

## Didattica delle lingue corpus-based

Elisa Corino

**Abstract** Competences required to a language teacher today go far beyond the simple knowledge of a foreign language, thus implying a proficient literacy in all the multiple aspects of the influence wielded by the web on the language itself, dealing with computer mediated communication, its emerging textual genres and the ongoing transformation of language related to the influence of technologies. In such a context being familiar with the tools offered by linguistic research is a priority for foreign language teachers, in order to observe, analyze and find new cues and opportunities for their didactics. This contribution is aimed at presenting tools and activities, showing how corpora can intervene in language classroom practice, thus being actively used by both teachers and students.

**Sommario** 1. La rete: ambiente e mezzo privilegiato dell'uso di una lingua. – 2. Corpora e il web as a corpus. – 3. Corpora e glottodidattica. – 4. Conclusioni.

### 1 La rete: ambiente e mezzo privilegiato dell'uso di una lingua

La rete è attualmente il più grande deposito di informazione linguistica esistente, ma è anche uno degli ambienti e dei mezzi privilegiati dell'uso di una lingua, lo spazio entro il quale sia gli usi funzionali sia gli usi creativi del linguaggio sono esercitati con sempre maggior frequenza. E in quanto tale essa è una fonte preziosa di dati linguistici multiformi, sempre aggiornati e a 'portata di click'.

L'affermarsi di nuove modalità di comunicazione - e in particolare con l'avvento della Comunicazione Mediata da Computer (CMC) - ha aperto nuove prospettive e interessi di ricerca in campo linguistico, dallo studio delle email alla chat, dai newsgroup a Twitter, dai blog a Facebook, si sono moltiplicati gli studi di pragmatica e testualità, oltre che le osservazioni analitiche delle varietà di lingua legate al mezzo, al registro, alla tipologia di utenti coinvolti.

Come più volte notato (da Koch, Oesterreicher 1990 a Corino 2007 a Fiorentino 2013) la lingua che si usa nel web presenta tratti di ibridazione

Il presente contributo è stato presentato in forma embrionale come comunicazione orale in occasione del XXI Convegno Nazionale ILSA: *approcci innovativi e buone pratiche dell'insegnamento linguistico*, Firenze 24 novembre 2012.

tra oralità e scrittura, un fenomeno intrinsecamente legato alla forma di comunicazione, avvenga essa in scambi sincroni (chat) o asincroni (mail, newsgroup, blog).

Lo studio della CMC ha visto una prima fase in cui erano per lo più aspetti macrolinguistici ad attirare l'attenzione: analisi legate al lessico emergente, alla spontaneità (vera o presunta) del discorso, alla presenza di elementi quali interiezioni, ideofoni, espressioni gergali o volgari, emoticons (certamente tentativi di rendere alcuni tratti del discorso orale, cosa che ha per l'italiano valenza particolare, visto lo sviluppo diacronico diversificato che lingua orale e scritta hanno seguito), ma che colgono solo una dimensione stilistico-espressiva superficiale, forse importante ma non esaustiva.

I linguisti hanno quindi spostato il loro campo di indagine verso il cambiamento della scrittura - e della lingua - in tutti i suoi molteplici significati, in relazione al cambiamento dei contesti e dei formati in cui viene prodotta grazie alle nuove applicazioni informatiche.

La comunicazione elettronica ha quindi dato origine per le proprie peculiarità a nuovi ambiti e nuove modalità d'uso della lingua, caratterizzati fondamentalmente dalla coesistenza, oltre che di aspetti di concezione del messaggio tipici alcuni dello scritto e altri del parlato, di tratti linguistici tradizionalmente ascritti alcuni a varietà formali e altri a varietà informali. Dal punto di vista concettuale si assiste all'oscillazione tra tratti di immediatezza e di distanza; l'immediatezza comunicativa è legittimata in particolare dai casi di comunicazione sincrona nelle chat e nei newsgroup ad alto indice dialogico-comunicativo (Corino 2007; Corino, Onesti 2013); la distanza è però chiaramente presente per la 'non compresenza fisica degli interlocutori', per l'uso di un filtro che non consente che si realizzi un feedback simultaneo nelle interazioni o che si rispetti l'adiacenza dei turni.

Alla dimensione più prettamente linguistica si somma poi una dimensione più sociale e pragmatica che si riflette nella definizione di un certo senso di comunità che i partecipanti condividono, ovvero quella rete di comunicazione online organizzata e autodefinita per interesse o scopo comune, che ha sede nella piazza virtuale. La percezione di appartenere ad un gruppo ben definito (familiarità ed interesse comune) induce una libertà di movimento ed espressione che favorisce la scomparsa di una lingua spesso fatta di perifrasi, introduzioni all'argomento o lunghe spiegazioni e contemporaneamente la comparsa di modalità di interazione, formule di saluto e commiato, routine, che diventano caratteristiche di quel particolare gruppo e contribuiscono a dare la cifra delle relazioni all'interno della stessa comuni.

Arno Scholz (2003) fa inoltre notare che la volatilità e l'instabilità dei generi testuali elettronici sono condizioni che predispongono ad una certa noncuranza verso le norme della lingua scritta. Se però le caratteristiche della scrittura digitale favoriscono una certa libertà, non bisogna tuttavia dimenticare che le scelte stilistiche e normative dipendono di gran lunga

dallo scrivente e molto meno dal mezzo, e il controllo dello scritto è forse proprio un metodo per veicolare un certo desiderio di ufficialità e serietà, che valorizzi in qualche modo l'autorevolezza del messaggio.

Recenti pubblicazioni (cfr. tra gli altri Campagna, Garzone et al. 2012) hanno messo in evidenza come le scelte degli utenti del web non abbiano solo influenza sulla lingua in quanto sistema regolato da norme e realizzato in usi e strutture morfosintattiche specifiche, ma abbiano anche ricadute a livello testuale, con l'istituzione di nuovi generi che, pur condividendo caratteristiche della testualità 'tradizionale', presentano peculiarità tutte nuove e legate ai molti processi di co-costruzione e multi-autorialità del testo multimediale in rete, ai suoi legami esterni con altri testi e materiali, alla sua granularità e fruibilità non lineare.

Accanto alle numerose problematiche diamesiche ruotano poi altri aspetti non meno importanti che investono invece la dimensione linguistica più ampia che travalica i confini virtuali del mezzo e che riguardano gli effetti che la lingua della rete ha sulla lingua comune, soprattutto scritta, e quale vantaggio didattico si può ricavare dall'osservazione della lingua in rete.

## 2 Corpora e il web as a corpus

Barbera (2013, p. 18) descrive un corpus come una raccolta di testi (scritti, orali o multimediali) o parti di essi in numero finito in formato elettronico trattati in modo uniforme (ossia tokenizzati ed addizionati di markup adeguato) così da essere gestibili ed interrogabili informaticamente; se (come spesso) le finalità sono linguistiche (descrizione di lingue naturali o loro varietà), i testi sono perlopiù scelti in modo da essere autentici e rappresentativi.

Questa definizione 'tecnica' elude l'impatto sostanziale che l'uso di un corpus ha sui contenuti della ricerca, e infatti Barbera stesso (2013, p. 18) coglie il cuore della questione integrando la dicitura formale con una riflessione sui contenuti linguistici di tali raccolte, richiamando la differenza saussuriana tra 'langue e parole':

*Linguisticamente*, inoltre, un corpus è una raccolta di atti di *parole*, e dai fatti di *parole* raccolti in un corpus si può risalire ai loro correlati stati di *langue*, anche se certamente non tutti gli elementi di una *langue* saranno contenuti in un corpus: è l'uso testimoniato dai corpora, anzi, che fonda la *langue*, anche se i corpora, essendo per definizione finiti, ne rappresenteranno solo un sottoinsieme.

Ecco allora che i corpora diventano lo strumento principe, l'osservatorio attrezzato per fornire un quadro della lingua autenticamente usata da parlanti reali in contesti reali e per godere dell'illimitata e piena fruibilità

di tali contesti, fino ad arrivare ai testi interi in cui si essi collocano.

Le posizioni sulla natura del *web as a corpus* sono molteplici e contrastanti. Kilgarriff e Grafenstette (2003, p. 334) affermano con certezza che «The Web is clearly a multilingual corpus». Altri come Sinclair (2005) e Barbera (2013) ritengono invece che il World Wide Web non sia un corpus, poiché le sue dimensioni sono sconosciute e in costante aggiornamento, inoltre non è stato progettato con intenti linguistici.

Quello che è certo è che l'esplorazione delle risorse web come 'mega corpus' risponde all'insufficienza quantitativa della base di dati per affrontare problematiche linguistiche specifiche sempre più complesse, e al sempre più rapido 'invecchiamento' dei materiali rispetto al continuo evolversi del linguaggio (anche in relazione alle nuove tecnologie e a nuovi mezzi di comunicazione legati alla rete).

Barbera (2013) osserva che in realtà l'etichetta 'Web as a corpus' è oggi utilizzata in contesti differenti, come unico contenitore di almeno due situazioni distinte: (1) il materiale del web reso corpus in un determinato taglio temporale, considerando le informazioni di un insieme molto ampio di testi ma comunque finito e stabile; (2) l'idea di elaborare le informazioni su materiale 'aperto', sulla rete in continuo movimento, non creando un vero e proprio corpus ma applicando ai dati *tools* di estrazione e *crawling*.<sup>1</sup>

Il rischio di attingere indiscriminatamente dal web per la creazione di corpora - e di corpora da sfruttare in chiave didattica - è quello di perdere il controllo sul materiale immesso. Si pensi ad esempio alla profusione di errori ortografici e grammaticali che proliferano in rete e alle numerose occasioni in cui si assiste alla loro conseguente auto- ed eterocorrezione che, se possono suscitare una certa ilarità all'interno della comunità virtuale, non giovano certo a chi è in fase di apprendimento.

