

Quale italiano?

Una proposta per l'insegnamento dell'italiano
a stranieri tra standard e neostandard

Anna Zingaro

Prefazione di Michele Prandi

Studi Interdisciplinari su Traduzione, Lingue e Culture

Bologna
University Press

Studi Interdisciplinari su Traduzione, Lingue e Culture

46

Studi Interdisciplinari su Traduzione, Lingue e Culture

Collana a cura del Dipartimento di Interpretazione e Traduzione (DIT)
dell'Alma Mater Studiorum – Università di Bologna, sede di Forlì.

La Collana, fondata nel 2004, raccoglie le pubblicazioni scientifiche dei suoi afferenti e degli studiosi che operano in ambiti affini a livello nazionale e internazionale.

A partire da una riflessione generale sul tradurre come luogo di incontro e scontro tra lingue e culture, la Collana si propone di diffondere e rendere disponibili, a livello cartaceo e/o su supporto elettronico, i risultati della ricerca in molteplici aree, come la linguistica teorica e applicata, la linguistica dei *corpora*, la terminologia, la traduzione, l'interpretazione, gli studi letterari e di genere, il teatro, gli studi culturali e sull'umorismo.

Le pubblicazioni della Collana sono approvate dal Dipartimento, sentito il motivato parere di almeno due esperti qualificati esterni.

Il/la responsabile della Collana è il/la Direttore/rice del DIT, cui si affianca un comitato scientifico internazionale che varia in relazione alle tematiche trattate.

Quale italiano?

Una proposta per l'insegnamento
dell'italiano a stranieri
tra standard e neostandard

Anna Zingaro

Prefazione di Michele Prandi

Bologna
University Press

Progetto Open Access Consorzio Alphabet

Il volume beneficia di un contributo alla pubblicazione da parte del Dipartimento di Interpretazione e Traduzione dell'Alma Mater Studiorum - Università di Bologna

Fondazione Bologna University Press
Via Saragozza 10 – 40123 Bologna
tel. (+39) 051 232 882
fax (+39) 051 221 019

www.buponline.com
email: info@buponline.com

Quest'opera è pubblicata sotto licenza CC-BY-4.0

ISSN: 2283-8910
ISBN: 979-12-5477-343-7
ISBN online: 979-12-5477-344-4
DOI 10.30682/sitlec46

Grafica: Alessio Bonizzato
Impaginazione: Sara Celia

Prima edizione: novembre 2023

La libertà nella lingua deve venire dalla perfetta scienza e non dall'ignoranza. La quale debita e conveniente libertà manca oggigiorno in quasi tutti gli scrittori. Perché quelli che vogliono seguire la purità e l'indole e le leggi della lingua, non si portano liberalmente, anzi da schiavi. Perché non possedendola interamente e fortemente, e sempre sospettosi di offendere, vanno così legati, che pare camminino fra le uova.

Giacomo Leopardi, Zibaldone, [704] 28 febbraio 1821

Sommario

11 **Prefazione**

Michele Prandi

17 **Introduzione**

23 **Struttura dell'opera e principali riferimenti teorici**

27 **1. Le varietà dell'italiano contemporaneo**

27 1.1. Brevi cenni storici

29 1.2. Gli assi di variazione: le tante fisionomie della lingua

33 1.3. Lo spazio linguistico italiano

34 1.3.1. Italiano standard

36 1.3.2. Italiano popolare (o “dei semicolti”)

38 1.3.3. Italiano regionale

40 1.3.4. Italiano dell'uso medio

43 1.4. L'italiano dell'uso medio: verso un nuovo standard?

64 1.5. La lingua ritrovata

71 **2. I criteri di un nuovo standard linguistico**

71 2.1. Come definire lo standard linguistico

73 2.2. Il criterio della funzionalità

82 2.3. Il criterio dell'accettabilità sociale

83 2.4. Applicazione dei criteri della funzionalità e dell'accettabilità sociale

88 2.5. Conclusioni sull'applicazione dei criteri della funzionalità e dell'accettabilità sociale

91	3. I corpora nella didattica delle lingue
91	3.1. Alcune nozioni di linguistica dei corpora
97	3.2. Metodo per l'identificazione dei fenomeni tramite corpus
99	3.3. Tipologie di corpora esistenti
103	3.4. Il web come corpus
105	3.5. Corpora nella linguistica cognitiva e nella <i>construction grammar</i>
106	3.6. Corpora e apprendimento delle lingue
109	3.7. Uso dei corpora per la creazione di materiali didattici
117	3.8. L'uso di materiali autentici
121	4. Creare un'unità didattica: obiettivi e metodi
122	4.1. Domande della ricerca
123	4.2. Corpus selezionato per l'indagine: PAISÀ
128	4.3. Criteri di individuazione del corpus da utilizzare e confronto tra corpora
137	4.4. Progettazione delle stringhe di ricerca
138	4.4.1. Il linguaggio CQL
144	4.4.2. Elaborazione delle stringhe di ricerca
165	4.5. Metodo per il salvataggio e il trattamento dei risultati dell'interrogazione del corpus
166	4.6. Metodo per la progettazione e sperimentazione delle unità didattiche
167	4.6.1. Selezione dei testi
169	4.6.2. Valutazione di alcune piattaforme di <i>e-learning</i>
173	4.6.3. Progettazione delle unità didattiche
181	4.7. Questionario sulle opinioni degli studenti
187	5. Frequenza dei fenomeni grammaticali indagati
188	5.1. Calcolo della frequenza dei fenomeni indagati
203	5.2. Criteri di selezione dei fenomeni indagati

209	6. Descrizione e sperimentazione delle unità didattiche
210	6.1. I tratti linguistici selezionati
211	6.2. Descrizione delle unità
215	6.2.1. Unità didattica sui verbi riflessivi “affettivi”
221	6.2.2. Unità didattica sul congiuntivo
227	6.3. Esiti della sperimentazione
230	6.3.1. Esiti delle attività finali sui verbi riflessivi e sull’uso del congiuntivo vs indicativo
238	6.3.2. Riflessioni finali sugli esiti delle sperimentazioni
243	6.4 Opinioni sulle attività svolte e sulla piattaforma
244	6.4.1. Unità didattica sui verbi riflessivi “affettivi”: opinioni sulle attività
250	6.4.2. Unità didattica sui verbi riflessivi “affettivi”: opinioni sulla piattaforma <i>Wizer.me</i>
252	6.4.3. Unità didattica sul congiuntivo: opinioni sulle attività
257	6.4.4. Unità didattica sul congiuntivo: opinioni sulla piattaforma <i>Wizer.me</i>
259	Conclusioni
265	Bibliografia
299	Sitografia
303	Appendice 1
309	Appendice 2
319	Appendice 3

PREFAZIONE

Michele Prandi

già Professore ordinario di Linguistica, Università degli Studi di Genova

Il libro nasce da una ricerca alimentata da due fonti: da un'esperienza pluriennale dell'autrice nell'insegnamento dell'italiano come lingua seconda (L2) a studenti stranieri, e da una riflessione sui limiti degli strumenti didattici, e in particolare delle grammatiche, a disposizione dei docenti e degli studenti di L2. Nell'insegnamento dell'italiano ai non nativi, la domanda inaggirabile è quale varietà di italiano possa essere insegnata come uno standard condiviso e affidabile. I manuali disponibili, d'altra parte, si limitano a riproporre lo standard di italiano tramandato dalle grammatiche scolastiche normative, che presenta limiti funzionali evidenti. Siccome si fonda su usi scritti, prevalentemente letterari e comunque elitari, ha una portata limitata ed è disancorato dagli usi effettivi; per questo lascia il docente senza una guida e disorienta lo studente che si confronta quotidianamente con un uso vivo dalla fisionomia completamente diversa.

Dopo aver preso atto di questa situazione, la ricerca confluita in questo volume si propone un obiettivo duplice. In primo luogo, si propone

di elaborare strumenti didattici basati su dati linguistici autentici in grado di offrire al discente non madrelingua un modello di italiano utilizzabile e affidabile in grado di coprire l'intera tastiera degli usi. Il perseguimento di questo primo obiettivo, tuttavia, presuppone che sia disponibile uno standard di italiano al tempo stesso realistico, documentato negli usi effettivi, e in grado di fornire una guida sicura al docente e all'allievo. In realtà, uno standard al tempo stesso rinnovato e adatto allo scopo non è al momento disponibile, ma deve essere messo a punto con criteri espliciti e affidabili. Identificare un neostandard in possesso di questi requisiti è il primo obiettivo che il volume si propone di affrontare, ed è il tema sul quale vorrei proporre al lettore una mia riflessione.

Negli ultimi decenni, i linguisti hanno certamente messo a fuoco un divario crescente tra uno standard pensato per un numero di parlanti limitato e per usi ristretti, e gli usi effettivi che si sono imposti da quando, a partire dall'Unità d'Italia ma soprattutto nel secondo dopoguerra, il numero dei parlanti e dei contesti d'uso si è moltiplicato, mettendo a nudo l'inadeguatezza del modello tradizionale. In particolare, due studi classici di Sabatini (1985) e di Berruto (1987) si sono proposti di identificare i tratti caratterizzanti di una forma rinnovata di italiano che i due studiosi chiamano rispettivamente *italiano dell'uso medio* e *neostandard*. Le due etichette mettono l'accento ciascuna su una delle due dimensioni indissociabili del problema. L'etichetta *italiano dell'uso medio* registra una costellazione di fenomeni innovativi diffusi tra i parlanti che travolgono letteralmente la norma tradizionale, spinti in primo luogo da ragioni funzionali. L'etichetta *neostandard* sposta invece l'attenzione sull'esigenza di un rinnovamento dello standard condiviso di riferimento. La sfasatura tra la severità della norma e le ragioni dell'uso, se impone di mettere in discussione lo standard canonizzato dalle grammatiche pedagogiche, impegna con altrettanta forza a mettere a punto uno standard alternativo compatibile con l'uso. Sulla strada dell'elaborazione di un nuovo standard, il censimento delle forme che si sono imposte nell'uso medio negli ultimi decenni è solo un primo passo, essenziale ma non sufficiente.

Uno standard non si limita a recepire l'esistente, ma deve essere in grado di adempiere alla sua funzione di dare la sicurezza linguistica a chi lo usa e un metro sicuro a chi valuta gli usi. Ora, per rispondere a

queste funzioni, uno standard deve essere un modello di uso accettato come tale, e quindi chiuso alle forme che marchiano chi le usa, nel giudizio medio, come parlante di una varietà substandard. Se misurati con l'ideale di uno standard rinnovato ma affidabile come una bussola, i dati documentati dall'«uso medio» appaiono disomogenei. Nella lista dei tratti identificati da Sabatini (1990), ad esempio, troviamo certamente risorse funzionali essenziali, tradizionalmente condannate come deviazioni dalla norma, come la frase scissa, la dislocazione e l'anacoluti. Tuttavia, troviamo anche deviazioni prive di vantaggi funzionali, come ad esempio l'uso del *ci* con il verbo *avere* in costrutti come *non c'ho tempo*. È evidente che, in vista della messa a punto di un neostandard, gli usi devianti rispetto allo standard tradizionale non devono essere semplicemente accettati, ma devono essere vagliati alla luce di criteri espliciti. L'individuazione di questi criteri e la loro applicazione a ciascuna delle forme innovative attestate dall'«uso medio» rappresenta ai miei occhi il contributo più significativo del volume.

Il primo criterio di selezione è funzionale. Un esempio illuminante di applicazione di questo criterio è la valutazione diversa dell'estensione del pronome *gli* alla terza persona plurale, al posto di *loro*, e alla terza persona femminile singolare, al posto di *le*. La prima estensione ha una giustificazione funzionale indiscutibile. I pronomi di terza persona plurale sono gli unici che non conoscono l'alternanza tra forma atona e forma tonica, destinata alla posizione di fuoco, per l'oggetto indiretto: l'unica forma *loro*, in effetti, è sempre tonica e occupa sempre la posizione di fuoco. Alle forme *Ho scritto a lui*, *Gli ho scritto*, ad esempio, corrispondono le forme *Ho scritto a loro*, *Ho scritto loro*. L'adozione della forma *gli* è funzionale perché, offrendo all'espressione dell'oggetto indiretto plurale la forma atona, colma una vistosa lacuna del paradigma e arricchisce la tastiera di opzioni offerte al parlante: *Ho scritto a loro*, *Gli ho scritto*. Lo stesso criterio, viceversa, vede l'estensione della forma *gli* al femminile singolare al posto di *le* come una caduta di registro non compensata da vantaggi funzionali, e quindi idealmente da escludere dallo standard.

Il criterio funzionale, che è dirimente quando la forma innovativa è in grado di ampliare il repertorio di opzioni offerte al parlante, perde efficacia quando la forma tradizionale e la forma innovativa sono entrambe prive di vantaggi funzionali. In questo caso, l'autrice propone

di adottare un criterio indipendente e complementare, basato sulla condivisione sociale: in assenza di vantaggi funzionali, la conservazione di una forma standard è preferibile all'adozione di una forma che espone il parlante alla censura della comunità. L'esempio più significativo analizzato nel volume è quello del modo congiuntivo nelle frasi subordinate. Contrariamente a quanto sostenuto dalla tradizione grammaticale, il congiuntivo non porta nella frase un valore modale di non realtà che ne giustifichi l'uso. Più in generale, non ha un valore funzionale proprio perché la sua selezione è controllata dall'esterno: dal verbo reggente nelle subordinate argomentali, e dalla congiunzione nelle subordinate non argomentali. Per questo, da un lato l'abbandono del congiuntivo in favore dell'indicativo non comporterebbe un danno funzionale; d'altro canto, per le stesse ragioni, la sostituzione non porterebbe nessun vantaggio in grado di controbilanciare la censura sociale.

Al criterio della condivisione sociale si possono opporre due ordini di obiezioni. In primo luogo, si potrebbe obiettare che il criterio ricalca il modello della grammatica prescrittiva. Il modello proposto, tuttavia, sfugge all'obiezione perché si basa su presupposti incommensurabili. A differenza della precettistica tradizionale, il criterio della condivisione sociale non impone imperativi categorici ma propone imperativi ipotetici: per esempio, non afferma "Non usare l'indicativo nella forma dell'irrealtà del periodo ipotetico – *Se lo sapevo non venivo*" – ma "Se usi l'indicativo nella forma dell'irrealtà del periodo ipotetico, il registro della tua forma di espressione si abbassa e, in certe occasioni, ti espone a una censura sociale". Nel momento in cui suggerisce una norma, il criterio lo fa scommettendo sulla consapevolezza e sulla responsabilità del parlante che compie le sue scelte come un soggetto attivo. Lo stesso docente che è pronto a usare la forma substandard del periodo ipotetico dell'irrealtà in una conversazione tra amici non la userebbe durante una prova di concorso. Per la stessa ragione, sa che non deve insegnare a un apprendente straniero una varietà di italiano priva del congiuntivo, che impoverirebbe la sua tastiera espressiva senza vantaggi funzionali e lo esporrebbe a una censura sociale sicura. Questo non esclude che, in una fase avanzata dell'apprendimento, quando il grado di consapevolezza raggiunto dall'allievo lo mette in grado di compiere scelte motivate, lo stesso docente lo possa sensibilizzare alle variazioni

di registro in relazione al contesto. In secondo luogo, la condivisione sociale è un dato dinamico, le cui condizioni possono cambiare nel tempo. Lo standard condiviso di una lingua romanza come il francese documenta un uso del congiuntivo più ridotto del nostro. La forma standard del periodo ipotetico, ad esempio, assomiglia più alla forma substandard dell'italiano che alla forma standard, e non è escluso che la deriva del francese possa prefigurare la norma di una fase futura dell'italiano. All'obiezione, ovviamente, non si può dare una risposta *a priori*. Qualsiasi standard è per sua natura soggetto a mutamenti perché l'evoluzione delle preferenze dei parlanti per una forma (*Se lo sapevo non venivo; Se lo avessi saputo non sarei venuto*) non sono prevedibili. L'unica strada che il linguista può ragionevolmente percorrere è il monitoraggio degli usi documentati e della loro ricezione sociale e la disponibilità a rivedere lo standard in funzione della condivisione effettiva.

Se l'argomentazione, come ritengo, è corretta, l'interesse del volume va oltre la domanda che l'ha fatto nascere, e cioè quale italiano insegnare agli apprendenti di L2. La domanda alla quale il volume propone una risposta, in effetti, ha una portata più ampia: quale potrebbe essere il profilo di una varietà neostandard di italiano realistica, funzionale e socialmente condivisa in questo inizio di millennio.

INTRODUZIONE

Tra le obiezioni più comuni che l'insegnante di lingua italiana per stranieri si sente rivolgere in aula quando si affrontano alcune strutture grammaticali c'è: "Però, ho sentito che gli italiani dicono questo [...]: è corretto?", dove da quel *però* traspare la perplessità suscitata dal contrasto tra una norma grammaticale insegnata e interiorizzata e l'uso quotidiano della lingua. Questo accade tipicamente nel contesto di apprendimento dell'italiano L2¹, in cui chi apprende l'italiano è esposto direttamente alla lingua viva, parlata nel mondo circostante.

¹ Quando una lingua non materna viene appresa nel paese in cui è parlata dai suoi parlanti nativi, è definita "lingua seconda", o L2, per differenziarla dalla lingua materna (L1). Ad esempio, una persona cinese che frequenta un corso di italiano in Italia, sta studiando l'italiano come "lingua seconda" (L2). Quando invece una lingua non materna viene appresa in un paese diverso da quello in cui è parlata dai suoi parlanti nativi, si parla di "lingua straniera" (LS). Ad esempio, se una persona italiana frequenta un corso di inglese in Italia, sta studiando l'inglese come "lingua straniera" (LS).

Questa esposizione, oltre ad arricchire chi apprende, però può anche suscitare dubbi insolubili sull'uso linguistico, che spesso vengono portati in aula, sperando che a scioglierli definitivamente sia l'insegnante, in veste – per così dire – di oracolo linguistico.

Questi dubbi sull'uso possono derivare, ad esempio, dal fatto che la competenza linguistica degli apprendenti non sia ancora abbastanza elevata da includere anche quella specifica struttura linguistica che crea loro difficoltà. Un altro scenario possibile è che gli apprendenti si imbattono in strutture linguistiche per le quali non trovano riferimenti teorici nei manuali per stranieri (né, di fatto, li troverebbero nelle grammatiche destinate agli italofoeni nativi) o che in qualche modo contrastano con le norme linguistiche in essi contenute.

Quest'ultimo scenario si verifica, ad esempio, nel caso di strutture linguistiche che, pur essendo ampiamente utilizzate dai parlanti nativi, si collocano al di fuori della varietà riconosciuta come standard dell'italiano e che pertanto o non compaiono nei manuali di lingua italiana L2/LS, oppure sono oggetto di una trattazione molto limitata.

Negli ultimi due secoli la lingua italiana ha subito, infatti, un processo di trasformazione a livello morfologico, sintattico e lessicale. Prima dell'Unità d'Italia (1861), l'italiano era la lingua usata da un numero ristretto di persone e coincideva grosso modo con la varietà parlata in Toscana, mentre nel resto del paese era utilizzata solo nella burocrazia e dai ceti sociali più elevati.

Dopo aver assunto il ruolo di lingua nazionale, questa lingua, fino ad allora utilizzata principalmente nella scrittura di registro formale e con uno standard tendente all'aulico, dovette estendere il proprio raggio d'azione alle esigenze pratiche della comunicazione nella vita quotidiana. Fu così che, sotto l'influenza delle varietà linguistiche parlate nelle diverse regioni, l'italiano subì una graduale trasformazione che fece emergere vari tratti linguistici già esistenti da secoli e ampiamente utilizzati nel parlato, ma considerati scorretti secondo la grammatica tradizionale che codificava l'italiano scritto.

Censurati per secoli dai grammatici, ma maggiormente rispondenti alle esigenze di una comunicazione reale, parlata e scritta (cfr. D'Achille, 1990), questi tratti sono giunti fino ai giorni nostri, coesistendo con le relative forme standard, e ispirando nel 1985 uno studio di Francesco Sabatini, intitolato *L'italiano dell'uso medio: una realtà tra le va-*

rietà linguistiche italiane. Questo studio aveva l'obiettivo di elencare e descrivere i tratti linguistici caratteristici di una nuova varietà standard della lingua nazionale, che erano ormai entrati anche nello scritto di una certa formalità, sebbene alcuni di essi storicamente fossero percepiti come una deviazione dalla norma codificata dalle grammatiche. La conclusione di questo studio fu che nella scelta della varietà linguistica da trasmettere nell'insegnamento dell'italiano bisognasse tenere conto dei mutamenti subiti dalla lingua, riproponendo, così, la necessità di definire uno standard linguistico.

A questo proposito, se è vero che, come messo in evidenza da Diadori, Palermo e Troncarelli (2015, p. 227) “per il docente di italiano L2 il problema fondamentale è quello di evitare l'eccessivo scollamento tra lingua utilizzata in aula e quella effettivamente presente nelle situazioni comunicative quotidiane”, allo stesso tempo è necessaria una precisazione. Ossia, nell'avvicinarsi il più possibile alla lingua parlata nella vita quotidiana, è necessario anche avere degli strumenti che permettano di distinguere che cosa possa essere considerato come parte integrante della nuova varietà standard e che cosa vada collocato, invece, nel substandard.

Da questa riflessione hanno origine alcuni degli interrogativi che si pone chi insegna l'italiano a stranieri: *quale* italiano insegnare? Quale varietà viene effettivamente proposta dai manuali attualmente in commercio? Come fornire un supporto a chi insegna e a chi apprende l'italiano che possa favorire una percezione realistica delle stratificazioni sociolinguistiche della lingua usata al giorno d'oggi?

Questa ricerca nasce dalla necessità di dare una risposta a queste domande, proponendosi come un'opera destinata principalmente agli insegnanti di italiano L2/LS e a chi studia per diventarlo e anche a chi applica metodi di ricerca *corpus-based* (Tognini-Bonelli, 2001) allo studio della lingua.

L'affermarsi di un “italiano dell'uso medio” è un segnale inequivocabile di come lo standard tradizionale, tramandato e consacrato dalle grammatiche, sia stato nel corso del tempo percepito come inadeguato sul piano funzionale da parte dei parlanti, che hanno iniziato ad adottare forme non standard, perché erano alla ricerca di funzioni che la norma codificata dell'italiano letterario non era in grado di assicurare. Tuttavia, limitarsi a prendere atto dell'esistenza delle forme linguistiche che costituiscono l'italiano dell'uso medio in quanto ormai diffuse

nell'uso, non è una soluzione definitiva e vale la pena di chiedersi: utilizzando queste forme linguistiche (e quindi non rispettando lo standard), i parlanti aggiungono effettivamente alla comunicazione qualcosa che lo standard tradizionale non metteva loro a disposizione o provocano unicamente un abbassamento del registro linguistico che li espone alla censura sociale (ossia, la percezione del *non saper parlare l'italiano*)?

Sulla base di queste riflessioni, verranno proposti alcuni criteri metodologici che permettano di riesaminare i tratti linguistici elencati da Sabatini nel 1990 come costitutivi dell'italiano dell'uso medio. L'obiettivo è di determinare quali di questi tratti linguistici siano da considerarsi effettivamente come parte di una possibile varietà neostandard e, di conseguenza, vadano trasmessi nell'insegnamento della lingua come parte della nuova norma e quali, invece, possano essere semplicemente menzionati come forme colloquiali substandard in cui è possibile imbattersi.

Una volta definiti i confini del nuovo standard linguistico e delle forme che devono essere insegnate, e data la scarsa presenza di approfondimenti sul neostandard nella manualistica per stranieri attualmente in commercio (cfr. Słapek, 2016)², si è fatta strada l'idea di progettare da zero delle unità didattiche su alcuni tratti dell'italiano neostandard, partendo dalla lista dei 14 fenomeni linguistici classificati da Sabatini (1990), e di sperimentarle in aula.

A questo proposito verrà fatta una riflessione sui contenuti grammaticali della manualistica per stranieri attualmente in commercio. Ciò permetterà di determinare in che misura venga dato spazio al neostandard, quali fenomeni linguistici siano stati finora esclusi o scarsamente trattati dalle pubblicazioni esistenti e se, tra questi ultimi, ci siano dei fenomeni linguistici per i quali possa essere rilevante e utile la creazione *ad hoc* di materiali didattici, al fine di dare una percezione realistica della varietà neostandard utilizzata al giorno d'oggi.

² Come verrà in seguito approfondito nella trattazione, la ricerca esposta da Słapek nel 2016 non è incentrata sul neostandard, bensì traccia una panoramica generale degli argomenti grammaticali presenti nella manualistica per stranieri. Dalla consultazione di tali dati ho potuto osservare l'eventuale presenza dei tratti del neostandard e dare avvio a questa ricerca.

La selezione dei fenomeni linguistici su cui incentrare le unità didattiche e la progettazione di queste ultime rappresenta il punto di intersezione, in questo studio, tra la linguistica italiana, la glottodidattica e la linguistica dei corpora, un ambito di ricerca che si è sviluppato negli ultimi anni del ventesimo secolo all'interno degli studi sull'uso della lingua e sulla variazione (Biber, Conrad, Reppen, 1998; Biber, 2012). I corpora sono delle raccolte di grandi dimensioni di testi orali o scritti prodotti in contesti comunicativi reali (ad esempio, raccolte di articoli di giornale, libri, registrazioni di discorsi, ecc.). Tali testi sono conservati in formato elettronico e possono essere consultati tramite l'uso di strumenti informatici (cfr. Baroni, 2010).

Quanto più un corpus sarà esteso e composto da fonti variegate, tanto più potrà essere considerato rappresentativo di una certa lingua. Di conseguenza, la sua consultazione permetterà di ricavare un modello del funzionamento della lingua, verificandone le tendenze generali su base quantitativa, osservando, ad esempio, quali siano le parole più utilizzate, come le parole si combinino le une alle altre e quali combinazioni di parole siano più comuni.

Uno dei metodi di ricerca principalmente utilizzati nella linguistica dei corpora è definito *corpus-based*, ossia basato su corpus, in cui i dati estratti da un corpus vengono utilizzati per verificare una teoria, con l'obiettivo di convalidarla, confutarla o perfezionarla. Nel contesto di questo lavoro, l'approccio *corpus-based* consente innanzitutto di verificare la diffusione dei tratti linguistici dell'italiano dell'uso medio nei testi scritti di media formalità degli ultimi vent'anni e, di conseguenza, di reperire un numero elevato di testi che contengono i tratti grammaticali oggetto di studio, da poter poi utilizzare per la progettazione da zero di unità didattiche.

Nella scelta dei fenomeni grammaticali su cui incentrare la progettazione delle unità didattiche si terrà conto non solo del criterio della frequenza statistica, ma anche e soprattutto di due criteri metodologici che verranno proposti nell'opera per riesaminare i tratti linguistici elencati da Sabatini nel 1990 e stabilire quali di essi vadano trasmessi nell'insegnamento. Inoltre, verranno presi in considerazione anche i contenuti grammaticali della manualistica per stranieri attualmente in commercio (cfr. Słapek, 2016), al fine di individuare i fenomeni finora esclusi dalle pubblicazioni esistenti e

per i quali possa quindi essere rilevante e utile la creazione di materiali didattici.

Una volta identificati i fenomeni grammaticali su cui incentrare la progettazione, i testi che li contengono da utilizzare per progettare le unità didattiche e la piattaforma di *e-learning* sulla quale collocarle, si descriverà la struttura delle unità didattiche prodotte.

A questo proposito, si proporrà un metodo per la progettazione di unità didattiche nelle quali verrà fornito un inquadramento teorico non solo dell'uso standard di alcuni fenomeni grammaticali, ma anche del loro uso nella varietà neostandard, che altrimenti, in assenza di un riferimento teorico, potrebbe essere percepito dagli apprendenti come errato o inspiegabile. Questo metodo di progettazione costituisce un'alternativa alle modalità generalmente adottate nella manualistica, ossia la breve nota di approfondimento, talvolta a piè di pagina, in cui si informano gli apprendenti dell'uso "colloquiale", oppure una trattazione più estesa, che però non fornisce un confronto tra uso standard e neostandard. A titolo esemplificativo verranno illustrate le attività di due unità didattiche prodotte e verranno commentati gli esiti della loro sperimentazione in alcuni atenei esteri.

Come emergerà con maggiore chiarezza nell'esposizione, si ritiene dunque che uno degli aspetti innovativi della presente ricerca risieda proprio nel fornire un metodo sia per la selezione dei fenomeni linguistici, sia per la progettazione di unità che integrino la trattazione dell'uso standard e neostandard.

Tutte queste riflessioni, che si inseriscono nel dibattito sullo standard della lingua italiana, prendono spunto dalle esigenze emerse nell'insegnamento della lingua italiana a stranieri presso il Dipartimento di Interpretazione e Traduzione dell'Alma Mater Studiorum - Università di Bologna (Campus di Forlì), nello specifico nell'attività di tutorato e nell'insegnamento di *Lingua Italiana L2*. Un pensiero particolare va, quindi, alle studentesse e agli studenti che dal 2009 a oggi, portandomi i loro dubbi in classe, mi spingono a interrogarmi costantemente su come funziona la lingua italiana e come io possa trasmetterla al meglio a loro.

STRUTTURA DELL'OPERA E PRINCIPALI RIFERIMENTI TEORICI

Questo studio si basa sulla rielaborazione della mia tesi di dottorato (Zingaro, 2021), di cui rappresenta un'evoluzione teorica per quanto riguarda i contenuti dei capitoli 1 e 2.

La prima parte (capitoli 1-3) fornisce un inquadramento teorico di linguistica italiana e linguistica dei corpora. Nel primo capitolo, dopo aver dato alcuni brevi cenni sulla storia della lingua italiana (D'Achille, 2003; De Mauro, 2011; Dardano, 2020 a c. di; Frosini, 2020 a c. di; Palermo, 2020), verrà tracciata una panoramica delle varietà linguistiche, dando particolar rilievo all'italiano dell'uso medio/neostandard (Sabatini, 1985, 1990; Berruto, 1987). In particolare, verranno esaminati nel dettaglio i fenomeni grammaticali che caratterizzano l'italiano dell'uso medio. La descrizione sarà integrata sia con gli studi sui singoli fenomeni (si citano a titolo di esempio Simone, 1993; Cortelazzo, 2000a; Tavoni, 2002; D'Achille, Proietti, Viviani, 2005; D'Achille, 2010; Renzi, 2012), sia con gli studi che hanno fornito un contributo fondamentale nel dibattito sulla ristrutturazione dello standard lingui-

stico italiano (ad esempio: Sabatini, 1985, 1990, 1997, 2016; Berruto, 2012 [1987]; Nencioni, 1987; Castellani, 1991; Sobrero, 1992; Dardano, 1994; Lorenzetti, 2002; D'Achille, 2003, 2016; Tesi, 2005; Renzi, 2012; Ondelli, Romanini, 2018; Prandi, 2020).

Come già affermato, la necessità di definire un modello linguistico di riferimento è un tema centrale in questa ricerca, poiché ha delle implicazioni nella scelta della varietà linguistica da trasmettere nell'insegnamento dell'italiano. Per queste ragioni, nel capitolo 2 verranno proposti alcuni criteri metodologici che permettano di riesaminare i tratti linguistici elencati da Sabatini nel 1990 come costitutivi dell'italiano dell'uso medio e determinare quali di essi possano effettivamente essere ritenuti parte della nuova varietà standard e quali, invece, siano da ritenersi substandard, seppur diffusi.

A ciò si accompagna la necessità di reperire dei testi che permettano agli apprendenti di avere una percezione della lingua usata al giorno d'oggi. Pertanto, nel capitolo 3 verranno illustrate le ragioni per le quali i corpora possono rappresentare delle valide risorse per l'apprendimento della lingua e, nel caso specifico, per il reperimento di testi autentici che contengano i fenomeni grammaticali dell'italiano dell'uso medio. Dopo aver illustrato alcuni concetti base della linguistica dei corpora (McEnery, Wilson, 1996; Bieber *et al.*, 1998; McEnery, Hardie, 2001; Kilgarrieff, Grefenstette, 2003; Baroni, 2010; McEnery, Hardie, 2012), verranno presentati alcuni ambiti di applicazione dei corpora alla didattica delle lingue, dedicando particolare attenzione al loro utilizzo per la realizzazione di materiali didattici. Per quest'ultimo ambito si farà riferimento alle pubblicazioni contenute in Sinclair (2004), Granger, Meunier (2008), Meunier, Granger (2008), Aijmer (2009) e Hinkel (2011), agli atti dei convegni *Teaching and Language Corpora* (TaLC) e alla bibliografia ivi contenuta, raggruppando gli studi per aree tematiche. In particolare, per l'italiano L2/LS si farà riferimento a Corino, Marello (2009), Merlo (2009), Kennedy, Miceli (2010), Guidetti, Lenzi, Storchi (2012), Lyding *et al.* (2013) e Ivanovska-Naskova (2018).

La seconda parte di questo studio (capitoli 4-6) è dedicata all'esposizione degli obiettivi e dei metodi della ricerca. Dopo aver esposto le domande alla base di questa ricerca, verranno brevemente presentati i vari corpora dell'italiano contemporaneo scritto attualmente dispo-

nibili. Successivamente verranno esposti i criteri di selezione utilizzati per individuare il corpus più idoneo per il reperimento di testi che contengano i fenomeni grammaticali in questione e che possano essere riutilizzati per la produzione di materiali didattici.

Una volta individuato il corpus, si illustrerà il metodo adottato per la sua interrogazione tramite la progettazione di alcune stringhe di ricerca in linguaggio Corpus Query Language (Evert, 2009; Nesi, Thompson, 2014), utilizzate per ricercare i fenomeni grammaticali elencati da Sabatini nel suo studio del 1990.

Infine, verrà presentato il metodo adottato per la progettazione e sperimentazione delle unità didattiche e per la realizzazione di un questionario per la raccolta delle opinioni degli studenti sulle attività svolte. Di quest'ultimo si descriveranno la struttura, il contenuto e le caratteristiche principali della piattaforma *online* su cui è stato collocato.

Per la progettazione delle unità didattiche si farà riferimento alla *Second Language Acquisition Theory* di Krashen del 1982, alla didattica acquisizionale (Vedovelli, 2003; Vedovelli, Villarini, 2003) e al modello dell'unità didattica ripreso e approfondito da Danesi (1988, 1998) da un punto di vista neurolinguistico. Verrà messa in luce l'importanza di strutturare le attività sulla base dell'ordine in cui si attivano i due emisferi cerebrali (Balboni, 1994; Danesi, 1998, p. 152 e sgg.; Balboni, 2002, 2008). Per l'elaborazione del questionario si farà riferimento agli studi di Nunan (1989), Coonan (2000) e Pozzo (2008).

Nel capitolo 5 si illustreranno i procedimenti adottati per il calcolo della frequenza all'interno del corpus dei fenomeni grammaticali oggetto di studio, al fine di determinare quali di essi siano i più ricorrenti. Questo procedimento permette di individuare quali tra i fenomeni linguistici elencati da Sabatini (1990) siano ora maggiormente utilizzati nella produzione scritta di media formalità dagli italofoeni nativi e la cui conoscenza è necessaria agli apprendenti di lingua italiana L2/LS al fine di raggiungere una competenza linguistica vicina all'uso reale della lingua. L'analisi quantitativa sarà supportata dalla presentazione dei dati in tabelle e dalla loro rappresentazione grafica.

Una volta identificati i fenomeni grammaticali su cui incentrare la progettazione delle unità, i testi (estratti dal corpus) in cui ricorrono questi fenomeni e che verranno utilizzati per progettare queste unità didattiche e la piattaforma di *e-learning* sulla quale collocarle, nel ca-

pitolo 6 si descriverà la struttura delle unità didattiche prodotte. Nella loro progettazione si terrà conto sia della collocazione dei fenomeni grammaticali selezionati tra i livelli del Quadro Comune Europeo di Riferimento per la conoscenza delle lingue (QCER) (Consiglio d'Europa 2018, 2002), sia dei fattori neurolinguistici che regolano l'ordine in cui i contenuti linguistici vengono acquisiti dalla mente umana e che, pertanto, sono alla base dell'acquisizione di una lingua (cfr. capitolo 4). Verranno, inoltre, descritti gli esiti delle sperimentazioni delle unità didattiche presso i dipartimenti di italianistica di alcuni atenei europei, che saranno poi confrontati con i dati raccolti tramite un questionario sulle opinioni dei partecipanti. In particolare, nel questionario verranno presi in considerazione i seguenti fattori: la lingua materna degli apprendenti (L1), la conoscenza di altre lingue straniere (LS), la durata dello studio pregresso della lingua italiana, la familiarità dei partecipanti con gli argomenti trattati, sia dal punto di vista dell'uso standard che dell'uso neostandard e, infine, la presenza nella L1 e nelle altre LS conosciute di argomenti grammaticali simili.

Infine, nelle Conclusioni verranno proposte alcune riflessioni sulla validità dell'approccio metodologico proposto in questa ricerca, mirata a promuovere una percezione realistica delle stratificazioni sociolinguistiche della lingua usata al giorno d'oggi e fornire degli strumenti di supporto a chi insegna e a chi apprende l'italiano.

1. LE VARIETÀ DELL'ITALIANO CONTEMPORANEO

Dopo alcuni brevi cenni sulla storia della lingua italiana (cfr. D'Achille, 2003; De Mauro, 2011; Dardano, 2020 a c. di; Frosini, 2020 a c. di; Palermo, 2020), verranno trattati gli aspetti sociolinguistici che permettono di identificare le varietà della lingua italiana e le differenze d'uso sulla base degli assi di variazione (Coseriu, 1973; Mioni, 1983; Berruto, 1993a e Berruto, 1993b).

1.1. Brevi cenni storici

Come tutte le varietà romanze, l'italiano deriva dal latino volgare, in particolare dalle varietà della lingua latina parlate dalle diverse popolazioni italiane. Tali varietà si distaccavano dal latino classico-letterario, con forti variazioni di carattere geografico e legate ai gruppi sociali di appartenenza, attraverso gli influssi ricevuti dai substrati linguistici locali delle diverse regioni in cui il latino era stato importato. La presenza di numerose varietà linguistiche locali causò sin dal Trecento l'apertura

di una disputa in ambito letterario, definita *questione della lingua*, nella quale ci si interrogava su quale lingua utilizzare nella penisola italiana. L'origine della disputa viene fatta risalire al *De vulgari eloquentia* (1303-04), nel quale Dante, alla ricerca di un *volgare illustre* per la lingua poetica, individuò 14 dialetti, ma giunse alla conclusione che nessuno di essi fosse adeguato a divenire il canone letterario. Tale ideale linguistico poteva essere elaborato solo in una corte ideale, nella quale ciascun letterato italiano portasse i contributi migliori dal proprio dialetto.

La disputa si prolungò per secoli, vivendo una delle sue fasi più acute nel Cinquecento anche in seguito all'avvento della stampa, che fece sorgere la necessità pratica di utilizzare una varietà linguistica omogenea in tutta la penisola italiana. Fu così che nel Cinquecento alcune strutture linguistiche del fiorentino letterario di Dante, Petrarca e Boccaccio, quali la fonologia, la morfologia, numerosi aspetti della sintassi e il vocabolario di base (Dardano, 2020), vennero prese a modello dell'uso scritto della lingua da parte della linea classicistica dei grammatici, capeggiati da Pietro Bembo, che animarono la fase più acuta della *questione della lingua*.

Tra le tappe importanti della *questione della lingua* nella codificazione dello standard si ricordano la pubblicazione nel 1525 delle *Prose della volgar lingua* di Bembo, in cui come modelli per la redazione di opere letterarie vennero proposti Francesco Petrarca per la poesia e Giovanni Boccaccio per la prosa, poi la pubblicazione della prima edizione del *Vocabolario degli Accademici della Crusca* nel 1612, alla quale fecero seguito varie edizioni rivedute. Dunque, il modello di lingua codificato non fu il fiorentino parlato, ma “il toscano urbano della classe colta di Firenze” (Galli de' Paratesi, 1984, p. 60), ossia una varietà scritta letteraria con alcune influenze da altre varietà volgari e dal latino.

Da ciò è possibile trarre due importanti conclusioni. In primo luogo, non tutte le caratteristiche del fiorentino parlato furono ammesse nella codificazione dello standard, caratteristica che influirà nel corso dei secoli nella percezione di distanza tra lo standard linguistico e la lingua comunemente in uso. Inoltre, il fiorentino si impose sugli altri dialetti non grazie ad un potere politico in grado di unificare il territorio e di imporre una politica linguistica, bensì grazie all'affermazione del modello proposto da Bembo sulla base del valore letterario degli autori che lo utilizzarono, al prestigio economico e artistico di Firenze

e anche ad alcune caratteristiche strutturali che lo collocavano in una posizione centrale tra gli altri dialetti italo-romanzi.

La varietà fiorentina, che solo a partire dal Settecento venne identificata con il termine *italiano*, fu usata fino all'Unità d'Italia nel 1861 nello scritto colto, il che favorì una certa stabilità e conservazione delle strutture linguistiche. Con l'unificazione del paese si crearono gradualmente le condizioni affinché l'italiano divenisse una lingua nazionale, non solo scritta, ma anche parlata, non più una seconda lingua affiancata ai diversi dialetti, ma la lingua materna della popolazione: l'alfabetizzazione della popolazione dovuta all'insegnamento scolastico, i cambiamenti nelle condizioni sociali, economiche e culturali, le migrazioni interne al territorio nazionale e, soprattutto, la diffusione dapprima della radio e poi nel corso della seconda metà del Novecento della televisione. Tutti questi fenomeni portarono l'italiano a contatto con fasce sempre più ampie della popolazione e ad allargare sempre più i suoi ambiti d'uso, raggiungendo anche la lingua orale. Tutto ciò ha innescato il processo da cui deriva l'attuale repertorio linguistico del territorio italiano, alla cui natura complessa, caratterizzata da diverse dimensioni, saranno dedicate le sezioni successive.

1.2. Gli assi di variazione: le tante fisionomie della lingua

All'interno del repertorio linguistico italiano è possibile individuare una serie di varietà le cui differenze sono legate, come verrà chiarito di seguito, alla loro diffusione nello spazio e nel tempo, al canale di trasmissione del messaggio, al suo contenuto, alla situazione comunicativa, agli strati e gruppi sociali a cui appartengono gli interlocutori, eccetera.

Nello specifico, i parametri osservabili in una lingua in un dato momento storico, sistematizzati da Eugenio Coseriu (1973), sono tre: la diatopia, relativa allo spazio, in base alla quale una lingua ha caratteristiche diverse a seconda delle aree geografiche in cui è usata, la diastratia, legata alla condizione sociale dei parlanti e influenzata anche dal genere, dall'età, dalla classe sociale, dal livello economico e di istruzione, e la diafasia, legata alla situazione comunicativa, da cui deriva la scelta di un registro linguistico formale o informale (cfr. Coseriu, 1973, pp. 135-152). A questi parametri, che determinano la

variazione linguistica sincronica, si aggiungono la diacronia, relativa ai cambiamenti nell'uso della lingua legati al passare del tempo e, infine, la diamesia, termine coniato da Alberto Mioni (1983, pp. 508-510), che identifica la variazione di una lingua a seconda del mezzo o canale adottato per comunicare.

Quest'ultima non è una semplice opposizione bipolare tra scritto e parlato, ma un continuum scritto-parlato che include una serie di forme intermedie che vanno, attraverso diverse gradazioni, dal polo dello "scritto-scritto", caratterizzato dal rispetto delle norme linguistiche, al "parlato-parlato", nella terminologia di Nencioni (1983 [1976]), passando attraverso una serie di forme intermedie. Con la diffusione dei nuovi mezzi di comunicazione di massa, l'opposizione tra parlato e scritto è stata, infatti, aggiornata per includere testi che presentano caratteristiche sia dello scritto che del parlato, che rientrano nell'italiano trasmesso (Sabatini 1982, 1997) e che spaziano dal "parlato trasmesso" dei mezzi di comunicazione a distanza (radio, televisione, cinema) allo "scritto trasmesso" delle comunicazioni su Internet, tramite la posta elettronica e, successivamente, i vari servizi di messaggistica istantanea che hanno caratteristiche assimilabili a quelle del parlato (cfr. lo *scritto a distanza* vs il *parlato a distanza* in D'Achille, 2003, pp. 216-220; il concetto di *e-taliano* di Antonelli (2016, pp. 139-199), successivamente ripreso in Gheno (2017, pp. 38; 41-102; 2019, pp. 82-94)), includendo anche lo "scritto per essere detto come se non fosse scritto" (Lavinio, 1986, p. 16) dei copioni cinematografici.

La diamesia è, quindi, una dimensione di variazione complessa, "particolarmente necessaria in una situazione linguistica come quella italiana" (Lorenzetti, 2002, p. 22), nella quale la lingua scritta, codificata dalle grammatiche, è stata identificata per secoli con lo standard, in contrapposizione alla lingua d'uso comune, sia nello scritto che nel parlato.

Ed è proprio nell'ambito del dibattito sull'italiano dell'uso medio (Sabatini, 1985) o neostandard (Berruto, 1987) e sull'italiano popolare che si sviluppa la proposta di Mioni di una nuova dimensione di variazione. Interrogandosi su quale posizione occupasse la dimensione diamesica nello studio dell'italiano popolare, Mioni propone per lo studio di questa varietà linguistica di superare l'opposizione tra "scritto = formalità" vs "parlato = informalità", che si realizza solo in casi

limitati della comunicazione, e di considerare invece separatamente la relazione che la lingua scritta e l'oralità hanno con i diversi registri linguistici, ossia, con la diafasia.

La proposta originaria di Mioni di un nuovo asse di variazione autonomo è accolta e ulteriormente sviluppata da Berruto nel 1985 e nei suoi studi successivi (Berruto, 1987, 1993 e 2011b; Berruto e Cerruti, 2014, pp. 147-148), portando ad una vera e propria ridefinizione dello spazio linguistico, in cui la diamesia è da considerarsi trasversale rispetto alle altre dimensioni e dalle quali non può essere completamente isolata, in particolare nel caso della diafasia, se non per specifiche esigenze di studio.

A questo proposito, Berruto (1987, p. 22, n. 12) specifica che, sebbene di solito un registro molto formale possa coincidere con l'uso scritto e viceversa, ci sono anche situazioni in cui il parlato può essere più formale dello scritto: questo avviene perché la scelta dell'uso orale o grafico dipende dalla natura del mezzo utilizzato, mentre il codice parlato e il codice scritto dipendono dalla pianificazione, dalla formalità della situazione ecc.

Ad esempio, un messaggio in una chat è considerato un testo scritto a causa del mezzo utilizzato per la sua produzione (la tastiera del cellulare o del computer), ma la sua natura di testo scritto non lo rende automaticamente una produzione formale. Quel messaggio può avere, ad esempio, caratteristiche del parlato e risultare meno formale di una produzione orale vera e propria prodotta in situazioni di maggiore ufficialità. Questo è possibile poiché il grado di pianificazione dei contenuti e di formalità dipendono da una varietà di fattori, ad esempio il tipo di rapporto esistente tra mittente e destinatario, i loro gruppi sociali di appartenenza, la loro provenienza geografica, eccetera. Per comprendere questo aspetto si immagini, ad esempio, come una stessa notizia di sport, cronaca o politica possa essere comunicata in uno scambio in chat informale tra due utenti con un rapporto di familiarità e come la stessa notizia possa venire annunciata oralmente in un contesto formale da un giornalista televisivo.

Questo tema della natura trasversale della diamesia viene ripreso da Berruto nel 1993: superando la sua stessa proposta del 1987, nel frattempo adottata dalla comunità scientifica italiana, di considerare la diamesia come un asse di variazione autonomo e aggiuntivo rispetto

agli altri precedentemente teorizzati da Coseriu, Berruto arriva a elaborare una nuova articolazione dello spazio di variazione che vede rientrare la diamesia nell'area della diafasia, considerandole inscindibili¹:

All'interno della dimensione diafasica si situa tuttavia un'importante altra sottodimensione di variazione. [...] Si tratta della differenziazione fra uso parlato e uso scritto della lingua, dipendente per molti aspetti dalle caratteristiche strutturali e di realizzazione che il mezzo o canale impone alla codificazione del messaggio linguistico; spesso ci si riferisce a questo genere di variazione, che a rigore si situerebbe all'interno della dimensione diafasica, come a una quarta dimensione fondamentale di variazione, la 'variazione diamesica'. (Berruto, 1993b, p. 93)

Da questo breve *excursus* è emerso come l'introduzione della dimensione diamesica nello spazio variazionale abbia avviato, fin dalla sua nascita, un lungo dibattito sulla sua definizione e sull'evoluzione di un dubbio relativo in particolare al suo rapporto con la diafasia. Le perplessità non riguardano tanto le differenze tra scritto e parlato, quanto come esse possano essere descritte su un asse di variazione, come è stato sottolineato prima da Holtus (1984) e poi da Radtke (1992). Una possibile proposta di definizione dell'asse diamesico è la seguente:

Poiché il repertorio [linguistico] nel suo complesso, ancor prima delle singole dimensioni di variazione, è definito come continuum pluridimensionale con addensamenti, anche la diamesia, una volta inserita in tale contesto, dovrà presentare caratteri adeguati. Il continuum diamesico non è orientato, perché le varietà non si dispongono da un estremo sociolinguisticamente 'alto' a uno 'basso'; né polarizzato, allo stesso modo

¹ Si vedano a questo proposito anche i successivi studi (Berruto, 2011b; Berruto e Cerruti, 2014, pp. 147-148).

in cui lo sono quello diafasico o diastatico. (Pistoiesi, 2015, p. 42)

Dunque, poiché l'uso scritto e l'uso dell'oralità non sono legati in maniera fissa alle variazioni diafasiche, ma possono essere impiegati in una grande varietà di realizzazioni della comunicazione con diversi gradi di formalità (vale l'esempio sopra citato della notizia di cronaca scritta tra amici in una chat o comunicata ufficialmente da un giornalista radiofonico/televisivo), le varietà diamesiche sono da considerarsi trasversali rispetto alle altre dimensioni, dalle quali non possono essere completamente isolate, se non per specifiche esigenze di studio (cfr. Berruto, 1993a e 1993b).

Per quanto riguarda le dimensioni di variazione, la loro combinazione definisce, come già accennato in precedenza, lo spazio linguistico italiano:

All'interno di questo spazio si sono individuate aree corrispondenti a varietà di italiano relativamente ben individuabili e definibili. Tra esse l'italiano standard, l'italiano "neostandard" (o "comune" o "medio"), l'italiano popolare e l'italiano regionale sono, per così dire, varietà complete, dotate di peculiarità a tutti i livelli. (Lorenzetti, 2002, p. 23)

A queste varietà si farà riferimento nelle sezioni successive.

1.3. Lo spazio linguistico italiano

Nelle seguenti sezioni si descriveranno le varietà dello spazio linguistico italiano. Il punto di partenza saranno i due poli estremi, ossia l'italiano standard e l'italiano popolare/dei semicolti. Successivamente la trattazione di sposterà sull'italiano regionale e, infine, sulla varietà linguistica a cui Berruto (1987) attribuisce la dicitura di italiano neostandard, a cui, dopo alcune premesse di carattere terminologico, verrà dato particolare rilievo coerentemente agli scopi di questa trattazione.

1.3.1. Italiano standard

Come già accennato precedentemente, il dibattito teorico sul concetto di standard linguistico è in corso almeno dal Trecento, cambiando nel corso dei secoli con il ciclico ritorno della cosiddetta *questione della lingua*. Ulrich Ammon (1986, 2003) identifica lo standard come:

1. codificato e descritto esplicitamente da un insieme di regole normative e da testi di riferimento (opere letterarie modello, grammatiche, dizionari);
2. sovraregionale, cioè non limitato ad una specifica area geografica;
3. elaborato, ossia possiede tutti i mezzi e le risorse linguistiche (vocabolario, strutture grammaticali e testuali, ecc.) per poter essere utilizzabile in tutti i domini, anche quelli specialistici;
4. in uso da parlanti appartenenti agli strati sociali superiori;
5. invariante, come conseguenza della codificazione normativa grammaticale;
6. utilizzabile in tutti gli usi della lingua scritta.

Tale definizione è successivamente ripresa da Lorenzetti (cfr. 2002, pp. 25-28), che aggiunge le seguenti caratteristiche in base alle quali, inoltre, lo standard è:

7. la varietà sociolinguistica normale, ossia la meno marcata;
8. la varietà di riferimento, che rappresenta la norma (e in quanto tale garantisce l'invarianza a cui fa riferimento Ammon);
9. la varietà utilizzabile in forma orale in tutti i contesti, da qualunque parlante per qualsiasi argomento.

La moderna questione della lingua, che si ripropone sin dal Cinquecento, fino ad Alessandro Manzoni (Koch, 2014), si interroga non solo su quale sia la varietà di riferimento, ma sulla possibilità che questa varietà esista. Il dibattito linguistico si è sviluppato secondo due correnti riassumibili come segue:

- a) l'italiano standard esiste e va identificato per la pronuncia con l'italiano fiorentino contemporaneo, privo di alcuni tratti prettamente locali, per la grammatica nell'italiano ottocentesco post-manzoniano;
- b) l'esistenza in sé di uno standard è dubbia². Nel caso specifico dell'i-

² In natura non esistono varietà di lingua standard, poiché esse sono il risultato di un'azione esplicita di codificazione di regole grammaticali e normazione. Lo

taliano standard, la sua grammatica di riferimento, presentata nel dettaglio in Serianni (1988), si può definire nel complesso come un “fiorentino emendato” (Galli de’ Paratesi, 1984, p. 57), quindi una varietà che non coincide con nessuna delle varietà effettivamente usate dalla comunità linguistica all’interno del territorio nazionale. Come precisato da Berruto (2010):

L’italiano standard in effetti non ha mai, fin dalla codificazione cinquecentesca, coinciso esattamente con il fiorentino, e sin dal Seicento ha accolto, data anche la mancanza fra il tardo Cinquecento e l’avanzato Ottocento di un centro preminente che imponesse una norma, innovazioni di varia provenienza. La distanza dal fiorentino si è ancora accresciuta dopo l’Unità d’Italia, nonostante i tentativi puristici di imporre il fiorentino moderno come modello, in particolare per la pronuncia³.

Rispetto a quest’ultimo punto, come riportato da Lorenzetti (2002, pp. 27-28), proprio nelle aree che identificano lo standard si rilevano tratti in regresso, rispetto ai quali si potrebbe affermare che lo standard non esiste dal punto di vista della pronuncia, poiché nessun parlante ha acquisito nativamente la pronuncia fiorentina emendata (es. la distinzione tra *e* e *o* aperte e chiuse e tra *s* e *z* sorde e sonore non si è mai imposta al di fuori della Toscana e del centro), che è di solito appresa tramite corsi di dizione da parte di parlanti professionisti (attori, presentatori radiofonici e televisivi). Un discorso simile vale per la grammatica, poiché la varietà codificata coincide per la maggior parte con l’italiano scritto formale, ma non riflette le varietà parlate,

standard si basa, dunque, inevitabilmente su un certo grado di artificialità (cfr. Berruto, 2010). La formazione di una varietà standard è, infatti, un processo sociale e culturale che deriva dall’azione congiunta di più fattori (Ammon, 2003): (a) parlanti e scriventi professionali, che producono testi modello; (b) autorità normative in ambito linguistico (per es., accademie) che sono la fonte di istruzioni e correzioni; (c) codici linguistici (i manuali descrittivo-normativi di riferimento); (d) esperti di lingua, che descrivono e giudicano le produzioni linguistiche.

³ Versione html, pertanto priva di pagine.

escludendo tratti morfosintattici effettivamente presenti nella lingua contemporanea.

Il prestigio della varietà standard normativa perde dunque terreno, riducendo la propria influenza allo scritto formale, mentre “lo spazio perduto dallo standard normativo viene guadagnato dalle altre varietà, che diventano così più accessibili, meno censurate, più capaci di espansione” (Lorenzetti, 2002, p. 28). Lo standard normativo esclude, infatti, tratti morfosintattici effettivamente presenti e funzionali nella lingua contemporanea e che hanno trovato gradualmente spazio nella varietà linguistica descritta da Sabatini (1985, 1990) come italiano dell'uso medio, a cui sarà dedicata la sezione 1.3.4.

La sezione successiva sarà dedicata alla varietà collocata sul polo opposto allo standard: l'italiano popolare o dei semicolti.

1.3.2. Italiano popolare (o “dei semicolti”)

L'italiano popolare, o “italiano dei semicolti” (Bruni, 1984; D'Achille, 1994), è la varietà, sia parlata che scritta, tipicamente utilizzata dalle fasce meno istruite della popolazione in alternativa al dialetto. È utilizzata da parlanti per i quali il dialetto è la lingua materna, e l'apprendimento dell'italiano è stato limitato ai pochi anni di istruzione scolastica ed è documentata in contesti quali la scrittura di diari, lettere indirizzate alle autorità pubbliche e lettere dal fronte durante la Prima e la Seconda guerra mondiale. Pur essendo una varietà sia scritta che parlata, prima della diffusione del telefono era una varietà per lo più scritta, il che non implica che gli utenti fossero in grado di dominare la correttezza formale della lingua, essendo questi, anzi, caratterizzati dal punto di vista sociolinguistico proprio dal basso grado di istruzione.

L'italiano popolare è il “modo di esprimersi di un incolto che, sotto la spinta di comunicare e senza addestramento, maneggia quella che ottimisticamente si chiama la lingua ‘nazionale’, l'italiano” (De Mauro, 1970, p. 49). In seguito, Manlio Cortelazzo descrisse l'italiano popolare come “il tipo di italiano imperfettamente acquisito da chi ha per madrelingua il dialetto” (Cortelazzo, 1972, p. 11). Si tratta, quindi, una varietà linguistica “di contatto”, poiché come l'italiano regionale (si veda la sezione seguente) deriva dall'incontro tra lingua e dialetto e quindi presenta fenomeni di interferenza linguistica. Si colloca in basso lungo l'asse diastratico (Berruto, 1993a e 2012 [1987]).

Dal punto di vista della sua diffusione, “Non ci sono dubbi che un it. pop. veramente unitario non esista. [...] La marcatezza diatopica in Italia è preliminare a tutte le altre dimensioni di variazione” (Berruto, 2012 [1987]: 130). Si tratta, dunque, di una varietà fortemente caratterizzata in diatopia, poiché presenta un’elevata frequenza di caratteri regionali.

