

Primjena korpusa u učenju talijanskog jezika

Končurat, Natali

Undergraduate thesis / Završni rad

2018

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Pula / Sveučilište Jurja Dobrile u Puli**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:137:538032>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-11-25**



Repository / Repozitorij:

[Digital Repository Juraj Dobrila University of Pula](#)



Sveučilište Jurja Dobrile u Puli

Università Juraj Dobrila di Pola

Fakultet za interdisciplinarne, talijanske i kulturološke studije

Facoltà di studi interdisciplinari, italiani e culturali

NATALI KONČURAT

L'USO DEI CORPORA NELL'INSEGNAMENTO DELL'ITALIANO

ZAVRŠNI RAD / TESI DI LAUREA

Pola, 2018

Sveučilište Jurja Dobrile u Puli
Università Juraj Dobrila di Pola
Fakultet za interdisciplinarne, talijanske i kulturološke studije
Facoltà di studi interdisciplinari, italiani e culturali

NATALI KONČURAT

L'USO DEI CORPORA NELL'INSEGNAMENTO DELL'ITALIANO

Završni rad / Tesi di laurea

JMBAG / N. MATRICOLA: 0269086085

Redoviti student / Studente regolare: Natali Končurat

Studijski smjer / Corso di laurea: Talijanski jezik i književnost / Lingua e letteratura italiana

Predmet / Corso: Korpusna Lingvistika / Lingustica dei corpora

Mentorica / Relatrice: Doc. dr. sc. Ivana Lalli Pačelat

Pula, rujan 2018. / Pola, settembre 2018



IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisana Natali Končurat kandidat za prvostupnika talijanskog jezika i književnosti ovime izjavljujem da je ovaj Završni rad rezultat isključivo mogega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na objavljenu literaturu kao što to pokazuju korištene bilješke i bibliografija. Izjavljujem da niti jedan dio Završnog rada nije napisan na nedozvoljen način, odnosno da je prepisan iz kojega necitiranog rada, te da ikoji dio rada krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili radnoj ustanovi.

Student

U Puli, _____, _____ godine



IZJAVA
o korištenju autorskog djela

Ja, Natali Končurat dajem odobrenje Sveučilištu Jurja Dobrile u Puli, kao nositelju prava iskorištavanja, da moj završni rad pod nazivom L'uso dei corpora nell'insegnamento dell'italiano koristi na način da gore navedeno autorsko djelo, kao cjeloviti tekst trajno objavi u javnoj internetskoj bazi Sveučilišne knjižnice Sveučilišta Jurja Dobrile u Puli te kopira u javnu internetsku bazu završnih radova Nacionalne i sveučilišne knjižnice (stavljanje na raspolaganje javnosti), sve u skladu s Zakonom o autorskom pravu i drugim srodnim pravima i dobrom akademskom praksom, a radi promicanja otvorenoga, slobodnoga pristupa znanstvenim informacijama.
Za korištenje autorskog djela na gore navedeni način ne potražujem naknadu.

U Puli, _____ (datum)

Potpis

INDICE

| | |
|--|----|
| 1. Introduzione | 1 |
| 2. La linguistica dei corpora..... | 2 |
| 3. Il corpus – definizione e tipologia..... | 3 |
| 4. I corpora dell'italiano..... | 5 |
| 5. L'utilizzo dei corpora nell'insegnamento | 7 |
| 6. Insegnare le lingue con l'uso dei corpora | 8 |
| 6.1 Le liste di frequenza..... | 9 |
| 6.2 Le collocazioni..... | 9 |
| 6.3 Le concordanze | 10 |
| 6.4 Word Sketch..... | 10 |
| 7. Esempi e proposte d'uso dei corpora per gli apprendenti dell'italiano | 10 |
| 7.1 Esempi di concordanze | 11 |
| 7.1.1 Esempio di creazione di un esercizio in base alle concordanze..... | 16 |
| 7.1.2 Esempi di collocazioni | 17 |
| 7.1.3 Esempi di Word Sketch..... | 20 |
| 8. Conclusione..... | 25 |
| 9. Bibliografia | 26 |
| 10. Riassunto..... | 29 |
| 11. Sažetak | 30 |
| 12. Summary | 31 |

1. Introduzione

La linguistica dei corpora e i corpora sono presenti nella glottodidattica da più di vent'anni ormai. I corpora vengono considerati da una parte risorse utili per la stesura di materiali didattici, dall'altra parte risorse da usare direttamente nell'insegnamento in classe o nell'apprendimento autonomo da parte degli studenti.

Lo scopo di questa tesi è di analizzare e spiegare le potenzialità e gli eventuali limiti dell'uso dei corpora nell'insegnamento e nell'apprendimento delle lingue, sia materne che straniere. In particolare, la tesi è concentrata sull'utilizzo dei corpora nell'insegnamento e nell'apprendimento della lingua italiana.

La parte introduttiva della tesi è dedicata alla parte teorica nella quale vengono chiariti alcuni concetti di base della linguistica dei corpora. Dopo aver spiegato cosa sia la linguistica dei corpora e cosa siano i corpora in generale, segue una rassegna dei tipi di corpora e dei corpora della lingua italiana scritta. Un'attenzione particolare è rivolta ai tipi di corpora che vengono abitualmente usati nell'insegnamento e a quelli la cui compilazione è mirata esclusivamente all'insegnamento della lingua italiana. Si passa poi ad una descrizione breve dei corpora della lingua italiana esistenti e accessibili in rete come pure degli strumenti utilizzati correntemente nella loro interrogazione. Successivamente vengono illustrati i procedimenti più semplici d'interrogazione dei corpora che non richiedono ulteriori competenze e conoscenze da parte degli utenti, ma che possono venir usati subito sia in classe dall'insegnante che dagli apprendenti senza l'aiuto del docente. Segue un'introduzione teorica sullo sviluppo dei metodi di insegnamento e apprendimento che fanno uso dei corpora (come ad esempio apprendimento guidato da dati e apprendimento per scoperta). Vengono chiarite di seguito le differenze tra l'impiego indiretto e l'impiego diretto dei corpora nell'insegnamento. È proprio quest'ultimo ad essere il centro d'interesse di questa tesi. Si ricorda infatti che l'intento di questa tesi è proprio verificare in prima persona, da parte di un apprendente di italiano, l'utilità concreta dell'impiego diretto dei corpora nell'apprendimento della lingua italiana.

L'ultima parte della tesi è perciò dedicata all'analisi concreta del possibile utilizzo dei corpora da parte degli apprendenti presentando esempi concreti dei risultati delle concordanze, collocazioni e Word Sketch. Gli stessi procedimenti d'interrogazione sono stati eseguiti su corpora diversi e su elementi linguistici di vario tipo appunto per poter illustrarne la loro eventuale utilità. Gli esempi analizzati seguiti dal commento da parte dell'apprendente possono servire da spunto per i futuri docenti e per i futuri apprendenti interessati all'uso dei corpora nell'apprendimento autonomo. La tesi si conclude riassumendo i possibili vantaggi e svantaggi dell'uso dei corpora per

l'insegnamento e apprendimento della lingua italiana rivolgendo l'attenzione anche alle possibili difficoltà che un utente potrebbe riscontrare e fornendo alcuni consigli di come affrontarle per poter sfruttare al massimo le potenzialità che vengono offerte dai corpora.

2. La linguistica dei corpora

La linguistica dei corpora generalmente si potrebbe definire come un insieme di metodi per le indagini linguistiche. Però non tutti sono d'accordo con questa definizione. Ci sono quelli che ritengono che la linguistica dei corpora sia una disciplina linguistica indipendente e sono sostenitori del "corpus as theory" (Leech 1992, Tognini-Bonelli 2001, Teubert 2005 e altri) e quelli che la ritengono solo una metodologia applicabile allo studio della lingua (Gries 2010, Hardie e McEnery 2012, McEnery et al. 2006 e altri). Per il dibattito sullo status della linguistica dei corpora si veda di più in McEnery e Hardie (2012: 147).

Sassi e Ceccotti (2001: 1) spiegano che si tratta di un settore della linguistica computazionale che "si pone l'obiettivo di ottenere da grandi quantità di testi in machine readable form le più svariate informazioni, utili ad un ampio spettro di utenti potenziali: studenti, docenti, studiosi, scrittori, lessicografi, linguisti, critici letterari, traduttori, sociologi, politologi, ecc".

Anche questo è un'idea che non è accettata da parte degli studiosi recentemente, perché anche se agli inizi dello sviluppo della linguistica dei corpora essa faceva parte della linguistica computazionale con l'evolversi delle due discipline, le discipline si sono allontanate ed è meglio considerarle come due discipline assistenti. Si veda di più in McEnery e Hardie (2012: 227-229).

Nella linguistica dei corpora si distinguono due approcci nella ricerca linguistica: quella basata sui corpora (*corpus-based*) e quella guidata dai corpora (*corpus-driven*). Possiamo dire che i due approcci rispecchino le due visioni diverse della linguistica dei corpora. Il primo approccio usa i corpora come ausilio, per trovare esempi e per validare le ipotesi partendo da precedenti presupposti teorici mentre il secondo approccio elabora analisi e descrizioni linguistiche guidate solo dai dati trovati nei testi (cfr. Hardie e McEnery 2012).

Come abbiamo visto non esiste una definizione condivisa e accettata riguardante la natura e lo status della linguistica dei corpora come disciplina linguistica, però quello con cui sono tutti d'accordo è che si basa sui corpora e che quindi privilegia lo studio della lingua basato sui testi autentici rispetto alle intuizioni dei singoli linguisti. Questi testi autentici sono raccolti nei corpora.

3. Il corpus – definizione e tipologia

Quando si parla della linguistica dei corpora il primo concetto chiave che bisogna spiegare è il corpus. Il corpus viene definito di solito come “una collezione di testi autentici conservati e consultabili in formato elettronico che rappresentano il campione di un lingua o di una sua varietà” (Lalli Paćelat 2017). Infatti, come afferma Baroni (2001):

“da un punto di vista statistico, un *corpus* è un campione estratto dalla popolazione di tutti i testi prodotti in una certa lingua, in un certo periodo, in un certo registro, ecc., sulla base del quale possiamo trarre conclusioni che si applicano alla popolazione campionata nel suo insieme”.

Riassumiamo dicendo che si tratta di “una raccolta strutturata di testi in formato elettronico che si assumono rappresentativi di una data lingua o di un suo sottoinsieme, mirata ad analisi di tipo linguistico” (Spina 2011: 64).

Il tipo di analisi che consente un corpus dipenderà da molti fattori. Per cui, prima di tutto è importante conoscere la struttura dei corpora e la sua estensione per poter valutare la sua rappresentatività e di conseguenza la validità dei dati ricavati dal corpus.

La realizzazione di un corpus elettronico deve essere pertanto ben pianificata e seguire le indicazioni per la costruzione dei corpora per poter soddisfare gli standard della linguistica dei corpora. La realizzazione di un corpus può essere frazionata in tre periodi differenti secondo Spina (2001: 64): “la fase progettuale, che comporta una serie di scelte tra cui quella del tipo di corpus, dei testi da selezionare e della loro estensione, la fase dell'acquisizione vera e propria dei dati e quella della loro codifica”.

Infatti è il modo in cui un corpus viene codificato e annotato a determinarne l'uso. Mentre i corpora della prima generazione non erano annotati, oggi quasi tutti i corpora sono riccamente annotati e lemmatizzati.

Per annotazione si intende “l'aggiunta di informazioni di carattere linguistico ed extralinguistico ai testi che compongono un corpus” Nosedà (2015/2016: 27). Tamburini (2000: 75) spiega che

“annotare grammaticalmente un corpus significa, data la sequenza di parole che lo compongono, assegnare ad ognuna di esse la corretta parte del discorso, risolvendo tutte le ambiguità intrinseche nel processo di classificazione grammaticale del linguaggio.”

Oltre a questo tipo di annotazione che viene chiamato POS-tagging (part-of-speech tagging), esistono anche annotazioni di altre categorie grammaticali.

La lemmatizzazione, invece secondo Bolasco (1999: 191) viene definita come: “riconoscimento della categoria grammaticale di una parola, che produce la riconduzione della forma grafica al lemma di appartenenza”.

Proprio il fatto di essere annotato e lematizzato fa distinguere “un corpus, da un semplice insieme di testi, rendendolo particolarmente adatto ad analisi linguistiche piuttosto elaborate” (cfr. Nosedà 2015/2016: 27).

Esistono vari tipi di corpora e varie tipologie. Come specifica Viganò (2000: 117) i “*corpora* si possono classificare a seconda degli scopi che l’investigazione degli stessi si propone”.

Secondo Lalli Paćelat (2016) per la glottodidattica è importante distinguere “i corpora generali o di riferimento da quelli specialistici, i monolingui da quelli multilingui, quelli scritti dagli orali o misti, e i corpora dei madrelingua dai corpora degli apprendenti”.

L’autrice spiega che i corpora generali o di riferimento contengono centinaia di milioni di parole e “si considerano rappresentativi di una determinata lingua in generale, mentre i corpora specialistici sono di dimensioni più piccole e vengono costruiti per uno scopo predeterminato” (Lalli Paćelat 2016).

I *corpora specialistici* sono, invece molto utili solamente nell’apprendimento e nell’insegnamento delle lingue speciali perché “interrogando tali corpora, gli apprendenti e i docenti “con un paio di click del mouse possono individuare le principali caratteristiche lessico-grammaticali di una varietà linguistica (terminologia di base, espressioni frequenti, strutture grammaticali tipiche ecc.)” Lalli Paćelat (2016).

