

L'USO DEI CORPORA DI ITALIANO PARLATO DA  
NON NATIVI NELLA STESURA DEI QUESTIONARI  
DI ACCETTABILITÀ LINGUISTICA:  
IL CASO DEL LIPS

1. INTRODUZIONE

La redazione dei questionari linguistici, che ad alcuni potrebbe sembrare una questione triviale ed ancillare alla ricerca linguistica, se fatta senza adottare un metodo scientifico, potrebbe avere un impatto rilevante sulle risposte elicitate. Tuttavia, non di rado accade che la redazione di un questionario linguistico sia prettamente funzionale alle specifiche domande della ricerca e che non tenga conto né delle proprietà del lessico come frequenza d'uso e disponibilità, né della competenza linguistica dei destinatari del questionario. Riportiamo qui a titolo esemplificativo una frase tratta da un questionario di accettabilità della selezione dell'ausiliare in italiano: «un nembo è/ha rimasto sopra il campo» (Laws 2010: 15). Sebbene il questionario fosse destinato ai parlanti nativi, la scelta di un termine tecnico-specialistico o letterario come 'nembo' al posto dei sostantivi meno marcati (e più trasparenti) come 'nube' o 'nuvola', potrebbe confondere e perfino distrarre i rispondenti dallo svolgimento del test, compromettendo potenzialmente la qualità dei dati raccolti.

Oltre a questi casi di scelte infelici di unità lessicali, che per lo più non hanno un carattere sistematico e dunque si potrebbero facilmente evitare, si riscontrano nume-

rosi casi di scelte lessicali dettate dalle esigenze della ricerca. Nel già citato articolo di Laws si fa un tentativo di individuare ulteriori sottoclassi semantiche all'interno del gruppo dei verbi inaccusativi in italiano, il quale comporta l'inclusione di nuovi verbi nei questionari di accettabilità linguistica, oltre a quelli tradizionalmente trattati nella letteratura scientifica. Così, accanto alle parole appartenenti al lessico fondamentale come 'arrivare' o 'nascere', compaiono i verbi come 'germogliare' e 'avvizire' (Laws 2010: 14), caratterizzati da una bassa frequenza d'uso e una bassa disponibilità. Se nel caso dei rispondenti italofoeni questi verbi non dovrebbero creare problemi, la situazione si complica nel momento in cui bisogna interpellare i rispondenti non nativi, alcuni dei quali potrebbero semplicemente ignorare i termini meno frequenti.

Dando per scontato che le scelte lessicali 'infelici' si dovrebbero evitare *a priori*, in questa sede ci occuperemo del secondo tipo di scelte lessicali, vale a dire di quelle motivate dalle esigenze della ricerca. In questo contributo considereremo dunque le potenzialità dell'uso del Nuovo vocabolario di base della lingua italiana (De Mauro 2016) insieme al Lessico dell'Italiano Parlato da Stranieri (Gallina 2015) nella redazione di un questionario di accettabilità della selezione dell'ausiliare in italiano L2 dei parlanti russofoeni, con lo scopo di ridurre al minimo l'interferenza di fattori lessicali con la qualità dei dati elicitati.

Sebbene nell'ambito italiano esista la tradizione ormai consolidata di utilizzare i vocabolari di base come punto di riferimento nella redazione dei materiali per l'insegnamento e per la valutazione della competenza linguistica (Chiari/De Mauro 2012: 23), non ci risulta che questa prassi venga estesa anche alla stesura di questionari linguistici.

Negli ultimi anni gli studi che indagano l'interazione tra la ricerca sui *learner corpora* e *Second Language Acquisition* (Le Bruyn/Paquot 2021) si sono concentrati principalmente su questioni teoriche, approfondendo temi come l'influenza cross-linguistica oppure il ruolo di universali linguistici, tralasciando l'utilizzo pratico dei *corpora* per la redazione di stimoli linguistici. Pur non negando l'importanza teorica di questi argomenti, con il presente contributo ci riproponiamo di focalizzarci sull'aspetto lessicale nella redazione dei questionari linguistici.