Tra le caratteristiche si cita qui a titolo esemplificativo (cfr. D’Achille, 2003, pp. 206-207; Berruto 2011a) una forte tendenza alla semplificazione che investe sintassi, morfologia, fonologia e lessico dello standard. A questo proposito, la nozione di semplificazione è intesa in senso lato:

Con il termine di “semplificazione” si designano, al di là del valore corrente intuitivo del termine [...]: a) una riduzione all’essenziale [...] eventualmente misurabile in termini di (un minor) numero di morfemi e di (una minore) articolazione strutturale interna; b) il richiedere meno regole, vale a dire un minor numero di regole ed eventualmente un’applicazione diversa, con ridotto numero di passaggi, nel processo di generazione e trasformazione, dalla struttura profonda o forma logica alla struttura superficiale [...]; c) l’abbandono della lingua alla “naturalzza”, a tendenze evolutive latenti all’interno del sistema e miranti a una ‘ottimizzazione’ e a una minore marcatezza. (Berruto, 1983, p. 40)

Ad esempio: la semplificazione del paradigma dell’articolo con la riduzione del numero delle sue forme e l’estensione dell’uso di alcune di esse: *il zio, dei amici, il spettatore, i altri, un sbaglio*; la regolarizzazione di paradigmi verbali (*dassi* vs *dessi*, *venghino* vs *vengano*, *persuàdere* vs *persuadère*, *soddisfava* vs *soddisfaceva*, *vadi/vadino* al congiuntivo presente, *misimo* come prima plurale del passato remoto di *mettere*, *potiamo* come prima persona plurale del presente di *potere*, *bevavamo* come imperfetto di *bere* ecc.), la riduzione di un intero paradigma di pronomi di terza persona al solo *ci* (*dicci* al posto di *dì a noi*, *dì loro*, *dille*, *digli*), uso del possessivo di terza persona *suo* anche per il plurale (gli amici con il *suo* [= loro] cane) e altri fenomeni (cfr. Berruto, 2011a; Lorenzetti, 2002, pp. 33-34); alcune

realizzazioni che mostrano lo scarso controllo ortografico, quali l'assenza della percezione dei confini tra le parole (*lamico* vs *l'amico*), mancanza o utilizzo improprio dell'acca, delle doppie e degli accenti eccetera. La semplificazione riguarda anche il lessico, caratterizzato da termini generici e da un vocabolario ridotto.

Rispetto all'italiano dell'uso medio, a cui si farà riferimento successivamente, questa varietà è caratterizzata da una maggiore influenza dei dialetti e da una più elevata frequenza di elementi rifiutati dallo standard e da tratti linguistici idiosincratici. Sebbene si manifestino con frequenza diversa, condivide con il neostandard colloquiale alcune tendenze: ad esempio, la già citata tendenza a regolarizzare i paradigmi verbali è più frequente nell'italiano dei semicolti (le forme *venghi*, *venghino* sono esclusive dell'italiano dei semicolti, mentre *persuādere* e *soddisfava* si realizzano anche nell'italiano colloquiale). Al di là di queste lievi variazioni,

ciò che distingue più nettamente l'italiano popolare dall'italiano colloquiale e informale è lo status socio-linguistico: l'italiano colloquiale e italiano informale non sono vere e proprie varietà, ma piuttosto registri dell'italiano neostandard, che possono essere scelti dai parlanti per adattarsi a determinate situazioni. (Lorenzetti, 2002, p. 34)

La sezione successiva sarà dedicata all'italiano regionale e alle caratteristiche che lo distinguono dall'italiano popolare o dei semicolti.

1.3.3. Italiano regionale

L'italiano regionale è stato identificato tra le varietà dell'italiano da Giovan Battista Pellegrini (1960), che ha articolato il repertorio linguistico italiano in quattro varietà: italiano letterario, italiano regionale, dialetto regionale (o *koiné* dialettale) e dialetto locale.

L'italiano regionale è dunque una varietà parlata in una determinata area geografica, nata dal contatto con la lingua di prestigio (analogamente alla derivazione dei dialetti italiani dal latino volgare), molto influenzata dal dialetto e che si differenzia sia dallo standard sia dalle altre varietà regionali per la presenza di elementi locali:

Storicamente è l'esito dell'adattarsi alle singole realtà dialettali preesistenti – e che normalmente erano l'unico modo di comunicare – di una lingua 'piovuta dall'alto': dalla scuola, dalle scritture, dall'amministrazione, poi in tempi più recenti anche dal cinema, dalla radio, dalla TV. (Poggi Salani, 2010)⁴

Con la già citata diffusione dell'italiano come lingua nazionale, nelle diverse aree italiane si è verificato un rafforzamento delle varietà standard regionali che, pur essendo diatopicamente marcate (ossia, presentando caratteristiche che le differenziano dalle altre varietà geografiche), non sono percepite come scorrette e sono utilizzate anche dai parlanti più colti. Inoltre, l'italiano regionale include sia le varietà apprese direttamente come lingua materna, sia quelle sviluppate da dialettofoni nativi.

Per quanto riguarda il rapporto tra l'italiano popolare e italiano regionale, entrambi sono varietà locali dell'italiano; tuttavia, mentre l'italiano popolare è particolarmente influenzato dall'estrazione socio-economica e culturale dei parlanti (diastria), l'italiano regionale è una varietà legata unicamente alla provenienza geografica dei parlanti (diatopia): "la marcatezza regionale e quella sociale si sommano nel dare luogo [...] a tanti italiani regionali popolari. Un it. popolare specifico sarà sempre anche un italiano regionale" (Berruto, 2012 [1987], p. 132). Rispetto all'italiano standard, le diverse varietà di italiano regionale sono tutte varianti geografiche, mentre le diverse varietà di italiano popolare sono varianti sia geografiche che sociali (Mengaldo, 1994, pp. 108-110). Riassumendo, si può affermare, quindi, che l'italiano popolare è un sottoinsieme più marcato dell'italiano regionale.

Un tratto essenziale nella regionalità linguistica è l'intonazione, ossia l'insieme dei tratti soprasegmentali dell'enunciato. Come nota Lorenzetti (2002, p. 30), è rarissimo trovare pronunce del tutto prive di tratti regionali, anche presso parlanti colti e in situazioni formali. Nella pronuncia "l'italiano regionale colto coincide, di fatto, con l'italiano neostandard parlato, mentre la regionalità nella grammatica, nella sintassi e nella scelta delle parole è spesso controllabile (e dunque, se il

⁴ Versione html, pertanto priva di pagine.

caso, evitabile) da parte dei parlanti che abbiano un certo grado di istruzione” (ivi, p. 30).

Un accento troppo marcatamente regionale è invece ritenuto substandard (Berruto, 2011a). Inoltre, il grado di regionalità di un testo non è determinato solo dalla quantità di elementi regionali che presenta, ma anche dalla valutazione sociale attribuita ad essi (ossia, il fatto che alcuni tratti siano percepiti come un regionalismo più marcato). Il lessico rappresenta un fattore di forte diversificazione regionale: ad esempio, tutto il vocabolario della vita quotidiana è soggetto a interferenze dei dialetti, poiché molti parlanti danno una morfologia italiana a lessemi dialettali, creando così dei geosinonimi che variano di area in area, alcuni dei quali entrano nel repertorio nazionale. Sul piano morfologico e sintattico vi sono, invece, meno tratti largamente condivisi e linee di tendenza comuni a tutti gli italiani regionali. In tal senso è possibile affermare che vi sia uno scambio reciproco tutt'ora in corso nell'italiano standard e le varietà regionali.

Infine, la sezione successiva sarà dedicata a un'altra varietà, derivata da un processo di ristandardizzazione dell'italiano contemporaneo, che sarà al centro della presente trattazione.

1.3.4. Italiano dell'uso medio⁵

Come già affermato precedentemente nella sezione 1.1, la diffusione dell'italiano in sempre maggiori ambiti d'uso, fino a diventare la lingua nazionale, è un processo iniziato gradualmente con l'Unità d'Italia e che si è intensificato nella seconda metà del Novecento, grazie soprattutto all'istruzione obbligatoria, alla diffusione dei mezzi di comunicazione di massa e ai cambiamenti sociali, descritti da De Mauro (2011, 2014) e da Poggi Salani (2012), dovuti all'industrializzazione e alle migrazioni interne.

⁵ Sebbene nella descrizione delle varietà che compongono lo spazio linguistico italiano venga utilizzata la dicitura di italiano neostandard, coniata da Berruto nel 1987, in questa sede si è scelto di utilizzare la dicitura di “italiano dell'uso medio” quando la varietà linguistica viene affrontata da un punto di vista puramente descrittivo dei tratti linguistici che la compongono. La dicitura di “neostandard” verrà utilizzata in una fase successiva dell'argomentazione con il valore di standard rinnovato. Questa alternanza terminologica verrà adeguatamente affrontata nel corso della trattazione.

All'origine del bisogno comune di una lingua *passe-partout*, valida su tutto il territorio nazionale, vi sono vari elementi (De Mauro, 2011, p. 201; Poggi Salani, 2012, pp. 127-129). Innanzitutto, le migrazioni interne, dal Sud e dalle Isole verso il Centro-Nord più industrializzato, dagli anni '50 del Novecento in poi: la presenza di immigrati nelle nuove sedi ha creato la necessità di italianizzazione anche presso gli abitanti originari di quelle stesse sedi, dove il dialetto era risultato sufficiente solo finché la popolazione era stata geograficamente omogenea. Altri fattori sono i grandi mezzi di diffusione della lingua parlata, ossia la radio, il cinema sonoro dal 1930 e la televisione dal 1954, con spettacoli e trasmissioni di massa, ma anche i dischi e il telefono, l'obbligo scolastico a 14 anni (legge 31 dic. 1962), il servizio militare obbligatorio e, infine, la facilità di spostamenti e scambi sul territorio nazionale al di là della migrazione, che includono anche il servizio militare obbligatorio, che ha contribuito allo spostamento di una fetta della popolazione verso regioni distanti dal proprio luogo d'origine.

In questa fase di evoluzione, la varietà standard, fino ad allora utilizzata per lo più nello scritto formale, dovette adattarsi a tante esigenze pratiche della comunicazione nella vita quotidiana. Di conseguenza, hanno incominciato, così, a emergere nell'uso quotidiano della lingua dei tratti linguistici già esistenti da secoli, ma sino ad allora sanzionati dalla grammatica tradizionale, che "via via "risalgono" dai livelli inferiori (sub-standard)" (Sobrero, 1992a), entrando nella nuova varietà diffusa a livello nazionale.

Il substandard a cui fa riferimento Sobrero (1992a) è composto da quegli elementi che, pur essendo ampiamente usati dai parlanti, sono esclusi dalla norma riconosciuta e codificata, sono principalmente tipici dell'uso parlato informale spontaneo e rientrano nelle varietà diastratiche e diafasiche basse, nell'italiano regionale marcato e nell'italiano popolare.

Il substandard non ha prestigio nella comunità linguistica ed è in genere oggetto di giudizi negativi, quali "cattivo italiano", "non è corretto", "sbagliato" e "dialettale" (Galli de' Paratesi, 1984, pp. 143-205), che rispecchiano la sua sanzione come errore nella scuola. Non essendo soggetto alle norme che inquadrano la lingua standard, il substandard è molto variabile, poiché in esso trovano espressione la naturalezza, le strategie dei parlanti finalizzate alla semplificazione, ad esempio in

quelle aree della lingua in cui lo standard presenta irregolarità, eccezioni e paradigmi complessi, soprattutto nella morfologia, e strategie di adattamento pragmatico e funzionale al contesto basate sull'utilizzo di strutture sintattiche influenzate dalla situazione comunicativa e dai partecipanti all'interazione, ad esempio un lessico più generico, collegato all'uso della deissi (si vedano le già citate sezioni dedicate all'italiano regionale e popolare; cfr. Berruto, 1990, 2011b; Simone, 1991). Per tornare alla citazione di Sobrero (1992a), è per queste ragioni che spesso il substandard contiene in sé quelli che potrebbero essere dei potenziali mutamenti linguistici e può svelare tendenze e caratteristiche del sistema linguistico che non emergono attraverso lo standard, ma che in alcuni casi "risalgono" dai livelli inferiori. Il processo appena descritto è comune a tutte le tradizioni linguistiche, dove accanto alla norma tende a formarsi una varietà caratterizzata da maggiore flessibilità, tipica dei registri informali del parlato, che accetta una serie di tratti un tempo sanzionati.

In ambito italiano a questa varietà Sabatini (1985) ha assegnato il nome di *italiano dell'uso medio*, denominazione che indica la media formalità diafasica sia del parlato che dello scritto in cui è accettato, mentre Berruto ha parlato di *neostandard*, attribuendo questa denominazione a quell'insieme di

costrutti, forme e realizzazioni che non erano presentati nel canone ammesso dalle grammatiche e dai manuali, o che, quando vi erano menzionati, lo venivano in quanto costrutti, forme o realizzazioni del linguaggio popolare o familiare o volgare, oppure regionali, e quindi da evitare nel bel parlare o scrivere. (Berruto, 2012 [1987], p. 73)

L'affermarsi di un "italiano dell'uso medio" è un segnale inequivocabile dell'inadeguatezza sul piano funzionale dello standard tradizionale, tramandato dalle grammatiche. Tuttavia, è necessario chiedersi se le forme linguistiche incluse in questa varietà diano tutte un effettivo apporto funzionale alla lingua, giustificando così il distacco dalla norma grammaticale tradizionale. In altre parole, se utilizzando queste forme linguistiche (e quindi non rispettando lo standard), i parlanti

aggiungano effettivamente alla comunicazione qualcosa che lo standard tradizionale non metteva loro a disposizione o se provochino unicamente un abbassamento del registro linguistico che li espone alla censura sociale.

Uno standard condiviso dalla comunità dei parlanti dovrebbe, infatti, dare sicurezza a chi parla e criteri di valutazione affidabili ai destinatari. Mentre la presenza di varianti regionali e di registro, precedentemente descritte, non rappresenta un ostacolo per l'insegnamento dell'italiano L2, la questione dell'italiano dell'uso medio rappresenta, invece, una questione aperta. La tendenza all'inclusione di tratti linguistici ampiamente diffusi, seppur censurati dalla norma, rende l'italiano dell'uso medio una varietà certamente più aperta rispetto allo standard alle spinte funzionali provenienti dall'allargamento dell'uso dell'italiano come lingua nazionale. Tuttavia, allo stesso tempo, questa sua peculiarità lascia intravedere un problema irrisolto: fino a che punto e entro quali limiti l'italiano dell'uso medio, oltre a costituire una varietà più inclusiva dello standard tradizionale, fornisce anche delle linee guida sicure, che garantiscono la correttezza d'uso sia a chi usa la lingua, sia a chi ne valuta la produzione?

Lasciando aperta la domanda, che sarà affrontata successivamente, nella sezione successiva verranno passati in rassegna i tratti che Sabatini nei suoi studi del 1985 e del 1990 identifica come i tratti caratterizzanti dell'italiano dell'uso medio.

1.4. L'italiano dell'uso medio: verso un nuovo standard?

Come anticipato precedentemente, Sabatini aveva in un primo tempo (1985) descritto l'*italiano dell'uso medio*, definendolo come una varietà linguistica diffusa su tutto il territorio nazionale, tipica dell'uso parlato e scritto di media formalità. Le sue caratteristiche linguistiche sono identificabili in una lista di 35 tratti fonologici, morfologici e sintattici diffusi su tutto il territorio italiano, "usati da persone di ogni ceto e ogni livello di istruzione; [...] [che] non sono limitati al discorso "orale-non pianificato", ma risultano pienamente funzionali anche per un discorso "scritto-pianificato", purché non decisamente formale" (Sabatini 1985, p. 171) e che rappresentano la diffusione e "l'accettazione, nell'uso parlato e scritto di media formalità, di un tipo di lingua che si differenzia dallo "standard" ufficiale" (ivi, p. 155).

Questi fenomeni linguistici sono ritenuti sufficienti da Sabatini per identificare una varietà della lingua nazionale “a cui si addice – meglio di altre denominazioni più riduttive o più vaghe, come “italiano colloquiale” o “colloquiale-informale” – il nome di “italiano dell’uso medio parlato e scritto”, e, per brevità, ITALIANO DELL’USO MEDIO” (ivi, p. 171).

Negli studi di Sabatini l’approccio rispetto allo statuto dell’italiano dell’uso medio oscilla tra un atteggiamento puramente descrittivo – ossia, il linguista prende atto di certi fenomeni documentati dall’uso – e l’idea che gli usi attestati rappresentino un’alternativa allo standard tradizionale, che si evidenzia quando scrive che l’italiano dell’uso medio “si candida ad occupare, dopo secoli di ostracismo, il baricentro dell’intero sistema linguistico italiano o per lo meno a condividere con lo standard ufficiale il crisma della norma” (ivi, p. 175).

Successivamente, in un suo studio del 1990, Sabatini rielaborò la lista, riducendola ai 14 tratti linguistici⁶ di seguito descritti.

1) *Lui, lei, loro* in funzione di soggetto.

Sabatini (1985, p. 159) a questo proposito scrive che “l’uso di *egli, ella, essa, essi, esse* è ristretto al parlato che possiamo definire “celebrativo” e alle scritture di tipo argomentativo e situazionale”, mentre *lui, lei, loro* in funzione di soggetto “sono ormai la norma in ogni tipo di parlato, anche formale, e nelle scritture che rispecchiano atti comunicativi reali”. Tali forme in funzione di soggetto hanno dei precedenti rintracciabili nel fiorentino del XIII secolo, si attestano con una certa frequenza nel XIV secolo, affermandosi maggiormente nel secolo successivo. “La sostituzione di *egli* con *lui* è avvenuta attraverso una lunga fase di convivenza, cioè di uso alternativo, anche se non indifferente, delle due forme: tra queste, *egli* teneva la posizione stilistica alta, *lui* quella bassa” (Renzi, 2012, p. 28)⁷. A questo proposito sempre Renzi:

⁶ Le denominazioni di ciascun fenomeno grammaticale sono tratte dalla lista del 1990. Allo scopo di facilitarne la comprensione, in alcuni casi alla denominazione originale è stata aggiunta una formulazione semplificata, che sarà poi utilizzata nel prosieguo di questo lavoro per ragioni di brevità e maggiore chiarezza. Per un confronto tra la lista del 1990 e la precedente del 1985, si rinvia a Zingaro (2021, pp. 52-53).

⁷ A questo proposito Renzi (2012, pp. 26-27) rimarca quanto segue: “Il cambiamento linguistico avviene attraverso alcune fasi. Poniamo che una lingua possieda

La concorrenza con *egli* dura da circa sei secoli: al tempo di Dante un *lui* soggetto al posto di *egli* sarebbe stato inconcepibile, nel Quattrocento la *Grammatica* di Leon Battista Alberti (1440 circa) lo propone invece come forma grammaticale, il che vuol dire che si usava correntemente a Firenze almeno già da qualche tempo. Nel 1525 Pietro Bembo nelle *Prose della volgar lingua* (III, 16) riesce a farlo regredire, almeno nell'uso colto, per un bel po'. (ivi, p. 114)

L'uso di *lui*, *lei*, *loro* come pronomi soggetto fu, infatti, condannato da Bembo, che raccomandava l'uso delle forme antiche dei pronomi, e ciò influenzò notevolmente la produzione di gran parte degli scrittori dell'epoca, ma non di tutti. Essendosi ormai affermate nel parlato, queste forme non scomparvero mai del tutto dal sistema linguistico, mostrando, invece, dal XVII secolo una graduale ascesa nella narrativa, che raggiunse il suo culmine con la revisione de *I Promessi Sposi*. In quest'opera Manzoni scelse o il soggetto sottinteso, conformemente alla tendenza moderna ad uno scarso utilizzo di pronomi soggetto, o la sostituzione delle forme soggetto con le forme oblique (cfr. Seriani, 1989a, pp. 190-192). I grammatici mantennero in proposito comunque un atteggiamento cauto, che cambiò nel corso del tempo.

L'avanzata di *lui* è avvenuta attraverso tappe successive, cioè occupando diverse posizioni sintattiche, da quella predicativa (per es. *Io non sono lui*), nella quale già il Bembo ammetteva la legittimità della forma obliqua, al pronome in isolamento (per es. *Chi è stato? Lui*), alla

due registri, uno alto, accurato e uno basso, spontaneo [...]. Le fasi di cambiamento linguistico possono essere rappresentate così: 1) nel contesto in cui si trova A, comincia a penetrare anche la nuova forma B, originariamente usata solo in un contesto diverso; 2) come conseguenza di 1), adesso A e B coesistono e si fanno concorrenza tra di loro, ponendosi generalmente a due livelli di registro diversi: la vecchia forma A si situa nel registro alto, la nuova forma B in quello più corrente, basso; 3a) A si indebolisce, e dopo un certo periodo di esistenza precaria, speso nello strato iperletterario della lingua, scompare; 3b) nella concorrenza tra A e B, A resiste allo sfidante B: quest'ultima si indebolisce e scompare”.

combinazione con un altro pronome o nome (per es. *Io e lui*, *Gianni e lui*) fino a raggiungere la posizione detta canonica, quella in cui il pronome precede direttamente il verbo (per es. *Lui dice che...*), che è l'ultima posizione a essere conquistata. (Renzi, 2012, p. 28)

Nella *Grande Grammatica Italiana di Consultazione* (Renzi, Salvi, Cardinaletti, 2001-2) si presentano come forme principali *lui*, *lei*, *loro* e le altre come forme alternative, definendole “pronomi rari nella lingua parlata, impiegati prevalentemente in alcune varietà della lingua scritta (scientifica, burocratica, letteraria) o in varietà della lingua parlata stilisticamente alte” (Cordin, 2001, I, p. 549). Inoltre, si mette in evidenza anche come l'uso di *lui*, *lei*, *loro* non risponda solo a criteri stilistici, di registro o normativi, ma in molti casi si tratti di una scelta obbligata sintatticamente: “I pronomi *egli*, *ella*, *esso*, *essa*, *essi*, *esse* pur non essendo clitici, hanno proprietà particolari che li distinguono dagli altri pronomi liberi” (ivi, p. 550), poiché non possono essere usati nei seguenti casi: in coordinazione con altri sintagmi nominali: **Egli e Mario/Mario e *egli vanno a Roma domani*; in isolamento, senza l'esplicitazione del verbo di cui sono soggetto: *Chi viene? *Essi*; se tra il pronome personale e verbo si frappongono altri elementi lessicali: **Essi, che sono arrivati per ultimi, non vogliono partire subito*; in posizione postverbale: *È arrivato *egli* e in funzione enfatica: **Egli è partito, non Piero; Proprio/solo *egli mi ha tradito*. Ciò trova conferma nello studio di Renzi (2012) e in alcune grammatiche per l'università e per la scuola secondaria superiore, citate a seguire a titolo di esempio, quali Salvi e Vanelli (2004), Fogliato (2015), Antonelli e Picchiorri (2016) e Serianni, Della Valle e Patota (2016).

In aggiunta a quanto esposto in precedenza (cfr. Cordin, 2001), Fogliato (2015) precisa l'obbligatorietà dell'uso di *lui/lei/loro* quando si vuole mettere in evidenza il soggetto (*questo lo sostiene lui*), con gli avverbi e le congiunzioni *anche*, *proprio* ecc. (*neanche lui sa qualcosa, nemmeno lei si vestirà elegante*), con il verbo al gerundio o al participio (*arrivato lui, ci saremo tutti; parlando lei, tutti pendevano dalle sue labbra*). Antonelli e Picchiorri (2016), invece, oltre a ribadire la caduta in disuso delle forme soggetto, sottolineano come però molto spesso sia preferibile omettere del tutto il pronome (Ø) o sostituirlo con un

nome (Esempi: *Manzoni lavorò alla seconda edizione dei Promessi sposi per molti anni. In questo periodo, egli si dedicò al romanzo quotidianamente* → *Manzoni lavorò alla seconda edizione dei Promessi sposi per molti anni. In questo periodo, (Ø) si dedicò al romanzo quotidianamente.* → *Manzoni lavorò alla seconda edizione dei Promessi sposi per molti anni. In questo periodo, lo scrittore...*). Ciò lascia intendere come l'eredità delle forme soggetto originarie sia stata raccolta non solo dalle forme oblique, ma anche dal pronome sottinteso.

Sulla base di quanto esposto, poiché l'uso dei pronomi *lui/lei/loro* in funzione di soggetto è ormai indiscusso nell'italiano contemporaneo e, dunque, non sussiste più alcun dubbio sulla reperibilità di tali forme nello scritto di media formalità/informalità, il fenomeno è stato escluso dalla ricerca basata su corpus dei tratti elencati da Sabatini (1990) (cfr. capitolo 5).

2) *Gli* come forma dativale onnivale, ossia l'impiego di *gli* nella forma dativale al posto di *le* (a lei) e *loro* (a loro).

L'uso generalizzato del pronome *gli*, per entrambi i generi al posto delle altre forme pronominali di terza persona singolare e plurale è definito da Sabatini "accettabile nell'uso medio, soprattutto parlato, ma anche scritto. S'incontra spesso nei giornali e nelle riviste; è quasi normale nella narrativa" (1985, p. 158). A giustificazione di quest'uso generalizzato, Sabatini ne richiama le origini storiche. Questa forma pronominale, originariamente *li*, deriva dalle forme del dativo latino singolare e plurale (*ILLI* e *ILLIS*), valide per tutti e tre i generi e questo uso indifferenziato per genere e numero rimase attivo nell'intero sistema linguistico italo-romanzo (ossia, dialetti e lingua letteraria) come pronome atono e, dunque, utilizzabile come clitico nelle varie posizioni. "*Li* era originariamente anche la forma del f., essendo *ILLI* l'unica forma sg. Ma accanto a *li* si è sviluppato il clitico *le* da una forma latina analogica, *ILLAE*" (Vanelli, 1999, p. 115). Oltre alla restrizione dell'ambito d'uso con l'emergere di *le*, a partire dal XII secolo si affiancò l'uso di *loro*, ritenuto un gallicismo, avente originariamente un uso sia tonico (come nell'italiano moderno, in cui il suo uso senza preposizione è quasi esclusivamente postverbale, esempio: *dico loro*) che atono proclitico. Con il passare del tempo questo uso proclitico è stato abbandonato, per cui l'unica forma clitica rimase *gli*, nel suo

uso generalizzato per il maschile e il femminile fino al XVI secolo. Successivamente Manzoni ripristinò *gli* per il plurale maschile al posto di *loro*, ma non per il femminile singolare e plurale. Da ciò Sabatini suggerisce che l'uso di *gli* nel parlato come forma unica corrisponda al riemergere del suo antico uso generalizzato. Serianni nella sua grammatica (1989b, p. 249) afferma che

un punto molto delicato è costituito dall'uso della forma dativale di 6a persona: *gli* o *loro*? L'atono *gli* [...] – largamente attestato in tutti i secoli della nostra lingua – appartiene al registro familiare; il parlato formale e la massima parte dello scritto (tecnico-scientifico, letterario e in una certa misura anche giornalistico) preferiscono la forma *loro*.

Per quanto riguarda l'uso di *gli* per *le*, Serianni lo segnala invece come “decisamente da evitare anche nel parlato colloquiale” (*ibidem*) a prescindere dai precedenti letterari illustri da lui stesso menzionati (Boccaccio, Machiavelli, Carducci, Verga e Moravia). Tornando sul tema anni dopo, Serianni ha rilevato che *loro* è una forma “ormai minoritaria anche nello scritto” (2006, p. 12), ma che “la distinzione *gli* maschile / *le* femminile si è addirittura consolidata nel corso degli ultimi decenni” (ivi, p. 24), costituendo, quindi, un’“evoluzione apparentemente bloccata nonostante la prevedibilità del percorso” (*ibidem*), ossia nonostante la possibilità che *gli* potesse arrivare a sostituire anche la forma *le*, come avvenuto per *loro*.

Sulla base di ciò, si può giungere alla conclusione che l'affermazione dell'uso di *gli* al posto di *loro* come forma unica per il plurale sia avvenuta anche sulla base di motivazioni sintattiche: contrariamente a *loro*, *gli* presenta il vantaggio di fungere sia da proclitico (*gli dico*) che da enclitico (*digli*), caratteristiche condivise con forma *le* (esempi: *le dico*; *dille*). Dunque, è possibile ipotizzare che l'uso di *gli* al posto di *le* sia ancora percepito come tratto substandard (come confermato, ad esempio, dallo studio di Ondelli e Romanini (2018)⁸, a cui si farà rife-

⁸ Nel loro studio hanno sottoposto ad un campione di 150 persone, provenienti da tutto il territorio nazionale e prive di competenze specifiche nelle questioni

rimento in seguito), proprio perché non vi è alcun vantaggio sintattico nella sostituzione di *le* con un'altra forma pronominale avente le stesse caratteristiche e dunque si tratterebbe di una deviazione del tutto arbitraria rispetto alla norma.

3) La combinazione di una preposizione e di un articolo partitivo, ossia l'impiego del partitivo preceduto da preposizione (*con degli amici*), come alternativa agli aggettivi indefiniti (*alcuni, taluni, certi*) o espressioni composte: *con olio o con un po' d'olio; con amici o con alcuni (o certi) amici; su un gruppo di passanti*" (Sabatini, 1985, pp. 159-160). Circa questo tratto grammaticale, si riporta la descrizione fatta da Sabatini nella prima versione della lista del 1985:

Sconsigliata dalle grammatiche, è in realtà di uso frequentissimo proprio nella lingua media (non in quella molto formale, né in quella regionale o popolare): Condiscilo con dell'olio crudo; È andato in gita con degli amici che non conosco. [...] Nella varietà più alta e più bassa s'incontrano altre soluzioni, come l'eliminazione del partitivo o la sua sostituzione con altri tipi (alcuni, taluni, certi, espressioni composte): con olio o con un po' d'olio; con amici o con alcuni (o certi) amici; su un gruppo di passanti. (Sabatini, 1985, pp. 159-160)

4) La dislocazione dell'elemento tematico a sinistra o a destra della parte rematica della frase, con la ripresa funzionale atona, ossia l'impiego di dislocazioni a sinistra (*quel libro l'ho già letto*), a destra (*l'ho già letto, quel libro*) della parte rematica della frase con la ripresa pronominale atona, che può giungere fino a costruzioni con uso pleonastico

linguistiche, dei brevi estratti di articoli tratti da *Corriere della Sera* e da *Repubblica* (2015-2016), contenenti alcuni dei tratti grammaticali elencati da Sabatini (1985) e Berruto (1987) nella definizione dell'italiano dell'uso medio/ neostandard. L'obiettivo di tale ricerca era di indagare il grado di accettabilità percepita dai rispondenti rispetto a tali tratti. A questo proposito, l'uso del pronome dativo maschile *gli* in sostituzione del femminile *le* è generalmente rifiutato dai partecipanti.

della particella pronominale, come a me mi piace leggere, di pane non ne ho più (ossia, il doppio partitivo), e all'anacoluto: per esempio, quel libro non ne voglio sapere.

Queste strutture rientrano nella classe che i linguisti definiscono “costruzioni sintatticamente marcate”, cioè quelle frasi in cui l'ordine delle parole non segue la sequenza prevista dal sistema linguistico di riferimento, che nel caso dell'italiano è la sequenza SVO, ossia soggetto – verbo – oggetto diretto – altri sintagmi, detta anche “ordine non marcato” (cfr. Gossen, 1954; Berruto, 1985b; Benincà, Salvi, Frison, 1988; Berretta, 1995 e Ferrari, 2003) e che sono un tratto caratteristico dell'oralità (Renzi, 2012, p. 41).

Nella versione del 1985 della lista Sabatini definisce questo costrutto come “frase segmentata, cioè con tematizzazione (a sinistra o a destra) del dato “noto” assunto come “tema”, e ripresa di esso mediante un pronome nella frase che predica l’“informazione nuova”, cioè il “rema”” (ivi, p. 162). La sua funzione è l'enfasi, ossia la focalizzazione dell'informazione che rappresenta il dato “nuovo” nell'atto comunicativo (ivi, p. 161).

Quando un costituente della frase viene spostato a sinistra rispetto alla sua posizione non marcata, si parla di “dislocazione a sinistra”⁹. Il costituente dislocato presenta un legame con il resto della frase tramite una ripresa pronominale (*Le donne, io proprio non le capisco*, in Cignetti,

⁹ La dislocazione a sinistra rientra tra le “costruzioni sintatticamente marcate a sinistra” (cfr. Ferrari, 2003, pp. 145 e sgg.) ossia quelle costruzioni in cui un costituente della frase, che normalmente avrebbe una posizione postverbale, occupa invece una posizione preverbale. Tra queste vi sono, oltre alla già citata dislocazione, anche le costruzioni con tema sospeso come l'anacoluto (esempio: *Elisabetta, non penso più a lei da tempo*), ove tra l'elemento marcato a sinistra e il resto dell'enunciato non vi è un legame sintatticamente esplicito, e la topicalizzazione (esempio: *MIA CUGINA ho visto [non mia sorella]*), che pone l'enfasi (intonativa, nel parlato) sull'elemento dislocato (gli esempi sono tratti da Cignetti, 2006, pp. 207 e sgg.). Benincà, Salvi e Frisson (1988, pp. 141-143) aggiungono anche “l'anteposizione anaforica”, che si verifica quando il costituente dislocato non è né ripreso da un pronome, né enfatizzato (esempio: *la stessa proposta fece poi il partito di maggioranza*). Tra le costruzioni marcate vanno inserite anche le frasi scisse (cfr. il punto 13 della lista di Sabatini, 1990 e Roggia, 2006) e le frasi presentative (De Cesare, 2006), ad esempio: *c'è una ragazza che ha scritto un annuncio*.

2006, p. 208, vs *io proprio non capisco le donne*, secondo l'ordine non marcato) e/o tramite la presenza di una preposizione, nel caso dell'oggetto indiretto (esempio: *A Elisabetta, non parlo da secoli, ibidem*).

In generale, è utilizzata con l'intento di creare nel testo una gerarchia delle informazioni, realizzata attraverso l'ordine marcato dei costituenti della frase. Lambrecht (1994) individua gli ambiti d'uso principali della dislocazione a sinistra: presentare un costituente della frase come l'elemento su cui verte il contenuto dell'enunciato, imporre un nuovo tema, mettendo in risalto un costituente e attirando l'attenzione dell'interlocutore su di esso e, infine, a risolvere casi di ambiguità, quando vi è più di un costituente che potrebbe fungere da punto focale del discorso (si veda anche Berruto, 1985b). In aggiunta, la dislocazione a sinistra può essere utilizzata come strumento di coesione testuale, in particolare quando è abbinato a espressioni anaforiche, come i dimostrativi: "Segue e cura – fin dai primi giorni – l'educazione della primogenita [...]. E di questa esperienza stende un resoconto [...]" (Soletti, in Cignetti, 2006, p. 212).

La frequenza d'uso nella lingua scritta varia a seconda del genere testuale: rare nei testi specialistici (Mortara Garavelli, 2001, p. 86 e sgg.), sono molto diffuse nella prosa giornalistica (Ferrari, 2003, p. 148), nei testi pubblicitari (*Oil of Olaz. L'ho scelto da sempre*, in Cignetti, 2006, p. 212). In aggiunta, oltre a svolgere le funzioni di tematizzazione già menzionate, la dislocazione a sinistra sta ormai sostituendo la forma passiva, che è percepita dai parlanti come di registro troppo elevato (Renzi, 2012, p. 42).

Nel caso, invece, in cui un costituente della frase è dislocato a destra rispetto alla sua posizione nell'ordine non marcato della frase si parla di "frasi dislocate a destra" (della vasta bibliografia sul tema si citano: Sornicola, 1981; Benincà, Salvi, Frisson, 1988; Cresti, 2000; Ferrari, 2003). Se l'elemento dislocato è un oggetto diretto, è presente normalmente un pronome clitico che anticipa l'elemento dislocato (esempio: *io proprio non le capisco, le donne*). Questa struttura serve a introdurre informazioni aggiuntive con un tono, per così dire, "confidenziale" rispetto al contenuto dell'elemento dislocato, che è qualcosa su cui verte già il discorso.

Per quanto riguarda gli ambiti d'uso, la dislocazione a destra ha "funzioni meno informativo-coesive, per dir così, e più emotivo-espressive,

quali l'isolamento di un tema in funzione di appendice (*detachment* o *extra-position*: Ferrari, 1999); il ripensamento (*afterthought* o *self-repair*: Tomlin, 1986; Berretta, 1986, 1992); ammiccamento o confidenzialità (*camaraderie*) tra interlocutori (Berruto, 1986; Ferrari, 2008)” (Rossi, 2016, p. 847).

In particolare, dallo studio di Rossi (2016), che prende in esame sia la prosa giornalistica, sia la narrativa, esaminando rispettivamente gli articoli presenti nel sito www.repubblica.it dal 10/1/2014 all'indietro e i 100 romanzi vincitori e finalisti del Premio Strega dal 1947 al 2006, emerge che gli ambiti d'uso della dislocazione a destra sono molto più variegati di quelli finora attribuiti a questa struttura negli studi precedenti:

- a) usi grammaticalizzati o in via di grammaticalizzazione (ossia, usi ormai integrati o in via di integrazione nella norma e, dunque, non più percepiti come forme marcate. Ad esempio, i costrutti completivi espliciti dislocati come *lo sol/sai/sa che* e *lo vedi che* sono le forme marcate dei costrutti completivi *sol/sai/sa che* e *vedi che*; il costrutto *chi lo sa se* vs *chi sa se*, ecc.);
- b) autocorrezione e glossa esplicativa (Rossi, 1999, 2005);
- c) ancoramento deittico (ossia, accentuazione della forza deittica degli enunciati. Esempio: *Ma la vedi lí, la porta? aveva aggiunto indicando con un gesto veloce l'uscita verso il ballatoio. La vedi? Apri, esci, chiudi e non tornar più.* (Testori in Rossi, 2016, p. 853);
- d) ancoramento emozionale (ripetizione) (si veda l'esempio precedente);
- e) modalità interrogativa (coinvolgimento dell'interlocutore, ammiccamento) (si vedano l'esempio precedente e il seguente: *Ecco, lo vedi a cosa serve la tua superbia? Lo vedi?*, Testori, in *ivi*, p. 854);
- f) messa in evidenza del centro di interesse comunicativo ed emotivo: a questo proposito vengono citati vari esempi tratti dalla pubblicità, tra cui *Noi, lo facciamo così / da cinquant'anni* pubblicità del 2013 di un panificio di Messina, ove l'elemento dislocato a destra (*noi, lo facciamo così, [il pane]*) è di fatto dislocato iconicamente nell'immagine sul cartellone;
- g) coesione, una funzione, ritenuta negli studi precedenti a quello di Rossi esclusiva della dislocazione a sinistra (esempio: *Io non lo vedo quasi mai il mio papà, perché egli lavora tutta la settimana a Torino*

e viene a trovarmi solo la domenica. Mio papà è un uomo molto alto e molto ordinato; e quando arriva, la domenica, mi deve sgridare perché se fossi da solo non mi laverei mai (Romano, in *ivi*, p. 855).

Sebbene la dislocazione a destra sia “più diffusamente avvertita come struttura ridondante, e dunque sanzionata negativamente dal sistema scolastico” (Ondelli, Romanini, 2018, p. 193), dallo studio di Rossi (2016, pp. 850-852) emerge che la narrativa è molto più incline alla dislocazione a destra, mentre invece la prosa giornalistica

preferisce la [dislocazione a sinistra] e soprattutto la frase scissa, in quanto tecniche predilette per il dinamismo comunicativo e di introduzione o riproposizione di nuovi temi nel discorso (Bonomi, 2002, pp. 212-215 e 321-323; Gatta, 2014, p. 339). Fatte queste precisazioni, sul computo totale la [dislocazione a destra] è oltre 2/3 della [dislocazione a sinistra], dunque la percentuale risulta invertita rispetto ai dati raccolti finora da Berruto [1985b, 1986] e altri.

Come già anticipato, Sabatini (1990, p. 97) inserisce all'interno della tipologia delle dislocazioni anche le strutture come *a me mi piace leggere*, *di pane non ne ho più* (ossia, il doppio partitivo) e l'anacoluto¹⁰:

- a) Le strutture come *a me mi* si basano sulla ripetizione della forma pronominale in forma tonica (*a me*) e atona (*mi*). In queste il

¹⁰ Entrando nel dettaglio, nella versione del 1985 Sabatini riunisce sotto un'unica categoria quattro procedimenti di natura sintattica che contribuiscono al “fenomeno dell’“enfasi” tramite la tematizzazione dell’elemento noto a destra o a sinistra (p. 161): la frase scissa (cfr. punto 13 dell’elenco del 1990), la posposizione del soggetto al predicato (punto 11), la frase segmentata, ossia la dislocazione a destra e a sinistra, includendo anche le costruzioni come *a me mi* e il doppio partitivo (punto 4), e infine l’anacoluto (punto 4). Il motivo di questo accorpamento è dato dalla condanna generalizzata da parte delle grammatiche tradizionali delle costruzioni con ripresa pronominale del tema, definite “pleonastiche”, e delle strutture con inizio “è a lui che”, definite come “francesismi”, una posizione che, di fatto, distoglie l’attenzione di chi insegna la lingua dal fenomeno della messa in prospettiva comunicativa (*ibidem*). Sulla condanna generalizzata delle grammatiche rispetto a questi costrutti, si veda anche Prandi (2020, pp. 174-175).

primo elemento (*a me*) equivale a *per quanto riguarda me* e rappresenta il tema sul quale poi si svolge il discorso o rema (*mi piace*). Nella presente trattazione saranno classificate all'interno del fenomeno delle dislocazioni, nello specifico come dislocazioni a sinistra dell'elemento pronominale e saranno oggetto di ricerca tutte le forme pronominali toniche del sistema pronominale, precedute dalla preposizione *a* e seguite dalla loro forma atona (*a me mi, a te ti, a lui gli...* ecc.) (cfr. capitolo 4).

- b) Il fenomeno del doppio partitivo (esempio: *di pane non ne ho più*) rientra anch'esso nella categoria della dislocazione a sinistra e consiste nell'anticipazione di un sintagma preposizionale, la cui testa è costituita dalla preposizione *di* (esempio: *di pane*), successivamente ripreso attraverso la particella pronominale *ne*. Sabatini (1985: 163) segnala la ricorrenza di questo fenomeno con alcuni verbi, soprattutto *fare, parlare, sapere, pensare, occuparsi, preoccuparsi, interessarsi, essere convinto* (*Di lui non me ne parlare; Di questa faccenda non ne voglio più sapere niente*).
- c) L'anacoluto (esempio: *Giorgio, non gli ho detto nulla*), dal greco *anakóluthon* [*schêma*], "privo di un seguito" è detto anche tema sospeso. È un sintagma nominale non integrato, destinato a ricevere una posizione e un ruolo con la ripresa pronominale. Qui il tema è solo un'enunciazione, un *nominativus pendens* che non ha un raccordo sintattico ma solo semantico con il rema, di cui, come riportato da Sabatini (*ibidem*), sono rintracciabili esempi anche in Dante.

5) Il *che* polivalente, sia con valore temporale, sia con altri valori (spaziali, personali).

Sabatini (1985, p. 164) nota come il *che*, originariamente pronome relativo con il significato di *di cui, in cui, a cui*, si sia poi trasformato in un connettivo generico con molte funzioni, il cui uso "viene accolto con maggiore larghezza" (*ibidem*, cfr. Fiorentino, 2010):

[Mentre] nell'italiano standard la congiunzione *che* è tipica di alcune frasi subordinate, quali le oggettive, le soggettive, le dichiarative e le relative (nei casi in cui l'antecedente del pronome relativo è un soggetto o un

complemento oggetto), al di fuori dell'uso standard il *che* viene utilizzato anche come forma pronominale generica e come introduttore generico di subordinate che nell'italiano standard prevedono, invece, congiunzioni subordinanti semanticamente più precise. (Zingaro, 2018)

Riallacciandosi agli studi di Alisova (1972, pp. 254-268), Cortelazzo (1976, pp. 93-98), Sornicola (1981, pp. 61-74), Sabatini (1985, p. 164) articola la polivalenza della congiunzione *che* in quattro categorie principali:

- a) il *che* con valore temporale, ossia come equivalente dei più formali *in cui*, *dal momento in cui*, *nel momento in cui*. Tra gli esempi citati da Sabatini: *La sera che ti ho incontrato; è un'ora che ti aspetto;*
- b) il *che* che congiunge le due parti di una frase scissa (cfr. punto 13 della lista del 1990): *è qui che ci siamo incontrati;*
- c) il *che* con apparente funzione di soggetto o oggetto, contraddetta da una successiva forma pronominale che ha la funzione di complemento indiretto: *Quel mio amico che gli hanno rubato la macchina;*
- d) il *che* sostitutivo di una congiunzione finale, consecutiva o causale: *aspetta che te lo spiego; vieni, che ti pettino.*

Le grammatiche, come riportato da Sabatini (ivi, p. 180), definiscono "abnorme" questo tratto grammaticale. Successivamente D'Achille (1990, p. 212) ne ammette nella norma il suo uso temporale, condannandone, invece, gli usi come relativo obliquo.

6) Il connettivo *per cui* (derivato per ellissi da *motivo, fatto, ragione per cui*, oppure per generalizzazione di casi con normali antecedenti nominali) in funzione di congiunzione causale-consecutiva.

L'uso di questo nesso relativo come congiunzione consecutiva risulta attestato sin dal Trecento (Sabatini, 1985, p. 165).

7) Il semplice *cosa?* interrogativo diretto e indiretto, ossia il suo uso al posto di *che cosa?*

Rispetto a questo fenomeno, Sabatini spiega che l'interrogativo *che cosa* ha perso terreno rispetto all'affermazione a livello nazionale del

solo *cosa*, di provenienza settentrionale, e del solo *che*, di provenienza meridionale e predominante da Roma in giù (ivi, p. 165).

8) Forme d'ingresso o connettivi testuali come *allora*, *comunque* e anche *ma*, *e*, *o* iniziali di discorso o di enunciato. A questo proposito si vedano gli studi di Sabatini (1997) e Mandelli (2006).

9) L'indicativo invece del congiuntivo nei seguenti cinque casi: dopo i verba putandi, ossia i verbi di opinione (*credo che hai ragione*), dopo una dichiarativa negativa (*non credo che hai torto*); in un'interrogativa indiretta (*gli chiesi se poteva aiutarmi*); nel periodo ipotetico dell'irrealità (*se me lo dicevi, ci pensavo io*)¹¹; nelle relative restrittive (*sei l'unico che parla arabo*).

Nella versione della lista del 1985 di Sabatini si precisa che già all'epoca si registrava una notevole tendenza alla sostituzione del congiuntivo con l'indicativo nei costrutti sopra indicati (ivi, pp. 166-167), giudicati normali nel parlato di media formalità e aventi tutti alle spalle una secolare tradizione letteraria: in particolare il caso delle ipotetiche ha dei saldi precedenti storici in Dante, Machiavelli e Manzoni (ivi, p. 167). Questi precedenti storici, uniti alle numerose norme d'uso del congiuntivo, che non chiariscono del tutto l'alternanza con l'indicativo, hanno dato vita ad un'annosa questione che "ricorre più frequente non solo nelle discussioni di linguisti e insegnanti ma anche nelle quotidiane riflessioni metalinguistiche" (Marani, 2016, p. 28) e che ha trasformato l'uso del congiuntivo in un argomento largamente studiato. Già nel 1987 Berruto affermava che "la 'morte' o 'sparizione' del congiuntivo" era diventata "un *topos*, a proposito delle vicende odierne del sistema verbale" (p. 70), proseguendo nel corso dei decenni successivi con vari studi che, invece, ne hanno dimostrato la resistenza¹².

¹¹ Si veda a questo proposito lo studio di Mazzoleni (1992).

¹² Già Serianni (1986, pp. 59-61) aveva constatato la resistenza del congiuntivo, in particolare nelle frasi complete, nella narrativa rosa e in fotoromanzi e fumetti. Restando nell'ambito della narrativa, la scarsa sostituzione del congiuntivo con l'indicativo è documentata anche in Bonomi (1996; 1997), Berisso (2000), Bonomi e Mauroni (2003), De Santis (2003) e nella letteratura per l'infanzia (Pesce, 2003). Circa i quotidiani, cartacei e online, la buona tenuta del congiuntivo è verificata da Bonomi (2002), mentre Antonelli (2007, p. 130)

In risposta a una lettera pubblica, scritta dagli alunni di una scuola di Treviso, in cui si chiedeva ai giornalisti di rispettare la grammatica, lo stesso Sabatini invitò alunni e insegnanti ad approfondire la storia della lingua, invece di chiudersi in “gabbie di norme rigide”, così da scoprire che l’alternanza del congiuntivo con l’indicativo è “una tendenza praticata da secoli”¹³, sulla quale sono stati prodotti numerosi studi, con l’intento di fare chiarezza nell’insegnamento dell’italiano sia a italofoeni nativi¹⁴ e sia a stranieri. Per quanto riguarda l’insegnamento della lingua italiana a stranieri si vedano, ad esempio, gli studi di Moneti 1988, Malinowska 1996, Maiden 2005, Manili 2015 e Marani 2016. Questi ricercatori hanno posto in particolare la loro attenzione sulle questioni relative a norma e uso e all’impiego del congiuntivo da parte degli studenti, nel tentativo di trovare una soluzione alla difficoltà messa in evidenza dal già citato studio di Manili, che afferma che “l’uso del congiuntivo nell’italiano contemporaneo non offre agli apprendenti non nativi indizi chiari e univoci delle sue funzioni” (2015, p. 154), in particolare in riferimento al tema dell’alternanza congiuntivo/indicativo.

Nella riflessione sull’alternanza dell’uso dell’indicativo e del congiuntivo, questa ricerca si riallaccia a quanto espresso da Prandi (2002). Sebbene la posizione tradizionale circa il valore del congiuntivo sia univoca, ossia che l’opposizione tra indicativo e congiuntivo sia da ricondursi principalmente all’opposizione tra realtà e non realtà¹⁵, Pran-

ha notato un ritorno significativo all’uso del congiuntivo nei testi delle canzoni dei 15 anni precedenti.

¹³ Dall’articolo *Un appello dalle elementari per la difesa del congiuntivo* (*La Stampa*, 20/02/2002), in Chiappini, De Filippo (2014, p.12), https://cdn3.scuolabook.it/Uploaded/loescher_b4442e_preview/b4442e_preview.pdf.

Tutti i siti web citati nel volume sono stati consultati il 13/09/2023.

¹⁴ Si vedano gli atti del convegno *Intorno al congiuntivo*, a c. di L. Schena, M. Prandi, M. Mazzoleni (2002) e la bibliografia ivi contenuta.

¹⁵ A questo proposito, come riportato da Ondelli e Romanini (2018, p. 197), la scelta di utilizzare nelle subordinate l’indicativo al posto del congiuntivo può basarsi su un’interpretazione semantica (indicativo come “modo della certezza” e congiuntivo come “modo dell’incertezza”: cfr. Schneider, 1999) oppure sintattica (congiuntivo come espressione di dipendenza, oppure usato per innalzare il registro. Si vedano a questo proposito Sgroi, 2013; Lepschy e Lepschy, 1988, p.

di nella sua argomentazione mette in evidenza come, fattualmente, la sola presenza del congiuntivo non è condizione sufficiente per esprimere la non realtà di una proposizione (esempio: *Ho ottenuto che Luca si assentasse dal lavoro per un'ora*, ove l'oggetto di *ottenere* è qualcosa che è realmente accaduto) e, viceversa, l'uso dell'indicativo non determina necessariamente la fattualità della proposizione (esempio: *Ho sognato che mia suocera aveva vinto un miliardo al lotto*, ove il contenuto della proposizione oggettiva – il contenuto del sogno – non è reale). Il tema verrà ripreso e approfondito nel capitolo 2.

10) La concordanza a senso tra il soggetto collettivo singolare seguito da un'espressione partitiva e il predicato plurale (*un milione di elettori non hanno votato*).

Quando il soggetto della frase è costituito da un nome collettivo (gruppo, serie, quantità, infinità, decina, dozzina, parte, classe, folla, gente, maggioranza, ecc.), il predicato spesso ha la forma del plurale, vale a dire è concordato "a senso". Questa concordanza è più frequente quando il collettivo è accompagnato dal partitivo al plurale: *Una quantità di uccelli si alzarono in volo*. [...] Ma anche senza la successiva specificazione partitiva al plurale, il nome collettivo a volte è seguito dal predicato al plurale: si sente dire, e talvolta si scrive, *Un centinaio partono domani; la terza C (sottinteso classe) vanno a Venezia; Venite, gente*. (Sabatini, 1985, p. 167)

Si tratta di un fenomeno con antichissimi precedenti, nel latino, in altre lingue romanze e non, e in italiano, per il quale Sabatini (*ibidem*) fa riferimento ad esempi tratti da Dante, Boccaccio e Manzoni.

11) La posposizione del soggetto rematico al predicato (ossia il soggetto in posizione rematica rispetto al predicato), talvolta con discordanza

204 e Renzi, 2012, pp. 52-53, secondo i quali la scelta congiuntivo-indicativo è di tipo stilistico).

di numero con questo (*non ci sono giornali e non c'è giornali, da cui si passa anche a: niente giornali*)¹⁶.

Come già anticipato per le dislocazioni, in italiano l'ordine non marcato dei costituenti della frase è soggetto + verbo + oggetto (SVO). Tale ordine è parzialmente modificabile, dunque è possibile mettere in rilievo il soggetto, o altri argomenti del verbo, per attirare su di esso l'attenzione di chi ascolta, presentandolo come l'elemento nuovo o inatteso nel discorso. Questa evidenziazione può essere realizzata in vari modi¹⁷, ad esempio sintatticamente invertendo l'ordine SV in VS: *ho preso io la decisione*. A questo proposito: "Se nella frase *Mario canta* l'informazione nuova è in *canta* (potenzialmente contrapposto a *balla, suona, ecc.*), nella frase *Canta Mario* l'informazione nuova è concentrata in *Mario* (potenzialmente contrapposto a *Fabrizio, Luigi, ecc.*)" (Sabatini, 1985, p. 162).

La possibilità di spostamento del soggetto cambia a seconda delle tipologie di frasi, poiché per alcune di esse la costruzione ordinaria, non marcata, si realizza proprio tramite la posizione post-verbale del soggetto. Per citare alcuni esempi, in presenza di un participio assoluto¹⁸ il normale ordine non marcato della frase prevede che questo sia seguito dal soggetto (esempio: *Morto Napoleone* e non *Napoleone morto*). Vi sono poi le frasi che contengono verbi che rientrano nella classe dei verbi intransitivi detti *inaccusativi*. Tale classe di verbi può essere coinvolta in frasi in cui la posizione post-verbale del soggetto rappresenta

¹⁶ La denominazione corretta di questo fenomeno (posposizione del soggetto rematico al predicato), a cui si fa riferimento in questa trattazione, è contenuta nello studio di Sabatini del 1990, p. 225.

¹⁷ In alternativa all'inversione dell'ordine dei costituenti possono essere utilizzate strategie lessicali, facendo precedere il costituente da focalizzatori come *anche, proprio, perfino* (esempio: *proprio io ho preso la decisione*), e/o prosodiche. In quest'ultimo caso, il costituente della frase focalizzato viene pronunciato con enfasi, accentuandolo e separandolo dal resto della frase tramite una pausa o un innalzamento (esempio: *IO ho preso la decisione*; oppure: *io [pausa] ho preso la decisione*). Tali strategie sono, chiaramente, anche cumulabili.

¹⁸ Come indicato da De Roberto (2010) le subordinate participiali assolute sono quelle in cui il soggetto è esplicitato e non coincide con quello della proposizione principale. Esempio: *arrivato Gianni, la festa si animò* (cfr. Egerland, 1996 e Loporcaro, Seiler, 2009).

la normale costruzione non marcata; seguono esempi tratti dal corpus PAISÀ: *occorre capacità, caparbietà e competenza; serve più umiltà; esiste un atto del Sindaco che nomina 5 membri del CSA*. In presenza di verbi inaccusativi il soggetto ha le caratteristiche dell'oggetto diretto, poiché si trova dopo il verbo, all'interno del sintagma verbale:

Si tratta di frasi presentative, costituite cioè dalla presentazione di un evento, dunque dal solo predicato. Esse, a differenza delle frasi predicative, non sono articolabili in soggetto e predicato. A questa particolarità logico-semantica se ne associa una di tipo sintattico: secondo alcune teorie in queste frasi il soggetto è un argomento interno al sintagma verbale (mentre nelle frasi predicative è argomento esterno). (Palermo, 2011)

Un esempio è: [SV[VÈ arrivata] + [SNAnna]], ove SN è il sintagma nominale e SV è il sintagma verbale.

Le frasi passive e i costrutti introdotti da *si* impersonale (o *passivante*) possono avere un soggetto postverbale percepito come non marcato (Sansò, 2003, p. 103). Alcuni esempi tratti da Palermo 2011: *furono presi provvedimenti eccezionali; si commettono molti errori per inesperienza; si verificano molti incidenti per la distrazione del guidatore*. Invece, la costruzione è più limitata nella frase interrogativa diretta, che tende ad avere un ordine fisso dei costituenti (Patota, 1990; Palermo 1997, 2011).

Fatta questa necessaria premessa, il fenomeno inserito in elenco da Sabatini riguarda invece i casi in cui lo spostamento del soggetto in posizione post-verbale costituisce una strategia di prospettiva comunicativa e, dunque, viene percepito come marcato rispetto all'ordine SVO. Anche questo fenomeno ha dei precedenti storici: l'inversione del soggetto come strategia di messa in rilievo si afferma stabilmente in italiano dal XVII secolo; nell'italiano antico e in quello letterario almeno fino alla metà del XIX secolo il soggetto era quasi sempre espresso e posposto al verbo nelle frasi interrogative (Patota, 1990). Una traccia di ciò è ancora visibile nella lingua contemporanea in contesti formali e cerimoniali (*lo giurate voi?; vuoi tu prendere per legittima sposa la qui presente...?*).

Per quanto riguarda, invece, l'uso di *niente* in funzione di aggettivo (esempio: *niente soldi*), questo fenomeno, assimilato al soggetto in posizione rematica nella presente lista del 1990, era invece trattato separatamente, in un punto a se stante della lista del 1985 (punto 31):

L'uso di *niente* in funzione di aggettivo (già attestato in epoca antica) permette di realizzare un tipo di espressione partitiva-negativa particolarmente efficace e di largo impiego: [...] *In questa marmellata niente coloranti, niente conservanti* (così spesso si legge sulle confezioni in commercio) (ivi, p. 168)

Per le caratteristiche generali della categoria grammaticale del soggetto, si veda anche Palermo 2011.

12) uso ridondante di un pronome clitico di oggetto indiretto (esempio: *mi bevo un caffè*) che coincide, di fatto, con il dativo etico, per indicare una più forte partecipazione affettiva o di interesse.

Quest'uso detto "costruzione riflessiva apparente o di affetto" è frequentissimo (con i verbi *mangiare* e *bere* (e loro sinonimi) e con altri che indicano azioni o atteggiamenti implicanti effetti sulla persona del soggetto: *Luca si è mangiato mezza torta; Verso le 11 mi bevo un caffè; Valerio si vede un film in televisione; Aldo si gode la vacanza; Stamattina mi sono fatto una splendida passeggiata.* (ivi, pp. 167-168)

A questo proposito, si osserva come la definizione "verbo in forma pronominale per indicare partecipazione affettiva", coniata da Sabatini nel 1985, possa risultare lievemente equivoca, poiché in realtà il fenomeno si basa sull'uso di un clitico di oggetto indiretto che coincide, di fatto, con il dativo etico.

Questa struttura indica la partecipazione o il coinvolgimento emotivo di una persona rispetto a un'azione o a una circostanza indicata dal predicato; è sempre espresso da un pronome atono e non è necessario

ai fini della compiutezza sintattico-grammaticale dell'enunciato (Salvi, 1988, pp. 65-66).

La denominazione “dativo etico” deriva dalla grammatica latina, in cui era possibile l'uso del pronome personale al dativo per indicare la partecipazione o il coinvolgimento emotivo di una persona rispetto a un'azione (esempio: *quid mihi Tulliola agit?*, ossia “che cosa mi combina la piccola Tullia?”, in Cicerone).