Se si prende in considerazione la lingua in cui sono scritti i testi che compongono, il corpus, bisogna fare una distinzione tra *corpora monolingui* e *corpora bilingui*. Tra i corpora bilingui si distinguono generalmente due categorie: *i corpora comparabili* e *i corpora paralleli*. I corpora comparabili sono composti da testi originali in due lingue diverse raggruppati secondo determinati criteri di somiglianza. I corpora paralleli contengono testi originali in una lingua e le rispettive traduzioni in una o più lingue e “consentono un’analisi contrastiva interessante sia per gli apprendenti che per i docenti” (Lalli Paćelat 2016).

Concludiamo questo capitolo indicando ancora un tipo di corpus interessante per la glottodidattica e precisamente il *learner corpus* o corpus degli apprendenti. Si tratta di corpora formati da testi prodotti da parlanti non nativi, molto spesso da persone che stanno apprendendo una lingua straniera (cfr. Granger et al. 2002). Osservando e analizzando gli errori più frequenti, spiega Lalli Paćelat (2016), gli insegnanti possono prevedere quali difficoltà potranno avere gli apprendenti di una data lingua e, tenendone conto, preparare attività e materiali didattici mirati. Si veda di più in McEnery e Xiao (2010: 372).

4. I corpora dell'italiano

Dopo aver spiegato cosa sia un corpus, come viene compilato e preparato per le ricerche linguistiche e dopo aver esposto in modo dettagliato i tipi di corpus che possono avere un impiego diretto o indiretto nell'insegnamento e apprendimento delle lingue, passiamo alla presentazione dei corpora della lingua italiana.

Per ragioni storiche, spiega Baroni (2010) bisogna menzionare “i lavori pionieristici di padre Roberto Busa [...], e il primo vero e proprio corpus di riferimento della lingua italiana, che portò alla pubblicazione del Lessico di frequenza della lingua italiana contemporanea (LIF = Bortolini, Tagliavini e Zampolli 1971).” Il LIF, continua Baroni (2010) è stato usato per fare la lista dei lemmi che sono entrati nel Vocabolario di base della lingua italiana.

Nell'elenco di Baroni (2010) si trova ancora Il Corpus e lessico di frequenza dell'italiano scritto (CoLFIS) (Laudanna et al. 1995¹), un corpus di riferimento di piccole dimensioni, ma compilato usando criteri alti di campionamento. Per una panoramica completa dei principali corpora disponibili per lo studio dell'italiano si vedano Cresti e Panunzi (2013).

In seguito verranno presentati solo quei corpora dell'italiano che potrebbero essere interessanti e d'aiuto agli apprendenti d'italiano L2 e ai loro insegnanti.

Un corpus di riferimento dell'italiano scritto della dimensione di 100 milioni di parole liberamente consultabile in rete è il Corpus di italiano scritto (CORIS/CODIS) che si aggiorna “tramite un corpus di monitoraggio inglobato con cadenza triennale²” Rossini Favretti (2000). Per ulteriori informazioni si veda Rossini Favretti (2000) e Rossini Favretti et al. (2002).

Un altro corpus che si propone come corpus di riferimento dell'italiano contemporaneo, scritto e parlato è Il Perugia Corpus (PEC). Si tratta di un corpus di piccole dimensioni, però liberamente consultabile e con un'interfaccia di consultazione molto avanzata. Per ulteriori informazioni si veda Spina (2014).

Un corpus di dimensioni maggiori liberamente consultabile in rete che contiene testi di quotidiani italiani pubblicati tra il 1985 e il 2000 è La Repubblica³ (Baroni et al. 2004).

¹Ulteriori informazioni disponibili sul sito <http://www.istc.cnr.it/material/database/colfis/> (consultato il 25/6/2018)

²Ulteriori informazioni riguardo al corpus e l'interrogazione del corpus sono disponibili sul sito http://corpora.dslo.unibo.it/coris_ita.html (consultato il 25/6/2018)

³Il corpus è oggi consultabile dal sito:

https://corpora.dipintra.it/public/run.cgi/first?corpname=repubblica&reload=1&iquery=&queryselector=iqueryrow&lemma=&phrase=&Word=&char=&cql=&default_attr=Word&fc_lemWord_window_type=both&fc_lemWord_wsize=5&fc_lemWord=&fc_lemWord_type=all&usesubcorp=&fsca_text.id=&fsca_text.Wordcount= (consultato il 5/7/2018)

L'ItWaC è il più grande corpus utilizzabile per la lingua italiana. Baroni (2010) spiega che il corpus ItWaC “riflette la recente tendenza a costruire corpora raccogliendo testi web con procedure automatiche”. L'ItWaC appartiene alla collezione di corpora WaCky (Baroni et al. 2009: 210), alla quale appartengono corpora di varie lingue, incluso quello della lingua croata hrWaC (Ljubešić e Erjavec 2011). Tutti i corpora sono liberamente accessibili agli utenti tramite un'interfaccia di consultazione molto avanzata⁴. Il corpus ItWaC appartiene alla famiglia itTenTen che viene spiegato come “un corpus italiano composto da testi raccolti dall' internet”⁵. Si veda di più sul itTenTen in Jakubíček e al. (2013) e in Suchomel e Pomikálek (2012).

Esistono, inoltre, altri corpora che sono meno usati nell'insegnamento dell'italiano perché che comunque vanno menzionati come ad esempio i *Corpora e Lessici di Italiano Parlato e Scritto – CLIPS* (Leoni 2003)⁶, Il Corpus di Italiano Televisivo (CiT) (Spina 2001)⁷.

Oltre a questi corpora italiani precedentemente elencati la cui compilazione non era mirata esclusivamente all'insegnamento e all'apprendimento dell'italiano, ma allo studio dell'italiano in genere, presenterò in seguito due corpora che sono stati usati nel processo di insegnamento della lingua italiana come lingua straniera e precisamente il corpus *VALICO e PAISÀ*. Ambedue i corpora trovano l'uso tra gli insegnanti e tra gli studenti per l'apprendimento della lingua italiana. Quindi il corpus *Valico* ha trovato il suo posto principale tra gli studenti della lingua italiana. “Il corpus è costruito secondo criteri di bilanciamento legati alle L1 di provenienza degli apprendenti, agli stimoli iconici usati per elicitare i testi, ai livelli di interlingua dichiarati”⁸.

Oltre a *Valico* esiste anche il progetto *Paisà* (Piattaforma per l'Apprendimento dell'Italiano Su corpora Annotati). “Il progetto si rivolge in particolar modo agli emigrati di seconda generazione, che hanno l'italiano come lingua materna ma ne fanno un uso molto limitato, e di terza generazione, che parlano l'italiano come seconda lingua (L2)”⁹.

⁴ Il corpus è consultabile sul sito: <https://www.Sketchengine.eu/itwac-italian-corpus/> (consultato il 09/6/2018)

⁵ <https://www.sketchengine.eu/ittenten-italian-corpus/#toggle-id-3> (consultato il 11/6/2018)

⁶ Per ulteriori informazioni si consulti il sito: <http://www.clips.unina.it/it/> (consultato il 09/6/2018)

⁷ Per ulteriori informazioni si consulti il sito: http://www.culturitalia.info/ARCHIVIO/s_spina/cit/cosa.html (consultato il 10/6/2018)

⁸ Il corpus è consultabile sul sito: <http://www.valico.org/valico.html> (consultato il 28/5/2018)

⁹ Il corpus è consultabile sul sito: <http://www.lingue.unibo.it/it/ricerca/progetto-paisa> (consultato il 29/5/2018)

5. L'utilizzo dei corpora nell'insegnamento

Lo sviluppo della linguistica dei corpora ha portato grandi cambiamenti in diversi campi della linguistica, inclusa la glottodidattica. La linguistica dei corpora viene usata nella glottodidattica da più di vent'anni.

Parlando dell'uso dei corpora nell'insegnamento e nell'apprendimento delle lingue, normalmente si distinguono l'impiego diretto e indiretto (Römer 2008).

L'impiego indiretto è rivolto ai ricercatori e ai compilatori di materiali didattici. I corpora vengono usati nella stesura di vocabolari, grammatiche, libri d'insegnamento, manuali e altri materiali didattici, come pure nella loro valutazione. Grazie al fatto che contengono testi autentici e che offrono l'informazione sulla frequenza sono importanti nella scelta e nella valutazione del materiale linguistico da inserire nei testi didattici. Infatti i corpora offrono ai docenti "una consultazione rapida di milioni di testi prodotti in contesti comunicativi reali nei quali poter verificare o trovare il miglior esempio di quello che intendono insegnare" (Paćelat 2016).

I corpora non si usano soltanto nella compilazione dei vari materiali didattici, i corpora trovano impiego diretto in classe nelle attività didattiche proposte dai docenti oppure nello studio autonomo della lingua da parte degli apprendenti. L'impiego diretto è quindi rivolto sia ai docenti che agli apprendenti. Infatti, i corpora consentono di osservare l'uso concreto di una lingua e di "verificarne tendenze generali su base statistica" (Lüdeling e Kytö 2008). I corpora aiutano gli apprendenti "nella ricerca della parole o dell'espressione adeguata in base al loro contesto, usando l'analisi delle concordanze" (Riccardi 2003: 27), ma anche delle collocazioni o del WordSketch.

Secondo Leech (1997: 23) l'uso diretto dei corpora include: "insegnare", "insegnare a sfruttare" e "sfruttare per insegnare". McEnery e Xiao (2010: 370) precisano che insegnare significa insegnare la linguistica dei corpora come qualsiasi altra disciplina della linguistica. Inoltre, il docente deve saper utilizzare i corpora cioè sfruttarli per l'insegnamento delle lingue, ma deve anche saper insegnare e spiegare agli apprendenti come si usano i corpora, affinché anche gli apprendenti possano approfittare del loro uso (cfr. McEnery e Xiao 2010: 370).

McEnery e Xiao (2010:370) sottolineano se *insegnare* e *sfruttare per insegnare* sono dedicati soltanto agli studenti di linguistica, *insegnare a sfruttare* è rivolto agli studenti di tutte le materie che coinvolgono lo studio e l'apprendimento delle lingue e che dovrebbero beneficiare del cosiddetto *data-driven learning* (DDL)".

Il concetto di *data-driven learning* (DDL) è stato coniato agli inizi degli anni Novanta da Johns (1991) "per descrivere le procedure con le quali le concordanze di particolari fenomeni linguistici possono essere analizzate dagli apprendenti per inferire e verificare generalizzazioni riguardanti il

loro uso seguendo un approccio induttivo o deduttivo” (Guidetti et al.2012). Nell’approccio induttivo gli apprendenti sono guidati dai dati del corpus e apprendono per scoperta aspetti della lingua. Il termine “apprendimento per scoperta” è stato invece coniato da Bernardini (2004).

Nel tipo di approccio deduttivo, secondo (Guidetti et al. 2012) gli apprendenti usano i corpora per verificare “in prima persona le regole che hanno appreso e con ciò consolidando e/o perfezionando le loro conoscenze”. Boulton (2008: 1) ritiene inoltre che l'apprendimento guidato dai dati (DDL) implica che gli studenti possano accedere a dati linguistici sufficienti e che possono scoprire le regole e imparare da soli. Secondo Boulton (2008: 1) l’uso del DDL ha molti benefici, perché promuove l’autonomia dell’apprendente e lo fa riflettere sugli aspetti della lingua. Ma nonostante tutti i vantaggi e benefici dell’uso diretto dei corpora in classe Boulton (2008: 1) osserva che i corpora non vengono usati da tutti. Mukherjee (2006: 11) ritiene che probabilmente gli insegnanti e gli studenti siano semplicemente inconsapevoli delle potenzialità che i corpora possono avere nell’insegnamento. McEnery e Xiao (2010: 365) concludono che l’uso dei corpora nell’insegnamento e nell'apprendimento delle lingue è stato fin’ora “più indiretto che diretto” forse perché l’uso diretto dei corpora nell’insegnamento delle lingue dipende da una serie di fattori tra cui, ad esempio, il livello di competenza degli apprendenti, i limiti di tempo e delle aule d’insegnamento, il livello di competenza e le abilità informatiche da parte degli insegnanti, ecc. Sono infatti proprio gli insegnanti ad avere un ruolo fondamentale, in quanto l’esplorazione dei corpora da parte degli studenti deve essere all’inizio mediata dall’insegnante perché è lui che illustra l’utilizzo di determinati strumenti di gestione e interrogazione dei corpora (Viganò 2011: 124). Man mano che lo studente acquisisce maggiore familiarità con gli strumenti e le tecniche della linguistica dei corpora e di conseguenza maggiore autonomia nella ricerca, il ruolo di guida dell’insegnante diventa meno importante e l’apprendente diventa una sorta di ricercatore (cfr. Bernardini 2004, Viganò 2011: 124, Lalli Pačelat 2016).

Si veda di più sul *data-driven learning* (DDL) in Chambers (2010) e invece per sapere come può essere usato il DDL nell’ambito dell'apprendimento delle lingue si veda in Gilquin e Granger (2010).

6. Insegnare le lingue con l'uso dei corpora

Come illustrato nel capitolo precedente, il corpus rappresenta una risorsa preziosa per gli insegnanti dando loro l’opportunità di vedere a quali parole e a quali strutture linguistiche dare la priorità nell’insegnamento. Infatti in base alla frequenza potranno scegliere il materiale linguistico più adatto all’insegnamento dei loro apprendenti (cfr. Corino 2014: 235). *Quali aggettivi di solito*

accompagnano un determinato nome? Quali avverbi seguono solitamente un determinato verbo e quali preposizioni? Quali sono le espressioni più frequenti? sono solo alcune delle domande a cui un docente trova risposta nei corpora.

L'utilizzo dei corpora nell'apprendimento della lingua e precisamente le concordanze e le collocazioni permettono all'apprendente, come specificato da Corino (2014: 239), di individuare i "potenziali diversi significati delle parole e dei sintagmi ricorrenti". L'apprendente potrà trovare anche le collocazioni tipiche di certi contesti, osservare strutture diverse a seconda dei registri e dei tipi di testo nei quali compaiono (cfr. Corino 2014: 239). Inoltre, l'insegnante inserendo in classe attività didattiche incentrate sui corpora stimola gli studenti al lavoro di gruppo.