### 3 Corpora e glottodidattica

Nell'ambito dell'insegnamento delle lingue la *corpus linguistics* (linguistica dei corpora) è ormai entrata a buon diritto tra le discipline che più contribuiscono ad arricchire in modo rilevante le ricerche di altri più 'tradizionali' settori (lessicografia, grammaticografia, manualistica) aprendo nuove prospettive sullo studio delle varietà delle lingue, soprattutto grazie alla quantità di dati informatizzati - immediatamente fruibili e interrogabili - che mette a disposizione dei linguisti.

L'importanza dei corpora in prospettiva glottodidattica emerge quando corpora bilanciati o, con più rischio, la rete stessa (*web as a corpus*) vengono utilizzati come testimoni dell'uso della lingua, come insieme di

1 Il *crawling* è un'operazione di analisi e filtro di dati, nel caso specifico di contenuti del Web.

parlanti (scriventi) nativi a disposizione a tutte le ore, pronti a far vedere come si dice *X* in una L2. Usare la rete come corpus o usare corpora bilanciati, o comunque ‘filtrati’ come gli archivi dei giornali, permette di far notare aspetti quantitativi (enucleare le parole e le strutture più frequenti della lingua target e dare ad esse la priorità nell’insegnamento) e aspetti qualitativi, collocazionali. Quali aggettivi solitamente s’accompagnano a tal nome? Quale avverbio compare solitamente con il tal verbo? Come è costruito di solito un verbo? Con oggetti diretti o indiretti e quali preposizioni compaiono negli oggetti indiretti? Un corpus elettronico è in grado di rispondere in modo veloce a simili domande, attingendo ad una notevole mole di dati, non altrimenti disponibili nei manuali, ed utilizzando delle maschere di interrogazione mirate (*query*) per incrociare gli elementi di maggiore interesse.

Usare dei corpora non solo per la ricerca linguistica, ma anche per la formazione di studenti e insegnanti o per l’approfondimento da parte di appassionati di lingue e delle loro mutevoli forme, consente di riflettere davvero su come apprendenti e nativi scrivono, sulla reciprocità degli errori o sulla comunanza di errori fatti ad es. da inglesi e tedeschi o da francesi e spagnoli. È un modo fruttuoso di praticare le nozioni teoriche dei corsi di linguistica generale o italiana (o di linguistica inglese, francese, ecc.) e di didattica delle lingue moderne, perché si ha esperienza diretta dei piani linguistici (sempre un po’ astratti quando non si vedono concretamente attraverso gli errori) e di cosa vuol dire sovraestendere un morfema o attuare una strategia di evitamento, ci si abitua ad individuare gli errori che sono errori solo in prospettiva transfastica o solo in termini di accettabilità e ad interrogarsi sul concetto di norma di una lingua... si impara insomma a ‘smontare scientificamente’ il giocattolo-lingua. E viceversa, attraverso i corpora annotati per parti del discorso, insegnanti, glottodidatti e linguisti possono più efficacemente riuscire a spiegarci come passare ‘dall’altra parte’, quella in cui il giocattolo si (ri)monta.

Uno degli usi dei corpora è estrarre dei modelli di lingua a partire da liste di parole o combinazioni di parole che descrivono come queste siano correlate, come vengono usate le une con le altre, e quanto frequenti esse sono in certo contesto o settore disciplinare (*Language for Specific Purposes*).

Saper usare le collocazioni tipiche di una parola è ciò che distingue l’apprendente avanzato dal principiante e lo sforzo mnemonico che tale abilità richiede può esser motivato soltanto tramite la testimonianza offerta da corpora o dalla rete come corpus. Un tempo solo gli specialisti consultavano le concordanze a stampa, ora si trovano in rete dati linguistici visibili in modalità KWIC (*Key Word In Context*) con grande facilità. Certo l’insegnante ha un impegnativo compito: far capire agli studenti che non ci si può fidare del primo sito che si incontra nella rete e che bisogna, come con ogni fonte di informazione, sapere che cosa si cerca e i metodi per trovarlo. Oggi però l’insegnante ha a disposizione strumenti informatici migliori e finestre

di interrogazione più amichevoli, può anche sfruttare dizionari monolingui e bilingui della L2 come primo terreno di allenamento per i propri studenti (con la modalità di interrogazione tutto testo) prima di farli navigare in Google o presentare loro dei corpora veri e propri.

Specificamente importanti per l'insegnante di lingua straniera sono i corpora di apprendenti (*learner corpora*). Se nel 1998 Leech poteva osservare che la ricerca sui corpora di apprendenti doveva ancora affrontare degli ostacoli sia sul piano dell'allestimento stesso dei *corpora*, sia sul piano del riconoscimento scientifico, a quindici anni di distanza possiamo affermare che i corpora di apprendenti, e la ricerca su di essi basata, hanno conquistato e consolidato la loro presenza negli ambienti della linguistica applicata. Sono diventati uno strumento vitale e ampiamente diffuso, non solo in ambito anglosassone, dove la disciplina è nata, ma anche nello studio dell'apprendimento di lingue diverse dall'inglese, non ancora però nella pratica didattica.

Come già facevano notare Barbera e Marelo (2004) e Marelo (2011) introducendo VALICO,<sup>2</sup> il corpus di apprendenti di italiano L2 liberamente accessibile in rete, tre sono i principali destinatari di un corpus di apprendenti: gli insegnanti, i linguisti applicati, i linguisti. A questi ultimi i *learner corpora* si propongono come strumenti dinamici e ricchi di dati su una particolare varietà di lingua che può essere comparata con altre varietà; per i linguisti applicati i *corpora* di apprendenti sono utili in quanto consentono di sviluppare efficaci materiali didattici, basati sull'analisi delle reali difficoltà riscontrabili nella produzione di coloro che studiano una lingua come lingua straniera. Si pensi, ad esempio, alle sezioni dedicate alla prevenzione degli errori più frequentemente commessi da apprendenti di inglese come lingua straniera: tali sezioni, apparse di recente nei dizionari per apprendenti di inglese (cfr. Immagine 1), sono appunto il frutto anche dell'indagine *corpus-based*, cioè a partire dai *corpora* di apprendenti.

Il MEDAL 2009 (MacMillan English Dictionary for Advanced Learners), ad esempio, contiene dei box con 'avvertimenti' rivolti agli studenti, che approfondiscono questioni lessicali o morfosintattiche emerse come particolarmente problematiche durante l'analisi del corpus. Di recente anche alcuni libri di testo, tra questi manuali di lingua inglese per la scuola secondaria, hanno integrato suggerimenti rivolti a gruppi specifici di apprendenti, mettendoli in guardia da strutture 'pericolose'.

2 *Varietà Apprendimento Lingua Italiana Corpus Online*, cfr. <http://www.valico.org>.

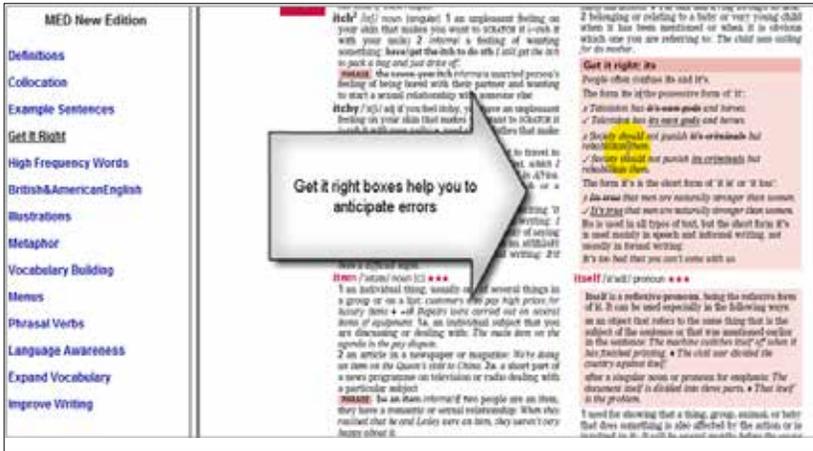


Figura 1. Esempio di focus per gli apprendenti tratti da analisi di corpus, MEDAL 2009

Infine, i *learner corpora* si rivelano preziosi per i docenti. Gli insegnanti di lingua in formazione possono sfruttare tali strumenti come risorsa di autoformazione relativamente alle caratteristiche delle varietà di apprendimento. I docenti già formati vi trovano materiali per la costruzione di esercizi a partire dal *corpus*, soprattutto se questo ha anche un corpus appaiato che contiene testi di parlanti nativi elicitati a partire dagli stessi stimoli dati agli apprendenti stranieri.

Corpora di libri di testo permettono poi l’analisi della varietà di lingua alla quale gli apprendenti sono esposti e, se comparati con corpora in L1, possono facilitare lo sviluppo di materiali più efficaci.