## 2. QUESTIONARI DI ACCETTABILITÀ LINGUISTICA

### 2.1 Questionari di accettabilità come strumento della ricerca sperimentale

I questionari di accettabilità linguistica fanno parte dello strumentario di quell'area della linguistica che viene comunemente definita sintassi sperimentale (Sprouse 2007: 123). Il questionario di accettabilità si redige (1) in vista delle specifiche domande di ricerca ed (2) è destinato a un concreto gruppo di rispondenti; questi due elementi costituiscono i punti di riferimento cardine nella redazione di un questionario. Le domande di ricerca sono necessariamente il punto di partenza, e determinano sia la struttura generale del questionario (il numero e il tipo di stimoli linguistici,

l'ordine e il modo della loro presentazione ai rispondenti, la presenza di eventuali distrattori ecc.) sia le caratteristiche che devono avere i suoi destinatari (il loro livello di competenza linguistica, struttura demografica, *background* socioculturale ecc.). Allo stesso tempo, anche le caratteristiche dei rispondenti sopraelencate influenzano il tipo di stimoli linguistici da includere in un questionario: è evidente che non tutti gli stimoli somministrabili a parlanti nativi possano essere somministrati anche a parlanti L2. Possiamo notare che la redazione di un questionario linguistico richiede non solo un'attenta valutazione delle domande di ricerca, ma esige anche un occhio di riguardo per le competenze dei rispondenti.

Di seguito, dopo una breve presentazione delle domande di ricerca sulle quali si basa il questionario analizzato, tratteremo più in dettaglio i criteri che ci hanno guidato nella selezione di unità lessicali da includere nel questionario in modo da adeguarlo alle competenze dei rispondenti non nativi.

## 2.2 Questionario di accettabilità della selezione dell'ausiliare italiano

Il presente contributo nasce all'interno di una ricerca dottorale sull'acquisizione delle marche di intransitività scissa in italiano da parte di apprendenti russofoni e illustra il processo di redazione di un questionario di accettabilità della selezione dell'ausiliare italiano nei tempi composti.

Il questionario analizzato si rifà al modello adottato inizialmente nello studio pionieristico sulla selezione dell'ausiliare di Antonella Sorace (Sorace 1992) e successivamente adottato da Jacqueline Laws (Laws 2010). La somministrazione del nostro questionario era strumentale all'elicitazione dei dati su un eventuale rapporto tra le proprietà semantiche e aspettuali dei verbi intransitivi italiani e la selezione dell'ausiliare operata dagli apprendenti russofoni. Dunque, era necessario che i verbi inseriti nel questionario fossero intransitivi e che rappresentassero ognuna delle sette classi semantiche dell'ASH (*Auxiliary Selection Hierarchy*, Sorace 2000: 863). Queste sette classi semantiche includono i verbi intransitivi che indicano: (1) cambiamento di luogo, (2) cambiamento di stato, (3) continuazione di stato, (4) esistenza di stato, (5) processo non controllato, (6) movimento atelico e (7) processo controllato.<sup>1</sup> Inoltre, le classi semantiche sono organizzate lungo il gradiente ASH, partendo dai verbi inaccusativi che scelgono categoricamente l'ausiliare 'essere' (1) fino ad arrivare ai verbi inergativi che scelgono categoricamente l'ausiliare 'avere' (7), con in mezzo i

---

<sup>1</sup> Tuttavia, non tutte le classi semantiche dell'ASH rappresentano delle unità coese e indivisibili, all'interno di almeno alcune di esse è possibile individuare delle sottoclassi semantiche il cui comportamento morfosintattico potrebbe differire dal comportamento della loro classe di appartenenza. L'esempio più evidente della complessità interna di queste classi semantiche è quello della classe verbale che indica il cambiamento di stato (2), la quale potrebbe essere ulteriormente suddivisa in quattro sottoclassi: (2a) *inherently telic*; (2b) *appearance*; (2c) *internally-caused*; (2d) *directed motion*. Una trattazione più dettagliata di questo argomento si trova in (Laws 2010).

verbi che mostrano diversi gradi di variabilità nella selezione dell'ausiliare (2-6).