Inoltre, si tratta per lo più di attività che usano un approccio induttivo e richiedono una riflessione e un alto grado di partecipazione da parte degli apprendenti.

La gran parte dei corpora è oggi disponibile ad un vasto pubblico grazie a degli strumenti di gestione e d'interrogazione dei corpora facili da usare. Sono questi strumenti di gestione e interrogazione dei corpora ad offrire liste di frequenza, concordanze, collocazioni e recentemente anche il Word Sketch, come verrà spiegato nei capitoli seguenti.

6.1 Le liste di frequenza

Secondo Spina (2001: 106) "un computer è facilmente in grado di distinguere, ordinare e contare gli elementi che compongono un qualsiasi testo". Spina (2001: 106) continua spiegando che tre operazioni e precisamente il reperimento, l'apprendimento e il conteggio "combinare con una serie di parametri di tipo statistico, costituiscono la base di ogni analisi quantitativa del linguaggio".

Sono appunto le liste di frequenza ad offrire indicazioni "preziose nel campo del *syllabus design* e nella preparazione dei materiali didattici, in particolare modo per la scelta del lessico e della sintassi da affrontare fin dai primi stadi di apprendimento della lingua straniera" (Guidetti et al. 2012). È chiaro che in base alle liste di frequenza l'insegnante può scegliere quali sono gli elementi solitamente più usati.

6.2 Le collocazioni

Le collocazioni vengono definite nella linguistica dei corpora diversamente rispetto alle altre discipline linguistiche. Per collocazioni, nella linguistica dei corpora, si intende generalmente la co-occorrenza statisticamente significativa di due o più parole (Lalli Pačelat 2016). Molto simile, anche Viganò (2001: 120) spiega che "il termine collocazione si riferisce a un qualunque tipo di relazione sintagmatica tra parole che co-occorrono". Spina (2001: 136) invece aggiunge che l'analisi delle collocazioni "oltre che su considerazioni e metodi di tipo statistico, si appoggia

sull'osservazione di alcuni fenomeni che ricorrono in molte delle lingue più diffuse”.

6.3 Le concordanze

Secondo Spina (2001: 133) le concordanze sono “un sofisticato metodo di ricerca e visualizzazione di parole chiave all'interno del loro contesto, che permette di individuare le regolarità, anomalie, usi più o meno ricorrenti di determinati segmenti di lingua”. Quindi si tratta dei risultati di ricerca di una parola nel suo contesto visualizzate generalmente nel formato più usato quello della KWIC (*Key-Word-In-Context*).

Nella Figura 1. sono rappresentate alcune concordanze della parola *casa* dimostrando come è possibile espandere ogni riga della concordanza per ricevere il contesto nel quale la parola viene utilizzata.

The image shows a list of KWIC (Key-Word-In-Context) search results for the word "casa". Each result consists of a line of text with "casa" highlighted in red. A yellow tooltip is overlaid on one of the results, showing a double downward arrow icon and expanded context. The tooltip text includes: "< previous possono impiegare in qualche lavoro. Gli stabilimenti dove il lavoro è organizzato, tengono gli uomini lontani da casa per quasi tutta la giornata, con orari e stipendi spesso penosi. Anche le mamme compiono spesso lavori fuori casa . Così i figli vivono nelle strade, in piccole bande, finché non hanno l'età sufficiente per lavorare. </p><p> A questa triste situazione si può rimediare mediante le scuole cristiane, in cui si insegna gratuitamente e solo < next >".

Figura 1. Esempio d'espansione del contesto per la parola *casa* nel corpus ItWaC

6.4 Word Sketch

Una ricerca più avanzata è quella del *Word Sketch* che è possibile fare con i corpora che usano l'interfaccia di Sketch Engine. Efrati (2008: 38) spiega che Word Sketch “costruisce riassunti automatici del comportamento grammaticale di una parola in un corpus” e continua chiarendo che “piuttosto che osservarla in una finestra arbitraria di testo, la funzione Word Sketch permette di osservare le relazioni grammaticali in cui la parola si manifesta.”

7. Esempi e proposte d'uso dei corpora per gli apprendenti dell'italiano

Dopo aver spiegato le tecniche della linguistica dei corpora solitamente più usate nell'apprendimento e insegnamento, passiamo ora alla parte pratica di questa tesi.

Esistono vari tipi di corpora che possono essere utili nell'insegnamento. Noi abbiamo scelto di usarne alcuni proprio per dimostrare il loro uso e le loro eventuali potenzialità e limiti che possono avere nell'impiego diretto nell'insegnamento e apprendimento dell'italiano. Di solito nell'apprendimento si utilizzano i corpora che sono disponibili online o su Sketch Engine¹⁰ e quindi in questa tesi abbiamo scelto di presentare gli esempi presi dai seguenti corpora itWaC, itTenTen, Perugia Corpus (PEC) e il corpus PAISÀ.

Nei capitoli successivi presentiamo diversi esempi delle analisi delle concordanze, delle collocazioni e dei Word Sketch commentando la loro eventuale utilità nell'insegnamento e nell'apprendimento dell'italiano.

7.1 Esempi di concordanze

Come già spiegato nel capitolo 6.3 le concordanze permettono di visualizzare la parola con il suo contesto e di individuare delle regolarità o delle anomalie nel loro uso come anche di trovare gli usi tipici e più frequenti di una parola (cfr. Spina 2001: 133). Vediamo ora, alcuni esempi di concordanze e della loro eventuale utilità nell'insegnamento dell'italiano.

I primi tre esempi sono dedicati al verbo *dare* (Figura 2, 3 e 4). Seguono tre tipi di concordanze ricavati da tre corpora diversi, ItWaC, Perugia Corpus (PEC) e il corpus PAISÀ. Il primo è un web corpus, il secondo si propone di essere un corpus di riferimento dell'italiano scritto e parlato, mentre l'ultimo è stato realizzato per apprendenti di italiano.

| | | | |
|---------------|--|---------|--|
| innomedima... | dunque chiesto: "Chi sei? Perché possiamo | dare | una risposta a coloro che ci hanno mandato. Che |
| innomedima... | con Giuseppe. </p><p> [44]A dire il vero, Gesù stesso ha | dato | la prova che un profeta non ottiene |
| innomedima... | , nella sinagoga e si alzò a leggere. 17 Gli fu | dato | il rotolo del profeta Isaia; apertolo trovò il |
| innomedima... | cose? E che sapienza è mai questa che gli è stata | dato | ? E questi prodigi compiuti dalle sue mani? 3 Non |
| innomedima... | : "Signore, perché non posso seguirti ora? | Darò | la mia vita per te!". 38 Rispose Gesù: "Darai la |
| innomedima... | ora? Darò la mia vita per te!". 38 Rispose Gesù: " | Darai | la tua vita per me? In verità, in verità ti dico: |
| chiacchier... | (navetta principale); nelle catene l'anima è | dato | dalla navetta principale mentre il corpo è dato |
| chiacchier... | data dalla navetta principale mentre il corpo è | dato | dalla navetta secondaria! Per questo si dice |
| chiacchier... | navetta. Potremo osservare che l'anima è | dato | dalla navetta principale mentre il corpo è dato |
| chiacchier... | data dalla navetta principale mentre il corpo è | dato | dalla navetta secondaria! </p><p> Come attaccare con |
| evabasile... | deciso di ufficializzare le nostre attività, | dandoci | veste di associazione culturale: ero nel |
| agenziahel... | , la comunicazione, il lavoro di squadra, come | dare | direttive, creare un team vincente sono alcune |
| rai.it | ultimo, ma non per ultimo </p><p> ", la manifestazione | dà | spazio alla Rete, con il Future Web Festival una |
| rai.it | sfuggire e ricomparire parole e immagini, e | dando | a tutto ciò un forte impatto sonoro, una |
| italianpai... | nelle anguste misure di una tela, dove nulla è | dato | al caso e dove masse e colori ridistribuiscono |
| gregoriano... | più prestigiosi cori a livello internazionale | darsi | appuntamento, come negli anni scorsi, nei |
| soarecords... | non vedrò mai e che sicuramente mi | dà | meno emozioni... In un paio di occasioni mi sono |
| soarecords... | avevano già i blog che emettevano effluvi e ti | davano | il buongiorno alla mattina. È un po' un modo di |
| soarecords... | Skinhead, Oi! bénd pluriristampata che ha | dato | indubbiamente lustro all'italica scena e che |
| soarecords... | all'italica scena e che in qualche modo ha | dato | il "15" al recupero della struttura della |

10 Si veda di più su Sketch Engine su <https://www.sketchengine.eu/user-guide/user-manual/corpora/by-language/italian-text-corpora/> (consultato il 11/6/2018)

Figura 2. Concordanze del verbo *dare* prese dal corpus ItWaC

La Figura 2 riporta i risultati delle concordanze del verbo *dare*, che come si può notare, è un verbo molto frequente. In base alle concordanze lo studente potrà notare che si tratta di un verbo transitivo come viene visto nella prima riga: [...] *Perché possiamo dare una risposta a coloro* [...].che forma il passato prossimo soltanto con l'ausiliare *avere*, visibile nelle ultime due righe. Lo studente noterà che viene usato anche l'ausiliare *essere*, però che si tratta di forme passive. Inoltre, lo studente potrà osservare le possibili forme del verbo *dare* e analizzarle o semplicemente individuarle. A seconda del livello di competenza linguistica e grammaticale degli studenti e dell'obbiettivo didattico è l'insegnante che dovrà decidere il tipo di attività da proporre in classe. Osservando il verbo *dare* nel suo contesto e le parole che lo circondano lo studente noterà delle espressioni utili come ad esempio *dare appuntamento*, *dare il buongiorno* e molte altre.

Oltre a corpus italiano ItWaC ho deciso di mostrare il verbo *dare* anche nel Perugia Corpus (PEC) e PAISÀ.

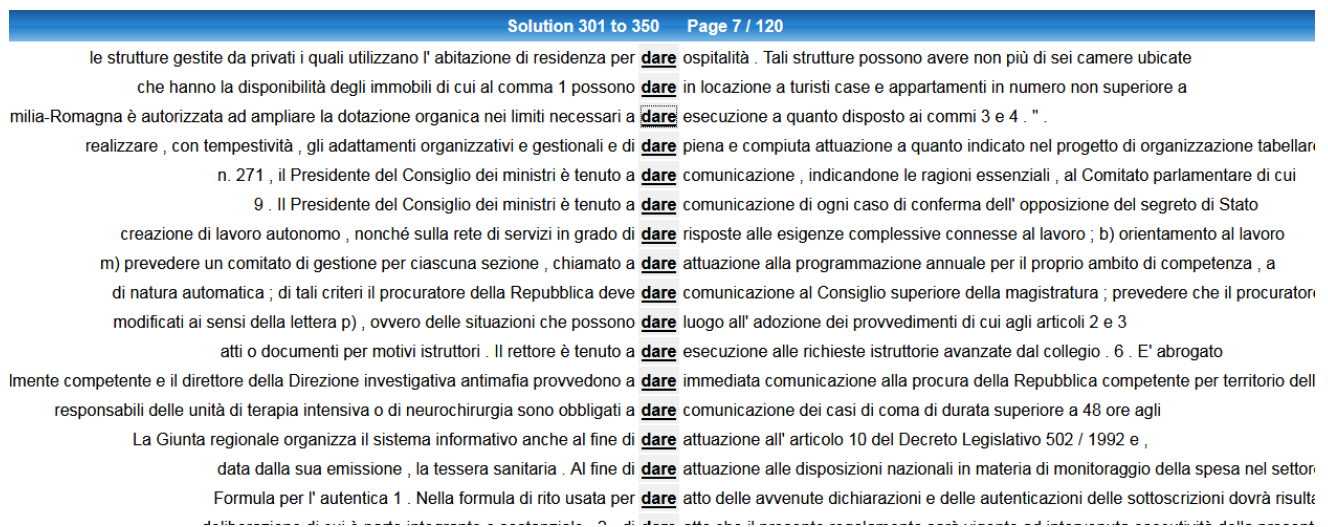


Figura 3. Concordanze del verbo *dare* prese dal Perugia Corpus

La Figura 3 riporta i risultati delle concordanze per il verbo *dare* generate nel Perugia Corpus (PEC). Il Perugia Corpus (PEC), si ricorda, viene considerato un corpus di riferimento dell'italiano contemporaneo. Si veda di più nel capitolo 4. Osservando gli esempi presentati nella Figura 3 uno studente può facilmente individuare alcuni usi frequenti del verbo *dare* nelle situazioni nelle quali l'azione del verbo, che porta un significato generico, è determinata dal sostantivo che lo segue. Troviamo così tra gli esempi riportati nella Figura 3 locuzioni come *dare comunicazione*, *dare*

attuazione, dare luogo, dare atto ecc. Il docente può usare questi esempi per creare esercizi mirati all'arricchimento lessicale. Ad esempio per formazione della parole: *dare comunicazione /comunicare* e *dare attuazione/attuare* ecc.