In italiano l'equivalente del dativo etico si realizza utilizzando i pronomi atoni. In particolare, quando sono utilizzati i pronomi di prima e quarta persona *mi* e *ci*, questa struttura “serve a enfatizzare la partecipazione emotiva del parlante, come effetto del riferimento deittico al soggetto enunciativo” (Cignetti, 2010¹⁹). E ancora:

Un valore affettivo-intensivo, atto a segnalare una più attiva e sentita partecipazione del soggetto all'azione, affine a quello del dativo etico propriamente detto, può essere esteso agli impieghi di pronomi atoni pleonastici accompagnati a un verbo transitivo, in frasi come *bersi una birra, farsi una passeggiata, farsi quattro risate* e sim.; oppure in casi in cui sono presenti riferimenti a parti del corpo (*soffiarsi il naso*), ad attività biologiche (*asciugarsi le lacrime*) o a indumenti che riguardano la sua persona (*togliersi il cappello*). (ivi; si veda anche Serriani (1988, p. 250))

13) La frase scissa (*sei tu che lo vuoi*)

La frase “scissa” (o “spezzata”) in due frasi, di cui la prima, col verbo *essere*, mette in forte rilievo il “nuovo”, mentre la seconda contiene il “noto”. [...] Questo tipo di frase non solo porta al massimo grado l'enfasi sul nuovo, ma amplia la durata dell'enunciato e quindi raggiunge l'effetto di quello “spezzettamento” dell'informazione che facilita la ricezione. (Sabatini, 1985, p. 163).

¹⁹ Versione online, pertanto priva di numeri di pagina.

Si tratta di una strategia sintattica che ha lo scopo di mettere in evidenza il punto di maggiore interesse del parlante e che contiene la maggiore quantità di informazione nuova. Tale segmento frasale su cui si focalizza l'attenzione si definisce *focus* o fuoco (si vedano gli studi di Benincà, Salvi, Frison, 1988; Berretta, 2002; D'Achille, Proietti, Viviani, 2005; Roggia, 2009).

La costruzione si impone a partire dal Settecento, sotto l'influenza del francese (Roggia, 2009), mentre ci sono poche attestazioni per la lingua antica. Nell'uso contemporaneo, come fa notare Renzi (2012, p. 47), la frase scissa viene utilizzata nella lingua scritta colta anche senza la funzione di mettere l'elemento che regge il verbo *essere* a contrasto con un altro elemento (esempio: *sei tu che lo vuoi, non lui!*), ma semplicemente per metterlo in rilievo (esempio: *sei tu che lo vuoi!*). È utilizzata anche in costruzioni interrogative, in cui l'introduttore viene focalizzato a inizio frase (esempi: *quand'è che torni? Con che cosa hai scritto? Dov'è che abiti?*, cfr. Renzi, 2012, pp. 47-48) e alcuni esempi di queste sono stati reperiti anche nell'italiano antico (Roggia, 2009). È interessante notare che tra gli esempi citati da Sabatini (1985; 1990, p. 98) c'è solo la tipologia della frase scissa esplicita, costituita dal verbo *essere* + [altro elemento grammaticale] + *che*; manca, invece, la tipologia implicita, basata su *a* + verbo all'infinito (esempio: *è lui a telefonare*).

14) ci attualizzante con *avere* e con alcuni altri verbi (*non c'ho tempo; non ci capisco niente*).

Nella versione del 1985 della lista si faceva riferimento all'uso attualizzante in presenza dei verbi *essere* e *avere* e di altri verbi:

originariamente con valore di avverbio di luogo 'qui' (dal lat. *ecce hic*), ha un uso larghissimo in unione con i verbi *essere* e *avere* (non con valore di ausiliari) e con altri verbi. Essa ha in gran parte perduto il suo significato originario: la sua funzione è quella di rinforzo semantico e fonico alle forme verbali. Con il verbo *essere* il *ci* [...] ha un effetto più propriamente e solo "attualizzante". (ivi, p. 160)

A questo proposito, Sabatini precisa che l'uso del *ci* è normale e obbligatorio quando si descrive un evento specifico del quale implicitamente sono richiamati aspetti materiali e localizzabili (*Oggi c'è sciopero dei giornali*) e quando il verbo *essere* è usato nel significato di "esistere", anche senza un implicito riferimento concreto ad un luogo (*C'è della gente che ama perder tempo*).

Per quanto riguarda il suo uso con il verbo *avere*, "è più evidente ancora la funzione puramente attualizzante del *ci* originariamente avverbio di luogo. Nell'uso orale della lingua esistono casi in cui il *ci* è obbligatorio per evitare ambiguità: a una domanda come *hai l'ombrello?*, Si risponde *ce l'ho* e non *l'ho*" (*ibidem*). Ciò avviene per evitare che la risposta si confonda con *lo* (pronomi o articolo), analogamente a *ce l'hai non ce l'ha*, per evitare ambiguità di suono con *la* o *non la*. La stessa domanda spesso si formula come: *c(i) hai l'ombrello?* O *ce l'hai l'ombrello?* "Sulla base di casi obbligatori come questi, si spiega l'uso (diffusissimo in tutte le regioni d'Italia e tra parlanti colti) delle espressioni *c(i) ho fame*, *c(i) ho freddo*, *non c(i) ho tempo*, *c(i) hai ragione e simili*" (ivi, pp. 160-161). Sull'uso con il verbo *avere* si veda in particolare lo studio di Moro (1998), che traccia un parallelismo tra l'uso di *averci* ed *esserci*, arrivando a paragonare a livello funzionale la struttura *c'hanno molti libri* con *ci sono molti libri*.

1.5. La lingua ritrovata

Come è evidente, si tratta di forme ormai parte dell'italiano contemporaneo che fino a non molto tempo fa erano accettate solo a livello di varietà diatopiche e "basse" (cfr. Sabatini, 1985): "tratti sub-standard vengono attratti nella sfera dello standard, dando origine a quell'italiano neo-standard che abbiamo considerato come una delle varietà cardine nell'architettura dell'italiano contemporaneo" (Berruto, 2012² [1987], p. 67). Si tratta, dunque, del risultato di un processo di graduale ampliamento di ciò che è compreso nella norma, come affermato da Sobrero, che nel 1992 descrive così questa varietà, per la quale utilizza il termine proposto da Berruto:

[Il neo-standard] è diffuso nelle classi medio-alte e nella parte più acculturata della popolazione, ed è re-

alizzato nel parlato più che nello scritto. L'etichetta di neo-standard si riferisce al fatto che su questo livello, oggi in piena evoluzione, troviamo un gran numero di forme che via via "risalgono" dai livelli inferiori (sub-standard): prima relegate nell'area delle forme "colloquiali" (o, come dicevano i vocabolari, "triviali"), ora si diffondono e sono accettate nella lingua nazionale. Lo standard così, a sua volta, estende i propri confini. (Sobrero, 1992a, p. 5)

La sintassi dell'italiano contemporaneo presenta, dunque, alcuni mutamenti dovuti spesso a quella che è stata definita come la "pressione del parlato" sulla lingua scritta, che ha portato ad una "forte spinta verso la semplificazione" (Ramat, 1993, p. 35) con una riduzione delle possibilità offerte dal sistema e un generale alleggerimento delle strutture sintattiche (cfr. D'Achille, 1993, pp. 162-163). A questo proposito Ondelli e Romanini (2018, p. 187) affermano che l'italiano dell'uso medio può essere definito come

il risultato dell'accoglienza o della riammissione nei testi scritti di tratti linguistici tipici dell'italiano parlato (sulla persistenza del parlato nella scrittura nella storia della lingua italiana cfr. D'Achille 2006). Questa permeabilità si acutizza ulteriormente all'inizio del nuovo millennio, quando Antonelli (2011) registra una nuova fase di oralizzazione e abbassamento dei registri in tutti i tipi di testo proprio nel momento in cui, grazie all'avvento dei nuovi media, la scrittura torna a svolgere un ruolo centrale per la lingua nazionale anche nella comunicazione quotidiana.

Secondo D'Achille (2003, p. 33), l'individuazione di questa varietà tra gli assi di variazione ha permesso di descrivere il repertorio della lingua italiana in maniera più completa:

L'italiano dell'uso medio e il neostandard sono caratterizzati da fatti morfosintattici e lessicali che non sempre

raccontano delle effettive novità, spesso si tratta di fenomeni già documentati in testi del passato, ma censurati o ignorati dalle grammatiche [...], [che] si sono progressivamente diffusi, tanto da apparire ormai del tutto normali non solo nel parlato, ma anche in certi tipi di testi scritti.

Infatti, tali tratti linguistici non sono innovazioni assolute dell'italiano contemporaneo, anzi molti di essi sono attestati da vari secoli e in alcuni casi si può risalire fino al latino tardo; risultano ampiamente presenti nei volgari del territorio italiano anche in epoche molto antiche e se ne trova spesso un riscontro nelle altre lingue romanze (Sabatini, 1985, p. 178), sebbene nella maggior parte dei casi non sia possibile risalire alle “prime attestazioni”, come si fa nel lessico (Renzi, 2012, p. 40).

Ciò trova conferma nel titolo del saggio con cui Sabatini nel 1990 ritorna sulla questione, ossia, *Una lingua ritrovata: l'italiano parlato*, ove l'aggettivo *ritrovata* fa riferimento al fatto che questa varietà non si basa sull'inserimento di nuovi elementi, bensì sul recupero di tratti grammaticali che, in realtà, esistono da secoli e coesistono, seppur in secondo piano, con la norma grammaticale imposta dalla tradizione che ha le sue radici in Bembo e che li ha sanzionati:

[L]a “novità” dell'italiano dell'uso medio riguarda sostanzialmente la validità della *norma*, non le caratteristiche profonde del *sistema*. Per dirla con i termini di E. Coseriu, vari tratti dell'italiano dell'uso medio erano già presenti da tempo nel *sistema* (o insieme di sistemi) che è alla base della lingua italiana ed erano stati accolti anche nelle varie *norme* scritte regionali (a volte di diffusione anche nazionale), ma non furono accolti in quella particolare norma, definibile come *supernorma*, che dal secolo XVI in poi ha dominato l'uso standard della lingua italiana: la norma letteraria di tipo bembesco, alla quale, in ultima analisi, si sono attenute le codificazioni grammaticali. (Sabatini, 1985, p. 178, dove a seguire si trova il dettaglio dei tratti più antichi).

Sebbene non vengano mai esplicitamente espresse le ragioni che hanno portato alla riduzione dell'elenco originario da 35 fenomeni linguistici (1985) a 14 (1990), è ipotizzabile che in una prima fase sia stato fatto un elenco più ampio, che raggruppasse tutte le forme presenti nell'uso comune e che storicamente erano percepite come una deviazione dalla norma codificata dalle grammatiche.

Tuttavia, già negli anni '80 del Novecento l'uso di alcune di queste forme poteva essere considerato ormai come un dato di fatto e non costituiva più un elemento di novità o di estraneità (cfr. Ondelli, Viale, 2010; Ondelli, Romanini, 2018, p. 188), dato che stavano entrando anche nello scritto di una certa formalità a livello nazionale, sebbene non fossero state codificate esplicitamente nella grammatica. Di conseguenza, la lista del 1990 può essere considerata come una seconda riflessione, con l'obiettivo di focalizzare l'attenzione sui soli fenomeni identificativi della nuova varietà panitaliana.

Il concetto dell'*italiano dell'uso medio* di Sabatini non fu esente da critiche e precisazioni, tra cui si citano in particolare le obiezioni di Castellani, secondo il quale nell'insieme dei tratti elencati non c'è "nulla che possa servire alla definizione d'una varietà nazionale d'italiano diversa dall'italiano normale o italiano senza aggettivi" (Castellani, 1991, p. 256), riprese anche da Dardano (1994, pp. 370-374) e Tesi (2005, p. 229). Oltre a questi, numerosi studiosi si sono occupati di precisare le caratteristiche di questa varietà, tra cui si citano Tavoni (2002), Simone (1993) e D'Achille (2010) per quanto riguarda la ristrutturazione del sistema verbale, Cortelazzo (2000a, pp. 16-21) sui pronomi, Renzi (2012, pp. 37-76) in particolare sulle dislocazioni e D'Achille (2010) e D'Achille, Proietti, Viviani (2005) sulla frase scissa.

Tuttavia, a prescindere dalle diverse denominazioni date a questa varietà e dalle successive elaborazioni delle sue caratteristiche, è ormai accettata la convinzione che l'italiano abbia vissuto un processo di riconfigurazione da cui è derivato un registro medio della lingua distinto dalla norma, ossia un *italiano dell'uso medio*, e distinto dallo standard tradizionale e che tale cambiamento non possa essere ignorato.

Una volta preso atto delle forme innovative attestate nell'uso medio, occorre domandarsi se tutti i tratti documentati dall'uso medio vadano realmente accettati come tratti di un nuovo standard.

L'identificazione dei tratti grammaticali che costituiscono questa varietà ha delle conseguenze anche sull'insegnamento dell'italiano, in particolare come lingua straniera all'estero, una questione che viene sollevata più volte da Sabatini sin dagli studi che descrivono l'italiano dell'uso medio, in cui si riflette sul fatto che "i discenti puntano ad acquisire una competenza (dapprima passiva, poi attiva) innanzitutto sul piano della lingua dell'uso medio, parlato e scritto" (Sabatini, 1985, p. 180). Successivamente Sabatini (1990, p. 90) coglie a questo proposito un'analogia: la varietà d'italiano a cui tendono gli apprendenti stranieri, più vicina al parlato che alla norma, può essere considerata lo specchio del cambiamento linguistico che la lingua vive – sebbene in maniera più inconsapevole – in Italia "e che consiste nell'affermarsi sempre più largo, nei più diversi contesti e nelle sfere sia del parlato che dello scritto, di un modello di lingua italiana fortemente comunicativa al tempo stesso unitaria" (*ibidem*), in contrasto con il cosiddetto "scolastichese" (Cortelazzo, 2000b, 2012), ossia di una lingua ricalcata sulla norma toscana, ma che di fatto si rivelerà nel corso del tempo distante dalla varietà usata dalla comunità dei parlanti.

Come precisato da Sabatini stesso, la scelta del verbo *affermarsi*, e non *nascere*, è dovuta al fatto che l'italiano parlato di carattere nazionale a cui fa riferimento non è una creazione recente, ma include alcuni fenomeni morfosintattici dalla storia secolare, (si citano a titolo di esempio gli studi di Nencioni, 1987 e D'Achille, 1990) erroneamente percepiti come mutamenti linguistici recenti. Nel corso del tempo tali tratti sono stati "smarriti", poiché sottoposti a censura diretta o alla semplice omissione da parte delle grammatiche specialistiche e, soprattutto, scolastiche: "A volte, più che la censura, nuoce l'omissione: anche questa rivela la mancanza di spessore nella considerazione della lingua" (Sabatini, 1985, p. 180). Analogamente Cortelazzo (1984) ha affermato che la tradizione scolastica "prescrittivo-repressiva" rischiasse di condannare la lingua italiana all'ostracismo, rallentandone l'evoluzione spontanea.

Dal divario tra l'italiano codificato nelle grammatiche e l'italiano parlato nasce la necessità per l'insegnante di italiano di fare riferimento ad una norma e, allo stesso tempo, la difficoltà di come stabilirla, dati i continui cambiamenti nella lingua. Sul tema interviene anche Sobrero (1992b) che, richiamando il dibattito avvenuto tra la fine degli anni '80 e gli anni '90 tra Sabatini e Castellani, fa suo il punto di vista

di quest'ultimo. Castellani, prendendo in esame la lista del 1985 dei tratti dell'italiano dell'uso medio, era giunto alla seguente conclusione: ad eccezione dei casi in cui questi tratti coincidano con quei tratti dell'italiano normale che nel corso del tempo sono stati omessi o non approfonditamente descritti dalla maggior parte delle grammatiche scolastiche, essi si riducono a delle semplici tendenze o a delle doppie possibilità all'interno della norma. E su questo Sobrero conclude che si tratterebbe, tutto sommato, di semplici variazioni di registro:

A mio avviso (dopo le ovvie considerazioni storiche o funzionali, che servono ad inquadrare il problema) si dovrebbero applicare, ad ognuno dei registri, criteri legati alla sincronia, all'attualità. Mi sembrano in particolare determinanti due parametri: della rilevanza statistica e della stigmatizzazione sociale. In base a questi criteri, per lo scritto formale si dovrebbero comunque accettare: (a) le forme che hanno una presenza e una frequenza statisticamente molto rilevanti nella produzione linguistica scritta e orale; (b) le forme che, già colpite da censura per la loro appartenenza al livello colloquiale o dialettale, ora sono generalmente accettate. E, naturalmente, si dovrebbero respingere le forme che non hanno questi requisiti. In questo modo, ad esempio, si accetterebbe senza tentennamenti il *lui* soggetto (criterio a) e la frase segmentata (tipo *i soldi te li ho dati*) (criterio b); ma si rifiuterebbe il *codesto* (a) e la costruzione del tipo *la valigia che ci ho messo i libri* (b). In modo un po' diverso ci si potrebbe regolare nell'orale di registro formale, e in modo ancor più permissivo nei registri colloquiali. (Sobrero, 1992b, p. 153)

Dunque, in entrambi i casi, seppur declinata attraverso i due aspetti della rilevanza statistica e della accettazione sociale, la questione si basa sul determinare la preesistente affermazione di tali fenomeni, sia essa di carattere quantitativo o sociale. Tale punto di vista, tuttavia, conserva una certa vaghezza sulla questione della funzione linguistica, ossia, quali strutture diano un effettivo apporto alla comunicazione, che viene data come premessa all'inizio del brano sopra citato ("dopo le ovvie

considerazioni storiche o funzionali”) (*ibidem*), ma di fatto non viene affrontata esplicitamente.

L'insegnamento dell'italiano L2 offre a questo proposito un banco di prova: se è indispensabile proporre a un apprendente straniero la dislocazione a sinistra di un complemento oggetto, come *La moto, l'ho parcheggiata in strada*, è sensato proporre come standard una forma come *C'ho la moto*?

Questo esempio ci conduce al nucleo di questa trattazione, ossia la necessità di identificare criteri di discriminazione espliciti, controllabili e affidabili, che permettano di riesaminare i tratti dell'italiano dell'uso medio e di separare le forme in grado di entrare in uno standard rinnovato dalle forme che condannano i loro utenti alla censura sociale senza presentare alcun vantaggio funzionale. Non è possibile fare riferimento a “tratti sub-standard [che] vengono attratti nella sfera dello standard, dando origine a quell'italiano neo-standard che abbiamo considerato come una delle varietà cardine nell'architettura dell'italiano contemporaneo” (Berruto, 2012² [1987], p. 67) senza porsi il problema di quali tratti siano in grado di affrontare questa promozione allo standard.

Infine, al termine di questa esposizione, è necessario anche definire la questione della varietà linguistica dal punto di vista terminologico. Sabatini e Berruto hanno studiato lo stesso fenomeno, ossia la nuova varietà linguistica standard, partendo da due prospettive diverse, rispettivamente la prospettiva dello storico della lingua e del sociolinguista, che non entrano in contrasto, bensì si integrano a vicenda. Per sottolineare anche dal punto di vista terminologico quanto appena esposto, in quest'opera verrà utilizzata la denominazione data da Sabatini di *italiano dell'uso medio* per fare riferimento in blocco a tutti i fenomeni attestati nell'uso e la denominazione *italiano neostandard* di Berruto per fare riferimento a quei tratti che meritano effettivamente di entrare in uno standard che sia al tempo stesso adeguato alla varietà degli usi, funzionale e socialmente condiviso e accettato.

Nel capitolo successivo, a questo proposito, verranno proposti dei criteri espliciti e affidabili che permettano di riesaminare l'elenco dei tratti dell'italiano dell'uso medio di Sabatini (1990) e stabilire quali tra queste strutture linguistiche ormai diffuse possano essere effettivamente ritenute parte di un nuovo standard linguistico funzionalmente adeguato e socialmente accettato.

2. I CRITERI DI UN NUOVO STANDARD LINGUISTICO

Dopo aver preso atto dei fenomeni linguistici il cui uso è ormai documentato, è necessario fare un passo successivo, ossia definire criteri espliciti e affidabili che permettano di riesaminare l'elenco dei tratti dell'italiano dell'uso medio di Sabatini (1990) e stabilire quali tra queste strutture linguistiche ormai diffuse possano essere effettivamente ritenute parte di un nuovo standard linguistico funzionalmente adeguato e socialmente accettato. Tale ricerca verrà svolta tenendo conto delle implicazioni che la definizione di uno standard ha sull'insegnamento di una lingua, nel caso specifico dell'italiano.

2.1. Come definire lo standard linguistico

Come già espresso precedentemente, dal divario tra l'italiano codificato nelle grammatiche e l'italiano parlato nasce la necessità per l'insegnante di italiano di fare riferimento ad una norma e, allo stesso tempo, la difficoltà di come stabilirla, dati i continui cambiamenti nella lingua e

le diverse posizioni adottate dagli studiosi circa quali fenomeni siano da considerarsi parte del nuovo standard.

Sobrero (1992b) e Castellani (1991), richiamando il dibattito avvenuto tra la fine degli anni '80 e gli anni '90 tra quest'ultimo e Sabatini, ritengono che per poter stabilire quali fenomeni linguistici siano costitutivi del nuovo standard sia necessario determinare se i fenomeni linguistici siano statisticamente diffusi e se siano percepiti come socialmente accettabili dai parlanti.

Tali punti di vista, tuttavia, meritano alcune riflessioni. Data per assodata la diffusione di questi fenomeni linguistici, basarsi esclusivamente sulla rilevanza statistica porterebbe a considerare come parte integrante dello standard qualunque fenomeno linguistico avente un'ampia diffusione. La frequenza statistica di una struttura, infatti, indica solo che essa è usata dai parlanti e che vada tenuta d'occhio come segnale di un possibile cambiamento in atto o tendenziale della lingua, ma ciò non implica affatto che il suo uso sia auspicabile nei registri di medio-alta formalità coperti dallo standard.

Inoltre, applicare il criterio della diffusione statistica permette di osservare la questione solo dal punto di vista dei parlanti, che usano ampiamente determinate strutture linguistiche e – eventualmente – le ritengono socialmente accettabili, ma in questo modo si tralascia il punto di vista del linguista e dell'insegnante di lingue, ossia: queste strutture linguistiche danno un effettivo apporto funzionale alla lingua tanto da poter essere classificate come parte dello standard? Il loro uso evita un involontario abbassamento del registro linguistico e, di conseguenza, evita che i parlanti vengano tacciati di trascuratezza formale o scarsa conoscenza della lingua? In altre parole: è davvero *consigliabile* insegnarle?

Esemplificando, l'uso di *gli* al posto di *le* o dell'indicativo al posto del congiuntivo (due esempi chiave per l'argomentazione nelle pagine successive) può avvenire ad esempio per distrazione, oppure a causa di una scarsa conoscenza dei meccanismi della lingua (anche per imitazione di parlanti nativi che a loro volta hanno delle competenze linguistiche limitate) o, più raramente, può essere motivato dalla consapevolezza distaccarsi volontariamente dalla norma per cambiare il registro stilistico della frase e produrre un determinato effetto pragmatico.

Quest'ultimo caso, tuttavia, presuppone una consapevolezza linguistica molto elevata, che permetta di scegliere se usare o meno determinati

fenomeni linguistici che non rientrano nello standard sulla base di una valutazione della situazione comunicativa, del suo livello di (in)formalità, degli interlocutori e del registro linguistico. È chiaro che si tratti di un livello di consapevolezza che l'apprendente (straniero o italofono nativo) non ha plausibilmente ancora raggiunto.

Di conseguenza, è necessario fornire un modello linguistico sicuro, che permetta innanzitutto di evitare che gli apprendenti causino un involontario abbassamento del registro linguistico e si espongano al giudizio sociale degli interlocutori a causa della loro trascuratezza formale, e che permetta eventualmente in una fase più avanzata del processo di apprendimento di *scegliere consapevolmente* di deviare dalla norma linguistica per produrre un determinato effetto pragmatico, ad esempio per dare una particolare “coloritura” alla lingua in uso tramite l'uso di forme substandard.

Partendo da queste riflessioni, la proposta di questo studio per la definizione dello standard linguistico è di passare in rassegna i tratti dell'italiano dell'uso medio alla luce di due criteri: il criterio della funzionalità e il criterio dell'accettabilità sociale, sulla base rispettivamente degli studi di Prandi (2020) e di Ondelli e Romanini (2018), che verranno qui di seguito brevemente ripercorsi.

2.2. Il criterio della funzionalità

A distanza di quarant'anni dagli studi di Sabatini, Prandi (2020, pp. 174-175) nella sua grammatica tocca il tema della censura generalizzata delle grammatiche rispetto ad alcuni tratti dell'italiano dell'uso medio (Sabatini, 1985; 1990). Ampliando la riflessione, Prandi nota che, sebbene tutti i fenomeni linguistici dell'italiano dell'uso medio siano accomunati dal fatto di distaccarsi dalle norme della tradizione prescrittiva, in realtà sia possibile tracciare una distinzione tra fenomeni linguistici funzionali e non funzionali. Questo punto può essere spiegato attraverso gli esempi di due fenomeni che illustrano il criterio funzionale: le strutture marcate di frase e la forma pronominale *gli* al posto di *le*.

Prandi parte dal caso delle strutture marcate di frase, come la dislocazione a sinistra, la dislocazione a destra, la frase scissa e il tema sospeso, delle quali nota che sono generalmente assenti o scarsamente trattate

nelle grammatiche, perché considerate tradizionalmente caratteristiche del parlato informale, o addirittura dell'improvvisazione, e causa di un abbassamento del registro. Tuttavia, queste strutture rappresentano degli strumenti indispensabili della prospettiva comunicativa, di cui la lingua non può fare a meno se non al prezzo dell'impossibilità di esprimere le gerarchie di peso comunicativo dall'interno della frase. La dislocazione a sinistra e a destra e la frase scissa permettono rispettivamente di tematizzare o focalizzare un costituente della frase. Entrambe, in un certo senso, lo mettono in rilievo, ma con esiti opposti. Dunque, proprio perché apportano un effettivo valore funzionale alla lingua, Prandi ritiene che sia necessario riaffermare con decisione che le frasi marcate non costituiscono una caduta di registro.

Ciò detto, se da un lato la critica all'omissione e alla censura nelle grammatiche fatta da Sabatini e la riflessione di Prandi pongono l'attenzione sulla tendenza errata a rifiutare in blocco tutto ciò che differisce dalla norma prescrittiva, dall'altro ciò non deve indurre, per contro, a trarre conclusioni affrettate circa l'accettabilità di tutti i tratti dell'italiano dell'uso medio.

Questo punto di vista riprende quanto già affermato da Renzi (2012, p. 39), che mette in guardia dal pericolo di ricadere sia nell'errore dei grammatici classici, che per secoli hanno inserito tra i loro scopi la tutela della buona lingua, la tutela dagli errori e, dunque, dal cambiamento, sia nell'errore opposto, ossia di credere che tutti gli errori di oggi saranno le regole di domani. Tuttavia, mentre la conclusione di Renzi sulla questione è solo di tipo evolucionistico, ossia alcune innovazioni si imporranno, mentre altre perderanno terreno e verranno eliminate, l'approccio di Prandi, invece, poggia saldamente sul criterio della funzionalità o mancata funzionalità delle forme linguistiche, da cui è possibile trarre dei criteri per l'insegnabilità dei tratti.

Come messo in evidenza da Prandi (2020, pp. 174-175), nell'elenco dei tratti distintivi dell'italiano dell'uso medio (Sabatini, 1985; 1990) sono presenti sia strutture linguistiche funzionalmente indispensabili, come la dislocazione a sinistra e a destra e la frase scissa, sia strutture linguistiche che, invece, infrangono la norma grammaticale senza – o con scarsi – vantaggi funzionali, portando semplicemente ad un abbassamento del registro linguistico. Un esempio di innovazione entrata nell'"uso medio" per la quale il criterio di funzionalità è dirimente, in

quanto permette di distinguere due innovazioni all'apparenza simili ma di valore opposto, è l'affermazione di *gli* in alternativa alla forma singolare femminile *le* e alla forma plurale *loro*. Contrariamente a *loro*, *gli* presenta il vantaggio di fungere non solo da enclitico (*digli* vs *dì loro*) ma anche da proclitico (*gli dico* vs l'impossibilità dell'uso *loro dico*). Questa caratteristica colma una lacuna nel paradigma delle forme pronominali enclitiche, che non ha una forma per la terza persona plurale. Di conseguenza, l'innovazione rappresenta un'evoluzione funzionale del sistema pronominale, che permette di includere questo fenomeno linguistico nel neostandard. Al contrario, nell'uso della forma pronominale *gli* al posto di *le* non vi è alcun vantaggio funzionale; si tratterebbe di una deviazione del tutto arbitraria rispetto alla norma. Serianni nella sua grammatica (1989b, p. 249) lo segnala come "decisamente da evitare anche nel parlato colloquiale" a prescindere dai precedenti letterari illustri da lui stesso menzionati (Boccaccio, Machiavelli, Carducci, Verga e Moravia), ad ulteriore conferma che il solo criterio della diffusione statistica non compensa il criterio della funzionalità.

Sulla base di quanto esposto da Prandi, è possibile ricavare questo primo criterio. Data una struttura linguistica che si distacca dalla norma grammaticale tradizionale:

- se il suo uso produce un effettivo apporto funzionale alla lingua (ad esempio, è funzionale a esprimere un certo concetto o raggiungere un determinato scopo pragmatico), allora quella struttura linguistica va trasmessa nell'insegnamento della lingua italiana come parte integrante della nuova varietà standard;
- se, invece, il suo uso apporta zero o scarsi vantaggi funzionali, si tratta di una forma che dal punto di vista funzionale non è necessaria, poiché è solo un'alternativa arbitraria a una forma standard perfettamente funzionante. In quel caso quella struttura linguistica può essere menzionata nell'insegnamento come una variante substandard che, per quanto possa esser diffusa, provoca di fatto un abbassamento del registro linguistico che espone l'apprendente a una censura sociale.

Tuttavia, per alcuni tratti dell'italiano dell'uso medio il criterio funzionale non risulta dirimente. Il caso più evidente è quello dell'uso dell'indicativo al posto del congiuntivo, che merita alcune precisazioni.

Il congiuntivo si giustifica sul piano funzionale quando si trova in una frase indipendente. In questo caso ha un valore definito, poiché esprime varie gradazioni della non-realtà nel contesto di ordini e auguri, in alternanza con l'imperativo: *volesse il cielo!*; *venga pure*, eccetera. Va però precisato che il suo uso in italiano, rispetto al latino, è marginale. Nelle frasi subordinate, viceversa, il congiuntivo è un tratto grammaticale che contribuisce fortemente alla fisionomia dell'italiano e che innalza il livello dell'espressione, ma che – contrariamente all'opinione comune – non ha un valore funzionale. Nella tradizione grammaticale, il congiuntivo è da sempre contrapposto all'indicativo come modo della non-realtà.¹ Tuttavia, come suggerisce Prandi (2002, pp. 32-39), di cui a seguire si citano alcuni degli esempi utilizzati, ad un'analisi più accurata, il presupposto per cui la sola presenza del congiuntivo in una frase basti a conferire un valore modale di non-realtà si rivela fuorviante rispetto alle reali condizioni d'uso del congiuntivo, che necessitano di una visione molto più ampia della questione. Per spiegare questo punto, verranno illustrati alcuni esempi.

Nelle frasi complete soggettive e oggettive, la realtà o l'irrealtà (nei suoi diversi gradi) della subordinata è determinata dal contenuto del verbo principale e non dalla presenza del congiuntivo. Ad esempio, in *Suppongo che tu sia al corrente delle nostre decisioni* il contenuto incerto della subordinata è dato dalla semantica del verbo principale *supporre* e non dalla presenza del congiuntivo. La funzione del congiuntivo, dunque, è quella di codificare una relazione grammaticale in sé vuota di contenuto: la relazione tra il verbo principale e il suo contenuto in posizione di oggetto. Per le frasi come quella citata nell'esempio si può affermare quanto segue:

In caso di alternanza tra indicativo e congiuntivo in presenza di una stessa accezione di uno stesso verbo, cambia il registro stilistico dell'espressione, ma non cambia il valore modale della completiva, dato che non cambiano né il contenuto del verbo principale né la

¹ Cfr. le grammatiche Battaglia, Pernicone (1951), Sorrento (1951), Dardano, Trifone (1985), Fogarasi (1969) [1983], Sensini (1990) e gli studi di Wandruszka (1991), Giorgi, Pianesi (1997).

relazione grammaticale, che sono gli unici fattori pertinenti. (Prandi, 2002, p. 38)²

Ciò si evince anche da alcuni esempi citati dall'autore nel suo stesso studio (ivi, pp. 38-39): confrontando l'esempio precedente, *Suppongo che tu sia al corrente delle nostre decisioni*, con *Gli inquirenti suppongono che il suo compito era quello di mantenere [...]*, emerge che la scelta dell'indicativo è giustificata da una differenza di registro, ma non influisce sul contenuto della subordinata. In aggiunta, si prendano in esame i seguenti esempi: 1) *Spero che Giovanni venga*; 2) *Spero che Giovanni verrà*; 3) *Spero che Giovanni viene*. Dall'esempio 1 all'esempio 3 si assiste ad un graduale abbassamento del registro, tuttavia l'assenza del congiuntivo non cancella il contenuto del verbo *sperare*: dato come incerto nel caso 1, non diviene certo nel caso 3.

Una riflessione simile può essere fatta circa i verbi principali che esprimono emozioni: *Temo che Giovanni sia partito* vs *Mi dispiace che Giovanni sia partito*: "Posto che *temere* e *dispiacere* reggono entrambi il congiuntivo, ad esempio, il contenuto della proposizione oggettiva di *temere* è dato come non reale, mentre il contenuto della proposizione soggettiva di *dispiacere* è presupposto come vero" (*ibidem*). Ciò ovviamente non significa che il congiuntivo presente nei due esempi abbia valori opposti, bensì che la presenza del congiuntivo non influisce sul valore reale o non reale del contenuto della

² Ciò si contrappone ai casi, elencati da Serianni (1989, p. 556) e ripresi da Schmitt-Jensen (1970), di verbi per i quali l'alternanza della reggenza dell'indicativo e del congiuntivo determina un cambiamento di accezione di significato. A questo proposito Prandi (2002, p. 39, nota 3): "*Ammettere*, ad esempio, regge l'indicativo con il valore constativo di "rendersi conto, riconoscere" e il congiuntivo con il valore di "fare un'ipotesi": *Ammetto che hai fatto il tuo dovere*; *Ammettiamo che non piova*. Un'alternanza analoga si ha con l'uso constativo e l'uso direttivo dei verbi di comunicazione: *Gli ho telefonato che era ora di partire*. *Gli ho telefonato che partisse quanto prima*. In alcuni casi l'alternanza è coerente con una differente disposizione del soggetto enunciante: *Ti ho detto che Luigi è partito*; *Dicono che Luigi sia partito*. Tutte le volte che la reggenza ammette un'alternanza coerente con una distinzione modale, la distribuzione del modo non è arbitraria ma motivata dall'affinità elettiva di cui si è detto", ossia l'indicativo come modo della realtà e il congiuntivo come modo dell'irrealtà.

completiva. I due esempi appaiono incoerenti solo se si pretende di giustificare i due valori opposti della subordinata non a partire dal contenuto di verbi principali distinti, ma dalla presenza del modo congiuntivo, secondo quanto prescritto tradizionalmente dalla norma:

Se ammettiamo che il valore modale della proposizione completiva è controllato dal verbo indipendentemente dalla presenza del congiuntivo, il problema rappresentato dai predicati emozionali e valutativi, e da predicati come *ottenere* semplicemente si dissolve. [...] Di fronte a esempi di questo genere, la domanda pertinente non è: “Che differenza di valore c’è tra congiuntivo e indicativo?”, ma “Fino a che punto le nostre orecchie sono disposte a accogliere costruzioni che usano l’indicativo invece del congiuntivo in dipendenza da certi verbi, come in *Spero che Giovanni viene?*”. Le mie stridono, ma non è questo il punto. Resta il fatto che le costruzioni più avventurose dell’italiano popolare non influiscono sullo statuto modale della completiva, alla sola condizione che la relazione grammaticale tra verbo e completiva rimanga trasparente. (Prandi, 2002, p. 39)

Proseguendo con gli esempi, nelle frasi subordinate completive sia il modo verbale (congiuntivo/indicativo), sia il valore modale (reale o non reale) della frase completiva sono imposti dal verbo reggente: ci sono, infatti, verbi che reggono il congiuntivo e altri che reggono l’indicativo, e non c’è una correlazione significativa tra il contenuto del verbo e il modo della subordinata. La conseguenza è che sono attestate tutte le combinazioni tra modo e valore modale, ad esempio:

- a) uso del congiuntivo in una frase dal valore modale non reale: *Non ho ottenuto che Chiara fosse assunta;*
- b) uso del congiuntivo in una frase dal valore modale reale: *Ho ottenuto che Chiara fosse assunta;*
- c) uso dell’indicativo in una frase dal valore modale reale: *Ho saputo che sei stata assunta;*
- d) uso dell’indicativo in una frase dal valore modale irreali: *Ho sognato che mia suocera aveva vinto un miliardo.*

Si può ipotizzare che il congiuntivo acquisti un valore di non-realtà nei pochi casi di frasi subordinate complete in cui il verbo reggente permette al parlante di scegliere tra congiuntivo e indicativo, ossia nel caso dei verbi del dire e verbi del pensiero. Ad esempio, in *Dicono che Matteo è scappato di casa* chi sceglie consapevolmente l'indicativo mostra di credere al contenuto della diceria, mentre chi sceglie consapevolmente il congiuntivo (*dicono che Matteo sia scappato di casa*) ne prende le distanze. Tuttavia, non è detto che l'alternanza dei modi indicativo/congiuntivo sia sistematicamente correlata ad una scelta consapevole di attribuire il valore modale di realtà o irrealtà, ma può dipendere da altri fattori. Possiamo immaginare due casi: un parlante che usa il congiuntivo per ragioni di registro e sottoscrive la realtà del fatto, e un parlante che usa l'indicativo senza sottoscrivere la realtà, perché il congiuntivo non è nel suo repertorio attivo.

Nel caso delle subordinate non complete, dette anche margini (Longacre, 2007), che collegano due processi indipendenti, come ad esempio la causa all'effetto o un'azione al fine, sia il valore modale di realtà o non-realtà, sia il modo verbale indicativo o congiuntivo sono controllati dalla congiunzione che realizza il collegamento tra i processi. Nelle subordinate non complete il congiuntivo ricorre:

- dopo congiunzioni e locuzioni finali (come *affinché, perché*) o eccezzuative (come *a meno che*);
- dopo congiunzioni concessive (*benché, sebbene*): a questo proposito Prandi fa notare che la relazione concessiva è coerente solo se i contenuti della frase principale e della secondaria sono dati entrambi come reali (per esempio in *Sebbene piova, partiremo per la montagna*) e che le uniche congiunzioni concessive che codificano la realtà della frase subordinata di fatto sono quelle che reggono il congiuntivo: *sebbene, quantunque, benché*: un'ulteriore conferma della compatibilità del congiuntivo con la realtà. Viceversa, nel caso dell'uso della congiunzione *anche se*, che regge l'indicativo, la realtà della subordinata non è codificata, ma è un'opzione da accertare caso per caso, cfr. Mazzoleni, 1991 (2001);
- nelle frasi temporali è previsto l'uso del congiuntivo (*usciamo prima che piova*) o indicativo (*siamo usciti dopo che è piovuto*), a seconda che l'attenzione sia focalizzata sull'incertezza del futuro (congiuntivo) o sulla realtà del passato (indicativo). Sebbene casi

come questo sembrano ricondurre allo schema tradizionale dell'indicativo come modo della realtà e il congiuntivo come modo dell'irrealtà, di fatto questi esempi ci mostrano semplicemente che in certi casi l'uso del congiuntivo coesiste con l'espressione della non-realtà, ma in molti altri ciò non avviene, come nei casi precedentemente descritti. Se ci fosse una correlazione codificata (congiuntivo = non-realtà), questa si verificherebbe sempre, senza eccezioni;

- nelle ipotetiche della possibilità e dell'irrealtà: nell'italiano standard, l'alternanza tra indicativo e congiuntivo nel periodo ipotetico separa, all'interno della non realtà, due gradi diversi di possibilità: possibilità di grado alto (*Se Anna è in casa possiamo a salutarla*) e possibilità di grado basso (*Se Anna fosse in casa potremmo passare a salutarla*). Invece, nel registro medio-basso l'imperfetto indicativo sostituisce il congiuntivo per la possibilità di grado basso e impossibilità / irrealtà. Di fatto, il registro alto distingue l'irrealtà e due gradi di possibilità, mentre il registro basso distingue direttamente l'irrealtà dalla possibilità. Ancora una volta, non c'è correlazione sistematica tra congiuntivo e non realtà (Prandi, 2002, p. 41);
- le frasi relative possono essere utilizzate per restringere il campo nell'identificare un referente. Potremmo pensare che la scelta tra l'indicativo e il congiuntivo segnali se il referente è identificato (*Cerco il libro che mi hai regalato*, ossia cerco *quel* libro e non un altro) o solo ipotizzato: *Non trovo nessuno che abbia voglia di venire al cinema con me*). Tuttavia, come precisato da Prandi (2012, pp. 119-121) la correlazione tra l'uso dell'indicativo o del congiuntivo per esprimere la certezza o l'ipotesi di una certa proprietà non è regolare: *È il libro più bello che ho letto quest'anno* vs. *È il libro più bello che abbia letto quest'anno*. In entrambi i casi, il libro a cui ci si riferisce è comunque identificato con certezza e la scelta tra indicativo e congiuntivo si riduce ad una questione di registro. Inoltre, non si può dare per scontato che chi parla scelga consapevolmente tra indicativo e congiuntivo: l'indicativo potrebbe essere utilizzato per esprimere una possibilità da un parlante che non possiede nel suo repertorio attivo l'alternanza indicativo/congiuntivo. "Ancora una volta una questione di valore grammaticale si capovolge in una differenza di registro" (ivi, p. 120).

In sintesi, l'ipotesi di Prandi è la seguente: nei casi in cui l'uso del congiuntivo è imposto, il contenuto di realtà o di gradi diversi di possibilità della frase dipende in effetti non dalla sua presenza nella frase, ma da fattori diversi, quali il contenuto del verbo principale (*Mi spiace che ti abbiano/hanno rubato la bici*) o della congiunzione (*sebbene piova/piove*) (Prandi, 2002). Il congiuntivo ha invece un suo valore solo nei casi in cui il suo uso si basa su una scelta libera del parlante, che può alternarlo all'indicativo per cambiare il registro stilistico della frase. Dunque, l'opposizione indicativo-congiuntivo non andrebbe più ricondotta sempre alla sola opposizione tra codifica della realtà e dell'irrealtà, secondo la posizione tradizionale, bensì bisognerebbe tenere conto anche dei casi in cui la sua scelta è un'opzione del parlante, e di conseguenza, il suo valore si esprime nel contenuto del messaggio (cfr. i già citati Lepschy e Lepschy, 1988, p. 204 e Renzi, 2012, pp. 52-53, secondo i quali la scelta tra congiuntivo e indicativo è di tipo stilistico). Concludendo, usare l'indicativo al posto del congiuntivo nelle frasi argomentali e nei margini non comporta alcuna perdita di funzione. La scelta tra indicativo e congiuntivo non è di sistema, ma di repertorio: nel caso in cui venga usato l'indicativo si verifica un abbassamento di registro, ma di fatto nessuna conseguenza funzionale sullo statuto modale (realtà/non-realtà) della subordinata. Tuttavia, ciò non è nemmeno una ragione sufficiente per eliminare del tutto il congiuntivo dallo standard a favore dell'indicativo, poiché anche la stessa scomparsa del congiuntivo non avrebbe di fatto alcun vantaggio funzionale, oltre al fatto di eliminare una delle caratteristiche più tipiche della lingua italiana.

Il caso del congiuntivo mostra che l'applicazione del criterio funzionale per riesaminare i tratti dell'italiano dell'uso medio è dirimente solo quando la forma innovativa è in grado di ampliare il repertorio di opzioni offerte al parlante, ma perde efficacia quando la forma tradizionale e la forma innovativa sono entrambe prive di vantaggi funzionali. Sulla base di ciò, vale la pena chiedersi: un'eventuale rinuncia al congiuntivo a favore dell'indicativo sarebbe neutra, oppure esporrebbe il parlante alla censura sociale, senza che da questa scelta derivino vantaggi funzionali?

Questo interrogativo apre lo scenario verso la proposta di un ulteriore criterio indipendente e complementare, da applicare nell'identificazione del neostandard, ossia il criterio dell'accettabilità sociale.

2.3. Il criterio dell'accettabilità sociale

La necessità di applicare anche un criterio che tenesse conto della percezione di accettabilità da parte della comunità dei parlanti era già emersa nello studio di Sobrero (1992b, p. 153), in cui accanto alla diffusione statistica dei fenomeni linguistici si proponeva il criterio della stigmatizzazione sociale.

Successivamente, il tema è stato ripreso e applicato in alcuni studi che hanno indagato tramite dei sondaggi il grado di accettabilità di alcuni tratti dell'italiano dell'uso medio, in particolare Ondelli e Romanini (2018) e Grandi (2018, 2019). Ondelli e Romanini (2018) hanno sottoposto a un campione di 150 persone provenienti da tutto il territorio nazionale, prive di competenze specifiche nelle questioni linguistiche, dei brevi estratti di articoli pubblicati tra il 2015 e il 2016 da *Corriere della Sera* e *Repubblica*, contenenti alcuni dei tratti elencati da Sabatini (1985) e Berruto (1987). Oltre alla valutazione del grado di accettabilità su una scala di cinque descrittori che spaziano dalla totale inaccettabilità alla piena accettabilità, con tre gradi intermedi (accettabile nel parlato formale, nel parlato informale e nel parlato o nello scritto di media formalità) è stato chiesto loro un commento libero sulla loro qualità linguistica.

Sulla base dei risultati ottenuti è possibile suddividere i tratti dell'italiano dell'uso medio in tre macrocategorie: quelli che non sono oggetto di alcuna censura da parte dei partecipanti o sono oggetto di segnalazioni minoritarie; quelli che invece vengono segnalati con costanza come non accettabili e, infine, quelli per i quali le valutazioni sull'accettabilità sono molto variabili, sulla base del co-testo in cui sono inseriti o delle caratteristiche diastratiche dei partecipanti.

Senza entrare ora nel dettaglio dei risultati ottenuti, che verranno inseriti nella sezione successiva di questa trattazione, in generale si rileva che la norma linguistica interiorizzata dai partecipanti tramite la formazione scolastica ha una forte influenza sui loro giudizi: gran parte delle opposizioni dei partecipanti sono assimilabili ad alcune regole caratteristiche dell'italiano scolastico, che hanno contribuito a formare il loro modello linguistico di riferimento in termini di percezione della correttezza grammaticale e della neutralità di registro³.

³ Alla base della decisione di prendere come punto di riferimento in questa trattazione solo lo studio di Ondelli e Romanini (2018) vi è l'eterogeneità del gruppo

2.4. Applicazione dei criteri della funzionalità e dell'accettabilità sociale

Sulla base di quanto esposto finora, la proposta metodologica di questo volume è di riesaminare i tratti dell'italiano dell'uso medio (Sabatini, 1990) innanzitutto sulla base del criterio della funzionalità, al fine di tracciare una distinzione tra i casi in cui il distacco dalla norma grammaticale tradizionale produce o meno un effettivo apporto funzionale alla lingua e, di conseguenza, se può essere considerata o meno parte integrante della nuova varietà standard. Inoltre, al fine di avere un riscontro circa la percezione dell'accettabilità sociale di tali fenomeni linguistici, la distinzione tra fenomeni funzionali e non funzionali verrà integrata con i risultati ottenuti dalla ricerca di Ondelli e Romanini (2018). Ne deriva la seguente classificazione:

Fenomeni linguistici neostandard: il loro uso produce un effettivo apporto funzionale alla lingua (ad esempio, raggiungere un determinato scopo pragmatico), dunque possono essere considerati parte dello standard linguistico e, di conseguenza, essere inseriti tra i contenuti da trasmettere nell'insegnamento:

- la costruzione dei verbi con il cosiddetto dativo etico per indicare una più forte partecipazione affettiva o di interesse (*mi bevo un caffè*);
- la frase scissa (*sei tu che lo vuoi*). L'accettazione di questa struttura linguistica trova conferma anche nella ricerca di Ondelli e Romanini (2018), in cui le segnalazioni sull'inaccettabilità del fenomeno sono davvero minoritarie;

di rispondenti coinvolti nella loro ricerca e l'interesse a conoscere la percezione di accettabilità da parte di persone prive di competenze specifiche nelle questioni linguistiche. A differenza di questo studio, invece, l'indagine di Grandi (2018, 2019) ha coinvolto un gruppo di circa 200 rispondenti, composto tuttavia solo da studenti dell'Università di Bologna. A titolo informativo, tra i tratti che in assoluto mostrano i valori di accettabilità più alti nelle ricerche di Grandi sopra nominate vi sono il futuro epistemico, l'uso di *loro* e *lui* come soggetto, il passato prossimo al posto del passato remoto e il *ne* ridondante. Il gruppo di informanti che accetta questi tratti tende anche a rifiutare in modo omogeneo *a me mi* e le varie manifestazioni del *che* polivalente, ad eccezione parziale di quello con funzione temporale.

- dislocazioni a sinistra (*quel libro l'ho già letto*), a destra (*l'ho già letto, quel libro*) della parte rematica della frase con la ripresa pronominale atona, che può giungere fino a costruzioni con uso pleonastico della particella pronominale, tipo *di pane non ne ho più* (ossia, il doppio partitivo). Nella ricerca di Ondelli e Romanini (2018) le frasi con la dislocazione a sinistra sono oggetto solo di segnalazioni minoritarie (maggiori segnalazioni sull'inaccettabilità sono state riscontrate per la dislocazione a destra): in base alle risposte dei partecipanti sembra che questi fenomeni non vengano più percepiti come forme caratteristiche del parlato o di registri informali;
- l'anacoluto (*Giorgio, non gli ho detto nulla*) da punto di vista funzionale può essere un modo semplice per tematizzare un costituente della frase che non è un complemento oggetto;
- *gli* come forma dativale al posto di *loro* (*a loro*), per le ragioni precedentemente menzionate del vantaggio funzionale di *gli* come proclitico ed enclitico. È uno dei fenomeni linguistici per i quali le valutazioni sull'accettabilità da parte dei partecipanti alla ricerca di Ondelli e Romanini (2018) sono molto variabili: l'uso di *gli* al posto del plurale *loro* risulta più accettabile per i giovani tra i 18 e i 25 anni, ma il loro numero è pari alla metà dei partecipanti adulti che lo segnalano, invece, come inaccettabile.

Fenomeni linguistici substandard: il loro uso apporta zero o scarsi vantaggi funzionali, poiché si tratta di un'alternativa arbitraria a forme linguistiche standard perfettamente funzionanti. Sono scelte che espongono alla censura sociale chi le usa senza che ci sia un apporto funzionale. In questi casi, la struttura linguistica può essere menzionata nell'insegnamento come una variante substandard che, per quanto possa esser diffusa, provoca solo un abbassamento del registro linguistico:

- l'impiego di *gli* al posto di *le* (*a lei*), è una deviazione del tutto arbitraria rispetto alla norma, essendo *le* perfettamente funzionante come enclitico. Ciò trova conferma anche nella ricerca di Ondelli e Romanini (2018), dalla quale emerge che l'uso di *gli* al posto di *le* è generalmente rifiutato dai partecipanti perché percepito come errato dalla loro sensibilità sociolinguistica;

- il *che* polivalente con valore di pronome relativo generico (*la persona che ho parlato ieri*) con valore di subordinante generico (causale, concessivo ecc., ad esempio, *Comprali che ce li mangiamo subito*), che rientra tra i fenomeni privi di un apporto funzionale alla lingua segnalati da Prandi (2020) ad eccezione del valore temporale (*il giorno che ti ho incontrato*), che può essere accettato come forma standard (D'Achille, 1990, p. 212; Prandi, 2020). Ciò trova una parziale conferma nella ricerca di Ondelli e Romanini (2018), in cui gli intervistati rifiutano in blocco l'intero fenomeno, non solo nei suoi usi prettamente substandard (ossia, come pronomi relativi generici), ma anche nel suo valore temporale, benché considerato ormai accettabile nello standard. Rispetto a quest'ultimo punto è necessario, però, tenere in considerazione il fatto che la maggioranza degli intervistati che lo sostituiscono con *in cui* rientra nella fascia di età compresa tra i 50 e i 65 anni. L'uso del *che* in funzione consecutivo-esplicativa è segnalato solo da un numero limitato di partecipanti al sondaggio;
- costruzioni con uso pleonastico della particella pronominale, tipo *a me mi piace leggere* (che rientrano nel caso della dislocazione dell'elemento tematico a sinistra della parte rematica della frase) e l'uso combinato di *ne* e del pronome relativo (esempio: *il libro del quale ne abbiamo appena discusso*), quest'ultimo segnalato da Prandi (2020) tra i fenomeni privi di un apporto funzionale alla lingua;
- *ci* attualizzante con il verbo *avere* seguito da un sostantivo in posizione di oggetto (*c'ho freddo, c'ho caldo, c'ho fame, c'ho sete, non c'ho tempo*) in cui l'uso del *ci* è ridondante e non è necessario alla codifica del possesso, che già è espresso da *avere*. Questo uso del *ci* rientra tra i fenomeni privi di un apporto funzionale alla lingua segnalati da Prandi (2020). A questo proposito, nella ricerca di Ondelli e Romanini (2018) il *ci* attualizzante è sempre rifiutato in combinazione con il verbo *avere* (esempio: *Io c'ho un problema grosso*);
- l'uso dell'indicativo al posto del congiuntivo nelle frasi subordinate, siano esse complete o marginali. Le ragioni della sua classificazione nel substandard e le conseguenti riflessioni sull'insegnamento dell'italiano verranno illustrate qui di seguito.

Il congiuntivo è un tratto grammaticale che contribuisce fortemente alla fisionomia dell'italiano e che innalza il livello dell'espressione, ma che nelle frasi subordinate non ha un valore funzionale, poiché, come già esposto, il contenuto della non-realtà o della realtà è affidato, rispettivamente, al significato del verbo reggente (*Mi spiace che ti abbiano rubato la bici*) – o della congiunzione (*sebbene piova/piove*) (Prandi, 2002). In questi casi cade, dunque, il criterio funzionale.

Tuttavia, a consigliare di mantenere l'uso del congiuntivo è il criterio della censura sociale, che emerge in maniera particolarmente evidente nel caso dei verbi fattivi (ad esempio, verbi di atteggiamento o giudizio come *Mi dispiace che il tempo si sia guastato*; *Mi sorprende che Piero abbia superato l'esame*): se al posto del congiuntivo si usasse l'indicativo, non si perderebbe il valore modale della frase, perché legata al significato intrinseco del verbo principale, ma si otterrebbe una formulazione di registro sociolinguistico nettamente più basso. Gli unici casi in cui il congiuntivo si basa su una scelta affidata al parlante si verificano in presenza di verbi come *dire, credere, pensare*. Qui la scelta può essere motivata dalla consapevolezza di utilizzare o meno il congiuntivo per mostrare di aderire o di distaccarsi a ciò che si pensa o si dice, per cambiare il registro stilistico della frase o, più banalmente, a causa di una scarsa conoscenza dei meccanismi della lingua.

Infine, integrando questa classificazione con i dati della ricerca di Ondelli e Romanini (2018), l'alternanza indicativo/congiuntivo è uno dei fenomeni linguistici per i quali le valutazioni sull'accettabilità da parte degli intervistati sono molto variabili. In sintesi, la sostituzione dell'indicativo al congiuntivo viene percepita come una variazione di registro, nella cui valutazione si richiede di tenere conto del contesto sintattico (in interrogative indirette, dopo verbi che esprimono incertezza, con verbi di opinione e nelle relative restrittive).

Riassumendo, per quanto riguarda la sostituzione dell'indicativo al congiuntivo, in assenza di chiari vantaggi funzionali tanto della forma standard quanto della sua alternativa non-standard, la conservazione della forma standard è preferibile all'adozione di una forma che espone il parlante alla censura della comunità. Nell'insegnamento dell'italiano andrebbero esplicitati quei casi, sopra elencati, in cui lo standard ammette la scelta tra indicativo e congiuntivo come se fosse funzionale a esprimere l'opposizione certezza/possibilità o realtà/irrealtà. In que-

sto contesto andrebbe precisato che è possibile imbattersi, nei registri più informali, nell'uso dell'indicativo al posto del congiuntivo, che si riallaccia alla tendenza generale dell'italiano parlato e dello scritto come mimesi del parlato a ridurre le funzioni del congiuntivo in favore dell'indicativo. Tuttavia, questo uso indiscriminato dell'indicativo, oltre a portare ad un abbassamento del registro linguistico e alla percezione di trascuratezza formale (voluta o meno che sia), in alcuni casi specifici può causare ambiguità (si vedano i casi delle relative restrittive).

Fenomeni linguistici entrati in passato nello standard: per quanto riguarda tutti gli altri fenomeni linguistici tratti dalla lista di Sabatini (1990) che non sono stati inseriti nelle due precedenti categorie, il loro uso non è ormai più percepito come in contrasto con la norma grammaticale, né è più messa in discussione la loro appartenenza allo standard:

- *lui, lei, loro* in funzione di soggetti;
- la combinazione di una preposizione e di un articolo partitivo, ossia l'impiego del partitivo preceduto da preposizione (*con degli amici*);
- il connettivo *per cui* (derivato per ellissi da *motivo, fatto, ragione per cui*, oppure per generalizzazione di casi con normali antecedenti nominali) in funzione di congiunzione causale-consecutiva;
- il soggetto in posizione rematica rispetto al predicato (*parla lui*), come struttura funzionale alla focalizzazione del soggetto (esempio: in *Il treno arriva, il treno* è tema: stiamo parlando del treno vs *arriva il treno*: il treno è fuoco);
- il semplice *cosa?* interrogativo diretto e indiretto, ossia il suo uso al posto di *che cosa?*;
- forme d'ingresso, come i connettivi testuali come *allora, comunque* e le congiunzioni *ma, e, o* iniziali di discorso o di enunciato;
- il *che* polivalente con valore temporale (eventualmente accettabile nello standard, D'Achille, 1990);
- la concordanza a senso tra il soggetto collettivo singolare seguito da un'espressione partitiva e il predicato plurale (*un milione di elettori non hanno votato*);
- *ci* attualizzante con il verbo *avere* seguito da un pronome diretto atono in posizione di oggetto (l'uso delle forme *ce l'ho, ce li ho, ce*

le ho – al posto di *l'ho*, *li ho*, *le ho* – è ormai consolidato e indiscusso e poggia le sue basi sull'uso originario di *ci* come avverbio di luogo). Tale accettazione riguarda solo il pronome atono e non il nome focale o sintagma nominale / pronome tonico. Ciò è dovuto, probabilmente, al fatto che il pronome clitico è – potremmo dire – foneticamente evanescente e si lega per eufonia al *ci*.

