in termini didattici anche se risultano preziose risorse rispetto ai corpora L1 dal momento che essi possono essere suddivisi per livello di competenza, avere un focus su pattern atipici ed errori e dunque offrire ai docenti un riferimento quantitativo solido a partire da dati linguistici autentici e rilevanti per elaborare materiali didattici mirati, personalizzati ed efficaci (Corino, 2021).

In entrambe le tipologie di corpora vediamo, dunque, che diviene necessaria e indispensabile una triangolazione tra docenti, corpora e discenti. I primi attori, i docenti, acquisiscono competenze nella progettazione di attività basate sul DDL e che parallelamente promuovono l'autonomia degli apprendenti, sviluppandone la capacità di utilizzare i corpora in modo indipendente. Corpora che devono essere progettati e ottimizzati specificamente per un uso didattico in classe, tenendo conto delle esigenze degli studenti, degli obiettivi pedagogici, dei vari livelli di competenza linguistica. Gli apprendenti, poi, attori nel processo del DDL che sono chiamati a sviluppare competenze di "imparare a imparare": devono acquisire infatti una consapevolezza metacognitiva che li renda in grado di utilizzare i corpora in modo strategico e autonomo. Questo implica non solo l'abilità di interpretare i dati linguistici forniti dai corpora, ma anche di riflettere criticamente sulle proprie pratiche di apprendimento, integrandole con strumenti digitali e metodologie basate sull'analisi dei pattern linguistici. Tale consapevolezza favorisce l'adozione dei corpora come risorsa abituale per l'apprendimento sfruttando i corpora per identificare pattern linguistici e coinvolgere gli studenti in attività di scoperta, analisi e confronto approfondito tra usi linguistici differenti all'interno della stessa lingua.

Altra area importante da trattare è costituita invece dalle esperienze di DDL con studenti di italiano L1. Appartengono a questa categoria tutte quelle esperienze che vedono dei *native speakers* essere apprendenti della propria L1 e necessitare di una formazione linguistica specifica mirata ad un processo di meta-riflessione cruciale non solo per l'analisi delle regole linguistiche generali, ma, soprattutto per l'acquisizione e l'utilizzo di codici linguistici specialistici e tecnici. Esiste, infatti, la necessità di ridurre «difficoltà legate alle sfumature specialistiche di certi significati o collocazioni che spesso allontanano il lessema dalla sfera semantica di alto uso per spostarne invece il centro verso il sottocodice» (Corino, 2022, p. 294). L'efficacia dell'utilizzo del DDL con studenti L1 sta dunque nel permettere loro di scoprire e analizzare dati linguistici e nel dare supporto nell'uso specialistico di certi termini.

Si vedano, ad esempio, esperienze che attraverso il DDL supportano gli studenti all'acquisizione e all'uso di linguaggio tecnico matematico. Tali esperienze mettono in luce come studenti L1 che conducono indagini linguistiche su un corpus appositamente progettato svolgono attività basate su corpora, sviluppano competenze linguistiche e matematiche e diventano più consapevoli dell'importanza del linguaggio tecnico che utilizzano. L'approccio DDL fornisce infatti agli studenti una chiave linguistica di accesso al contenuto specialistico andando a superare quello che spesso si configura come ostacolo e dando agli studenti che si avvicinano al lessico disciplinare della matematica la possibilità di accedervi in maniera pratica interattiva e non trasmissiva incrementando la motivazione di questi ultimi e permettendo sistemi di valutazione formativa.

In sintesi, seppur l'approccio DDL nell'insegnamento dell'italiano possa favorire lo sviluppo di competenze linguistiche autonome, migliorando la consapevolezza grammaticale e lessicale degli studenti e apportando un complessivo impatto positivo sull'apprendimento linguistico (Boulton, Cobb, 2017), per svilupparsi pienamente esso richiede una maggiore maggior collaborazione tra ricercatori e insegnanti. Solo in questo modo, sarà possibile sfruttare appieno il potenziale per migliorare la competenza linguistica e rendere l'apprendimento più efficace e significativo. Per far ciò si può auspicare, da un lato, un crescere della ricerca sul DDL nell'insegnamento dell'italiano sia sul piano quantitativo, estendendo i tempi di raccolta dei dati, sia qualitativo, con l'inclusione di campioni ancor più rappresentativi, dall'altro una sempre maggiore integrazione dell'uso dei corpora nella prassi didattica, in modo da poterne conseguentemente valutare gli effetti sulle dinamiche di apprendimento andando a rinforzare un mutuo scambio tra comunità di ricerca e comunità di insegnamento.

7. UNA RISORSA EDUCATIVA APERTA PER L'USO DEI CORPORA NELLA CLASSE DI ITALIANO: IL PROGETTO

Per colmare il divario tra ricerca e didattica in relazione all'uso dei corpora nella classe di italiano, e quindi per promuoverne gli usi e diffonderne i potenziali benefici, è stato sviluppato un progetto di ricerca mirato alla creazione e valutazione di una risorsa educativa aperta (OER). Sulla scia di quanto è già stato realizzato in ambito prevalentemente anglosassone (cfr. § 5), il progetto si prefigge di far emergere la risorsa da un contesto di collaborazione e scambio tra ricercatori e insegnanti, in cui i confini stessi dei due ruoli arrivano a sfumarsi¹⁸. In tal senso, ci si prefigge di imparare "insieme" proprio per integrare le diverse conoscenze e competenze, impiegate per il raggiungimento di un medesimo scopo: migliorare l'apprendimento linguistico.

Se da un lato il DDL nasce e si sviluppa in larghissima parte all'interno di ambienti accademici, dediti alla linguistica dei corpora e alle sue connessioni con il mondo della didattica delle lingue, dall'altro è solo attraverso l'interazione regolare con docenti di lingua, che insegnano la lingua quotidianamente, che la realizzazione di una risorsa educativa può essere possibile. Inoltre, per superare la dominanza della lingua inglese (Lanvers, 2024), anche nei materiali educativi relativi all'insegnamento corpus-based di lingue diverse dall'inglese, la risorsa sarà interamente in lingua italiana, oltre che liberamente accessibile online.

Il progetto prevede le seguenti fasi principali:

1. rilevazione dei bisogni formativi in relazione all'uso dei corpora in classe per l'apprendimento linguistico;
2. progettazione della risorsa educativa aperta, in relazione ai risultati della rilevazione di cui al punto 1;
3. creazione, impiego in classe, e successivo perfezionamento di attività DDL;
4. realizzazione della risorsa educativa aperta, con successiva messa in rete;
5. valutazione dell'impatto della risorsa educativa aperta, in contesti educativi specifici.

La sezione che segue tratterà la fase 1 del progetto, illustrandone i risultati e le implicazioni per la realizzazione della risorsa.

8. RILEVAZIONE DEI BISOGNI FORMATIVI SU CORPORA E APPRENDIMENTO DELL'ITALIANO

8.1. Metodo

8.1.1. Obiettivi e contenuti dello strumento di rilevazione

Per rilevare i bisogni formativi in relazione all'uso dei corpora in classe per l'apprendimento linguistico è stato costruito un questionario, basato su un insieme di domande aperte e domande chiuse. L'obiettivo ampio del questionario è stato quello di intercettare una platea di rispondenti variegata, a vario titolo coinvolta in attività di insegnamento / apprendimento della lingua italiana, e dunque in grado di interagire proficuamente nella realizzazione di una OER utile per tutti. Per questo motivo, la prima

¹⁸ Significativo, in tal senso, il titolo del simposio internazionale *Teacher Researchers and Researcher Teachers: Bridging the Research-Practice Gap*, tenutosi alla Libera Università di Bolzano il 25 ottobre 2024 (programma del simposio accessibile al link: <https://www.unibz.it/en/events/teacher-researchers-and-researcher-teachers>).

domanda del questionario, impostato in modo tale da essere comunque anonimo, ha riguardato l'identità del rispondente. Come si potrà notare dalla Tabella 6, varie sono state le opzioni fornite, insieme a un campo "Altro" in cui inserire possibilità ulteriori rispetto a quelle fornite.

Tabella 6. *Rilevazione dei bisogni formativi su corpora e apprendimento dell'italiano – Item 1*

1. Chi sei?	
<input type="radio"/>	Una persona che insegna italiano L2/LS
<input type="radio"/>	Una persona che sta studiando per insegnare L2/LS
<input type="radio"/>	Una persona che insegna italiano L1
<input type="radio"/>	Una persona che svolge attività di ricerca
<input type="radio"/>	Una persona che svolge attività editoriali
<input type="radio"/>	Altro

In corrispondenza della voce "Altro" è stata inserita una ramificazione contenente la domanda "Di che cosa ti occupi?", per raccogliere informazioni in merito alle attività dei rispondenti non esplicitate all'interno delle opzioni fornite. Nella parte successiva, sono stati inseriti due item per chiedere ai rispondenti di selezionare il proprio grado di bisogno formativo rispetto a una serie di temi e il proprio grado di preferenza rispetto a una serie di modalità per sviluppare conoscenze e competenze nell'ambito di quei temi. A ciascuno dei temi e delle modalità è stata associata una scala di Likert di quattro valori: nullo, basso, medio, alto. I temi sono elencati nel Tabella 7, mentre le modalità nel Tabella 8.

Tabella 7. *Rilevazione dei bisogni formativi su corpora e apprendimento dell'italiano – Item 2*

2. Cosa vuoi imparare? Seleziona il tuo grado di bisogno formativo relativo ai seguenti temi:	nullo	basso	medio	alto
Descrizione di corpora utilizzabili in classe.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Descrizione di principi da seguire per la creazione di attività didattiche basate su corpora in modalità digitale e cartacea.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Descrizione di software per la creazione di corpora propri da usare in classe.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Come creare attività con corpora per la scoperta di regolarità e regole d'uso della lingua.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Come creare attività con corpora per migliorare le competenze orali e interazionali.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Come creare attività con corpora per migliorare le abilità di scrittura.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Come creare attività con corpora per DSA e BES.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Come creare attività con corpora per bambini.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Come usare i corpora per attività che riguardano l'italiano L1.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Come usare i corpora per attività che riguardano l'italiano L2.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Come usare i corpora per l'apprendimento di linguaggi specialistici.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Teorie linguistiche e pedagogiche sulle potenzialità dei corpora in classe.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Studi recenti sull'uso dei corpora in classe, in riferimento all'insegnamento della lingua italiana.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Studi recenti sull'uso dei corpora in classe, in riferimento all'insegnamento di lingue diverse dall'italiano.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Tabella 8. *Rilevazione dei bisogni formativi su corpora e apprendimento dell'italiano – Item 3*

3. Come vuoi imparare? Seleziona il tuo grado di preferenza in relazione alle seguenti modalità:				
	nullo	basso	medio	alto
Con un vademecum di base che spieghi come creare attività basate su corpora.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Osservando filmati di lezioni in cui vengono usati i corpora.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Attraverso video-tutorial, in cui viene spiegato come usare un corpus.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Attraverso una raccolta sistematica e aggiornata nel tempo sugli studi che riguardano l'uso dei corpora in classe.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Attraverso filmati di testimonianza da parte di docenti che usano già i corpora in classe.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Attraverso attività già realizzate e sperimentate, da scaricare e riutilizzare.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

A seguire, sono state inserite le seguenti tre domande a risposta aperta:

4. Quali sono le maggiori difficoltà che immagini di avere nell'uso dei corpora a fini didattici?
5. Hai altri suggerimenti per la realizzazione della risorsa?
6. Se desideri ricevere aggiornamenti sulla realizzazione della risorsa, inserisci il tuo indirizzo di posta elettronica.