Nella Figura 4 possiamo vedere i risultati delle concordanze ricevute per lo stesso verbo, però in un corpus diverso.

| | | |
|--|------|--|
| meridionale vuole a tutti i costi divincolar- si , per | dare | piena libertà alle sue passioni . |
| si ha la pretesa della completezza , ma si vuole | dare | almeno un' idea delle varietà di tecniche , problemi e |
| e di attrazione con gli altri amminoacidi , necessarie a | dare | la forma sferica alla parte proteica dell' emoglobina costituita da |
| Se le rape non dovessero | dare | abbastanza umore da coprire il tutto (ma succede raramente |
| amento o integrazione è accettato ed anzi gradito , per | dare | uno strumento operativo utile a chi si trova a dover- |
| lmente per evitare interpretazioni campanilistiche e per | dare | una base di clausole comuni ai partecipanti del negozio giuridico |
| è necessario se ne aggiunga un terzo in grado di | dare | origine a interazioni idrofobiche o di van del Waals . |
| tipo e possono essere utilizzati in più colori , per | dare | un aspetto diverso alla parete . |
| di sono quindi investiti con forza della responsabilità di | dare | un seguito ad una volontà di cambiamento espressa in modo |
| ondo è usanza comune festeggiare la fine dell' anno e | dare | il benvenuto all' anno che viene , augurando- si prosperità |
| ività industriale che in maniera concreta contribuisse a | dare | solievo alla disoccupazione che rappresenta un problema atavico per Corigliano |
| Ospedale di Belgrado rifiuta di | dare | un bimbo alla propria madre . |
| L' Istituto di neonatologia di Belgrado ha rifiutato di | dare | un bimbo neonato a sua madre , poiché questa non |
| come una « azione terroristica » e ha promesso di | dare | la caccia ai responsabili . |
| la polizia , la moto non si sarebbe fermata a | dare | la precedenza alla vettura che procedeva a velocità normale , |

Figura 4. Concordanze per il verbo *dare* prese dal corpus PAISÀ.

Esempi interessanti dell'uso del verbo *dare* si possono trovare anche nel corpus PAISÀ, come riportato nella Figura 4. Si tratta di un corpus realizzato appositamente per l'apprendimento dell'italiano nell'ambito del omonimo progetto PAISÀ (Piattaforma per l'Apprendimento dell'Italiano Su corpora Annotati). Si veda di più nel capitolo 4.

Siccome il corpus ItWaC e ItTenTen16 sono per ampiezza i più grandi corpora della lingua italiana e sono disponibili su Sketch Engine, come precedentemente spiegato nel capitolo 6 ho deciso di usarli per il resto degli esempi che presenterò in questa tesi. Il motivo di questa scelta è di natura pratica. Infatti, Sketch Engine ha un'interfaccia semplice per gli utenti però permette allo stesso tempo di fare anche ricerche avanzate.

Un elemento, spesso, problematico e l'articolo determinativo *lo* e il suo uso (Figura 5).

| | |
|---------------|---|
| ilpost.it ... | Channel Islands al largo della California, lo Stretto di Cook in Nuova Zelanda, lo Stretto di |
| ilpost.it ... | , lo Stretto di Cook in Nuova Zelanda, lo Stretto di Tsugaru in Giappone, lo Stretto di |
| ilpost.it ... | Zelanda, lo Stretto di Tsugaru in Giappone, lo Stretto di Gibilterra (il più corto, di circa 13 |
| villaodino... | fra i migliori della provincia di Venezia, dove lo Charme e l'ospitalità si danno appuntamento |
| hotelfedai... | DEL PAN. </p><p> Una bella passeggiata in quota, sotto lo sguardo del Vernel per poi finire abbracciati |
| disinforma... | cereali come frumento, orzo , segale e avena. Lo studio sostiene che un maggior uso del |
| milano.it ... | questa Associazione. Roberto Mattei pratica lo Yoga da oltre 35 anni. E' Consigliere e |
| hotelbeaus... | la sua privilegiata posizione geografica, che lo colloca sul mare, ma con alle spalle le dolci |
| astori.it ... | dei dati personali degli utenti che lo consultano. </p><p> Si tratta di un'informativa che è |
| astori.it ... | in risposta, il codice numerico indicante lo stato della risposta data dal server (buon fine |
| chiesacatt... | nel quale si è compiaciuto" (Mt 3,17), ma lo ha consegnato all'umanità con un gesto di amore |
| chiesacatt... | gli uomini sanno dire la parola giusta. Se lo sapessero si risparmierebbero la miseria e l' |
| cinaoggi.i... | artistico cinese contemporaneo e con lo sviluppo di un clima socio/culturale idoneo il |
| istitutopa... | maggior numero possibile di pubblicazioni che lo riguardano personalmente, anche in relazione |
| virgilio.i... | a supporto di questa istituzione (una persona lo segue ancora attualmente una volta a settimana |
| colliniato... | per dare forma alle idee che rendono famoso lo stile italiano nel mondo. La continua |
| grandespir... | . Anche lui emana qualcosa: come e che cosa io non lo so, ma è così. lo l'ho vissuto. Uno sciamano deve |

Figura 5. Concordanze per l'articolo *lo* prese dal corpus ItWaC

La Figura 5 riporta i risultati delle concordanze per l'articolo determinativo *lo*. Ho scelto questo esempio perché all'inizio dello studio della lingua italiana gli apprendenti fanno molti errori con gli articoli determinativi e in particolare con l'uso dell'articolo maschile singolare *lo*.

In base alle concordanze che ho ricevuto lo studente sarà in grado di notare che l'articolo determinativo *lo* viene utilizzato con la *s* impura (come per es. *lo studio*, *lo sguardo*, *lo stato*, *lo sviluppo*, ecc.) come pure con la semiconsonante *i* e *j* come nell'esempio: ([...]Roberto Mattei pratica lo Yoga oltre a 35 anni [...]). L'insegnante potrebbe chiedere agli studenti di provare a formulare la regola dell'uso dell'articolo *lo* in gruppi.

| | | | |
|---------------|---|--------------|---|
| drzap.it | sexso e' bello, piacevole e fa bene alla salute. | Spiega | l'opuscolo Informativo distribuito a |
| anonaguid... | esposizione simbolica dedicata a Guelfo, come | spiegano | gli organizzatori, allestita in due sedi poco |
| settimocel... | molto semplice: nella guida di seguito, ti sarà | spiegato | passo per passo e con figure illustrate, come |
| adjora.it ... | piuttosto spinosa. </p><p> Argomento difficile da | spiegare | e (forse per questo) sottovalutato, o |
| linux-maga... | stesso. </p><p> I più letti I più commentati Dal Forum </p><p> Ti | spieghiamo | come fanno i pirati informatici a entrare nelle |
| linux-maga... | ? Fai tutto con la voce! Linux Magazine ti | spiega | come fare 498282 visite </p><p> Nonostante le |
| linux-maga... | illeggibili pendrive e hard disk USB </p><p> Ecco | spiegato | come confinare l'esecuzione e l'uso di |
| aviscalabr... | . "Tre 'A' costituiscono la mia ispirazione - ha | spiegato | Biagio Cutri'- e sono l'Arte, l'Avis e gli |
| repubblica... | , direttore generale di Vodafone Italia, | spiega | gli obiettivi e i tempi per rendere gli utenti |
| repubblica... | agli sviluppi della tecnologia". Prima ha | spiegato | che è cambiata la rete e anche l'atteggiamento |
| anticocaff... | informativa si pone come obiettivo quello di | spiegare | in dettaglio quali sono i tipi di cookie che |
| danielesal... | su http://www.linuxfromscratch.org , che | spiega | quali sono le modalità per la realizzazione |
| parmaest.i... | trasformatori per caricabatteria - | spiega | Pallini indicando vari macchinari sparsi a |
| spietati.i... | /pergamena lasciata dai Costruttori che forse | spiega | come sia possibile lasciare Ember e tornare in |
| piccolipri... | d'arte ci siano questioni che non si possono | spiegare | , misteriose, "magiche". </p><p> Note dell'autore Per |
| petroltecn... | consumo posizionate in tutta la casa. </p><p> Come | spiega | ancora Antonio, le funzionalità della sua casa |
| rodiola.it... | e tendono ad aumentare con l'età: questo | spiegherebbe | come mai la IPB colpisca soprattutto 'vittime' |
| emilydicki... | la compiutezza, Franklin il dato cronologico) | spiegano | la diversità di numerazione e datazione. Nell' |
| teatrodeis... | Riccardo Rossi in Consigli d'amore che ci " | spiega | " i trucchi per vivere al meglio tutte le fasi di |
| chiesacatt... | avverrà nella nuova chiesa di S. Carlo - | spiega | il parroco don Giancarlo Mignardi - e siamo |

Figura 6. Concordanze per il verbo *spiegare* prese dal corpus ItWaC

La Figura 6 riporta i risultati delle concordanze per il verbo *spiegare*, che è un verbo regolare con una caratteristica particolare e spesso difficile agli apprendenti principianti di italiano. Pur avendo una coniugazione regolare, nel coniugare i verbi che finiscono in *-care* e *-gare* bisogna tener conto delle regole di scrittura per i suoni [k] e [g]. Proprio dalle concordanze scelte lo studente potrà notare questa particolarità e dagli esempi arrivare alla regola, come contenuto didattico nuovo o come ripetizione. Inoltre, facendo l'interrogazione del corpus per lemmi si riceve il paradigma completo del verbo esaminato e gli apprendenti analizzando gli esempi possono individuare le varie desinenze dei verbi per differenti forme verbali.

| | | | |
|----------------|--|--------------|--|
| icsedeglia... | i tre anni entro il 31.12.2016; </p><p> devono essere | iscritti | i bambini che compiono i sei anni entro il |
| icsedeglia... | i sei anni entro il 31.12.2016 ; </p><p> possono essere | iscritti | i bambini che compiono i sei anni di età entro il |
| icsedeglia... | per le iscrizioni </p><p> per i genitori che intendono | iscrivere | i figli alle scuole primarie di Flaibano e |
| icsedeglia... | . </p><p> Invito rivolto ai genitori dei bambini che si | iscriveranno | per la prima volta alla scuola dell'infanzia . </p> |
| icsedeglia... | incontreranno i genitori degli alunni | iscritti | alla classe prima delle scuole primarie di |
| agenziagel... |] </p><p> Stiamo preparando le nuove offerte Speciali, | iscriviti | alla Newsletter per riceverle in anteprima. </p> |
| hotelbella... | . </p><p> Se vuoi ricevere tutte le news e le offerte | iscriviti | alla nostra newsletter. *I campi indicati con |
| 100asa.it ... | a .. </p><p> diritto di recesso RMA con sostituzione </p><p> | iscriviti | alla nostra Newsletter </p><p> ed ottieni subito un |
| astegianoe... | attività a partire dal 1° settembre 2015. </p><p> | iscriviti | alla nostra newsletter: </p><p> Aeroporto Valerio |
| etul.it | 100% sicuro. Non c'è bisogno di essere | iscritti | a paypal. </p><p> Spedizioni a prezzo fisso 9.90 € per |
| acquisition... | ti offre la tua città e quali attività promuove? | iscriviti | ad Acquisti On Line, il sito dove con un semplice |
| acquisition... | . Andando su www.acquisitionline.it potrai | iscriverti | alla newsletter della tua città, così ogni |
| eltra.it | e produzione di sensori di posizione dal 1985 </p><p> | iscriviti | alla nostra Newsletter per ricevere |
| palazzoLeo... | reale su offerte e promozioni dell'hotel può | iscriversi | alla newsletter ufficiale del ROSS Hotels |
| hotelcarac... | insieme ai vostri cuccioloni pelosi! </p><p> | iscrivetevi | alla nostra newsletter per essere sempre |
| centrostud... | sulle risposte al tuo commento </p><p> E' possibile | iscriversi | alla newsletter per ricevere un riepilogo |
| hse24.it | che cerchi Usa il nuovo strumento di ricerca! </p><p> | iscriviti | alla Newsletter e riceverai in Anteprima tutte |
| squash.it ... | , e le offerte dello Store di Squash.it... | iscriviti | alla nostra Newsletter! </p><p> SQUASH.it: Il punto di |
| latuabelle... | . Ricco di vitamina B6. Per saperne di più </p><p> | iscriviti | alla nostra newsletter </p><p> Radiofrequenza contro |

Figura 7. Concordanze per il verbo *iscriversi* prese dal corpus ItWaC

La Figura 7 riporta i risultati delle concordanze per il verbo *iscriversi*. Il verbo *iscriversi* è un verbo riflessivo, si tratta di verbi, come lo spiegano Dardano e Trifone (1997: 297) le cui “voci verbali sono precedute dalle particelle pronominali mi, ti, si, ci, vi.” Proprio questa particolarità diventa evidente allo studente negli esempi riportati. In base alle concordanze che ho ricevuto lo studente vedrà che le particelle si possono unire insieme al verbo (come per esempio *iscriviti alla classe*, *iscriviti ad Acquisti On Line*, ecc.) L'altra cosa lo studente noterà è che il verbo *iscriversi* di solito viene utilizzato con la preposizione *a*.

7.1.1 Esempio di creazione di un esercizio in base alle concordanze

Gli insegnanti con l'uso dei corpora possono creare vari tipi di esercizi. Mostriamo qui l'esempio di un esercizio a scelta multipla usando le concordanze. I docenti hanno la possibilità di scrivere da soli in base alla loro esperienza gli usi non corretti dell'unità linguistica in questione o possono trovare gli esempi nei corpora degli apprendenti.

Riportiamo in seguito l'esempio dell'esercizio nel quale gli studenti devono scegliere la risposta esatta aiutandosi con le concordanze.