Una volta stabiliti i criteri 'qualitativi' che descrivevano le tipologie dei verbi da includere nei questionari, bisognava selezionare unità lessicali concrete che non solo rispettassero i criteri sopra menzionati, ma che allo stesso tempo fossero comprensibili ai rispondenti non nativi.

Il questionario era destinato a due gruppi di rispondenti – un gruppo di controllo, composto da parlanti nativi italiani, e un gruppo di apprendenti russofoni (suddivisi a loro volta in più sottogruppi in base al loro livello di competenza in italiano). Dunque, era necessario che a entrambi i gruppi di rispondenti, a prescindere dalla loro competenza in italiano, fossero somministrati gli stessi stimoli linguistici. Tenendo in considerazione che la competenza in italiano dei rispondenti russofoni variava tra i livelli A2 e C1, era importante che la difficoltà degli stimoli somministrati fosse calibrata sul livello più basso di competenza, in modo da garantire che tutti i rispondenti non nativi fossero in grado di comprendere gli stimoli presentati. In questo modo, nella selezione dei verbi da includere nel questionario, accanto a criteri qualitativi (appartenenza a una delle sette classi semantiche dell'ASH), dovevamo considerare anche criteri quantitativi. Nella definizione di questi ultimi siamo partiti dalla stessa idea che sta alla base del Nuovo vocabolario di base della lingua italiana di Tullio De Mauro (NVdB), che integra le parole di alta disponibilità (AD) con le parole ad alta frequenza d'uso, le quali includono il lessico fondamentale (FO) e di alto uso (AU). Questo approccio si basa da un lato su calcoli statistici che permettono di individuare le unità lessicali che hanno un'elevata frequenza d'uso, dall'altro lato sui risultati di esperimenti psicolinguistici sulla comprensione del lessico da parte dei parlanti nativi, fornendo un'immagine completa non solo dell'uso, ma anche della rappresentazione mentale del lessico (cfr. Chiari/De Mauro 2012: 23).

Tuttavia il NVdB, fondandosi sugli usi e sulle intuizioni dei parlanti nativi, non può costituire l'unica fonte cui attingere per selezionare le unità lessicali da includere nel questionario destinato anche a parlanti non nativi. Se dovessimo considerare solo il lessico fondamentale del NVdB – che comprende circa 2000 lemmi – scopriremmo che è quasi due volte più ampio della lista lessicale riportata nel Profilo della lingua italiana per il livello A2 del QCER (Quadro comune europeo di riferimento), composta dai soli 1037 lemmi, di cui circa 150 sono verbi.<sup>2</sup> Per questo motivo, al fine di ottenere dati sull'effettiva competenza lessicale degli apprendenti non nativi di italiano abbiamo deciso di consultare i *corpora* di italiano L2 che analizzeremo nella sezione successiva.

---

2 Il Profilo della lingua italiana è consultabile sul sito dell'Università per Stranieri di Perugia al seguente link: [https://www.unistrapg.it/profilo\\_lingua\\_italiana/site/liste\\_lessicali\\_a2.html](https://www.unistrapg.it/profilo_lingua_italiana/site/liste_lessicali_a2.html) (ultimo accesso: 14/03/2024).

### 3. LEARNER CORPORA IN CONTESTO ITALIANO

#### 3.1 Cos'è un corpus?

Sebbene esistano tanti modi per definire un *corpus*, è possibile adottare una definizione lata che ne descrive le caratteristiche principali come una raccolta di testi autentici (incluse le trascrizioni del parlato) che possono essere informaticamente trattati e sono selezionati per essere rappresentativi di una lingua o una varietà linguistica (McEnery *et al.* 2006: 6). Questa definizione si applica anche ai *learner corpora*, ovvero i *corpora* basati su testi prodotti da apprendenti non nativi; un *corpus* di italiano L2 dunque dovrebbe essere rappresentativo delle varietà di apprendimento dell'italiano, vale a dire contenere una quantità adeguata delle effettive produzioni orali e/o scritte di apprendenti non nativi di italiano.