A differenza di tutti gli altri campi, gli ultimi due sono stati impostati come compilabili in maniera facoltativa.

8.1.2. *Analisi dei dati*

I dati raccolti attraverso Microsoft Forms sono stati analizzati attraverso una serie di grafici a barre divergenti per quanto riguarda gli item basati su scala Likert, e attraverso un'analisi del contenuto tematico per quanto riguarda le domande a risposta aperta. Prima di procedere con l'analisi dei dati, le risposte del questionario sono state controllate manualmente, al fine di rilevare eventuali incongruenze. Da questo controllo è stata rilevata la presenza di due serie di risposte identiche, inviate a distanza di pochi secondi. Trattandosi, dunque, di un probabile errore tecnico nella trasmissione dei dati, una delle due serie di risposte è stata eliminata.

Limitatamente ai dati ottenuti dagli item basati su scala Likert, è stato calcolato l'alpha di Cronbach (o coefficiente alpha), per misurare il grado di coerenza interna di questa parte del questionario, e dunque la sua affidabilità e riproducibilità (Dörnyei, Dewaele, 2023: 94)¹⁹. Sono state, poi, calcolate le distribuzioni di risposta relative ai due tipi di

¹⁹ Secondo gli autori, e in linea con gli standard internazionali, il coefficiente di Cronbach, che va da 0 a 1, dovrebbe raggiungere una soglia minima di 0.60, e viene considerato ottimo se supera la soglia dello 0.80.

rispondenti più presenti all'interno del campione, ovvero docenti di italiano L2/LS attualmente in servizio e studenti che si stanno formando per diventarlo.

Le risposte date alla domanda aperta obbligatoria “Quali sono le maggiori difficoltà che immagini di avere nell'uso dei corpora a fini didattici?” sono state analizzate attraverso i criteri dell'analisi tematica, che mira a individuare e analizzare le regolarità concettuali presenti all'interno di dati qualitativi (Clarke, Braun, 2017). Nello specifico, sono stati creati dei codici lessicali, corrispondenti a concetti chiave, a partire dall'analisi delle risposte fornite dai rispondenti. Ciascun codice è stato, poi, associato a un tema più generale. L'annotazione di ciascuna porzione di risposta con un codice e un tema ha poi condotto a una quantificazione per ciascuna delle due tipologie di annotazione, analizzate, in seguito, anche sul piano qualitativo.

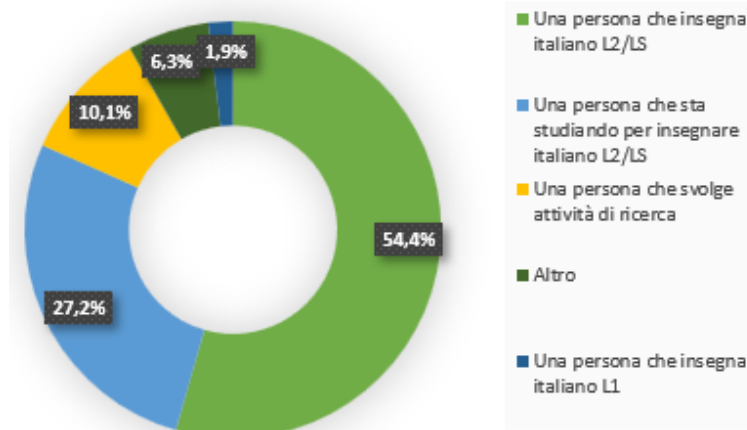
Infine, le risposte date alla domanda aperta facoltativa “Hai altri suggerimenti per la realizzazione della risorsa?” sono state analizzate manualmente considerando, inizialmente, il numero totale di risposte fornite, poi le risposte negative o similari, e infine, approfondendo quelle articolate, con relativa analisi.

8.2. Risultati e discussione

8.2.1. Il campione dei rispondenti

Il campionamento dei rispondenti al questionario è avvenuto in prevalenza nell'ambito di attività di formazione destinate a docenti, ed è per questo motivo che è proprio tale categoria a risultare preponderante. Come si può notare nel Grafico 1, infatti, i rispondenti al questionario rientrano, nel 54,4% dei casi (n=86) nella categoria delle persone che insegnano italiano L2/LS. Questa categoria è poi seguita da quella delle persone che stanno studiando per insegnare italiano L2/LS (27,2%; n=43), delle persone che svolgono attività di ricerca (10,1%; n=16), delle persone che insegnano italiano L1 (1,9%; n=3), e delle persone che rientrano in altre categorie (6,3%). All'interno di quest'ultimo gruppo, alla domanda aggiuntiva “Di che cosa ti occupi?”, i rispondenti hanno risposto come segue: insegnamento scuola primaria (n=2), insegnamento scuola media superiore (n=1), insegnamento – non meglio specificato (n=1), traduzione e interpretariato (n=1), insegnamento e ricerca sulla didattica dell'italiano L2 (n=1), insegnamento e ricerca sulla didattica dell'italiano L2, traduzione, attività editoriali (n=1), studio per insegnamento di e ricerca su italiano L2 (n=1), non specificato (n=2).

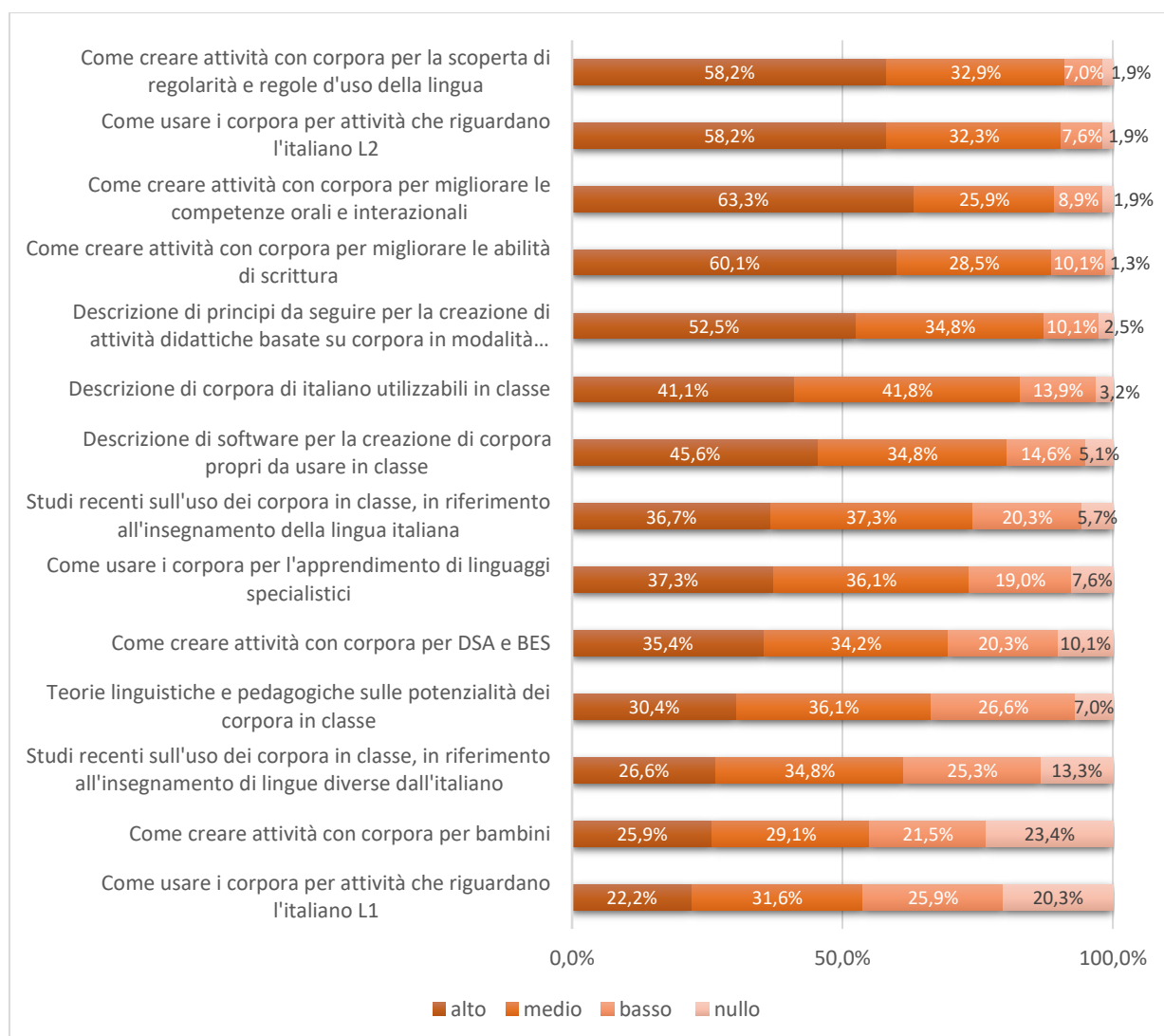
Grafico 1. *Il campione dei rispondenti*



8.2.2. Cosa imparare

La prima sezione del questionario contenente item a scala Likert riguarda la domanda “Cosa vuoi imparare? Seleziona il tuo grado di bisogno formativo relativo ai seguenti temi”, con a seguire una lista di 14 item. L’alpha di Cronbach per questa sezione del questionario è risultata essere 0.88, indicando un ottimo livello di affidabilità. La distribuzione delle risposte è rappresentata nel Grafico 2. Il grafico considera l’intero campione di rispondenti e segue un ordine decrescente rispetto al grado medio-alto del bisogno formativo.