- a) *Anche se Anna è innamorata con Marco, a lei piace un altro ragazzo.*
- b) *Anche se Anna è innamorata di Marco, a lei piace un altro ragazzo.*
- c) *Anche se Anna è innamorata per Marco, a lei piace un altro ragazzo.*
- d) *Anche se Anna è innamorata in Marco, a Lei piace un altro ragazzo.*

Per dare la risposta giusta lo studente dovrà osservare gli esempi presentati nella Figura 8.

| | | |
|--|------------|--|
| Lei , segretamente | innamorata | di lui , si lascia sedurre senza opporre resistenza . |
| Nel frattempo Ayumi , da sempre | innamorata | di Mayama , scopre che questo si è messo a |
| sia nata una nuova passione poiché afferma che Rita è | innamorata | solo di Mento e non di Steve Dayton . |
| relazione con christopher finisce perché Lorelai capisce che lei è | innamorata | solamente di Luke , l' ha sempre amato e lo |
| al cast , nel ruolo di Carmen , la nuova | innamorata | di Spencer . |
| si per la quarta volta , ma da sempre follemente | innamorata | di Nick ; i gemelli Jeremy , " adolescente " |
| Edie , ancora | innamorata | di Carlos , riesce a sottrarre a Gabrielle i documenti |
| Ponzu ha rivelato che è stata | innamorata | di lui , ma Cross era talmente ossessionato da Satan |
| Inoltre ci è stato rivelato che era | innamorata | di Cross ed sembra apparentemente unir- si a lui , |
| lla figlia del reggente Carlotta Aglae di Borbone-Orléans , follemente | innamorata | di lui . |
| trio , vengono aiutati da Jeanette , una stramba ragazza | innamorata | di Charlie . |
| posto di Derek e scopre che anche lei è segretamente | innamorata | di lui . |
| nata e di cui è senza saper- lo da sempre | innamorata | , e con lui e con Cookie partecipa a molte |
| la peggio , o evitando la professoressa Xavier che è | innamorata | di lui . |
| Nella terza stagione , invece , è | innamorata | follemente di Ned e perciò fa di tutto affinché Suzie |

Figura 8. Concordanze per *innamorata* nel corpus PAISÀ. ¹¹

Dopo aver osservato gli esempi presentati e analizzato i contesti in cui compare il participio passato, l'apprendente sarà in grado di trarre da solo una conclusione in base agli esempi presentati.

Questo era solo un esempio di come si possono creare esercizi con i corpora. Lo stesso tipo di esercizio può venir creato anche usando i risultati delle collocazioni e dei Word Sketch.

¹¹ http://www.corpusitaliano.it/en/access/simple_interface.php (consultato il 11/6/2018)

7.1.2 Esempi di collocazioni

Come ho già specificato nel capitolo 6.2 nella linguistica dei corpora la collocazione viene intesa come la cooccorrenza statisticamente significativa di due segmenti. In seguito presenterò il possibile uso e l'eventuale utilità delle collocazioni nell'apprendimento e insegnamento della lingua italiana.

| | count | count | I-score | MI | logDice |
|--------------------|--------|-----------|---------|-------|---------|
| P N grazie | 12,369 | 1,989,489 | 95.500 | 2.823 | 5.826 |
| P N unico | 10,707 | 1,043,156 | 94.617 | 3.546 | 5.823 |
| P N gioia | 9,413 | 299,093 | 94.312 | 5.162 | 5.822 |
| P N attraverso | 12,677 | 2,213,514 | 95.320 | 2.704 | 5.817 |
| P N svolta | 9,411 | 316,671 | 94.142 | 5.080 | 5.817 |
| P N l | 11,975 | 1,818,860 | 94.827 | 2.905 | 5.814 |
| P N vostro | 10,010 | 681,702 | 94.063 | 4.062 | 5.813 |
| P N personale | 11,023 | 1,276,473 | 94.309 | 3.297 | 5.812 |
| P N spettacolo | 9,692 | 512,508 | 93.874 | 4.427 | 5.809 |
| P N bambini | 11,148 | 1,379,502 | 94.105 | 3.201 | 5.805 |
| P N stessi | 10,201 | 839,909 | 93.694 | 3.789 | 5.801 |
| P N molte | 10,128 | 824,452 | 93.440 | 3.805 | 5.795 |
| P N evento | 10,160 | 860,794 | 93.294 | 3.747 | 5.790 |
| P N caccia | 9,049 | 204,852 | 93.234 | 5.651 | 5.790 |
| P N Come | 11,655 | 1,758,119 | 93.651 | 2.915 | 5.788 |
| P N fosse | 10,614 | 1,149,435 | 93.222 | 3.393 | 5.786 |
| P N gruppo | 11,254 | 1,533,106 | 93.388 | 3.062 | 5.785 |
| P N energia | 9,876 | 714,742 | 93.059 | 3.975 | 5.785 |
| P N natura | 10,393 | 1,022,428 | 93.135 | 3.532 | 5.785 |
| P N Nel | 12,750 | 2,451,335 | 93.843 | 2.565 | 5.779 |
| P N pubblico | 11,119 | 1,485,763 | 93.067 | 3.090 | 5.778 |
| P N coloro | 9,698 | 640,275 | 92.766 | 4.107 | 5.777 |
| P N suggerimenti | 8,771 | 110,799 | 92.614 | 6.493 | 5.771 |
| P N annuncio | 8,825 | 148,478 | 92.552 | 6.080 | 5.769 |
| P N indicazione | 8,854 | 168,363 | 92.523 | 5.903 | 5.768 |
| P N propri | 9,951 | 850,609 | 92.263 | 3.735 | 5.763 |
| P N Una | 12,232 | 2,235,169 | 92.843 | 2.639 | 5.761 |
| P N puoi | 9,413 | 533,302 | 92.191 | 4.328 | 5.761 |
| P N sicuramente | 9,612 | 656,738 | 92.155 | 4.058 | 5.760 |

Figura 9. Collocazioni del verbo *dare* prese dal corpus ItWaC

La Figura 9 riporta i risultati delle collocazioni per il verbo *dare*. Analizzando le collocazioni ricevute l'apprendente avrà la possibilità di arricchire il proprio vocabolario grazie alle locuzioni come dare gioia, dare l'annuncio, dare suggerimenti, dare indicazioni, dare una svolta e altre.

| | | | | | |
|-------------------|-------|-----------|--------|--------|-------|
| P N vestiti | 120 | 102,174 | 10.821 | 6.364 | 4.403 |
| P N cena | 240 | 288,055 | 15.226 | 5.868 | 4.403 |
| P N invitati | 126 | 112,250 | 11.082 | 6.298 | 4.397 |
| P N live | 135 | 126,481 | 11.463 | 6.226 | 4.396 |
| P N techno | 59 | 8,228 | 7.665 | 8.974 | 4.395 |
| P N Carnevale | 95 | 64,696 | 9.652 | 6.686 | 4.391 |
| P N estate | 382 | 514,847 | 19.169 | 5.701 | 4.386 |
| P N amiche | 107 | 84,093 | 10.228 | 6.479 | 4.385 |
| P N comincia | 185 | 207,188 | 13.384 | 5.968 | 4.381 |
| P N carnevale | 73 | 32,437 | 8.489 | 7.302 | 4.365 |
| P N falò | 60 | 12,181 | 7.723 | 8.432 | 4.359 |
| P N vivo | 270 | 347,778 | 16.129 | 5.766 | 4.358 |
| P N country | 67 | 24,354 | 8.142 | 7.592 | 4.345 |
| P N sirtaki | 52 | 279 | 7.210 | 13.674 | 4.344 |
| P N Milly | 54 | 3,912 | 7.340 | 9.919 | 4.337 |
| P N iniziato | 259 | 340,809 | 15.791 | 5.736 | 4.321 |
| P N sorridere | 79 | 46,317 | 8.813 | 6.902 | 4.316 |
| P N ragazzi | 583 | 881,277 | 23.624 | 5.535 | 4.307 |
| P N faceva | 198 | 245,989 | 13.821 | 5.818 | 4.298 |
| P N canteranno | 51 | 1,544 | 7.138 | 11.177 | 4.294 |
| P N assieme | 222 | 288,898 | 14.623 | 5.752 | 4.287 |
| P N fisarmonica | 59 | 15,461 | 7.652 | 8.064 | 4.286 |
| P N inizia | 262 | 357,541 | 15.871 | 5.683 | 4.282 |
| P N successi | 112 | 105,265 | 10.441 | 6.221 | 4.280 |
| P N felici | 119 | 117,174 | 10.755 | 6.154 | 4.279 |
| P N correre | 132 | 139,273 | 11.316 | 6.054 | 4.278 |
| P N concerto | 222 | 295,832 | 14.616 | 5.717 | 4.260 |
| P N pioggia | 155 | 181,828 | 12.241 | 5.901 | 4.258 |
| P N vederti | 56 | 12,440 | 7.459 | 8.302 | 4.255 |
| P N 21.30 | 70 | 37,535 | 8.302 | 7.031 | 4.242 |
| P N veder | 70 | 37,615 | 8.302 | 7.028 | 4.241 |
| P N ridono | 53 | 8,234 | 7.263 | 8.818 | 4.240 |
| P N sudare | 54 | 10,239 | 7.328 | 8.530 | 4.236 |
| P N sulle | 1,316 | 2,204,416 | 35.410 | 5.387 | 4.236 |
| P N tamburi | 56 | 13,883 | 7.456 | 8.144 | 4.234 |

Figura 10. Collocazioni del verbo *ballare* prese dal corpus ItWaC

La Figura 10 riporta i risultati delle collocazioni che riguardano il verbo *ballare*. Si può vedere che il verbo *ballare* viene associato agli elementi come per es. i *vestiti* di ballo, *amiche*, cioè a tutto ciò che ha a che fare con il divertimento, le feste ecc. Osservando le collocazioni lo studente potrà assumere e indovinare in quali contesti il verbo *ballare* normalmente appare.

Precedentemente abbiamo visto i risultati delle collocazioni per due verbi, adesso passiamo all'esempio delle collocazioni di un nome e precisamente di *carta* (Figura 11).

| | <u>count</u> | <u>count</u> | <u>t-score</u> | <u>mi</u> | <u>logrice</u> |
|----------------------|--------------|--------------|----------------|-----------|----------------|
| P N credito | 83,340 | 315,599 | 288.490 | 10.520 | 10.961 |
| P N forno | 27,152 | 235,428 | 164.521 | 9.325 | 9.430 |
| P N identità | 23,522 | 368,920 | 152.936 | 8.470 | 9.081 |
| P N stampata | 16,025 | 39,493 | 126.533 | 11.140 | 8.907 |
| P N foglio | 15,606 | 122,103 | 124.748 | 9.473 | 8.763 |
| P N carta | 20,266 | 630,709 | 141.562 | 7.481 | 8.622 |
| P N sulla | 56,647 | 5,178,052 | 234.095 | 5.927 | 8.218 |
| P N pagamento | 12,601 | 517,923 | 111.424 | 7.080 | 8.036 |
| P N regola | 10,394 | 253,342 | 101.504 | 7.834 | 8.024 |
| P N assorbente | 8,478 | 16,793 | 92.043 | 11.455 | 8.018 |
| P N carte | 10,492 | 281,530 | 101.936 | 7.695 | 8.007 |
| P N su | 75,517 | 8,668,005 | 269.133 | 5.598 | 7.991 |
| P N igienica | 7,879 | 12,295 | 88.738 | 11.799 | 7.919 |
| P N cartone | 7,687 | 64,688 | 87.542 | 9.368 | 7.814 |
| P N teglia | 7,328 | 45,541 | 85.508 | 9.805 | 7.770 |
| P N gioco | 14,475 | 1,124,391 | 118.632 | 6.162 | 7.766 |
| P N fogli | 7,414 | 73,451 | 85.951 | 9.133 | 7.751 |
| P N mazzo | 6,815 | 35,523 | 82.475 | 10.059 | 7.678 |
| P N prepagata | 6,490 | 8,415 | 80.541 | 12.066 | 7.644 |
| P N penna | 6,698 | 71,615 | 81.684 | 9.023 | 7.606 |
| P N parati | 6,125 | 7,733 | 78.244 | 12.105 | 7.562 |
| P N giocare | 7,953 | 335,333 | 88.503 | 7.043 | 7.551 |
| P N plastica | 6,825 | 182,915 | 82.215 | 7.697 | 7.498 |
| P N diritti | 9,805 | 724,327 | 97.705 | 6.234 | 7.497 |
| P N bianca | 6,749 | 213,438 | 81.685 | 7.458 | 7.446 |
| P N vini | 6,755 | 238,288 | 81.667 | 7.300 | 7.419 |
| P N pezzo | 6,857 | 265,895 | 82.229 | 7.164 | 7.411 |
| P N Servizi | 6,579 | 240,117 | 80.578 | 7.251 | 7.379 |
| P N intestata | 5,395 | 12,781 | 73.419 | 11.197 | 7.372 |
| P N bancomat | 5,329 | 18,601 | 72.954 | 10.638 | 7.346 |
| P N circolazione | 6,070 | 186,681 | 77.479 | 7.498 | 7.324 |
| P N mano | 10,430 | 1,115,825 | 100.163 | 5.700 | 7.299 |
| P N Carta | 5,619 | 129,186 | 74.650 | 7.918 | 7.281 |
| P N costituzionale | 5,774 | 162,897 | 75.601 | 7.623 | 7.280 |
| P N tavola | 6,121 | 243,399 | 77.677 | 7.128 | 7.272 |
| P N debito | 5,661 | 185,472 | 74.796 | 7.407 | 7.225 |

Figura 11. Collocazioni della parola *carta* prese dal corpus ItWaC

La Figura 11 riporta i risultati delle collocazioni della parola *carta*. Come prima cosa l'apprendente vedrà quali sono i campi e i contesti nei quali la parola *carta* viene generalmente utilizzata. Grazie ai risultati gli studenti possono collegare la parola *carta* con il sistema bancario come per es. *credito*, *pagamento*, *bancomat*, *prepagata*, *debito*, ecc. Poi dai candidati riportati nell'elenco *igienica*, *assorbente*, *teglia* e *forno* traspare il suo significato in casa. *Giochi*, *tavola*, *mano*, *mazzo* possono venir ricollegati alle carte da gioco e gioco di carte. Il significato di documento viene individuato dalla collocazione *identità*. Poi *foglio*, *stampata* e *penna* rivelano il significato di foglio. Gli esempi esposti servono al docente non solo per spiegare i vari contesti d'uso, ma anche per spiegare i vari significati che una parola può assumere in vari contesti.