Lo schema seguente (McEnery/Gabrielatos 2005) ben riassume le connessioni tra *corpus linguistics* e didattica:

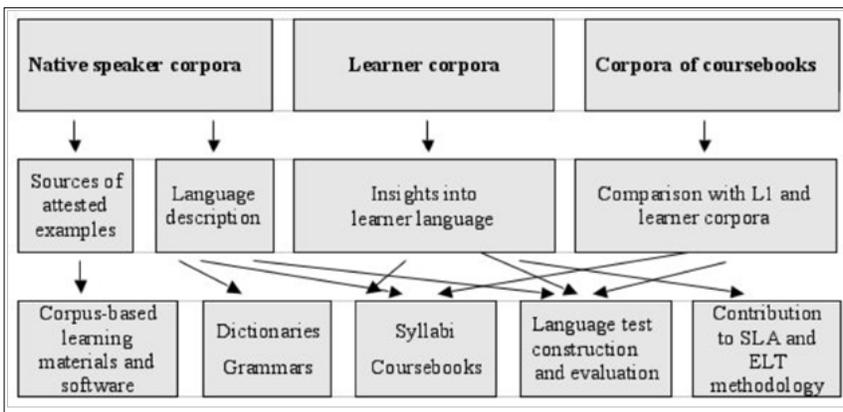


Figura 2. Connessioni tra *corpus linguistics* e didattica, McEnery/Gabrielatos 2005

Lo studio dell'italiano usato, ad esempio, nei newsgroup potrebbe inoltre fornire utili riscontri sull'italiano scritto informale in continua evoluzione, con una preminente utilità per lo studio della lingua in rete e per l'analisi dell'interazione tra norma e sistema nella lingua italiana standard e neo-standard, pur con le dovute cautele imposte dai possibili errori ortografici o dalle numerose sviste legate alla velocità di scrittura.

### 3.1 Corpora e insegnamento: strumenti e proposte

Coniugare corpora e insegnamento delle lingue significa sciogliere il binomio insegnamento/apprendimento in due livelli paralleli ma dotati di denominatore comune: l'aula di lezione e i suoi attori. Da una parte c'è lo sfruttamento dei corpora da parte del docente, nel processo di autoformazione e nel momento della programmazione e creazione di materiali; dall'altra ci sono gli studenti e la possibilità di integrare risultati e strumenti della corpus linguistics in un processo di apprendimento attivo e parzialmente autonomo.

Il potenziale dei corpora come fonte per lo sviluppo di materiali didattici è enorme, sulla scorta di quanto fatto dai lessicografi negli ultimi vent'anni, e risulta oggi ancor più necessario che anche gli autori di manuali ed esercizi facciano riferimento ai dati emersi dagli studi di *corpus linguistics*, per definire lo spettro lessicale da includere nei sillabi (le liste di frequenza danno importanti indicazioni) o le strutture morfosintattiche nelle loro più attuali e diffuse varietà d'uso (integrando le norme della grammatica prescrittiva con gli usi reali osservati in contesti specifici).

Dal punto di osservazione speculare, quello delle ricadute sulla formazione dei discenti, coinvolgere gli studenti in ricerche 'linguistiche' può renderli più attivi nella formazione della propria conoscenza e stimolare processi cognitivi superiori, oltre che incoraggiare una certa autonomia, contribuendo così all'auspicato passaggio dalle BICS (*Basic Interactive Communicative Skills*) alle CALP (*Cognitive Academic Language Proficiency*).<sup>3</sup>

Tra le possibili attività da proporre in classe l'insegnante può guidare una ricerca linguistica basata su concordanze che confermi regole e predizioni o che porti gli studenti a fare ipotesi e trarre conclusioni dai fatti di lingua. Il docente sarà in questi casi un facilitatore dell'apprendimento, più che il detentore del sapere, il discente avrà invece un ruolo attivo di primo piano

3 Non ci soffermiamo in questa sede sulla declinazione delle attività in relazione ai processi cognitivi innescati da ciascun compito; rimandiamo, tra i numerosi testi di riferimento, a Cummins 2000.

nello sviluppo e organizzazione della conoscenza e, come nota Schmidt (1990), «what language learners become conscious of - what they pay attention to, what they notice - influences and in some ways determines the outcome of learning».

I benefici di un approccio che metta la figura del discente al centro del processo educativo sono dibattuti e da tempo chiariti dalla pedagogia e dalla glottodidattica, il valore aggiunto di attività corpus-based è invece dato dall'autenticità dei materiali che derivano da (con)testi reali, piuttosto che da materiali creati a tavolino per scopi didattici. Lavorare per WordSketches (cfr. *infra* § 3.2), per concordanze e per KWIC, permette allo studente di determinare i potenziali diversi significati delle parole, i sintagmi ricorrenti e le collocazioni tipiche di certi contesti, strutture diverse in relazione a registri, mezzi, tipi testuali differenti. Agli studenti si può chiedere di:

- esaminare l'ordine delle parole,
- ricavare regole grammaticali da un numero cospicuo di esempi,
- definire i contesti d'uso di termini sinonimici (con sfumature semantiche e implicazioni connotative diverse),
- comparare scritto e parlato o diversi generi testuali.

Interessante, ad esempio, è notare il trattamento riservato all'inglese *any*, ai suoi significati e costruzioni sintattiche, nei libri di testo in uso nelle scuole italiane: la spiegazione consueta che si offre a studenti di livello A1-B1 è pressappoco «*any* si usa in sostituzione di *some* per indicare una certa quantità in frasi negative e interrogative». Ricerche corpus-based, tuttavia, hanno sottolineato che il 50% degli usi di *any* ha luogo in frasi affermative, il 40% in negative e solo il 10% in interrogative (Mindt 1998). Alla luce di tali risultati è necessario riconsiderare l'opportunità di offrire agli studenti una spiegazione che di fatto risulta parziale e non rispecchia gli usi effettivi della lingua nella realtà. L'English Vocabulary Profile, esito per la lingua inglese del progetto promosso dal Consiglio d'Europa<sup>4</sup> per 'riempire' di contenuti la cornice delineata dal QCER, inserisce l'uso del quantificatore con significato di 'qualsiasi' già a livello A1.

Appare chiaro che le due entrate di *any* corrispondono a due significati diversi del quantificatore e dagli esempi si evincono i tre pattern sintattici che ad essi corrispondono.

4 Il progetto del Consiglio d'Europa sui referenziali coinvolge molti dei paesi europei e ha portato alla creazione di Profili per molte delle lingue studiate nello spazio Comune, oltre al Profilo inglese (<http://vocabulary.englishprofile.org>) per l'italiano cfr Spinelli, Parizzi 2010, *Profilo della lingua italiana*, La Nuova Italia; per il tedesco Profilo Deutsch, Langenscheidt; per lo spagnolo Plan curricular, liberamente accessibile sul sito dell'Istituto Cervantes [http://cvc.cervantes.es/ensenanza/biblioteca\\_ele/plan\\_curricular/default.htm](http://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/default.htm).

The screenshot shows the 'English Vocabulary Profile' website. The main heading is 'English Vocabulary Profile'. Below it, there are two circular icons: one showing a person looking at a screen and another showing a keyboard. The page is divided into a left sidebar and a main content area. The sidebar includes options for 'British English' and 'American English', a 'Choose level' section with radio buttons for A1, A1-A2, A1-B1, A1-B2, A1-C1, A1-C2, A2 only, B1 only, B2 only, C1 only, and C2 only, and a search bar with the text 'any'. The main content area shows the word 'any' with its phonetic transcription /'ni:/ and an 'Outline view' button. The word is categorized as a 'DETERMINER ; PRONOUN'. It has two main uses: 'SOME' and 'NOT IMPORTANT WHICH'. The 'SOME' section includes a definition, dictionary examples, and a learner example. The 'NOT IMPORTANT WHICH' section includes a definition, dictionary examples, and a learner example. The page also features a navigation menu in the sidebar with links to 'English Profile Home', 'About English Vocabulary Profile', 'Word of the Week', 'Help', and 'Feedback'.

Figura 3. Any nell'English Profile

A questo proposito Krieger (2003) propone un esercizio induttivo che prende spunto dalle occorrenze di *any* contenute nel Corpus of Spoken Professional American English, partendo dai dati selezionati dall'insegnante: gli studenti dovranno individuare le tre diverse costruzioni sintattiche che ruotano intorno al quantificatore, osservarne la frequenza, ipotizzarne i significati, analizzare collocazioni e *chunk* linguistici legati alla parola in grassetto.

Read through the following lines taken from a concordance of the word *any*.

This is going to be a test like **any** other test, like, for example working with you.. If there are **any** questions about how we're going to and I didn't receive **any** materials for the November meeting and it probably won't make **any** difference. I mean, that's the next You can do it **any** way you want.

Do you want to ask **any** questions? make **any** comments?  
I don't have **any** problem with that. I'm just saying  
if they make **any** changes, they would be minor changes.  
I think we ought to use **any** kind of calculator. I think that way  
I see it and it doesn't make **any** sense to me, but I can take that

What conclusions can you draw about the use of *any*?  
What are the three main uses of *any* in order of frequency?

Any 1:  
Any 2:  
Any 3:

Certo lavorare con i corpora in classe non è semplice: la tecnologia non raggiunge ancora capillarmente ogni aula di una scuola o di un istituto, i laboratori di informatica sono sovraffollati, le connessioni lente, e infine bisogna tenere conto anche del target di studenti che si ha di fronte: si tratta di attività ancora poco integrate nella routine del processo educativo, insomma strumenti per molti, ma non per tutti. La fruizione e lo sfruttamento dei corpora devono essere accuratamente pianificate in base al livello degli studenti, alle loro inclinazioni, tenendo conto di intelligenze e stili cognitivi diversi, e naturalmente con una particolare attenzione alle tematiche disciplinari da trattare.