Grafico 2. *Bisogni formativi rispetto a diverse tematiche del DDL - campione complessivo*



Come si può notare, al primo posto troviamo “Come creare attività con corpora per la scoperta guidata di regolarità e regole d’uso della lingua”, con il 58,2% di rispondenti che hanno selezionato di avere un “alto” grado di bisogno formativo rispetto a questa area. Questo risultato farebbe pensare a una generale sensibilità della maggior parte dei rispondenti rispetto alle peculiarità degli usi reali e concreti della lingua. A seguire, troviamo un interesse spiccato verso l’uso dei corpora nell’ambito dell’insegnamento dell’italiano L2, il che riflette senz’altro lo sbilanciamento interno del campione e, a

seguire, con la percentuale di rispondenti più alta fra tutte, ovvero il 63,3%, un interesse per la creazione di attività con corpora per migliorare le competenze orali e interazionali. Si tratta di un dato significativo. Gli usi dei corpora nella didattica delle lingue, che finora si sono osservati su scala mondiale, sono stati legati in misura preponderante allo sviluppo delle abilità orali solo in una fase iniziale, per poi estendersi nel tempo su aspetti legati al lessico e alle combinazioni di parole (Boulton, Vyatkina, 2021: 79). Per quanto riguarda la situazione italiana, l'attenzione allo sviluppo delle competenze orali e interazionali nell'uso didattico dei corpora è stata finora molto limitata. Ciò è probabilmente dovuto alla scarsissima disponibilità di corpora orali e multimediali, per l'italiano così come per le altre lingue, notoriamente di più difficile costruzione rispetto ai corpora di testi scritti.

La pubblicazione, in tempi recenti, di corpora orali di italiano, come il KiParla (Mauri *et al.*, 2019) e il CIP (Spina *et al.*, 2021), potrebbe incrementare l'attenzione, ma anche le opportunità, di chi si dedica o si vorrebbe dedicare all'uso di questa tipologia di corpora in attività DDL. Ciò si allineerebbe ai bisogni formativi rilevati nel nostro studio, andando dunque a colmare una lacuna che senz'altro è attualmente presente nel panorama del DDL per l'apprendimento dell'italiano.

A seguire, troviamo un bisogno formativo importante legato all'uso dei corpora per migliorare le abilità di scrittura, che raggiunge il 60,1% del grado "alto". In questo caso, i corpora attualmente disponibili per la lingua italiana possono sicuramente costituire una base valida per lo sviluppo di materiali DDL e per l'esplorazione autonoma da parte degli apprendenti, soprattutto in attività di autocorrezione delle proprie produzioni scritte (e.g., Forti, Spina, 2019). Un interesse più contenuto, ma pur sempre rilevante, si osserva anche per la formazione relativa alla modalità digitale e a quella cartacea che possono caratterizzare gli usi dei corpora in classe, nonché per la descrizione di corpora utilizzabili in classe e di software per creare corpora propri, per gli studi recenti relativi all'uso dei corpora nella didattica dell'italiano, e per quelli relativi all'impiego di corpora per l'apprendimento di linguaggi specialistici.

Un interesse progressivamente più ridotto si rileva, invece, per l'area dei BES e DSA, per le teorie linguistiche e pedagogiche a fondamento del DDL, per gli studi recenti relativi al DDL per lingue diverse dall'italiano, per l'uso dei corpora con bambini, e per l'uso dei corpora per l'apprendimento dell'italiano L1. Se gli ultimi due aspetti potrebbero rappresentare un effetto dello sbilanciamento interno del campione, che contiene una prevalenza di insegnanti di italiano L2 e studenti che studiano per diventare insegnanti di italiano L2, gli altri aspetti potrebbero essere interpretati come una conseguenza del fatto che certe aree e certi contesti dell'apprendimento sono ancora poco associabili al DDL. Ciò è dovuto anche al fatto che molte aree del DDL sono ancora in fase di sviluppo, come nel caso della sua applicabilità ad apprendenti con BES e DSA o a bambini.

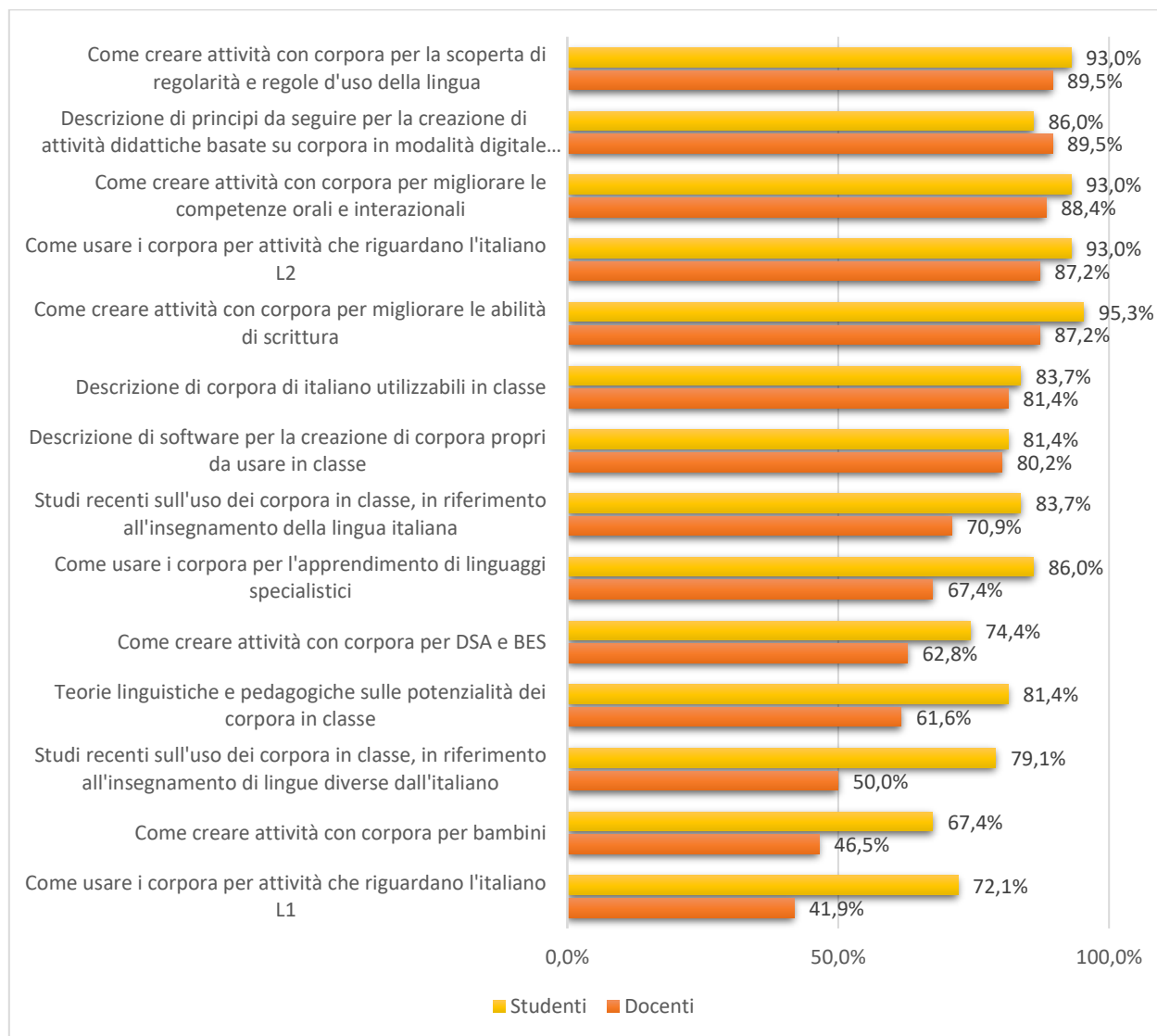
Il Grafico 3 mostra un confronto dei dati percentuale relativi ai gradi di bisogno medio-alti percepiti dagli insegnanti di italiano L2 rispetto a quelli percepiti da studenti iscritti a corsi di laurea che stanno studiando per diventare insegnanti di italiano L2. Le differenze maggiori che questo confronto permette di osservare riguardano i seguenti item:

1. "Come usare i corpora per attività che riguardano l'italiano L1";
2. "Studi recenti relativi all'uso dei corpora in classe, in riferimento all'insegnamento di lingue diverse dall'italiano";
3. "Come creare attività con corpora per bambini".

In ognuno di questi casi, sono gli studenti a mostrare un grado di bisogno formativo superiore rispetto a quello degli insegnanti, rispettivamente del +30,2%, del +29,1% e del +20,9%. Se da un lato il *bias* derivante dalla natura dei campioni crea inevitabilmente un'influenza sulla natura delle risposte date, nel caso degli studenti, che risultano iscritti a un corso di laurea che mira a formare docenti di italiano a stranieri, questo *bias* sembra

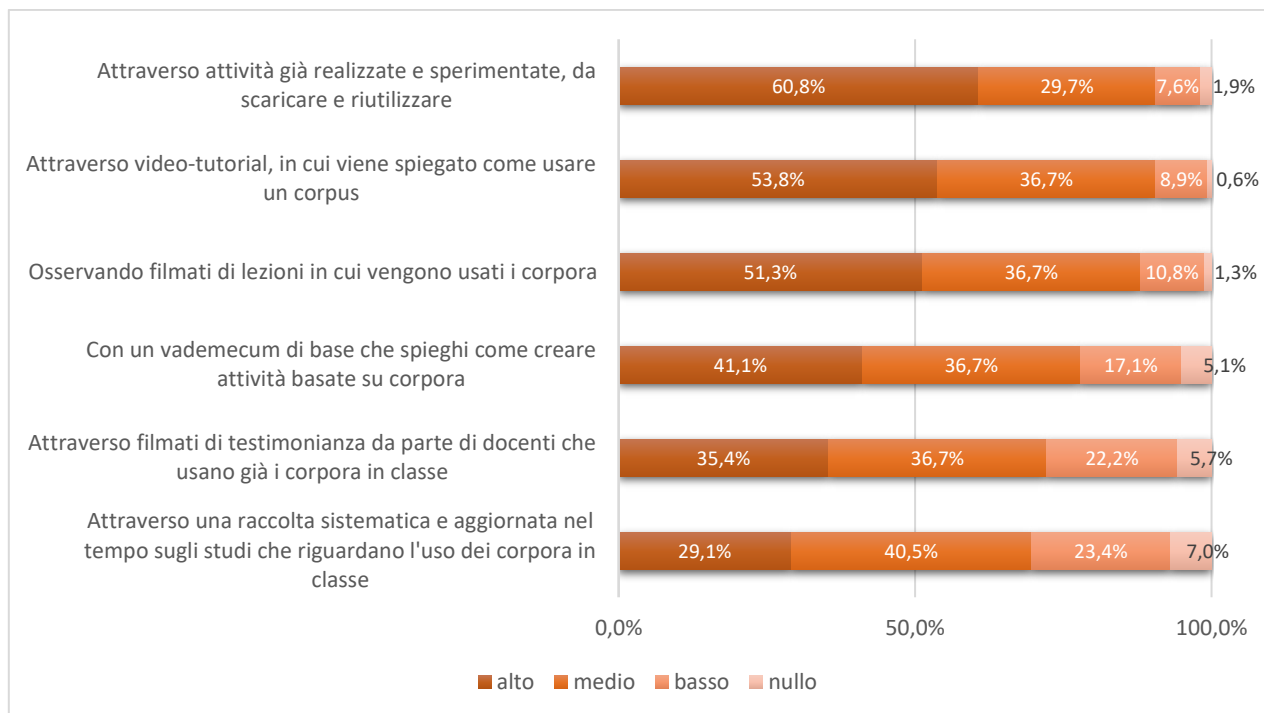
ridotto, rendendoli più aperti a temi del DDL non strettamente connessi al loro percorso accademico.