7.1.3 Esempi di Word Sketch

Come già spiegato nel capitolo 6.4 la funzione di Word Sketch permette di osservare le relazioni grammaticali in cui una parola si manifesta e può aiutare sia il docente che lo studente a trovare ottimi esempi d'uso della parola presa in analisi. Si veda nella Figura 12 l'esempio dei risultati del Word Sketch per la parola *oro*.

oro (noun) Alternative PoS: [orare verb](#) (freq: 28,172)
Italian Web 2016 (itTenTen16) freq = [400,280](#) (68.25 per million)

| verbs with "oro" as object | verbs with "oro" as subject | modifiers of "oro" | noun modifiers of "oro" | nouns modified by noun "oro" |
|---|--|--|---|---|
| 9.16 | 0.86 | 12.99 | 0.02 | 0.61 |
| placcare + 1,167 9.91 placcato oro | saiwa 40 8.55 gli oro saiwa | color + 4,380 10.08 color oro | metallo 8 8.96 visto 9 3.73 numero 22 1.94 l' oro numero 200 | foglia + 178 10.87 la foglia oro |
| colare + 1,162 9.79 per oro colato | luccicare 37 8.17 oro luccicante | giallo + 5,285 9.90 in oro giallo | | quotazione + 132 10.60 la quotazione oro |
| comprare + 2,987 8.59 compro oro | scintillare 36 7.62 oro scintillante , | colato + 1,520 9.84 per oro colato | | colore + 552 10.46 il colore oro |
| rodere + 685 8.46 in oro rosa | placcare 20 7.56 , color oro placca logo , | zecchino + 1,426 9.77 in oro zecchino | | colorazione 55 9.20 nelle colorazioni oro , |
| conquistare + 2,218 8.19 | cio 19 7.27 e ' tutto oro cio ' che luccica | olimpico + 2,341 9.51 oro olimpico | | distretto 57 8.93 distretto oro faringeo mal |
| vincere + 2,475 7.61 | rosè 17 7.27 oro rosè , | rosa + 1,799 8.96 in oro rosa | | targa 41 8.91 la targa oro |
| valere + 1,098 7.03 che vale oro | argentare 17 7.25 oro argento | massiccio + 888 7.98 d' oro massiccio | | sterlina 35 8.86 . La sterlina oro |
| glitter + 147 7.01 glitter oro | 18k 16 7.24 oro 18k con | bianco + 4,825 7.96 in oro bianco | | riflesso 33 8.63 dai riflessi oro |
| estrarre + 384 6.82 estrarre l' oro | simboleggiare 23 7.11 . L' oro simboleggia | puro + 1,936 7.77 oro puro | | ombretto 32 8.54 un ombretto oro |
| tempestare + 114 6.63 d' oro tempestata di | asi 14 6.98 targa oro asi | nero + 2,907 7.59 dell' oro nero | | tonalità 38 8.47 tonalità oro e |
| aggiudicare + 260 6.32 | impresiosire 10 6.49 | blu + 906 7.48 oro blu " | | prezzo + 104 8.42 il prezzo oro |
| vendere + 774 6.27 | paralimpico 11 6.49 l' oro paralimpico | colloidale + 166 6.65 oro colloidale | | fiamma 26 8.34 delle fiamme oro |
| rubare + 249 6.27 | zecchino 9 6.41 | pallido + 207 6.60 oro pallido | | finitura 29 8.33 nelle finiture oro |
| satinare + 106 6.27 oro satinato | splendere 25 6.29 | | | |
| fondere + 138 6.18 oro fuso | oro splendente | | | |
| filare 88 6.09 | | | | |

Figura 12. Word Sketch per la parola *oro* presa dal corpus ItTenTen16 – prima parte

La Figura 12 illustra una parte dei risultati generati dal Word Sketch. Il Word Sketch genera una lista di verbi frequenti dove appare la parola *oro* come oggetto (*placcare, colare, comprare, rubare, vincere, ecc.*), poi verbi con i quali *oro* fa da sogetto (*luccicare, scintillare, placcare, splendere, simboleggiare, ecc.*). Segue l'elenco dei più frequenti modificatori della parola *oro* (*color, giallo, colato, ecc.*), *oro* in posizione di modificatore e nomi che modificano *oro* (*foglia, quotazione, colore, ecc.*).

Si veda nella Figura 13 la seconda parte dei risultati generati da Word Sketch per la parola *oro*.

| prepositional phrases with nouns | | | "oro" is a ... | | | "oro" and/or ... | | | ... is a "oro" | | |
|----------------------------------|-------|------|---------------------------------------|-----|-------|-------------------------|--------|-------|--------------------------------------|----|-------|
| "oro" al | 8,622 | 2.15 | metallo | 83 | 11.61 | argento + | 12,725 | 12.63 | monnezza | 11 | 10.30 |
| "oro" del | 6,495 | 1.62 | . L' oro è un metallo | | | oro e argento | | | monnezza è oro | | |
| "oro" della | 5,590 | 1.40 | investimento | 18 | 9.67 | diamante + | 1,222 | 9.70 | preferito | 11 | 10.03 |
| "oro" per | 5,428 | 1.36 | oro è un investimento | | | oro e diamanti | | | mio preferito è oro e fiori di | | |
| "oro" di | 4,946 | 1.24 | oro | 13 | 9.42 | bronzo + | 1,049 | 9.26 | consiglio | 12 | 9.81 |
| "oro" in | 4,473 | 1.12 | oro è oro | | | platino + | 805 | 9.15 | consigli sono oro | | |
| "oro" a | 3,625 | 0.91 | conduttore | 10 | 9.40 | oro , platino | | | colore | 80 | 9.61 |
| "oro" nel | 2,941 | 0.73 | moneta | 11 | 8.92 | oro + | 842 | 8.81 | colore è giallo oro | | |
| "oro" alla | 2,750 | 0.69 | colore | 34 | 8.73 | pietra + | 1,019 | 8.76 | oro | 13 | 9.42 |
| "oro" nella | 2,188 | 0.55 | oro è il colore | | | oro e pietre preziose | | | oro è oro | | |
| "oro" con | 2,140 | 0.53 | materia | 10 | 7.97 | rame + | 625 | 8.60 | capo | 9 | 9.15 |
| "oro" da | 1,242 | 0.31 | bene | 35 | 7.72 | oro , rame | | | silenzio | 21 | 8.99 |
| "oro" dell' | 928 | 0.23 | l' oro è un bene | | | incenso + | 496 | 8.44 | il silenzio è oro | | |
| "oro" all' | 644 | 0.16 | simbolo | 30 | 7.67 | oro , incenso e mirra . | | | parola | 14 | 6.94 |
| "oro" sul | 544 | 0.14 | oro è simbolo | | | gioiello + | 626 | 8.39 | parole sono oro | | |
| "oro" su | 542 | 0.14 | materiale | 20 | 7.65 | oro e gioielli | | | tempo | 30 | 6.93 |
| "oro" dal | 338 | 0.08 | oro è un materiale | | | petrolio + | 412 | 7.98 | il tempo è oro | | |
| "oro" sulla | 240 | 0.06 | elemento | 16 | 5.75 | metallo + | 438 | 7.78 | acqua | 18 | 6.85 |
| "oro" dalla | 211 | 0.05 | . L' oro è un elemento | | | gemma + | 270 | 7.56 | L' acqua è l' oro | | |
| "oro" ad | 201 | 0.05 | protagonista | 8 | 4.73 | oro e gemme | | | Non | 8 | 5.70 |
| "oro" da di | 119 | 0.03 | pronominal possessors of "oro" | | | medaglia + | 293 | 7.48 | protagonista | 8 | 4.97 |
| "oro" nell' | 111 | 0.03 | 0.59 | | | prezioso + | 223 | 7.30 | adjective predicates of "oro" | | |
| "oro" sotto | 50 | 0.01 | tuo + | 530 | 2.64 | oro e preziosi | | | 0.03 | | |
| "oro" tra | 44 | 0.01 | il tuo oro | | | avorio + | 224 | 7.27 | puro | 8 | 7.12 |
| "oro" dopo | 42 | 0.01 | vostro + | 230 | 2.25 | oro e avorio | | | presente | 25 | 4.17 |
| | | | il vostro oro | | | smalto + | 203 | 7.07 | l' oro è presente | | |
| | | | proprio + | 259 | 1.21 | oro e smalto | | | bello | 13 | 3.64 |
| | | | vendere il proprio oro | | | moneta + | 225 | 7.06 | importante | 8 | 3.23 |
| | | | loro + | 240 | 0.77 | record + | 185 | 6.92 | disponibile | 8 | 2.11 |
| | | | il loro oro | | | oro e record | | | | | |
| | | | suo + | 674 | 0.61 | rosso + | 155 | 6.63 | | | |
| | | | il suo oro | | | oro e il rosso | | | | | |

Figura 13. Word Sketch per la parola *oro* presa dal corpus ItTenTen16– seconda parte

La Figura 13 riporta la seconda parte dei risultati di Word Sketch per la parola *oro*. Come si può vedere, oltre a quello nominato precedentemente nella Figura 12, Word Sketch genera elenchi di aggettivi possessivi e qualificativi che accompagnano la parola *oro* (*tuo, vostro, ecc. puro, presente, ecc.*), di preposizioni che lo seguono (*al, del, della, ecc.*) e di parole che stanno con la parola *oro* in relazione congiuntiva “e” (*argento, diamante, bronzo, ecc.*). Infatti, si ricevono informazioni interessanti riguardanti la parola esaminata, in questo caso l’*oro*, che non sono informazioni di tipo puramente linguistico (*metallo, investimento, moneta, ecc.*).

Vediamo ora un altro esempio interessante quello della parola *bagaglio* (Figura 14).

bagaglio *(noun)*
Italian Web 2016 (itTenTen16) freq = 92,013 (15.68 per million)

| verbs with "bagaglio" as object | verbs with "bagaglio" as subject | modifiers of "bagaglio" | noun modifiers of "bagaglio" | nouns modified by noun "bagaglio" |
|---------------------------------|----------------------------------|---------------------------------|------------------------------|-----------------------------------|
| 25.69 | 0.93 | 29.81 | 0.04 | 2.38 |
| arricchire + <u>1,987</u> 9.17 | portabagaglio <u>13</u> 8.94 | vano + <u>1,469</u> 9.42 | farmaco <u>8</u> 8.38 | deposito + <u>749</u> 12.13 |
| imbarcare + <u>316</u> 8.41 | Vano bagagli portabagaglio | vano bagagli | incontro <u>7</u> 6.52 | un deposito bagagli |
| sistemare + <u>521</u> 7.82 | pesare <u>17</u> 6.51 | smarrito + <u>284</u> 8.02 | ragazzo <u>9</u> 6.02 | ritiro + <u>207</u> 11.25 |
| disfare + <u>175</u> 7.73 | dirigere <u>24</u> 6.40 | bagagli smarriti | all <u>7</u> 2.80 | al ritiro bagagli |
| disfare i bagagli | i bagagli ci dirigiamo verso | esperienziale + <u>212</u> 7.61 | | franchigia + <u>158</u> 11.09 |
| caricare + <u>574</u> 7.72 | viaggiare <u>26</u> 5.88 | bagaglio esperienziale | | la franchigia bagaglio |
| ritirare + <u>522</u> 7.53 | Il bagaglio viaggia | culturale + <u>5,232</u> 7.51 | | trasporto + <u>225</u> 9.53 |
| trasportare + <u>467</u> 7.53 | arricchire <u>9</u> 4.74 | bagaglio culturale | | il trasporto bagagli |
| depositare + <u>468</u> 7.42 | partire <u>20</u> 4.19 | ingombrante + <u>214</u> 7.40 | | smistamento <u>44</u> 9.31 |
| depositare i bagagli | i bagagli partiamo | bagagli ingombranti | | allo smistamento bagagli |
| accrescere + <u>438</u> 7.30 | rispettare <u>7</u> 4.10 | lessicale + <u>155</u> 7.14 | | cofano <u>38</u> 9.04 |
| accrescere il proprio bagaglio | superare <u>24</u> 4.06 | bagaglio lessicale | | del cofano bagagli |
| posare + <u>209</u> 7.16 | bagaglio supera | conoscitivo + <u>262</u> 7.07 | | nastro <u>44</u> 9.01 |
| ampliare + <u>445</u> 6.96 | arrivare <u>75</u> 3.74 | bagaglio conoscitivo | | al nastro bagagli |
| ampliare il proprio bagaglio | bagagli arrivano | scarico + <u>136</u> 6.49 | | consegna <u>47</u> 8.99 |
| smarrire + <u>124</u> 6.82 | costare <u>15</u> 3.65 | carico e scarico bagagli | | la consegna bagagli |
| bagagli smarriti | bagaglio costa | voluminoso <u>79</u> 6.27 | | eccedenza <u>23</u> 8.41 |
| preparare + <u>1,065</u> 6.49 | uscire <u>19</u> 3.46 | bagagli voluminosi | | per l' eccedenza bagaglio |
| preparare i bagagli | i bagagli usciamo | pesante + <u>417</u> 5.95 | | controllo + <u>107</u> 8.38 |
| Franchigia <u>61</u> 6.39 | rientrare <u>9</u> 3.30 | tecnico + <u>1,640</u> 5.77 | | al controllo bagagli |
| - Franchigia bagaglio | pagare <u>8</u> 2.77 | bagaglio tecnico | | scontrino <u>22</u> 8.34 |
| acquisire + <u>478</u> 5.97 | aspettare <u>10</u> 2.71 | medico + <u>460</u> 5.63 | | delle scontrino bagagli |
| stivare <u>43</u> 5.86 | seguire <u>14</u> 2.58 | assicurazione medico bagaglio | | riconsegna <u>21</u> 8.28 |
| stivare bagagli | Per <u>7</u> 2.48 | esperenziale <u>42</u> 5.62 | | la riconsegna bagagli |
| spedire + <u>116</u> 5.83 | contenere <u>22</u> 2.20 | bagaglio esperenziale | | etichetta <u>22</u> 8.13 |
| recuperare + <u>377</u> 5.82 | restare <u>19</u> 2.18 | incluso + <u>105</u> 5.60 | | etichetta bagaglio |
| riporre <u>81</u> 5.74 | i bagagli restano | e bagaglio inclusi | | compartimento <u>12</u> 7.43 |
| riporre i bagagli | aumentare <u>7</u> 2.07 | appresso <u>59</u> 5.60 | | compartimento bagagli |