2.5. Conclusioni sull'applicazione dei criteri della funzionalità e dell'accettabilità sociale

Partendo dai criteri della funzionalità e dell'accettabilità sociale è possibile determinare l'appartenenza delle forme linguistiche dell'uso medio a uno standard condiviso e insegnabile. Questo schema può essere applicato non solo ai fenomeni finora considerati, ma anche a qualunque innovazione linguistica che entri nell'uso e si candidi a entrare nello standard.

Le riflessioni finora fatte non negano che l'elaborazione di concetti come *italiano neostandard* o *italiano dell'uso medio* abbia dato risposta ad un'esigenza esistente da secoli e più volte affiorata nel corso della storia della lingua italiana, ossia la revisione dello standard ereditato dai secoli passati, quando il pieno accesso alla lingua italiana era limitato a poche persone colte, in ambiti d'uso molto ristretti. Tuttavia, come già anticipato, l'inclusione nelle liste di tutto ciò che si distaccava dalle norme grammaticali ha portato, di fatto, all'accostamento di risorse funzionali essenziali a forme che invece non solo non apportano alcun vantaggio funzionale, ma producono una mera caduta di registro, esponendo l'apprendente alla censura sociale. Se da un lato l'italiano dell'uso medio può essere ormai considerato lo standard dal punto di vista della frequenza statistica, il giudizio diventa più difficile quando si prendono in considerazione i concetti di correttezza grammaticale e neutralità di registro.

Di conseguenza, un possibile approccio nell'insegnamento della lingua potrebbe essere quello di affrontare i fenomeni grammaticali “ritrovati” – per usare la scelta lessicale di Sabatini – presentandoli come diffusi tra i parlanti nativi e come segnali di un cambiamento tendenziale dell'italiano (in accordo con quanto già proposto da Sobrero, 1992b), avendo cura, però, di distinguere i fenomeni che deviano dalla norma grammaticale tradizionale dando un effettivo apporto funzionale alla

lingua e quelli che, ferma restando la loro diffusione, di fatto non apportano alcun vantaggio funzionale. Questi ultimi rappresentano una deviazione del tutto arbitraria rispetto alla norma, il cui uso produce solo un calo di registro, esponendo quindi il parlante alla censura sociale (Prandi, 2020). In caso di assenza di chiari vantaggi funzionali tanto della forma standard quanto della sua alternativa non-standard (si veda il caso già citato dell'uso del congiuntivo vs. indicativo), la conservazione di una forma standard è preferibile all'adozione di una forma che espone il parlante alla censura della comunità.

Concludendo, dopo aver fornito una descrizione dei tratti linguistici della lista di Sabatini del 1990 e aver discusso la questione della necessità di un modello linguistico di riferimento e delle implicazioni che ciò ha per la didattica dell'italiano, nel prossimo capitolo verranno illustrate le ragioni per le quali i corpora possono rappresentare delle valide risorse per l'apprendimento della lingua e per uno studio sulla diffusione dei tratti linguistici dell'italiano dell'uso medio.

3. I CORPORA NELLA DIDATTICA DELLE LINGUE

Dopo aver descritto nei capitoli precedenti le varietà dello spazio linguistico italiano, dando particolare rilievo all'italiano dell'uso medio, e dopo aver fornito una descrizione dei tratti linguistici che lo caratterizzano, in questo capitolo verranno illustrati alcuni concetti base della linguistica dei corpora e le ragioni per le quali i corpora possono rappresentare delle valide risorse per l'apprendimento della lingua. A questo proposito verranno presentati alcuni ambiti di applicazione dei corpora alla didattica delle lingue, dedicando particolare attenzione all'uso dei corpora per la realizzazione di materiali didattici.

3.1. Alcune nozioni di linguistica dei corpora

Il punto di partenza sarà chiarire cosa siano i corpora, per i quali Baroni (2010)¹ elabora la seguente definizione:

¹ www.treccani.it/enciclopedia/corpora-di-italiano_%28Enciclopedia-dell%27Italiano%29.

I cosiddetti *corpora* (sing. *corpus*) linguistici sono collezioni, per lo più di grandi dimensioni, di testi orali o scritti prodotti in contesti comunicativi reali (per es., registrazioni di discorsi o articoli di giornale), conservati in formato elettronico e spesso corredati di strumenti di consultazione informatici.

Tale definizione si richiama a quella precedentemente elaborata da McEnery e Wilson (1996):

In principle, any collection of more than one text can be called a corpus... But the term “corpus” when used in the context of modern linguistics tends most frequently to have more specific connotations than this simple definition provides for. These may be considered under four main headings: sampling and representativeness, finite size, machine-readable form, a standard reference. (McEnery, Wilson, 1996, p. 21 in Kilgariff, Grefenstette, 2003, p. 333)

Dunque, se in senso ampio qualunque raccolta di testi può essere considerata un corpus, nel contesto della linguistica moderna una raccolta di testi è definita un corpus se riveste il ruolo di oggetto di studio linguistico o letterario (“A corpus is a collection of texts when considered as an object of language or literary study”, Kilgariff, Grefenstette, 2003, p. 334) e se è stata realizzata sulla base di alcuni criteri, quali: campionamento e rappresentatività, poiché i testi devono rappresentare un campione dell’uso di una determinata lingua o di una sua varietà; dimensioni definite, ossia la quantità indicativa di testi da includere nel corpus, che viene stabilita nelle fasi iniziali che precedono la realizzazione del corpus stesso. Un’eccezione in tal senso è rappresentata dai corpora di monitoraggio (cfr. sezione 3.3), i quali vengono periodicamente aggiornati con nuovi testi; formato elettronico, poiché tutti i corpora moderni sono in formato digitale e, infine, il concetto del riferimento standard: un corpus deve rappresentare non solo un punto di riferimento per lo studio della lingua o varietà che rappresenta, ma anche un modello rispetto al quale è possibile fare un confronto sia tra

corpora diversi, sia con gli studi successivi su un determinato ambito: una volta definita la metodologia impiegata per la raccolta e l'analisi dei dati di un certo corpus, gli eventuali nuovi risultati ottenuti su argomenti correlati possono essere confrontati direttamente con i risultati già pubblicati negli studi precedenti, senza la necessità di un ricalcolo.

Tornando al legame esistente tra corpora, campionamento e rappresentatività, sempre Baroni (2010) chiarisce:

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. Per es., analizzando un *corpus* di articoli di giornale pubblicati in Italia negli anni Novanta del Novecento possiamo formulare generalizzazioni che, se il *corpus* è sufficientemente grande e vario, in genere sono valide per i testi giornalistici italiani del periodo scelto.

Dunque, un corpus mira idealmente a essere un campione rappresentativo di una certa lingua, intesa come produzione parlata e scritta. La conseguenza è che le proprietà linguistiche dei testi che contiene (ad esempio, la diffusione delle unità lessicali o delle strutture sintattiche, ecc.) possono essere generalizzate all'intera popolazione linguistica, ossia ad una certa lingua, o varietà linguistica, nel suo insieme. Nello specifico:

In corpus linguistics, the population is the entirety of a given language, or of a given variety of language. We attempt to make generalisations about a population, say of all English newspaper language or of all spoken British English, based on a corpus that contains only a sample of that population. (McEnery, Hardie, 2012, p. 248)

Un campione è, invece, solo una sezione della popolazione:

The word sample may also be used in its statistical sense by corpus linguists. In this latter sense, it means a group of cases taken from a population that will, hopefully, represent that population such that findings from the sample can be generalised to the population. (ivi, p. 250)

Pertanto, quanto più un corpus sarà esteso e composto da fonti variegate, tanto più sarà rappresentativo di una certa lingua e, di conseguenza, la sua consultazione permetterà di osservare l'uso effettivo di quella lingua, verificandone tendenze generali su base statistica (Lüdeling, Kytö, 2008-2009). La selezione di campioni di testi rappresentativi di una lingua dipende dal tipo di lingua che il corpus deve rappresentare:

A corpus seeks to represent a language or some part of a language. The appropriate design for a corpus therefore depends upon what is meant to represent. Representativeness of the corpus, in turn, determines the kind of research questions that can be addressed and the generalizability of the results of the research. (Bieber *et al.*, 1998, p. 246)

La costruzione di un corpus dipende, dunque, da ciò che esso si prefigge di rappresentare (ad esempio, la varietà scritta/parlata, un registro particolare ecc.) e la sua rappresentatività determina a sua volta il tipo di domande di ricerca per le quali il corpus può essere utilizzato (in un corpus rappresentativo della lingua parlata potrà essere ricercato, ad esempio, l'uso degli intercalari nel parlato). A questo proposito, per quanto riguarda gli scopi con cui vengono utilizzati i corpora, Kilgariff e Grefenstette (2003: 343) citano i seguenti ambiti di applicazione:

One use of a corpus is to extract a language model: a list of weighted words, or combinations of words, that describe (1) how words are related, (2) how they are used with each other, and (3) how common they are in

a given domain. Language models are used in speech processing to predict which word combinations are likely interpretations of a sound stream, in information retrieval to decide which words are useful indicators of a topic, and in machine translation to identify good translation candidates.

Questa definizione lascia intendere che attraverso l'uso dei corpora sia possibile ricavare un modello del funzionamento della lingua semplicemente osservando, ad esempio, quali siano le parole più utilizzate, come le parole si combinino le une alle altre e quali combinazioni di parole siano più comuni. Tali modelli linguistici trovano applicazione in vari ambiti: citandone a titolo di esempio alcuni che hanno delle analogie con la presente ricerca, l'estrazione di elementi linguistici dai corpora è applicata anche agli studi di lessicografia (cfr. Wunderlich, 2018) e fraseologia (cfr. Cosme, Gilquin, 2008; Coxhead, 2008; Mudraya *et al.*, 2008; Babych, 2012; Orenha-Ottaiano, 2012; Seracini, 2018 e gli studi di Pecman, 2008, Wible, 2008 e Alhujaylan, 2014, che si collocano in entrambi gli ambiti sopra nominati) e come supporto nell'apprendimento delle lingue (cfr. Curado-Fuentes, 2012; Charkova, Charkova, 2014; Vališová, 2014; Brezina, Gablasova, Marín Cervantes, 2018; Curry, Goodman, 2018; Hadley, Hadley, 2018; Picoral *et al.*, 2018; Tono, 2018; in particolare per l'italiano L2/LS: Corino, Marellò, 2009; Merlo, 2009; Kennedy, Miceli, 2010; Guidetti, Lenzi, Storch, 2012; Lyding *et al.*, 2013 e Ivanovska-Naskova, 2018).

L'uso dei corpora come supporto nell'apprendimento delle lingue verrà indagato nella sezione 3.6, in cui verranno presentati degli studi che propongono delle strategie per favorire l'uso guidato o diretto dei corpora da parte degli apprendenti (dalle operazioni più semplici, come le liste di frequenza e la ricerca delle concordanze, alla padronanza di modalità di ricerca più avanzate) allo scopo di introdurre i corpora come strumento di studio per formulare e testare ipotesi sul funzionamento della lingua.

La consultazione di una grande quantità di dati contenuta in un corpus, come descritto da McEnery e Hardie (2012, p. 2), è resa possibile solo attraverso alcuni strumenti informatici. A questo proposito, come verrà qui di seguito illustrato nella citazione di McEnery e Hardie

(2012), sebbene nel corso della storia della linguistica dei corpora vi siano stati casi di risorse in formato cartaceo, o le cui dimensioni ridotte abbiano consentito una consultazione non elettronica, attualmente, come ad esempio nella definizione di Baroni (2010) sopra menzionata, si fa riferimento principalmente a testi in formato elettronico:

It should be noted that the term text here denotes a file of machine-readable data. [...] Although we have said that corpus linguistics always uses machine-readable text, in fact, historically, much work was undertaken on corpora held in paper form; for example, Fries (1952) produced a grammar of English based upon such a corpus. Also, while it is true that much research using corpus methods [...] uses corpora of millions of words, there are others studies such as those of Ghadessy and Gao (2001) and McEnery and Kifle (2001) which, appropriately, use smaller, specialised corpora that might conceivably have been analysed by hand and eye. Nonetheless, despite the exceptions, the generalisations above characterise much of the work that can reasonably be described as corpus linguistics. (McEnery, Hardie, 2012, pp. 2-3)

Vi sono vari strumenti informatici che mettono a disposizione diverse modalità di ricerca. Limitando gli esempi alle sole funzioni utilizzate nell'ambito di questa ricerca, si ricordano la *concordanza* e la *lista di frequenza*. Queste ultime rappresentano, come affermato da McEnery e Hardie (*ibidem*), due tipi di analisi che rivestono pari importanza nella linguistica dei corpora, ossia l'analisi qualitativa e quantitativa, alle quali si farà riferimento (cfr. capitolo 5) nell'ambito di questa ricerca.

Nello specifico, con la funzione concordanza la parola cercata può essere visualizzata nel contesto in cui è stata utilizzata (dalla singola frase, fino ad allargare il contesto ad un intero paragrafo o all'intero testo in cui è contenuta, a seconda delle impostazioni dello strumento di ricerca in uso), mentre con la *lista di frequenza* è possibile ottenere una lista di tutte le parole contenute in un corpus, per ciascuna delle quali viene indicata la frequenza, ossia il numero di volte in cui ricorre.

Ad esempio, i corpora possono essere utilizzati per indagare la frequenza d'uso di una parola (ad esempio, in un dato contesto oggetto di ricerca, quanto è frequente la parola *questione* rispetto a *problema*?), per individuare le collocazioni tipiche in cui questa ricorre (ad esempio, a quali verbi si accompagna la parola *questione*? *Sollevare una questione*, *sottoporre una questione a qualcuno*, ecc.) e per cogliere la prosodia semantica² sulla base del contesto d'uso. Un corpus elettronico è in grado di dare risposta a queste domande tramite alcune modalità di ricerca, tra cui quelle sopra citate. Dopo aver, dunque, definito cosa sia un corpus e avere descritto in linee generali alcuni possibili usi, nella sezione successiva si illustreranno brevemente i due principali metodi di applicazione dei corpora allo studio del linguaggio, al fine di identificare il metodo più adeguato alla realizzazione degli obiettivi della presente ricerca.

3.2. Metodo per l'identificazione dei fenomeni tramite corpus

Come illustrato nella sezione precedente, la linguistica dei corpora è un ambito di ricerca che si è sviluppato negli ultimi anni del ventesimo secolo all'interno degli studi sulla variazione e l'uso della lingua (Biber, Conrad, Reppen, 1998; Biber, 2012) impiegando principalmente due metodi di ricerca.

Nel primo, definito *corpus-based* (Tognini-Bonelli, 2001), ossia basato su corpus, i dati estratti da un corpus vengono utilizzati per verificare una teoria, con l'obiettivo di convalidarla, confutarla o perfezionarla. In questo approccio la linguistica dei corpora è intesa, dunque, come metodo per lo studio del linguaggio.

The term corpus-based is used to refer to a methodology that avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were

² Descrive il modo in cui alcune parole possono essere percepite con connotazioni positive o negative, poiché tendono a ricorrere assieme ad altre parole che appartengono ad una determinate area semantica. "[A] word may be said to have a particular semantic prosody if it can be shown to co-occur typically with other words that belong to a particular semantic set" (Hunston, Francis, 2000, p. 137).

formulated before large corpora become available to inform language study. [...] Corpus-based linguists adopt a 'confident' stand with respect to the relationship between theory and data in that they bring with them models of language and descriptions which they believe to be fundamentally adequate, they perceive and analyse the corpus through these categories and sieve the data accordingly. (ivi, pp. 65-66)

Allo studio del linguaggio con approccio *corpus-based* si oppone l'approccio *corpus-driven*, ossia guidato dal corpus, di tipo induttivo. Non vi è una teoria a priori; i dati sono osservati e studiati nel loro contesto, in modo tale che "parlino da sé" e sia possibile derivarne una teoria sull'uso della lingua. Tale approccio rifiuta la concezione della linguistica dei corpora come metodo e afferma che la sola fonte per la teoria del linguaggio sia il corpus stesso:

[i]n a corpus-driven approach the commitment of the linguist is to the integrity of the data as a whole, and descriptions aim to be comprehensive with respect to corpus evidence. The corpus, therefore, is seen as more than a repository of examples to back pre-existing theories or a probabilistic extension to an already well-defined system. The theoretical statements are fully consistent with, and reflect directly, the evidence provided by the corpus. [...] The theory has no independent existence from the evidence and the general methodological path is clear: observation leads to hypothesis leads to generalisation leads to unification in theoretical statement. (Tognini-Bonelli, 2001, pp. 84-85)

La dicotomia *corpus-based* e *corpus-driven* divide la ricerca in due approcci distinti, sebbene McEnery e Hardie (2001, p. 151) rifiutino di considerarla una vera e propria distinzione binaria, sostenendo che vi siano casi in cui l'utilizzo del corpus si collochi a metà strada tra entrambi gli approcci ("the corpus-based versus corpus-driven distin-

ction implies a dichotomy where there is actually a sliding scale”) e che a separare tali approcci vi sia in realtà solo la diversa concezione del *ruolo* ricoperto dal corpus (come metodo per verificare/confermare/confutare una teoria preesistente o come fonte sulla cui base formulare una teoria) piuttosto che il loro *uso* effettivo:

we would argue that the different schools of corpus linguistics are not reliably distinguished – or, alternatively, that their nature is not optimally communicated – by the corpus-based versus corpus-driven distinction, or by the ‘top-down’ versus ‘bottom-up’ distinction with which it is often equated. Rather, it is the contrasting stances on the conceptual status of the corpus and of corpus linguistics – as having theoretical status versus as a linguistic methodology – that truly separates the two schools. (*ibidem*)

Per tali ragioni McEnery e Hardie trovano più adeguata la distinzione *corpus-linguistics-as-method* vs *corpus-linguistics-as-theory* rispetto a *corpus-based* vs *corpus-driven* (ivi, p. 6).

Sulla base di tali riflessioni, l’approccio da utilizzare per la presente ricerca è di tipo *corpus-based* o *corpus-linguistics-as-method*, ove un corpus di lingua italiana è utilizzato come risorsa di verifica dell’ipotesi circa l’effettiva reperibilità – e in quale misura – dei tratti dell’italiano dell’uso medio nei testi scritti di media formalità. Dopo aver illustrato i due principali metodi di applicazione dei corpora allo studio del linguaggio, nella sezione successiva si illustreranno brevemente le principali tipologie di corpora esistenti.

3.3. Tipologie di corpora esistenti

Una prima distinzione tracciabile tra le diverse tipologie di corpora esistenti si basa sulla lingua dei testi raccolti. Nello specifico, esistono corpora monolingui e corpora bilingui (o multilingui), i quali sono ulteriormente distinguibili in corpora paralleli e corpora comparabili. I corpora monolingui contengono testi redatti in un’unica lingua, il che non esclude che possano rappresentarne le diverse varietà esistenti,

ad esempio nel caso di un corpus di lingua inglese che includa la varietà britannica, americana ecc.: “So the International Corpus of English [...], for example, is a large monolingual corpus – it represents one language, English, though it allows linguists to compare and contrast a number of international varieties of that language” (McEnery, Hardie, 2012, p. 18).

Si definiscono, invece, corpora multilingui le risorse che coinvolgono tre o più lingue, mentre le risorse che coinvolgono solo due lingue vengono convenzionalmente definite bilingui (ivi, p. 19).

Inoltre, i corpora che coinvolgono più di una lingua possono essere ulteriormente classificati come segue, sulla base delle teorie di Baker (1993, p. 248; 1995; 1999), McEnery e Wilson (2001, p. 70) e Hunston (2002, p. 15) riportate in McEnery e Hardie (2012, p. 19): si definiscono corpora paralleli quelle risorse che contengono dei testi sia nella loro lingua originaria, sia nelle relative traduzioni in una o più lingue. Nel caso in cui le unità linguistiche dei testi (ad esempio le frasi) nella loro lingua originaria vengano collegate alle unità linguistiche dei testi tradotti, le risorse vengono definite corpora paralleli allineati. Ad esempio, l’Opus Corpus³ è il risultato di un progetto di raccolta e allineamento di testi tradotti tratti dal web, con l’aggiunta di annotazioni linguistiche, allo scopo di fornire un corpora paralleli liberamente accessibili, come ad esempio l’EuroParl⁴, costituito dagli atti del Parlamento Europeo dal 1996 al 2011.

Si parla, invece, di corpora paragonabili nel caso in cui vi siano due o più corpora monolingui (ad esempio: un corpus monolingue italiano e un corpus monolingue di qualunque altra lingua, tra i quali non vi sia un rapporto “testo di partenza – traduzione”) che siano tra loro paragonabili sulla base di variabili quali ad esempio la tipologia testuale, il dominio/argomento, la data di produzione, la varietà linguistica di riferimento, il registro o altro. I corpora paragonabili consentono di confrontare due o più lingue rispetto allo stesso genere testuale o tema. Un’ulteriore distinzione tracciabile tra le diverse tipologie di corpora è quella tra corpora di riferimento e corpora specialistici:

³ <http://opus.nlpl.eu>

⁴ <http://statmt.org/europarl>.

Un corpus di riferimento si propone come campione rappresentativo di una lingua in tutti i suoi aspetti: scritto e parlato, registri e varietà d'uso, ecc. Questo obiettivo però non è mai pienamente raggiungibile, per ragioni sia pratiche (la raccolta dei dati richiederebbe costi enormi) che teoriche (quanti e quali registri e varietà andrebbero campionati? in che proporzione? ecc.). Tuttavia, più un corpus è composto da fonti variegata, più esso sarà rappresentativo di una lingua in generale. (Baroni, 2010)

La denominazione dipende, dunque, dal fatto che tale tipologia di corpora si propone come risorsa di riferimento per la descrizione di una certa lingua ed è, dunque, trasversale rispetto alle diverse varietà che la compongono. Ciò non esclude che un corpus di riferimento possa essere organizzato in vari sottocorpora specialistici, rappresentativi delle diverse varietà della lingua.

A questo proposito, i corpora specialistici hanno, appunto, l'obiettivo di rappresentare particolari varietà ristrette di una lingua, come ad esempio i corpora per lo studio di domini linguistici settoriali (ad esempio, il gergo di determinati settori professionali) o i corpora di testi che rientrano esclusivamente in una specifica tipologia testuale, eccetera. Di conseguenza, poiché il loro scopo è di tracciare un quadro di una varietà linguistica specifica, i corpora specialistici sono caratterizzati da fonti meno variegata rispetto ai corpora di riferimento e da un'elevata omogeneità linguistica e sono queste due caratteristiche che ne determinano il grado di rappresentatività.

Ulteriori tipologie esistenti sono i corpora sincronici, che contengono testi appartenenti tutti ad una stessa finestra temporale e, pertanto, descrivono un particolare stadio del linguaggio; al contrario i corpora diacronici contengono testi che appartengono a diverse finestre temporali e, quindi, descrivono i cambiamenti di una lingua attraverso il tempo. Infine, vi è la distinzione tra corpora di monitoraggio e corpora bilanciati.

The monitor corpus approach, proposed most notably by John Sinclair, seeks to develop a dataset which grows

in size over time and which contains a variety of materials. The relative proportions of the different types of materials may vary over time. (McEnery, Hardie, 2012, p. 6)

Dunque, un corpus di monitoraggio è una raccolta di testi che cresce nel tempo attraverso l'introduzione di nuovi testi selezionati secondo gli stessi criteri usati per costruire la raccolta iniziale. Di conseguenza, poiché vengono aggiornati nel corso del tempo, i corpora di monitoraggio sono adatti a seguire le evoluzioni che avvengono in una lingua.

In contrast to monitor corpora, balanced corpora, also known as sample corpora, try to represent a particular type of language over a specific span of time. In doing so, they seek to be balanced and representative within a particular sampling frame which defines the type of language, the population that we would like to characterise. The population is the notional space within which language is being sampled. So, for example, if we want to look at the language of service interactions in shops in the UK in the late 1990s, the sampling frame is clear – we would only accept data into our corpus which represents service interactions in UK shops in the 1990s. (ivi, p. 8)

I corpora bilanciati (*balanced corpora* o *sample corpora*) includono, dunque, testi che rappresentano una varietà linguistica specifica “fotografata” in un preciso confine temporale. A tale scopo, per costruire un corpus bilanciato è necessario definire accuratamente la popolazione linguistica di interesse, tracciandone i confini spaziali e temporali (quali testi sono inclusi o esclusi dalla popolazione) e la tipologia dei testi (l'articolazione in strati della popolazione). La differenza tra corpora di monitoraggio e corpora bilanciati può essere sintetizzata come segue:

Two broad approaches to the issue of choosing what data to collect have emerged: the monitor corpus ap-

proach (see Sinclair, 1991, pp. 24-6), where the corpus continually expands to include more and more texts over time; and the balanced corpus or sample corpus approach (see Biber, 1993 and Leech, 2007), where a careful sample corpus, reflecting the language as it exists at a given point in time, is constructed according to a specific sampling frame. (ivi, p. 6)

A conclusione di questa panoramica sulle principali tipologie di corpora esistenti, nella sezione successiva verrà descritto l'uso del web come corpus.

3.4. Il web come corpus

Kilgarriff e Grefenstette (2003) hanno proposto l'approccio del *Web as a Corpus*, ossia dell'uso del web come un enorme corpus di testi. "The web is clearly a multilingual corpus" (Kilgarriff, Grefenstette, 2003, p. 334): il web può, infatti, essere utilizzato sia per ottenere dati di tipo quantitativo sulla frequenza d'uso delle espressioni linguistiche tramite i motori di ricerca (ad esempio, quante pagine web si ottengono attraverso la ricerca di una data parola), sia come una fonte inesauribile di testi per costruire corpora. In quest'ottica, il web è un corpus che presenta numerosi vantaggi: è una risorsa enorme, facilmente accessibile, oltre che gratuita, e variegata negli stili, nelle tipologie testuali e nei contenuti.

Sebbene tanto per i corpora quanto per il web il punto di partenza sia il medesimo, ossia un'enorme raccolta di dati in continua crescita, in realtà l'approccio del *Web as a Corpus* presenta alcuni limiti (Kilgarriff, Grefenstette, 2003, pp. 344-345), come evidenziato anche da McEnery e Hardie (2012, p. 7).

Innanzitutto, nel caso in cui per *web as a corpus* si intenda nello specifico l'interrogazione del web tramite l'uso di motori di ricerca commerciali⁵, vi è la mancanza di replicabilità delle ricerche: i risultati

⁵ In contrapposizione al caso in cui i testi vengono scaricati dal web e poi processati per la costruzione di un corpus da consultare tramite specifici programmi di interrogazione.

delle ricerche effettuate tramite motori di ricerca commerciali, come ad esempio *Google*, *Yahoo!* ecc., non sono sempre affidabili, poiché dipendono dalle specifiche caratteristiche di tali motori. Uno stesso tipo di ricerca può, infatti, condurre a risultati diversi sia tramite motori di ricerca differenti, sia tramite un unico motore di ricerca a causa degli algoritmi di indicizzazione e di recupero delle pagine e della presenza di duplicati, ragion per cui i dati di frequenza risultano inaffidabili e non paragonabili. Inoltre, i risultati corrispondono al numero di pagine web reperite (“pagina contenente *x*”) e non all’effettiva frequenza dell’espressione linguistica ricercata.

Sebbene i motori di ricerca siano degli eccellenti strumenti dal punto di vista della loro funzione primaria, ossia la ricerca di informazioni, essi possono risultare inadeguati alle esigenze dei linguisti:

The default means of access to the Web is through a search engine such as Google. Although the Web search engines are dazzlingly efficient pieces of technology and excellent at the task they set for themselves, for the linguist they are frustrating. (Kilgarrieff, Grefenstette, 2003, p. 344)

Le ragioni citate dai due studiosi sono le seguenti (ivi, p. 345). Il web presenta l’impossibilità sia di fare ricerche in sottoinsiemi di testi, poiché il contenuto dei testi presenti nel web non è suddiviso per generi testuali, sia di fare ricerche tramite criteri prettamente linguistici, ad esempio fare ricerche per parti del discorso (sostantivi, verbi, aggettivi, ecc.) e lemmi ecc., senza contare che il contesto fornito per ciascuna occorrenza è molto ridotto (circa 10 parole su *Google*). Infine, il web include testi qualitativamente disomogenei dal punto di vista dell’accuratezza nella pianificazione testuale, molti dei quali non sono privi di errori di ogni genere.

Dunque, sebbene il web sia di fatto un immenso corpus (“The answer to the question “Is the web a corpus?” is yes.” (ivi, p. 334)), le caratteristiche dei motori di ricerca non lo rendono adeguato ad un suo utilizzo come corpus linguistico, ma solo come fonte di testi a cui attingere per la costruzione di corpora da consultare tramite specifici programmi di interrogazione. I vantaggi principali di questo approccio

sono, ad esempio, l'offerta illimitata di testi attuali e appartenenti a generi e domini diversi, in cui sono rintracciabili i vari registri linguistici, la disponibilità immediata dei testi e il fatto che siano reperibili in formato digitale.

3.5. Corpora nella linguistica cognitiva e nella *construction grammar*

Sulla base di quanto illustrato, i corpora rappresentano delle risorse importanti per la conoscenza della lingua, la cui consultazione tramite specifici programmi di interrogazione permette, ad esempio, di valutare la frequenza d'uso di elementi lessicali, costruzioni grammaticali e molto altro, fornendo preziose fonti di informazioni sull'evoluzione e lo stato corrente della lingua in uso. A questo proposito, si ricorda come l'utilizzo di corpora come base di dati linguistici reali, utili all'analisi e alla teorizzazione linguistica, venga applicato in ambiti quali la linguistica cognitiva e la *construction grammar*. Come evidenzia Bazzanella (2014), la linguistica cognitiva si affianca ad un ricorso sempre maggiore ai corpora, poiché le strutture linguistiche emergono dall'uso stesso della lingua, secondo i cosiddetti modelli basati sull'uso (*usage-based models*):

An important consequence of adopting the usage-based thesis is that there is no principled distinction between knowledge of language and use of language (competence and performance in Generative Grammar terms), since knowledge of language *is* knowledge of how language is used. (Evans, 2007, p. 217)

Inoltre, come riportato da Newman (2011) in una riflessione sullo stato attuale della linguistica cognitiva, in particolare sulle sue metodologie, Geeraerts (2006, p. 29) fa riferimento alla tendenza crescente della linguistica cognitiva a sottolineare la sua essenza di linguistica basata sull'uso ("growing tendency of Cognitive Linguistics to stress its essential nature as a usage-based linguistics"), una tendenza che si inserisce chiaramente nella ricerca basata su corpora, come dimostrano alcune pubblicazioni che illustrano metodi basati su corpora in am-

bito cognitivo-linguistico, tra cui Gries, Stefanowitsch (2006), Stefanowitsch, Gries (2006), Lewandowska-Tomaszczyk, Dziwirek (2009). Analogamente, per quanto riguarda la *construction grammar*, Hilpert (2013) afferma che i dati recuperati dai corpora possono rivelare dettagli precisi riguardo ai cambiamenti linguistici di tipo costruttivo (ad esempio, i cambiamenti relativi alla frequenza di una costruzione linguistica, alla sua forma e alla sua funzione), facilitando l'analisi quantitativa e, nel contempo, mostrando come la linguistica dei corpora abbia influenzato in modo significativo la comprensione dei cambiamenti linguistici di tipo costruttivo (cfr. gli approcci *corpus-based* alla grammatica costruttiva in Gries, Yoon (2016); Hoffmann, Trousdale (2013) e Gries, Wulff (2005) sull'applicazione dei corpora allo studio delle costruzioni linguistiche degli apprendenti di una lingua straniera).

Dopo aver illustrato alcuni concetti base della linguistica dei corpora e le ragioni per le quali i corpora possano rappresentare delle valide risorse per la conoscenza della lingua, nella sezione successiva si illustreranno alcuni ambiti di applicazione dei corpora alla didattica delle lingue, dedicando particolare attenzione all'uso dei corpora per la realizzazione di materiali didattici.

3.6. Corpora e apprendimento delle lingue

In questa sezione verrà affrontato il tema dell'applicazione dell'uso dei corpora all'insegnamento delle lingue. Verranno esposti brevemente gli ambiti in cui tale applicazione ha luogo, dando particolare rilievo all'utilizzo dei corpora per il reperimento di testi da utilizzare per la progettazione di attività didattiche. A tale scopo verranno citati alcuni studi nei quali si propone l'uso dei corpora per la realizzazione di risorse lessicografiche e grammaticali a supporto dell'apprendimento linguistico; si escluderà invece dalla trattazione l'impiego dei corpora nella traduzione. Infine, si affronterà il tema dell'importanza dell'autenticità degli esempi linguistici impiegati nella didattica, in contrapposizione agli esempi creati *ad hoc*, e della varietà di lingua da trasmettere agli apprendenti.

L'applicazione dell'approccio *corpus-based* alla linguistica e all'insegnamento delle lingue ha assunto un'importanza sempre maggiore tra la

fine degli anni '80 e i primi anni '90 del secolo scorso (McEnery, Xiao, 2005, p. 364), con i primi studi sul tema di Sinclair (1991) e Johns (1991). Inoltre, a partire dal 1994 la serie internazionale di conferenze biennali *Teaching and Language Corpora* (TaLC), giunta al momento della scrittura alla sua quattordicesima edizione, testimonia il forte interesse per l'uso dei corpora nell'insegnamento delle lingue.

Nel corso degli anni, come affermato da Römer (2008), si è sviluppato dunque un rapporto bidirezionale tra la linguistica dei corpora e la glottodidattica: da un lato la linguistica dei corpora ha permesso di ricavare dati sull'uso della lingua e ha fornito strumenti tecnologici, quali programmi per la creazione di corpora, liste di frequenza, concordanze e corpora di grandi dimensioni, che sono utili a chi studia una lingua; dall'altro lato la glottodidattica ha incentivato, ad esempio, la produzione dei *learner corpora*, che comprendono produzioni orali e/o scritte degli apprendenti di una lingua straniera.

Con l'obiettivo di indagare il rapporto tra corpora e apprendimento, gli atti del convegno TaLC del 2002 (Aston, Bernardini, Stewart, 2004), furono divisi in tre possibili ambiti di applicazione dei corpora alla didattica: 1) *corpora by learners*: i corpora composti da materiali prodotti dagli apprendenti e finalizzati, ad esempio, allo studio dell'interlingua (Selinker, 1972) e dell'"errore", con l'obiettivo di identificare le caratteristiche della lingua degli apprendenti e utilizzarle come punti di partenza per riprogrammare l'attività didattica e favorire l'acquisizione linguistica; 2) *corpora for learners*: corpora progettati per permettere ai docenti e, più generalmente, a coloro che creano materiali didattici, di trasmettere una descrizione più accurata della lingua oggetto di studio e, pertanto, di decidere cosa debba essere appreso; 3) *corpora with learners*: i corpora sono considerati degli strumenti utili ad acquisire conoscenze e competenze linguistiche tramite il loro uso diretto da parte degli studenti, con o senza l'aiuto dei loro insegnanti. Tale tripartizione degli ambiti d'uso ricalca parzialmente quella elaborata da Leech (1997): a) uso indiretto dei corpora nell'insegnamento, che prevede la consultazione dei corpora al fine di ricavare esempi autentici e tipici dell'uso di elementi linguistici, nonché di informazioni sulla loro frequenza d'uso. Tutto ciò è finalizzato allo sviluppo di materiali e test linguistici e alla compilazione di pubblicazioni di riferimento, siano esse grammatiche o dizionari, poiché l'uso di corpora per-

mette di tenere conto delle variazioni di registro e delle differenze tra lingua scritta e parlata e, pertanto, di ottenere descrizioni grammaticali più accurate rispetto a quelle comunemente fornite dalle grammatiche (McEnery, Xiao, 2005); 2) uso diretto dei corpora nell'insegnamento, che si articola, secondo la tripartizione di Leech (1997) in *teaching about*, *teaching to exploit* e *exploiting to teach*:

“Teaching about” means teaching corpus linguistics as an academic subject like other sub-disciplines of linguistics such as syntax and pragmatics. [...] “Teaching to exploit” means providing students with “hands-on” know-how, as emphasized in McEnery, Xiao and Tono (2006), so that they can exploit corpora for their own purposes. Once the student has acquired the necessary knowledge and techniques of corpus-based language study, the learning activity may become student centred. “Exploiting to teach” means using a corpus-based approach to teaching language and linguistics courses (e.g., sociolinguistics and discourse analysis), which would otherwise be taught using noncorpus-based methods. (McEnery, Xiao, 2011, p. 370)

3) ulteriori sviluppi orientati all'insegnamento: corpora di linguaggi per scopi specifici (LSP), corpora incentrati sullo sviluppo della prima lingua (L1) e corpora degli apprendenti di seconda lingua (L2). Sulla base di tali suddivisioni delle possibili applicazioni dei corpora alla didattica, è possibile affermare che la presente ricerca, avente tra gli obiettivi l'uso di corpora come fonte da cui ricavare testi autentici sui quali incentrare la progettazione di materiali didattici, si colloca nell'ambito dell'uso indiretto dei corpora nell'insegnamento, secondo la terminologia di Leech (1997). Partendo da tale presupposto, nella sezione seguente verranno citati alcuni studi nei quali si affronta l'uso dei corpora per la realizzazione di risorse a supporto dell'apprendimento linguistico.

3.7. Uso dei corpora per la creazione di materiali didattici

Come già illustrato nella sezione precedente, a partire dalla metà degli anni '90 sono stati svolti numerosi studi circa l'uso con finalità didattiche di testi o singoli elementi linguistici (lessico, verbi, fraseologie ecc.) estratti dai corpora. In particolare, in questa trattazione si farà riferimento alle pubblicazioni contenute in Sinclair (2004), Aijmer (2009), Granger, Meunier (2008), Meunier, Granger (2008), Hinkel (2011) e agli atti dei convegni TaLC a partire dalla pubblicazione del 2004 di Aston, Bernardini e Stewart fino al 2018, raggruppando gli studi per aree tematiche.

Tra gli studi che gettano le basi per la futura progettazione di materiali didattici si citano Cosme e Gilquin (2008), che sottolineano le ricadute positive che l'analisi contrastiva dei dati ricavati da un corpus bilingue inglese-francese sull'uso delle preposizioni *with* e *avec* può avere nell'insegnamento delle lingue, nella lessicografia bilingue e nella compilazione di dizionari. Alle stesse conclusioni giungono anche Mudraya *et al.* (2008) nel loro studio comparativo *corpus-based* dei *phrasal verbs* inglesi e delle loro espressioni equivalenti in russo, di cui evidenziano le possibili applicazioni all'insegnamento delle lingue, alla linguistica contrastiva e all'estrazione di lessico multilingue. Babych (2012) descrive l'impiego di uno strumento di ricerca basato sul Corpus Workbench (Christ, 1994a) per reperire, analizzare e valutare strutture lessico-grammaticali aventi la funzione di connettori testuali che occupano la posizione iniziale di paragrafo o frase. Dalla consultazione di alcuni corpora in lingua inglese, russa e ucraina sono emerse delle liste di connettori che hanno permesso di ampliare dell'80% le tradizionali liste di connettori testuali presenti nelle grammatiche, dimostrandone la limitatezza originaria.

Vališová (2014) presenta un corpus specializzato costituito da manuali di lingua ceca come lingua straniera e illustra come il confronto della lingua usata nei manuali con la lingua usata nei corpora generalisti apra delle riflessioni sulla lingua insegnata agli apprendenti e sui possibili miglioramenti dei materiali didattici esistenti.

Particolarmente frequente risulta l'applicazione dei corpora alla descrizione grammaticale e all'insegnamento delle collocazioni e della fraseologia. Coxhead (2008, p. 156) nel suo studio sulla preparazione dei

materiali didattici dedicati all'inglese accademico (*English for Academic Purposes*, EAP) cita lo studio di Jones e Haywood (2004), nel quale, tra le varie attività proposte per l'apprendimento della fraseologia inglese, vi sono anche delle lezioni incentrate sulla produzione scritta, in cui agli studenti vengono fornite delle espressioni fisse ricorrenti (*target formulaic sequences*) estratte da un corpus composto da testi tratti da manuali di scrittura accademica. Successivamente, viene chiesto di analizzare quelle espressioni ricorrenti nel contesto più ampio del testo da cui sono tratte e, infine, di svolgere degli esercizi di completamento (*fill in the gap*). Lo studio di Pecman (2008) illustra, invece, la creazione di un corpus parallelo inglese-francese finalizzato a reperire la fraseologia tipica dell'EAP in ambito scientifico e tecnologico e a creare un dizionario elettronico bilingue della fraseologia. Wible (2008, p. 176) illustra lo sviluppo dello strumento *Collocator*, basato sui materiali del *British National Corpus* e finalizzato all'impiego in aula: durante la lettura di testi sul web, gli studenti possono ottenere delle collocazioni semplicemente evidenziando delle parole del testo e cliccando poi sull'icona del *Collocator*, che mostrerà le collocazioni che interessano le parole evidenziate. Orenha-Ottaiano (2012) presenta alcune attività, esercizi e materiali basati su corpus che si concentrano su collocazioni e destinati ad apprendenti di L1 portoghese. Alhujaylan (2014) propone attività basate su collocazioni estratte dal corpus Collins WordBanks, mentre Seracini (2018) propone attività sulla fraseologia estratta da un corpus specialistico composto da testi di area economica tradotti dall'italiano all'inglese da studenti universitari del settore dell'*international management*.

Circa il potenziamento dell'uso del lessico, Wunderlich (2018) illustra la creazione di elenchi di termini estratti da un corpus specialistico costituito da articoli accademici del settore dell'informatica aziendale a favore degli apprendenti della disciplina, mentre Tono (2018) propone lo sviluppo di risorse per l'insegnamento delle lingue basate su corpora e sul Quadro Comune Europeo di Riferimento per le Lingue, implementabili attraverso alcuni sistemi di e-learning, tra cui si citano a titolo di esempio un'app per smartphone per la creazione di flash-card lessicali e un'applicazione di assistenza alla traduzione che include il rilevamento di errori.

Per il potenziamento delle abilità di lettura e comprensione testuale si citano Curado-Fuentes (2012), che presenta la progettazione di test e

attività di comprensione di testi scritti basati su parole chiave estratte da due corpora di testi giornalistici in lingua inglese e spagnola e Hadley e Hadley (2018), che propongono attività di lettura mirate al potenziamento della conoscenza lessico-grammaticale ricettiva e ad una maggiore velocità di lettura di apprendenti giapponesi utilizzando materiali estratti da un corpus costituito dai testi della collana Oxford Bookworms Graded Readers.

Nell'ambito della formazione dei docenti circa la realizzazione di materiali didattici derivati da corpora, Charkova e Charkova (2014) illustrano come effettuare su corpora ricerche di carattere morfologico, lessicale e grammaticale e come utilizzare i testi estratti dai corpora per produrre materiali didattici. Si citano, infine, i seminari svolti nell'ambito del TaLC13 da Curry e Goodman (2018), focalizzati sull'impatto che i dati estratti dal Cambridge Learner Corpus (CLC) di lingua scritta possono avere sullo sviluppo di materiali didattici, mentre Pico-ral, Staples, Shin e Swatek (2018) e Brezina, Gablasova e Marín Cervantes (2018) illustrano delle procedure di creazione di attività didattiche basate su dati estratti da corpora e riguardanti, rispettivamente, l'intertestualità negli scritti degli apprendenti di lingua inglese L2 e l'insegnamento della sociolinguistica.

Per quanto riguarda l'applicazione dei corpora alla didattica dell'italiano L2/LS, nel corso degli anni Duemila si è registrata una forte crescita di corpora di vario genere, come riportato negli studi di Ivanovska-Naskova (2018), Barbera (2015, a c. di), Panunzi, Cresti e Gregori (2013) e Baroni (2010). Di particolare importanza è lo sviluppo di svariati *learner corpora* costituiti da testi prodotti dagli apprendenti di italiano come L2 o LS (cfr. Ivanovska-Naskova, 2018, p. 73, Giacalone Ramat, Chini e Adorno, 2013, p. 179). Tuttavia, nonostante questa tendenza positiva, esiste un divario tra i corpora dell'italiano e il loro utilizzo nell'insegnamento della lingua e gli studi su questo tema sono ancora scarsi (cfr. Ivanovska-Naskova, 2018, p. 73). Al di là degli studi di carattere generale sulle potenzialità dei corpora di Corino (2014) e Viganò (2011), le ricerche che illustrano proposte ed esperienze didattiche per l'impiego in aula dei corpora dell'italiano L2/LS o per la formazione dei docenti sono poco numerose (cfr. Ivanovska-Naskova, 2015).

Di seguito si citano a titolo di esempio alcuni studi, tutti caratterizzati dalla proposta dell'uso diretto dei corpora da parte degli apprendenti,

dando particolare rilievo nell'esposizione a quelli che presentano alcune analogie con la presente ricerca, ossia che in aggiunta propongono l'uso dei corpora come fonte per la produzione di materiali didattici. Lo studio di Guidetti, Lenzi e Storchi (2012) propone una serie di strategie per portare progressivamente gli apprendenti dell'italiano come LS ad un utilizzo sempre più autonomo dei corpora. Dopo aver tracciato una panoramica sui principali utilizzi dei corpora nella glottodidattica, viene proposto l'utilizzo di tre risorse giudicate dagli autori di facile utilizzo e caratterizzate da un'interfaccia *user-friendly*, ossia *La Repubblica Corpus*, il *Corpus LIP* e il programma di interrogazione di corpora *Sketch Engine* (per ciascuno di essi si rinvia al capitolo 4). Nello studio viene proposto un percorso graduato che parte con attività basate sull'uso dei corpora guidato dal docente, fino ad arrivare ad un uso autonomo come strumento di studio per ricercare la soluzione a quesiti sul funzionamento della lingua e che richiedono la padronanza di diversi stili di interrogazione dei corpora. Partendo con la produzione di liste di frequenza e concordanze, che permettono di identificare delle strutture linguistiche ricorrenti, gli apprendenti assumono gradualmente il ruolo di apprendenti-ricercatori che, sulla base dei dati raccolti dalla consultazione del corpus, traggono delle generalizzazioni sul funzionamento della lingua, consolidando e/o perfezionando così le loro conoscenze.

Kennedy e Miceli (2010) propongono l'uso del corpus CWIC (*Contemporary Written-Italian Corpus*) come risorsa utile al potenziamento della competenza scritta degli studenti universitari. Partendo dalle difficoltà riscontrate da studenti di livello intermedio di italiano LS nell'uso dei corpora come supporto all'apprendimento linguistico, le due studiose hanno progettato un percorso di apprendimento mirato allo sviluppo delle capacità di ricerca su corpus che sono necessarie per formulare e testare ipotesi sull'uso della lingua. Il percorso ha la durata di un semestre; nella fase iniziale l'uso dei corpora viene introdotto come risorsa di supporto alla produzione scritta da affiancare all'uso tradizionale del dizionario bilingue, per poi essere applicato gradualmente per la risoluzione di dubbi e problemi grammaticali sempre più specifici. Attingendo alle intuizioni dei tre partecipanti al *case study* in questione, viene definito il percorso di approccio alla consultazione dei corpora da replicare in successive sperimentazioni e vengono indivi-

duati alcuni perfezionamenti da apportare rispetto al percorso iniziale. A differenza degli studi precedenti, la ricerca di Corino e Marellò (2009) è incentrata sull'uso diretto dei *learner corpora*. Nello specifico si propone l'uso di VALICO⁶ (Varietà Apprendimento Lingua Italiana Corpus Online), un corpus di testi scritti da apprendenti stranieri di italiano, e di VINCA⁷ (Varietà di Italiano di Nativi Corpus Appaiato). Entrambe le risorse fungono da strumenti di supporto alla formazione di futuri insegnanti di italiano L2/LS e all'insegnamento della morfologia sintassi italiana a stranieri.

Una prima applicazione del corpus VALICO è la sua consultazione mirata all'individuazione degli errori più frequenti commessi dagli apprendenti stranieri. La selezione nella maschera di ricerca di alcuni dati sociolinguistici, quali la L1 degli apprendenti, l'anno di studio della lingua italiana, la conoscenza di altre lingue ecc. permette di creare dei sottocorpus con i testi prodotti dalla popolazione selezionata. In tal modo è possibile analizzare le caratteristiche dei testi ottenuti e ottenere dati funzionali alle esigenze della formazione di insegnanti e studenti di italiano L2/LS.

Oltre alla consultazione del corpus, anche la sua stessa costruzione è stata utilizzata con finalità di formazione degli studenti di Didattica delle Lingue Moderne e Didattica della Lingua Italiana come LS dell'Università di Torino: collaborando alla trascrizione dei testi, alla correzione manuale e all'attribuzione di PoS si sono imbattuti negli errori più frequenti e sono stati spinti a riflettere sulle cause di quegli errori.

Inoltre, è stata proposta la consultazione del corpus VALICO da parte di studenti che svolgono la formazione in didattica dell'italiano con l'obiettivo di individuare nei testi che compongono il corpus gli errori commessi nella produzione scritta in italiano e, in una seconda fase, basare su queste osservazioni la progettazione di alcuni test linguistici a scelta multipla con distrattori basati sugli errori più frequenti.

Infine, vengono fatti alcuni brevi cenni alla produzione di attività didattiche basate sui testi del corpus di italofoeni VINCA: la sperimentazione da parte di apprendenti sia italiani che stranieri di un'attività di

⁶ <http://www.valico.org>.

⁷ <http://www.valico.org/vinca.html>.

riordino delle sequenze dei testi, finalizzata all'elaborazione di un'attività di verifica per apprendenti romeni (Merlo, 2009) e delle attività di comprensione testuale destinate a studenti italofoeni nativi e statunitensi (Corino, Marellò, 2009).

Questo aspetto relativo alla produzione di attività didattiche basate su testi tratti da corpora risulta, invece, maggiormente sviluppato dalle ricerche di Lyding *et al.* (2013) e Ivanovska-Naskova (2018), a cui verrà dato ora particolare rilievo, date le analogie con la presente ricerca: in entrambi i casi sono stati progettati dei materiali didattici per l'insegnamento dell'italiano a stranieri basate sull'uso del corpus PAISÀ. Lo studio di Lyding *et al.* (2013) dopo aver illustrato le principali caratteristiche e funzioni del corpus PAISÀ, presenta un'unità di lavoro dedicata ad apprendenti adolescenti di lingua italiana, la cui competenza linguistica si colloca tra i livelli A2-B1, incentrata sull'uso diretto da parte degli apprendenti dell'interfaccia di PAISÀ e delle sue funzionalità. I contenuti tematici dell'unità includono il saper parlare e scrivere delle proprie preferenze musicali, mentre dal punto di vista grammaticale l'unità è incentrata sull'uso dei comparativi e superlativi e su alcuni tratti di fonetica.

Dopo aver dato delle nozioni generali su cosa sia un corpus, viene chiesto agli apprendenti di ricercare con la modalità semplice e avanzata dell'interfaccia le diverse forme del comparativo e del superlativo. Successivamente viene proposta la lettura di un testo tratto dal corpus e viene chiesto agli apprendenti di ricercare nel corpus delle occorrenze di alcuni tratti tipici della scrittura di registro informale sul web presenti nel testo (ad esempio, acronimi, abbreviazioni, troncamenti ecc.) Infine, si richiede di creare un sottocorpus con le occorrenze reperite e di riscrivere le frasi trovate con un registro più formale. L'unità propone anche un'attività sul lessico: attraverso la funzione Word Cloud è possibile visualizzare quali parole co-occorrono con la parola ricercata sulla base della loro frequenza e, partendo dai risultati ottenuti, si può richiedere agli studenti di raggruppare le parole sulla base di relazioni semantiche.

Lo studio di Ivanovska-Naskova (2018), invece, è incentrato sull'uso dapprima guidato e infine diretto di corpora (tra cui PAISÀ) applicato allo studio dei connettivi concessivi italiani nell'insegnamento dell'italiano a stranieri. Viene presentata una proposta didattica sperimentale

tata nell'ambito del corso di Sintassi dell'italiano, destinato a studenti di livello B2 nel corso di laurea quadriennale in Lingua e Letteratura italiana presso l'Università di Skopje. Tra gli obiettivi delle attività proposte vi è quello di abituare gli studenti all'uso incrociato di più risorse elettroniche, ossia corpora monolingui, corpora paralleli e dizionari elettronici; nello specifico le risorse utilizzate sono: il corpus PAISÀ (cfr. capitolo 4), un corpus parallelo italiano-macedone e, per quanto riguarda i dizionari, il Nuovo De Mauro e il Sabatini Coletti.

Le attività proposte, di cui di seguito si darà una breve descrizione, prevedono una prima fase in cui si richiede agli studenti di riflettere su alcuni testi tratti dal corpus PAISÀ (che è la parte dello studio che presenta analogie per la presente ricerca, poiché il corpus viene utilizzato come fonte di testi su cui costruire le attività didattiche) e una seconda fase che prevede l'uso diretto del corpus PAISÀ da parte degli studenti. La parte introduttiva è mirata a far familiarizzare gli studenti con la costruzione della subordinata concessiva, le possibili traduzioni in macedone e il legame esistente tra la relazione concessiva e quella avversativa. Nelle attività iniziali si richiede di individuare le strutture concessive in alcuni brevi testi tratti dal corpus PAISÀ e, in particolare, di focalizzare l'attenzione sulla varietà di connettivi utilizzati per esprimere la concessione (coniunzioni e locuzioni concessive), sulla posizione della frase concessiva nella sintassi del periodo e sul tipo di frase concessiva (concessiva verbale – esplicita o implicita – o nominale).

In una seconda fase vengono proposte attività che prevedono l'uso diretto di PAISÀ da parte degli studenti: data una lista di connettivi concessivi, si richiede di ordinarli in base alla frequenza d'uso nel corpus; in un'altra attività viene chiesto di ricercare nel corpus degli esempi contenenti alcuni connettivi, in modo da verificare le intuizioni circa l'uso di questi connettivi.

Successivamente viene proposto un uso combinato di dizionari monolingui e del corpus PAISÀ: in alcune attività in una prima fase si richiede di consultare il dizionario De Mauro per individuare le funzioni principali dei connettivi *per quanto* e *malgrado* e, successivamente, si richiede di ritrovare tali funzioni in alcune frasi tratte da PAISÀ che contengono i connettivi sopra menzionati. La stessa tipologia di attività viene svolta con *anche se*, consultando però il dizionario Sabatini Coletti. In quest'ultimo caso, oltre a fornire degli esempi tratti

da PAISÀ, viene chiesto agli studenti di tradurne alcuni, allo scopo di stimolare la riflessione su analogie e differenze tra il sistema dei connettivi italiani e quelli macedoni.

Nella parte conclusiva viene proposta un'attività basata interamente sull'uso diretto del corpus PAISÀ da parte degli studenti, a cui si richiede di ricercare del costruito *pure* + gerundio.

Infine, vengono forniti degli esempi di frasi concessive italiane con le corrispettive traduzioni in macedone tratte da un corpus parallelo (Ivanovska-Naskova, 2014) e si richiede di individuare le strutture concessive nelle due lingue e riflettere sulla loro forma.

Come è evidente, la maggior parte degli studi citati riguardano principalmente l'estrazione di elementi linguistici dai corpora al fine di facilitare l'approccio da parte degli apprendenti con il lessico e la fraseologia, gettare le basi per la compilazione di dizionari e grammatiche o per fornire un supporto nell'apprendimento delle lingue. Solo in alcuni degli studi precedentemente citati, che includono anche l'insegnamento dell'italiano a stranieri, i corpora sono utilizzati come fonte per la produzione di materiali didattici o come strumenti tecnologici a supporto della didattica. Ciò che, però, emerge in generale è la necessità di far riferimento ai contenuti dei corpora al fine di sensibilizzare gli apprendenti circa alcune strutture linguistiche e fornire loro strumenti metodologici e/o tecnologici per l'analisi dell'italiano, oltre che la necessità di ottenere descrizioni grammaticali più accurate rispetto a quelle generalmente fornite dalle grammatiche e dai manuali, poiché gli esempi sono tratti da produzioni autentiche della lingua e non da esempi creati *ad hoc*.

A questo proposito, Wible (2008, p. 179) mette in evidenza due modalità di utilizzo del web a supporto della didattica delle lingue, il secondo dei quali è quello che coincide con uno degli obiettivi della presente ricerca, ossia il reperimento di testi autentici da reimpiegare nella progettazione di attività didattiche:

to language learning, there are roughly two ways of constructing the Web as a language learning environment. First, the Web can be used as a location for storing and publishing content designed explicitly for language learning. Content of this sort would include

online language lessons or activities. [...] A second way of viewing the Web as a source for language learning is to see it as a massive archive of pre-existing content that has not been designed for the purpose of language learning and teaching but which learners and teachers can exploit for that purpose.

Il riferimento a testi preesistenti che non sono stati progettati per gli scopi dell'apprendimento linguistico rinvia alla questione dell'autenticità dei testi, che è uno dei temi teorici e metodologici fondamentali nell'uso dei corpora con finalità didattiche, di cui si tratterà nella sezione successiva.

3.8. L'uso di materiali autentici

Nella definizione di materiale autentico, Bonvino (2004, p. 26); afferma che “un materiale può considerarsi autentico se si tratta di un testo che non è stato concepito per studenti stranieri, ma al contrario se nasce per un pubblico di parlanti nativi”. Si tratta, dunque, della pluralità di materiali prodotti dalla comunità dei parlanti della L1, “creati per i madrelingua, quindi non a scopo didattico” (Begotti, 2006, p. 14), e classificabili, sempre secondo Begotti, come cartacei, audio, video e multimediali (ivi, pp. 15-18). L'uso di materiali autentici nella didattica delle lingue seconde e straniere è oggi ampiamente diffuso e consolidato, “in quanto testimonianze vive e spontanee della cultura e della civiltà [...] e delle relative manifestazioni” (Comodi, 1995, p. 10). Si vedano a questo proposito gli studi di Breen (1985), Taylor (1994), Widdowson (2000), Stubbs (2001) Amor (2002) e, nell'ambito dell'italianistica, Harrich (2011), Santipolo (2014) e i già citati Comodi (1995) Begotti (2006 e 2011). L'autenticità dei materiali fu, inoltre, anche il tema principale della *keynote session* di Sinclair e Widdowson al TaLC 2002 “corpora and language teaching tomorrow”⁸, sulla base di due articoli di Sinclair (1991) e Widdowson (1991) che avevano sollevato l'interesse sul tema.

⁸ Si fa riferimento ai discorsi (non pubblicati) tenuti da John Sinclair e Henry Widdowson nell'ambito del TaLC 2002.

Il concetto di materiale autentico, in contrapposizione a quello di materiale didattico, fu introdotto da Wilkins (1976, p. 79), sebbene la spinta ad utilizzare esempi linguistici autentici abbia iniziato a farsi strada sin dagli anni '50, con lo sviluppo dell'approccio comunicativo nella glottodidattica. Quest'ultimo ha posto l'accento in tutte le sue declinazioni teoriche (metodo situazionale, nozionale-funzionale, metodo naturale di Krashen ecc.) sull'esigenza di esporre gli apprendenti della lingua a situazioni comunicative reali e, dunque, ad un contatto diretto con la lingua in uso.

Sempre a partire dagli anni '50 nell'ambito della linguistica Firth si esprime negativamente circa i manuali di lingua, sostenendo che molti degli esempi linguistici impiegati nella didattica fossero senza senso ("just nonsense") dal punto di vista semantico (1957, p. 24). Questa opinione fu ripresa in seguito da Sinclair (1991: 5), che fa riferimento ad un presupposto, ossia "the absurd notion that invented examples can actually represent the language better than real ones", a cui fanno eco Glisan e Drescher (1993, p. 24): "if grammar is to be taught for communicative purposes, the structures presented should reflect their use in current-day native speaker discourse".