Grafico 3. *Bisogni formativi medio-alti rispetto a diverse tematiche del DDL - studenti vs. docenti*



8.2.3. *Come imparare*

La seconda sezione del questionario contenente item a scala Likert riguardanti la domanda “Come vuoi imparare? Seleziona il tuo grado di preferenza relativo alle seguenti modalità”, con a seguire una lista di 6 item. L’alpha di Cronbach per questa sezione del questionario è risultata essere 0.70, indicando un buon livello di affidabilità. Il Grafico 4 illustra la distribuzione dei gradi “alto”, “medio”, “basso”, e “nullo” relativi alle preferenze dei rispondenti, seguendo, anche in questo caso, un ordine decrescente dei gradi medio-alti.

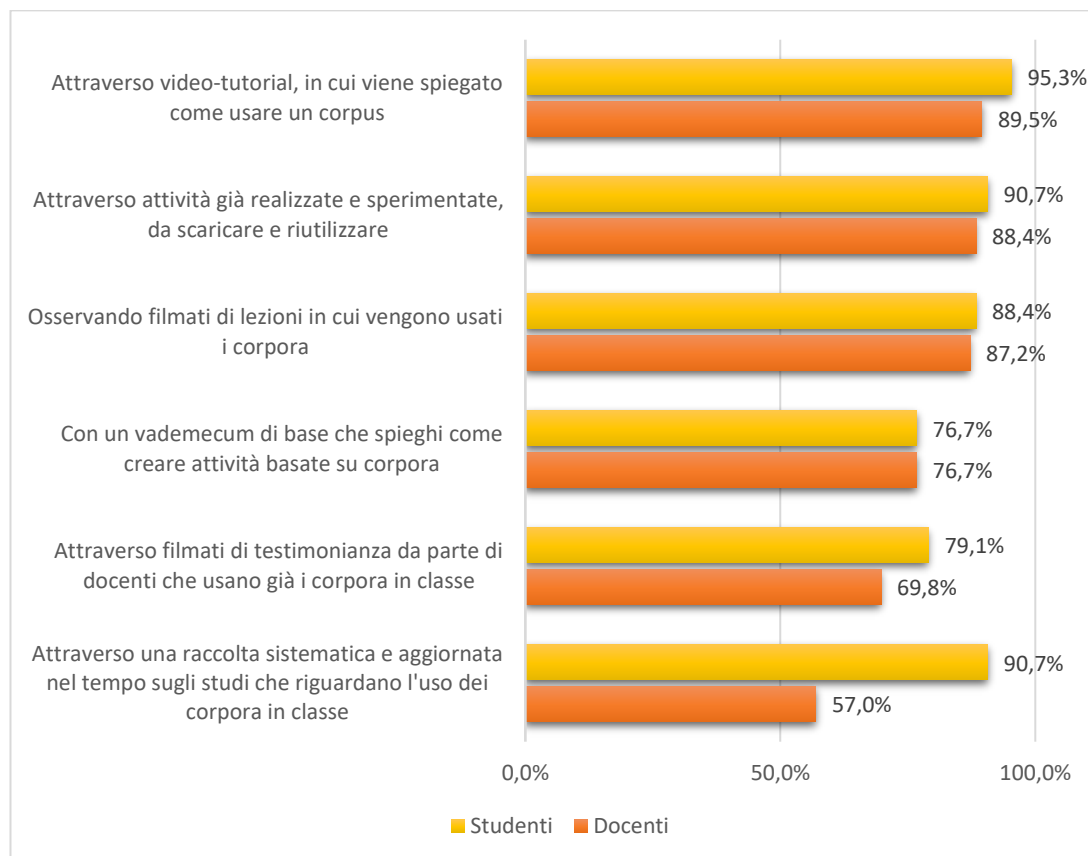
Grafico 4. *Bisogni formativi rispetto alle modalità di svolgimento della formazione - campione complessivo*

Come si può notare, la modalità preferita è quella che riguarda la possibilità di avere a disposizione attività già realizzate e sperimentate, da scaricare e riutilizzare. Complessivamente, questa modalità raggiunge il 60,8% delle preferenze alte, e il 90,5% delle preferenze medio-alte. A seguire, troviamo la preferenza per video-tutorial, in cui venga spiegato e dimostrato come usare un corpus (53,8% delle preferenze alte) e la preferenza per l'osservazione di filmati di lezioni in cui si usano i corpora (51,3% delle preferenze alte). I video-tutorial in cui viene spiegato come usare un corpus in classe sono diversi e facilmente accessibili su canali come YouTube. La maggior parte di queste risorse, tuttavia, sono in lingua inglese e riguardano la didattica della lingua inglese. Mancano, allo stato attuale, risorse di questo tipo specificamente indirizzate a persone coinvolte nella didattica dell'italiano. Lo stesso discorso vale per filmati che consentano l'osservazione di lezioni in cui vengono usati i corpora. L'osservazione costituisce una delle basi fondamentali della formazione dei docenti, così come degli studenti interessati a intraprendere questa professione. Se le attività di osservazione rappresentano già una prassi consolidata in tal senso, integrare le pratiche di osservazione con la realizzazione di filmati di lezioni DDL ne rappresenterebbe un naturale sviluppo. A seguire troviamo la possibilità di osservare filmati di testimonianza da parte di docenti che già usano i corpora in classe. La modalità meno preferita dal campione, invece, riguarda la possibilità di disporre di una raccolta sistematica e aggiornata nel tempo sugli studi che riguardano l'uso dei corpora in classe. Per questa modalità viene comunque indicato un grado di preferenza medio-alta complessiva del 69,6%, tutt'altro che trascurabile. Tuttavia, rispetto alle altre modalità proposte, risulta la meno preferita, probabilmente per il suo carattere meno immediato rispetto alle altre.

Anche per questa sezione del questionario, è stata svolto confronto tra insegnanti di italiano L2 e studenti che stanno studiando per diventarlo. Focalizzando, nuovamente, l'attenzione soltanto sui bisogni medio-alti, il Grafico 5 illustra i risultati per i due gruppi. Anche in questo caso, si notano alcune differenze. La più rilevante è quella che riguarda

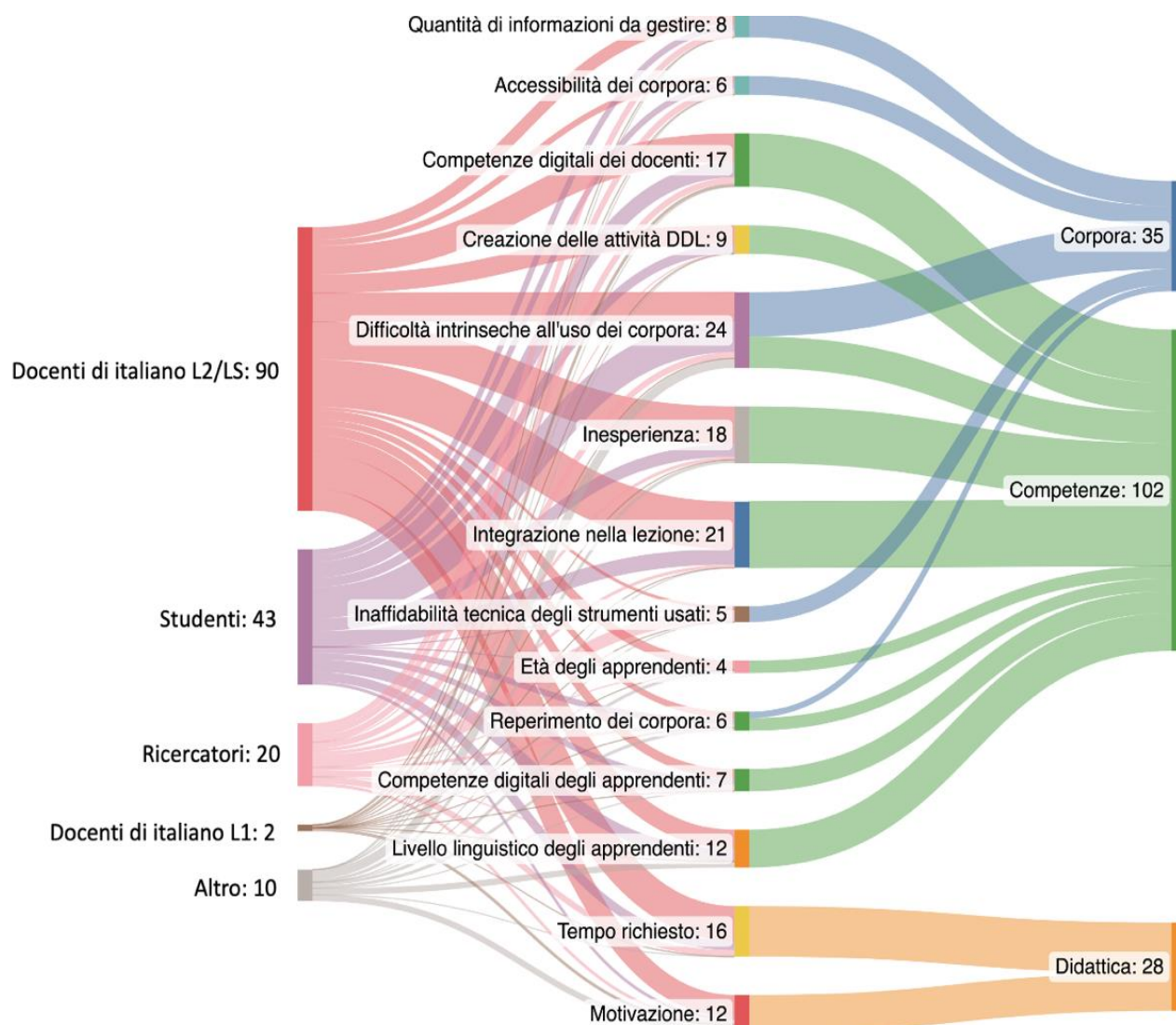
proprio la possibilità di disporre di una raccolta sistematica e aggiornata nel tempo sugli studi che riguardano l'uso dei corpora in classe. Gli studenti indicano una preferenza medio-alta del 90,7%, mentre gli insegnanti del 57,0%. La differenza risultante è del 33,7%. Tutte le altre voci riportano dati piuttosto bilanciati.