Nel lasso di tempo trascorso dalla seconda metà degli anni '80 del Novecento sono stati realizzati diversi *corpora* di italiano L2 che rientrano per la maggior parte nella definizione lata proposta sopra. Tuttavia, questi *corpora* variano sia per la tipologia di testi utilizzati nella loro composizione (orali, scritti o misti), sia per il numero e tipo di apprendenti interpellati (contesto di apprendimento, lingue di partenza ecc.). Stando all'informazione riportata sul sito *Learner corpora around the world* – un ampio *database* digitale che tiene traccia dei *learner corpora* realizzati in tutto il mondo – al 4 aprile del 2023 esistevano 10 *corpora* redatti con i testi prodotti da apprendenti la cui unica lingua target era l'italiano.<sup>3</sup> Di questi 10 *corpora* ne abbiamo dovuto escludere 7, in quanto la L1 degli apprendenti interpellati non includeva il russo. Di seguito passeremo in rassegna i restanti 3 *corpora* di italiano L2, specificando i motivi che ci hanno guidato nella selezione del *corpus* LIPS che abbiamo utilizzato nella redazione del nostro questionario di accettabilità.

#### 3.2 Corpora orali di italiano L2

Partendo dai *corpora* orali di italiano L2, non possiamo non menzionare l'ormai leggendaria *Banca dati di italiano L2 del Progetto di Pavia* (Andorno 2001) – il primo tentativo di creare una raccolta di produzioni orali in italiano L2. Il *corpus* di Pavia, che comprende le registrazioni del parlato di 20 apprendenti dell'italiano in contesto spontaneo,<sup>4</sup> è alla base di una serie di ricerche sull'acquisizione dell'italiano, alcune delle quali sono riportate in un volume curato da Giacalone Ramat (Giacalone Ra-

3 La lista completa dei *learner corpora* è consultabile sul sito: <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html> (ultimo accesso: 14/03/2024).

4 Anche se i 14 apprendenti su 20 hanno ricevuto anche qualche input scolastico formale, la sua scarsità insieme al «comportamento linguistico non normativo» degli apprendenti spinge gli autori a considerarli alla stessa stregua degli apprendenti esposti al solo input spontaneo (cfr. Giacalone Ramat 2003: 31).

mat 2003). Allo stesso tempo va notato che questo *corpus* comprende i dati raccolti da parlanti nativi di un numero ridotto di lingue madre (11), spesso tipologicamente molto distanti dall'italiano (malese, morè, arabo, tigrino, cinese)<sup>5</sup> e anche la quantità dei dati raccolti da ciascun apprendente varia dalle 2 alle 10 ore di registrazione (cfr. Giacalone Ramat 2003: 28-33). Le particolarità della composizione del *corpus* di Pavia sono dovute al fatto che nella sua creazione sono spesso confluite ricerche che partivano da interessi diversi. Queste considerazioni, insieme alla quantità ridotta dei dati raccolti rispetto ad altri *corpora*, ci costringe ad escludere il *corpus* di Pavia dalla nostra trattazione.

Il LIPS (Lessico dell'Italiano Parlato da Stranieri) è un *corpus* orale di circa 670 mila *token*. Il *corpus* LIPS è costituito da 1426 testi tratti dalle prove orali di esame CILS (Certificazione di Italiano come Lingua Straniera), sostenute da circa 700 candidati con lingue madre diverse (incluso il russo), di tutti i livelli di competenza (dall'A1 al C2). Le prove orali sono state svolte sia in Italia sia all'estero, nel periodo che spazia tra il 1993 e il 2006 (cfr. Gallina 2015: 55).

### 2.3 *Corpora scritti di italiano L2*

Passando ai *corpora* scritti consideriamo qui i casi di due *corpora* di italiano L2: il CAIL2 (Corpus di Apprendenti di Italiano L2) e il recentissimo *corpus* CELI (Certificati di Lingua Italiana).