A questo proposito, Römer (2004) si interroga su quale varietà linguistica venga proposta agli apprendenti di una lingua tramite i manuali, proponendo di confrontare il cosiddetto "school English", ossia la varietà utilizzata nei manuali, con l'inglese della vita quotidiana ("real English"):

to help us answer two crucial questions related to language teaching: "Do we teach our pupils authentic English, i.e., do we confront them with the same type of English we are likely to be confronted with in natural communicative situations?" and "What can we do to improve EFL teaching materials?" (Römer, 2004, p. 152)

A conclusione di tale riflessione, Römer si schiera a favore dell'uso di frasi tratte da situazioni comunicative reali in contrapposizione all'uso di testi creati *ad hoc*: "I would like to claim that there is a need for more authentic, naturally produced, non-invented examples in EFL

teaching” (ivi, p. 155), illustrando come la didattica delle lingue possa trarre beneficio dall’uso di testi autentici.

Innanzitutto, uno dei vantaggi degli esempi reali è che essi si sono verificati davvero, coerentemente all’affermazione di Sinclair sulla necessità della didattica delle lingue di presentare solo esempi reali (“present real examples only”), poiché la lingua non può essere inventata, ma solo catturata: “language cannot be invented; it can only be captured” (1997, p. 31).

Inoltre, sempre secondo Römer (2004, p. 154) l’uso dei testi autentici favorirebbe la naturalezza dell’espressione linguistica degli apprendenti: “Another reason why we might want to replace invented examples with real ones is that we may “hinder the development of fluency by excluding data samples that fluent native speakers actually say” (de Beaugrande, 2001, p. 39)”. Tale affermazione trova conferma in Mauranen (2004: 208): i corpora composti da testi autentici (nello studio in questione, i corpora della lingua orale) “can achieve high authenticity, serve as communication aids, and provide irreplaceable models of the target language”.

Pertanto, sviluppare materiali didattici basati sulla lingua autentica significa avvicinare i tratti linguistici insegnati a quelli che si osservano nell’uso della lingua (Römer, 2004, p. 161) ed è proprio questo aspetto a motivare sempre più i docenti a inserire l’uso dei corpora nella didattica, come affermato da Aston, Bernardini e Stewart (2004, p. 12):

Authenticity [...] refers to a piece of text being “attested”, having occurred as part of genuine communicative (spoken or written) interactions. According to Hoey (this volume), exposure to authentic data is crucial since “only authentic data can preserve the collocations, the collocations, semantic associations of the language (p. 37). Indeed, it is this belief that motivates more and more teachers to introduce corpora into their classrooms.

Sulla base di queste riflessioni, si ritiene che i corpora possano essere utili ai fini di questa ricerca come fonti per il reperimento di testi autentici che contengano i fenomeni dell’italiano dell’uso medio e che fungano da base per la progettazione di attività didattiche, come verrà illustrato successivamente nei capitoli 5 e 6.

4. CREARE UN'UNITÀ DIDATTICA: OBIETTIVI E METODI

Dopo aver descritto nel capitolo precedente alcuni concetti base della linguistica dei corpora e dopo aver illustrato alcuni ambiti di applicazione dei corpora alla didattica delle lingue, tra cui la realizzazione di materiali didattici, in questo capitolo verranno descritti gli obiettivi e i metodi della ricerca. Innanzitutto, verranno presentate le domande alla base di questa ricerca. A seguire, verrà presentato il corpus utilizzato per ricercare i fenomeni linguistici dell'italiano dell'uso medio e verranno illustrati i criteri applicati per la sua selezione tra altri corpora dell'italiano contemporaneo scritto attualmente disponibili, a cui si farà brevemente riferimento. Successivamente si illustrerà il metodo adottato per l'interrogazione del corpus tramite la progettazione di alcune stringhe di ricerca, di cui verrà data una descrizione. Infine, verrà presentato il metodo adottato per la progettazione e sperimentazione delle unità didattiche e per la realizzazione di un questionario per la raccolta delle opinioni degli studenti sulle attività svolte, di cui si descrive-

ranno la struttura, il contenuto e le caratteristiche principali della piattaforma online su cui è stato collocato.

4.1. Domande della ricerca

Come illustrato nei capitoli precedenti, la diffusione della lingua italiana su tutto il territorio nazionale, fino a diventare la lingua ufficiale dello stato, è frutto di un lungo processo iniziato gradualmente con l'Unità d'Italia e che si è intensificato nella seconda metà del Novecento, soprattutto tramite l'istruzione obbligatoria e i mezzi di comunicazione di massa. La varietà standard della lingua, fino ad allora impiegata unicamente nella scrittura di registro formale, dovette adattarsi alle esigenze pratiche della comunicazione nella vita quotidiana. Sotto la pressione della lingua parlata la varietà standard subì una graduale trasformazione che fece emergere tratti linguistici nuovi o già esistenti da secoli ma sanzionati dalla grammatica tradizionale, in quanto ritenuti substandard.

Sebbene tali tratti linguistici siano ampiamente utilizzati dalla comunità dei parlanti (Sabatini, 1990; 2016), la loro trattazione nei manuali per l'insegnamento della lingua italiana a stranieri è ancora limitata: ciò si evince sia da una mia osservazione dei dati raccolti da Słapek nel 2016 sugli argomenti grammaticali presenti in 32 pubblicazioni delle maggiori case editrici operanti nel settore dell'insegnamento della lingua italiana a stranieri (AlmaEdizioni, CDL Edizioni, Edilingua, Eli, Guerra Edizioni, Hoepli, Le Monnier e Loescher), sia da un successivo approfondimento da me condotto sui manuali pubblicati dal 2018 al 2021 (per entrambi i punti, cfr. capitolo 5.2).

Poiché “per il docente di italiano L2 il problema fondamentale è quello di evitare l'eccessivo scollamento tra lingua utilizzata in aula e quella effettivamente presente nelle situazioni comunicative quotidiane” (Diadori, Palermo, Troncarelli, 2015, p. 227), è emersa la necessità di progettare da zero dei materiali didattici mirati a favorire l'approccio di chi apprende con alcuni fenomeni linguistici del neostandard. Pertanto le domande della ricerca sono le seguenti:

1. Quali sono i corpora rappresentativi della lingua italiana scritta di media formalità degli ultimi 20 anni in cui è possibile verificare la diffusione dei tratti linguistici identificati da Sabatini (1990) come costitutivi dell'italiano dell'uso medio?

2. Sulla base dell'interrogazione del corpus prescelto, quali sono i tratti linguistici più frequenti e, dunque, più diffusi nei testi autentici rappresentativi della varietà scritta di media formalità?
3. A fronte all'attuale presenza limitata, se non assenza, dei tratti dell'italiano dell'uso medio nella manualistica L2/LS (come si deduce dai dati raccolti da Słapek nel 2016 e dal mio successivo approfondimento sul tema), su quali di essi è rilevante progettare dei materiali per la didattica dell'italiano L2/LS?
4. Dopo aver progettato e sperimentato i materiali in aula, quali sono le opinioni dei partecipanti alla sperimentazione rispetto all'approfondimento dell'uso neostandard dei tratti linguistici selezionati?

Come è stato illustrato nei punti 1 e 2, il primo obiettivo è identificare un corpus di lingua italiana scritta in cui ricercare i tratti linguistici identificati da Sabatini (1990). Tale corpus verrà utilizzato per verificare l'ipotesi circa l'effettiva reperibilità – e in quale misura – dei tratti dell'italiano dell'uso medio nei testi scritti di media formalità.

Pertanto, nella sezione seguente verrà descritto il corpus ritenuto più idoneo per gli scopi della ricerca. Successivamente, verranno motivate le ragioni per le quali possa essere considerato preferibile rispetto ad altri corpora sincronici dell'italiano scritto attualmente disponibili. Di questi verrà fornito un elenco alfabetico e verranno brevemente messe in luce le caratteristiche che ne hanno determinato l'esclusione. È bene precisare che questa breve panoramica non si prefigge l'obiettivo di coprire la totalità dei corpora di lingua italiana esistenti, bensì di presentare una selezione di alcune risorse ritenute potenzialmente idonee agli scopi della ricerca.

4.2. Corpus selezionato per l'indagine: PAISÀ

È un corpus realizzato nell'ambito del progetto PAISÀ (Piattaforma per l'Apprendimento dell'Italiano Su corpora Annotati) che ha coinvolto dal 2009 al 2012 l'Università di Bologna, il CNR di Pisa, l'Accademia Europea di Bolzano e l'Università di Trento, avente l'obiettivo di facilitare l'accesso a una grande quantità di testi scritti in italiano contemporaneo da parte degli utenti del web che desiderano migliorare le proprie conoscenze linguistiche e culturali (cfr. Lyding *et al.*,

2014, p. 36). Nello specifico, i destinatari principali del progetto sono gli emigrati sia di seconda generazione, che pur avendo l'italiano come L1 ne fanno un uso molto limitato, sia di terza generazione, per i quali l'italiano è una seconda lingua (L2)¹.

I testi che costituiscono il corpus PAISÀ sono rappresentativi dell'uso contemporaneo della lingua italiana, selezionati in maniera automatica da internet tra settembre e ottobre del 2010² e liberamente disponibili, poiché protetti da licenze Creative Commons di tipo *CC-Attribuzione*, *CC-Attribuzione-Condividi allo stesso modo*, *CC-Attribuzione-Non commerciale*, o *CC-Attribuzione-Non commerciale-Condividi allo stesso modo*³. A questo proposito:

The main novelty of the PAISÀ web corpus is that it exclusively draws on Creative Commons licensed data, provides advanced linguistic annotations with respect to corpora of comparable size and corpora of web data, and invests in a carefully designed query interface, targeted at different user groups. In particular, the integration of richly annotated language content with an easily accessible, user-oriented interface makes PAISÀ a unique and flexible resource for language teaching. (Lyding *et al.*, 2014, p. 36)

La raccolta dei testi dal web è avvenuta attraverso due procedimenti (cfr. *ibidem*, pp. 37-38). Il primo, ispirato al progetto *WaCky* (*Web-as-Corpus Kool Ynitiative*; Baroni, Bernardini, 2006; Baroni *et al.*, 2009), è definito *seeds crawling* (letteralmente, “raccolta di semi”) che con-

¹ Cfr. www.corpusitaliano.it/it/contents/PAISÀ.html.

² “All documents contained in the PAISÀ corpus date back to Sept/Oct 2010” (Lyding *et al.*, 2014, p. 37). Contrariamente a quanto si potrebbe dedurre da tale affermazione, i testi non *risalgono* al 2010, bensì sono stati *raccolti* tra settembre e ottobre 2010, come è stato chiarito dal confronto diretto con due dei ricercatori coinvolti nel progetto PAISÀ (Sara Castagnoli e Marco Brunello). Questi ultimi hanno chiarito che non sono state poste restrizioni temporali nella raccolta di testi disponibili con licenze *Creative Commons*, tra i quali i più recenti risalgono al 2010.

³ Per approfondire le caratteristiche di ciascuna licenza, si rinvia a: <http://creativecommons.org/licenses/>.

sta di due fasi. Nella prima fase, tramite BootCaT, un programma di costruzione di corpora, è stato utilizzato il motore di ricerca *Yahoo!* per ricercare nel web combinazioni casuali di parole di uso comune (ad esempio: cane+casa), che rappresentano i “semi” (*seeds*) da cui far partire la ricerca. Nello specifico, le combinazioni di parole utilizzate come stringa di ricerca in *Yahoo!* sono 50.000 coppie di parole tratte dal Vocabolario di Base della Lingua Italiana di Tullio De Mauro (1997 [1980]). Nel fare ciò è stata limitata la raccolta alle pagine dal dominio .it e con le licenze *Creative Commons* precedentemente menzionate. I risultati ottenuti hanno permesso di identificare le URL dei documenti da scaricare per costruire il corpus⁴. Successivamente sono state rimosse dalla lista delle URL tutte le pagine che il motore di ricerca aveva classificato erroneamente come protette da licenze *Creative Commons* (individuate in base ad una *black list* di siti realizzata manualmente durante la realizzazione delle versioni precedenti e sperimentali del corpus). In base all’osservazione di Baroni e Ueyama (2006), le ricerche tramite parole appartenenti al vocabolario di base conducono a blog e a testi di carattere più personale.

Il secondo procedimento consiste nell’integrazione all’interno del corpus di altri testi protetti da licenza *Creative Commons* tratti dalle versioni italiane di Wikipedia, Wikinews, Wikisource, Wikibooks, Wikiversity e Wikivoyage. Come riportato in Lyding *et al.* (2014, p. 38), ciò è stato reso possibile dall’iniziativa della Wikimedia Foundation nel 2009 di rendere pubblici i propri contenuti dapprima dietro licenza GNU e in seguito *CC-Attribuzione-Condividi allo stesso modo*, permettendo così di inserire nel corpus una grande quantità di testi delle versioni italiane dei siti sopra nominati.

Successivamente si è proceduto a scaricare il contenuto degli indirizzi URL, a ripulire le pagine attraverso strumenti KrdWrd⁵ o, nel caso

⁴ Né in Lyding *et al.* (2014) né sulla pagina web del corpus risulta che in una fase successiva sia stato fatto il crawling a partire dalle pagine trovate.

⁵ “The KrdWrd Project ran from 2008 to 2011. The mission statement was: *provide tools and infrastructure for acquisition, visual annotation, merging and storage of web pages as parts of bigger corpora; develop a classification engine that learns to automatically annotate pages, and provide visual tools for inspection of results.* Basically, it was an infrastructure for research into web page cleaning” (cfr. <https://krdwr.github.io>).

specifico delle pagine di Wikipedia, attraverso il Wikipedia Extractor in combinazione con uno script che permette di separare i singoli documenti. Dopodiché sono stati rimossi dai contenuti i file vuoti e i file contenenti meno di 150 parole.

I procedimenti illustrati hanno permesso di realizzare un corpus costituito da 250 milioni di *token*⁶ e i testi che lo compongono hanno una lunghezza media di 637 parole.

Per quanto riguarda invece le fonti dei testi raccolti:

The corpus contains approximately 388,000 documents. [...] The documents come from several web sources which, at the time of corpus collection, provided their content under Creative Commons license [...]. About 269,000 texts are from Wikimedia Foundation projects [...]. The remaining 119,000 documents come from guide.supereva.it (ca. 19,000), italy.indymedia.org (ca. 10,000) and several blog services from more than another 1,000 different sites (e.g., www.tv-blog.it (9,088 pages), www.motoblog.it (3,300), www.ecowebnews.it (3,220), and www.webmasterpoint.org (3,138)). (Lyding *et al.*, 2014, p. 37)

Ai testi sono poi stati aggiunti diversi livelli di annotazione sia sul piano linguistico, sia sul piano dei metadati. Dal punto di vista linguistico, i testi sono stati innanzitutto annotati – secondo il formato CoNLL (Computational Natural Language Learning) (Buchholz, Marsi, 2006) – rispetto alle parti del discorso (*POS-tagging*, ossia *part-of-speech tagging*)⁷ tramite l'ILC-POS-Tagger del CNR di Pisa, ossia un software che permette di analizzare un testo ed ottenerne l'analisi grammaticale. Per fare ciò, è stato utilizzato il TANL *fine-grained pos*

⁶ “A token is any instance of a particular wordform in a text; comparing the number of tokens in the text to the number of types of tokens – where each type is a particular, unique wordform – can tell us how large a range of vocabulary is used in the text” (McEnery e Hardie, 2012, p. 50).

⁷ Associazione di specifici attributi informativi, in questo caso di carattere morfosintattico, agli elementi che compongono un testo elettronico (cfr. Barbera, 2015, p. 120).

tagset (ossia, una lista di parti del discorso). Successivamente, i testi sono stati sottoposti a lemmatizzazione (che consiste nell'associare a ogni parola la sua forma base con cui è elencata nel dizionario, esempio: *gatti* > *gatto*), *parsing* (ossia, l'analisi sintattica, tramite il DeSR *Dependency Parser* dell'Università di Pisa, a cui è stato applicato l'ISSTANL *Dependency Tagset*) e, infine, sono state aggiunte le informazioni morfosintattiche e le relazioni di dipendenza.

Le annotazioni sul piano dei metadati (argomento, funzione e genere testuale) sono state svolte in base alla letteratura su generi e funzioni testuali nel web (ad esempio, Santini, 2005; Sharoff, 2006; Rehm *et al.*, 2008; Santini *et al.*, 2010; Sharoff, 2010). Quest'ultima classificazione è stata realizzata sia per favorire il giusto bilanciamento nella composizione del corpus rispetto a argomenti, funzioni e tipologie testuali ed eventualmente intervenire per bilanciarlo scaricando in maniera mirata ulteriori testi, sia per fornire agli utenti una vasta gamma di varietà testuali e consentire modalità di ricerca avanzate (cfr. Borghetti, Castagnoli, Brunello, 2011). Inoltre, a ogni documento del corpus è stato attribuito l'indirizzo URL sorgente, il che permette di risalire alle pagine web da cui sono stati tratti i testi, qualora queste siano ancora attive.

Il corpus è liberamente consultabile online⁸ attraverso una maschera di ricerca che prevede tre opzioni: ricerca semplice, avanzata e tramite il linguaggio di interrogazione *Corpus Query Processor (CQP)*. Inoltre, è possibile scaricare gratuitamente il corpus, sia nella versione annotata che nella versione non annotata, e consultarlo tramite altri software di interrogazione. Come verrà successivamente spiegato, per la presente ricerca è stata adottata la soluzione di scaricare la versione annotata e consultarla esternamente in seguito ad alcune difficoltà tecniche riscontrate con l'uso dell'interfaccia originale.

Nella sezione successiva verranno illustrate le ragioni che hanno portato all'individuazione del corpus PAISÀ come il più idoneo per gli scopi della presente ricerca attraverso un confronto con altre risorse ritenute potenzialmente idonee.

⁸ www.corpusitaliano.it/it/access/simple_interface.php.

4.3. Criteri di individuazione del corpus da utilizzare e confronto tra corpora

Per identificare il corpus da utilizzare, sono stati presi in esame alcuni corpora di lingua italiana che, tra quelli attualmente disponibili, risultano pertinenti rispetto all'obiettivo di documentare empiricamente la diffusione dei tratti linguistici identificati da Sabatini (1990) nei testi rappresentativi della varietà scritta contemporanea di media formalità. Ciò è in linea con quanto affermato da McEnery e Hardie (2012, p. 2): “the corpus data we select to explore a research question must be well matched to that research question”.

Pertanto, la selezione si è concentrata esclusivamente sui corpora sincronici dell'italiano scritto, ossia sulle raccolte di testi scritti prodotti all'interno della stessa finestra temporale e che, quindi, consentano la descrizione di uno stadio ben preciso della lingua: nel caso specifico, della varietà scritta dell'italiano ricorrente nell'uso comune dei parlanti nativi negli ultimi vent'anni. Di conseguenza, sono stati esclusi i corpora diacronici, utili ad uno studio dell'evoluzione dei fenomeni linguistici nel corso di un arco di tempo più ampio, i corpora dell'italiano parlato e i *learner corpora* dell'italiano L2/LS.

Nella valutazione dei corpora ai fini della presente ricerca sono stati presi in considerazione i seguenti parametri: 1) le date di pubblicazione dei testi presenti nei corpora, privilegiando i testi prodotti negli anni Duemila fino ai nostri giorni, allo scopo di ottenere materiali rappresentativi dell'uso attuale della lingua; 2) gli eventuali aggiornamenti dei contenuti e la cadenza con cui vengono svolti; 3) le dimensioni generali della risorsa⁹; 4) i generi testuali a cui appartengono i testi; 5) i

⁹ A questo proposito Sinclair (1996, p. 6): “The default value of Quantity is large. [...] The whole point of assembling a corpus is to gather data in quantity”. Affermazioni concordi sono reperibili, ad esempio, anche in Svartvik (1992); Marengo (1996); Bolwler, Pearson (2002), Mukherjee (2002), McEnery (2003) e Baker, Hardie, McEnery (2006). Per quanto riguarda, invece, le dimensioni minime che un corpus deve avere affinché sia adeguato a un'analisi statistica “Come indicazione generale, un corpus di 15.000 occorrenze si può considerare di “piccola dimensione”; un corpus di 50.000-100.000 occorrenze di “media dimensione” e un corpus maggiore di 200.000 occorrenze di “grande dimensione”. Un corpus sufficientemente grande (oltre le 500.000 occorrenze) può costituire una base per la costruzione di un lessico di frequenza rappresentativo di un linguaggio purché i

criteri di accessibilità della risorsa (liberamente accessibile online o sottoposta a restrizioni di registrazione ed eventuale abbonamento). Un particolare punto a favore è stato considerato il rilascio del corpus con licenze *Copyleft*, in vista di un possibile reimpiego nella prassi didattica dei risultati delle ricerche.

Di seguito verrà fornito un elenco alfabetico dei corpora sincronici dell'italiano scritto attualmente disponibili e inizialmente presi in esame per lo studio della varietà scritta dell'italiano neostandard, di cui verrà data una breve descrizione e una panoramica riassuntiva nell'Appendice 1¹⁰:

- a) *Araneum Italicum* (Benko 2014a, 2014b, 2016): ha due versioni, denominate *Maius* e *Minus*, costituite rispettivamente da 1.2 miliardi di token e 119 milioni di token, che rappresentano le dimensioni standard delle versioni *Maius* e *Minus* di tutti i corpora Aranea. Non sono state reperite informazioni circa la catalogazione in base a generi testuali o aree tematiche dei testi contenuti nei corpora e non sono presenti informazioni circa l'eventuale rilascio di licenze d'uso dei testi raccolti;
- b) *Athenaeum Corpus* (Barbera, 2007): è un corpus dell'italiano scritto accademico, realizzato con testi prodotti dall'Università di Torino, tratti principalmente dal magazine universitario *L'Ateneo*, dalla newsletter *Dall'Università*, da circolari ed e-mail amministrative, catalogati in base a generi testuali e aree tematiche. Il corpus consta di 306 mila token e dispone della licenza *Creative Commons Share Alike*. Tuttavia, data la varietà di lingua che si propone di os-

testi siano abbastanza rappresentativi della sua eterogeneità". (Giuliano, La Rocca, 2008, p. 208). Sinclair (2005, online, senza numero di pagina) sostiene in proposito che "The minimum size of a corpus depends on two main factors: 1) the kind of query that is anticipated from users; 2) the methodology they use to study the data. There is no maximum size". Il che lascia intendere che "the sustainability of the sample depends on the specific study that is undertaken, and that there is no such thing as the best, optimum, sample size as such" (de Haan, 1992, p. 3).

¹⁰ I dati si riferiscono al biennio 2017-2018 in cui è stata svolta dapprima la ricerca del corpus più idoneo e, successivamente, la sua consultazione. Per una descrizione dettagliata di ciascuna risorsa, si rinvia a Zingaro (2021, pp. 93-111), mentre l'Appendice 1 fornisce una panoramica riassuntiva del confronto delle diverse risorse.

servare (scritto accademico) e le ridotte dimensioni, questa risorsa non risulta adeguata agli scopi di questa ricerca.

- c) *CLIPS* (Corpora e Lessici dell'Italiano Parlato e Scritto)¹¹: sono reperibili informazioni sul solo corpus di italiano parlato, mentre per la sezione Lessici e Scritto dalla pagina del progetto si viene reindirizzati ad una pagina vuota dell'Istituto di Linguistica Computazionale di Pisa, il che non consente di valutare l'adeguatezza o meno della risorsa rispetto agli scopi della ricerca;
- d) *COLFIS* (Corpus e Lessico di Frequenza dell'Italiano Scritto) (Laudanna *et al.*, 1995): è un corpus bilanciato, i cui contenuti rispecchiano le tendenze di lettura degli italiani ricavate da un'indagine ISTAT del 1993. Include testi tratti dai quotidiani *la Repubblica*, *La Stampa*, *Il Corriere della Sera* del periodo 1992-1994, periodici e libri (inclusi anche quelli letti per motivi scolastici o professionali). È composto da 3,7 milioni di *token*. Poiché i dati inseriti nel corpus risalgono esclusivamente agli anni '90, il suo utilizzo condurrebbe a risultati non più attuali e inadeguati agli scopi della presente ricerca;
- e) *CorDIC* (Corpora Didattici Italiani di Confronto)¹², pubblicati in rete a cura di Lorenzo Gregori e Alessandro Panunzi (Università di Firenze): i testi del corpus CorDIC-scritto sono classificati in 5 tipologie (*arti*, *burocratico*, *creativo*, *economia* e *stampa*), ciascuna delle quali è costituita all'incirca dallo stesso numero di parole e dallo stesso numero di testi. Il corpus contiene circa 500.000 parole. Non è stato possibile valutarne appieno la conformità agli obiettivi del progetto, poiché riguardo ai contenuti, non sono state reperite informazioni né circa le date in cui sono stati prodotti, né circa il loro status del *copyright*;
- f) *CorIS/CoDIS* (CORpus di Riferimento dell'Italiano Scritto) (Rosini Favretti, Tamburini, De Santis, 2002): raccolta di testi selezionati come rappresentativi dell'italiano attuale, classificabili, come affermato sulla stessa pagina di riferimento del progetto, come testi appartenenti all'italiano "scritto-scritto" (Nencioni, 1976), ossia testi scritti nei quali non vi è alcun intento di avvicinarsi allo

¹¹ www.clips.unina.it/it.

¹² corporadidattici.lablita.it.

stile del parlato. Il nucleo originario è collocabile approssimativamente tra gli anni '80 e '90 del Novecento. È aggiornato con cadenza biennale (la più recente, al momento dello svolgimento di questa ricerca, nel 2017), inserendo nuovi materiali in un corpus di monitoraggio inglobato al suo interno (Rossini Favretti, Tamburini, De Santis, 2002). È strutturato in sette sottocorpora: sei che rappresentano le diverse varietà dell'italiano scritto (stampa, narrativa, prosa accademica, prosa giuridico-amministrativa, miscellanea ed *ephemera*) e il già nominato corpus di monitoraggio. È costituito da 150 milioni di *token*. Dal *CorIS* è derivato il *CORpus Dinamico dell'Italiano Scritto (CoDIS)*, così definito proprio perché può essere dinamicamente adattato alle diverse esigenze di ricerca. È costituito dai sei sottocorpora delle varietà dell'italiano scritto presenti nel *CorIS* ma, a differenza di quest'ultimo, offre agli utenti la possibilità non solo di selezionarne contemporaneamente più di uno per volta, ma anche di sceglierne le dimensioni funzionali agli scopi di ricerca fra le quattro dimensioni indicate. Dal punto di vista del copyright dei testi, in Rossini Favretti, Tamburini, De Santis (2002) vengono messe in evidenza le difficoltà legate all'acquisizione dei diritti al riutilizzo dei testi raccolti attraverso complesse trattative con gli editori, che non sempre hanno dato i loro frutti, nonostante la garanzia di un uso non commerciale, funzionale agli scopi della ricerca accademica, e l'adozione di misure di sicurezza contro la pirateria. Non essendo reperibili al momento della scrittura di questo capitolo ulteriori informazioni sull'esito di tali trattative, si può dedurre che alcuni editori non abbiano concesso i permessi e che, pertanto, è probabile che alcuni testi o siano stati esclusi dal corpus, o vi siano stati inseriti, ma privi delle licenze d'uso;

- g) *ItTenTen* (Jakubiček *et al.*, 2013): è un corpus che fa parte del set di corpora *TenTen Corpus Family* (Lexical Computing e Università Masaryk, Repubblica Ceca). I punti di forza di questa risorsa sono la grande quantità di dati contenuti (5,8 miliardi di *token*) e la loro attualità (2016). Va precisato che, sebbene sia composto soprattutto di testi recenti, caricati nell'arco di tempo previsto dal progetto, è possibile reperire al suo interno anche testi antichi o meno recenti (per es. brani de *La Divina Commedia* o *Il Piacere*

- di D'Annunzio). Inoltre, come nel caso dei corpora *Aranea*, non sono state reperite informazioni riguardanti la catalogazione dei testi in base a generi testuali o aree tematiche di appartenenza, né circa l'eventuale rilascio di licenze d'uso dei testi raccolti;
- h) *itWaC* (progetto *WaCky - Web-as-Corpus Kool Ynitiative*; Baroni, Bernardini, 2006; Baroni *et al.*, 2009): è un corpus molto variegato a suo interno, realizzato con testi tratti dal web raccolti attraverso il già citato procedimento definito *seeds crawling*. I *seed* utilizzati, ossia i termini che si ritengono rappresentativi di un certo ambito¹³, erano costituiti da parole a media frequenza tratte in parte dal corpus *la Repubblica* (cfr. sezione *i*, a seguire) e in parte da una lista di parole appartenenti al vocabolario italiano di base. La scelta di queste due fonti si basa sull'osservazione (Baroni, Ueyama, 2006) che le ricerche tramite parole tratte da giornali hanno come risultato testi appartenenti alla sfera pubblica, mentre le ricerche tramite parole appartenenti al vocabolario di base conducono a blog e a testi di carattere più personale, il che permette di ottenere un corpus variegato al suo interno. Il corpus contiene 2 miliardi di *token*, ma non risulta pienamente adeguato agli scopi di ricerca sotto due aspetti fondamentali: essendo stato realizzato tra il 2005 e il 2007, i suoi contenuti risultano più datati rispetto a quelli di altre analoghe risorse citate in questa trattazione; inoltre, l'impossibilità di accertare lo status del copyright dei suoi contenuti non consentirebbe un possibile reimpiego con finalità didattiche dei materiali in essa contenuti;
- i) *la Repubblica* (Baroni *et al.*, 2004): è un corpus di prosa giornalistica, che contiene gli articoli delle annate dal 1985 al 2000 del quotidiano omonimo da cui trae il nome. I testi sono stati taggati per genere testuale (reportage e commento)¹⁴ e per argomento (cultura, economia, istruzione, meteo, notizie, politica, religione, scienza, società e sport)¹⁵. Contiene 326 milioni di *token* (Baroni

¹³ "Terms that are expected to be typical of the domain of interest" (Baroni e Bernardini, 2004, p. 1313).

¹⁴ Nel corpus sono utilizzate le denominazioni in inglese: news report e comment.

¹⁵ Come sopra: culture, economics, education, weather, news, politics, church, science, society and sport.

et al., 2004), tuttavia presenta alcune criticità: i suoi contenuti provengono tutti da un'unica fonte, rientrano all'interno della sola prosa giornalistica, inoltre risultano ormai datati rispetto agli scopi di questa ricerca e, infine, non sono stati rilasciati con licenza *Copyleft*;

- j) *LexALP* (Lyding *et al.*, 2006): si tratta di corpora monolingue paralleli che coprono il francese, l'italiano il tedesco e lo sloveno, basati su circa 3000 documenti legali degli stati dell'arco alpino (Austria, Italia, Francia, Germania, Slovenia e Svizzera), mirato all'armonizzazione dei linguaggi giuridici utilizzati negli otto sistemi giuridici sopra nominati. I generi testuali in cui rientrano i materiali contenuti non sono adeguati agli scopi della presente ricerca;
- k) *LIF* - Lessico di frequenza della lingua italiana contemporanea (Bortolini *et al.*, 1971): è stato il primo grande progetto di realizzazione di un lessico di frequenza per la lingua italiana non incentrato in maniera specifica su un singolo autore o su testi letterari. I testi contenuti nel corpus si collocano tra il 1947 e il 1968, sono tratti da 5 fonti diverse (teatro, romanzi, cinema, periodici, sussidiari) con l'obiettivo di rappresentare diversi generi testuali, alcuni dei quali con varietà di lingua vicine al parlato. Il corpus contiene 500.000 parole. A causa della datazione non più attuale dei suoi testi e del fatto che la risorsa non sia disponibile per la consultazione né offline né online, il LIF è stato valutato come inadatto agli scopi della presente ricerca;
- l) *NUNC* (*Newsgroups UseNet Corpora*) (Barbera, Marello, 2008): è parte di un insieme di corpora generici e specialistici in cinque lingue, tra cui l'italiano. *NUNC italiano* è un web corpus generico, realizzato con pagine web di conversazioni scritte tratte da "newsgroups" di cinema, cucina, fotografia e motori presenti in rete negli anni 2002-2004. Il corpus è suddiviso in due sezioni, costituite rispettivamente da 127 milioni e 109 milioni di *token*. A determinarne, tuttavia, l'esclusione dalla presente ricerca è l'innaturalità dei materiali contenuti, che si collocano tra il 2002 e il 2004, nonché la complessa questione relativa alle licenze d'uso, ove le licenze *Copyleft* si applicano solo all'intera risorsa, ma non ai singoli testi contenuti;

- m) *PEC* (Perugia Corpus) (Spina, 2014): realizzato con l'obiettivo di costituire un riferimento per l'italiano contemporaneo, nelle varietà scritte e parlate, è costituito da 10 sezioni, che corrispondono ad altrettanti generi testuali. Ogni genere è, poi, a sua volta suddiviso in una serie di sottogeneri, per un totale di 43 tipologie testuali. Nello specifico, per l'italiano scritto sono disponibili i seguenti generi, qui elencati per ordine decrescente rispetto alle dimensioni (*token*): testi tratti dal web (blog, forum, chat, social network e, infine, testi di riferimento tratti da Wikipedia), letteratura, saggi, scritto scolastico, amministrativo e accademico. Questa rappresentatività delle diverse varietà della lingua scritta e parlata costituisce un punto di forza di questa risorsa. Il suo impiego va, tuttavia, escluso a causa delle sue dimensioni contenute (26 milioni di *token*), nettamente inferiori a quelle di altre risorse già citate, e della datazione dei materiali, che si collocano tra il 1990 e il 2012. Quest'ultimo aspetto potrebbe determinare, come nel caso dei già citati corpora CorIS e CoDIS (Cfr. sezione f), la presenza di effetti diacronici sulla variazione lessicale, entrando in contrasto con gli obiettivi di uno studio sincronico del neostandard. Inoltre, la scelta di questo corpus comporterebbe l'impossibilità di reimpiegarne i contenuti: "Non è invece prevista, per motivi di copyright, la disponibilità dei testi che compongono il corpus" (Spina, 2014, p. 357, n. 5);
- n) *RIDIRE* (Panunzi *et al.*, 2014): è un corpus di testi estratti dal web italiano tra il 2009 e il 2013, suddiviso in dodici subcorpora corrispondenti ciascuno a un dominio funzionale, in cui la lingua è caratterizzata dalla funzione con cui viene usata (informazione, economia e business, amministrazione e legislazione) o semantico, in cui la lingua è caratterizzata dall'insieme degli argomenti per cui viene usata (sport, architettura e design, cucina, cinema, musica, moda, arti visive, religione, letteratura e teatro) (cfr. Budinich, 2016, p. 32). È stato realizzato con una procedura di *web-crawling* "mirato", in cui al processo di *crawling* (cfr. la *sezione dedicata all'itWaC*) sono stati applicati criteri di bilanciamento modellati su quelli adottati per il British National Corpus (Cresti, Panunzi, 2013, p. 63), con l'obiettivo di mostrare l'utilizzo della lingua italiana nel web in vari contesti. Contiene circa 1,5 miliardi di *token*.

I punti di forza di questa risorsa sono sicuramente le sue dimensioni, la varietà dei domini coperti, i materiali sufficientemente recenti (la raccolta è terminata nel 2013), seppur non attualissimi, e la libera accessibilità in rete. Tuttavia, alcuni dei subcorpora disponibili (ad esempio, amministrazione e legislazione, economia e *business*, religione ecc.) è ipotizzabile che contengano tipologie di testi caratterizzati da un registro linguistico più elevato rispetto alla media formalità che caratterizza il neostandard e l'impossibilità di selezionare solo alcuni subcorpora, escludendone altri, appare poco funzionale agli scopi della ricerca.

Al fine di illustrare le ragioni per le quali il corpus PAISÀ possa essere considerato preferibile rispetto alle altre risorse citate, si farà riferimento alle seguenti tabelle (4.1 e 4.2) e all'Appendice 1, che permettono di effettuare un rapido confronto delle risorse incrociando i dati raccolti. Poiché al centro di questa ricerca vi è la varietà neostandard attualmente in uso, il criterio principale in questa fase finale di confronto tra le risorse citate sarà l'attualità dei materiali. Pertanto, innanzitutto nella tabella 4.1 verranno riportati i primi cinque corpora contenenti i materiali più attuali¹⁶. In secondo luogo, per le cinque risorse più attuali verrà preso in considerazione il rapporto attualità/dimensioni (tabella 4.2). Infine, l'Appendice 1 riassumerà tutti i dati raccolti, al fine di fornire un quadro generale.

Osservando la tabella 4.1, dal punto di vista dell'attualità al primo posto vi è **CorIS/CoDIS**, il cui ultimo aggiornamento risale al 2017. Nel suo complesso i suoi contenuti rappresentano una notevole variazione diacronica, essendo stati raccolti dagli anni '80 in poi. Dal punto di vista delle dimensioni, CorIS/CoDIS non si distingue per l'estensione, poiché occupa la quinta posizione con 150 milioni di *token* (cfr. tabella 4.2). Risulta, inoltre, sprovvisto delle licenze *Copyleft*. Questo aspetto costituirebbe un ostacolo nell'eventuale reimpiego con finalità didattiche o di ricerca dei testi contenuti nei corpora. Per tali ragioni

¹⁶ Ove disponibili, sono state riportate le informazioni relative alle date di *pubblicazione* dei testi. In caso contrario, sono state riportate solo le informazioni relative al periodo in cui è stata svolta la *raccolta* dei testi, che consentono perlomeno di fissare con certezza un limite oltre il quale non sono stati raccolti (e dunque prodotti) i testi che sono entrati a far parte del corpus.

le risorse Aranea, pur presentando alcuni punti di forza, sono state valutate come non pienamente adeguate agli scopi della presente ricerca. Al secondo posto rispetto all'attualità dei materiali vi è **ItTenTen**, con l'ultima versione risalente al 2016, che presenta l'estensione maggiore rispetto a tutte le risorse esaminate (5,8 miliardi), tuttavia non essendo state reperite informazioni sulle fonti dei testi che lo compongono, né sullo status del copyright, la risorsa è stata valutata come non pienamente adeguata agli scopi della presente ricerca.

La stessa riflessione è applicabile all'**Araneum Italicum Maius**, che si colloca al terzo posto per l'attualità dei materiali (i più recenti risalgono al 2014) e al quarto per le dimensioni (1,2 miliardi), per il quale non sono disponibili informazioni sulle fonti dei testi che lo compongono, né circa le licenze d'uso dei testi.

Al quarto posto, dal punto di vista dell'attualità si colloca il corpus **RIDIRE**, per il quale la raccolta di testi si è conclusa nel 2013. Pur occupando la seconda posizione rispetto alle dimensioni (1,5 miliardi), presenta due criticità, ossia l'impossibilità di escludere dalle impostazioni di ricerca i singoli subcorpora non adeguati allo studio della varietà dell'uso medio e l'assenza, anche in questo caso, di informazioni sullo status del *copyright* dei testi.

Con la sua quinta posizione rispetto all'attualità dei materiali (i più recenti risalgono al 2010) e la quarta posizione per le dimensioni (250 milioni) ad una prima osservazione il corpus **PAISÀ** non emerge rispetto alle altre risorse già citate, più estese e aggiornate fino ai nostri giorni. Tuttavia, i fattori che ne hanno determinato la scelta vanno compresi alla luce degli obiettivi finali della ricerca, tra cui il reperimento di testi da reimpiegare per la produzione di materiali didattici per l'italiano L2/LS. Questi fattori includono innanzitutto la presenza di licenze *Creative Commons*, che ne consentono il reimpiego dei risultati nella successiva fase di progettazione delle attività didattiche. In secondo luogo, sono state prese in considerazione le finalità didattiche per l'apprendimento dell'italiano con cui è stato progettato. Infine, la datazione dei suoi materiali¹⁷, pur non essendo attualissima, è sufficientemente recente da non compromettere lo studio sull'italiano contemporaneo.

¹⁷ Richiamando quanto già espresso in questo capitolo, i testi più recenti contenuti nel corpus risalgono al 2010.

Dopo aver identificato il corpus PAISÀ come la risorsa di riferimento per la ricerca, nelle sezioni seguenti si descriveranno la procedura di progettazione delle stringhe di ricerca funzionali all'interrogazione del corpus e la scelta dei materiali utili per la progettazione delle unità didattiche e della piattaforma online su cui collocare tali unità.

Tabella 4.1: Classificazione dei corpora in base all'attualità dei contenuti

Corpora	Date dei contenuti
1. CorIS/ CoDIS	Prodotti dagli anni '80 al 2017
2. ItTenTen	Scaricati dal web nel 2016
3. Araneum Italicum (Maius e Minus)	Scaricati dal web nel 2014
4. RIDIRE	Scaricati dal web dal 2009 al 2013
5. PAISÀ	Scaricati dal web nel 2010

Tabella 4.2: Classificazione dei corpora in base alle dimensioni

Corpora	Dimensioni della risorsa (n.ro di token)	Date dei contenuti
1. ItTenTen	5.8 miliardi	Scaricati dal web nel 2016
2. RIDIRE	1,5 miliardi	Scaricati dal web dal 2009 al 2013
3. Araneum Italicum Maius	1,2 miliardi	Scaricati dal web nel 2014
4. PAISÀ	250 milioni	Scaricati dal web nel 2010
5. CorIS/ CoDIS	150 milioni	Prodotti dagli anni '80 al 2017

4.4. Progettazione delle stringhe di ricerca

Una volta selezionato il *corpus* PAISÀ come risorsa di riferimento per la ricerca, le attività successive si sono concentrate sulla sua consultazione, al fine di reperire al suo interno testi contenenti i tratti linguistici elencati da Sabatini (1990). Nello specifico, ciò ha comportato l'elaborazione di stringhe di ricerca in linguaggio CQL (cfr. 4.7.1) funzionali al reperimento nel *corpus* delle occorrenze dei suddetti tratti linguistici (cfr. 4.7.2), l'interrogazione del *corpus* e il salvataggio dei

risultati ottenuti (cfr. cap. 5) e, infine, il calcolo della loro ricorrenza all'interno del corpus, al fine di determinare i fenomeni linguistici più frequenti (cfr. cap. 5).

4.4.1. Il linguaggio CQL

CQL (*Corpus Query Language*) è un linguaggio di interrogazione di corpora che permette di ricercare strutture grammaticali o lessicali complesse (Evert, 2009), funzione indispensabile per il reperimento dei tratti linguistici elencati da Sabatini (1990). È stato sviluppato dall'*Institut für Maschinelle Sprachverarbeitung* dell'Università di Stoccarda nei primi anni '90, dove è stato impiegato nel programma di interrogazione di corpora IMS Open Corpus Workbench (Christ, 1994a; 1994b). Un'estensione del linguaggio originale, rispetto al quale presenta alcune differenze¹⁸, è utilizzata sia nel programma di interrogazione di corpora Sketch Engine (Rychlý, 2007, p. 69) che nella sua versione *open source* NoSketch Engine: quest'ultima è stata impiegata nell'ambito di questo progetto per la consultazione del corpus PAISÀ, in seguito ad alcune difficoltà tecniche riscontrate con l'uso dell'interfaccia originale.

Sia Sketch Engine, sia NoSketch Engine mettono a disposizione diverse modalità di ricerca. La modalità di ricerca base è *concordance*, con la quale è possibile inserire nella barra di ricerca una parola o una breve sequenza di parole. Nel caso in cui si inserisca il lemma di una parola (esempio: *andare*), verranno reperite le occorrenze di tutta la sua coniugazione, mentre digitando una forma flessionale (esempio: *vai*) si otterranno le occorrenze solo di quella specifica forma inserita.

Nella modalità di ricerca avanzata è, invece, possibile scegliere tra le seguenti opzioni: *lemma*, *phrase*, *word*, *character* e *CQL*.

Con l'opzione *lemma*, inserendo nella maschera di ricerca la forma base di una parola, si otterranno le occorrenze di tutte le sue forme flessionali. Richiamando l'esempio sopra menzionato, digitando un verbo all'infinito, verranno reperite le occorrenze di tutta la sua coniugazione; lo stesso vale inserendo un aggettivo o un sostantivo al singolare o una preposizione semplice con cui si possano ottenere delle preposizioni articolate.

¹⁸ www.sketchengine.eu/documentation/corpus-querying, dove si descrive la versione realizzata tramite Manatee 2.122, risalente all'aprile 2015.

L'opzione *phrase* permette di reperire una sequenza di parole nell'ordine esatto in cui sono state digitate.

Word troverà le occorrenze di una parola che rispettano la grafia esatta di come è stata digitata. Tale procedura si utilizza quando si ricerca una parola considerandola solo come una determinata sequenza di grafemi e, dunque, indipendentemente dalla categoria grammaticale a cui appartiene.

L'opzione *Character* troverà tutte le parole che contengono il carattere o la sequenza di caratteri digitati. Ad esempio, digitando *sch* si otterranno le occorrenze di tutte le parole che contengono tale sequenza (esempi: *schizzo*, *rischio*, ecc.).

Infine, l'opzione *CQL* si basa sull'uso del *Corpus Query Language* per effettuare ricerche più complesse attraverso alcuni attributi (*word*, *lemma*, *tag*, *lembo* e *lowercase*), a cui possono combinarsi le espressioni regolari (*regular expressions*) e i caratteri jolly (*wild cards*). Di seguito verrà data una breve descrizione solo delle modalità di ricerca utilizzate nella progettazione delle stringhe.

Le opzioni *Word* e *Lemma*, già citate tra le varie modalità di ricerca, possono essere utilizzate anche tramite CQL secondo queste due modalità: ad esempio, la stringa `[word="gatto"]` o semplicemente l'uso delle virgolette come in "gatto" permettono di identificare tutte le occorrenze della parola ricercata rispettando la grafia esatta in cui è stata digitata, senza dover impostare la ricerca tramite la categoria grammaticale (in questo caso come sostantivo). Le stesse considerazioni valgono per la modalità di ricerca *Lemma* in CQL: la stringa `[lemma="andare"]` consente di reperire le occorrenze di tutta la sua coniugazione, senza dover impostare la ricerca tramite la categoria grammaticale *verbo*.

L'attributo *tag* contiene informazioni sulla parte del discorso di ciascun *token* e di solito anche altre informazioni grammaticali o morfologiche come numero, genere, tempo, ecc. I *tag* vengono assegnati automaticamente da un *tagger*. L'uso del *tag* per la ricerca consente di trovare tutte le occorrenze etichettate con la stessa parte del discorso (esempio: `[tag="A"]` identifica gli aggettivi), come verrà illustrato nel dettaglio più avanti.

Le espressioni regolari sono un insieme di simboli speciali che possono essere utilizzati per ricercare modelli anziché caratteri specifici,

ad esempio per trovare tutte le parole che iniziano, contengono o finiscono in una specifica sequenza di caratteri. Un esempio, che è stato spesso utilizzato nelle stringhe di ricerca prodotte è (?i). Questa espressione regolare indica la funzione *case insensitive*, che serve a reperire occorrenze con lettera maiuscola e minuscola e quindi i casi in cui la parte del discorso ricercata può trovarsi a inizio frase o al suo interno.

La *wild card*, invece, è un carattere che all'interno di una stringa rappresenta altri caratteri o sequenze di caratteri. Un esempio è l'uso dell'asterisco (*), che rappresenta genericamente un numero indefinito e variabile di caratteri.

Per apprendere il linguaggio CQL e l'uso del programma NoSketch Engine sono stati utilizzati il manuale di Nesi e Thompson (2014) e la documentazione presente sul sito di Sketch Engine¹⁹, che hanno consentito di familiarizzare con le diverse modalità di interrogazione del *corpus* e di elaborare delle stringhe di ricerca funzionali al reperimento di strutture grammaticali complesse. Per quest'ultima attività è stata necessaria, inoltre, la consultazione del *tagset* morfosintattico (ISST TANL *tagset*)²⁰ e del *tagset* delle dipendenze sintattiche (ISST TANL DEPTagset)²¹ utilizzati specificamente per l'annotazione del corpus PAISÀ. L'ISST TANL è un sistema di annotazione per parti del discorso: attraverso un set di 14 *tag* generici (*coarse-grained pos tag*) e 37 *tag* specifici (*fine-grained pos tag*) a ogni *token* contenuto nel corpus vengono associate informazioni relative alle categorie morfosintattiche a cui esso appartiene. Ad esempio, nell'espressione “il mio libro” a *mio* corrisponde sia il *tag* generico che identifica in generale tutti gli aggettivi, ovvero [cpostag="A"] oppure [postag="A.*"], sia il *tag* specifico [postag="AP"], che identifica gli aggettivi possessivi.

Alle parti del discorso possono, inoltre, essere associate delle caratteristiche morfosintattiche (*morpho-syntactic features*, il cui *tag* è *feats*), di cui di seguito a titolo di esempio si riportano alcune tipologie: genere (*tag=gen*), numero (*tag=num*), persona (*tag=per*), modo (*tag=mod*)

¹⁹ www.sketchengine.eu/documentation/cql-basics.

²⁰ www.corpusitaliano.it/static/documents/POS_ISST-TANL-tagset-web.pdf.

²¹ <http://www.italianlp.it/docs/ISST-TANL-DEPTagset.pdf>.

e tempo (*tag=tmp*)²². Ciascuna di esse è ulteriormente esplicitata da una sottocategoria di caratteristiche che forniscono informazioni più dettagliate sulle parti del discorso, a cui corrisponde un set di *tag*. Ad esempio, nella seguente stringa, che identifica i sostantivi plurali, la caratteristica morfosintattica del numero (*feats=num*) è ulteriormente esplicitata dal plurale (*p*): [*postag="S.*" & feats=".*num=p.*"*]. L'uso dell'ISST TANL DEPTagset ha permesso di progettare delle stringhe di ricerca che tengano conto delle relazioni di dipendenza sintattica. Tra queste si citano a titolo di esempio le relazioni di soggetto, oggetto e di verbo principale, che sono state utilizzate per ricercare i diversi casi di soggetto in posizione rematica rispetto al predicato e del verbo in forma pronominale per indicare partecipazione affettiva, a cui si farà riferimento nel dettaglio nella sezione 4.6.2 (cfr. i punti 7c-h e il punto 9 dell'elenco delle stringhe prodotte).

Quest'ultimo *tagset* si è rivelato una risorsa fondamentale per poter interrogare il corpus rispetto ad alcuni fenomeni grammaticali che potremmo definire più complessi, poiché coinvolgono la sintassi del periodo e non la semplice co-occorrenza di parti del discorso all'interno della stessa frase. Tuttavia, come verrà precisato sia nella sezione 4.6.2, sia nel capitolo 5, è bene fare alcune riflessioni su alcune scelte compiute nella formulazione delle stringhe stesse. È necessario precisare che i risultati ottenuti dall'utilizzo di stringhe che fanno uso dell'annotazione sintattica di ricerca dipendono necessariamente dalla formulazione stessa delle stringhe. Nell'ambito di questa ricerca la progettazione delle stringhe si è basata su un'annotazione per parti del discorso (POS, *parts-of-speech*), che non sono sempre lo strumento ideale per ricercare fenomeni sintattici, specialmente se questi coinvolgono elementi della frase a lunga distanza. Per questi ultimi occorrerebbe fare ricerche basate su un'annotazione per analisi sintattica (*parsing*): ciò è tecnicamente possibile attraverso il linguaggio CQL, ma sarebbe stato necessario progettare delle stringhe di ricerca notevolmente più complesse.

Nel caso, ad esempio, della ricerca delle diverse tipologie di soggetto in posizione rematica rispetto al predicato, le stringhe massimizzano

²² Le caratteristiche qui descritte non sono standard, bensì si applicano al *tagset* utilizzato nel corpus PAISÀ.

le possibilità che la parte del discorso che segue il verbo sia effettivamente il soggetto del verbo che occorre a inizio frase, ma non lo garantisce, dato che imporre tramite il linguaggio CQL la ricerca di un determinato soggetto legato ad un determinato verbo è un'operazione molto complessa. È il caso, ad esempio del punto 7e della lista che verrà di seguito esaminata, ossia il fenomeno del soggetto in posizione rematica rispetto a verbi composti con l'ausiliare *avere*, senza l'oggetto diretto tra participio e soggetto (esempio: *ha cancellato lei la pagina*; cfr. sezione 4.6.2). Una prima versione della stringa includeva anche i casi di frasi con soggetto ellittico, ad esempio *ho visto Marco*, in cui il soggetto ellittico è *io*, ma la funzione di soggetto all'interno del corpus è stata assegnata a *Marco*. Ciò ha permesso di ipotizzare che il *tagger* abbia assegnato la funzione di soggetto (`deprel="subj"`) a qualunque sostantivo o pronome collocato subito dopo il verbo: una procedura che non sempre ha condotto a risultati corretti.

Inoltre, sono state riscontrate alcune criticità del tagset nel caso 7d-Soggetto in posizione rematica rispetto a verbi composti con l'ausiliare *essere* (esempio: *Sono state moltissime le reazioni*): nella produzione della stringa è emerso che il *tagset* non fa distinzione tra participi passati o presenti, dunque la ricerca è stata impostata sul solo *tag* del modo participio. Sempre nella ricerca di questo fenomeno grammaticale sono state riscontrate delle discrepanze tra alcuni *tag* relativi ai tempi passati dei verbi, riportati nella guida al *tagset*, e quelli poi risultati effettivamente in uso nel corpus, a cui si farà riferimento nel dettaglio nella sezione 4.6.2. In altri casi, pur non essendo stati riscontrati dei limiti nell'uso del linguaggio CQL, è stato necessario impostare le stringhe trovando un equilibrio tra la creazione di stringhe ampie, che identificassero tutti i casi che rientrano nel fenomeno indagato, e stringhe più mirate e restrittive, che non includessero troppo "rumore", ossia casi che, pur presentando delle analogie strutturali con il fenomeno indagato, non fossero inerenti. Un esempio è il punto 12c-*Che* pseudorelativo (una sottocategoria del *che* polivalente. Esempio: *li vedo che scendono*. Cfr. sezione 4.6.2), ove la prima versione della stringa²³ includeva tra i ri-

²³ La prima versione della stringa mirata a reperire i casi di *che* pseudorelativo è tratta da Budassi (2017:165), Rispetto all'originale presenta le sole modifiche relative al *tagset* in uso in PAISÀ:

sultati anche molte occorrenze del *si* impersonale (esempio: *si vede che hanno qualche domanda*), che chiaramente non rientrano nel fenomeno indagato. Di conseguenza, nella seconda fase di progettazione della stringa si è scelto di escludere il pronome *si* dai pronomi ricercabili, trovando un compromesso tra la perdita di occorrenze corrette e l'esclusione di casi non inerenti²⁴.

Un altro esempio dell'esigenza di restringere il campo della ricerca è il caso 9- impiego del verbo in forma pronominale per indicare partecipazione affettiva (esempio: *mi guardo un film*), ove, utilizzando la prima versione²⁵ delle stringhe prodotte, la ricerca restituiva molte occorrenze che non rientrano nell'ambito del fenomeno in questione (esempi: *Mi faccio (del) male; mi sono fatto accompagnare; mi sono visto rimproverare; mi faccio prendere; mi sono visto costretto*, eccetera). Pertanto, allo scopo di ridurre il "rumore" tra i risultati, si è preferito restringere la ricerca, modificando la stringa e imponendo come elemento finale la presenza di un gruppo nominale col valore di oggetto²⁶, come verrà illustrato nel dettaglio nella sezione 4.6.2.

La riflessione derivante da quanto esposto, che verrà ripresa successivamente nel corso di questo capitolo e nel capitolo 5, è che le occorrenze ottenute dipendano notevolmente dalla formulazione delle stringhe

```
[postag="P.*"] [lemma="vedere|sentire|guardare|scorgere|notare|osservare|mirare|ascoltare|udire|intendere|percepire"]
[word="che"] [postag="V.*"]
```

²⁴ La seconda versione della stringa, modificata rispetto a Budassi (*ibidem*), con l'esclusione del pronome *si*:

```
[postag="P.*&lemma!="si"] [lemma="vedere|sentire|guardare|scorgere|notare|osservare|mirare|ascoltare|udire|intendere|percepire"]
[word="che"] [postag="V.*"]
```

Questa versione esclude dalla ricerca il pronome *si* tramite la seguente parte della stringa: `[postag="P.*&lemma!="si"]`.

²⁵ Un esempio della prima versione della stringa per la prima persona singolare dei verbi riflessivi "affettivi", con tempi verbali non composti:

```
[word="(?)mi"] [lemma="mangiare|leggere|bere|fare|guardare|vedere"& feats=".*num=s.*" & feats=".*per=1.*"]
```

²⁶ Un esempio della seconda versione della stringa per la prima persona singolare dei verbi riflessivi "affettivi", con tempi verbali non composti:

```
[word="(?)mi"] [lemma="mangiare|leggere|bere|fare|guardare|vedere"& feats=".*num=s.*" & feats=".*per=1.*"]
[postag="R.*|A.*|D.*"] {0,2} [postag="P.*|S.*" &deprel="obj"]
```

di ricerca, la quale a sua volta è in alcuni casi influenzata o dai limiti imposti dal tipo di annotazione su cui si basa la ricerca stessa, o dalla necessità di ricercare un equilibrio tra la copertura ad ampio spettro di un fenomeno grammaticale e il rischio di reperire anche grandi quantità di casi non inerenti. Alla progettazione stringhe di ricerca e alle scelte sottostanti si farà riferimento nella sezione 4.6.2.

4.4.2. Elaborazione delle stringhe di ricerca

Il punto di partenza per l'elaborazione delle stringhe di ricerca è stata la rilettura della lista dei 14 tratti linguistici elencati da Sabatini (1990). Da questa lista è stato escluso per le finalità della ricerca il primo fenomeno elencato, ossia l'uso dei pronomi *lui* e *lei* e *loro* in funzione di soggetto al posto di *egli/ella* ed *essi/esse*, ritenendo che tali forme siano ormai integrate nello standard:

L'uso di *lui*, *lei* e *loro* come pronomi soggetto è stato a lungo sanzionato dai grammatici del passato e oggetto di dibattito nella storia della grammatica italiana fin dal Cinquecento. Oggi non vi sono più dubbi sull'accettabilità di queste forme²⁷.

Di conseguenza, partendo dai 13 tratti linguistici rimanenti, sono state elaborate 57 stringhe di ricerca (*query*) in linguaggio CQL. Com'è evidente, il numero di stringhe è nettamente superiore ai tratti elencati. Come verrà di seguito illustrato, l'elaborazione di più stringhe si è resa necessaria principalmente in due casi: per i fenomeni linguistici articolati in sottofenomeni, che chiaramente vanno cercati separatamente, e nel caso in cui le specifiche caratteristiche sintattiche o morfologiche di un determinato fenomeno hanno inevitabilmente richiesto la progettazione di più stringhe di ricerca (si veda ad esempio il caso dei verbi in forma pro-

²⁷ EGLI/LUI, in *La grammatica italiana*, Istituto dell'Enciclopedia Italiana www.treccani.it/enciclopedia/egli-lui_%28La-grammatica-italiana%29, versione html, pertanto priva di pagine. Si faccia riferimento, a questo proposito, anche alle motivazioni linguistiche fornite da Francesco Sabatini in risposta ad un quesito sull'uso di lui/egli soggetto posto sulla rivista *La Crusca* per voi (n. 1, ottobre 1990, pp. 6-8), disponibile online: www.accademiadellacrusca.it/it/lingua-italiana/consulenza-linguistica/domande-risposte/soggetto.

nominale per indicare partecipazione affettiva, per i quali un'unica stringa di ricerca non avrebbe potuto coprire tutte le forme pronominali usate per l'intera coniugazione). Quindi, il numero di stringhe finali è in alcuni casi correlato linguisticamente alla presenza di sottofenomeni, in altri casi dall'oggettiva impossibilità di coprire con un'unica stringa di ricerca tutte le occorrenze di un fenomeno, a causa della sua complessità intrinseca. Ciò ha portato, quindi, il numero finale di fenomeni indagati a 26.