Grafico 5. *Preferenze medio-alte rispetto alle modalità di svolgimento della formazione – studenti vs. docenti*



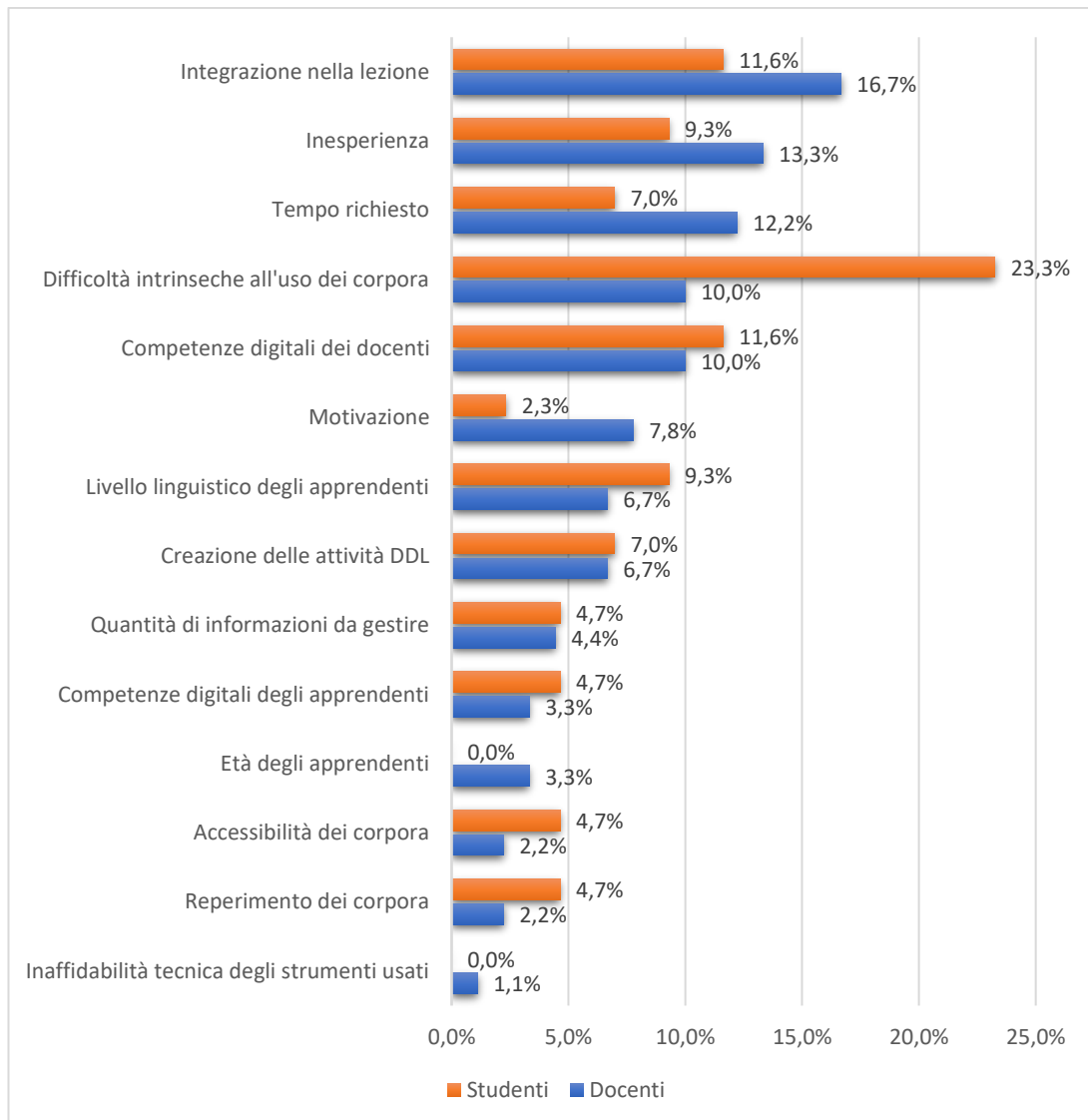
8.2.4. Potenziali difficoltà percepite

Con la domanda a risposta aperta “Quali sono le maggiori difficoltà che immagini di avere nell’uso dei corpora a fini didattici?” sono state raccolte 158 risposte, un numero pari al totale dei membri del campione coinvolto nell’indagine, trattandosi di una domanda a risposta obbligatoria. In seguito all’analisi degli enunciati raccolti, si arriva ad ottenere un totale di 189 porzioni di testo codificate tramite l’associazione di codici tematici, sulla base di quanto illustrato nel § 8.1.2. Nel diagramma di Sankey riportato nel Grafico 6 viene mostrata la quantificazione dei codici, con raggruppamenti sia rispetto alla tipologia di partecipanti, sia al tema in cui i diversi codici confluiscono. La quantificazione numerica riportata accanto a ogni tipologia di partecipante si riferisce alle porzioni testuali associate a un determinato codice. Nel diagramma sono stati esclusi i codici “Non specificato”, “Nessuna”, “Non sa”, “Riferimento vago” e “Non pertinente” (complessivamente 24 occorrenze), giungendo dunque a un totale di 165 porzioni di testo rappresentate. Il Grafico 7 riporta la distribuzione percentuale dei codici all’interno dei due gruppi di partecipanti prevalenti, ovvero studenti che si stanno formando per diventare docenti di italiano L2/LS e docenti di italiano L2/LS in servizio.

Grafico 6. *Potenziati difficoltà percepite: campione complessivo*

Il tema maggiormente rappresentato risulta essere quello relativo alle competenze. In 102 casi su 165, infatti, il campione attribuisce alla carenza delle necessarie competenze la principale difficoltà potenziale nell'usare i corpora in classe. All'interno di questo tema, troviamo innanzitutto la difficoltà di integrare attività svolte con i corpora nel contesto della lezione o unità didattica. Nel Grafico 7 osserviamo che questo tipo di difficoltà è maggiormente percepita dai docenti (16,7% vs. 11,6%).

Questa difficoltà potenziale emersa, legata ad un'integrazione efficace delle attività coi corpora alla lezione, rimarca l'importanza di progettare un ambiente condiviso che sia in grado di proporsi come risorsa aperta, destinata alla formazione e al perfezionamento delle competenze dei docenti di lingua a questo riguardo. A seguire, troviamo i codici "Inesperienza" (18 occorrenze) e subito dopo "Competenze digitali dei docenti" (17 occorrenze). Nel primo caso, sono i docenti in servizio a prevalere (13,3% vs. 9,3%), mentre nel secondo caso i valori sono più simili (studenti: 11,6% vs. docenti: 10,0%).

Grafico 7. *Potenziati difficoltà percepite: studenti vs. docenti (codici)*

I dati raccolti in riferimento all'“Inesperienza” riguardo all'uso dei corpora e dei necessari strumenti implicati, ci mostrano indirettamente che gli insegnanti e gli aspiranti insegnanti di italiano L2 coinvolti nell'indagine non hanno avuto modo di beneficiare di una formazione al riguardo nel loro passato formativo. Persiste una grande distanza tra questa categoria professionale e la pratica coi corpora; inoltre, i risultati della ricerca scientifica appaiono ancora molto lontani dalla pratica didattica come diverse ricerche sia in ambito internazionale che nazionale evidenziano (Forti, Spina, 2019). Riguardo al codice “Competenze digitali dei docenti”, il quadro di riferimento europeo per le competenze digitali dei docenti e dei formatori, il DigCompEdu (Bocconi *et al.*, 2018), costituisce un documento essenziale, in grado di guidare lo sviluppo di tali competenze, anche in riferimento all'uso di strumenti specifici come, ad esempio, i software necessari all'uso dei corpora.

Il “Livello linguistico degli apprendenti” rappresenta una difficoltà percepita che ricorre 12 volte. Nella maggior parte dei casi i rispondenti immaginano la difficoltà che incontrerebbero se utilizzassero i corpora in classi dal livello linguistico tendenzialmente basso. Nei casi restanti, le possibili criticità vengono ricondotte alle operazioni di

adattamento al livello target degli apprendenti. Questa problematica viene segnalata sia dai docenti in servizio che da quelli in formazione, sebbene con qualche differenza percentuale (studenti: 9,3% vs. docenti: 6,7%). La difficoltà di condurre attività DDL in classi dal livello linguistico tendenzialmente basso è una percezione fondata, che si lega alle ridotte evidenze empiriche provenienti dalla ricerca DDL, che tuttavia, come dimostrato, non ne limitano le potenzialità (Crosthwaite, 2020). L'adattamento al livello richiede infatti al docente uno sforzo aggiuntivo ma che – sulla scia di quanto si commenterà più avanti per l'analisi del codice “Tempo richiesto” – non solo rientrerebbe appieno tra le sue “funzioni creative”, ma gli darebbe anche la possibilità di espandere le sue competenze legate alla progettazione di materiale proprio grazie all'uso dei corpora (Michelini, 2022).

I codici rimanenti che rientrano nel tema “Competenze” riportano un numero di occorrenze minore a 10: “Competenze digitali degli apprendenti” (7 occorrenze); “Età degli apprendenti” (4 occorrenze, parzialmente assimilabili alla percepita difficoltà del livello linguistico); “Creazione delle attività DDL” (9 occorrenze). I codici “Reperimento dei corpora” e “Difficoltà intrinseche all'uso dei corpora” si suddividono in due temi: da un lato quello delle competenze, dall'altro quello dei corpora.