La prima sfida che si trova a dover affrontare un insegnante che voglia integrare i corpora nella sua prassi didattica è la scelta del materiale sul quale strutturare riflessioni e attività. Il web in quanto corpus è una dedalo intricato e spesso disorientante; la lingua, con tutte le varietà che propone, i generi testuali nei quali si realizza, le dicotomie intrinseche alla sua forma mutevole, è ancora di più un terreno insidioso sul quale muovere i propri passi.

L'insegnante per primo deve imparare ad orientarsi e a selezionare, assicurandosi che il corpus sia utile in un certo contesto di insegnamento/apprendimento e che sia rappresentativo della varietà di lingua sulla quale si è deciso di strutturare il proprio intervento didattico. Proporre a studenti di livello intermedio (diciamo approssimativamente intorno al B1) l'intero BNC<sup>5</sup> può essere ottimistico da parte dell'insegnante e demotivante per parte degli allievi, data l'ampiezza e la varietà dei materiali contenuti; inoltre una mole di dati tanto variegati potrebbe comunque portare a risultati insoddisfacenti se si è alla ricerca di questioni specifiche, in particolare legate al registro. Sottoporre invece la sezione del COCA<sup>6</sup> creata a partire da articoli di medicina a studenti avanzati (B2/C1) di materie scientifiche può al contrario essere stimolante, senza dimenticare l'esistenza di strumenti di facile fruizione che permettono di creare corpora *ad hoc* o di generare concordanze a partire da testi selezionati e immessi dagli stessi utenti, libri di testo, articoli, siti web (i cosiddetti *concordancer*, i WordSmith Tools ne sono un esempio,<sup>7</sup> ma anche software come SketchEngine).<sup>8</sup>

Infine usare un corpus, per quanto semplice l'interfaccia possa essere e per quanto i nativi digitali siano pronti per le classi 2.0, non è un'operazione che uno studente può fare da solo: affinché raggiunga una certa autonomia è necessario che venga guidato e istruito sulle modalità di interrogazio-

5 British National Corpus, la più grande raccolta di dati linguistici di inglese britannico (100 milioni di parole), <http://www.natcorp.ox.ac.uk/>.

6 Corpus of Contemporary American English (450 milioni di parole), <http://corpus.byu.edu/coca/>.

7 <http://www.lexically.net/wordsmith/>.

8 [the.sketchengine.co.uk](http://the.sketchengine.co.uk).

ne e di lettura dei dati; usare strumenti linguistici vuol dire innanzitutto avere una discreta padronanza della terminologia metalinguistica (si deve sapere cos'è una collocazione e, prima ancora, cos'è un nome, un verbo, una preposizione), usare strumenti linguistici legati al dominio informatico può voler dire avere una - almeno - minima competenza d'uso di operatori logici o di semplici espressioni regolari.



Figura 4. FOTO in itTenTen

I linguaggi di interrogazione per linguisti sono spesso filtrati da interfacce che consentono una maggiore usabilità anche a utenti inesperti, così la stessa interrogazione condotta su un corpus come il NUNC-Photo<sup>9</sup> (figura 5) può dare risultati affidabili e coerenti con l'intenzione iniziale.

La sintassi di ricerca sarà

[pos='VER.\*'] []{0,1} [word= 'foto']

ovvero:

[verbo generico per modo e tempo] [da 0 a 1 parola - presunto articolo] [la parola foto]

9 La sezione dedicata alla fotografia all'interno dei NUNC - *Newsgroup UseNet Corpora*, creati dall'Università di Torino ([http://www.corpora.unito.it/photo/nunclit\\_pho.php](http://www.corpora.unito.it/photo/nunclit_pho.php)).

Ma un'interfaccia più amichevole, come quella che si mostra sotto, consentirà una maggiore facilità di accesso ai dati, permettendo a chi cerca di scrivere l'espressione regolare semplicemente cliccando sui pulsanti predisposti.

The screenshot shows a web browser window with the URL [www.corpora.unibo.it/photo/cgornide/](http://www.corpora.unibo.it/photo/cgornide/). The page title is "Corpus 'NUNC - Fotografia', ricerca linguistica - indice/indietro". Below the title, there are search parameters: "Contesto: 20 parole", "Risultati: 20", and "Attributi da mostrare: parola, lemma, pos". A search box contains the example "[word='va']". Below the search box, there is a message: "dunque invia la richiesta che hai formulato!".

On the left side, there are several columns of buttons for selecting search criteria:

- Punteggiatura:** inizio parola, fine parola, oppure, non, e non, e, Quantità: almeno uno, zero o più, uno o zero, da... a
- Pos:** una parola qualsiasi, il lemma, la parola, la parol, un aggettivo, un nome, un avverbio, un segno interpuntivo, un articolo, un nome proprio, un numero
- Verbi:** un verbo, un presente, un infinito, un part. pass, un futuro, un imperfetto, un imperativo, un condizionale, un gerundio
- Pronomi:** un pronome, un riflessivo, un indefinito, un p. personale, un p. relativo, un possessivo, un dimostrativo, un interrogativo

On the right side, there is a list of 20 matches. The first five matches are visible:

- 1 I c. Spero non si presenti alla porta del lab vestito solo di quello . ROTFL , se dovesse accadere **faci delle foto** -) A g - " La vita senza contropartita è tutto quello che ancora mi esalta . Non riuscirete a
- 2 e non è poco . Quando mi hai raccontato alcune delle esperienze che hai vissuto in Irlanda , mi hai **ricordato le foto** . Quando ho rivisto le foto mi hanno ricordato le tue parole e le sensazioni che mi hai descritto .
- 3 impegno che ci metto inizia a portare miglioramenti , poi che non sono capolavori lo so , però riuscire a **are foto** vedibili mi fa piacere . Le altre sinceramente le ho trovate un po' bruttine , molto posate o casuali o
- 4 ad es . a Filippo o ad Ezio ( giusto per citarne due , e non a caso ) . **Riprendendo la foto** della tua dolce metà , se è vero che ci hai messo tempo per prendere consapevolezza dei difetti in essa
- 5 post TQT Re : fotopride " Virgus " <ADDRESS@virgilio.it> ha scritto nel <news>news:ar2t7\$5to\$1@iacerta.fiscalinet.it</news> ... scusa ragazzi ????!!? entro domani bisogna **inviare le foto** per il prossimo ... il tema era sopravvissuti giusto ? : ciao No . La scadenza del nuovo è venerdì

Figura 5. Interrogazione NUNC Photo

### 3.2 Concordanze e word sketches

Di particolare utilità sono corpora gestiti da software quali Sketch Engine (SkE, <http://www.sketchengine.co.uk/>) e RIDIRE (Risorsa Dinamica di Rete Italiana, <http://www.ridire.it>).

Il primo è un software (accessibile con una versione di prova della durata di trenta giorni) che permette di generare dei cosiddetti *word sketches*, dei riassunti schematici del pattern sintattico delle parole e del loro comportamento collocazionale. Si tratta di uno strumento creato dalla ricerca in linguistica computazionale anglosassone e inizialmente rivolto ai lessicografi, usato, ad esempio, per la creazione del Macmillan English Dictionary (2002), ma oggi mette a disposizione degli utenti strumenti per processare corpora di un gran numero di lingue (oltre a inglese, francese, tedesco, spagnolo e italiano è possibile immettere materiali di bulgaro, estone, ecc.). Con SkE è possibile interrogare corpora già esistenti come il BNC per l'inglese o corpora creati appositamente per il motore di ricerca computazionale come itTenTen, corpus di 3.1 miliardi di parole generato tramite un processo di *web crawling*; inoltre è possibile inserire personalmente testi e generare corpora personalizzati o 'dragare' il web grazie a WebBootCat, che consente di inserire parametri

specifici per creare raccolte di linguaggi specialistici a partire da materiali online. Una volta scelto o creato il corpus, si può disporre di tesauri, pattern lessicali e sintattici intorno alle parole richieste, oltre a utili analisi statistiche sulle similitudini o differenze d'uso che riguardano quasi-sinonimi.

All'insegnante uno strumento del genere può essere utile per creare raccolte di testi perfettamente aderenti ai bisogni, ai livelli, agli interessi dei propri allievi, per mostrare loro con un supporto visuale come in effetti funziona una lingua e quali sono le relazioni tra le parole e le varie parti del discorso.

Riprendiamo il collocato FOTO e interroghiamo il *word sketch* (Figura 6a/b) corrispondente con tutto l'intorno della parola 'foto': aggettivi che la seguono, sintagmi preposizionali, verbi con i quali colloca... dall'occorrenza è naturalmente poi possibile espandere i contesti e risalire a un vero e proprio testo e all'URL dal quale il documento è stato tratto (Figura 7).