Qui di seguito si riporta per ciascun fenomeno la descrizione delle stringhe di ricerca prodotte, di cui viene fornita una breve descrizione. L'elenco completo di tutte le stringhe è, inoltre, disponibile in appendice (cfr. Appendice 2).

1) impiego di *gli* nella forma dativale al posto di *le* (a lei) e *loro* (a loro). Allo scopo di reperire occorrenze in cui *gli* può trovarsi a inizio frase o al suo interno, nella stringa è stata inserita la già menzionata espressione regolare (?i), che permette di ottenere occorrenze sia con lettera maiuscola, sia con lettera minuscola. A seguire vi è un verbo, al fine di individuare tutti i casi in cui *gli* si trova in posizione preverbale (esempi: *gli ho detto*; *gli rispondo*, ecc).

Poiché si è imposto nella stringa che *gli* fosse seguito immediatamente da un verbo e ciò può verificarsi solo nel caso in cui esso funga da pronome (esempio: *gli dico* = pronome + verbo) e non da articolo determinativo maschile plurale (esempio: *gli alberi*), si è ritenuto superfluo progettare una stringa di ricerca più complessa, in cui *gli* venisse indicato specificamente come pronome clitico (esempio, [postag="PC" & word="(?)gli"]). Per tale ragione, la sua ricerca è stata impostata come semplice parola [word="(?)gli"].

La discriminazione tra i casi in cui il pronome *gli* si riferisce secondo lo standard a individui di genere maschile (esempio: *ho visto Marco e gli ho detto...*) o, nel registro colloquiale, a individui di genere femminile singolare (esempio: *ho parlato ad Anna e gli ho detto...*) o di genere maschile/femminile plurale al posto di *loro* (esempio: *ho chiamato i miei amici/ le mie amiche e gli ho detto...*), è stata fatta manualmente, tramite la consultazione delle occorrenze ottenute e la lettura del contesto circostante. Segue la stringa prodotta:

[word="(?)gli"] [postag="V.*"]

2) impiego del partitivo preceduto da preposizione (esempio: *con degli amici*). Per questo fenomeno è stata progettata una stringa in cui una preposizione semplice, indicata dal *tag* E, purché diversa da *prima* e *dopo*, sia immediatamente seguita da una preposizione articolata (*tag* = EA) composta da *di* + articolo:

```
[postag="E" & word!="prima|dopo"] [postag="EA" & lemma="di"]
```

È necessaria una precisazione. *Prima* e *dopo* non fanno parte delle preposizioni semplici (*di, a, da, in, con, su, per, tra, fra*), bensì delle cosiddette preposizioni improprie, ossia parole appartenenti ad altre categorie grammaticali, come aggettivi, verbi o avverbi, usate anche in funzione di preposizioni. La necessità di impostare l'esclusione di *prima* e *dopo* dalla stringa di ricerca è dovuta al fatto che il *tagger* usato per il corpus PAISÀ non distingue tali categorie di preposizioni. Pertanto, l'esclusione di *prima* e *dopo* permette di evitare il reperimento di occorrenze come “*prima dopo di + articolo*” (esempio: *prima del concerto, dopo dei pasti*, ecc.) che non costituiscono casi di partitivo, ma semplici espressioni di tempo;

3) impiego di dislocazioni a sinistra (*quel libro l'ho già letto*), a destra (*l'ho già letto, quel libro*) della parte rematica della frase con la ripresa pronominale atona e **costruzioni ad esse assimilabili**, tipo: *di pane non ne ho più, a me mi piace leggere*, con uso pleonastico della particella pronominale. Nella progettazione delle stringhe si è tenuto conto solo della dislocazione a sinistra dell'elemento pronominale (*a me mi*) e del cosiddetto “doppio partitivo” (*di pane non ne ho più*) e non di tutti i casi di dislocazione a sinistra e a destra di altri elementi grammaticali, come il gruppo nominale preceduto o meno da preposizione (a sinistra: *il libro, lo ha preso Giacomo; a Giorgio gli porto una bottiglia di vino*; a destra: *lo ha preso Giacomo, il libro; gli porto una bottiglia di vino, a Giorgio*) o il verbo (*piovere, non è piovuto, per imparare, Carla impara*) eccetera. Nel compiere tale scelta si è tenuto conto dell'obiettivo finale del progetto, ossia la progettazione di materiali didattici destinati a non-italofoni nativi. In questa prospettiva, si è ritenuto che i casi di dislocazione dell'elemento pronominale potessero risultare più produttivi dal punto di vista della progettazione didattica²⁸ e

²⁸ A conferma della crescente attenzione dedicata a questo fenomeno linguistico nella progettazione di materiali didattici, si rinvia alla sezione 5.2 di questo volume.

più utili dal punto di vista degli utenti finali delle unità didattiche. Ciò è dovuto alla complessità per un parlante non nativo di un raddoppiamento della forma pronominale (forma tonica + atona, ad esempio: *a me mi piace*), rispetto invece agli altri casi di dislocazione sopra citati (esempio: *quel libro l'ho già letto; l'ho già letto, quel libro*, ecc.), nei quali il significato della frase comunque è facilmente deducibile dal contesto. Sulla base di tali motivazioni sono state elaborate le seguenti stringhe di ricerca:

3a) doppio partitivo (esempio: *di pane non ne ho più*): per questo fenomeno è stata progettata una stringa in cui la preposizione *di*, scritta con lettera maiuscola o minuscola: 1) non è mai preceduta da qualunque tipo di sostantivo o aggettivo, per evitare occorrenze in cui la preposizione *di* sia parte di un'espressione che esprime il possesso (esempio: *casa di Anna*) o la materia (esempio: *scatola di cartone*) o di un comparativo (esempi: *più grande di; maggiore/minore di*); 2) è immediatamente seguita da qualunque tipo di sostantivo (purché diverso da *fatto, seguito, conseguenza, norma, solito*, per evitare occorrenze come: *di fatto, di seguito*, ecc.), poi dalla particella pronominale *ne* e, infine, da un verbo di qualunque tipo. Ne deriva:

```
[postag!="S.*|A.*"] [word="( ?i) di"] [postag="S.*"
& word!="fatto|seguito|conseguenza|norma|solito"]
[word="ne"] [postag="V.*"]
```

3b) dislocazione a sinistra dell'elemento pronominale (*a me mi, a te ti*, ecc.). In questo caso è stato necessario progettare 8 stringhe di ricerca per le 7 forme pronominali toniche precedute dalla preposizione *a* e seguite dalla forma atona (*a me mi, a te ti, a lui gli, a lei le, a noi ci, a voi vi, a loro gli*), più un'ottava forma che corrisponde ad un'alternativa della forma *a lei le*, ossia *a lei gli* (esempio: *a lei gli dico*), tipica delle varietà substandard. Di seguito si riporta la prima delle 8 stringhe progettate, ossia la preposizione *a*, scritta con lettera maiuscola o minuscola, seguita immediatamente da *me* e *mi*, rinviando all'Appendice 2 per una visualizzazione completa del fenomeno:

```
[word="( ?i) a"] [word="me"] [word="mi"]
```

4) impiego del connettivo *per cui* (derivato per ellissi da *motivo/fatto/ragione per cui*, oppure per generalizzazione di casi con normali ante-

cedenti nominali) **in funzione di congiunzione causale-consecutiva**. È stata progettata la seguente stringa, in cui *per*, scritto con lettera maiuscola o minuscola, è stato ricercato come semplice parola seguita immediatamente da *cui*:

```
[word="( ?i)per" ] [word="cui" ]
```

5) impiego di *cosa*? interrogativo diretto e indiretto al posto di *che cosa*? Per il fenomeno in questione sono state progettate due stringhe.

Nel caso dell'**interrogativa diretta** è stato imposto che la stringa inizi con qualunque segno di punteggiatura ("F"), per delimitare la fine della frase precedente; in apertura di frase può esservi o meno²⁹ un elemento grammaticale, purché sia diverso da qualunque aggettivo ("A"), determinante ("D")³⁰, articolo determinativo o indeterminativo ("R"), per evitare di reperire occorrenze simili a *bella cosa, questa/ quell'al/alcuna/ tale/ qualsiasi cosa, la cosa, una cosa*, ecc. e, allo stesso tempo, non impedire il reperimento di casi in cui *cosa* sia preceduto, ad esempio, da congiunzioni, avverbi ecc. (Esempi: *ma cosa dici? E ora cosa fai?* ecc.). Immediatamente dopo vi è *cosa*, scritto maiuscolo o minuscolo (a seconda che sia posto ad apertura di frase o meno); a seguire è stato imposto che non vi sia immediatamente un punto interrogativo, ma un numero di parole che spaziano da 0 a 10 e, infine, il punto interrogativo. Segue la stringa:

```
[cpostag="F" ] [postag!= "A.*|D.*|R.*" ]? [word="( ?i)cosa" ] [word!="\?" ] {0,10} [word="\?" ]
```

Nel caso dell'**interrogativa indiretta** (esempio: *mi domando cosa...*) in apertura della stringa vi sono i lemmi dei verbi *chiedere* e *domandare*; qualora tra questi verbi e *cosa* siano presenti degli elementi grammaticali (ad esempio: *mi chiedo allora cosa...*), la cui presenza opzionale è simboleggiata dal punto interrogativo, si impone che questi siano diversi sia dai segni di punteggiatura (per evitare di reperire esempi di discorso diretto), sia da *che* (per evitare l'interrogativo standard *che cosa*). Segue la stringa:

```
[lemma="chiedere|domandare" ] [cpostag!="F" & word!="che" ]? [word="( ?i)cosa" ]
```

²⁹ La presenza opzionale è rappresentata dal simbolo del punto interrogativo.

³⁰ Per la voce "determinante", si rinvia alla guida: www.corpusitaliano.it/static/documents/POS_ISSST-TANL-tagset-web.pdf.

6) impiego di *e*, *ma*, e di connettivi testuali come *allora*, *comunque* in posizione iniziale di discorso o di enunciato. In questi casi si è imposto che la stringa inizi con un punto o un punto e virgola, seguiti dalla congiunzione o dal connettivo testuale in questione con lettera maiuscola o minuscola (a seconda che siano preceduti o meno dal punto):

- a) `[word="\.|\;"] [word="(?)E"]`
- b) `[word="\.|\;"] [word="(?)Ma"]`
- c) `[word="\.|\;"] [word="(?)Allora"]`
- d) `[word="\.|\;"] [word="(?)Comunque"]`

7) soggetto in posizione rematica rispetto al predicato, talvolta con discordanza di numero con questo (esempi: *non ci sono soldi*; o anche il tipo: *niente soldi!*). Data la complessità del fenomeno in questione, che si articola in numerosi sottocasi, è stato necessario elaborare 8 stringhe di ricerca, partendo dai casi più particolari, caratterizzati dalla discordanza di numero tra soggetto e predicato:

7a) *niente* seguito da un sostantivo plurale (esempio: *niente soldi*): è stata elaborata una stringa in cui la parola *niente* è seguita da un sostantivo di genere plurale:

`[word="niente"] [postag="S.*" & feats=".*num=p.*"]`

7b) sostantivo plurale seguito da *non ce n'è* (esempio: *soldi non ce n'è*): è stata elaborata una stringa in cui un sostantivo di genere plurale è seguito da *non + ce + n'* e, infine, dal verbo *essere* alla terza persona singolare:

`[postag="S.*" & feats=".*num=p.*"] [word="non"]
[word="ce"] [word="n`"] [lemma="essere" &
feats=".*num=s.*" & feats=".*per=3.*"]`

7c) Soggetto in posizione rematica rispetto a verbi non composti (esempio: *parla il presidente*): nella costruzione della stringa sono stati esclusi tutti i casi di frasi che nella costruzione non marcata richiedono il soggetto in posizione post-verbale (per il dettaglio, si rinvia al capitolo 2)³¹.

³¹ Trattandosi di una stringa mirata a ricercare le occorrenze di soggetti in posi-

Pertanto, è stato imposto che la stringa inizi con un punto, per delimitare la fine della frase precedente; in apertura di frase vi è un verbo con lettera maiuscola, che sia diverso da *essere*, *occorrere*, *servire*, *avere* e *esistere*, avente la funzione di verbo principale e coniugato al modo indicativo. A seguire possono essere presenti in quantità compresa tra 0 e 2 un articolo determinativo o indeterminativo (R.*) e/o un aggettivo (A.*) o, in alternativa all'articolo, un determinante (D.*) (ad esempio: *questo*, *codesto*, *quello*, *quale*, *alcuno*, *certo*, *tale*...) e, infine, un sostantivo o un pronome di qualunque tipo ("S.*|P.*") che ricopra il ruolo di soggetto:

```
[word="\." ] [postag="V.*" & word=" [A-Z] .+" &
lemma!="essere|occorrere|servire|avere|esistere"
& deprel="ROOT" & feats=".*mod=i.*"]
[postag="R.*|A.*|D.*"] {0,2} [postag="S.*|P.*" &
deprel="subj"]
```

Esempi di occorrenze reperibili con questo tipo di stringa (per pura praticità verranno presentati solo casi con soggetto al singolare): parla (V.*) Luca (S.*); parla (V.*) lui (P.*); interviene (V.*) il/un (R.*) deputato (S.*); interviene (V.*) il/un (R.*) nostro (A.*) deputato (S.*); recita (V.*) il/un (R.*) simpatico (A.*) artista (S.*); gioca (V.*) questo/quel (D.*) calciatore (S.*), gioca (V.*) questo/quel (D.*) bravo (A.*) calciatore (S.*), ecc. Come si evince dagli esempi, la formulazione di questa stringa non consente, ad esempio, di ricercare occorrenze in cui il sostantivo che funge da soggetto sia preceduto da più di un aggettivo (esempio: *è arrivato il mio carissimo amico*, ove *mio* è un aggettivo possessivo e *carissimo* è un aggettivo qualificativo). Le occorrenze reperite presentano, infatti, un solo aggettivo, sia esso possessivo, qualificativo o di altro tipo. Pertanto, è bene ricordare che i risultati ottenuti dall'utilizzo di questa stringa di ricerca dipendono necessariamente dalla for-

zione rematica rispetto a verbi non composti, i verbi *essere* e *avere* sono stati necessariamente esclusi dalla stringa per evitare di ottenere occorrenze in cui tali verbi fungono da ausiliari dei tempi composti. Inoltre, sono stati esclusi dalla ricerca anche i verbi *occorrere*, *servire* (che possono essere usati entrambi impersonalmente, mentre nel caso del verbo *servire*, esso può essere usato anche transitivamente) ed *esistere*, perché rientrano nella classe dei verbi intransitivi detti *inaccusativi* (cfr. capitolo 2). L'esclusione di questi verbi è frutto di una seconda fase di elaborazione della stringa, al fine di evitare di includere nei risultati le frasi inaccusative, che ricorrevano nella prima versione.

mulazione della stringa stessa come già anticipato nella sezione 4.6.1 in riferimento ad altri casi³².

7d) Soggetto in posizione rematica **rispetto a verbi composti con l'ausiliare *essere*** (esempio: *Sono state moltissime le reazioni*): valgono le stesse considerazioni fatte per la stringa precedente, a differenza della quale si è imposto che il verbo sia composto dalle voci del lemma del verbo *essere*, seguite dal participio passato di qualunque verbo, e abbia come soggetto un qualunque sostantivo o pronome:

```
[word="\." ] [lemma="essere"] [postag="V.*" &
feats="*.mod=p.*" ] [postag="R.*|A.*|D.*" ] {0,2}
[postag="P.*|S.*" & deprel="subj"]
```

Sebbene nella descrizione del fenomeno si sia fatto correttamente riferimento al participio passato (es. è *arrivata Anna*), è necessario precisare che, al momento della realizzazione di questa ricerca, nel corpus PAISÀ i participi non risultano taggati come “passati” o “presenti”, bensì solo come participi: la *feature* *tmp=r* prevista nella guida al *tagset* per il passato (*past tense*), e dunque applicabile anche al participio passato, in realtà non risulta attribuita ad alcun elemento linguistico. Ad esempio, il participio *andato* risulta taggato come: *V/Vnum=s|mod=p|gen=m*, ossia come un verbo avente numero singolare, modo participio e genere maschile. Tale considerazione vale per tutti i casi successivi in cui verrà menzionato un participio.

Per avere un riscontro circa l'utilizzo del *tag* per i tempi passati è stata fatta inoltre una verifica inserendo nella maschera di ricerca alcuni verbi al passato prossimo, passato remoto e imperfetto (esempio: è *andato*; *andai*; *andava*). Ne risulta che: non esiste un *tag* specifico che identifica il passato prossimo, poiché l'ausiliare e il participio sono sempre stati taggati separatamente (Esempio: è=*V/VA/num=s|per=3|mod=i|ten=p*;

³² Ad un successivo riesame della stringa prodotta, si è valutato che possa eventualmente essere implementata nel seguente modo:

```
[word="\." ] [postag="V.*" & word="[A-Z].+" & lemma!="essere|occorrere|servire|avere|esistere" & deprel="ROOT" &
feats="*.mod=i.*" ] [postag="R.*|D.*" ] ? [postag="A.*" ] {0,}
[postag="S.*|P.*" & deprel="subj"]
```

Questa intuizione, giunta al termine del presente studio, potrà essere il punto di partenza per ulteriori sviluppi della ricerca sul tema.

andato=*V/V num=s; mod=p; gen=m*); i verbi al passato remoto sono taggati con *ten=s*, anche se questa voce non è riportata nell'elenco della guida al *tagset*, mentre, contrariamente a quanto riportato nella guida, i verbi all'imperfetto sono taggati con *ten=i* (e non *tmp=i*). Data quest'ultima discrepanza tra il *tag* corretto dell'imperfetto reperito nel corpus (*ten=i*) e quello erroneamente riportato nella guida (*tmp=i*), è stato fatto un ultimo tentativo di ricerca del participio passato, modificando il *tag* *tmp=r*, previsto nella guida per il tempo passato, in *ten=r*. Tuttavia, quest'ultimo non è stato reperito tra i risultati, confermando l'osservazione originaria sull'assenza di distinzione tra participi presenti e passati.

7e) Soggetto in posizione rematica rispetto a verbi composti con l'ausiliare avere, senza l'oggetto diretto tra participio e soggetto (esempio: *ha cancellato lei stessa la pagina*): valgono le stesse considerazioni fatte per la stringa precedente, a differenza della quale si è imposto che il verbo sia composto dalle voci del lemma del verbo *avere*, seguite dal participio passato di qualunque verbo. Inoltre, si è imposto che immediatamente dopo il verbo non vi sia la parola *che*. Questo accorgimento è frutto di una seconda fase di elaborazione della stringa, al fine di evitare di includere nei risultati i casi con soggetto ellittico, in cui o il verbo è immediatamente seguito dalla congiunzione subordinante *che* (esempio: *ho pensato che*), oppure vi è un complemento oggetto tra il verbo e la congiunzione *che* (esempio: *ho visto Marco che*). A fronte di questi casi, reperiti in una prima fase di ricerca con una versione differente della stringa, si è ipotizzato che, nella fase di *part-of-speech tagging*, in cui alle parole dei testi che compongono il corpus PAISÀ sono stati attribuiti i *tag* relativi alle parti del discorso, il lemmatizzatore abbia assegnato la funzione di soggetto (*deprel="subj"*) a qualunque sostantivo o pronome collocato subito dopo il verbo. In alcuni casi tale procedura si è rivelata corretta, ma nel caso delle frasi con soggetto ellittico la categoria di soggetto è stata assegnata a elementi della frase che in realtà erano un complemento oggetto (ad esempio, in *ho visto Marco* la funzione di soggetto è stata assegnata a *Marco*). Pertanto, è stata progettata la seguente stringa:

```
[word="\." ] [lemma="avere" ] [postag="V.*" &
feats=".*mod=p.*" ] [postag="R.*|A.*|D.*" ] {0,2}
[postag="P.*|S.*" & deprel="subj" & word!="che" ]
```

7f) Soggetto in posizione rematica rispetto a verbi composti con l'ausiliare avere, con l'oggetto diretto tra participio e soggetto (esempio: *ha suscitato sdegno l'articolo che racconta...*): valgono le stesse considerazioni fatte per la stringa precedente, a differenza della quale si è imposto che tra il verbo composto e il soggetto vi sia un gruppo nominale con la funzione di oggetto (*deprel="obj"*):

```
[word="\." ] [lemma="avere"] [postag="V.*" &
feats=". *mod=p.*" ] [postag="R.*|A.*|D.*" ] {0,2}
[postag="S.*" & deprel="obj" ] [postag="R.*|A.*|D.*" ]
{0,2} [postag="P.*|S.*" & deprel="subj" & word!="che"]
```

7g) Soggetto in posizione rematica rispetto a verbi passivi: valgono le stesse considerazioni fatte per le stringhe precedenti, a differenza delle quali si è imposto di ricercare le voci dei lemmi dei verbi *essere*, *venire*, *andare* (che sono i tre ausiliari possibili per la costruzione della forma passiva, cfr. Grandi, 2011), seguite opzionalmente dal participio dei soli verbi *essere* e *andare*³³ e poi da qualunque verbo al participio. In chiusura della stringa vi è come sempre un gruppo nominale in funzione di soggetto, per la cui descrizione si rinvia alle stringhe precedenti. Tutto ciò permette di reperire casi come: *fu istituito il Commissario distrettuale* (tempo non composto dell'ausiliare *essere*); *vennero svolte numerose prove* (tempo non composto dell'ausiliare *venire*); *va garantita la centralità* (tempo non composto dell'ausiliare *andare*); *sono state invitate le autorità* (tempo composto dell'ausiliare *essere*); *è andato distrutto il tempio* (tempo composto dell'ausiliare *andare*).

```
[word="\." ] [lemma="essere|venire|andare"]
[lemma="essere|andare" & feats=". *mod=p.*" ] ?
[postag="V.*" & feats=". *mod=p.*" ]
[postag="R.*|A.*|D.*" ] {0,2} [postag="P.*|S.*" &
deprel="subj" ]
```

7h) Soggetto in posizione rematica in frase subordinata (esempio: *non cambia nulla, se resta l'unilateralismo*): è stato imposto che la stringa inizi con una congiunzione subordinante con lettera non maiuscola;

³³ L'uso del verbo *venire* come ausiliare (esempio: *viene distrutto*) è limitato solo ai tempi semplici del verbo *venire* stesso (esempio: *viene/veniva/venga/venisse distrutto*) e non ai suoi tempi composti (è *venuto distrutto*) (Sabatini, 2004, pp. 8-9).

successivamente vi è un verbo che non sia al modo infinito, seguito da altri elementi linguistici (ossia, eventuali altre parole che compongono la frase) in quantità variabile da 1 a 2. A seguire possono essere presenti in quantità compresa tra 0 e 2 un articolo determinativo o indeterminativo e/o un aggettivo o, in alternativa all'articolo, un determinante e, infine, con il ruolo di soggetto può esservi un sostantivo di qualunque tipo o un pronomine di qualunque tipo:

```
[postag="CS" & word!=" [A-Z] .*" ] [postag="V.*" &
feats!=" . *mod=f.*" ] {1,2} [postag="R.*|A.*|D.*" ] {0,2}
[postag="P.*|S.*" & deprel="subj"]
```

8) impiego della frase scissa: il fenomeno della frase scissa può realizzarsi secondo due tipologie: 1) verbo *essere* + gruppo nominale (o altro costituente della frase) + *che* + verbo coniugato (esempio: *è lui che parla*); 2) verbo *essere* + gruppo nominale + *a* + verbo all'infinito (esempio: *è lui a parlare*). Sebbene in entrambe le versioni della lista dei tratti dell'italiano dell'uso medio Sabatini (1985, p. 163; 1990, p. 98) sia presente solo la tipologia della frase scissa costituita dal verbo *essere* + gruppo nominale (o altro costituente della frase) + *che* e, invece, manchi la tipologia basata su *a* + verbo all'infinito, nell'impostazione della ricerca su corpus si è scelto di tenere conto di entrambe le tipologie e dei vari costituenti della frase che possono essere inseriti nel focus della scissa, ossia: nel caso della costruzione della frase scissa con *che*, i costituenti possibili sono il gruppo nominale, il gruppo nominale preceduto da preposizione (sintagma preposizionale), l'avverbio e il verbo nella sua forma nominale (ossia all'infinito); nel caso della costruzione della frase scissa con *a* + infinito, invece, c'è il solo gruppo nominale. Pertanto sono state progettate le seguenti stringhe di ricerca:

8a) verbo *essere* + gruppo nominale + *che* + verbo coniugato: è stato imposto che la stringa inizi con un segno di punteggiatura e/o una congiunzione per evitare che il verbo *essere* sia preceduto da un altro gruppo nominale di cui costituirebbe la copula. Esempi: *Questa è la donna che ha telefonato*, ove il verbo *essere* funge da copula vs la frase scissa: *Poiché/Ma/Ed è la donna che ha telefonato* (e non qualcun altro). A seguire vi è un verbo che rientra nelle voci del lemma del verbo *essere*, dopo il quale possono essere presenti in quantità compresa tra

0 e 2 un articolo determinativo o indeterminativo e/o un aggettivo o, in alternativa all'articolo, un determinante; vi è poi qualunque tipo di pronome o sostantivo, seguito da *che*. Esempi: *è Anna che parla*; *è Firenze che ospita la mostra* (verbo *essere* + qualsiasi tipo di sostantivo, sia esso nome proprio o nome comune); *sono loro che parlano* (verbo *essere* + pronome personale); *è questol/quello che parla* (verbo *essere* + pronome dimostrativo); *è mio padre che parla* (verbo *essere* + aggettivo possessivo + sostantivo); *è un/l'amico che parla* (verbo *essere* + articolo determinativo/indeterminativo + sostantivo); *è ill/un mio amico che parla* (verbo *essere* + articolo determinativo/indeterminativo + aggettivo possessivo + sostantivo); *è questol/quel mio amico che parla* (verbo *essere* + aggettivo dimostrativo + aggettivo possessivo + sostantivo). Segue la stringa:

```
[postag="F.*|C.*"] [lemma="essere"]
[postag="R.*|A.*|D.*"] {0,2} [postag="P.*|S.*"]
[word="che"] [postag="V.*"]
```

8b) verbo *essere* + verbo all'infinito + *che* + verbo coniugato: è stato imposto che la stringa inizi con un segno di punteggiatura o una congiunzione; a seguire vi è un verbo che rientra nelle voci del lemma del verbo *essere*, dopo del quale vi è un verbo al modo infinito, seguito da *che* e da qualunque verbo coniugato. Esempio: *è decidere che è fondamentale*. Segue la stringa:

```
[postag="F.*|C.*"] [lemma="essere"] [postag="V.*"&
feats=".*mod=f.*"] [word="che"] [postag="V.*"]
```

8c) verbo *essere* + avverbio + *che* + verbo coniugato: valgono le stesse considerazioni fatte per la stringa precedente, a differenza della quale il costituente della frase che è stato dislocato è l'avverbio. Esempio: *è qui che dovete stare*. Segue la stringa:

```
[postag="F.*|C.*"] [lemma="essere"] [postag="B.*"]
[word="che"] [postag="V.*"]
```

8d) verbo *essere* + gruppo nominale preceduto da preposizione + *che* + verbo coniugato: in questo caso il verbo che rientra nelle voci del lemma del verbo *essere* è seguito da qualunque tipo di preposizione, dopo della quale possono essere presenti in quantità compresa tra 0 e 2 un aggettivo o un determinante; infine, vi è qualunque tipo di pronome o sostantivo seguito da *che* e da un verbo coniugato. Di seguito

verrà fornito un esempio per ciascuna delle casistiche reperibili tramite questa stringa:

1. verbo *essere* seguito da una preposizione, un sostantivo, *che* e un verbo: *è a casa che dobbiamo entrare*;
2. verbo *essere* seguito da una preposizione, un aggettivo, un sostantivo, *che* e un verbo: *è nella mia casa che dobbiamo entrare* (aggettivo possessivo); *è nella grande casa che dobbiamo entrare* (aggettivo qualificativo);
3. verbo *essere* seguito da una preposizione, un determinante, un sostantivo, *che* e un verbo: *è in questa casa che dobbiamo entrare*; *è in quella casa che dobbiamo entrare*;
4. verbo *essere* seguito da una preposizione, un determinante, un aggettivo, un sostantivo, *che* e un verbo: *è in questa grande casa che dobbiamo entrare*.

Segue la stringa:

```
[postag="F.*|C.*"] [lemma="essere"] [postag="E.*"]  
[postag="A.*|D.*"] {0,2} [postag="P.*|S.*"]  
[word="che"] [postag="V.*"]
```

8e) verbo *essere* + gruppo nominale + *a* + verbo al modo infinito:

è stato imposto che la stringa inizi con un segno di punteggiatura e/o una congiunzione coordinante; a seguire vi è un verbo che rientra nelle voci del lemma del verbo *essere*, dopo il quale possono essere presenti in quantità compresa tra 0 e 2 un articolo determinativo o indeterminativo e/o un aggettivo o, in alternativa all'articolo, un determinante; poi vi è qualunque tipo di pronome o sostantivo seguito dalla preposizione *a* e da un verbo al modo infinito. Esempi: *è Anna a parlare*; *è Firenze a ospitare la mostra* (verbo *essere* + qualsiasi tipo di sostantivo, sia esso nome proprio o nome comune); *sono loro a parlare* (verbo *essere* + pronome personale); *è questo/quel a parlare* (verbo *essere* + pronome dimostrativo); *è mio padre a parlare* (verbo *essere* + aggettivo possessivo + sostantivo); *è un/l'amico a parlare* (verbo *essere* + articolo determinativo/indeterminativo + sostantivo); *è il/un mio amico a parlare* (verbo *essere* + articolo determinativo/indeterminativo + aggettivo possessivo + sostantivo); *è questo/quel mio amico a parlare* (verbo *essere* + aggettivo dimostrativo + aggettivo possessivo + sostantivo). Segue la stringa:

```
[postag="F.*|C.*"] [lemma="essere"]
```

```
[postag="R.*|A.*|D.*"] {0,2} [postag="P.*|S.*"]
[word="a"] [postag="V.*" & feats=".*mod=f.*"]
```

9) impiego del verbo in forma pronominale (per indicare partecipazione affettiva) (esempi: *mi mangio la pizza, mi sono bevuto il caffè*). Per poter ricercare nel corpus questo fenomeno grammaticale è stato necessario suddividere la progettazione delle stringhe di ricerca in due macroaree, ossia l'impiego del verbo in forma pronominale con i tempi non composti (esempio: *mi bevo/ berrò/ bevi ecc. il caffè*) e composti (esempio: *mi sono bevuto/ ero bevuto/ mi sarei bevuto ecc. il caffè*) dei principali "verbi con uso intensivo, (o verbi di affetto)" (Ježek, 2011)³⁴, quali: bersi (*mi bevo una birra*), farsi (*mi faccio una mangiata*), guardarsi (*mi guardo un film*), leggersi (*mi leggo un romanzo*), mangiarsi (*mi mangio un panino*) e vedersi (*mi vedo la partita*). Pertanto, sono state progettate 12 stringhe di ricerca: 6 stringhe corrispondenti alle 6 forme pronominali riflessive (*mi, ti, si, ci, vi, si*) in presenza di tempi non composti dei verbi (esempi: *mi bevo/berrò/bevi (ecc.) il caffè*) e altre 6 stringhe per i casi dei tempi composti dei verbi (esempi: *mi sono bevuto/ero bevuto/ mi sarei bevuto (ecc.) il caffè*).

Di seguito si riportano con finalità di esempio solo due stringhe, rinviando all'Appendice 2 per una visione completa di tutte le stringhe progettate e utilizzate nella ricerca di tutte le forme pronominali riflessive.

Stringa per la prima persona singolare con i tempi non composti dei verbi:

```
[word="(?)mi"] [lemma="mangiare|leggere|bere|fare|guardare|vedere"& feats=".*num=s.*" & feats=".*per=1.*"]
[postag="R.*|A.*|D.*"] {0,2} [postag="P.*|S.*"
&deprel="obj"]
```

È stato imposto che la stringa inizi con il pronome personale (in questo caso *mi*) scritto con lettera maiuscola o minuscola; successivamente vi sono le voci di lemmi dei verbi *mangiare, leggere, bere, fare, guardare, vedere* coniugate alla prima persona singolare. A seguire è stata imposta la presenza del gruppo nominale: possono essere presenti in quantità compresa tra 0 e 2 un articolo determinativo o indeterminativo e/o un aggettivo o, in alternativa all'articolo, un determinante e, infine, con

³⁴ Versione html, pertanto priva di pagine.

il ruolo di oggetto (“obj”) può esservi un sostantivo di qualunque tipo o un pronome di qualunque tipo, quest’ultimo inserito per reperire anche casi come *mi mangio questo; mi bevo tutto*, eccetera.

Stringa per la prima persona singolare con i tempi composti dei verbi:

```
[word="(?)mi"] [lemma="essere"& feats=".*num=s.*"
& feats=".*per=1.*"] [lemma="mangiare|leggere|
bere|fare|guardare|vedere"& feats=".*mod=p.*"]
[postag="R.*|A.*|D.*"] {0,2} [postag="P.*|S.*"
&deprel="obj"]
```

Valgono le stesse considerazioni fatte per la stringa precedente, a differenza della quale si è imposto che il verbo sia composto dalle voci del lemma del verbo *essere* alla prima persona singolare, seguite dal participio passato dei verbi sopra nominati, e che il verbo abbia come oggetto il gruppo nominale precedentemente descritto. La scelta di imporre in entrambe le stringhe che dopo il verbo vi sia il gruppo nominale con valore di oggetto è nata dall’esigenze di restringere il campo della ricerca: in una precedente versione delle stringhe, che terminavano con il solo verbo (a seconda dei casi, tempo semplice o composto) la ricerca restituiva molte occorrenze che non rientravano nell’ambito dell’uso del verbo in forma pronominale per indicare partecipazione affettiva, come ad esempio: *Mi faccio (del) male / mi sono fatto (del) male, non mi faccio niente/ non mi sono fatto niente, mi sono fatto/visto + infinito (mi sono fatto accompagnare; mi sono visto rimproverare), mi faccio/vedo + infinito (mi faccio prendere; mi vedo arrivare), mi sono visto + participio (mi sono visto costretto), mi sono guardato da + articolo + infinito (mi sono guardato dal disfare), mi guardo da + articolo + infinito (mi guardo dal disfare), mi sono guardato (tutt’)attorno/intorno*, eccetera. Pertanto, allo scopo di ridurre il “rumore” tra i risultati, si è preferito restringere la ricerca.

10) impiego della concordanza a senso tra il soggetto collettivo singolare seguito da un’espressione partitiva e il predicato plurale (esempio: *un milione di elettori non hanno votato*): per questo fenomeno è stata progettata una stringa che prevede un sostantivo singolare a cui seguono la preposizione *di*, un sostantivo plurale e, infine, un verbo di cui si considerano solo le persone plurali del lemma e il cui modo sia diverso dal participio. L’esclusione del participio permette

di eliminare dalla ricerca tutte le occorrenze simili a *un centinaio di prodotti venduti*, che non rientrano nella casistica della concordanza a senso tra un soggetto singolare e un predicato plurale, poiché in casi di questo tipo il participio semplicemente funge da aggettivo. Ciò permette di ottenere casi come: *un centinaio di persone hanno partecipato*. Segue la stringa:

```
[postag="S.*" & feats=".*num=s.*"] [word="di"]
[postag="S.*" & feats=".*num=p.*"] [postag="V.*" &
feats=".*num=p.*" & feats!=".*mod=p.*"]
```

11) impiego del *ci* attualizzante con *avere* (non *c'ho tempo*; non *c'ho voglia*): è stato necessario elaborare 4 stringhe di ricerca al fine di includere tutte le possibili realizzazioni del fenomeno:

11a) “C’ + verbo *avere* + un sostantivo” per reperire casi come *c'ho tempo/voglia*. Segue la stringa:

```
[word="(?)c`"] [lemma="avere"] [postag="S.*"]
```

11b) “Ci + verbo *avere* + un sostantivo” per reperire casi come *ci ho tempo/voglia*. Segue la stringa:

```
[word="(?)ci"] [lemma="avere"] [postag="S.*"]
```

11c) “C’ + verbo *avere* + preposizione *da* + verbo” per reperire casi come *c'ho da fare*. Segue la stringa:

```
[word="(?)c`"] [lemma="avere"] [word="da"]
[postag="V.*"]
```

11d) “Ci + verbo *avere* + preposizione *da* + verbo” per reperire casi come *ci ho da fare*. Segue la stringa:

```
[word="(?)ci"] [lemma="avere"] [word="da"]
[postag="V.*"]
```

12) impiego del *che* polivalente: per poter restringere il campo di tutti gli usi di *che* e avvicinarsi il più possibile all'uso del *che* polivalente come forma generica dei pronomi relativi o delle congiunzioni subordinanti, sono state utilizzate le stringhe progettate da Budassi (2017), a cui sono state apportate alcune modifiche necessarie per uniformarle

al *tagset* in uso per la consultazione del corpus PAISÀ³⁵. Nella seguente descrizione si riutilizzerà la terminologia impiegata da Budassi (*ibidem*) nella definizione delle diverse tipologie di *che* polivalente:

12a) *Che* esplicativo/consecutivo; *che* causale: come espresso da Budassi (2017, p. 160), “i ‘che’ polivalenti di tipo esplicativo consecutivo (*Vieni che ti pettino*) e causale (*Vai a dormire che ne hai bisogno*) esprimono un legame tra due frasi”. La stringa, prodotta da Budassi nello studio sopra citato e adattata con poche modifiche al *tagger* del corpus PAISÀ, si basa sull'esclusione dei verbi che introducono una frase dichiarativa o una completiva:

```
[postag="V.*" & lemma!="ricordare|dimostrare|raccontare|parere|sembrare|sentire|dire|confidare|convenire|sperare|vedere|volere|sapere|scrivere|rispondere|scoprire|ribadire|promettere|mettere|insistere|dichiarare|osservare|sostenere|immaginare|temere|ritenere|premettere|augurare|pensare|pretendere|calcolare|lasciare|ripetere|scommettere|riconoscere|significare|vedere|prevedere|constatare|supporre|importare|capire|trovare|decidere|precisare|aggiungere|ammettere|notare|credere|chiedere|aspettare|considerare|bisognare|occorrere|avvenire|bastare|essere"] [word="che"] [postag="V.*"]
```

La stringa consente, indubbiamente, di restringere il campo d'indagine, ma si è rivelata poco produttiva, perché tra i risultati reperiti nel corpus si trovano soprattutto le frasi oggettive e le relative standard. Tra gli esempi reperiti figurano: *vieni che ti spalmo la crema* (esplicativo-consecutivo); *potevano fare tutte le guerre che volevano che a lui gli facevano un baffo* (causale).

12b) *Che* consecutivo-presentativo (con avverbi di dubbio): l'esempio inizialmente fornito per questa categoria da Budassi (2017), a sua volta tratto da Sornicola (1981), è il seguente: *Io sono una donna tranquilla che sto in casa, lavoro*. Casi come questo si collocano nel *continuum* esistente tra le due categorie di *che* polivalente: il *che* introduttore generico di frase subordinata, giustappositivo, e il *che* introduttore di frasi relative restrittive (Budassi, 2017, pp. 155-156). La progettazione

³⁵ Per le differenze tra le versioni originali e quelle impiegate in questa ricerca si rinvia a Budassi (2017, pp. 147-172).

della stringa di ricerca per la tipologia “consecutivo-presentativo” si è svolta attraverso varie fasi di raffinazione (cfr. *ivi*, pp. 162-164), che conducono alla seguente riflessione:

[...] la percentuale di polivalenza del ‘che’ rintracciabile con la [versione precedente della stringa] sale vertiginosamente se, nello slot immediatamente successivo a [tag=“CHE”], vengono ricercati alcuni tipi di avverbi, la maggior parte dei quali appartiene al gruppo degli avverbi di dubbio. (*ivi*, p. 163)

Budassi è, pertanto, approdato ad una stringa che ha come elemento iniziale un segno di punteggiatura, per delimitare la fine della frase precedente; segue la parola *che*, seguita da alcuni avverbi di dubbio, da eventuali altri elementi linguistici nel numero di 0 o 1 e, infine, da un verbo. La stringa è stata riutilizzata in questa ricerca, apportando delle modifiche necessarie per uniformarla al *tagset* in uso per la consultazione del corpus PAISÀ (esempio: [word=“che”] vs [tag=“CHE”] dell’originale) ed impostando che la parola *che* possa iniziare con lettera maiuscola o minuscola (ossia, (?i)) a seconda che sia a inizio frase o meno:

```
[word="\.|\;|\!|\?|\,|\" [word="(?)che\" [word="se|forse|chissà|eventualmente|probabilmente|poi|magari\" []
{0,1} [postag="V.*"]
```

Ciò ha consentito di reperire nel corpus PAISÀ casi come: *Io mi rendo conto di una cosa. Che se devo fare questa vita per altri dieci anni, una vita da trottola, girando tutto il giorno da un posto all’altro, rischio di sfinirme (sic).*

12c) *Che* pseudorelativo: l’esempio inizialmente fornito per questa categoria da Budassi (2017: 155), a sua volta tratto da Berruto (2012), è il seguente: *Li vedo che scendono.*

[Questa frase] è molto utile per individuare la struttura all’interno della quale potrebbe essere rintracciato il che polivalente pseudorelativo. Essa potrebbe essere la seguente: Pn (pronome) – V1– che – V2, dove V2 è

concordato con Pn. Se anche il 'che' sembra svolgere funzione di soggetto relativo di V2, il suo valore è primariamente giustappositivo (proprio in valore di ciò è detto pseudorelativo). (ivi, p. 164)

La stringa elaborata da Budassi (*ibidem*) prevede un pronome seguito dalle voci del lemma dei verbi *vedere* e *sentire*, seguite da *che* e da un verbo ed è stata presa come modello per l'elaborazione di una nuova stringa che include i principali verbi di percezione (Cimaglia, 2011) e che è strutturata come segue: un pronome seguito dalle voci del lemma dei verbi *vedere*, *sentire*, *guardare*, *scorgere*, *notare*, *osservare*, *mirare*, *ascoltare*, *udire*, *intendere*, *percepire*, seguite da *che* e da un verbo. Ossia:

```
[postag="P.*"] [lemma="vedere|sentire|guardare|scorgere|notare|osservare|mirare|ascoltare|udire|intendere|percepire"] [word="che"] [postag="V.*"]
```

Tuttavia, come già anticipato nella sezione 4.6.1, tra i risultati ottenuti dall'uso di questa stringa vi erano molte occorrenze del *si* impersonale (esempio: *si vede che hanno qualche domanda*), che chiaramente non rientrano nel fenomeno indagato. La struttura indicata da Budassi (*ibidem*), infatti, prevede la presenza di un pronome in funzione di oggetto, seguito dal verbo principale (V1), *che* e da un secondo verbo (V2), che concorda con il pronome. È il caso del già citato *Li vedo che scendono*, ove il pronome *Li* concorda con *scendono* (V2). Da tale descrizione risulta evidente che i casi di *si* impersonale (esempio: *si vede che hanno qualche domanda*) non rientrino all'interno di questa struttura sintattica, poiché il pronome *si* concorda con verbo principale (Esempio: *si vede che hanno qualche domanda*, il pronome *si* concorda con *vede*).

Pertanto, si è scelto di escludere il pronome *si* dai pronomi ricercabili. Ciò ha premesso di eliminare le occorrenze del *si* impersonale e di reperire occorrenze sul modello di *Li vedo che scendono*, trovando un compromesso tra la perdita di occorrenze corrette e l'esclusione di casi non inerenti, come già anticipato nella sezione 4.6.1, dedicata all'uso di CQL. Segue la stringa modificata:

```
[postag="P.*"&word!="si"] [lemma="vedere|sentire|guardare|scorgere|notare|osservare|mirare|ascoltare|udire|intendere|percepire"] [word="che"] [postag="V.*"]
```

13) impiego dell'indicativo al posto del congiuntivo in alcune subordinate (dopo i verbi di opinione, dopo una dichiarativa negativa (*non credo che hai torto*), in un'interrogativa indiretta (*gli chiesi se poteva aiutarmi*) e nelle ipotetiche (*se me lo dicevi, ci pensavo io*) (cfr. Mazzoleni, 1992). La progettazione delle stringhe è avvenuta sulla base delle seguenti casistiche:

13a) dopo i verbi di opinione/sentimento: per questo fenomeno è stata progettata una stringa che prevede le voci dei lemmi dei principali verbi che esprimono opinione o sentimento siano seguite da *che*, da eventuali altri elementi linguistici in quantità variabile da 0 a 5, purché diversi dal punto (`[word!="\."`]) e, infine, da un verbo al modo indicativo (esempi: *penso che è giusto*; *mi dispiace che questa cosa è successa*). L'esclusione del punto dagli eventuali elementi linguistici presenti dopo *che* è un espediente adottato per evitare che, come era accaduto in una prima fase di progettazione della stringa, la ricerca restituisca anche casi in cui il verbo all'indicativo si trovi dopo il punto, ossia al di fuori della frase retta da verbi di opinione/sentimento (esempio: *Penso che questo sia giusto. Insisto, ove penso* è il verbo di opinione + *che* + 4 elementi linguistici, tra cui un verbo al congiuntivo e il punto, che verrebbero percepiti dal software come possibili elementi da includere, e infine un verbo al modo indicativo, che regge un'altra frase.) Segue la stringa:

```
[lemma="pensare|credere|ritenere|sembrare|parere|supporre|dubitare|immaginare|temere|sperare|dispiacersi|stupirsi|augurarsi|aspettarsi|volere"] [word="che"] [word!="\." ] {0,5} [postag="V.*" & feats=".*mod=i.*"]
```

13b) dopo i verbi di opinione/sentimento preceduti da negazione: valgono le stesse considerazioni fatte per la stringa precedente, a differenza della quale l'elemento iniziale è la negazione *non* (esempio: *Non penso che è giusto*). Segue la stringa:

```
[word="non"] [lemma="pensare|credere|ritenere|sembrare|parere|supporre|dubitare|immaginare|temere|sperare|dispiacersi|stupirsi|augurarsi|aspettarsi|volere|sapere"] [word="che"] [word!="\." ] {0,5} [postag="V.*" & feats=".*mod=i.*"]
```

13c) dopo la costruzione “verbo *essere* + aggettivo + *che*”: questo fenomeno include sia i casi di costruzioni impersonali (esempio: *è necessario che*), sia casi di espressioni di sentimento (esempio: *sono contenta che*). Pertanto, è stata progettata una stringa in cui le voci del lemma del verbo *essere* sono seguite da un aggettivo, da *che*, da eventuali altri elementi linguistici in quantità variabile da 0 a 5, purché diversi dal punto ([word!="\."]) e, infine, seguite da un verbo al modo indicativo:

```
[lemma="essere"] [cpostag="A"] [word="che"] [word!="\."]
{0,5} [postag="V.*" & feats=".*mod=i.*"]
```

13d) dopo la costruzione “verbo *avere* + paura/timore + *che*”: valgono le stesse considerazioni fatte per la stringa precedente, a differenza della quale l'elemento iniziale è il lemma del verbo *avere*, seguito dalle parole *paura* o *timore* (esempio: *ho paura che non ce la fai*). Segue la stringa:

```
[lemma="avere"] [word="paura|timore"] [word="che"]
[word!="\."]{0,5} [postag="V.*" & feats=".*mod=i.*"]
```

13e) nell'interrogativa indiretta: per questo fenomeno è stata imposta una stringa in cui i lemmi dei verbi *chiedere* e *domandare* sono seguiti da *se*, da eventuali altri elementi linguistici in quantità variabile da 0 a 5, purché diversi dal punto ([word!="\."]) e, infine, seguiti da un verbo al modo indicativo:

```
[lemma="chiedere|domandare"] [word="se"] [word!="\."]
{0,5} [postag="V.*" & feats=".*mod=i.*"]
```

13f) nelle ipotetiche: per questo fenomeno, che esprime l'ipotesi irreali, è stata impostata una stringa in cui si richiede che la congiunzione subordinante *se*, scritta con la lettera maiuscola o minuscola, possa essere seguita da eventuali altri elementi linguistici in quantità variabile da 0 a 5, purché diversi da *ne* e *stesso/stessa/stessi/stesse* (in modo da evitare costruzioni come *se ne* e *se stesso/stessa/stessi/stesse*, che erano comunque emerse tra i risultati di una precedente versione della stringa). Seguono poi un verbo al modo indicativo e al tempo imperfetto (i), eventuali altri elementi linguistici in quantità variabile da 0 a 5, purché diversi dal punto, e infine un altro verbo con le stesse caratteristiche sopra nominate. La stringa permette di individuare esempi assimilabili a *se lo sapevo, te lo*

dicevo, in cui un'ipotesi non più realizzabile è espressa tramite l'indicativo imperfetto e non tramite la combinazione di congiuntivo trapassato e condizionale passato (*se l'avessi saputo, te l'avrei detto*):

```
[postag="CS" & word="( ?i) se"]
[word!="ne|stess(o|a|i|e)"] {0,5} [postag="V.*"
& feats=".*mod=i.*" & feats=".*ten=[i].*"]
[word!="\." ] {0,5} [postag="V.*" & feats=".*mod=i.*" &
feats=".*ten=[i].*"]
```

Le stringhe finora descritte sono state utilizzate per l'interrogazione del *corpus* PAISÀ, che verrà descritta nel capitolo 5. Di seguito verrà data una breve descrizione della metodologia che è stata adottata nel salvataggio delle occorrenze ottenute dall'interrogazione del *corpus* al fine di ottenere dati affidabili.

4.5. Metodo per il salvataggio e il trattamento dei risultati dell'interrogazione del corpus

Le stringhe di ricerca precedentemente descritte sono state utilizzate per interrogare il *corpus* PAISÀ. Allo scopo di ottenere dati di frequenza affidabili, nella fase di salvataggio dei risultati dell'interrogazione del corpus è stata svolta una quantificazione basata su un campione scelto in modo casuale e poi ripulito manualmente. Nello specifico, dopo aver ottenuto le occorrenze dei fenomeni ricercati, si è proceduto al salvataggio in formato .txt dei risultati e alla creazione, per ciascun fenomeno, di tre campioni da 200 occorrenze l'uno, per un totale di 600 occorrenze da analizzare per fenomeno³⁶. La scelta di fissare il limite a 600 risponde principalmente alla necessità di disporre di un numero gestibile di occorrenze da visualizzare, dato l'elevato numero di stringhe di ricerca elaborate, nonché di avere un totale di partenza che consentisse di dividere i risultati in tre campioni uguali³⁷. Al fine

³⁶ Allo scopo di mettere da parte un ampio campionario da cui eventualmente attingere ulteriori occorrenze, nel caso in cui le prime 600 non avessero fornito risultati sufficienti, nella fase di salvataggio dei dati dal *corpus* sono state scaricate in formato .txt le prime 2000 occorrenze del totale fornito dal *corpus*; da queste sono state estratte le prime 600, che sono state suddivise nei tre campioni da 200 a cui si fa riferimento nel testo.

³⁷ Nei casi in cui il totale delle occorrenze reperite è risultato inferiore a 600 e

di ottenere dei campioni casuali, ossia che non seguissero l'ordine delle occorrenze fornito dal *corpus*, prima del salvataggio dei risultati è stata selezionata la funzione *shuffle*, che consente di rimescolare in ordine casuale i risultati ottenuti, evitando in tal modo anche che si verificassero eventuali raggruppamenti dei doppioni.

Da ciascun campione sono stati eliminati i doppioni e i casi non corrispondenti al fenomeno oggetto, tra cui vi sono sia casi contenenti errori di *tagging*, sia casi in cui il *tagging* è corretto, ma è la stringa di ricerca a trovare anche risultati non inerenti al caso in esame. Per citare alcuni esempi, alcuni errori di *tagging* sono stati rilevati nel caso del soggetto in posizione rematica rispetto a verbi composti con ausiliare *avere*, con l'oggetto diretto tra participio e soggetto (esempio: *ha suscitato sdegno l'articolo che racconta...*, ove *l'articolo* è il soggetto e *sdegno* è l'oggetto diretto posto tra il participio e il soggetto): tra i risultati della ricerca sono presenti anche casi in cui è stato taggato erroneamente come soggetto qualunque elemento linguistico di carattere nominale successivo al verbo (esempio: *ha fatto i compiti la bambina*, ove *i compiti* è erroneamente taggato come soggetto). Invece, nel caso del verbo in forma pronominale (esempio: *mi mangio il panino*) alla terza persona singolare e plurale sono stati reperiti numerosissimi casi di strutture impersonali (esempi: *si mangia bene qui*; *si mangiano i calamari*), che chiaramente non rientrano nel caso in questione.

Dopo aver dato una breve descrizione della metodologia adottata nel salvataggio delle occorrenze ottenute dall'interrogazione del *corpus*, nella sezione 4.6 verrà illustrato il metodo adottato per la progettazione e sperimentazione delle unità didattiche.

4.6. Metodo per la progettazione e sperimentazione delle unità didattiche

Richiamando gli obiettivi principali di tale progetto di ricerca, enunciati in apertura del capitolo, la selezione del corpus PAISÀ tra le varie

maggiore di 60, il totale è stato comunque diviso in tre campioni identici o che presentassero uno scarto di 1-2 unità di differenza; nei casi in cui il totale delle occorrenze ottenute è risultato invece inferiore a 60, i risultati sono stati trattati come un campione unico.

risorse attualmente disponibili è stata svolta con l'obiettivo di reperire dei testi contenenti tali fenomeni grammaticali, con i quali progettare delle unità didattiche per l'apprendimento dell'italiano L2.

A tale scopo, le stringhe di ricerca precedentemente illustrate sono state utilizzate per ricercare nel corpus i fenomeni grammaticali oggetto di studio, scaricare dei campioni dei risultati ottenuti ed esaminarli con il duplice obiettivo sia di ottenere degli esempi dell'uso reale di tali strutture in testi autentici, sia di calcolare la ricorrenza dei fenomeni ricercati. Il procedimento di interrogazione del corpus, esame dei risultati e calcolo della frequenza sarà dettagliatamente descritto nel capitolo 5; nelle sezioni seguenti, invece, si illustreranno le attività finalizzate alla produzione delle unità didattiche, ossia: la selezione dei testi, la progettazione delle unità didattiche e la selezione di una piattaforma *e-learning* sulla quale collocare le suddette unità didattiche.

4.6.1. Selezione dei testi

Dopo aver individuato quali tra i fenomeni grammaticali elencati da Sabatini (1990) siano i più frequenti nel corpus (cfr. capitolo 5), ne sono stati selezionati due sui quali incentrare le unità didattiche, al fine di avere un'unità incentrata su un fenomeno linguistico che risponde sia al criterio della funzionalità che dell'accettabilità sociale e un fenomeno su cui incide il criterio dell'accettabilità sociale, ossia: l'uso del verbo in forma pronominale riflessiva in senso affettivo e l'uso dell'indicativo al posto del congiuntivo (sulle motivazioni della scelta di tali fenomeni specifici rispetto a tutti quelli indagati, si rinvia ai capitoli 5 e 6).

Al fine di selezionare i testi da utilizzare nella progettazione delle unità didattiche, innanzitutto è stata svolta una rilettura delle occorrenze scaricate dal corpus PAISÀ per i due fenomeni grammaticali. Poiché il corpus permette di visualizzare il contesto testuale degli elementi linguistici ricercati, per ciascuna occorrenza del fenomeno oggetto di ricerca è stata attivata la visualizzazione di alcune righe di testo precedenti e successive alla posizione nella frase dell'elemento ricercato.

Tale procedura ha permesso di identificare in ciascun campione di occorrenze i casi ritenuti più interessanti sia dal punto di vista grammaticale, sia dal punto di vista tematico. Dal punto di vista grammaticale, il loro utilizzo nell'unità didattica avrebbe permesso di approfondire

aspetti particolari dell'uso di un determinato fenomeno linguistico; per quanto riguarda, invece, il punto di vista tematico, nella selezione dei testi si è tenuto conto anche del loro contenuto, avendo cura di selezionare temi generalmente trattati nella manualistica per stranieri rispetto ai livelli intermedi B1-B2 (cfr. capitolo 6) e che, pertanto, potessero risultare familiari agli apprendenti e, soprattutto, adeguati alle loro conoscenze e competenze linguistiche. Per quanto riguarda questi due aspetti, la riflessione è stata guidata dalla consultazione dei seguenti testi: 1) il QCER (Consiglio d'Europa 2018, 2002); 2) i sillabi della lingua italiana attualmente disponibili (cfr. capitolo 6); 3) i contenuti tematici trattati nei 32 manuali di italiano L2/LS citati nella relazione a convegno di Słapek (2016) e nel suo studio del 2017 sulla trattazione della flessione verbale nella suddetta manualistica e nel successivo approfondimento della ricerca da me condotto sui manuali pubblicati dal 2018 in poi (cfr. capitolo 5).

La consultazione delle risorse citate ai punti 1 e 2 ha consentito di avere un quadro completo degli argomenti grammaticali da inserire nelle unità didattiche (1) e della loro collocazione all'interno dei vari livelli previsti al QCER (2): ad esempio alcuni argomenti grammaticali sono caratterizzanti di un determinato livello e possono essere ripresi come approfondimento nei livelli successivi; la manualistica attualmente in commercio ha permesso di determinare quali tematiche vengano generalmente trattate nei livelli intermedi.

Una volta identificati i casi ritenuti più interessanti dal punto di vista grammaticale e tematico, per ciascuno di essi sono state visitate le rispettive pagine web originarie, così da poter leggere i testi integralmente e in modo più agevole. Ciò è stato reso possibile dal fatto che per ciascuno dei testi che compongono il corpus PAISÀ sono indicate tra le informazioni generali anche le pagine web da cui i testi sono stati scaricati.

Infine, una volta avuta una panoramica del contesto più ampio e/o dell'intero testo in cui era utilizzato un dato fenomeno grammaticale, per ciascun caso, in base alle specificità grammaticali o tematiche del testo in questione, si è valutato se per la progettazione delle unità fosse più opportuno utilizzare i testi in forma integrale o parziale (ad esempio, sotto forma di brani da leggere), o se utilizzare solo le singole frasi contenenti il fenomeno in questione (ad esempio, per creare degli

esercizi di completamento o delle liste, ecc.). Quest'ultima opzione è stata scelta nel caso di testi risultati inadatti al livello linguistico dell'unità (ad esempio, dal punto di vista del registro linguistico, del tema trattato, dei riferimenti culturali presenti ecc.).