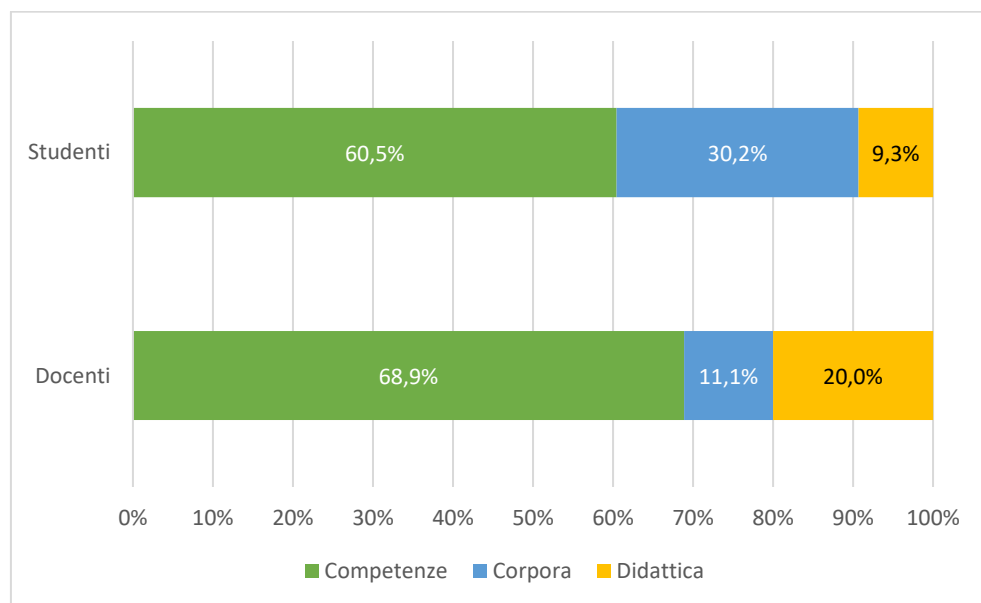
Il tema “Corpora” è quello più presente dopo “Competenze”, sia pure con un distacco molto netto, riportando 35 occorrenze. Il codice maggiormente presente è quello relativo alle “Difficoltà intrinseche all'uso dei corpora”, con 24 occorrenze, quasi equamente suddivise tra il tema “Corpora” e il tema “Competenze”. Confluiscono nel primo tema risposte come “non tutti i corpora sono adatti per l'uso in ambito didattico” (Id. 128), in cui si fa leva sui limiti di adattabilità pedagogica di alcuni corpora e, nel secondo, risposte come “scarse abilità metodologiche per utilizzare corpora in modo efficace, come la selezione e l'interpretazione dei dati, ecc.” (Id. 126), in cui invece il focus è sulle competenze di chi vorrebbe usare i corpora. Nel confronto tra docenti in servizio e studenti, questo codice si concentra in misura sensibilmente maggiore nella percezione espressa dagli studenti (studenti: 23,3%; docenti: 10,0%). Le principali criticità potenziali legate alle “Difficoltà intrinseche all'uso dei corpora” sono riconducibili a due tipi di contenuto principali: da una parte, a difficoltà legate alle modalità di interrogazione dei corpora per la ricerca delle informazioni desiderate; dall'altra, a criticità riguardanti un'interpretazione efficace dei dati ottenuti. Emerge dunque la necessità non solo di sviluppare una formazione efficace e periodica per favorire la familiarizzazione con l'uso dei corpora, ma anche quella di crearne di più agevoli da usare in ambito didattico. Infatti, nel tema “Corpora” confluiscono anche i codici “Quantità di informazioni da gestire” (8 occorrenze), “Accessibilità dei corpora” (6 occorrenze), “Inaffidabilità tecnica degli strumenti utilizzati” (5 occorrenze) e “Reperimento dei corpora” (2 occorrenze). Anche in quest'ultimo caso osserviamo una bipartizione: il codice confluisce nel tema “Corpora” nei casi in cui emerge una difficoltà legata alla scarsa disponibilità di corpora specifici per la un certo tipo di didattica (es. corpora multimediali), e nel tema “Competenze”, quando invece emerge una scarsa abilità nel reperire corpora esistenti e adatti alla didattica. Come si può vedere dal Grafico 7, le difficoltà percepite in relazione alla quantità di informazioni da gestire sono distribuite in misura piuttosto omogenea (studenti: 4,7%; docenti: 4,4%), e in misura esclusiva riguardo all'inaffidabilità tecnica, indicata solo dai docenti in servizio. Questo aspetto si lega all'uso dei corpora online (ovvero al DDL “computer-based”) e dunque anche alla necessità di disporre di infrastrutture informatiche adeguate.

Il tema che riporta un numero complessivo minore di occorrenze, ma comparabili al tema “Corpora”, è quello della “Didattica”, con 28 occorrenze. All'interno di questo tema confluiscono due codici: “Tempo richiesto” (16 occorrenze) e “Motivazione” (12 occorrenze). Nel primo caso, il riferimento è al tempo richiesto per la preparazione delle attività in DDL che, nella maggior parte dei casi, viene percepito il modo problematico,

soprattutto dai docenti in servizio (Grafico 7, 12,2%), e questo dato potrebbe allinearsi a quanto emerso nella categoria “Creazione attività DDL” che, seppure non abbia un numero di occorrenze cospicuo, è comunque presente. In altri casi, i rispondenti fanno invece riferimento al tempo necessario ad acquisire dimestichezza con gli strumenti e le procedure coinvolte nell’interrogazione dei corpora. Un rispondente introduce un punto di vista ben definito: “Dal punto di vista tecnico, nessuna, almeno dopo avere imparato a usare i corpora. Per un insegnante il problema è sempre uno solo, il rapporto impegno-stipendio, di solito sempre parecchio sbilanciato sul primo fattore [...]” (Id. 30). Il “tempo del docente” (Michelini, 2024) è infatti un tempo rapido, che include funzioni di tipo diverso, tra cui quella legata alla creazione e all’adattamento al livello di materiale didattico autentico, che gli appartiene in modo naturale. Per sostenere questa funzione, occorrerebbe riconoscere appieno la “funzione creativa” del docente, magari creando spazi formalizzati dedicati alla progettazione didattica, auspicabilmente condivisa, e dove possa anche trovare spazio una formazione mirata a supporto a supporto della progettazione di attività didattiche di qualità, anche grazie all’uso dei corpora. Riguardo alla difficoltà legata alla motivazione, anche in questo caso percepita soprattutto dai docenti in servizio (Grafico 7, 7,8%), l’ostacolo, da un lato, sarebbe costituito da alcune caratteristiche specifiche di certi tipi di apprendenti come ad esempio gli adolescenti e le possibili distrazioni in cui incapperebbero tramite l’uso di dispositivi tecnologici; dall’altro, i docenti fanno riferimento a loro stessi e alla possibile difficoltà di motivare gli studenti all’uso dei corpora. In un caso, tra le varie criticità legate alla motivazione si menziona il cambiamento delle abitudini di studio che l’uso dei corpora implicherebbe.

Il Grafico 8 mostra la distribuzione complessiva dei temi rispetto alle due categorie di partecipanti prevalenti, ovvero docenti in servizio e studenti.

Grafico 8. *Potenziiali difficoltà percepite: studenti vs. docenti (temi)*



Come si può notare, il bilanciamento maggiore si ha in relazione alle difficoltà potenziali percepite legate alle competenze necessarie, con lieve prevalenza nel gruppo dei docenti (+8,4%). A seguire troviamo quelle legate ai corpora, in cui prevalgono gli studenti (+19,1%), e infine quelle legate alla didattica, in cui, invece, sono maggiormente presenti i docenti (+10,7%). Questi dati sembrerebbero indicare una percezione parzialmente

diversa delle potenziali difficoltà legate all'uso dei corpora in classe, e potrebbero stimolare una riflessione su possibili percorsi formativi specifici per le due categorie.

8.2.5. *Suggerimenti*

In questa sezione, verrà analizzata la parte del questionario denominata “altri suggerimenti”. L'analisi di questa sezione è di fondamentale importanza poiché, indipendentemente dal gruppo a cui appartengono i rispondenti, può offrire una visione chiara riguardo alla conoscenza e all'uso dei corpora nella didattica. In particolare, essa consente di comprendere la consapevolezza degli intervistati su ciò che è possibile realizzare attraverso l'utilizzo dei corpora e di individuare eventuali necessità di approfondimento su questo tema. È importante sottolineare inoltre, che al questionario hanno risposto docenti, studenti e futuri docenti in formazione. Nel campione totale di 158 rispondenti, la sezione “altri suggerimenti” è stata compilata da 56 partecipanti. I commenti che, oltre alle risposte di tipo “no”, meritano attenzione verranno analizzati di seguito, considerando i fabbisogni di conoscenza più o meno approfonditi emersi.

Le risposte fornite dai docenti ammontano a 42 e si suddividono in due categorie principali: una riguarda le attività indirizzate agli studenti basate sull'uso dei corpora, l'altra riguarda le loro conoscenze dei corpora come docenti. Di seguito alcuni dei commenti più significativi in riferimento al presente studio:

1. Nessuno al momento, ho bisogno di maggiore approfondimento e studio.
2. La selezione dei corpora dovrebbe seguire il QCER (Quadro Comune Europeo di Riferimento per le Lingue) per aiutare gli insegnanti a scegliere i corpora più adatti.
3. Sarei interessata ai corpora orali, poiché gli studenti che vivono fuori dall'Italia hanno poche occasioni di ascoltare la lingua reale (cadenze dei parlanti, linguaggio parlato, ecc.).
4. Linguaggi specialistici per le persone che lavorano in ambiti specifici (contabilità, risorse umane, customer service, ecc.).
5. Sarebbe interessante interrogare i corpora per lavorare sugli aspetti fonetici.
6. Esplicitare le funzionalità nelle interrogazioni del corpus.
7. Fornire istruzioni adeguate anche per persone con limitate competenze tecnologiche.
8. Trovare un equilibrio tra corpora di natura diversa (audio-visivi, multimediali, cartacei, ecc.).

Le risposte 1, 6 e 7 evidenziano una chiara esigenza da parte dei docenti di una formazione più approfondita nell'uso dei corpora, sia per quanto riguarda l'aspetto teorico che pratico. Le risposte fornite dagli studenti e dagli studenti che si stanno formando per diventare docenti ammontano a 9. Tra queste, quelle degne di nota sono le seguenti:

9. Per la realizzazione della risorsa, bisogna innanzitutto definire gli obiettivi, cioè, stabilire le competenze linguistiche che si desidera sviluppare. È importante incoraggiare gli studenti a utilizzare il corpus in modo autonomo, sviluppando abilità critiche e metodologiche, e proporre

attività guidate come esercizi che favoriscano la comprensione della modalità di ricerca e di analisi del corpus.

10. Creare una piattaforma 'Insegna con i corpora', che permetta di creare percorsi personalizzati di Unità di Apprendimento (UdA) con i corpora e ottenere risorse finali.

Queste risposte, precise e puntuali, denotano una buona conoscenza dei corpora, delle loro potenzialità didattiche e una consapevolezza chiara riguardo ciò che sarebbe necessario fare per utilizzarli in modo efficace. Nella risposta n. 9 viene posta, inoltre, particolare attenzione alle pratiche volte a favorire autonomia dello studente nell'uso dei corpora, un aspetto che spesso viene trascurato.