Sketch Engine

user: Dr. elisa corino corpus: itTenTen10

Concordance  
Word List  
Word Sketch  
Thesaurus  
Find X  
Sketch-Diff  
Save  
Change options  
Clustering  
Sorting  
Gramrels  
MW links  
More data  
Less data

**foto** (noun) itTenTen10 freq = 384897 (125.1 per million)

n_modifier	4622	1.5	preN_V	105054	3.3	e_o	63316	1.1	AoNf	60115	1.2	postN_V	50443	2.7
gallery	79	7.96	scattare	7466	10.34	video	6434	9.36	disponibile	7065	9.75	scattare	4029	9.94
segnalamento	37	7.74	postare	1818	8.2	filmato	1459	8.55	segnalatico	691	8.46	ritrarre	476	7.82
taroccate	33	7.67	guardare	4908	8.19	autografo	504	7.82	bello	4624	7.54	ritoccare	221	7.0
satellitare	253	7.42	mandare	2528	8.1	disegno	1256	6.88	allegato	613	7.47	ingiallire	190	6.86
tessera	284	7.09	allegare	1119	7.88	descrizione	511	6.47	aereo	857	7.42	sciabilità	174	6.81
cineoperatore	20	6.97	pubblicare	3482	7.79	teste testo	859	6.44	spia	311	7.37	ricordare	1780	6.61
taroccata	18	6.86	caricare	925	7.69	cartolina	210	6.19	digitale	1072	7.34	pubblicare	1319	6.55
choc	39	6.8	inviare	2547	7.61	poster	167	6.11	compromettente	228	6.91	ansare	122	6.27
polaroid	13	6.04	vedere	8013	7.46	filmino	115	5.82	panoramico	310	6.88	osè	108	6.12
hot	21	5.97	mostrare	2054	7.17	intervista	470	5.81	inedito	431	6.74	appendere	138	6.1
pop-art	9	5.91	inserire	1546	7.13	ricetta	298	5.77	vincitore	281	6.71	postare	264	5.87
ritocco	19	5.32	visualizzare	575	6.91	illustrazione	178	5.75	rubato	229	6.71	incorniciare	97	5.84
reporter	17	5.14	mettere	3774	6.84	locandina	125	5.75	sexy	214	6.55	formare	494	5.82
glamour	12	5.07	stampare	509	6.83	news	159	5.63	ag	171	6.54	NON	97	5.81
safari	9	4.98	Le	1411	6.67	immagine	1320	5.63	splendido	502	6.48	tesserare	89	5.79
classificata	12	4.94	scaricare	570	6.38	Testo	95	5.6	vecchio	1254	6.42	mostrare	657	5.7

Figura 6a. Word Sketch FOTO

pp_di-i	18910	1.1	pp_del-x	15130	1.7	pp_della-x	9030	1.3	pp_in-i	5621	1.2	pp_a-i	6272	2.0
rito	925	7.86	backstage	47	6.27	gallery	20	6.05	alto	1524	8.94	fianco	440	6.84
copertina	496	7.61	cadavere	125	6.14	vacanza	334	5.71	bidini	44	7.06	destra	865	6.46
nudo	165	7.14	nipotino	35	5.64	modella	93	5.63	topless	30	6.85	corredo	67	6.25
scena	618	5.48	booklet	26	5.62	copertina	99	5.47	posa	114	6.66	lato	694	6.22
backstage	32	5.47	matrimonio	251	5.39	diva	22	5.43	copertina	133	5.95	sinistra	981	6.15
repertorio	112	5.43	raduno	46	5.22	nipotina	16	5.36	hp	17	5.87	sx	38	6.02
modella	59	5.42	bimbo	111	5.11	premiazione	50	5.36	gallery	21	5.8	colore	1065	6.01
gruppo	1740	5.39	ricercato	19	5.11	gita	29	5.36	jog	11	5.64	pag	18	5.91
reportage	38	4.81	plastico	22	5.04	gite	23	5.28	formato	213	5.6	raffica	34	5.74
cadavere	41	4.44	paparazzo	18	4.93	fidanzata	40	5.14	homepage	21	5.42	megapixel	16	5.58
modellino	15	4.35	compleanno	56	4.83	lapide	21	4.93	b	89	5.41	dx	25	5.49
moda	114	4.35	bolide	17	4.83	bimba	34	4.9	home	50	5.03	sin	80	4.64
ragazza	232	4.31	gattino	16	4.72	sfilata	28	4.89	anteprima	54	4.98	posa	19	4.09
paesaggio	126	4.3	disperso	15	4.7	tortura	51	4.86	bianco	48	4.93	ripetizione	15	3.8
kurt	14	4.28	gemello	23	4.68	serata	243	4.79	miniatura	18	4.91	risoluzione	48	3.55
bimbo	63	4.24	festino	14	4.59	salma	14	4.58	esclusiva	34	4.78	seno	15	2.43
bambina	73	4.22	retro	24	4.56	opel	8	4.51	allegato	92	4.78	grandezza	12	2.33
bill	13	4.2	arrestato	17	4.55	famigliola	9	4.5	bianconero	18	4.75	petto	8	2.31
bimba	26	4.17	reportage	29	4.55	torta	34	4.47	mutanda	16	4.72	pagina	102	2.29
famiglia	627	4.16	gatto	66	4.51	scocca	9	4.47	costume	76	4.21	schermo	23	2.14
donna	12	4.13	proiettore	19	4.5	tetta	14	4.43	risoluzione	75	4.19	testimonianza	25	2.03
neonato	26	3.99	duce	14	4.49	cattura	29	4.42	intimo	8	3.93	cura	52	1.79

Figura 6b. Word Sketch FOTO

The screenshot shows the Sketch Engine interface with the following elements:

- Header:** Sketch Engine logo, navigation links (About, Home, Settings, Change password, Log out), and a search bar.
- Search Results:** A list of concordance items for the word "foto". Each item includes a document ID (e.g., doc#1049609), a snippet of text, and a frequency count in parentheses (e.g., 33 (0.0 per million)).
- Left Sidebar:** A menu with options like "Concordance", "Word List", "Thesaurus", "Find X", "Sketch-Diff", "Save", "View options", "KWIC", "Sentence", "Sort", "Left", "Right", "Node", "Shuffle", "Sample", "Filter", "Frequency", "Node tags", "Node forms", "Collocations", and "ConcDex".
- Right Panel:** A detailed view of a selected concordance item, showing the word "foto" in context and associated metadata like "doc.lang\_filter\_score", "doc.timestamp", "doc.http\_date", "doc.http\_last\_modified", "doc.wordcount", "p.heading", and "p.class".

Figura 7. Concordanze FOTO

Il word sketch è molto dettagliato e qui ne abbiamo riportato solo una parte. Interessanti sono i riquadri che mettono in luce l'uso delle preposizioni in costruzioni in cui il termine viene usato: *foto DI rito* (pp\_di-i), *foto IN topless* (pp\_in-i), *foto AL microscopio* (pp\_al-x), ecc. Notiamo poi tra le occorrenze di *foto*+modificatore (n\_modifier) il frequentissimo *foto satellitare* o l'interessante - da un punto di vista diafasico - *foto taroccata*; rispetto ai verbi che precedono (preN\_V) o seguono (postN\_V) il nome troviamo *scattare*, *postare*, *guardare*. Scorrendo la lista emerge anche agli occhi degli apprendenti che *prendere* (frequente esito di interferenza dall'inglese) non è la scelta più frequente in italiano.

La rete è anche il punto di partenza per i corpora sui quali si basa RIDIRE, una collezione di più domini specialistici - architettura, cucina, amministrazione, politica - interrogabili separatamente, ma accomunati da una stessa interfaccia di ricerca. Come i corpora generati da WebBootCat, anche RIDIRE è il frutto di operazioni di *web crawling* e ha tra i suoi obiettivi, sostenuti anche dalla SILFI,<sup>10</sup> proprio quello di potenziare la competenza negli usi linguistici attraverso l'osservazione e l'interrogazione dei materiali in rete.

Supponiamo ad esempio che si voglia verificare quale verbo (si) colloca con il lemma *contratto* in campo amministrativo: sarà sufficiente selezionare il dominio di indagine nella finestra di ricerca (Figura 8) e poi restringere la ricerca con un filtro per PoS (Figura 9):

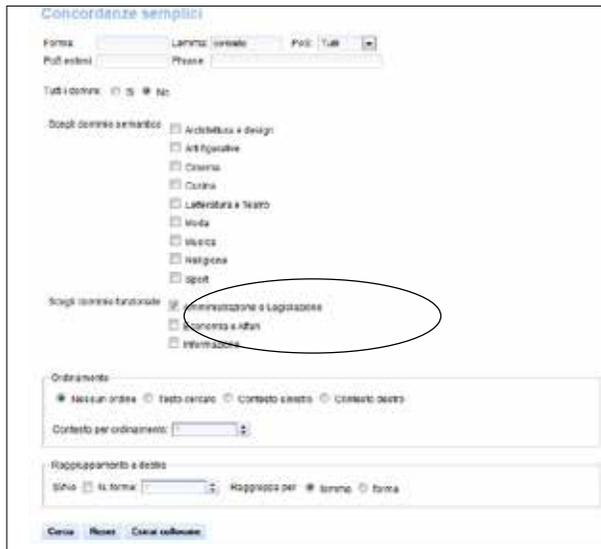


Figura 8. RIDIRE

10 Società Internazionale di Linguistica e Filologia Italiana.



Figura 9. RIDIRE Collocare

Selezionando l’occorrenza (type) desiderata è poi possibile risalire a tutte le occorrenze (token) (Figura 10) in cui essa è presente e, successivamente, anche all’intero testo o al sito internet dal quale il testo stesso è stato tratto (Figura 11).