Il CAIL2 (Bratankova 2015) è un *corpus* con un'ampiezza di circa 273 mila *token*, composto da 400 testi scritti, prodotti da altrettanti apprendenti di italiano L2 di livelli che variano dall'intermedio all'avanzato con 31 lingue madre diverse, incluso il russo. I dati del *corpus* CAIL2 – pubblicato nel 2015 – sono stati raccolti presso tre università: l'Università per Stranieri di Perugia, l'Università Carolina di Praga e l'Università della Boemia meridionale di České Budějovice.

Il CELI (Spina *et al.* 2022) è un *corpus* scritto che ha un'ampiezza totale di circa 600 mila *token*. Questo *corpus* è composto da più di 3 mila prove scritte di candidati adulti non nativi agli esami di certificazione di lingua italiana CELI di livello intermedio e avanzato (B1, B2, C1 e C2). Le prove d'esame sono state sostenute sia in Italia sia all'estero, il che ha permesso di creare un *corpus* bilanciato per nazionalità e lingua madre dei candidati – le nazionalità sono 104 e ad ognuno dei quattro livelli di certificazione è presente un certo numero di parlanti russofoni.

Come si evince da una breve descrizione delle principali caratteristiche dei 3 *corpora* di italiano L2 riportati sopra, i *corpora* CAIL2 e CELI, che si basano sulle produzioni scritte prodotte in parte da apprendenti di lingua madre russa, hanno fissato la soglia minima di competenza in italiano al livello intermedio – B1/CELI 2 rispet-

---

5 È degno di nota che nessuno degli apprendenti consultati parlava il russo e dunque le loro varietà di apprendimento potrebbero divergere sostanzialmente dalle varietà di apprendimento di parlanti russofoni.

tivamente – il che rende impossibile qualsiasi tentativo di indagare la competenza lessicale di apprendenti di livello A2. In questo modo il LIPS rappresenta l'unico *corpus* di italiano L2 che soddisfa tutti i criteri della nostra ricerca, comprendendo i testi prodotti anche da apprendenti russofoni con livelli di competenza in italiano tra l'A2 e il C1. Nell'ultima sezione di questo contributo presentiamo le principali risorse del *corpus* LIPS che abbiamo utilizzato nella stesura del nostro questionario di accettabilità.

#### 4. IL CORPUS LIPS COME RISORSA LESSICALE PER LA REDAZIONE DEI QUESTIONARI DI ACCETTABILITÀ

Nel primo capitolo di questo contributo abbiamo delineato i criteri qualitativi (intransitività, appartenenza a una delle sette classi dell'ASH) e quantitativi (frequenza d'uso) che ci hanno guidato nella selezione di unità lessicali da includere nel nostro questionario. A questo proposito le liste di frequenza e uso estratte dal *corpus* LIPS (e suddivise per livello di competenza) costituiscono uno strumento duttile che ci permette non solo di sapere quali parole vengono effettivamente utilizzate dai parlanti di un determinato livello di competenza, ma anche con quale frequenza.<sup>6</sup> Consultando dunque la lista di frequenza per il livello A2, troviamo che essa comprende 1307 lemmi, di cui 163 sono verbi. Escludendo da questa lista i verbi transitivi, nonché i verbi di sola forma riflessiva, otteniamo 66 verbi intransitivi che possiamo suddividere in base alla loro appartenenza alle sette classi semantiche del gradiente ASH (cfr. tab. 1)

Dopo aver suddiviso i verbi intransitivi in classi semantiche ci risulta che ognuna di esse è rappresentata da almeno due verbi, il che soddisferebbe i requisiti minimi posti dalle esigenze della nostra ricerca. Tuttavia, per testare l'ipotesi di Laws (Laws 2010) di una più complessa organizzazione interna delle classi semantiche, avremmo bisogno di aggiungere alla nostra analisi altri verbi che non sono presenti nella lista di frequenza del livello A2. Questo è particolarmente vero per due classi semantiche dell'ASH – (2) cambiamento di stato e (6) movimento atelico. Ricordiamo che la classe semantica che indica cambiamento di stato potrebbe essere ulteriormente suddivisa in quattro sottoclassi, delle quali la sottoclasse (2d) *directed motion* è rappresentata da un solo verbo 'salire' caratterizzato da un numero molto basso di occorrenze. Per quanto riguarda la classe semantica che indica movimento atelico, in essa sono del tutto assenti i verbi 'alternanti' che possono descrivere anche movimento telico.