Cosa emerge da queste risposte? In primo luogo, emerge chiaramente un bisogno di formazione da parte di coloro che hanno risposto al questionario ed hanno risposto alla sezione "altri suggerimenti". L'uso dei corpora per progettare una didattica efficace richiede non solo una solida conoscenza teorica del campo, ma anche competenze pratiche nell'utilizzo degli strumenti. Il bisogno di formazione riguarda non solo la conoscenza dei corpora e il loro utilizzo, ma anche la progettazione di attività DDL. Alcune risposte, come le risposte n. 3, n. 4 e n. 5, tra quelle fornite dai docenti, si riferiscono a settori specifici come l'oralità, i linguaggi specialistici e gli aspetti fonetici della lingua.

Questi temi risultano altrettanto rilevanti, poiché, per quanto riguarda l'oralità, ad esempio, si evidenzia l'importanza dell'aspetto pragmatico della lingua. Tale aspetto spesso trascurato o applicato a pochi contesti comunicativi, è tuttavia fondamentale per l'acquisizione di competenze chiave nel processo di sviluppo delle abilità comunicative, attraverso le quali è possibile osservare come e per quali scopi la lingua viene utilizzata. È indubbio che potere sfruttare attività DDL che vertano su questo aspetto si potrebbe fornire agli apprendenti tutte quelle chiavi atte alla conoscenza e anche alla padronanza di tutti quegli aspetti extralinguistici (sociali, ambientali e psicologici) che influenzano gli atti linguistici.

Per quanto concerne la necessità di strutturare attività DDL riferite ai linguaggi specialistici, anche questo tema merita un approfondimento. Esistono numerosi settori come, ad esempio, in ambito lavorativo e universitario, in cui apprendenti manifestano un bisogno di acquisire competenze linguistiche specifiche. Questo tipo di apprendimento consente di acquisire il lessico pertinente e rispondere alle necessità linguistiche legate alle specifiche esigenze professionali o accademiche. Relativamente agli aspetti fonetici, si evidenzia ancora una volta l'importanza della pragmatica e inoltre, della grammatica, sottolineando la necessità di focalizzarsi su quegli aspetti della lingua che, troppo spesso, vengono trascurati o trattati solo marginalmente, ma che, si manifestano chiaramente nell'apprendente durante il suo percorso di apprendimento.

Tutti questi aspetti denotano, senza dubbio, una certa consapevolezza, da parte dei docenti che hanno risposto, riguardo a elementi che coinvolgono la lingua. Si evidenzia la loro predisposizione a rilevare i bisogni linguistico-comunicativi degli apprendenti, riconoscendo la necessità di farlo in modo puntuale. Inoltre, emerge la volontà di intraprendere un percorso di insegnamento-apprendimento finalizzato ad ampliare le proprie conoscenze e capacità, per progettare attività basate su metodologie di insegnamento di tipo induttivo, metodologie che, come confermato da numerosi studi, si rivelano essere di maggiore successo.

Nonostante quanto evidenziato, si rileva quindi un bisogno di formazione, e ancora prima, di una conoscenza adeguata di che cosa siano i corpora. Questa peculiarità emerge anche dal fatto che molte risposte non sono state date o, quando fornite, si limitano a indicare la mancanza di suggerimenti concreti. I suggerimenti provengono da chi ha una

conoscenza almeno parziale della materia e intravede la possibilità di espandere le proprie competenze didattiche nell'insegnamento della lingua. Queste competenze, si ritiene sia necessario sottolineare, possono essere sviluppate attraverso la condivisione di attività pratiche che sono volte a favorire l'apprendimento esperienziale.

Infine, un aspetto emerso in una delle due risposte fornite dagli studenti che si stanno formando per diventare docenti evidenzia la necessità di creare una piattaforma per condividere attività già sperimentate e di successo. Questo sottolinea l'importanza di un lavoro basato sulla condivisione, il confronto e la ricerca sul campo, elementi imprescindibili per una buona pratica didattica.

9. CONCLUSIONI

La ricerca basata sui corpora sta acquisendo sempre più importanza in tutte le branche della linguistica, e le sue ramificazioni sono in fase espansiva. I principali progressi sono dati dal fatto che ci si renda conto che i corpora siano sì, raccolte di dati, ma che le analisi condotte su questi stessi dati possono assumere forme e combinazioni differenti, portando allo sviluppo di nuove teorie. Da qui consegue che l'approccio DDL sia una modalità di apprendimento e insegnamento molto preziosa, caratterizzata all'uso di uno strumento che crea un focus sulla lingua autentica, sull'apprendimento mirato di suoi numerosi aspetti. Tutto questo è altamente compatibile con gli obiettivi attuali e desiderabili, propri della pedagogia linguistica.

Questo studio ha messo in evidenza come i corpora possano essere fondamentali in due diverse aree: nella ricerca linguistica e le sue applicazioni, ma soprattutto nella ricerca di un approccio efficace da utilizzare per insegnare e apprendere una lingua. In questo caso, lo studio cerca di porre le premesse per la creazione e lo sviluppo di una risorsa educativa aperta, che possa essere una guida ben progettata, efficace e mirata, grazie ai feedback da parte dei fruitori della stessa. Già Breyer (2011), nel suo *Corpora in Language Teaching and Learning: Potential, Evaluation, Challenges*, traendo le sue conclusioni riguardo l'utilizzo dei corpora nella glottodidattica, indicava, tra i punti chiave, come fosse fondamentale il ruolo degli insegnanti nella diffusione e normalizzazione dell'uso dei corpora, ma sottolineava anche come il loro punto di vista a riguardo fosse ampiamente inesplorato. Da allora, gli studi su questo aspetto si sono susseguiti con una maggiore costanza, ma restano estremamente limitati nell'ambito della didattica dell'italiano.

Sappiamo quanto sia cruciale il fatto che qualunque approccio glottodidattico deve essere adottato e calibrato sulla tipologia dei discenti a cui si andranno ad insegnare i contenuti, in base alle condizioni generali di insegnamento. È per questo che la nostra fonte di conoscenza per approfondire l'argomento sono stati gli insegnanti stessi: di italiano L2-LS, aspiranti tali, ricercatori ed insegnanti di italiano come L1.

Gli ultimi paragrafi hanno avuto come focus l'analisi dei dati ottenuti. Abbiamo rilevato come al primo posto vi sia come interesse primario dei rispondenti quello di capire come creare attività con corpora per la scoperta guidata di regolarità e regole d'uso concrete della lingua; tra gli obiettivi quindi le abilità comunicative orali. Preferenza massima per quanto riguarda la modalità di apprendimento è stata data alla possibilità di avere a disposizione attività già realizzate e sperimentate, da scaricare e riutilizzare e, a seguire, videotutorial ed osservazioni di lezioni reali da cui attingere e in cui viene spiegato e dimostrato come usare un corpus.

Dall'analisi globale delle risposte raccolte, le potenziali difficoltà percepite riguardano il docente, l'apprendente e l'ente formativo. Il quadro riflette un gruppo di docenti e aspiranti tali non ancora pienamente formato a livello digitale; questo in riferimento non tanto alle competenze globali, quanto a quelle specifiche che entrerebbero in gioco

nell'interrogazione dei corpora. Queste sono motivo di timore soprattutto per il docente, che riconosce di non poter trasmettere senza prima saper padroneggiare.

Dei punti di attenzione possono quindi essere, sulla base dei dati raccolti e analizzati, a) creare una OER user-friendly, semplice e immediata da usare, b) differenziare le risorse per livelli e contesti di apprendimento, così da poter immediatamente fruire di un contenuto per le necessità del momento.

Prossimi passaggi fondamentali della ricerca presentata in questo studio saranno la progettazione della risorsa educativa stessa, seguita dalla creazione, dall'impiego e dal perfezionamento di attività, e la successiva messa in rete della risorsa, con futura valutazione del suo impatto in contesti educativi specifici.

RIFERIMENTI BIBLIOGRAFICI

- Aston G. (1996), "The British National Corpus as a language learning resource", in Botley S., Glass J., McEnery T., Wilson A. (eds.), *Proceedings of Teaching and language corpora 1996*, UCREL technical papers 9, UCREL, Lancaster, pp. 178-191.
- Bernardini S. (2003), "Corpora in the classroom: An overview and some reflections on future developments", in Sinclair J. (ed.), *How to Use Corpora in Language Teaching*, Benjamins, Philadelphia, pp. 15-38.
- Bocconi S., Earp J., Panesi S. (2018), *DigCompEdu. Il quadro di riferimento europeo sulle competenze digitali dei docenti. Istituto per le Tecnologie Didattiche, Consiglio Nazionale delle Ricerche (CNR), Roma.*
- Boulton A. (2016), "Integrating corpus tools and techniques in ESP courses", in *Asp*, 69, pp. 113-137.
- Boulton A., Cobb T. (2017), "Corpus use in language learning: A meta-analysis", in *Language Learning*, 67, 2, pp. 348-393.
- Boulton A., Vyatkina N. (2021), "Thirty years of data-driven learning: Taking stock and charting new directions over time", in *Language Learning and Technology*, 25, pp. 66-89.
- Breyer A. Y. (2011), *Corpora in Language Teaching and Learning: Potential, Evaluation, Challenges*, Peter Lang, Bern.
- Chambers A. (2019), "Towards the corpus revolution? Bridging the research-practice gap", in *Language Teaching*, 52, 4, pp. 460-475.
- Clarke V., Braun V. (2017), "Thematic analysis", in *The Journal of Positive Psychology*, 12, 3, pp. 297-298.
- Carter R., Hughes R., McCarthy M. (2000), *Exploring Grammar in Context: Upper-intermediate and Advanced*, Cambridge University Press, Cambridge.
- Carter R., McCarthy M. (2006¹), *Cambridge Grammar of English: A comprehensive Guide*, Cambridge University Press, Cambridge.
- Cesarini S. (1995), "L'uso degli schemi vuoti nella riflessione meta-linguistica", in *Educazione permanente*, 1-2, pp. 111-123.
- Clandfield L., Goldstein B., Hendra L. A. (2023 [2018]), *Evolve Level 2 Student's Book*, Cambridge University Press, Cambridge.
- Cobb T. (1999), "Applying constructivism: A test for the learner as scientist", in *Educational Technology Research & Development*, 47, 3, pp. 15-31.
- Corino E. (2019), "Data-driven Learning: tra lingue straniere e CLIL, tra ricerca e didattica", in *ELLE - Educazione linguistica. Language Education*, 8, 2, pp. 272-286.