Figura 10. RIDIRE Collocare

Domanda	Se	Se	Domanda	URL
	un divanetto di ingegneria prelevato nella cantina general del	<a href="#">colloca il controllo</a>	nella stanza (presumibilmente non ancora avuta) fai	<a href="#">Amministrazione_1_1</a>
	il pubblico del bando, con l'attore al tempo del	<a href="#">colloca il controllo</a>	contro complessivo sostenuto per il personale dipendente nel bilancio	<a href="#">Amministrazione_1_1</a>
	del qualificato come capogruppo, che sarà chiamata a	<a href="#">colloca il controllo</a>	in nome e per conto proprio e delle mandati,	<a href="#">Amministrazione_1_1</a>
	trattamento del risultato una attività di rilevazione economica oggetto di	<a href="#">colloca il controllo</a>	con una pubblica amministrazione, devono sempre applicarsi le regole	<a href="#">Amministrazione_1_1</a>
	... Si chiede pertanto di sapere se sarà possibile	<a href="#">colloca il controllo</a>	di co-co, per la scuola	<a href="#">Amministrazione_1_1</a>
	del qualificato come capogruppo, che sarà chiamata a	<a href="#">colloca il controllo</a>	in nome e per conto proprio e delle mandati,	<a href="#">Amministrazione_1_1</a>
	una "attività" e di analizzarla presso	<a href="#">colloca il controllo</a>	di collaborazione autonoma, indipendentemente dall'oggetto della prestazione,	<a href="#">Amministrazione_1_1</a>
	il "fatto stesso" di essere agente di	<a href="#">colloca il controllo</a>	di servizio con soggetti privati che si configurano come esenti	<a href="#">Amministrazione_1_1</a>
	0012160 qualificato come capogruppo, che sarà chiamata a	<a href="#">colloca il controllo</a>	in nome e per conto proprio e delle mandati,	<a href="#">Amministrazione_1_1</a>
	del qualificato come capogruppo, che sarà chiamata a	<a href="#">colloca il controllo</a>	in nome e per conto proprio e delle mandati,	<a href="#">Amministrazione_1_1</a>
	del servizio senza selezione di candidati, si rende opportuno	<a href="#">colloca il controllo</a>	di durata pluriennale, filtrato - che fa le	<a href="#">Amministrazione_1_1</a>
	0012160 qualificato come capogruppo, che sarà chiamata a	<a href="#">colloca il controllo</a>	in nome e per conto proprio e delle mandati,	<a href="#">Amministrazione_1_1</a>
	... qualificato come capogruppo, che sarà chiamata a	<a href="#">colloca il controllo</a>	in nome e per conto proprio e delle mandati,	<a href="#">Amministrazione_1_1</a>
	del qualificato come capogruppo, che sarà chiamata a	<a href="#">colloca il controllo</a>	in nome e per conto proprio e delle mandati,	<a href="#">Amministrazione_1_1</a>
	gli enti locali privati nel presente caso unico, passarsi	<a href="#">colloca il controllo</a>	di specializzazione ed attività di collaborazione, nonché convegni con	<a href="#">Amministrazione_1_1</a>
	trattamento del risultato una attività di rilevazione economica oggetto di	<a href="#">colloca il controllo</a>	con una pubblica amministrazione, devono sempre applicarsi le regole	<a href="#">Amministrazione_1_1</a>
	... qualificato come capogruppo, che sarà chiamata a	<a href="#">colloca il controllo</a>	in nome e per conto proprio e delle mandati,	<a href="#">Amministrazione_1_1</a>
	... (vedere) che l'Università con la quale si intende	<a href="#">colloca il controllo</a>	esiste la collaborazione sopra indicata - in mancanza di	<a href="#">Amministrazione_1_1</a>
	sono sussistenti misure cautelari e dovrebbe essere di durata temporanea di	<a href="#">colloca il controllo</a>	con la pubblica amministrazione si sono del tipo,	<a href="#">Amministrazione_1_1</a>
	... qualificato come capogruppo, che sarà chiamata a	<a href="#">colloca il controllo</a>	in nome e per conto proprio e delle mandati,	<a href="#">Amministrazione_1_1</a>

Figura 11. RIDIRE Collocate

Usare i corpora (e il *web as a corpus*) per fare didattica vuol dire fare una didattica che mette in luce aspetti testuali basandosi sui legami morfosintattici e lessicali tra le parole. Grazie agli strumenti offerti dalla *corpus linguistics*, molto usati in lessicografia, e in altri campi dell'indagine linguistica, ma fino ad ora poco sfruttati in termini di applicazioni all'insegnamento, usare i corpora (e il *web as a corpus*) per fare didattica è sicuramente una possibilità da tenere in considerazione e da promuovere alla luce dei vantaggi, ormai in varie sedi dimostrati, di uno *student-centered approach*.

### 3.3 Esercizi a scelta multipla corpus-based

Usare i corpora come strumento di formazione e usare corpora come strumento didattico sono due passaggi non necessariamente implicazionali, ma è auspicabile che l'uno preceda l'altro.

Interrogare un *learner corpus* può essere utile ad un insegnante per valutare scientificamente, con un metro quantitativo e qualitativo affidabile, la portata di un errore, validando o smentendo quelle che spesso sono solo sensazioni personali o convinzioni derivate dalla propria esperienza, certamente importanti, ma giocoforza limitate nel tempo e nello spazio. In seguito si può programmare un intervento didattico che faccia uso di corpora di lingua scritta o parlata da sottoporre agli studenti e creare batterie di esercizi mirati a praticare quegli aspetti della lingua che più creano problemi: un corpus annotato per POS (*part-of-speech*) ne rivelerà molti altrimenti non elicetabili, oltre che aprire nuovi orizzonti per l'autoapprendimento guidato da computer.

Riprendiamo, ad esempio, l'interrogazione sulla collocazione 'verbo + foto' per un apprendente di italiano L2/LS e verifichiamo quali sono le occorrenze nei testi prodotti dagli studenti di italiano nel *learner corpus* VALICO. Notiamo subito che circa un quinto dei risultati è frutto di un'interferenza interlinguistica e contiene un'errata scelta del verbo, *prendere* in luogo di *fare*, *scattare*.<sup>11</sup>

Intorno alla <oper> Torre </oper> c'è una grande piazza con collone antiche in mezzo dove si può <b>prendere splendide foto</b> . davanti ai al portoni portone ci sono cavalieri sopra le loro monture con vestiti dell'epoca almohade (
era un gruppo di turistici e una guida che parlava . Dopo , un uomo e una donna ha fatto <b>prendere una foto</b> di loro a un giovane . Nell'autobus , una ragazza stava filmava due uomini che parlavano dei telefoni
figlio ma lui è pigro e vuole ascoltare la musica . La famiglia è grande la mama e la papa <b>prendano il foto</b> insieme vicino alla ruins . # La protagonista è sul treno andando a venezia . Quando scende dal treno fa
la macchina del foto . Guardava sullo sguardo della macchina ed si ha visto nello specchio . Ha deciso per <b>prendere un foto</b> nello specchio , ma dopo lo ha fatto , ha realizzato che era l'ultimo foto . # La donna
città a piedi e a un autobus . E' possibile da la gente sono familia . La geta <b>hanno prendato molte foto</b> e film . La ragazza ha usato la macchina de foto nello autobus . La donna ha rotto una cosa
ancora cominciato a lavorare , lo uso soprattutto per ciarlare con gli amici e con la famiglia , <b>prendere e inviare delle foto</b> o controllare le mie e – mail . Potrebbe sembrare semplice dapprima ma se non avessi un cellulare , sarebbe

Sfruttare questi errori per costruire i distrattori di test a scelta multipla relativi a conoscenze morfosintattiche può essere utile per due motivi: obbliga l'insegnante a mettere a fuoco i piani su cui muoversi, a esplicitare quali sono gli errori gravi, a ragionare metalinguisticamente sulle costruzioni coinvolte e a provare a dare delle spiegazioni. Come fa osservare Colombo (2009) spesso gli elementi che non sembrano degni di nota svelano problemi inaspettati e evidenziano gli effettivi nodi caldi della grammatica dell'apprendente che l'insegnante forse non avrebbe mai associato ad una difficoltà possibile.

D'altra parte questo percorso offre agli studenti dei materiali che non sono costruiti a tavolino sulle ipotesi degli autori dei manuali, ma che rispecchiano effettivamente i processi di apprendimento del discente a livelli diversi e, soprattutto, con provenienza linguistica diversa.

È chiaro infatti che, anche a parità di livello di competenza linguistica in LS/L2, gli errori di uno studente ispanofono non sono quelli commessi da un sinofono, per una diversa distanza tipologica e geografica, per una

11 Tra gli altri verbi usati correttamente, oltre a *fare* (di cui però ci sono solo 6 occorrenze, a fronte di ben 5 di *prendere*), troviamo: *inviare*, *sviluppare*, *mandare*, *mostrare*.

tradizione nell'approccio all'apprendimento profondamente dissimile.

Marello (2009, p. 181) nota come i distrattori studiati a tavolino possono presentare distrattori discutibili se sottoposti al giudizio di parlanti nativi.

Quali sono le frasi corrette?

- a. Il bus fermare.
- b. Il bus ferma.
- c. Il bus fermasi.
- d. Il bus si ferma.