Figura 14. Word Sketch per la parola *bagaglio* presi dal corpus ItTenTen16 – prima parte

La Figura 14 riporta una parte dei risultati generati da Word Sketch dal corpus ItTenTen16. Si possono vedere i verbi più frequenti dove appare la parola *bagaglio* come oggetto (*arricchire*, *imbarcare*, *sistemare*, ecc.), poi i verbi con i quali fa da soggetto (*portabagaglio*, *pesare*, *dirigere*, ecc.). Segue l'elenco dei più frequenti modificatori (*vano*, *smarrito*, *esperenziale*, ecc.), *bagaglio* posizione di modificatore (*farmaco*, *incontro*, *ragazzo*, ecc.) e nomi che lo modificano (*deposito*, *ritiro*, *trasporto*, ecc.). Dai risultati si riceve una terminologia molto specializzata che potrà essere molto utile per coloro che devono trattare un argomento legato ai bagagli. Accanto a questo significato concreto nel campo del viaggio i risultati rivelano un uso frequente della parola *bagaglio* nel suo senso più astratto di *bagaglio culturale*, *esperenziale*, *lessicale*, ecc. Si veda nella Figura 16 la seconda parte dei risultati generati da Word Sketch per la parola *bagaglio*.

| prepositional phrases with nouns | | | pronominal possessors of "bagaglio" | | | "bagaglio" and/or ... | | | adjective predicates of "bagaglio" | | |
|----------------------------------|---------------|-------|-------------------------------------|--------------|------|------------------------------|------------|-------------|------------------------------------|-----------|-------------|
| "bagaglio" di | <u>12,731</u> | 13.84 | | 15.75 | | | | 5.36 | | | 0.08 |
| "bagaglio" a | <u>5,676</u> | 6.17 | proprio + | <u>3,601</u> | 5.00 | annullamento + | <u>119</u> | 9.04 | leggero | <u>9</u> | 7.33 |
| "bagaglio" in | <u>1,701</u> | 1.85 | il proprio bagaglio | | | , bagaglio e annullamento | | | pronto | <u>39</u> | 5.04 |
| "bagaglio" da | <u>1,310</u> | 1.42 | vostro + | <u>952</u> | 4.28 | check-in | <u>42</u> | 8.01 | bagagli sono pronti , | | |
| "bagaglio" per | <u>629</u> | 0.68 | il vostro bagaglio | | | bagagli , check-in | | | gratuito | <u>11</u> | 3.25 |
| "bagaglio" della | <u>529</u> | 0.57 | loro + | <u>1,837</u> | 3.70 | passaggero | <u>88</u> | 8.00 | disponibile | <u>8</u> | 2.12 |
| "bagaglio" al | <u>506</u> | 0.55 | il loro bagaglio | | | bagagli e passeggeri | | | | | |
| "bagaglio" del | <u>461</u> | 0.50 | mio + | <u>1,848</u> | 3.60 | valigia | <u>58</u> | 7.99 | | | |
| "bagaglio" da/di | <u>316</u> | 0.34 | il mio bagaglio | | | merce + | <u>142</u> | 7.87 | | | |
| "bagaglio" all' | <u>230</u> | 0.25 | nostro + | <u>2,081</u> | 3.42 | bagagli e merci | | | | | |
| "bagaglio" sul | <u>196</u> | 0.21 | il nostro bagaglio | | | pacco | <u>42</u> | 7.50 | | | |
| "bagaglio" nel | <u>188</u> | 0.20 | tuo + | <u>843</u> | 3.31 | bagagli e pacchi | | | | | |
| "bagaglio" nella | <u>177</u> | 0.19 | il tuo bagaglio | | | vivero | <u>31</u> | 7.30 | | | |
| "bagaglio" alla | <u>154</u> | 0.17 | suo + | <u>3,329</u> | 2.91 | bagagli , viveri | | | | | |
| "bagaglio" dell' | <u>109</u> | 0.12 | il suo bagaglio | | | bagaglio | <u>52</u> | 7.18 | | | |
| "bagaglio" con | <u>94</u> | 0.10 | | | | check | <u>19</u> | 6.85 | | | |
| "bagaglio" sulla | <u>83</u> | 0.09 | | | | reception | <u>27</u> | 6.82 | | | |
| "bagaglio" presso | <u>46</u> | 0.05 | | | | deposito bagagli , reception | | | | | |
| "bagaglio" durante | <u>35</u> | 0.04 | | | | colle collo | <u>21</u> | 6.72 | | | |
| "bagaglio" dalla | <u>32</u> | 0.03 | | | | bagagli o colli | | | | | |
| "bagaglio" su | <u>31</u> | 0.03 | | | | zaino | <u>27</u> | 6.51 | | | |
| "bagaglio" dal | <u>28</u> | 0.03 | | | | partenza | <u>65</u> | 6.45 | | | |
| "bagaglio" nell' | <u>17</u> | 0.02 | | | | dei bagagli e partenza | | | | | |
| "bagaglio" fuori | <u>15</u> | 0.02 | | | | ritardo | <u>48</u> | 6.42 | | | |
| "bagaglio" dall' | <u>13</u> | 0.01 | | | | bicicletta | <u>45</u> | 6.42 | | | |

Figura 15. Word Sketch per la parola *bagaglio* presi dal corpus ItTenTen16 – seconda parte

Vediamo ora, nella Figura 15 la seconda parte dei risultati Word Sketch per la parola *bagaglio*. In particolare i risultati delle parole che stanno con la parola *bagaglio* in relazione congiuntiva “e” (*annullamento, check-in, passeggero, ecc.*) confermano quello che abbiamo detto precedentemente e cioè che il Word Sketch offre all’utente, studente o docente che sia, un’ottima preparazione terminologica di base.

Segue un altro esempio di Word Sketch presentato nella Figura 16, adesso con un verbo.

dare (verb)
Italian Web 2016 (itTenTen16) freq = 5,147,831 (877.79 per million)

| objects of "dare" | 67.84 | subjects of "dare" | 6.03 | modifiers of "dare" | 8.30 | prepositional phrases with nouns | prepositions after "dare" | 8.16 | |
|-------------------|---------------|--------------------|------------|---------------------|-------------|----------------------------------|---------------------------|-----------------------|--------------|
| vita + | 195,777 10.52 | Dio + | 6,567 8.89 | tanto + | 8,747 8.18 | "dare" al | 108,787 2.11 | per + | 69,356 7.62 |
| risposta + | 97,821 9.70 | Signore + | 2,947 8.05 | dato tanto | | "dare" alla | 96,136 1.87 | da + | 77,583 7.60 |
| possibilità + | 82,415 9.42 | governo + | 2,406 7.48 | più + | 56,704 8.07 | "dare" da | 63,379 1.23 | a + | 100,702 7.05 |
| contributo + | 59,951 9.04 | consiglio + | 1,457 7.18 | così + | 18,438 8.00 | "dare" a | 57,719 1.12 | dare a | |
| via + | 55,474 8.90 | Gesù + | 1,645 7.10 | , dando così | | "dare" in | 54,085 1.05 | ad + | 18,623 6.98 |
| dato il via | | volta + | 2,230 7.08 | anche + | 56,855 7.95 | "dare" di | 42,509 0.83 | dare ad | |
| mano + | 51,982 8.79 | anno + | 2,589 7.07 | via + | 5,981 7.91 | "dare" del | 41,685 0.81 | in + | 65,900 6.54 |
| dare una mano | | fatto + | 1,311 6.96 | dato via | | "dare" dal | 39,168 0.76 | dato in | |
| luogo + | 50,834 8.78 | risultato + | 1,412 6.95 | subito + | 6,751 7.76 | "dare" dalla | 33,733 0.66 | di + | 55,461 6.35 |
| occhiata + | 46,535 8.74 | tempo + | 1,687 6.94 | sempre + | 17,579 7.55 | "dare" della | 29,649 0.58 | data di | |
| dare un'occhiata | | risposta + | 1,250 6.93 | finalmente + | 3,374 7.55 | "dare" per | 22,838 0.44 | senza + | 2,864 6.19 |
| spazio + | 48,632 8.63 | sondaggio + | 1,053 6.77 | dare finalmente | | "dare" all' | 20,645 0.40 | da di + | 1,676 5.70 |
| dare spazio | | sondaggi danno | | molto + | 16,326 7.42 | "dare" dall' | 10,448 0.20 | , data d'uscita | |
| nome + | 40,320 8.37 | natura + | 1,060 6.66 | su + | 20,277 7.34 | "dare" ad | 9,383 0.18 | fin + | 811 5.49 |
| notizia + | 36,715 8.33 | amministrazione + | 1,145 6.63 | solo + | 16,856 7.34 | "dare" nel | 6,448 0.13 | dato fin | |
| consiglio + | 35,916 8.33 | vita + | 1,135 6.60 | mai + | 6,368 7.31 | "dare" con | 6,020 0.12 | entro + | 902 5.48 |
| voce + | 36,092 8.30 | giudice + | 1,000 6.57 | ancora + | 12,333 7.29 | "dare" dell' | 4,916 0.10 | , data entro la quale | |
| dare voce | | regione + | 1,067 6.55 | troppo + | 4,812 7.25 | "dare" sul | 4,902 0.10 | su + | 4,108 5.42 |
| origine + | 34,216 8.28 | giorno + | 1,391 6.52 | dare troppo | | "dare" sulla | 3,526 0.07 | durante + | 1,046 5.38 |
| senso + | 33,994 8.22 | ragazzo + | 1,128 6.52 | dentro + | 2,886 7.19 | "dare" nella | 3,459 0.07 | dato durante | |
| dare un senso | | ragazzi hanno dato | | quindi + | 5,236 7.18 | "dare" nell' | 2,947 0.06 | de + | 651 5.29 |
| ragione + | 32,369 8.18 | uomo + | 1,328 6.49 | poi + | 5,351 7.16 | "dare" su | 1,856 0.04 | dopo + | 979 5.25 |
| inizio + | 30,237 8.11 | cosa + | 1,428 6.45 | addossare + | 2,142 7.15 | "dare" senza | 1,598 0.03 | dato dopo | |
| fastidio + | 29,230 8.08 | Comune + | 936 6.41 | dare addosso | | "dare" da di | 1,271 0.02 | fino + | 1,360 5.23 |
| informazione + | 32,072 7.99 | medico + | 889 6.36 | pure + | 2,902 6.99 | "dare" durante | 557 0.01 | dati fino a | |
| forma + | 29,709 7.97 | missione + | 925 6.35 | almeno + | 2,890 6.93 | | | attraverso + | 968 5.15 |
| idea + | 28,923 7.92 | Italia + | 1,026 6.32 | però + | 3,553 6.90 | | | con + | 9,865 4.91 |
| risultato + | 29,423 7.89 | Italia ha dato | | meno + | 3,401 6.90 | | | secondo + | 608 4.89 |
| indicazione + | 26,203 7.86 | | | sicuramente + | 2,433 6.66 | | | via + | 439 4.83 |
| prova + | 25,043 7.79 | | | proprio + | 5,050 6.66 | | | dato via | |

Figura 16. Word Sketch per il verbo *dare* presi dal corpus ItTenTen16

La Figura 16 riporta i risultati di Word Sketch per il verbo *dare* dal corpus ItTenTen16. Si ricorda che in questo capitolo sono stati presentati vari esempi di analisi del verbo *dare* e precisamente tre tipi di concordanze ricavati da tre corpora diversi, itWaC, Perugia Corpus (PEC) e il corpus PAISÀ. Il primo è un web corpus, il secondo si propone di essere un corpus di riferimento dell'italiano scritto e parlato, mentre l'ultimo è stato realizzato per apprendenti di italiano. Oltre ad aver sottoposto lo stesso verbo all'analisi delle concordanze di tre corpora diversi, per lo stesso verbo sono state generate le collocazioni e il Word Sketch. Il verbo *dare* non è stato preso perché fosse particolarmente interessante, ma per fare vedere come una stessa parola può venir analizzata in vari modi e in vari corpora.

Sarà il docente che dovrà saper valutare qual'è l'analisi adeguata per raggiungere lo scopo delle attività pianificate o del materiale didattico che sta preparando. Per fare la scelta giusta, come abbiamo cercato di spiegare in questa tesi, bisogna capire le principali tecniche di analisi della linguistica dei corpora e quali tipi di risultati sono in grado di generare come pure quale tipo di

corpus bisogna usare.

8. Conclusione

Questa tesi aveva lo scopo di analizzare e spiegare le potenzialità e gli eventuali limiti dell'uso dei corpora nell'insegnamento e nell'apprendimento delle lingue. In particolare, l'intento di questa tesi era di illustrare il modo in cui i corpora italiani di lingua scritta possono essere inseriti nell'insegnamento e nell'apprendimento dell'italiano.

Dopo una parte introduttiva teorica riguardante i concetti di base della linguistica dei corpora, segue una rassegna dei corpora italiani disponibili e adeguati allo studio della lingua come pure la spiegazione delle tecniche di analisi dei corpora utili all'insegnamento.

L'ultima parte, è la parte più importante in quanto è dedicata alla verifica personale da parte di un apprendente di italiano dell'utilità concreta dell'impiego diretto dei corpora nell'apprendimento della lingua italiana.