4.6.2. Valutazione di alcune piattaforme di *e-learning*

Parallelamente all'elaborazione delle unità didattiche si è svolta la ricerca di una piattaforma dedicata all'apprendimento in cui collocare i materiali progettati in formato elettronico. L'*e-learning* è un insieme di metodologie e strategie didattiche che vengono impiegate per la progettazione di un ambiente di apprendimento basato sull'uso del web e della multimedialità. Una sua possibile definizione è quella elaborata dall'osservatorio Anee (Associazione nazionale dell'editoria elettronica, dal 2001 Associazione dei Servizi e Contenuti Multimediali):

L'e-learning è una metodologia di insegnamento e apprendimento che coinvolge sia il prodotto sia il processo formativo. Per prodotto formativo si intende ogni tipologia di materiale o contenuto messo a disposizione in formato digitale attraverso supporti informatici o di rete. Per processo formativo si intende invece la gestione dell'intero iter didattico che coinvolge gli aspetti di erogazione, fruizione, interazione, valutazione. In questa dimensione il vero valore aggiunto dell'e-learning emerge nei servizi di assistenza e tutorship, nelle modalità di interazione sincrona e asincrona, di condivisione e collaborazione a livello di Community. Peculiarità dell'e-learning è l'alta flessibilità garantita al discente dalla reperibilità sempre e ovunque dei contenuti formativi, che gli permette l'autogestione e l'autodeterminazione del proprio apprendimento. (Liscia, 2004, p. 12)

Si tratta, dunque, di una metodologia che propone un'innovazione nella concezione e progettazione dei contenuti formativi, nella loro organizzazione e fruizione da parte degli utenti. Tra le risorse utilizzate per l'*e-learning* vi sono le piattaforme specificamente progettate per gestire attività didattiche online, che si basano su due sistemi: LMS (*Learning*

Management System) e LCMS (*Learning Content Management System*), ove il primo gestisce la distribuzione dei corsi online, l'iscrizione degli apprendenti e il tracciamento delle loro attività online, mentre il secondo è dedicato alla gestione dei contenuti. I due sistemi spesso cooperano per la realizzazione di un ambiente di apprendimento online.

I criteri che hanno guidato la selezione della risorsa *e-learning* da utilizzare per la presente ricerca sono:

- a. la struttura della piattaforma (struttura e grafica dell'interfaccia, tipologie di attività didattiche progettabili, la possibilità di caricare/scaricare contenuti multimediali, quali audio/video, immagini, creare collegamenti ad altri siti web ecc.);
- b. l'accessibilità (sia dal punto di vista tecnico, ossia la possibilità di accedere efficacemente al sito web sia da tecnologia fissa che da dispositivi mobili, quali *smartphone* e *tablet*, sia dal punto di vista economico, ossia se si tratta di risorse *open source* o a pagamento);
- c. il tipo di formazione erogata: in autoapprendimento (basata sull'offerta di contenuti e di un supporto minimale da parte di un tutor), assistita (che prevede sia momenti di studio individuale, sia momenti di interazione con un tutor) o collaborativa (impostata sull'interdipendenza tra i membri del gruppo-classe che si realizza attraverso attività basate sulla condivisione dei contenuti e collaborazione in cui il docente ricopre il ruolo di moderatore della comunità di apprendimento);
- d. la possibilità di realizzare le attività didattiche direttamente sulla piattaforma (invece di caricarne i singoli file) e il grado di personalizzazione con cui tale progettazione può avvenire.

Innanzitutto è stata svolta un'indagine finalizzata alla raccolta di informazioni sulle piattaforme di *e-learning* attualmente disponibili. Sulla base delle classiche fornite da *Capterra*³⁸, *Academy of mine*³⁹ e *Wikipedia*⁴⁰, risulta che le piattaforme più diffuse tra gli enti di formazione

³⁸ Capterra, LMS Software www.capterra.com/learning-management-system-software/#top-20.

³⁹ Academy Of Mine, *Best Free Learning Management Systems (LMS)*. www.academyofmine.com/12-best-free-learning-management-systems-lms

⁴⁰ Wikipedia, *Learning Management System* https://it.wikipedia.org/wiki/Learning_management_system.

al momento della scrittura di questo capitolo siano: *A-tutor*, *Blackboard*, *Chamilo*, *Claroline*, *Docebo*, *Dokeos*, *E-Front*, *Ilias*, *Moodle*, *Sakai* e *Schoology*⁴¹. Ciascuna risorsa è stata valutata sulla base dei criteri sopra menzionati, che hanno portato all'immediata esclusione di alcune di esse. Per quanto riguarda, ad esempio, l'accessibilità open source vs pagamento, *Blackboard* e *Docebo*, entrambe ampiamente diffuse e utilizzabili sia da tecnologia fissa che mobile, presentano costi che si aggirano intorno alle migliaia di euro l'anno; mentre invece *Dokeos* ed *E-Front* presentano versioni base con funzionalità molto ridotte rispetto alle versioni a pagamento.

Inoltre, alcune di esse sono concepite come una "libreria", ossia non consentono la progettazione di un'intera unità didattica online, bensì si basano principalmente sul caricamento da parte del docente di file contenenti le attività, i quali devono essere poi scaricati dagli studenti e ricaricati per la consegna al docente, con un evidente aggravio di tempo e perdita di immediatezza nell'uso della piattaforma. È il caso, ad esempio, della piattaforma *Moodle*, sebbene essa consenta entro centri limiti di svolgere attività direttamente online (lettura/produzione di testi in formato blog o wiki, utilizzo di file multimediali ecc.) e della piattaforma *Ilias*, un LMS aperto che offre un ambiente per l'organizzazione e strutturazione in corsi e classi dei contenuti liberamente condivisi con chiunque e non con i soli utenti del sistema.

Al di là di tali specificità, ciò che è emerso in generale dalla documentazione disponibile online sulle risorse sopra nominate è che esse siano state concepite principalmente per la creazione e gestione di interi corsi online, piuttosto che per la creazione di un set di attività didattiche raggruppabili all'interno di un'unità didattica. Poiché, come emergerà nel capitolo 6, per le finalità di questa ricerca è stata valutata la produzione di due sole unità didattiche focalizzate su due fenomeni grammaticali, l'utilizzo di piattaforme che presentano una certa complessità

⁴¹ A-tutor (<https://atutor.ca>), Blackboard (www.blackboard.com/it-it), Chamilo (<https://chamilo.org/it>), Claroline (<https://www.claroline.com/>), Docebo (<https://www.docebo.com/>), Dokeos (<https://www.dokeos.com/>), E-Front (<https://www.efrontlearning.com/>), Ilias (<https://www.ilias.de/>), Moodle (<http://moodle.org>), Sakai (<https://www.sakailms.org/>) e Schoology (<https://www.powerschool.com/classroom/schoology-learning/>).

nella gestione dei contenuti sarebbe risultato sproporzionato rispetto all'obiettivo in questione.

Di conseguenza, l'attenzione è stata rivolta verso strumenti che consentano la produzione di fogli elettronici interattivi. Consultando il sito *E-Learning industry*⁴² sono state valutate varie risorse di questo tipo sulla base dei criteri sopra citati e, infine, è stata selezionata la risorsa *Wizer.me*⁴³.

Definito come “in assoluto uno dei migliori strumenti per creare diverse tipologie di quiz ed esercizi interattivi online”⁴⁴, il fattore principale che ne ha determinato la selezione è che viene dato ampio spazio alla personalizzazione da parte del docente, sia dal punto di vista didattico, sia dal punto di vista grafico, nella progettazione delle attività all'interno della risorsa sotto forma di fogli di lavoro interattivi.

Nello specifico, si tratta di una piattaforma con registrazione gratuita che offre una vasta gamma di modelli di attività (ad esempio: inserimento di brani da leggere, creazione di domande aperte o chiuse, tabelle, esercizi di riempimento, esercizi di collegamento immagine-parola o collegamento tra parole, oltre alla possibilità di inserire collegamenti ipertestuali, come link a video o pagine web ecc.). Le consegne di ciascuna attività possono, poi, essere impostate non solo in forma scritta, ma anche sotto forma di file audio, favorendo lo svolgimento delle attività anche da parte di apprendenti con disturbi specifici di apprendimento (DSA) e, in tal modo, l'inclusività all'interno del gruppo-classe. Inoltre, offre numerose opzioni per sfondi, caratteri del testo, colori, nonché la possibilità di stabilire manualmente il punteggio da attribuire a ciascun quesito e la possibilità di attivare la funzione di correzione automatica delle attività, previo inserimento da parte del docente delle soluzioni dei diversi quesiti in fase di progettazione. Quest'ultimo aspetto ha costituito un ulteriore punto a favore, poiché permette sia

⁴² “It is the largest online community of eLearning professionals in the industry, and was created first and foremost as a knowledge-sharing platform to help eLearning professionals and instructional designers connect in a safe online community where they can stay up to date with the latest industry news and technologies, and find projects or jobs”. (<https://elearningindustry.com/>).

⁴³ <https://app.wizer.me/>.

⁴⁴ www.robartosconocchini.it/: sito che raccoglie informazioni sulle risorse didattiche digitali, nato come iniziativa informale di un ricercatore non affiliato.

di svolgere le attività in autoapprendimento, sia di abbreviare i tempi di correzione in aula. I docenti hanno il report dettagliato di tutte le prove degli studenti cliccando su *Answers*.

La piattaforma è accessibile sia tramite desktop, sia tramite cellulare. L'accesso da parte degli apprendenti avviene tramite la condivisione da parte del docente del link del foglio di lavoro creato: una volta copiato il link nel browser, si apre una maschera per il login con profilo studente e, dopo aver effettuato la registrazione, il foglio elettronico con le attività si apre automaticamente.

Oltre che per la progettazione dei fogli elettronici, la piattaforma può essere utilizzata anche per reperire materiali progettati da altri utenti, riutilizzarli ed eventualmente personalizzarli in base a esigenze diverse. L'accesso ai materiali da parte dell'intera community degli utenti di *Wizer.me* avviene di default. In alternativa, è possibile sottoscrivere un abbonamento annuale che consente la creazione di un account riservato: in quest'ultimo caso è l'utente che decide se condividere o meno i materiali all'interno della community. Nel caso della presente ricerca, si è optato per la sottoscrizione dell'abbonamento, al fine di poter preservare la riservatezza dei materiali didattici sperimentali.

4.6.3. Progettazione delle unità didattiche

La glottodidattica necessita di modelli operativi che permettano di realizzare le sue componenti teoriche (l'approccio) e pratiche (il metodo):

[l']approccio costituisce la filosofia di fondo di ogni proposta glottodidattica. L'approccio valuta e seleziona dati e impianti epistemologici dalle varie teorie e dalle varie scienze di riferimento, e li riorganizza secondo i parametri propri della glottodidattica, individuando le mete e gli obiettivi dell'insegnamento linguistico. Un approccio genera uno o più metodi che ne realizzano l'applicazione nelle varie situazioni. Nella storia della glottodidattica alcuni approcci sono stati definiti "metodi" (Comunicativo, Formalistico, Metodo diretto, Metodo naturale, Reading method, Silent Way, Strutturalistico, Suggestopedia). (Balboni, 1999, p. 5)

Tra i vari modelli operativi esistenti (cfr. Diadori, Semplici, Troncarelli, 2020) in questa trattazione si farà riferimento all'unità didattica, facendo successivamente una breve distinzione tra due delle varie tipologie di unità attualmente esistenti.

Nella seconda metà del Novecento viene gradualmente superato il modello operativo della lezione frontale nell'insegnamento delle lingue, con l'emergere di nuove teorie sull'apprendimento delle lingue. Dagli anni '70 prende il via una revisione dei modelli di insegnamento delle lingue moderne, ispirata in particolare alla psicologia della *Gestalt* (Koffka, 1935; Köhler, 1929), una corrente psicologica che descrive la percezione umana secondo la sequenza "globalità → analisi → sintesi" che va dal generale al particolare e che viene utilizzata anche per avviare una nuova concezione della didattica.

Poiché la percezione umana segue questa sequenza, emerge in quegli anni l'idea che anche il contatto da parte degli apprendenti con i contenuti di una disciplina di studio possa essere proposto in maniera più efficace utilizzando un percorso che dalla globalità passi all'analisi e infine alla sintesi.

Di fondamentale importanza in questo contesto fu l'elaborazione da parte di Krashen della *Second Language Acquisition Theory* (1982), basata su sei ipotesi inerenti ai processi attraverso i quali si impara una lingua. Di seguito si farà riferimento brevemente solo alle ipotesi strettamente legate alla modalità di presentazione dei contenuti linguistici agli apprendenti.

Innanzitutto Krashen identifica due processi fondamentali attraverso i quali si impara una lingua: l'apprendimento e l'acquisizione. L'apprendimento è un processo consapevole e razionale, generalmente tipico del percorso scolastico, che può portare a competenze linguistiche pressoché perfette, ma non durature, poiché si basa sulla memoria a breve termine. L'acquisizione, invece, è un processo inconscio, che porta a interiorizzare la lingua in maniera duratura e alla graduale capacità di comprendere e produrre frasi riutilizzando le nozioni assorbite.

Si basa su un percorso graduale che presenta molte tappe intermedie, che costituiscono l'interlingua (Selinker, 1972), ossia un sistema linguistico provvisorio, in continua evoluzione, relativo alla lingua oggetto di studio, che l'apprendente ricostruisce gradualmente, man mano che nuove regole e strutture si aggiungono alla sua competenza.

Il processo inconscio dell'acquisizione, che sfrutta l'intuizione e la memorizzazione naturale, è paragonabile alla maniera in cui il bambino impara la propria lingua madre (ipotesi di distinzione tra acquisizione e apprendimento). Da ciò si deduce che l'insegnante tramite la didattica debba puntare a favorire l'acquisizione duratura da parte degli apprendenti e il raggiungimento di tale obiettivo chiaramente è influenzato dalla modalità di presentazione dei contenuti.

Al fine di realizzare l'acquisizione linguistica, è necessario che le regole di una lingua siano acquisite in modo spontaneo, seguendo il cosiddetto "ordine naturale di acquisizione", che rappresenta l'ordine sequenziale con cui il cervello processa i contenuti linguistici, procedendo dalle strutture più semplici alle strutture più complesse (ipotesi dell'ordine naturale).

Krashen (1982, 1985) sintetizza l'ordine naturale attraverso la formula " $i+1$ ", che descrive l'input linguistico presentato con gradualità: rispetto all'input linguistico già noto (i), il nuovo input fornito deve avere un grado di difficoltà leggermente maggiore ($+1$). Solo in questo modo l'input linguistico risulta comprensibile, poiché viene presentato in un momento in cui l'apprendente è in grado di coglierne il significato e l'utilità (ipotesi dell'input comprensibile).

Infine l'input verrà successivamente immagazzinato nella memoria a lungo termine solo se è stato fornito in una situazione priva di ansia e paura di essere giudicato, in cui le difese psicologiche (il cosiddetto "filtro affettivo") siano abbassate (ipotesi del filtro affettivo).

L'ordine naturale è studiato e descritto dalla linguistica acquisizionale, una disciplina che si è inizialmente sviluppata nell'area anglosassone e successivamente nei vari paesi europei che hanno vissuto il fenomeno dall'immigrazione, tra cui la Germania, la Francia e in seguito anche l'Italia, come un nuovo settore di ricerca focalizzato sull'apprendente e sullo sviluppo della sua competenza linguistico-comunicativa, soprattutto nelle L2.

Studiare il funzionamento della facoltà di linguaggio e alcuni suoi principi coinvolti nel processo di costruzione della competenza linguistica ha dei risvolti nella didattica di L2, poiché consente nella progettazione di materiali didattici e nella prassi didattica di predisporre delle strategie che facilitino l'acquisizione linguistica. La prospettiva assunta in questa ricerca coinvolge sia la linguistica acquisizionale, sia

la glottodidattica, ricadendo all'interno della cosiddetta didattica acquisizionale (Vedovelli, 2003; Vedovelli, Villarini, 2003).

Dando brevemente delle definizioni di tali ambiti, si può affermare che la linguistica acquisizionale “studia processi e modelli relativi all'acquisizione di una lingua non materna” (Chini, 2005, p. 9), focalizzando la sua attenzione sull'apprendente e sul contesto comunicativo che gli fornisce gli input linguistici, mentre la glottodidattica è la scienza dell'educazione linguistica (Balboni, 1999) che “studia l'apprendimento e l'apprendente dalla prospettiva dell'insegnamento” (Vedovelli, 2003, p. 179). Dall'interesse per le ricadute didattiche degli esiti della ricerca sulla L2 ha avuto origine la didattica acquisizionale, “un modello teorico di didattica linguistica che si rapporti alle prospettive acquisizionali, rispetti le fasi di sviluppo della L2, non violi le sequenze implicazionali” (Vedovelli, 2003, p. 178; cfr. anche Rastelli, 2009, 2010)⁴⁵.

A questo proposito, è alla linguistica acquisizionale che si deve l'identificazione delle sopra menzionate sequenze implicazionali, note anche come sequenze di acquisizione o acquisizionali, ossia le tappe che costituiscono l'ordine naturale e che risultano comuni a tutti gli apprendenti di una determinata lingua, qualsiasi sia la loro lingua d'origine e indicano quali contenuti grammaticali siano propedeutici all'acquisizione di altri:

La linguistica acquisizionale fornisce delle indicazioni utili tramite appunto le sequenze acquisizionali, riscontrate per diversi ambiti specie morfosintattici dell'ita-

⁴⁵ Si vedano a questo proposito i seguenti riferimenti per la didattica acquisizionale: per la lingua italiana si segnalano il volume sull'apprendimento e la didattica dell'italiano a studenti universitari sinofoni (Rastelli, 2010, a c. di), Banfi (2003), Chini (2011, 2012a, 2012b, 2012c, 2016), lo studio di Della Putta, Visigalli (2010) sulla didattica del sintagma nominale italiano ad anglofoni, gli atti del convegno CIS del 2006 (Grassi *et al.*, 2008), gli atti dei convegni e le pubblicazioni del Centro di Italiano per Stranieri CIS dell'Università di Bergamo, dell'Associazione Italiana di Linguistica Applicata (AIItLA) e dalla rivista on-line Italiano LinguaDue. Sul piano internazionale si segnalano Ellis (1997), Doughty (2003), Hinkel (2005), Housen, Pierrard (2005), Véronique (2000, 2005) e Trévisiol-Okamura, Komur-Thilloy (2011).

liano L2 (come di altre L2). Esse attestano che soggetti con L1 diversa, studiati longitudinalmente per un periodo più o meno lungo, acquisiscono nello stesso ordine (stadio 1 > stadio 2 > stadio n) alcune regole, nozioni e sottosistemi di L2 (ad esempio, strutture negative, interrogative, morfologia verbale), pur percorrendo la sequenza con velocità diverse, in dipendenza da vari fattori (la maggiore o minore marcatezza della struttura; la maggiore o minore distanza genetica e strutturale dalla L1 dell'apprendente; fattori extralinguistici come età, motivazione, esposizione all'input, attitudine alle lingue, ecc.).

Un aspetto di tali stadi, interessante sia in chiave teorica che didattica, è la loro valenza implicazionale: il raggiungimento dello stadio 3 implica l'acquisizione delle strutture degli stadi 1 e 2, il che ha un chiaro risvolto diagnostico e prognostico, fornendo un punto di orientamento anche per la programmazione didattica, in considerazione della cosiddetta "insegnabilità". (Chini, 2016, p. 5)

Per quanto riguarda la lingua italiana, le sequenze acquisizionali finora scoperte riguardano la morfologia verbale e nominale, i pronomi clitici, la negazione, la connessione interproposizionale, la subordinazione, la modalità, gli ordini marcati e le strategie anaforiche (si vedano i capp. 3-6 di Giacalone Ramat, 2003 e una sintesi in Chini, 2005, pp. 94-105). Sono state identificate a partire dagli anni '80 nell'ambito del "Progetto di Pavia"⁴⁶, il più longevo ed esteso progetto di ricerca sull'italiano L2, che ha studiato e descritto le fasi dell'acquisizione dell'ita-

⁴⁶ Nato a metà degli anni '80, il progetto è stato coordinato dapprima da Anna Giacalone Ramat e dal 2003 da Giuliano Bernini. Attivo in una prima fase a Pavia, poi all'Università di Bergamo, ha coinvolto vari atenei (Pavia, Bergamo, Milano Bicocca, Udine, Vercelli, Verona, Siena Stranieri, Torino, poi Verona e Modena-Reggio), con l'obiettivo principale di individuare, attraverso lo studio delle produzioni – principalmente orali – degli apprendenti i principi che li guidano nella scoperta del rapporto fra forma (linguistica) e funzione nella lingua di arrivo.

liano L2⁴⁷. Si cita a titolo di esempio la sequenza dell'acquisizione di tempi e modi verbali dell'italiano, ove il segno > indica che ciò che si trova alla sua sinistra è appreso prima di ciò che si trova alla sua destra: *presente (e infinito) > (ausiliare) participio passato > imperfetto > futuro > condizionale > congiuntivo* (Banfi, Bernini, 2003, p. 90).

Le riflessioni che sono derivate in ambito didattico dalla scoperta delle sequenze acquisizionali sono le seguenti: è necessario insegnare le strutture linguistiche nell'ordine in cui si manifestano naturalmente nell'interlingua degli apprendenti, senza aspettarsi che le strutture che si trovano nelle tappe più avanzate della sequenza vengano usate correttamente prima che siano state apprese le strutture precedenti. È ciò a cui fa riferimento la *teachability hypothesis* (ipotesi della insegnabilità) formulata da Pienemann (1986), secondo la quale ha senso insegnare solo ciò che un apprendente può imparare in un certo momento del suo percorso acquisizionale, poiché "l'insegnamento può favorire l'acquisizione linguistica solo se l'interlingua è vicina al punto in cui la struttura da insegnare viene acquisita nella situazione naturale" (Pienemann, 1986, p. 313). Successivamente Pienemann (1998, 2005) fornisce un inquadramento teorico per tale ipotesi tramite la Teoria della Processabilità, con la quale spiega che a ogni fase dell'acquisizione l'apprendente può produrre e comprendere solo le strutture linguistiche che il suo "processore" linguistico (una sorta di processore mentale della lingua) riesce a elaborare in quel momento⁴⁸.

⁴⁷ Si citano a questo proposito i seguenti ambiti di ricerca e alcuni studi a titolo di esempio. Per l'italiano L2: le principali differenze tra le sequenze di acquisizione dell'italiano L1 e quelle dell'italiano L2 (Calleri *et al.*, 2003, pp. 221-228), la morfologia del nome L2 (Berretta, 1987; Lo Duca, 1990), il sistema del tempo, modo e aspetto verbale (Banfi, Bernini, 2003), la sintassi della frase: (Calleri *et al.*, 2003, pp. 236-245) e la sintassi del periodo: (Ferraris, 1999, pp. 53-62). Per l'italiano L1: la morfologia del nome: (Chini, 1995, pp. 129-151), il sistema di tempo, modo e aspetto verbale (Banfi, Bernini, 2003; Lo Duca, Solarino, 2006), la sintassi della frase: (Andorno, 2003, pp. 127-140, Andorno *et al.*, 2003, pp. 116-127), la sintassi del periodo: (Ferraris, 1999; Andorno 2003); (Ferraris, 1999, pp. 34-42).

⁴⁸ Tra i richiami teorici di tale teoria vi è il modello psicolinguistico di Levelt (1989) e la Grammatica Lessico-Funzionale di Bresnan (2001) e ha trovato conferma in alcune sperimentazioni didattiche sull'italiano L2 (Bettoni, Di Biase, 2008; Whittle, 2015; Di Biase, Bettoni, 2015).

Gli studi nell'ambito della linguistica acquisizionale andarono a innestarsi nel filone di ricerca in ambito glottodidattico, che, dagli anni '70 del Novecento in poi portò al modello di unità didattica elaborato da Freddi (1970, 1994, 1999) e Titone (1976), poi ripreso e approfondito da Danesi (1988, 1998) da un punto di vista neurolinguistico, mettendo in luce l'importanza di strutturare le attività sulla base dell'ordine in cui si attivano i due emisferi cerebrali (Danesi, 1998, p. 152 e sgg.) e, infine, da Balboni (1994, 2002, 2008), con un'evoluzione nella sua organizzazione interna. Nel modello di Freddi e Danesi alle tre fasi originarie della psicologia della *Gestalt* se ne aggiungono altre tre, andando a costituire il seguente percorso di fasi di apprendimento, collegabili a processi psicodidattici, che corrispondono a: motivazione → globalità → analisi → sintesi → riflessione → controllo/verifica. Tale modello gestaltico è stato applicato in questa ricerca nella progettazione delle unità didattiche.

Pertanto, in entrambe le unità didattiche progettate, le attività si svolgono seguendo la sequenza di fasi sopra descritta: *Contatto con il tema*, con attività di stimolo della motivazione e mirate a esprimere le conoscenze pregresse sul tema che verrà trattato; *Comprensione*, ossia la fase della globalità, in cui viene chiesto agli studenti di leggere uno o più testi e rispondere a delle domande di comprensione testuale; nella fase di *Analisi* si chiede di identificare nell/i testo/i alcuni elementi lessicali e sintattici; *Applicazione pratica*, ossia la fase della sintesi, in cui si chiede agli studenti di sistematizzare le strutture sintattiche e lessicali precedentemente individuate e reimpiegarle; *Riflessione*, in cui si chiede di elaborare o completare una definizione del loro uso. Infine, vengono proposte alcune attività di verifica dell'acquisizione dei contenuti grammaticali e pragmatici al centro dell'unità didattica. Nel modello di unità didattica elaborata da Freddi la successione delle diverse fasi dell'unità consente dunque, di realizzare

un armonico gioco di rimandi tra i due emisferi: dalla fase di motivazione, che integra le emozioni e la curiosità del cervello destro con l'analisi dei bisogni di quello sinistro, si passa ad una fase di globalità (modalità destra), cui segue l'analisi della lingua (modalità sinistra) per approdare alla sintesi conclusiva. (Balboni, 2002, p. 28)

Per quanto riguarda la durata dello svolgimento delle unità didattiche, Freddi (1994, p. 111) propone come valutazione temporale per ciascuna unità “4-6 ore di lezione circa a seconda degli obiettivi posti dall'unità stessa e della consistenza dei materiali da manipolare”; la durata aumenta a 6-8 ore nel caso di unità pensate per la scuola secondaria (Freddi, 1999, p. 148) o a 6-10 ore o oltre nel caso di unità basate su testi letterari o tecnico-specialistici (Balboni, 2002, p. 103).

Lo stesso Balboni nel 2002 ha rivisto l'idea di unità didattica, proponendone un'innovazione attraverso il modello operativo della rete di unità di apprendimento:

Un'unità d'*apprendimento* può durare pochi minuti o anche un'ora: è l'unità di misura secondo la quale lo studente percepisce il suo apprendimento: “oggi a scuola abbiamo fatto..., ho imparato a...”; un'unità *didattica* è invece una tranche linguistico-comunicativa più complessa, realizzata mettendo insieme eventi, atti, espressioni, strutture linguistiche legati da un contesto situazionale. (Balboni, 2002, pp. 105-106)

In questo caso il modello operativo si basa su un'unità didattica da 6-10 ore che al suo interno contiene una rete di unità di apprendimento di varia durata, che può essere svolta in maniera flessibile, senza alcuna sequenza obbligata delle attività in termini di ordine temporale o gerarchico (l'unica sequenza rispettata nella sua progettazione è quella acquisizionale “i+1”, che presenta le strutture linguistiche secondo l'ordine naturale di acquisizione, a cui si è fatto riferimento in precedenza).

Nel caso del presente progetto di ricerca, si è optato, invece per la produzione di due brevi unità didattiche sul modello di Freddi, aventi però una durata compresa tra 1 ora e 30 e 2 ore. Tale scelta è dovuta al fatto che la sperimentazione delle attività didattiche sull'uso della varietà neostandard si è svolta nell'ambito di accordi siglati con alcuni atenei esteri che hanno concesso la disponibilità all'interno dei propri corsi curriculari per massimo 2 o 3 incontri della durata di 1 ora e 30 ciascuno, fino ad un massimo di 3-4 ore totali.

Pertanto, è bene precisare che, nonostante la piena consapevolezza

dell'evoluzione che ha investito il modello dell'unità didattica (cfr. i già citati Diadori, Palermo e Troncarelli, 2015, pp. 285-312) – dall'originaria matrice gestaltica fino alla moderna unità di apprendimento – la scelta di progettare due brevi unità didattiche improntate sul modello originario gestaltico è dovuta al contesto specifico di sperimentazione, che non disponeva di una continuità temporale tale da permettere lo svolgimento di una rete di unità di apprendimento.

Di conseguenza, le brevi unità prodotte possono essere considerate come delle forme ibride tra l'unità didattica tradizionale, di cui rispettano la struttura ispirata alla bimodalità emisferica secondo i già citati studi di Danesi, e le unità di apprendimento, di cui colgono la struttura reticolare, dato che offrono al docente varie attività con diversi livelli di complessità, da realizzare nell'ambito di un'unica sessione di lavoro di 1 ora e 30 - 2 ore e non di un percorso superiore alle 3-4 ore. Dopo aver fornito i necessari riferimenti teorici e metodologici, per maggiori dettagli sulla progettazione delle singole unità si rinvia al capitolo 6, mentre la sezione successiva sarà dedicata alla descrizione di un questionario progettato per la raccolta delle opinioni dei partecipanti sulle attività svolte.

4.7. Questionario sulle opinioni degli studenti

Per raccogliere dati sulla popolazione coinvolta e sulla percezione personale relativa all'utilità delle attività svolte, all'acquisizione dei contenuti linguistici e all'adeguatezza della piattaforma di *e-learning* rispetto agli scopi della didattica, è stato elaborato uno strumento d'osservazione da somministrare ai partecipanti al termine di ciascuna lezione.

Tra i vari strumenti che consentono l'osservazione diretta della didattica delle lingue tramite una registrazione di dati strutturata, si è scelto di progettare uno strumento che integrasse la tipologia del questionario con quella della *checklist* (si vedano gli studi di Nunan, 1989; Coonan, 2000; Pozzo, 2008). Entrambi gli strumenti sono tra quelli maggiormente utilizzati nell'ambito della ricerca-azione per la raccolta quantitativa dei dati (Coonan, 2000, p. 25). In particolare, la *checklist* si basa su una selezione di voci prestabilite, si tratta dunque di uno strumento chiuso, volto a focalizzare l'attenzione su

aspetti specifici individuati dal ricercatore. Come espresso da Coonan (ivi, p. 37):

Una checklist può essere usata per investigare praticamente qualsiasi aspetto del processo d'insegnamento e d'apprendimento. In sostanza non è altro che un elenco di elementi che sono da segnare, scegliere, categorizzare oppure mettere in ordine. Può fungere sia come scheda di osservazione 'in diretta' (cfr. COLT) sia come strumento di riflessione 'a distanza' dove l'uso dello strumento non è legato direttamente ad un qualche aspetto osservato in tempo reale. [...] La checklist può anche avere incorporata una scala valutativa se viene richiesto di valutare l'item che si è osservato o sul quale si sta riflettendo.

Pertanto, si è deciso di strutturare lo strumento come segue:

1. domande aperte di tipo conoscitivo: lingua materna; altre lingue straniere conosciute oltre all'italiano; da quanto tempo è in corso l'apprendimento dell'italiano; eventuali conoscenze pregresse sul tema linguistico-grammaticale dell'unità ed eventuale presenza di fenomeni linguistici analoghi nella lingua madre e/o in altre lingue straniere conosciute;
2. 3 liste di controllo (*checklist*), in cui si richiede di attribuire alle voci dell'elenco un valore su una scala numerica da 0 a 3. La scala di valori è stata ulteriormente esplicitata da *emoticon*, in modo da adattare la scheda ai diversi stili di apprendimento (analitici o olistici) degli apprendenti:
 - 2.1. la prima contiene l'elenco delle attività svolte nell'unità e si chiede agli apprendenti di valutarle in base alla loro utilità percepita;
 - 2.2. la seconda contiene quattro affermazioni sulla percezione dell'acquisizione del tema grammaticale dell'unità, ordinate dal livello massimo al livello minimo della soddisfazione personale (Es., dopo queste attività ho un'idea più chiara dell'uso dei riflessivi).
 - 2.3. la terza contiene quattro affermazioni sulla soddisfazione degli studenti rispetto all'uso della piattaforma *e-learning*.

Sia la seconda che la terza lista offrono, inoltre, spazi per aggiungere apporti personali, affinché il campo di osservazione non sia ristretto alle sole voci prestabilite, come messo in evidenza da Coonan (*ibidem*):

[...] le checklist (come tutte le schede di osservazione) consistono di una serie di categorie pre-designate per la registrazione delle osservazioni per cui sono capaci di dare informazioni solo riguardo le categorie previste. In questo modo si corre il rischio di non 'vedere' aspetti interessanti proprio perché la scheda ci orienta in una precisa direzione che esclude moltissime cose. Ciò non toglie che se si è ben individuato un ambito preciso di indagine la checklist o scheda di osservazione ci permette di raccogliere delle informazioni a riguardo in maniera precisa e sistematica.

La formulazione dei contenuti dello strumento di osservazione si è basata sul concetto neurobiologico di *Stimulus Appraisal* (Schumann, 1999; Schumann *et al.*, 2004), ossia di valutazione dell'input. Tale concetto, applicato alla glottodidattica (cfr. Daloiso, 2009, 2015), permette di descrivere la motivazione all'apprendimento come un processo continuo del cervello di valutazione e confronto degli input ricevuti dal mondo esterno con le aspettative, i bisogni e i desideri. Se in base a tale confronto l'input verrà giudicato positivamente, verrà inviato dalla memoria di lavoro ai centri deputati alla memoria a lungo termine. Tale processo di valutazione dell'input è stato definito *stimulus evaluation check* da Scherer (1984, 2001, 2009) e si basa sui seguenti criteri:

1. motivazione e *novità* dell'input: il soggetto valuta la novità dell'input, la sua prevedibilità e la sua discrepanza rispetto ad aspettative e conoscenze pregresse;
2. motivazione e *piacevolezza intrinseca* dell'input: viene valutato il senso di piacere o di dispiacere provocato dall'input;
3. *pertinenza* dell'input rispetto ai bisogni e agli obiettivi: si valuta se l'input è adeguato al raggiungimento dei propri obiettivi formativi, culturali, sociali ecc.;
4. *realizzabilità* sulla base della comprensibilità dell'input e della sua adeguatezza rispetto alle capacità del soggetto;

5. *sicurezza psico-sociale*: il soggetto valuta se la situazione (nello specifico, la situazione comunicativa, nel caso della glottodidattica) può rafforzare la propria immagine sociale o comprometterla; si tratta della paura della “perdita della faccia” (Balboni, 2002), che può innescare il cosiddetto filtro affettivo (Krashen, 1982), ossia una difesa psicologica che il cervello innalza quando un soggetto ha paura di sbagliare e teme di compromettere la propria immagine.

È su questi criteri che si è basata la progettazione del questionario da somministrare ai partecipanti alla sperimentazione, nella consapevolezza che, come affermato da Daloiso (2009, 2015), nell'insegnamento e apprendimento delle lingue è necessario che l'input susciti emozioni positive, evitando situazioni che generino paura e ansia.

Come per le unità didattiche, anche per la progettazione e somministrazione del questionario è stata utilizzata una piattaforma online, nello specifico *Socrative*⁴⁹, il cui riferimento è stato reperito spesso in associazione alle piattaforme di *e-learning* prese in esame per la progettazione delle unità didattiche, perché consente di svolgere verifiche in classe scegliendo tra varie tipologie di quiz.

I fattori principali che ne hanno determinato l'adozione per la presente ricerca sono i seguenti. Innanzitutto, la possibilità di scegliere tra varie tipologie di quiz e di personalizzarne la struttura e la visualizzazione a seconda delle esigenze. Inoltre, vi è la possibilità per il docente di ricevere i risultati in tempo reale sul suo dispositivo durante lo svolgimento del test e non solo al termine, il che permette di monitorare l'andamento del quiz e di individuare a colpo d'occhio eventuali blocchi da parte dei partecipanti, funzione particolarmente utile nel caso di partecipanti italofoni non nativi che svolgono il test in lingua italiana. Una volta ottenuti i dati finali, la piattaforma offre una serie di funzioni che agevolano il lavoro di analisi dei dati, poiché consente di analizzare i report online dei dati ottenuti su un foglio di lavoro di Google o di impostare l'invio via e-mail in formato Ms-Excel, di effettuare il calcolo automatico delle percentuali delle risposte date nella classe e di realizzare automaticamente istogrammi. Infine, consente sia una visualizzazione a livello globale degli esiti dell'intera classe, sia di avere una visione dettagliata delle risposte dei singoli partecipanti, con

⁴⁹ <https://socrative.com>.

la possibilità di nascondere i nomi per preservarne l'anonimato.

In appendice (cfr. Appendice 3) si riportano gli strumenti di osservazione utilizzati per valutare rispettivamente l'unità didattica sui verbi riflessivi e quella sulla scelta tra congiuntivo e indicativo. Per le informazioni sugli esiti del questionario si rinvia al capitolo 6, mentre il capitolo successivo sarà dedicato alla descrizione dei risultati ottenuti dall'interrogazione del *corpus*.

5. FREQUENZA DEI FENOMENI GRAMMATICALI INDAGATI

Dopo aver descritto nel capitolo precedente l'elaborazione delle stringhe di ricerca in linguaggio CQL (cfr. sezione 4.4), in questo capitolo verranno esposti i risultati ottenuti dalla ricerca nel corpus PAISÀ dei tratti grammaticali dell'italiano dell'uso medio, al fine di determinare quali di essi siano i più ricorrenti.

Il calcolo della frequenza permette di individuare quali tra i tratti linguistici che, all'epoca della pubblicazione degli studi di Sabatini sul tema, erano ancora percepiti come in via di graduale inserimento nella norma, siano invece ora maggiormente utilizzati nella produzione scritta di media formalità o informalità dagli italofoeni nativi. A questo proposito, è possibile ipotizzare che alcuni tratti linguistici siano maggiormente utilizzati dai parlanti perché non più percepiti come scorretti e che, in quanto tali, siano ritenuti da loro accettabili in un registro non particolarmente "basso". Nella riflessione su cosa costituisca il neostandard, disporre di dati sulla diffusione di questi fenomeni permette a chi insegna l'italiano L2/LS di poter

valutare, nell'ottica del criterio della funzionalità e dell'accettabilità sociale (cfr. capitolo 2), quali di essi possano essere considerati parte integrante della nuova varietà standard e, dunque, insegnati, e quali invece vadano solo menzionati come varianti substandard, seppur diffuse.

5.1. Calcolo della frequenza dei fenomeni indagati

Allo scopo di determinare la diffusione dei tratti linguistici indagati, per ciascuno di essi sono state svolte le seguenti operazioni, che verranno di seguito illustrate prendendo l'esempio del fenomeno grammaticale del doppio partitivo:

1. innanzitutto, per ciascuno dei 3 campioni (cfr. sezione 4.5) è stato calcolato il *rapporto tra il numero di occorrenze salvate* dopo l'eliminazione di eventuali casi non inerenti e duplicati (ossia le occorrenze effettivamente corrispondenti al fenomeno grammaticale indagato) *e le 200 occorrenze iniziali del campione*. Tale procedura permette di verificare quanto sia efficace la stringa di ricerca, al fine di avere una stima più veritiera e affidabile della frequenza dei fenomeni indagati. Nello specifico, nel caso del doppio partitivo, per ciascun campione sono stati calcolati i seguenti rapporti: campione 1: $161/200=0,805$; campione 2: $167/200=0,835$; campione 3: $174/200=0,87$;
2. *media campione*: è stata calcolata la media dei tre rapporti ottenuti, determinando così la frequenza media del fenomeno per campione, ad esempio: $(0,805 + 0,835 + 0,87)/3=0,84$.

Al fine di rendere paragonabili i valori ottenuti per ciascuno dei fenomeni grammaticali indagati, sono state, inoltre, calcolate:

1. *frequenza stimata*: è data dal prodotto tra il totale delle occorrenze inizialmente ottenute per il fenomeno indagato tramite l'interrogazione del corpus PAISÀ e la media campione, ossia nel caso del doppio partitivo: $659 \cdot 0,84=553,56$;
2. *frequenza stimata per 1 milione di parole*: è ottenuta attraverso due passaggi: si calcola il prodotto tra la frequenza stimata e 1.000.000 e il risultato ottenuto si divide per il numero totale di *token* del

corpus PAISÀ, ossia: frequenza stimata*1.000.000/265.062.636¹.

Esempio: 553,56* 1.000.000/265.062.636.

Ciò è stato svolto nella consapevolezza che la scelta di avere tre campioni casuali su cui calcolare dapprima la frequenza stimata (al netto del “rumore”, ossia casi che, pur presentando delle analogie strutturali con il fenomeno indagato, non fossero inerenti) e, successivamente, la normalizzazione per milione di parole non rappresentino procedure standard. Infatti, generalmente per quantificare i casi inerenti al fenomeno ricercato rispetto al cosiddetto “rumore” le stime di frequenza si basano su un campione unico. Tuttavia, si è ritenuto che avere tre campioni casuali consentisse di disporre di una base di dati più eterogenea e, pertanto, di poter formulare delle affermazioni sulla frequenza dei fenomeni basate – si auspica – su una maggiore fondatezza.

Per ciascuna di queste frequenze si è proceduto anche al calcolo della *deviazione standard*, che fornisce una misura di come siano distribuiti i dati nel campione rispetto alla media, ossia permette di sapere se la media costituisce una rappresentazione affidabile dei dati. Quanto più il valore della deviazione standard è vicino a 0, tanto più i dati si avvicinano alla media e quindi si può affermare che c'è poca dispersione nel campione. Nell'interesse della sinteticità e degli scopi di questa trattazione, per il dettaglio delle procedure statistiche adottate e delle ragioni per cui si è scelto di calcolare la deviazione standard campione e non l'errore standard della media, si rinvia a Zingaro (2021, cap. 5). I valori della frequenza stimata per 1 milione di parole (e della relativa deviazione standard) di tutti i fenomeni grammaticali indagati sono stati utili per confrontare tra loro i fenomeni e ordinarli dal fenomeno meno frequente a quello più frequente all'interno del corpus (cfr. tabelle 5.1, 5.2 e 5.3).

I 26 casi esaminati, che includono 13 tratti della lista di Sabatini (1990) e i relativi sottofenomeni di alcuni di essi, sono stati suddivisi in tre gruppi in base alla loro frequenza stimata per milione di parole nel corpus PAISÀ. Ogni gruppo corrisponde a un terzile nella lista dei

¹ Il corpus PAISÀ è stato consultato tramite il programma di interrogazione NoSketch Engine (Rychlý, 2007, p. 69) e non tramite la sua interfaccia originale, in seguito ad alcune difficoltà tecniche riscontrate con quest'ultima. Il conteggio è quindi basato sul numero di *token* calcolato da NoSketch Engine.

fenomeni ordinata per frequenza, e dunque: 8 fenomeni a frequenza bassa, 9 fenomeni a frequenza media e 9 fenomeni a frequenza alta, riportati nelle seguenti tabelle riassuntive.

Tabella 5.1: Fenomeni a bassa frequenza

Fenomeni a bassa frequenza (0 < <i>f_q</i> < 3)	Frequenza stimata per 1M di parole
1. Soggetto in posizione rematica rispetto al predicato: <i>niente</i> + sostantivo pl. (esempio: <i>niente soldi</i>)	2,21
2. Dislocazione a sinistra: doppio partitivo (es.: <i>di pane non ne ho più</i>)	2,08
3. Impiego dell'indicativo al posto del congiuntivo: indicativo nelle ipotetiche (es.: <i>se lo sapevo, te lo dicevo</i>)	1,59
4. Dislocazione a sinistra: ripetizione del pronome in forma tonica e atona (es.: <i>a me mi, a te ti...</i>)	0,41
5. <i>Ci</i> attualizzante (es.: <i>non c'ho tempo</i>)	0,38
6. Impiego dell'indicativo al posto del congiuntivo: <i>avere paura/ timore che</i> (es.: <i>aveva paura che [...] era diventato sospetto</i>)	0,03
7. Soggetto in posizione rematica rispetto al predicato (in frasi subordinate) (es.: <i>non cambia nulla, se resta l'unilateralismo</i>)	< 0,01
8. Impiego dell'indicativo al posto del congiuntivo: verbi di opinione/sentimento (es.: <i>Pensavo che erano loro le migliori</i>)	< 0,01

Tabella 5.2: Fenomeni a media frequenza

Fenomeni a media frequenza (3 < <i>f_q</i> < 20)	Frequenza stimata per 1M di parole
1. Pronome <i>gli</i> vs <i>le/loro</i> (es. <i>ho parlato ad Anna e gli ho detto; ho chiamato i miei amici/ le mie amiche e gli ho detto</i>)	16,50
2. Soggetto in posizione rematica rispetto al predicato (in frasi principali) (es.: <i>Entra Andrea</i>) ²	15,96
3. Frase scissa con <i>a</i> (es.: <i>È lo stesso Dioniso a dirlo</i>)	12,77

² Per tutte le altre tipologie ricercate di questo fenomeno, si fa riferimento alle descrizioni delle stringhe di ricerca nel cap. 4 o ai contenuti successivi di questo capitolo.

4. <i>Che</i> polivalente (es.: <i>Vieni che ti spalmo la crema</i>) ³	9,46
5. Verbo in forma pronominale (es.: <i>io mi bevo</i>)	6,48
6. Impiego dell'indicativo al posto del congiuntivo: indicativo nell'interrogativa indiretta (es.: <i>mi chiedo se vi ha colpito</i>)	5,35
7. Concordanza a senso: soggetto collettivo sing. + <i>di</i> partitivo + predicato pl. (es.: <i>un milione di persone hanno visto</i>)	3,55
8. Impiego dell'indicativo al posto del congiuntivo: <i>non</i> + verbi di opinione/sentimento (es.: <i>non penso che è così chiaro</i>)	3,33
9. <i>Cosa</i> interrogativo vs <i>che cosa</i> nell'interrogativa indiretta (es.: <i>Cosa vuoi?</i>)	3,01

Tabella 5.3: Fenomeni ad alta frequenza

Fenomeni ad alta frequenza (fq > 20)	Frequenza stimata per 1M di parole
1. <i>Ma</i> a inizio frase (es.: <i>Ma non finisce qui!</i>)	344
2. <i>E</i> a inizio frase (es.: <i>E gli va malissimo</i>)	289,48
3. Il partitivo preceduto da preposizione (es.: <i>con degli amici</i>)	67,06
4. <i>Per cui</i> con valore conclusivo (es.: <i>Per cui ho rinunciato</i>)	52,28
5. <i>Comunque</i> a inizio frase (es.: <i>Comunque vi ringrazio tutte</i>)	32,31
6. Impiego dell'indicativo al posto del congiuntivo: verbo <i>essere</i> + aggettivo + <i>che</i> (es.: <i>È possibile che [...] è diventato un optional</i>)	26,73
7. <i>Cosa</i> interrogativo vs <i>che cosa</i> nell'interrogativa diretta (es.: <i>Mi chiedo cosa sia accaduto</i>)	26,01
8. Frase scissa con <i>che</i> (es.: <i>È il pubblico che decreta il successo</i>)	24,33
9. <i>Allora</i> a inizio frase (es.: <i>Allora Bright scatena un attacco</i>)	20,62

³ Si veda la nota precedente.

Per ciascun gruppo è stato realizzato un istogramma (cfr. grafici 5.1, 5.2 e 5.3) in cui l'altezza delle colonne blu corrisponde ai valori sopra citati della frequenza stimata per 1 milione di parole, e la lunghezza delle linee sottili nere sovrapposte alle colonne indica l'ampiezza della deviazione standard stimata.

Grafico 5.1: Fenomeni a bassa frequenza

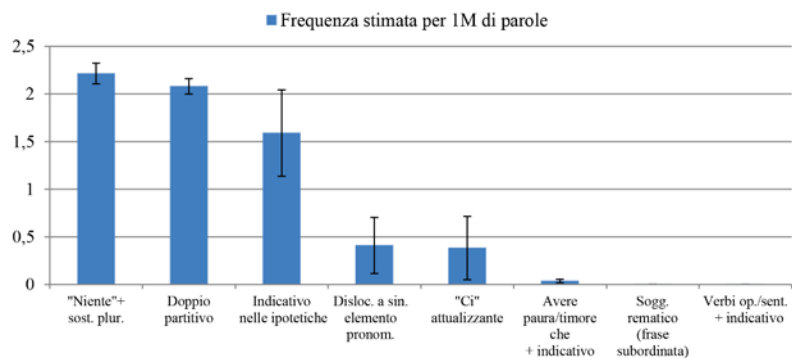


Grafico 5.2: Fenomeni a media frequenza

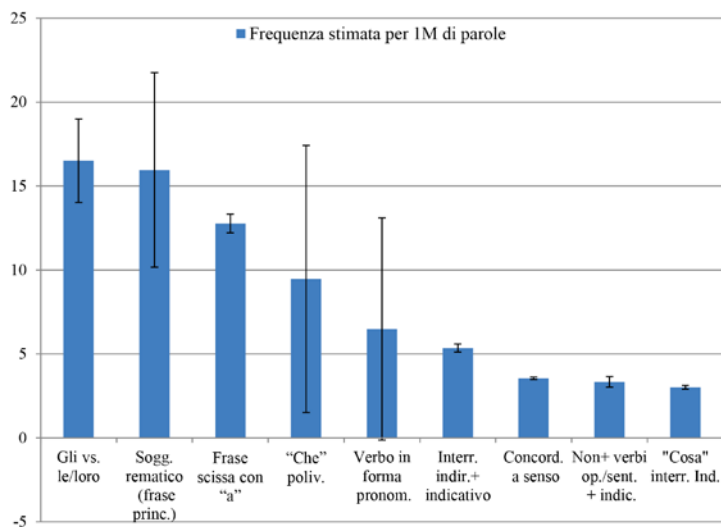
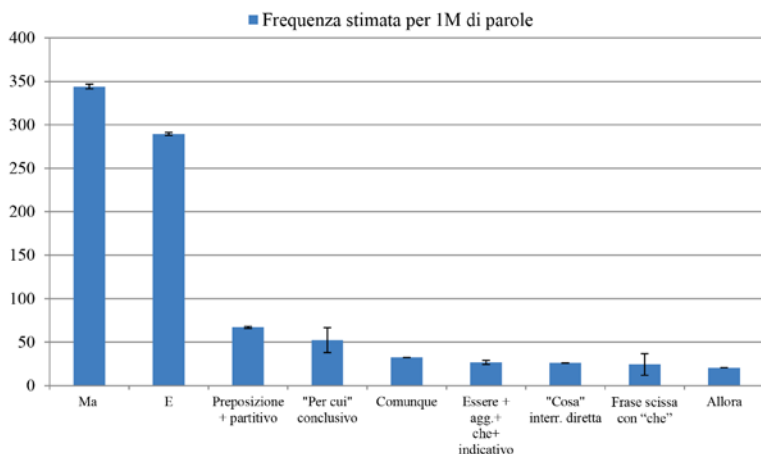


Grafico 5.3: Fenomeni ad alta frequenza

Osservando i grafici 5.1, 5.2 e 5.3 emerge immediatamente che alcuni fenomeni presentano un'elevata deviazione standard rispetto alla frequenza. Senza entrare in questa sede nel dettaglio dei dati ottenuti, per i quali si rinvia a Zingaro (2021, cap. 5), per ragioni di chiarezza e per fornire un parziale commento dei grafici, si precisa quanto segue. Per poter affermare che la frequenza ottenuta fosse un dato affidabile e determinare da cosa dipendesse una simile variazione dell'errore, per ciascun gruppo sono stati analizzati due casi che presentano un'elevata deviazione standard e due casi con poca variazione, prendendo in esame i seguenti aspetti: il numero di stringhe di ricerca prodotte, il numero di campioni ricavati da ciascuna stringa, il totale iniziale di occorrenze da cui è composto ciascun campione e il totale delle occorrenze salvate per ciascun campione.

In linea generale, i casi che presentano un valore elevato della deviazione standard (cfr. grafici 5.1, 5.2 e 5.3) possono essere suddivisi in:

1. variazione non casuale, alla cui base vi sono fattori ben identificabili: un numero di stringhe di ricerca superiore a 1 (ad esempio, 12 stringhe per ricercare il verbo in forma pronominale per indicare partecipazione affettiva, 4 per la frase scissa nella sua variante retta da *che*, 3 per il *che* polivalente e l'uso del *ci* attualizzante); la produttività delle stringhe di ricerca (ossia il

fatto che in alcuni casi le stringhe abbiano permesso di reperire un numero elevato di occorrenze, distribuibili su 3 campioni, e in altri casi meno, fino a costituire addirittura un campione unico); il numero di campioni ricavati da ciascuna stringa (3 campioni vs un campione unico); il totale iniziale di ciascun campione (200 unità o meno); la differenza esistente tra il totale iniziale di occorrenze reperite per ciascun campione e il totale delle occorrenze salvate al termine della ripulitura del campione dai casi non inerenti (ad esempio, 176 su 200), eccetera. In questi casi, i valori relativi alla media campione, deviazione standard campione eccetera sono il risultato di calcoli che hanno coinvolto collettivamente tutti i campioni ottenuti da tutte le stringhe di ricerca (a seconda dei casi 12, 4 o 3, come è stato descritto) elaborate per quel fenomeno, andando a costituire un unico campione caratterizzato, per l'appunto, da una notevole eterogeneità interna;

2. variazione casuale, ove non emerge una causa evidente per cui per alcuni fenomeni la frequenza sia alta e per altri bassa. In generale, ci si aspetta che la variazione sia casuale nei fenomeni a bassa frequenza per pure ragioni probabilistiche.

Queste sono le possibili cause della variazione riscontrate. Per contro, la media è stabile nei casi caratterizzati da una sola stringa di ricerca, i cui risultati sono distribuiti su 3 campioni da 200 unità e in cui la differenza tra il totale iniziale di occorrenze per ciascun campione e il totale delle occorrenze salvate al termine della ripulitura dei casi non inerenti è molto ridotta. Tra i fenomeni che presentano poca variazione si citano a titolo di esempio:

1. *Primo gruppo, fenomeni a bassa frequenza:*
 - 1.1. soggetto in posizione rematica (*niente* seguito da un sostantivo plurale): una sola stringa di ricerca con tre campioni da 200, i cui totali dei risultati salvati variano da 160 a 176;
 - 1.2. dislocazione a sinistra (uso del doppio partitivo): una sola stringa di ricerca con tre campioni da 200, i cui totali dei risultati salvati variano da 161 a 174;
2. *Secondo gruppo, fenomeni a media frequenza:*
 - 2.1. la frase scissa retta da *a*: una sola stringa di ricerca con tre campioni da 200, i cui totali dei risultati salvati variano da 166 a 180;

- 2.2. l'uso dell'indicativo nella frase interrogativa indiretta: una sola stringa di ricerca con tre campioni da 200, i cui risultati salvati variano da 159 a 174;
3. *Terzo gruppo, fenomeni ad alta frequenza:*
 - 3.1. il partitivo preceduto da preposizione semplice: una sola stringa di ricerca con tre campioni da 200, i cui totali dei risultati salvati variano da 165 a 170;
 - 3.2. l'uso dell'indicativo al posto del congiuntivo nella costruzione "verbo *essere* + aggettivo + *che* + indicativo" una sola stringa di ricerca con tre campioni da 200, i cui totali dei risultati salvati variano da 109 a 128.

Alla luce di quanto esposto finora, è stata svolta un'analisi più approfondita sui fenomeni che presentano un valore elevato della deviazione standard. In questa nuova fase di analisi, le operazioni di calcolo utili ad un'analisi quantitativa sono state ripetute separatamente per ciascuno dei sottofenomeni grammaticali dei fenomeni indagati e non più selezionando globalmente tutti i dati relativi alle varie stringhe prodotte per inquadrare un determinato fenomeno grammaticale. Ad esempio, analizzando separatamente i dati di frequenza relativi a ciascuna delle 8 stringhe di ricerca della dislocazione a sinistra dell'elemento pronominale (sottofenomeni: *a me mi*; *a te ti*; *a lui gli*; *a lei le*; *a lei gli*; *a noi ci*; *a voi vi*; *a loro gli*), che nella fase precedente erano stati conteggiati collettivamente, producendo un campione con notevole eterogeneità interna.

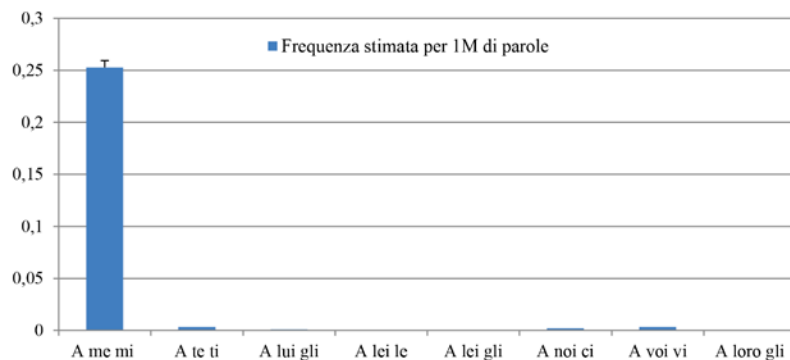
Senza entrare in questa sede nel dettaglio dei risultati ottenuti, in generale è possibile affermare che questo approfondimento ha permesso di identificare le cause della variazione all'interno dei campioni dei fenomeni grammaticali presi in esame e ha mostrato come la variazione si sia notevolmente ridotta dopo che le operazioni di calcolo utili ad un'analisi quantitativa sono state svolte separatamente per ciascuno dei sottofenomeni. Per una trattazione dettagliata, si rinvia a Zingaro (2021, pp. 164-176). Di seguito verranno, invece, brevemente descritti solo alcuni dei fenomeni linguistici la cui variazione di frequenza è ritenuta linguisticamente interessante (nello specifico uno per ciascun gruppo) e il caso dell'uso dell'indicativo al posto del congiuntivo, che si articola nei suoi sottofenomeni su tutti e tre i gruppi.

1. Dislocazione a sinistra dell'elemento pronominale (sottofenomeni: *a me mi*; *a te ti*; *a lui gli*; *a lei le*; *a lei gli*; *a noi ci*; *a voi vi*; *a loro gli*): osservando la tabella 5.4 e grafico 5.4 emerge che l'uso della prima persona *a me mi* è nettamente più frequente delle altre persone. In particolare, ricorre con il verbo *piacere* (*a me mi piace vivere alla grande*; *a me mi farebbe piacere*; *a me mi piace quella cosa*), con verbi che esprimono una percezione (*Ma a me mi era sembrato che già da molto tempo qualcosa stava cambiando*; *Insomma, a me mi sa tanto di flop questo cd!*) e di espressione di gusti e sentimenti (*a me mi fanno schifo le cose sporche*; *a me mi preoccupa*; *a me mi va bene anche alle dieci, le undici*; *a me mi fa venire il sangue al cervello*).