L'esercizio, come osserva Marello, è un test a scelta multipla con una consegna che induce a pensare vi sia più di una risposta corretta. Non è un buon test a scelta multipla perché presenta un distrattore discutibile. «La scelta B, infatti, divide i parlanti nativi: per alcuni è grammaticale, per altri è innaturale nella sua brevità ed accettabile solo come risposta a una domanda, come parte ellittica di una frase più ampia (ad es. alla domanda 'Il bus ferma a Porta Susa?' si risponde 'Il bus ferma, la metropolitana pure'). Sarebbe quindi un test con risposte troppo brevi e si presterebbe a contestazioni, evitabili se le scelte avessero tutte anche l'indicazione di dove il bus ferma» (Marello 2009, p. 181).

Partendo da enunciati veramente prodotti dagli apprendenti è meno facile correre il rischio di proporre scelte innaturali per cui è necessario immaginare contesti d'uso particolari.

Tra i progetti di ricerca legati a VALICO, l'elaborazione di test a scelta multipla corpus-based<sup>12</sup> calibrati per difficoltà, e offerti a gruppi specifici di apprendenti quanto a L1, è quello con ricadute glottodidattiche più immediate.

Come già osservava Marello (2009)<sup>13</sup> nella sua introduzione a esperimenti embrionali sulle scelte multiple corpus-based, la complessità degli errori contenuti nel corpus interessa per la possibilità che offre a studenti e docenti di riflettere sulla loro natura e sulla loro causa; ecco perché agli studenti viene chiesto non solo di scegliere la risposta corretta, ma anche di valutare la parziale correttezza dei distrattori, attribuendo anche ad essi un punteggio decrescente. Si tratta di una modalità di riflessione difficile perché inusuale, ma fruttuosa, perché permette di avviare discussioni metalinguistiche, aprendo una finestra sui processi mentali degli allievi, sulle loro convinzioni, sulla loro competenza (o 'incompetenza').

Nel database online inoltre è possibile accedere a brevi spiegazioni

<sup>12</sup> Sono disponibili su <http://www.valico.org>, corredati inoltre di alcune registrazioni di discussioni in classe.

<sup>13</sup> Per una disamina dopo cinque anni di sperimentazioni con pubblici diversi si veda Corino, E. Marello, C. (in stampa).

sulla natura dell'errore, glosse che non vogliono essere prescrittive, ma semplicemente descrittive al fine di richiamare le conoscenze degli allievi, lasciando loro la libertà di definirne la maggiore o minore gravità.

Si vedano ad esempio le seguenti scelte multiple costruite a partire da alcuni degli errori emersi dall'interrogazione sulla collocazione di 'V+foto':

Un uomo e una donna si sono fatti prendere una foto
Un uomo e una donna si hanno fatto prendere una foto di loro
Un uomo e una donna si sono fatti fare una foto
Un uomo e una donna ha fatto prendere una foto di loro

Seppur un caso limite al confine tra lessico e sintassi che rispecchia una casistica di errori estremamente diffusi tra gli apprendenti, la collocazione non è che uno dei problemi dell'esercizio presentato: le frasi proposte sono complicate da una costruzione causativa alla quale si aggiunge l'uso del clitico. È necessario quindi decidere a quale errore attribuire un maggiore grado di gravità: se C è la risposta corretta, A risulta errata solo in virtù della scelta del collocatore, mentre gli altri distrattori sono complicati anche dalla ridondanza pronominale (B) o dal mancato accordo soggetto-verbo (D).

Ecco come potrebbe essere attribuito il punteggio e come sono stati illustrati i distrattori:

Un uomo e una donna si sono fatti prendere una foto	la collocazione esatta è FARE una foto <b>-3 punti</b>
Un uomo e una donna si hanno fatto prendere una foto di loro	il SI passivante richiede l'ausiliare essere; di loro non è necessario, perché già contenuto nella particella proclitica <b>-2 punti</b>
Un uomo e una donna si sono fatti fare una foto	ESATTA <b>-4 punti</b>
Un uomo e una donna ha fatto prendere una foto di loro	la collocazione esatta è FARE una foto; il soggetto è plurale (uomo+donna); usando l'ausiliare avere è necessario completare la frase con un agente (fare fare una foto A QUALCUNO), meglio sarebbe la costruzione passiva <b>-1 punto</b>

Simile, perché ispirato alla stessa occorrenza nel corpus, l'esempio seguente; in questo caso è stato aggiunto anche l'agente:

Un uomo e una donna ha fatto prendere una foto di loro a un giovane	la collocazione esatta è FARE una foto; il soggetto è plurale (uomo+donna); è necessario usare un SI benefattivo al posto di 'di loro': <i>farSI fare una foto</i> ; l'ausiliare con il SI benefattivo è <i>essere</i> ; l'agente ( <i>un giovane</i> ) deve essere introdotto da DA
Un uomo e una donna si fanno fare una foto da un giovane	ESATTA
Un uomo e una donna si fa prendere una foto da un giovane	la collocazione esatta è FARE una foto; il soggetto è plurale (uomo+donna)
Una donna si fa prendere una foto con un uomo a un giovane	la collocazione esatta è FARE una foto; l'uso del SI benefattivo implica un agente ( <i>un giovane</i> ) che deve essere introdotto da DA

### 3.4 Il web e i registri

L'uso di dati autentici nell'insegnamento delle lingue mostra un ultimo, ma non meno interessante beneficio: lavorando su contesti realmente prodotti dai parlanti, si dedica necessariamente una rinnovata attenzione alle varietà di lingua.

In più sedi (Fiorentino 1997, Bosc 2009, Onesti 2010) è stata in particolare notata una significativa lacuna nel panorama della didattica e della manualistica per le lingue straniere, e per l'italiano in particolare: la mancanza di attenzione nell'educazione linguistica, in L1 come in L2, per le varietà diafasiche.

Si tratta di un vuoto determinante nella definizione di una competenza comunicativa a tuttotondo, soprattutto se si considera che la sua acquisizione è in genere tardiva e difficilmente raggiunge un livello di dettaglio e articolazione tale da permettere all'apprendente di gestire con precisione contesti di registro molto diversi.

Oltre a questioni lessicali, la competenza diafasica contempla la gestione di modalità di mitigazione, termini di cortesia, allocutivi che in situazioni di apprendimento guidato devono necessariamente essere introdotti e presentati

agli studenti. L'obiettivo di un buon insegnamento non deve essere solo la conoscenza delle regole di una lingua, ma anche la capacità di usare la lingua in modo appropriato alla situazione e alle norme socio-culturali, per non incorrere in situazioni in cui lo svantaggio sociolinguistico diventa determinante.

Onesti (2010) riprende alcune osservazioni di Stefano Rastelli e fa notare come nel caso della didattica delle forme di cortesia sia necessario introdurre un insegnamento precoce del pronome allocutivo formale *Lei*, solitamente collocato invece al fondo delle sequenze di apprendimento contemplate da molti manuali, e lo si deve fare proprio in virtù di un insegnamento funzionale ai bisogni dell'apprendente che spesso si trova a dover interagire in contesti formali.

Se si espongono gli studenti unicamente alla lingua dei manuali, essi conosceranno una lingua standard con alcuni caratteri dei contesti formali o informali, ma difficilmente avranno la competenza necessaria per raggiungere l'autonomia nel discernere linguaggi appartenenti a varietà diverse per usi contestuali diversi. Il rischio con apprendenti giovani, inoltre, è che il contatto con la rete e con strumenti di comunicazione mediata quali chat, Facebook, Twitter diano loro degli esempi di lingua che noi sappiamo essere validi solo in quel dato contesto, ma che non potrebbero certo essere utilizzati in conversazioni o testi scritti di altro genere.

Da queste osservazioni nasce l'anima didattico-applicativa del progetto VALERE (Varietà Alte di Lingue Europee in Rete, <http://www.progettovalere.org>) che si propone di riversare in attività ed esercizi quegli aspetti di pragmatica, testualità e morfosintassi legati alla diafasia che l'analisi linguistica può cogliere.

Una piattaforma web collegata a una sezione del corpus NUNC,<sup>14</sup> pensata con attività multimediali ed esercizi interattivi per l'apprendimento - per ora - dell'italiano con uno speciale focus sul registro alto, offre a docenti e studenti esercizi costruiti a partire da testi dei NUNC e consente di lavorare su collocazioni, modi di dire, espressioni idiomatiche molto comuni nell'interazione in rete.

Di fronte per esempio a un modo di dire di cui è difficile - anche per un docente madrelingua - rendere pienamente la sfumatura semantica, sarà il confronto di varie occorrenze di lingua autentica in contesti reali a chiarirne il significato. Oltre alla possibilità di *queries* online, si può ricorrere a batterie di esercizi basate su testi formali, dai semplici *cloze* a domande a scelta multipla e ad attività interattive.

Si veda ad esempio l'esercizio seguente che propone di integrare lacune in due brevi testi scegliendo l'integrazione da una scelta multipla. Sono di questo tipo gli esercizi tratti dal NUNC VALERE.<sup>15</sup>

14 Newsgroup UseNet Corpus, la sezione che costituisce VALERE è stata creata a partire da una selezione di newsgroup che rispondevano a criteri legati ad una connotazione diafasica alta.