<sup>6</sup> Le modalità di elaborazione delle liste di frequenza e uso sono descritte in (Gallina 2015: 67-68).

## IL CASO DEL LIPS

Cambiamento di luogo	Cambiamento di stato	Continuazione di stato	Esistenza di stato	Processo non controllato	Movimento atelico	Processo controllato
andare (99) <sup>7</sup> venire (24) arrivare (14) partire (5) ritornare (4) tornare (3) entrare (1) uscire (1)	<b>cambiare</b> (6) <b>cominciare</b> (6) <b>iniziare</b> (6) <b>riuscire</b> (5) <b>passare</b> (4) <b>diventare</b> (3) <b>succedere</b> (3) <b>finire</b> (2) <b>morire</b> (2) <b>salire</b> (2) <b>scendere</b> (2) <b>seguire</b> (2) <b>aprire</b> (1) <b>incominciare</b> (1) <b>mancare</b> (1) <b>migliorare</b> (1) <b>nascere</b> (1) <b>ricominciare</b> (1) <b>scadere</b> (1)	<b>continuare</b> (3) <b>rimanere</b> (1)	<b>essere</b> (328) <b>piacere</b> (72) <b>stare</b> (19) <b>abitare</b> (17) <b>vivere</b> (13) <b>sembrare</b> (7) <b>servire</b> (5) <b>interessare</b> (3) <b>convenire</b> (2) <b>dispiacere</b> (2) <b>esistere</b> (2) <b>bastare</b> (1) <b>importare</b> (1)	<b>sentire</b> (8) <b>suonare</b> (7) <b>funzionare</b> (4) <b>dormire</b> (3) <b>assistere</b> (2) <b>dipendere</b> (1) <b>godere</b> (1) <b>ridere</b> (1)	<b>viaggiare</b> (9) <b>camminare</b> (2) <b>girare</b> (1) <b>nuotare</b> (2) <b>passaggiare</b> (1) <b>saltare</b> (1)	<b>parlare</b> (33) <b>studiare</b> (30) <b>lavorare</b> (21) <b>pensare</b> (18) <b>giocare</b> (7) <b>cantare</b> (1) <b>chiacchierare</b> (1) <b>comunicare</b> (1) <b>sperare</b> (1) <b>vestire</b> (1)

**Tab. 1.** Distribuzione di verbi intransitivi tratti dalle produzioni di apprendenti non-nativi di italiano L2 di livello A2<sup>8</sup>

Considerando l'incompletezza del materiale lessicale disponibile per il livello A2 abbiamo dovuto consultare la lista di frequenza del livello successivo – B1, nella quale compare un solo verbo di movimento atelico alternante – ‘correre’ (32 occorrenze). Per trovare un altro verbo di questo tipo abbiamo dovuto consultare la lista di frequenza del livello ancora più alto – B2, dove abbiamo trovato il verbo ‘volare’ (8 occorrenze).

Infine, dopo aver consultato le liste di frequenza suddivise per livelli di competenza, abbiamo composto una lista di 23 verbi intransitivi da includere nel nostro questionario di accettabilità. Questa lista è composta principalmente dai verbi presenti nella lista di frequenza del livello A2 (20) con l'aggiunta di un verbo della lista B1 (‘correre’) e altri due verbi della lista B2 (‘volare’ e ‘cadere’).

<sup>7</sup> I numeri tra parentesi indicano il numero di occorrenze dei verbi nella lista di frequenza LIPS.

<sup>8</sup> Seguendo le convenzioni tipografiche adottate nel NvDB, usiamo neretto tondo per il lessico fondamentale, tondo chiaro per il lessico di alto uso e corsivo chiaro per il lessico di alta disponibilità.