- Corino E. (2022), “La Costituzione italiana, è ancora un testo facile?”, in Ferrari A., Lala L., Pecorari F. (a cura di), *L'italiano dei testi costituzionali: indagini linguistiche e testuali tra Svizzera e Italia*, Edizioni dell'Orso, Alessandria, pp. 293-318.
- Crosthwaite P. (ed.) (2020), *Data-driven learning for the next generation*, Routledge, London-New York.
- Curry N., Mark G. (2023), “Using Corpus Linguistics in Materials Development and Teacher Education”, in *Second Language Teacher Education*, 2, 2, pp. 187-208.
- Dörnyei Z., Dewaele J.-M. (2023³), *Questionnaires in Second Language Research. Construction, Administration, and Processing*, Routledge, London-New York.
- Farr F., Karlsen P. H. (2023), “DDL pedagogy, participants, and perspectives”, in Csomay, E., Jablonkai, R. R. (eds.), *The Routledge Handbook of Corpora and English Teaching and Learning*, Routledge, London-New York, pp. 329-343.
- Flowerdew L. (2009), “Applying corpus linguistics to pedagogy: a critical evaluation” in *International Journal of Corpus Linguistics*, 14, 3, pp. 393-417.
- Forti L. (2021), “Apprendere la grammatica attraverso il data-driven learning”, in Giunchi, P. Roccaforte M., *La grammatica tra acquisizione e apprendimento*, Carocci, Roma, pp. 100-113.
- Forti L. (2023), *Corpus Use in Italian Language Pedagogy. Exploring the Effects of Data-driven learning*, Routledge, London-New York.
- Forti L. (2024), “I corpora nella didattica dell'italiano L2. Una storia lunga trent'anni”, *Congresso DILLE 2024, Università della Basilicata (Campus di Matera), 16 – 18 maggio 2024* [presentazione a convegno]
- Forti L. (2024), “Proficiency-rated learner corpora: A promising resource for data-driven learning”, in *International Journal of Learner Corpus Research*, 10, 1, pp. 216-240.
- Forti L., Spina S. (2020), “Corpora for Linguists vs. Corpora for Learners. Bridging the Gap in Italian L2 Learning and Teaching”, in *ELLE - Educazione linguistica. Language Education*, 8, 2, pp. 349-362.
- Farr F., Karlsen P. H. (2023), “Corpus linguistics and Data Driven Learning (DDL): pedagogy, participants, perspectives”, in Jablonkai R., Csomay E. (eds.), *The Routledge Handbook of Corpora in English Language Teaching and Learning*, Routledge, London-New York.
- Friginal E. (2018), *Corpus linguistics for English teachers: new tools, online resources, and classroom activities*, Routledge, London-New York.
- Gabrielatos C. (2005), “Corpora and language teaching: just a fling or wedding bells?”, in *Teaching English as a Second Language-Electronic Journal*, 8, 4, pp. 1-35.
- Granger S. (2009), “The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation”, in Aijmer K. (ed.), *Corpora and Language Teaching*, Benjamins, Amsterdam-Philadelphia, pp. 13-33.
- Goulart L., Veloso I. (2023), *Corpora in English Language Teaching. Classroom Activities for Teachers New to Corpus Linguistics*, Open educational resource, Pressbooks: <https://pressbooks.pub/testbook123/>.
- Hanks P. (2008), “The lexicographical legacy of John Sinclair”, in *International Journal of Lexicography*, 21, 3, pp. 219-229.
- Jablonkai R., Csomay E. (2023), *The Routledge Handbook of Corpora and English Language Teaching and Learning*, Routledge, London-New York.
- Johns T. (1986), “Micro-concord: A language learner's research tool”, in *System*, 14, 2, pp. 151-162.
- Johns T. (1988), “Whence and whither classroom concordancing?”, in Bongaerts T., de Haan P., Lobbe S. (eds.), *Computer applications in language learning*, Foris, Dordrecht, pp. 9-27.

- Johns T. (1990), "From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning", in *CALL Austria*, 10, pp. 14-34.
- Johns T. (1991), "Should you be persuaded - two samples of data-driven learning materials", in *English Language Research Journal*, 4, pp. 1-16.
- Kennedy C., Miceli T. (2010), "Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource", in *Language Learning and Technology*, 14, 1, pp. 28-44.
- Lanvers U. (2024), *Language Learning beyond English: Learner Motivation in the Twenty-First Century*, Cambridge University Press, Cambridge.
- Lee H., Warschauer M., Lee J. H. (2018), "The Effects of Corpus Use on Second Language Vocabulary Learning: A Multilevel Meta-analysis", in *Applied Linguistics*, 40, 5, pp. 721-753.
- Leech G. (1997), "Teaching and Language corpora: a convergence", in Wichmann A., Fligelstone S., McEnery A., Knowles G. (eds.), *Teaching and Language Corpora*, Longman, London, pp. 1-24.
- Le Foll E. (2021³), *Creating Corpus-Informed Materials for the English as a Foreign Language Classroom. A step-by-step guide for (trainee) teachers using online resources*, Open Educational Resource: <https://pressbooks.pub/elenlefol/>.
- Le Foll E. (2023), "'Opening up' Corpus Linguistics: An Open Education Approach to Developing Corpus Literacy among Pre-Service Language Teachers", in *Second Language Teacher Education*, 2, 2, pp. 161-186.
- Leńko-Szymańska A. (2024), "Teacher Education for Pedagogical Uses of Corpora", In McCallum L., Tafazoli D. (a cura di), *The Palgrave Encyclopedia of Computer-Assisted Language Learning*, Palgrave, Cham.
- LDOCE = Longman Dictionary of Contemporary English (2016⁶), Pearson Education, Harlow.
- Lo Duca M. G. (2004 [1997]), *Esperimenti Grammaticali. Riflessioni e proposte sull'insegnamento della grammatica dell'italiano*, Carocci, Roma.
- Mauri C., Ballarè S., Gorla E., Cerruti M., Suriano F. (2019), "KIParla corpus: a new resource for spoken Italian", in Bernardi R., Navigli R., Semeraro G. (eds.), *Proceedings of the 6th Italian Conference on Computational Linguistics CLiC-it*, pp. 1-5.
- McCarthy M. (2008), "Accessing and interpreting corpus information in the teacher education context", in *Language Teaching*, 41, 4, pp. 563-574.
- McCarthy M., McCarten J., Sandiford H. (2004/2015²), *Touchstone. Level 4*, Cambridge University Press, Cambridge.
- McEnery T., Hardie A. (2012), *Corpus Linguistics: Method, Theory and Practice*, Cambridge University Press, Cambridge.
- McEnery T., Xiao A., Tono R. Y. (2006), *Corpus-Based Language Studies: An Advanced Resource Book*, Routledge, London-New York.
- McIntosh C. (2013⁴), *Cambridge Advanced Learner's Dictionary*, Cambridge University Press, Cambridge.
- Michellini E. (2022), "Expanding design skills for language teachers: a corpus-based web application for 'language for work' content creation in Italian L2", in Arnbjörnsdóttir B., Bédi B., Bradley L., Friðriksdóttir K., Garðarsdóttir H., Thouësny S., Matthew James M. J. (eds.), *Intelligent CALL, granular systems and learner data: short papers from EUROCALL 2022*, Research-publishing.net, pp. 266-272.
- Michellini E. (2024), *La formazione docente di italiano L2 in contesto migratorio: un sistema integrato per lo sviluppo di materiale didattico corpus-informed sulla lingua per il lavoro*, Tesi di dottorato, Università di Macerata.
- Mizumoto A., Chujo K. (2015), "A Meta-analysis of Data-driven Learning Approach in the Japanese EFL Classroom", in *English Corpus Studies*, 22, pp. 1-18.

- Mukherjee J. (2004), "Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany", in *Language and Computers*, 52, 1, pp. 239-250.
- O'Keefe A. (2021), "Data-driven learning - A call for a broader research gaze", in *Language Teaching*, 54, 2, pp. 259-272.
- Polezzi L. (1993), "Concordancing and the teaching of ab initio Italian language for specific purposes", in *ReCALL*, 5, 9, pp. 14-18.
- Poole R. (2018), *A Guide to Using Corpora for English Language Learners*, Edinburgh University Press, Edinburgh.
- Quirk R. (1960), "Towards a description of English Usage, in *Transactions of the Philological Society*, 59, 1, pp. 40-61.
- Quirk R., Greenbaum S., Leech G., Svartvik J. (1985²), *A Comprehensive Grammar of the English Language*, Pearson Longman, London-New York.
- Römer U. (2008), "Corpora and language teaching", in Lüdeling A., Kyt M. (eds.), *Corpus Linguistics. An International Handbook*, De Gruyter, Berlin-New York, pp. 112-130.
- Schmidt R. W. (2001), "Attention", in Robinson P. (ed.), *Cognition and second language instruction*, Cambridge University Press, Cambridge.
- Sinclair J. M. (1987), *Looking up: An account of the COBUILD project in lexical computing and the development of the Collins COBUILD English Language Dictionary*, Collins, London-Glasgow.
- Sinclair J. M. (1991), *Corpus, concordance, collocation*, Oxford University Press, Oxford.
- Spina S., Forti L., Zanda F. (2021), "Verso un corpus di riferimento dell'italiano parlato dialogico: il modello del BNC2014", in Gorla E., Ciccolone S. (a cura di), *Corpora di parlato nel panorama italo-romanzo. Verso l'individuazione di pratiche condivise*, numero monografico di RID - *Rivista Italiana di Dialettologia*, 44, pp. 89-106.
- Tognini Bonelli E. (2001), *Corpus linguistics at work*, Benjamins, Amsterdam-Philadelphia.
- Tomasello M. (2003), *Constructing a Language: A Usage-Based Theory of Language Acquisition*, Harvard University Press, Cambridge (MA) - London.
- Tavares Pinto P. (2023), *Using language data to learn about language: a teachers' guide to classroom corpus use*, Open educational resource, Pressbooks: <https://uq.pressbooks.pub/using-language-data/>.
- Tribble C. (2015), "Teaching and language corpora: Perspectives from a personal journey", in Leńko-Szymańska A., Boulton A. (eds.), *Multiple Affordances of Language Corpora for Data-driven Learning*, Benjamins, Amsterdam-Philadelphia.
- Vyatkina N. (ed.) (2020), *Incorporating corpora: Using corpora to teach German to English-speaking learners* [Online instructional materials], University of Kansas Open Language Resource Center: <https://opentext.ku.edu/corpora>.
- Vygotskij L. (2012 [1934]), *Thought and Language*, The MIT Press, Cambridge (MA).
- Viana V. (ed.) (2022), *Teaching English with Corpora*, Routledge, London-New York.