15 Bozzone Costa e Piantoni in [http://www.unibg.it/dati/corsi/13005/34997-Esercitazioni2010\\_Prima.pdf](http://www.unibg.it/dati/corsi/13005/34997-Esercitazioni2010_Prima.pdf)

Argomentare significa sostenere o **(1) demolire / confutare / controbattere / oppugnare** una tesi con argomenti efficaci. Lo scopo può essere volere convincere qualcuno a modificare il proprio pensiero, oppure **(2) sospingerlo / invogliarlo / indurlo / stuzzicarlo** a fare qualcosa. Nel primo caso si parla di argomentazione **(3) suadente / persuasiva / pervasiva / persuadente**, nel secondo deliberativa o pratica (Della Casa 1994, p. 450).

La retorica antica dava grande spazio all'insegnamento delle tecniche argomentative; un buon oratore doveva conoscere tutti gli espedienti della lingua e usarli al momento giusto, scegliendo gli argomenti più opportuni per coinvolgere ed emozionare l'uditorio. Un po' come fanno oggi i bravi avvocati. Perry Mason lascia sempre **(4) al discorso / alla prolusione / all'arringa / monologo** finale l'argomento determinante per vincere la causa, tenendo tutti col fiato sospeso fino in fondo.

Avere dei corpora ben preparati permette di lavorare sulla variazione lessicale che è possibile riscontrare in diafasia, di concentrarsi sui possibili sinonimi o quasi-sinonimi di un termine: un apprendente di italiano non può imparare efficacemente tale vocabolario in semplici liste di parole, al di fuori di un contesto. Nel caso specifico, l'analisi del contesto è fondamentale per il riconoscimento della situazione comunicativa, primo obiettivo in termini di consapevolezza diafasica. Il maggiore ostacolo per molti studenti risulta infatti essere, prima ancora che la produzione del termine corretto, il piano ricettivo, ovvero la consapevolezza del registro adatto alla circostanza (calata in una determinata cultura, ma soprattutto resa con un adeguato linguaggio, sia a livello lessicale, sia a livello morfosintattico).

L'esercizio seguente consente di lavorare su entrambi i piani: dapprima si richiede al discente di riconoscere il registro utilizzato nella frase; in seconda battuta è chiamato a completare la frase con uno dei tre termini dati. Si tratta di triadi quali *bugia / menzogna / balla*, che si riferiscono evidentemente allo stesso contenuto semantico, variando tuttavia il registro:

Completare ciascuna frase scegliendo tra le tre parole elencate per ciascun gruppo la più adatta al contesto e rispettando le regole morfo-sintattiche. Specificare poi il registro (basso, medio, alto) della frase così ottenuta:

**1) reggere; sopportare; tollerare**

Nessuno può ... il disprezzo altrui

REGISTRO: \_\_\_\_\_

Il ferito pareva ... l'anima con i denti

REGISTRO: \_\_\_\_\_

Quando dormo, non ... la luce.

REGISTRO: \_\_\_\_\_

**2) aver l'aria di; sembrare; parere**

Non mi ... proprio che tu stia ascoltando!

REGISTRO: \_\_\_\_\_

Ma ti .... il caso?

REGISTRO: \_\_\_\_\_

Quel tizio è sospetto: ... non essere lì per caso.

REGISTRO: \_\_\_\_\_

## 4 Conclusioni

Gli strumenti tradizionalmente usati dalla ricerca linguistica oggi sono stati raffinati a tal punto che possono utilmente essere impiegati da docenti e studenti nella pratica didattica. Se le case editrici fanno ormai da tempo uso di corpora per la costruzione di materiali, dai dizionari al lessico da inserire nei manuali, tali strumenti vengono ancora guardati con sospetto da molti insegnanti che li considerano dispendiosi in termini di tempo da sottrarre alla lezione tradizionale e complessi da usare. Introdurre nella pratica didattica (sia in fase di progettazione che nella prassi in classe) una riflessione linguistica a partire da materiali etichettati per parti del discorso e organizzati in raccolte ordinate, può diventare invece una risorsa importante per stimolare l'autonomia dello studente e la sua capacità di riflessione e costruzione della conoscenza. Gli esempi riportati non sono che un inizio, la strada è ancora lunga, ma ogni piccolo passo di sperimentazione è un grande passo per una didattica nuova e rivolta al futuro.

## Bibliografia

- Barbera, M. (2013). *Linguistica dei corpora e linguistica dei corpora italiana: Un'introduzione* [online]. Disponibile all'indirizzo: [http://www.bmanuel.org/man/Barbera\\_IntroduzioneCL\\_2013=Ver1-60.pdf](http://www.bmanuel.org/man/Barbera_IntroduzioneCL_2013=Ver1-60.pdf).
- Barbera, M.; Marello, C. (2004). «VALICO (Varietà di Apprendimento della Lingua Italiana Corpus Online): una presentazione». In: *Didattica e linguistica dell'italiano come lingua straniera*, II, 4, pp. 7-18.
- Bosc, F. (2009). «Il manuale: un testo difficile per i suoi destinatari». In: Bosc, F.; Mosca, S.; Onesti, C. (a cura di), *Conoscere l'italiano per studiare: Un percorso di auto-formazione*. DVD di aggiornamento docenti. Ufficio Scolastico Regionale Piemonte.
- Campagna, S.; Garzone, G.; Ilie, C.; Rowley-Jolivet, E. (eds.) (2012). *Evolving Genres in Web-mediated Communication*. Bern: Peter Lang.
- Corino, E. (2007). «NUNC est disputandum. Questioni metodologiche e aspetti della testualità». In: Barbera, M.; Corino, E.; Onesti, C. (a cura di), *Corpora e linguistica in rete*. Perugia: Guerra Edizioni, pp. 225-252.
- Corino, E.; Marello, C. (in stampa). *Syntax in multiple choice texts for foreign language learners*. Alessandria: Edizioni dell'Orso.
- Corino, E.; Onesti, C. (2013). «L'argomentazione nei gruppi di discussione italiani: registri formali in rete». In: Carmello, M.; Cacchione, A.; Iasci, M.L. (a cura di), *Unità e molteplicità in Italia, fra lingua e cultura*. Alessandria: Edizioni dell'Orso, pp. 33-68.
- Cummins, J. (2000). *Language, Power and Pedagogy*. Multilingual Matters LTD.
- Fiorentino, G. (1997). «Quale italiano parlano le grammatiche?». In: Calò,

- R.; Ferreri, S. (a cura di), *Il testo fa scuola: Libri di testo, linguaggi ed educazione linguistica*. Atti dell' VIII Convegno GISCEL. Firenze: La Nuova Italia, pp. 109-130.
- Fiorentino, G. (2004). «Scrittura elettronica: il caso della posta elettronica». In: Orletti, F. (a cura di). *Scrittura e nuovi media*. Roma: Carocci, pp. 69-112.
- Fiorentino, G. (2013). «L'italiano delle piazze virtuali». In: Carmello, M.; Cacchione, A.; Iasci, M.L. (a cura di), *Unità e molteplicità in Italia, fra lingua e cultura*. Alessandria: Edizioni dell'Orso, pp. 7-32.
- Kilgarriff, A.; Grafenstette, G. (2003). «Introduction to the web as corpus». *Computational Linguistics*, 29. pp. 333-347.
- Koch, P.; Oesterreicher, W. (1990). *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*. Tübingen: Max Niemeyer.
- Krieger, D. (2003). «Corpus Linguistics: What It Is and How It Can Be Applied to Teaching» [online]. *The Internet TESL Journal*, Vol. IX, 3, March. Disponibile all'indirizzo <http://iteslj.org/Articles/Krieger-Corpus.html> (2014-10-15).
- Leech, G. (1998). «Teaching in language corpora: a convergence». In: Knowles, G.; Mcenery, T.; Fligelstone, S.; Wichman, A. (eds.), *Teaching and language corpora*. Longman pp. 1-22.
- Marello, C. (2009). «Distrattori tratti da corpora di apprendenti di italiano LS/L2». In: Corino, E.; Marello C. (a cura di), VALICO. *Studi di linguistica e didattica*. Perugia: Guerra, pp. 177-193.
- Marello, C.; Colombo, S.; Allora A. (2011). «I corpora VALICO e VINCA: stranieri e italiani alle prese con le stesse attività scritte». In: Marschio, N.; De Martino D. (a cura di), *La Piazza delle lingue L'italiano degli altri*. Firenze: Accademia della Crusca, pp. 49-61.
- Onesti, C. (2010). «Varietà formali in rete. Peculiarità linguistiche e ricadute glottodidattiche». *Italiano LinguaDue*, 1.
- Schmidt, R. (1990). *Input, interaction, attention, and awareness: the case for consciousness-raising in second language teaching*. Paper prepared for presentation at Enpuli Encontro Nacional Professores Universitarios de Lengua Inglesa, Rio de Janeiro.
- Sinclair, J. (2005). «Corpus and Text – Basic Principles» [online]. In: Wynne, M. *Developing Linguistic Corpora: a Guide to Good Practice*, Oxbow Books, Oxford, pp. 1-16. Disponibile all'indirizzo <http://ahds.ac.uk/linguistic-corpora/> (2014-10-15).
- MEDAL (2009) (MacMillan English Dictionary for Advanced Learners), (McEnergy/Gabrielatos 2005).

### **Corpora e strumenti online**

<http://corpus.byu.edu/coca/>

<http://www.lexically.net/wordsmith/>

<https://the.sketchengine.co.uk>

[http://www.corpora.unito.it/photo/nunclit\\_pho.php](http://www.corpora.unito.it/photo/nunclit_pho.php)

<http://www.valico.org>

<http://www.ridire.it>

<http://www.progettovalere.org>