## 5. CONCLUSIONE

Nel presente contributo abbiamo descritto l'utilizzo del *corpus* LIPS nella selezione dei verbi da includere in un questionario di accettabilità linguistica. Abbiamo consultato diverse liste di frequenza per lemmi (dal livello A2 al livello B2) estratte dal *corpus* con l'obiettivo di individuare le unità lessicali che si avvicinino il più possibile alle competenze linguistiche di parlanti non nativi. Il nostro lavoro ha rivelato che le liste di frequenza rappresentano uno strumento utile e perfino indispensabile nella redazione di questionari di accettabilità destinati agli apprendenti di italiano L2, in quanto mettono a nostra disposizione il lessico effettivamente utilizzato da apprendenti di un determinato livello di competenza linguistica. Tuttavia, la relativa scarsità del materiale lessicale disponibile per i livelli di competenza più bassi (A2) ci costringe a cercare un equilibrio tra le esigenze della ricerca e la comprensibilità del questionario, attingendo con cautela al repertorio lessicale dei livelli di competenza più alti (B1 e B2).

## BIBLIOGRAFIA

- Andorno 2001 = Cecilia Andorno, *Banca dati di Italiano L2. Progetto di Pavia*, CD-Rom, Università di Pavia, Dipartimento di Linguistica.
- Bratankova 2015 = Leontyna Bratankova, *Le collocazioni Verbo + Nome in apprendenti di italiano L2*, Tesi di dottorato, Università per Stranieri di Perugia.
- Chiari/De Mauro 2012 = Isabella Chiari / Tullio De Mauro, *The new basic vocabulary of Italian: problems and methods*, in *Rivista di statistica applicata*, Vol. 22, 1, pp. 21-35.
- Corino/Marello 2017 = Elisa Corino / Carla Marello, *Italiano di stranieri. I corpora VALICO e VINCA*, Perugia, Guerra Edizioni.
- De Mauro 2016 = Tullio De Mauro, *Il Nuovo vocabolario di base della lingua italiana*, in *Internazionale*, <https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana> (ultimo accesso: 06/03/2024).
- Gallina 2015 = Francesca Gallina, *Le parole degli stranieri. Il Lessico dell'Italiano Parlato da Stranieri*, Perugia, Guerra Edizioni.
- Giacalone Ramat 2003 = Anna Giacalone Ramat (a cura di), *Verso l'italiano. Percorsi e strategie di acquisizione*, Roma, Carocci.
- Laws 2010 = Jacqueline Laws, *To 'be', and not to 'have': auxiliary selection in unaccusative verbs in Italian*, in Luke James O'Brien / Davide Simone Giannoni (a cura di), *Language Studies Working Papers*, 2, pp. 3-16.
- Le Bruyn/Paquot 2021 = Bert Le Bruyn / Magali Paquot (a cura di), *Learner Corpus Research Meets Second Language Acquisition*, Cambridge, Cambridge University Press.
- McEnery et al. 2006 = Tony McEnery, Richard Xiao, Yukio Tono, *Corpus-based language studies: An advanced resource book*, London-New York, Routledge.
- Sorace 1992 = Antonella Sorace, *Lexical conditions on syntactic knowledge: Auxiliary selection*

- in native and non-native grammars of Italian* [tesi di dottorato non pubblicata]. Edinburgh: University of Edinburgh.
- Sorace 2000 = Antonella Sorace, *Gradients in auxiliary selection with intransitive verbs*, in «Language», Vol. 76, pp. 859–890.
- Spina *et al.* 2022 = Stefania Spina / Irene Fioravanti / Luciana Forti / Valentino Santucci / Angela Scerra / Fabio Zanda, *Il corpus CELI: una nuova risorsa per studiare l'acquisizione dell'italiano*, in «Italiano Lingua Due», Vol. 14(1), pp. 116-138.
- Sprouse 2007 = Jon Sprouse, *Continuous Acceptability, Categorical Grammaticality, and Experimental Syntax*, in *Biolinguistics*, <https://bioling.psychopen.eu/index.php/bioling/article/view/8597> (ultimo accesso: 11/12/2024).