L'ultima parte della tesi è perciò dedicata all'analisi concreta del possibile utilizzo dei corpora da parte degli apprendenti e presenta esempi concreti dei risultati delle concordanze, collocazioni e Word Sketch. Ho scelto di presentare esempi di parole che mi sembravano interessanti ed eventualmente problematiche mostrando come l'apprendimento della lingua italiana possa diventare facile e divertente grazie ai corpora. Sono i corpora ad offrire all'apprendente tantissimi esempi d'uso di una parola in contesti comunicativi reali e in testi autentici in base ai quali l'apprendente può da solo o con l'aiuto dell'insegnante trovare soluzioni e trarre generalizzazioni e conclusioni. Sono stati presentati diversi esempi di analisi delle concordanze, delle collocazioni e dei Word Sketch usando vari corpora per illustrare la loro differenza nelle possibilità di ricerca, la loro adeguatezza e semplicità d'uso come ad esempio il corpus Valico, Paisà, Perugia corpus, ItWaC e ItTenTen16. Gli ultimi due si sono dimostrati semplicissimi nell'uso.

Un ruolo fondamentale nell'impiego diretto dei corpora nell'insegnamento o nell'apprendimento delle lingue, come abbiamo nominati più volte in questa tesi, ha l'insegnante che deve saper illustrare il funzionamento e spiegare l'utilità dei corpora agli apprendenti e stimolarli al loro uso, all'inizio mediato dall'insegnante e molto presto all'uso autonomo.

Speriamo che questa tesi possa servire da spunto ai futuri insegnanti di lingua italiana che sono pronti a sfruttare le potenzialità offerte dai corpora e dalle tecniche della linguistica dei corpora per arricchire e migliorare il proprio metodo d'insegnamento ma anche per i futuri apprendenti interessati all'uso dei corpora nell'apprendimento autonomo della lingua italiana.

9. Bibliografia

- Barbera, M. (2013). *Linguistica dei corpora e linguistica dei corpora italiana*. Milano: Qu.A.S.A.R., s.r.l.
- Baroni, M., Bernardini S., Ferraresi A. e Zanchetta E. (2008). *The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora*. Kluwer Academic Publishers. Preso da:http://wacky.sslmit.unibo.it/lib/exe/fetch.php?media=papers:wacky_2008.pdf (consultato il 09/6/2018).
- Baroni, M., Bernardini, S., Comastri, F., Piccioni, L., Volpi, A., Aston, G. e Mazzoleni, M. (2004). Introducing the la Repubblica corpus. A large, annotated, TEI(XML) compliant corpus of newspaper Italian. In Lino, M. T. et al. (a cura di) *Proceedings of the 4th international conference on language resources and evaluation LREC (1771-1774)*. Paris: ELRA European Language Resources Association.
- Baroni, M. (2010). Corpora di italiano. In Simone, R. (a cura di.), *Enciclopedia dell'italiano*, vol 1, (300-303). Roma: Istituto della Enciclopedia Italiana. Disponibile sul:[http://www.treccani.it/enciclopedia/corpora-di-italiano_\(Enciclopedia-dell'Italiano\)/](http://www.treccani.it/enciclopedia/corpora-di-italiano_(Enciclopedia-dell'Italiano)/) (consultato il 11/6/2018).
- Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In J. M. Sinclair (a cura di.), *How to use corpora in language teaching* (15–36). Amsterdam: John Benjamin
- Boulton, A. (2008). Looking (for) empirical evidence of data driven learning at lower Levels. In Lewandowska-Tomaszczyk, B. (a cura di) *Corpus Linguistics, Computer Tools, and Applications: State of the Art* (591-598). Frankfurt: Peter Lang.
- Bolasco, S. (1999). *Analisi multidimensionale dei dati. Metodi, strategie e criteri d'interpretazione*. Roma: Carocci edizioni.
- Bortolini, U., Tagliavini, C. e Zampolli, A. (1971). *Lessico di frequenza della lingua italiana contemporanea*. Milano: Garzanti.
- Ceccotti, M. L. e Sassi, M. (2001). *L'utilizzo didattico di corpora: proposte metodologiche*. Bari. Preso da: <http://www.ilc.cnr.it/CEG/progetto/pdf/Linguistica%20dei%20corpora.pdf> (consultato il 09/6/2018)
- Corino, E. (2014). Didattica delle lingue corpus-based. *ELLE*, vol. 3, num. 2, 231-258.
- Corino, E. e Marellò, C. (2009). Didattica con i corpora di italiano per stranieri. *Italiano linguadue*, vol.1, num.1., 279-285.

- Cresti, E. e Panunzi A. (2013). *Introduzione ai corpora dell'italiano*. Bologna: Il Mulino.
- Dardano, M. e Trifone, P. (1997). *La nuova grammatica della grammatica della lingua italiana*. Milano: Zanichelli editore.
- Efrati, V. (2008). *Breve guida all'uso di alcuni software per l'analisi testuale ed il trattamento automatico del linguaggio (TAL)*. Preso da: http://host.uniroma3.it/laboratori/triple/Risorse_Resources_files/Guida_ai_software_TAL_18_092008.pdf (consultato il 09/6/2018).
- Giuliano, L. e La Rocca, G. (2008). *L'analisi automatica e semi- automatica dei dati testuali: software e istruzioni per l'uso*. Milano: Universitarie di lettere economia diritto edizioni.
- Gries, S. T. (2010). Corpus linguistics and theoretical linguistics. A love hate relationship? Not necessarily.... *International Journal of Corpus Linguistics*, 15 (3), 327-343.
- Granger, S., Hung, J. e Petch-Tyson, S. (a cura di), (2002). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: Benjamins.
- Guidetti, M. G., Lenzi, G e Storch, S. (2012). Potenzialità e limiti dell'uso dei corpora linguistici per la didattica dell'italiano LS. *Supplemento alla rivista EL.LE*. Preso da: <https://www.italy.it/potenzialita-e-limiti-dell-80%99uso-dei-corpora-linguistici-la-didattica-dell-80%99italiano-ls> (consultato il 11/6/2018).
- Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., e Suchomel, V. (2013). The TenTencorpus family. In *7th International Corpus Linguistics Conference CL* (125-127).
- Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. In T. Johns e P. King (a cura di), *Classroom concordancing: ELR Journal*, 4 (1-16). Birmingham: Centre for English Language Studies, University of Birmingham.
- Lalli Paćelat, I. (2016). *I corpora nell'apprendimento e nell'insegnamento delle lingue*. In *L'apprendimento delle lingue straniere*. Bilingualism Matters@Rijeka. Preso da: <http://www.bilingualism-matters-rijeka.ffri.hr/it/bilinguismo/l-apprendimento-delle-lingue-straniere.html?tmpl=component&print=1&page=> (consultato il 09/6/ 2018).
- Lalli Paćelat, I. (2017). *Primjer primjene korpusa u nastavi jezika*. Bilingualism Matters@Rijeka Preso da: <http://bmri.uniri.hr/primjer-primjene-korpusa-u-nastavi-jezika/> (consultato il 09/6/2018).
- Laudanna, A. et al. (1995). Un corpus dell'italiano scritto contemporaneo dalla parte del ricevente. In S. Bolasco et al. (a cura di) *JADT 1995*. III. Giornate internazionali di analisi statistica dei dati testuali (103-109). Roma: Consiglio nazionale delle ricerche.

- Leech, G. (1992). Corpora and theories of linguistic performance. In Svartvik, J. (a cura di), *Directions in Corpus Linguistics: Proceedings of the Nobel Symposium 82, Stockholm, 4–8 August 1991*, (105–122). Berlin: Mouton de Gruyter.
- Leech, G. (1997). *Teaching and language corpora: A convergence*. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (a cura di) *Teaching and language corpora* (18-23). London: Longman.
- Leoni, F. A. (2003). *Il corpus CLIPS*. Preso da: http://www.clips.unina.it/it/documenti/presentazione_clips.pdf (consultato il 09/6/2018).
- Lüdeling, A. e Kytö, M. (a cura di) (2008). *Corpus linguistics. An international handbook*, Berlin: Mouton de Gruyter.
- McEnery, T. e Xiao, R. (2010). What corpora can offer in language teaching and learning. In E. Hinkel (a cura di), *Handbook of research in second language teaching and learning*. London/New York: Routledge.
- McEnery, T. e Hardie, A. (2012). *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.
- McEnery, T., Xiao, R. Z. e Tono, Y. (2006). *Corpus-based Language Studies: An Advanced Resource Book*. London: Routledge.
- Mukherjee, J. (2006). *Corpus linguistics and language pedagogy: The state of the art – and beyond*. Giessen: Justus Liebig University.
- Noseda, V. (2015/2016). *Corpora paralleli e linguistica contrastiva: ampliamento e applicazioni del corpus italiano-russo nel nacional' NYJ korpus russkoga jazyka*. Dottorato di ricerca in scienze linguistiche e letterarie. Milano: Università Cattolica del Sacro Cuore.
- Ljubešić, N. e Erjavec, T. (2011). hrWaC and slWaC: Compiling Web Corpora for Croatian and Slovene. In Habernal, I. e Matousek, V. (a cura di), *Text, Speech and Dialogue / Proceedings of the 14th International Conference (TSD 2011)*, LNCS 6836, (395–402). Heidelberg/Berlin: Springer Verlag.
- Riccardi, A. (2003). *Dalla traduzione all'interpretazione. Studi d'interpretazione simultanea*. Milano: LED.
- Rossini Favretti, R. (2000). Progettazione e costruzione di un corpus di italiano scritto: *CORIS/CODIS*. In R. Rossini Favretti (a cura di), *Linguistica e informatica. Multimedialità, corpora e percorsi di apprendimento*, (39-56). Roma: Bulzoni.
- Rossini Favretti, R., Tamburini, F. e De Santis, C. (2002). CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model. In Wilson, A., Rayson, P., & McEnery, T. (a cura di), *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*, (27–

38). Munich: Lincom–Europa.

Römer, U. (2008). Corpora and language teaching. In A. Lüdeling A. e M. Kytö (a cura di). *Corpus Linguistics. An International Handbook* . vol. 1, (112-130). Berlin: Mouton de Gruyter.

Spina, S. (2001). *Fare i conti con le parole: Introduzione alla linguistica dei corpora*. Perugia: Guerra edizioni.

Suchomel, V., e Pomikálek, J. (2012). Efficient web crawling for large text corpora. In *Proceedings of the seventh Web as Corpus Workshop (WAC7)* (39-43).

Tamburini, F. (2000). Annotazione grammaticale e lemmatizzazione di corpora in italiano. In Rossini Favretti, R. (a cura di), *Linguistica e informatica: multimedialità, corpora e percorsi di apprendimento* (57–73). Roma: Bulzoni.

Teubert, W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics*, 10 (1), 1–13.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Viganò, P. B. (2011). Il corpora e il loro sfruttamento in didattica. *Italiano linguadue vol.3, n.2.*, 115-128. Preso da: <https://riviste.unimi.it/index.php/promoitals/article/view/1918/2171> (consultato il 09/6/2018).

10. Riassunto

La linguistica dei corpora ha portato notevoli cambiamenti in varie discipline linguistiche, includendo la glottodidattica. Tuttavia, anche se lo stretto rapporto tra la linguistica dei corpora e la glottodidattica esiste da più di vent'anni, la linguistica dei corpora non è sufficientemente presente nell'apprendimento e nell'insegnamento delle lingue straniere, ad eccezione della lingua inglese. Lo scopo di questo lavoro è di presentare i vantaggi dell'utilizzo della linguistica dei corpora nell'apprendimento e nell'insegnamento delle lingue straniere. Dopo un'introduzione teorica necessaria per chiarire i concetti di base della linguistica dei corpora, segue una descrizione dell'impiego diretto e indiretto dei corpora nell'insegnamento. Siccome l'attenzione è rivolta in particolare all'impiego diretto dei corpora nell'insegnamento e nell'apprendimento della lingua italiana, la tesi offre una panoramica dei corpora della lingua italiana scritta che possono venir utilizzati nell'apprendimento della lingua. In fine vengono presentate e analizzate proposte concrete ed esempi d'impiego diretto dei corpora nell'insegnamento e nell'apprendimento della lingua italiana.

11. Sažetak

Korpusna lingvistika dovela je do velikih pomaka u raznim lingvističkim disciplinama uključujući i glotodidaktiku. Međutim, iako je do spajanja korpusne lingvistike i glotodidaktike došlo već prije dvadesetak godina, korpusna lingvistika nije dovoljno zastupljena u učenju i poučavanju stranih jezika, osim engleskoga jezika. Cilj je ovoga rada prikazati prednosti primjene korpusne lingvistike u učenju i poučavanju stranih jezika. Nakon teorijskoga uvoda u kojem su pojašnjeni osnovni pojmovi iz korpusne lingvistike, opisana je izravna i neizravna primjena korpusa u nastavi. Poseban je naglasak na izravnoj primjeni korpusa u nastavi talijanskog jezika, pa su stoga u ovome radu predstavljeni talijanski korpusi koji mogu biti korisni u učenju samoga jezika.

Prikazani su također i konkretni prijedlozi i primjeri izravne primjene korpusa u nastavi talijanskoga jezika.

12. Summary

Corpus linguistics has led to major shifts in various linguistics disciplines including glotodidactics. However, although the merge of corpus linguistics and glotodidactics has come already twenty years ago, corpus linguistics is not sufficiently represented in learning and teaching foreign languages, except of English.

The aim of this work is to show the advantages of applying the corpus linguistics in learning and teaching foreign languages. After the theoretical introduction where the basic terms of corpus linguistics are clarified, direct and indirect use of the corpora are described. Special emphasis is placed on the direct application of the corpora in teaching Italian language. Concrete proposals and examples of direct use of corpora in the teaching Italian language are shown and corpora of Italian language that can be useful in learning Italian are presented.