Tabella 5.4: Approfondimento di un fenomeno a bassa frequenza ($0 < f_q < 3$)

Dislocazione a sinistra: ripetizione del pronome in forma atona e tonica. Segue in corsivo il dettaglio dei sottofenomeni:	Frequenza stimata per 1M di parole ⁴
<i>A me mi</i>	0,25
<i>A te ti</i>	< 0,01
<i>A lui gli</i>	< 0,01
<i>A lei le</i>	< 0,01
<i>A lei gli</i>	0
<i>A noi ci</i>	< 0,01
<i>A voi vi</i>	< 0,01
<i>A loro gli</i>	< 0,01

⁴ NB: Il numero estremamente ristretto di occorrenze reperite per alcuni sottofenomeni grammaticali, quali *a te ti*, *a lui gli*, *a lei le*, *a lei gli*, *a noi ci*, *a voi vi* e *a loro gli* ha comportato la creazione di un campione unico per ciascuno di essi (cfr. sez. 5.1), rendendo, quindi, impossibile il calcolo della media e relativa deviazione standard. Per questi sottofenomeni è stato dapprima calcolato il rapporto tra il numero di occorrenze salvate dopo l'eliminazione di eventuali casi non inerenti e duplicati e le occorrenze iniziali del campione; successivamente questo valore è stato normalizzato per 1 milione di parole (ossia, moltiplicato per 1.000.000 e diviso 265.062.636).

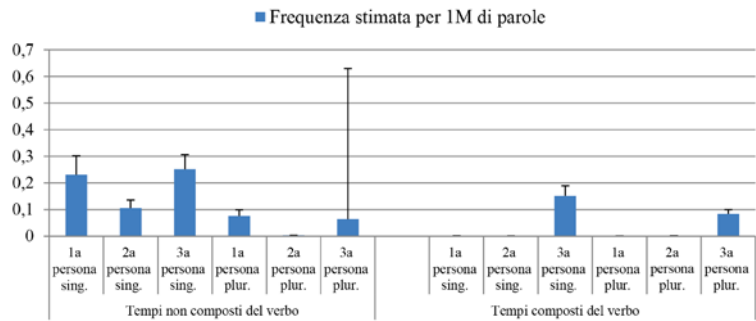
Grafico 5.4: Dislocazione a sinistra dell'elemento pronominale

2. Verbo in forma pronominale per indicare partecipazione affettiva (tabella 5.5 e grafico 5.5): le forme più frequenti risultano essere la terza, la prima e la seconda persona singolare dei tempi non composti dei verbi e la terza persona singolare dei tempi composti, di cui seguono esempi. In particolare, nel caso della terza e prima persona singolare dei tempi non composti, le occorrenze appaiono equamente distribuite tra i verbi *bere*, *fare*, *guardare*, *mangiare* e *vedere*: *Questo era un criceto normale, che si mangiava i suoi semi di girasole*; *Non è possibile che una persona che si beve una birra nel parco debba essere multata in questo modo*; [...] *che una persona media si guardi un programma dall'inizio alla fine*; *Se voglio vedere un film di altra pasta, mi guardo un film di Kubrick o di Nolan*; *Cosa ci perde la Rai se mi vedo i programmi tramite vlc?*; *Domani mi faccio il calendario*; *Mi faccio un giro*. Invece, nel caso della seconda persona singolare dei tempi non composti e la terza persona singolare dei tempi composti, prevale nettamente l'uso del verbo *fare*. Seguono gli esempi: *Invece del prato inglese ti fai l'orto e tieni un po'di galline*; *ti fai un giro nei siti*; *ti fai una bella mangiata qui a New York*; *ti fai un panino davanti alla tv*; *Stefano con al volante il fratello maggiore si è fatto un viaggio su una fiat 850*; *Si è fatto il giro del mondo*; *Si è fatta il caffè appena alzata*.

Tabella 5.5: Approfondimento di un fenomeno a media frequenza ($3 < fq < 20$)

Verbo in forma pronominale (es. <i>io mi bevo un caffè</i>). Segue in corsivo il dettaglio dei sottofenomeni: Tempi non composti del verbo	Frequenza stimata per 1M di parole ⁵
<i>1a persona sing.</i>	0,23
<i>2a persona sing.</i>	0,10
<i>3a persona sing.</i>	0,25
<i>1a persona plur.</i>	0,07
<i>2a persona plur.</i>	0
<i>3a persona sing.</i>	0,06
Tempi composti del verbo	
<i>1a persona sing.</i>	< 0,01
<i>2a persona sing.</i>	< 0,01
<i>3a persona sing.</i>	0,15
<i>1a persona plur.</i>	< 0,01
<i>2a persona plur.</i>	< 0,01
<i>3a persona plur.</i>	0,08

Grafico 5.5: Verbo in forma pronominale



⁵ Per i sottofenomeni del verbo in forma pronominale alla 2a persona plurale dei tempi non composti del verbo e alla 1a e 2a persona singolare e alla 1a e 2a persona plurale dei tempi composti del verbo valgono le stesse considerazioni fatte nella nota precedente.

3. *Fenomeno ad alta frequenza*: frase scissa retta da *che* (tabella 5.5 e grafico 5.5)⁶:

3.1. verbo *essere* + gruppo nominale + *che* + verbo coniugato, esempi: *è quest'approccio che persegue attualmente l'Istituto Europeo di Mimopedagogia; Furono questi leader che firmarono gli accordi; È il pubblico che decreta il successo di un artista*. Questa tipologia risulta essere la più frequente nell'uso.

3.2. verbo *essere* + gruppo nominale preceduto da preposizione + *che* + verbo coniugato: è interessante notare la ricorrenza di casi in cui il gruppo nominale è composto dal pronome o aggettivo dimostrativo *questo* in funzione di ripresa anaforica (ossia di forma linguistiche con cui "il parlante fa riferimento ad un referente al quale egli, nel suo discorso, ha già fatto riferimento con un'espressione antecedente" (Conte, 1999, p. 19)). Esempi: *È per questo motivo che a lungo andare giocare con squadre di Terza Categoria; Ed è per questo che nei porti si possono trovare tante razze diverse; È in questo senso che la giustizia diventa una virtù morale*;

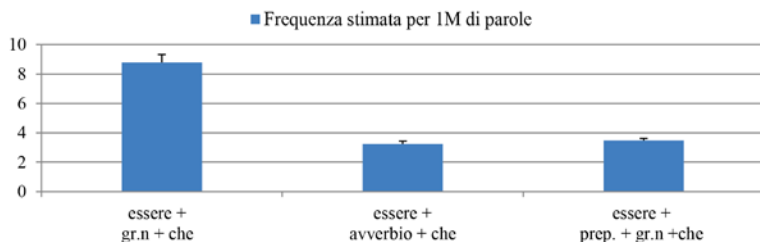
3.3. verbo *essere* + avverbio + *che* + verbo coniugato: *È così che viene a conoscenza del fatto; Fu lì che morì dopo venti anni; È qui che potrete ammirare il bel castello*.

Infine, in generale, per il fenomeno della frase scissa, va notato che la variante retta da *che* risulta molto più frequente della variante retta da *a* (*È lo stesso Dioniso a dirlo; Fu lui a portare con sé questo modo in Germania*).

Tabella 5.6: Approfondimento ad alta frequenza (fq > 20)

Frase scissa retta da <i>che</i> Segue in corsivo il dettaglio dei sottofenomeni:	Frequenza stimata per 1M di parole
<i>Verbo "essere" + gruppo nominale + che</i>	8,78
<i>Verbo "essere" + avverbio + che</i>	3,25
<i>Verbo "essere" + preposizione + gruppo nominale + che</i>	3,48

⁶ La stringa relativa al sottofenomeno "verbo *essere* + verbo all'infinito + *che* + verbo coniugato" (esempi: *è restare che dobbiamo fare; è votare che è importante* ecc.) è stata esclusa dall'analisi, poiché su 7 occorrenze scaricate dal corpus solo una corrisponde al fenomeno in esame.

Grafico 5.6: Frase scissa retta da *che*

Infine, per quanto riguarda l'uso dell'indicativo al posto del congiuntivo, i suoi sottofenomeni sono reperibili in ciascuno dei tre terzi. Questa differenza nella frequenza d'uso si ricollega ai dati della ricerca di Ondelli e Romanini (2018; cfr. cap. 2), da cui è emerso che l'alternanza indicativo/congiuntivo è uno dei fenomeni linguistici per i quali le valutazioni degli intervistati sull'accettabilità sono molto variabili. In sintesi, la sostituzione dell'indicativo al congiuntivo viene percepita come una variazione di registro, nella cui valutazione si richiede di tenere conto del contesto sintattico (in interrogative indirette, dopo verbi che esprimono incertezza, con verbi di opinione e nelle relative restrittive). Di seguito si elencano i diversi sottofenomeni in ordine decrescente di frequenza:

1. *essere* + aggettivo + *che* + indicativo: *È probabile che a lui si rifecce il Boccaccio; È possibile che governare SERIAMENTE l'Italia è diventato un optional;*
2. uso dell'indicativo nell'interrogativa indiretta: *Due anni fa le è stata posta la domanda se questo credito era stato recuperato o era andato perduto; Mi chiedo se vi ha colpito come ha colpito me;*
3. la negazione dei verbi di opinione: *Non penso che è così chiaro; Fino a 15 anni, non sapevo che c'era l'occupazione; Certa gente non vuole che vengano scoperti i mali dell'Italia;*
4. indicativo nelle frasi ipotetiche (*Se lo facevano anche un mese fa poteva avere molto successo*). Interessante notare anche un caso "misto" di uso corretto del condizionale nella principale e indicativo nella secondaria. *È un po' lungo ma se separavo tutto, temevo che poi si sarebbe generata confusione;*
5. *avere paura/timore che* + indicativo: non sono stati reperiti esempi con *avere timore*, ma solo con *avere paura*: *Aveva paura che quanto*

conquistato sulle strade era diventato sospetto; Avete paura che il Milan si conferma anche quest'anno campione d'Europa;

6. verbi di opinione/sentimento + indicativo: *Pensavo che erano loro le migliori a far musica; Chi crede che i problemi economici sono sempre risolvibili; Io spero che la censura riguarda tutti quei delitti stile CSI; Alcuni ritengono che l'Atto di abiura influenzò direttamente Thomas Jefferson.*

Infine, una breve riflessione sulla cercabilità dei fenomeni linguistici nel corpus. Ci sono due fenomeni che presentano un valore della deviazione standard elevato rispetto alla frequenza, per i quali si è deciso di non ricalcolare separatamente i risultati di ciascuna stringa, ma di lasciare invariato il calcolo collettivo dei dati relativi in un unico campione eterogeneo: si tratta del *che* polivalente e del soggetto in posizione rematica rispetto al predicato nelle frasi principali.

Innanzitutto, dal punto di vista linguistico, non ci sono ragioni per separare le occorrenze: per entrambi i fenomeni, le diverse stringhe di ricerca prodotte non identificano dei veri e propri sottofenomeni diversi tra loro, né dei cambiamenti morfologici all'interno di un fenomeno di per sé unitario (ad esempio, l'intera flessione dei verbi riflessivi con soggetto pronominale "affettivo": *mi mangio, ti mangi* ecc.) Seguono esempi.

Nello specifico, nel caso del *che* polivalente sono state prodotte 3 stringhe di ricerca per identificare: 1) il *che* esplicativo/consecutivo (*Vieni che ti spalmo la crema*) e il *che* causale (*Potevano fare tutte le guerre che volevano che a lui gli facevano un baffo*), 2) il *che* pseudorelativo (*Li vedo che agiscono come al solito; Li vedi che corrono?*) e 3) il *che* consecutivo-presentativo (con avverbi di dubbio) (*Io mi rendo conto di una cosa. Che se devo fare questa vita per altri dieci anni, una vita da trottola, girando tutto il giorno da un posto all'altro, rischio di sfinirme* (sic.)).

Per quanto riguarda il soggetto in posizione rematica rispetto al predicato nelle frasi principali, la ricerca di questo fenomeno grammaticale si basa su 5 stringhe, finalizzate a reperire le varie tipologie di predicato coinvolto, a seconda dei casi, con: 1) verbo non composto (*Entra Andrea*); 2) verbo composto con ausiliare *essere* (*È giunto il momento di fare un po' di teoria*); 3) verbo composto con ausiliare *avere* con l'oggetto tra il participio e il soggetto (*Ha suscitato sdegno l'articolo*); 4) verbo composto con ausiliare *avere* (senza oggetto) (*Ha cancellato lei stessa la*

pagina in estate); 5) verbo in forma passiva con *andare, essere, venire* (*fu istituito il Commissario distrettuale* (tempo non composto dell'ausiliare *essere*); *vennero svolte numerose prove* (tempo non composto dell'ausiliare *venire*); *va garantita la centralità* (tempo non composto dell'ausiliare *andare*); *sono state invitate le autorità* (tempo composto dell'ausiliare *essere*); *è andato distrutto il tempio* (tempo composto dell'ausiliare *andare*).

In sostanza, in entrambi i casi la scelta di conteggiare i dati collettivamente in un unico campione finale si giustifica alla luce del fatto che le varie stringhe di ricerca prodotte rappresentano diverse formulazioni sintattiche dello *stesso* fenomeno grammaticale, per il quale non è stato possibile progettare un'unica stringa che coprisse tutte le sue diverse formulazioni.

Ciò è dovuto anche al fatto che la progettazione delle stringhe si è basata su un'annotazione per parti del discorso (POS, *parts-of-speech*), che non sono sempre lo strumento ideale per ricercare fenomeni sintattici, specialmente se questi coinvolgono elementi della frase a lunga distanza. Per questi ultimi occorrerebbe fare ricerche basate su analisi sintattica (*parsing*) (cfr. Capitolo 4), che sono possibili attraverso il linguaggio CQL, ma per le quali sarebbe stato necessario progettare delle stringhe di ricerca molto più complesse.

La riflessione finale, derivante da quanto esposto finora in questo capitolo, è che i valori della frequenza e della deviazione standard dipendono massicciamente dalla formulazione delle stringhe di ricerca e dal tipo di annotazione (POS o *parsing*) su cui si basa la ricerca stessa. Ne consegue che alcune stringhe di ricerca sono più precise, mentre altre lo sono meno. Come già illustrato nel capitolo 4 (cfr. sezioni 4.6.1 e 4.6.2, dedicate al linguaggio CQL e alla progettazione delle stringhe), alcuni tratti grammaticali sono semplicemente più facilmente ricercabili tramite linguaggio CQL rispetto ad altri, perché legati a fenomeni sintattici e morfologici basati su delle sequenze fisse di elementi grammaticali che co-occorrono. Ad esempio, un conto è ricercare la ricorrenza di *a me mi*, o di un verbo di opinione seguito sempre da un verbo al congiuntivo, ben altro è riuscire a inquadrare un fenomeno come il *che* polivalente, che risulta più “evanescente” dal punto di vista sintattico e distribuzionale, dato che non ci sono categorie morfosintattiche specifiche con cui co-occorre.

A questo proposito vale la riflessione fatta da Budassi (2017, p. 157) circa la reperibilità del *che* polivalente, applicabile anche ad altri feno-

meni grammaticali la cui reperibilità risulta complessa: “La ricerca del ‘che’ polivalente in un corpus taggato unicamente per *part of speech* è difficoltosa. Il ‘che’ polivalente non co-occorre (in maniera univoca) con una qualunque altra parte del discorso con un alto livello di probabilità”. Pertanto risultati significativi possono essere ottenuti “non solo ricercando quanto normalmente è presente nel contesto di un ‘che’ polivalente, ma anche escludendo tutto ciò che è probabilmente presente nel contesto di un ‘che’ di altra natura” (*ibidem*). Sulla base di queste riflessioni, la sezione successiva sarà dedicata ai criteri applicati per la selezione dei fenomeni grammaticali sui quali costruire le unità didattiche.

5.2. Criteri di selezione dei fenomeni indagati

Al termine di queste riflessioni, è necessario precisare che, nella selezione dei fenomeni sui quali incentrare le unità didattiche, si è tenuto conto non solo della loro frequenza all’interno del corpus, ma anche di altri fattori.

Prendere in considerazione la frequenza è importante, perché consente di individuare quali tra i fenomeni linguistici che all’epoca della pubblicazione degli studi di Sabatini erano ancora percepiti come in contrasto con la norma grammaticale si siano gradualmente inseriti nell’uso linguistico di media formalità/informalità. Disporre di questi dati sulla diffusione permette a chi insegna l’italiano L2/LS innanzitutto di avere la percezione di quali fenomeni linguistici è più probabile che gli apprendenti di lingua italiana L2/LS trovino nei testi scritti. In secondo luogo, permette di valutare, nell’ottica del criterio della funzionalità e dell’accettabilità sociale (cfr. capitolo 2), quali di essi possano essere considerati parte integrante della nuova varietà standard e, dunque, inseriti nell’insegnamento, e quali invece vadano solo menzionati come varianti substandard attualmente diffuse, sulla cui futura evoluzione è possibile solo fare delle ipotesi. L’unica strada che il linguista può ragionevolmente percorrere è il monitoraggio degli usi documentati e della loro ricezione sociale.

Se si applicasse esclusivamente il criterio della rilevanza statistica, ciò porterebbe a considerare come parte integrante dello standard qualunque fenomeno linguistico avente un’ampia diffusione. Tuttavia, la frequenza statistica indica solo su quali fenomeni grammaticali sia più

rilevante per i linguisti porre l'attenzione come segnali di un possibile cambiamento in atto o tendenziale della lingua, ma ciò non implica affatto che il loro uso sia socialmente già accettato, accettabile o auspicabile nei registri coperti dallo standard.

Inoltre, basarsi sulla diffusione statistica permette di osservare la questione solo dal punto di vista dei parlanti, che usano ampiamente determinate strutture linguistiche perché le hanno riscontrate in precedenza, ma in questo modo si traslascia il punto di vista del linguista e dell'insegnante di lingue, ossia: queste strutture linguistiche danno un effettivo apporto funzionale alla lingua o sono solo delle alternative substandard a strutture già presenti nello standard e perfettamente funzionanti? Il loro uso evita un abbassamento del registro linguistico e, di conseguenza, evita che i parlanti vengano tacciati di trascuratezza formale o scarsa conoscenza della lingua?

In aggiunta ai criteri sopra menzionati, si è verificato quali fossero i fenomeni linguistici del neostandard trattati dalla manualistica per stranieri attualmente in commercio. Il punto di partenza è stata la relazione a convegno "Argomenti grammaticali nei manuali d'italiano L2 al livello C1/C2", tenuta da Ślapek (2016), che, pur non trattando in maniera esplicita dei tratti del neostandard, di fatto è risultata indirettamente utile alla presente ricerca. In questa relazione, oltre a trattare il tema che si evince dal titolo, Ślapek ha fatto riferimento in misura minore anche ai contenuti grammaticali inseriti nei livelli precedenti al C1/C2, fornendo così una panoramica degli argomenti grammaticali presenti nella manualistica in commercio al momento della presentazione della relazione. La disponibilità di Ślapek alla consultazione dei dati da lui raccolti sotto forma di tabelle con elenchi di argomenti grammaticali suddivisi per livello ha permesso di osservare se tra loro fossero presenti anche i tratti linguistici identificati da Sabatini come tipici del neostandard.

Su 32 pubblicazioni delle maggiori case editrici del settore della didattica dell'italiano per stranieri, quali AlmaEdizioni, CDL Edizioni, Edilingua, Eli, Guerra Edizioni, Hoepli, Le Monnier e Loescher, solo 2 affrontano alcuni tratti dell'italiano dell'uso medio: *Nuovo Magari C1-C2* della casa editrice Alma Edizioni, che tratta la frase scissa nel volume dedicato al livello C1, e *Affresco italiano C2* della casa editrice Le Monnier, che affronta la dislocazione a sinistra e a destra, la frase scissa, l'uso del *c'è* presentativo, l'uso di *lui*, *lei* e *loro* in funzione di

soggetto, l'uso di *gli* con i valori di *a loro* e *a lei*, l'uso di *cosa* come pronomi interrogativo in concorrenza con *che* e *che cosa* (dati tratti da Šlapek, 2016).

Per ottenere dati più attuali, ho ripreso questa ricerca consultando i manuali pubblicati dalle stesse case editrici sopra menzionate nell'arco di tempo successivo alla ricerca di Šlapek, ossia tra il 2018 e il 2022, selezionando dai cataloghi solo testi per adulti/giovani adulti e non per adolescenti e bambini. Ciò che emerge in generale è che nel corso del tempo le case editrici sembrano aver raccolto la sfida di introdurre maggiormente il neostandard. Si citano a titolo di esempio i seguenti manuali. *Al dente 1 A1* (2019, p. 232) edito da CDL fa riferimento all'uso del pronome diretto per dare enfasi (*Lo preparo io il tiramisù! Le compri tu le olive?*), senza però esplicitare che si tratti di dislocazione. Una soluzione analoga è adottata in *Nuovo Caffè Italia A1* (2020) edito da ELI, ove viene presentata indirettamente la dislocazione a sinistra attraverso alcuni esempi con il verbo *fare* sul tema della cucina (*Le tagliatelle le faccio io!*). Il tema della dislocazione viene ripreso esplicitamente nel volume *Al dente 4* di livello B2, in cui nell'unità 2 si fa riferimento alla dislocazione a sinistra. In *Via del corso B2* (2021, p. 236) edito da Edilingua viene brevemente presentata la costruzione "se + indicativo imperfetto + indicativo imperfetto" (*Se potevo, venivo anch'io con voi*) specificando che "Nella lingua parlata è molto comune" (*ibidem*). Si tratta, tuttavia, di una breve nota e non di un'intera unità di riflessione sulle differenze tra standard e neostandard. *Nuovo Espresso C2* (Alma Edizioni, 2019) affronta esplicitamente il tema "il pronomi affettivo (dativo etico)". Merita, infine, particolare attenzione *Bellissimo! B2/C1* (2018) edito da ELI, che nell'unità 1 presenta i seguenti contenuti: ordine marcato della frase: frasi scisse (nella variante retta da "che", ivi, p. 21); e dislocazioni: "la dislocazione a sinistra del verbo si usa per mettere in evidenza un fatto nuovo [...], la dislocazione a destra del verbo si usa per anticipare un fatto già conosciuto dai parlanti" (p. 22). Seguono vari esempi sia di dislocazione, sia di scissa retta da "che", con un buon approfondimento sul confronto tra ordine non marcato SVO e ordine marcato e, infine, l'unità 4 dedicata proprio ai fenomeni dell'italiano neostandard.

Al di fuori di questo elenco si segnala, inoltre, il manuale di lingua italiana per stranieri di Zamora, Alessandro e Hernández González

(2010), realizzato con l'obiettivo di far fronte alla scarsità di materiali didattici che includano i tratti dell'italiano colloquiale e di cui si dà una descrizione nello studio di Zamora e Alessandro (2011)⁷. Il manuale ha il pregio di offrire un punto di vista innovativo rispetto allo standard linguistico proposto e la sua consultazione è risultata utile sul piano metodologico. Tuttavia, l'opera si focalizza in maniera specifica sul registro colloquiale, che come descritto da Berruto (1987, pp. 139-140), costituisce il nucleo principale dell'italiano substandard e condivide alcune caratteristiche del neostandard, ma non coincide con quest'ultimo, dato che solo alcune sue forme e strutture meno marcate in termini di informalità sono passate e continuano a entrare a far parte dell'italiano dell'uso medio. Dunque, pur apprezzandone l'impostazione e gli obiettivi, è stato utilizzato solo come materiale di consultazione ed è escluso dall'analisi della manualistica a cui si è fatto precedentemente riferimento, perché il suo focus differisce da quello della presente ricerca, incentrata sul neostandard.

Dopo aver verificato nei manuali sopra menzionati quali fenomeni grammaticali della lista di Sabatini (1990) venissero affrontati e in quale modo (breve nota a piè di pagina, paragrafo dedicato o intera unità didattica), per ciascuno dei fenomeni esclusi si è valutato se fosse rilevante produrre *ad hoc* dei materiali didattici, sulla base dei criteri precedentemente esposti in questo capitolo e nel capitolo 2.

Inoltre, prima di procedere alla produzione di materiali didattici è necessario tenere conto anche di molti altri fattori legati alla glottodidattica, ad esempio i processi cerebrali alla base dell'acquisizione linguistica (la bimodalità emisferica, il cosiddetto "ordine naturale" e le sequenze di acquisizione, cfr. capitolo 4), da cui non si può prescindere nella presentazione dei contenuti linguistici agli apprendenti.

Ad esempio, se nell'ambito di questa ricerca si prendesse in considerazione il solo criterio della frequenza statistica per selezionare i feno-

⁷ In questo studio si pone l'accento più volte sul fatto che esistano pochi materiali mirati all'insegnamento e apprendimento colloquiale in classe: "Si miramos a la didáctica del italiano como lengua extranjera (I/LE) o lengua segunda (I/L2), observamos que son escasos los materiales que reflejan esta pluralidad y, especialmente, los que se plantean y abordan de forma sistemática la enseñanza y el aprendizaje del coloquial en el aula" (versione online, priva di numeri di pagina).

meni grammaticali a cui dedicare le unità didattiche, emergerebbe che la frase scissa retta da *che* (esempio: *è lui che l'ha detto*) è il secondo fenomeno più frequente nella consultazione del corpus. Nel caso in cui si volesse progettare dei materiali didattici su questo tema, sarebbe indispensabile tenere conto del fatto che l'uso di questa tipologia della frase scissa presuppone delle conoscenze di morfosintassi (pronomi relativo *che*, costruzione della frase, ordine marcato e non marcato dei costituenti ecc.) che non possono essere date per scontate prima di introdurre l'argomento.

Pertanto, esaminando i fenomeni ad alta e media frequenza nel corpus, si è fatta strada l'idea di creare due unità didattiche che potessero collocarsi a cavallo tra più livelli del Quadro Comune Europeo di Riferimento per la conoscenza delle lingue (QCER) (Consiglio d'Europa 2018, 2002) e che permettessero di coprirne delle grandi aree tematiche.

Per queste ragioni, sono stati selezionati l'uso dell'indicativo al posto del congiuntivo (che include sottofenomeni ad alta, media e bassa frequenza nel corpus e che consente di mostrare un fenomeno linguistico su cui incide particolarmente il criterio dell'accettabilità sociale) e l'uso del verbo in forma pronominale riflessiva in senso affettivo (es. io *mi* mangio un panino; fenomeno a media frequenza nel corpus, che risponde sia al criterio della funzionalità che dell'accettabilità sociale). Dal punto di vista glottodidattico, in entrambi i casi è stata valutata anche la potenziale produttività dell'argomento ai fini di un'unità didattica: circa quest'ultimo punto, come illustrato dai principali sillabi di riferimento della lingua italiana (cfr. capitolo 6, sez. 6.2), la trattazione dell'uso standard dei pronomi (e più nello specifico dei pronomi riflessivi) e dell'uso del congiuntivo si dispiega su più livelli del QCE. Ne consegue che gli argomenti vengano affrontati più volte nel percorso di apprendimento, sia nella fase di primo approccio con il tema, sia negli approfondimenti dei livelli successivi e che, quindi, dei temi così articolati possano rappresentare un oggetto di interesse, nonché una fonte di dubbi sull'uso della lingua, su più livelli, al progredire della competenza linguistica. Si ritiene, pertanto, che la produzione di materiali didattici su questi temi abbia una maggiore rilevanza rispetto alla produzione relativa ad un singolo argomento specifico collocato in un solo livello.

A questo proposito, ad esempio, Zamora e Alessandro (2011) in riferimento al già citato manuale sull'italiano colloquiale affermano che l'insegnamento/apprendimento del colloquiale debba essere svolto in modo graduale ma progressivo, unito allo studio dei tratti dello standard e del neostandard e legato alle situazioni o ai contesti comunicativi che li richiedono⁸.

Inoltre, si è considerato in prospettiva futura che proporre argomenti distribuibili su più livelli linguistici avrebbe permesso alle unità didattiche prodotte di integrarsi più agevolmente nella programmazione didattica dei dipartimenti coinvolti nella sperimentazione, a differenza di altri argomenti (è il caso, ad esempio, della frase scissa e delle dislocazioni) che, come già accennato, necessitano di conoscenze preliminari di sintassi e ordine dei costituenti che sarebbero state difficilmente appurabili prima della sperimentazione e del contatto con i gruppi-classe (contrariamente a quanto avverrebbe se per la sperimentazione delle unità si avesse a disposizione un intero corso di lingua). Altri fenomeni ad alta/media frequenza nel corpus, quali l'uso di *alloral comunque/ma* a inizio frase e l'uso dell'interrogativo *cosa* al posto di *che cosa*, sono stati scartati perché giudicati poco produttivi dal punto di vista della realizzazione di un'intera unità didattica.

Per queste ragioni sono stati selezionati i tratti linguistici sopra menzionati. La descrizione della struttura delle unità didattiche prodotte e degli esiti della loro sperimentazione sarà oggetto del capitolo 6.

⁸ “Desde nuestro punto de vista, estimamos que la enseñanza-aprendizaje del coloquial debe ser llevada a cabo de forma paulatina pero progresiva, combinada con el estudio de los rasgos del estándar y neostándard y vinculada con las situaciones o contextos comunicativos en las que se requieran y se exijan” (versione online, priva di numeri di pagina).

6. DESCRIZIONE E SPERIMENTAZIONE DELLE UNITÀ DIDATTICHE

Nei capitoli precedenti è stato illustrato l'approccio metodologico adottato per l'identificazione *corpus-based* dei fenomeni linguistici e per la scelta sia dei testi da includere nelle unità didattiche, sia della piattaforma di *e-learning* sulla quale collocare le attività (capitolo 4).

Sulla base di quanto precedentemente esposto, in questo capitolo verranno brevemente richiamati i due fenomeni linguistici selezionati per la progettazione delle unità, ossia l'uso del verbo in forma pronominale riflessiva in senso affettivo e l'uso dell'indicativo vs congiuntivo (cfr. capitolo 2).

Successivamente, verranno descritte le attività inserite nelle unità didattiche (sez. 6.2) e i loro esiti nelle sperimentazioni svolte presso i dipartimenti di italianistica di alcuni atenei europei (sez. 6.3).

Infine, verranno presentati gli esiti del questionario per la raccolta delle opinioni degli studenti sulle attività svolte (sez. 6.4).

6.1. I tratti linguistici selezionati

Come esposto nel capitolo 5, nella selezione dei fenomeni grammaticali sui quali incentrare le unità didattiche si è tenuto conto non solo della loro frequenza all'interno del corpus, ma anche della rilevanza sul piano glottodidattico, ossia per quali tratti linguistici potesse essere rilevante e utile creare un approfondimento sull'uso neostandard nell'ambito della didattica dell'italiano per stranieri.

Pertanto, i tratti linguistici selezionati attraverso queste valutazioni sono l'uso del verbo in forma pronominale riflessiva in senso affettivo (es. *io mi mangio un panino*), a cui in seguito per brevità si farà riferimento come verbi riflessivi “affettivi” e l'impiego dell'indicativo al posto del congiuntivo, la cui descrizione è stata affrontata nel capitolo 2 e di cui ora si riprenderanno solo alcuni punti salienti.

Per quanto riguarda il primo tratto, la “costruzione riflessiva apparente o di affetto” (Sabatini, 1985, p. 167) è utilizzata per indicare la partecipazione o il coinvolgimento emotivo di una persona rispetto a un'azione o a una circostanza indicata dal predicato. Il suo uso è frequente con i verbi *mangiare* e *bere* (e loro sinonimi) e con altri verbi che esprimono azioni o atteggiamenti che hanno effetti sulla persona del soggetto, anche con riferimento a parti del corpo o a indumenti che riguardano la sua persona. Si citano alcuni esempi tratti da Sabatini (*ibidem*) e Cignetti, 2010: *Luca si è mangiato mezza torta; verso le 11 mi bevo un caffè; Valerio si vede un film in televisione; Aldo si gode la vacanza; stamattina mi sono fatto una splendida passeggiata; farsi quattro risate* e casi simili.

La costruzione “riflessiva affettiva” è composta, dunque, da un verbo a cui si accompagna un pronome atono che “serve a enfaticizzare la partecipazione emotiva del parlante, come effetto del riferimento deittico al soggetto enunciativo” (Cignetti, 2010).

Per quanto riguarda, invece, l'alternanza del congiuntivo con l'indicativo, Sabatini (1985, 1990) la segnala nei seguenti cinque casi: dopo i verbi di opinione (*credo che hai ragione*), dopo una dichiarativa negativa (*non credo che hai torto*); in un'interrogativa indiretta (*gli chiesi se poteva aiutarmi*); nel periodo ipotetico dell'irrealtà (*se me lo dicevi, ci pensavo io*); nelle relative restrittive (*sei l'unico che parla arabo*).

Come già anticipato nel capitolo 2, questa ricerca si riallaccia a quanto espresso da Prandi (2002), secondo il quale il congiuntivo ha un suo valore

solo nei casi in cui il suo uso si basa su una scelta libera del parlante, che può alternarlo all'indicativo per cambiare il registro stilistico della frase. Dunque, l'opposizione indicativo-congiuntivo non andrebbe più ricondotta alla sola opposizione tra codifica della realtà e dell'irrealtà, secondo la posizione tradizionale (si veda la già citata interpretazione semantica, con l'indicativo come "modo della certezza" e congiuntivo come "modo dell'incertezza" cfr. Schneider, 1999), bensì bisognerebbe tenere conto anche dei casi in cui la sua scelta è un'opzione del parlante e, di conseguenza, il suo valore si esprime nel contenuto del messaggio (cfr. i già citati Lep-schy e Lepschy, 1988, p. 204 e Renzi, 2012, pp. 52-53, secondo i quali la scelta tra congiuntivo e indicativo è di tipo stilistico).

Dopo aver richiamato brevemente i tratti salienti dei fenomeni grammaticali sui quali sono state incentrate le unità, nella sezione seguente si illustrerà la struttura delle unità didattiche prodotte e gli esiti della loro sperimentazione presso il Dipartimento di Italianistica dell'Università di Breslavia (Polonia) dal 27 maggio al 31 luglio 2019 e presso i seguenti dipartimenti spagnoli: Dipartimento di Studi Romanzi, Francesi, Italiani e Traduzione dell'Università Complutense di Madrid dal 13 al 30 settembre 2019; Dipartimento di Filologia Francese, Romanza, Italiana e Araba dell'Università di Murcia per tutto il mese di ottobre 2019 e, infine, il Dipartimento di Traduzione e Interpretazione dell'Università di Malaga dal 13 al 24 gennaio 2020¹.

6.2. Descrizione delle unità

Di seguito verranno descritte le due unità didattiche, per i cui fondamenti teorici si rinvia al capitolo 4. Il punto di partenza nella progettazione è stata l'individuazione all'interno del Quadro Comune Europeo di Riferimento per la conoscenza delle lingue (QCER) del livello linguistico rispetto al quale progettare le unità didattiche, con i relativi obiettivi comunicativi e pragmatici. Il QCER è un sistema descrittivo elaborato dal Consiglio d'Europa (2018, 2002) nell'ambito del progetto *Language Learning for European Citizenship*, adottato dal 2001 sia per valutare, ed

¹ L'ultima fase di sperimentazione, prevista per il periodo di maggio-giugno 2020 presso l'Università della Slesia (Sosnoviec, Polonia), è stata cancellata in seguito alle misure di contenimento dell'epidemia di Covid-19.

eventualmente certificare, le abilità conseguite dagli apprendenti di una lingua, sia per programmare il livello di un insegnamento linguistico. Il QCER identifica tre ampie fasce di competenza a loro volta suddivise in due livelli ciascuna, per un totale di sei livelli complessivi:

A – livello base

- A1 – Livello di Contatto o Livello Base
- A2 – Livello di Sopravvivenza o Livello Elementare

B – livello autonomia

- B1 – Livello Soglia o Livello Intermedio Basso
- B2 – Livello di Progresso o Livello Intermedio Alto

C – livello padronanza

- C1 – Livello di Efficacia o Livello Avanzato
- C2 – Livello di Padronanza o Livello Competenza

Il QCER ha lo scopo di definire la competenza linguistico-comunicativa per i diversi livelli in riferimento alle abilità di ricezione e di produzione scritte e orali e alla conoscenza delle strutture linguistiche, del lessico, dei tratti della testualità, della pragmatica e della sociolinguistica. Tuttavia, esso non fornisce indicazioni sulle strutture grammaticali richieste per i diversi livelli. Tali informazioni possono, invece, essere ricavate consultando i sillabi di riferimento della lingua italiana, ossia dei testi che propongono degli elenchi delle strutture grammaticali richieste per ciascun livello di apprendimento linguistico del QCER. I sillabi sono compilati principalmente da enti certificatori e centri di formazione linguistica e sono “parte dell’attività curricolare che si riferisce alla specificazione e alla sequenziazione dei contenuti di insegnamento fatta in termini di conoscenze e/o capacità” (Ciliberti, 1994, pp. 99-100). Nello specifico, il sillabo di certificazione “elenca le competenze che il candidato deve possedere come acquisite, e saper utilizzare in modo produttivo, per poter essere considerato a un certo livello di competenza” (Arcangeli *et al.*, 2014, p. 14).

Pertanto, al fine di identificare innanzitutto in quali livelli vengano collocati l’uso standard dei verbi riflessivi e del congiuntivo e, in secondo luogo, se vi siano delle indicazioni sull’uso dei verbi riflessivi in forma “affettiva” e sulla scelta tra indicativo e congiuntivo in contesti informali, sono stati consultati sillabi mirati alla programmazione di un corso di lingua o alla certificazione per la lingua italiana, di seguito

elencati seguendo l'ordine della loro data di pubblicazione: Minciarelli, Comodi (2005), Lo Duca (2006), Barni *et al.* (2009), Benucci (2007, a c. di), Spinelli, Parizzi (2010, a c. di), Arcangeli *et al.* (2014), AIL: Accademia Italiana di Lingua (2015) e infine Cardillo, Vecchio (2015, 2016, 2017, 2018, a c. di).

La tabella 6.1 permette di avere un quadro generale dei livelli del QCER nei quali gli autori dei sillabi consultati inseriscono i tratti grammaticali in oggetto. Sotto ciascun livello sono riportate delle abbreviazioni corrispondenti ai vari sillabi, seguite da un numero tra parentesi, corrispondente alla ricorrenza della trattazione di tale argomento: dalla sua prima introduzione (1) ai successivi approfondimenti (esempi: (2) e (3)). Segue una legenda delle abbreviazioni per ciascun sillabo: **MC**: Minciarelli, Comodi (2005); **LD**: Lo Duca (2006); **Ba**: Barni *et al.* (2009); **Be**: Benucci (2007 a c. di); **SP**: Spinelli, Parizzi (2010, a c. di); **ADA**: Arcangeli *et al.* (2014), **AIL**: Accademia Italiana di Lingua (2015) e infine **PLIDA**: Cardillo, Vecchio (2015a, 2015b, 2016, 2017, 2018, a c. di).

Tabella 6.1: Livelli del QCER in cui gli autori dei sillabi inseriscono i tratti grammaticali in oggetto

	A1	A2	B1	B2	C1	C2
Pronomi riflessivi (uso standard)	SP (1) Be (1) ADA (1) AIL (1)	MC (1) LD (1) Be (2) SP (2) ADA (2) AIL (2)	MC (2) Ba (1) Be (3) ADA (3)	SP (2) Ba (2) PLIDA (1)		
Pronomi riflessivi "affettivi"			MC (1) Be (1)	LD (2)		
Congiuntivo (uso standard)		MC (1)	MC (2) LD (1) SP (1) Be (1) PLIDA (1) ADA (1) AIL (1)	LD (2) SP (2) Ba (1) Be (2) PLIDA (2) ADA (2) AIL (2)		
Indicativo vs congiuntivo nell'informalità			LD (1) SP (1)		MC (1) Be (1)	

Dalla tabella 6.1 emerge che 4 sillabi su 8 suggeriscono una prima introduzione nel livello A1 del tema dei pronomi riflessivi, che viene sviluppato appieno nel livello A2, con successivi approfondimenti nei livelli B1-B2. Le uniche eccezioni sono i casi di Barni (2009) e del PLI-DA, che menzionano i riflessivi rispettivamente solo al livello B1 e B2. Per quanto riguarda i pronomi riflessivi “affettivi”, vengono menzionati solo in tre sillabi: Minciarelli, Comodi (2005) e Benucci (2007, a c. di) inseriscono questo argomento al livello B1, mentre Lo Duca (2006) lo inserisce al livello B2, come prosecuzione dell’uso standard dei riflessivi trattato al livello A2.

Per quanto riguarda il congiuntivo, ad eccezione di Minciarelli e Comodi (2005), che propongono una prima introduzione delle strutture del congiuntivo nel livello A2, gli altri sillabi propongono l’introduzione delle forme del presente e del passato nel livello B1, per poi proseguire nel livello B2 con il congiuntivo imperfetto e trapassato, la concordanza dei tempi e il periodo ipotetico. In alcuni casi, tali ulteriori sviluppi vengono inseriti nei livelli C1 o C2; tuttavia, in questa sede sono state omesse queste informazioni dettagliate, poiché lo scopo di tale confronto è determinare il livello in cui le strutture vengono introdotte per la prima volta.

Per quanto riguarda, invece, la scelta dell’indicativo al posto del congiuntivo in contesti informali, il tema viene proposto soltanto da quattro sillabi, che si dividono in maniera equa tra l’introduzione di questo argomento al livello B1 (Lo Duca, 2006; Spinelli, Parizzi, 2010) e al livello C1 (Minciarelli, Comodi, 2005; Benucci, 2007, a c. di).

Pertanto, sulla base di tale confronto, le unità sono state progettate per essere adatte globalmente al livello B1. Nel caso del congiuntivo la trattazione è stata limitata ai tempi del presente e del passato, introducendo l’uso neostandard dell’indicativo al posto del congiuntivo come integrazione delle norme dell’uso standard. Invece, nel caso dei verbi riflessivi “affettivi” sono stati inseriti alcuni collegamenti con il livello A2, proponendo attività in cui si richiama la coniugazione dei verbi e la posizione del pronome clitico nella frase, che possono fungere da primo approccio per il livello A2 o da rinforzo o recupero per il livello B1. Sarà dunque possibile scegliere se svolgere tutte le attività proposte o se creare dei percorsi personalizzati adattabili al livello della classe. Anche i contenuti tematici delle unità sono stati scelti tenendo conto delle indicazioni per i livelli B1-B2 del QCER e dei vari sillabi.

Per facilitare i processi cognitivi degli apprendenti, è necessario elaborare delle tecniche che tengano conto dei meccanismi alla base dell'apprendimento delle lingue (cfr. Daloiso, 2009; 2015) e, che in tal modo, favoriscano l'acquisizione linguistica. A questo proposito, si cita qui di seguito il percorso proposto da Balboni (2002, pp. 137-138):

- a) *individuazione*: nei testi che costituiscono input gli allievi sottolineano, cerchiano, ricopiano ecc. le «regole» che costituiscono l'obiettivo di quell'unità di apprendimento;
- b) *creazione di ipotesi* sul funzionamento di quel dato meccanismo: ipotesi suggerite dal docente o guidate da uno schema del manuale e discusse eventualmente con i compagni;
- c) *fissazione*, sia attraverso attività di natura comportamentistica, quindi abbastanza ripetitive, sia per mezzo di giochi che, pur condividendo la natura del *pattern drills*, non ne abbiano le debolezze;
- d) *reimpiego* via via più creativo, dapprima con esercizi di applicazione e poi in attività comunicative;
- e) *riflessione* esplicita guidata dal docente: e mirata alla creazione di una rappresentazione mentale e della sua eventuale «trasposizione» in uno schema grafico, una tabella ecc.

Queste indicazioni sono state seguite nella progettazione delle unità didattiche, secondo lo schema già citato nella sezione 4.8.3 del capitolo 4. Di seguito verranno descritte entrambe le unità; per una loro visione completa, si rinvia a Zingaro (2021, pp. 309-369).

6.2.1. Unità didattica sui verbi riflessivi "affettivi"

Come anticipato nella sezione precedente, le attività sono state progettate per poter essere adatte sia ad un primo approccio con il tema grammaticale dell'uso dei riflessivi, nel caso del livello A2 - B1, sia come attività di ripasso per i livelli superiori. I testi utilizzati, sotto forma di interi brani o di singole frasi, sono tratti dal corpus PAISÀ, mentre le immagini utilizzate come illustrazione delle attività sono state scaricate da siti che mettono a disposizione immagini prive di copyright, gratuite e liberamente utilizzabili anche a fini commerciali². Richiamando le basi teoriche descritte nel capitolo 4, le unità sono

² Cfr. i siti Unsplash (<https://unsplash.com/license>), Burst (<https://burst.shopify.com/legal/terms>), Gratisography (<https://gratisography.com/license>).

state strutturate secondo le seguenti fasi di apprendimento, che permettono agli apprendenti di approcciarsi con le strutture linguistiche in maniera induttiva: motivazione, globalità, analisi, sintesi, riflessione e verifica.

Come già anticipato nel capitolo 4, sono state collocate sulla piattaforma *Wizer.me*, che consente un ampio margine di personalizzazione nella progettazione delle attività, inclusa l'attribuzione dei punteggi. A questo proposito, è stato stabilito di attribuire un punteggio pari ad 1 per ciascun *item* corretto, 0,5 per ciascun *item* corretto dal punto di vista del contenuto, ma contenente errori di ortografia (esempio: *andasi* vs *andassi*) e 0 per ciascun *item* del tutto scorretto (esempio: *andrei* vs *andassi*). Alla base di questa scelta c'è l'intento di dare preminenza alla competenza comunicativa e pragmatica rispetto alla pura correttezza formale: riprendendo l'esempio di cui sopra, se alla richiesta di usare un congiuntivo l'apprendente inserisce la risposta *andasi*, a tale risposta viene attribuito mezzo punto rispetto al totale, con l'intento di valorizzare comunque la capacità dell'apprendente di riconoscere la struttura grammaticale utile in quel contesto.

Motivazione

La prima pagina, intitolata *Mi preparo per uscire*, ha l'obiettivo di avvicinare gli apprendenti al tema dell'unità: l'immagine di una donna con l'accappatoio e l'asciugacapelli anticipa il tema che verrà trattato poi nel brano da leggere. Vengono presentati gli obiettivi generali (ossia l'arricchimento del lessico, il ripasso dei pronomi riflessivi e un approfondimento dell'uso dei riflessivi in senso intensivo-affettivo nell'italiano neostandard) e viene proposta un'attività iniziale per "rompere il ghiaccio", incentrata sull'uso "non affettivo" dei verbi riflessivi, in cui si richiede di collegare alcune immagini che raffigurano le azioni della vita quotidiana legate alla cura del corpo ai verbi che le descrivono (*radersi/farsi la barba, alzarsi, lavarsi, vestirsi, pettinarsi, truccarsi*). Trattandosi semplicemente di un'attività introduttiva, svolta oralmente *in plenum* e a cui è stata attribuita la correzione automatica, i risultati non sono stati inseriti in questa trattazione.

Globalità e analisi

La seconda pagina è intitolata *Che cosa mi metto?* e corrisponde alle fasi di globalità e analisi dell'unità didattica. Nella fase della globalità si propongono attività di comprensione testuale. Anche in questo

caso l'immagine di un armadio con dei vestiti richiama direttamente il tema del brano che gli apprendenti andranno a leggere, che si intitola *L'appuntamento*³. È tratto da un blog in cui una donna descrive tutte le azioni compiute prima di presentarsi ad un primo appuntamento. Si richiede agli apprendenti di leggere il testo per avere un'idea generale del tema, poi di svolgere un esercizio lessicale, a cui sono stati attribuiti 6 punti sul totale di 12 dell'intera pagina, in cui vengono forniti due elenchi di parole in colonna e viene chiesto di collegare alcune parole tratte dal brano, poste nella colonna di sinistra, con i relativi sinonimi nella colonna di destra. All'esercizio è stata attribuita la correzione automatica, impostando le soluzioni nella fase di progettazione dell'unità. Successivamente viene chiesto di rispondere a 3 domande aperte di comprensione del tema del brano, a ciascuna delle quali è stato attribuito un punto su 12.

L'ultima parte della pagina corrisponde alla fase di analisi, in cui si richiede di inserire in una tabella tutte le parole del brano che riguardano abbigliamento e accessori, prodotti per il corpo e i verbi che descrivono la cura del corpo, per un totale di 3 punti su 12 (cfr. Figura 6.1): quest'ultima attività crea le basi per l'approfondimento sui verbi riflessivi nella pagina successiva, seguendo un percorso di presentazione induttiva dei contenuti grammaticali come illustrato nel capitolo 4.

Figura 6.1: Completamento della tabella

Analisi
Trova nel testo tutte le parole che riguardano:

ABBIGLIAMENTO e ACCESSORI	PRODOTTI PER IL CORPO	VERBI che descrivono la cura del corpo

Riflessione

In questa fase viene chiesto agli apprendenti di riflettere sulle forme verbali riflessive finora incontrate. Viene innanzitutto proposto un esercizio di completamento della coniugazione di 3 verbi riflessivi, uno per ciascuna coniugazione del sistema verbale italiano (pari a 12 pun-

³ Adattamento del brano tratto da: <https://notimetolose.myblog.it/?s=appuntamento&submit=Cerca>.

ti). Trattandosi di un'unità destinata al livello B1, non è stata proposta la distinzione tra le due forme della terza coniugazione (*-ire*; *-isc-*), essendo questo un tema generalmente affrontato nei precedenti livelli di competenza linguistica del QCER.

Segue un'attività di ripasso delle regole sulla posizione dei pronomi diretti e indiretti all'interno della frase (pari a 7 punti), in cui, tramite una lista di domande a scelta multipla, viene chiesto agli apprendenti di esplicitare perché il pronome si trovi in una certa posizione all'interno della frase (cfr. Figura 6.2).

Infine, viene chiesto agli apprendenti di leggere alcune frasi contenenti dei verbi riflessivi (cfr. Figura 6.3), e di collocarle in una tabella che racchiude quattro usi possibili della forma riflessiva, tra cui il verbo riflessivo "affettivo" (in questo caso *Mi fumo una sigaretta*), usato colloquialmente per sottolineare la partecipazione psicologica di chi compie l'azione (attività pari a 4 punti).

Figura 6.2: Ripasso delle regole sulla posizione dei pronomi diretti e indiretti

La posizione dei pronomi

1. Nelle seguenti frasi ci sono degli esempi delle regole grammaticali sulla posizione dei pronomi. Scegli a quale regola corrispondono le frasi.

Esempio: cosa mi metto? Il pronome si trova prima del verbo

È questa la pietra che vuoi? L'ho trovata nella foresta. Prendila, è tua!

a) Prima del verbo b) Unito al verbo all'imperativo c) Unito al verbo all'infinito d) Unito al verbo al gerundio

e) Unito al verbo al participio.

Figura 6.3: Completamento della tabella sugli usi della forma riflessiva

Inserisci nella tabella gli esempi 1-6 che trovi qui di seguito.
NB: Ogni esempio appartiene ad una sola delle 4 categorie (a, b, c, d).

1. Mi metto la gonna;
2. Mi vesto;
3. Ci incontriamo;
4. Mi asciugo i capelli;
5. Mi fumo una sigaretta;
6. Mi trucco.

a) Azione reciproca tra più soggetti	b) azione su parti del corpo o vestiti/accessori che appartengono al soggetto	c) azione del soggetto su se stesso	d) il verbo non è un vero verbo riflessivo. Il pronome riflessivo viene usato colloquialmente solo per sottolineare la partecipazione psicologica di chi compie l'azione

Sintesi

Questa fase dell'unità include due pagine intitolate *Applicazione pratica 1* e *2*, dedicate al reimpiego attivo da parte degli apprendenti delle strutture grammaticali su cui si focalizza l'unità. Nell'*Applicazione pratica 1* vengono proposti due esercizi, per un totale di 19 punti, nei quali gli apprendenti devono completare delle frasi scegliendo tra i verbi riflessivi "affettivi" dati nella consegna. Nell'*Applicazione pratica 2* sono presenti due esercizi: nel primo si richiede di completare delle frasi con le voci del verbo *fare* alla forma riflessiva (10 punti), mentre nel secondo si richiede di coniugare al passato prossimo il verbo riflessivo all'infinito dato tra parentesi (14 punti).

Figura 6.4: Esempio di attività dell'Applicazione pratica 1



Figura 6.5: Applicazione pratica 2, primo esercizio

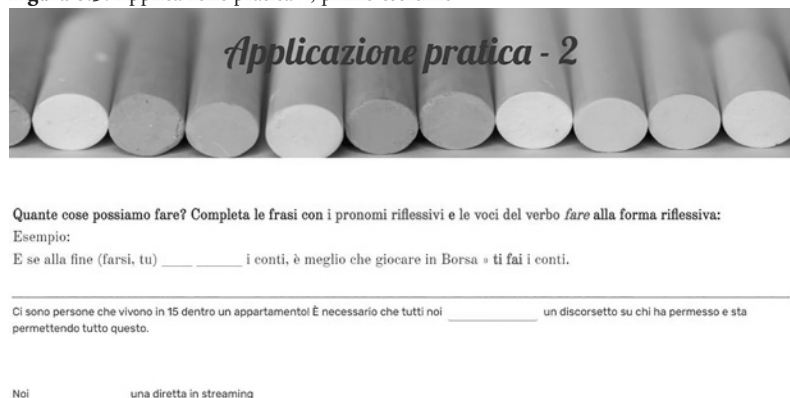


Figura 6.6: Applicazione pratica 2, secondo esercizio

Coniuga al passato prossimo il verbo riflessivo all'infinito che trovi tra parentesi.

NB: l'ausiliare dei verbi riflessivi è sempre *essere*.

TEMPO LIBERO

Esempio: Da poco (farsi, lui) _____ una vacanza in una beauty farm

>> Da poco **si è fatto** una vacanza in una beauty farm.

Lucia: "Ieri sera (guardarsi, io) _____ la registrazione del webcast "NET ed il test del software".

Tra il calore delle quattro mura di Giulio, le tre amiche (farsi) _____ i pop-corn e Giulio si è rilassata.

Marco: "Ieri sera (vedersi, io) _____ la puntata di Ballarò registrata martedì."

Verifica

Sono state progettate due attività finalizzate a verificare il raggiungimento degli obiettivi generali dell'unità: conoscere la coniugazione riflessiva dei verbi, conoscere i diversi tipi di azioni che il verbo riflessivo può descrivere e, infine, saper scegliere in base al registro formale o informale della comunicazione tra l'uso della forma non riflessiva o della forma riflessiva "affettiva".

La pagina di *Applicazione pratica 3* contiene un esercizio in cui, data una frase contenente un verbo riflessivo già coniugato, gli apprendenti devono selezionare da una lista di risposte a scelta multipla la funzione che il verbo riflessivo esercita in quella frase: azione reciproca (esempio: *ci incontriamo*), azione compiuta su parti del corpo o vestiti/accessori (esempi: *mi lavo i denti*, *mi metto il cappello*), azione che il soggetto compie genericamente su se stesso (senza la specificazione di parti del corpo o vestiti/accessori, ad esempio: *mi lavo*), oppure un'azione descritta da un verbo riflessivo "affettivo" (esempio: *mi mangio un panino*), per un totale di 15 punti.

Infine, la pagina di *Applicazione pratica 4* contiene un esercizio in cui viene chiesto agli apprendenti di completare delle frasi coniugando i verbi dati all'infinito nella consegna e, una volta identificato il verbo giusto, di scegliere in base al contesto formale o informale in cui è inserita ciascuna frase, tra la forma non riflessiva o la forma riflessiva "affettiva". A tale attività è attribuito un totale di 10 punti.

Figura 6.7: Scelta multipla

Applicazione pratica -3



Qual è la funzione del verbo riflessivo? Scegli la risposta giusta:

1) Mi lavo i capelli due o tre volte a settimana.

- ☐ a azione reciproca tra più soggetti
- ☐ b azione compiuta su parti del corpo/ vestiti/accessori del soggetto
- ☐ c azione del soggetto su se stesso
- ☐ d azione descritta da un verbo riflessivo "affettivo"

Figura 6.8: Attività di completamento e scelta della forma riflessiva o non riflessiva

Applicazione pratica - 4

Completa le seguenti frasi. Scegli, in base al contesto informale o formale, se usare la forma riflessiva-affettiva o la forma non riflessiva dei verbi dati:

bere - fare - fare - fare - guardare - guardare - leggere - leggere - leggere - vedere

NB: gli spazi vuoti possono contenere più di una parola.

1) L'insegnante dice: "Se (voi) _____ un po' di storia del movimento operaio italiano, scoprite che il problema della salute nelle fabbriche è stato sollevato per la prima volta proprio dagli stessi operai."

2) In un commento scritto su Facebook: "Se cerchi una soluzione, basta che _____ un giro su internet e avrai tutte le risposte ai tuoi dubbi."

Rinforzo/recupero

Seguono delle attività facoltative dedicate all'uso dei verbi riflessivi nei modi di dire o in espressioni metaforiche e alcune attività di produzione scritta e orale, i cui risultati esulano dal tema di questa trattazione.

6.2.2. Unità didattica sul congiuntivo

Motivazione

Per la fase di motivazione è stata mostrata una vignetta ironica sul tema dei dubbi sull'uso dell'indicativo vs congiuntivo e sull'istinto di cor-

reggere gli errori degli altri, che è stata utilizzata come attività orale di rottura del ghiaccio e condivisione di impressioni ed esperienze personali. In assenza di informazioni relative ai diritti di utilizzo e diffusione dell'opera, la vignetta non verrà riprodotta in questo volume e si rinvia al relativo link al profilo dell'autore su X⁴.

Globalità - analisi

Nelle fasi di globalità e analisi vengono proposte una serie di attività che permettono di familiarizzare con le regole d'uso del congiuntivo nella varietà standard. Le attività sono state progettate per poter essere adatte sia al livello B1, per un primo approccio con il tema del congiuntivo e della coniugazione del tempo presente e passato, sia come attività di ripasso per i livelli superiori, a seconda che si scelga di svolgere tutte le attività proposte o di creare percorsi personalizzati adattabili al livello della classe.

Nella prima pagina, come avvenuto per la fase di motivazione, come attività orale di rottura del ghiaccio e condivisione di impressioni ed esperienze personali è stata mostrata una vignetta ironica sulla tendenza comune a giudicare gli altri in base all'uso corretto o scorretto del congiuntivo. Anche in questo caso, in assenza di informazioni relative ai diritti di utilizzo e diffusione dell'opera, la vignetta non verrà riprodotta in questa trattazione e si rinvia al relativo link alla pagina che ospita l'immagine⁵.

Sempre nella prima pagina vengono proposti quattro brevi testi da leggere⁶ e viene chiesto di cercare al loro interno i verbi al congiuntivo e di classificare gli esempi trovati nei testi all'interno di una tabella che racchiude 4 ambiti d'uso: congiuntivo retto da verbi di opinione, congiuntivo retto da una frase principale con una forma impersonale, congiuntivo introdotto da una congiunzione subordinante e congiun-

⁴ Astutillo Smeriglia: https://twitter.com/a_smeriglia/status/1081098811567673344.

⁵ *I love Italian language*:

<https://iloveitalianlanguage.com/2018/04/07/chi-ha-paura-del-congiuntivo/>.

⁶ Adattamento brani tratti da:

https://it.wikipedia.org/wiki/Monaco_in_riva_al_mare;

https://it.wikipedia.org/wiki/Vita_su_Marte; http://guide.supereva.it/viaggi_safari/interventi/2003/03/131025.shtml (consultato il 20/03/2019 e non più reperibile); http://guide.supereva.it/case_editrici/interventi/2005/07/218030.shtml (come sopra).

tivo nelle interrogative indirette. All'attività corrisponde un totale di 4 punti.

Nella seconda pagina vengono presi in esame tutti i casi in cui nello standard il congiuntivo è obbligatorio, chiedendo gli apprendenti di completare delle frasi, coniugando al congiuntivo i verbi dati all'infinito per un totale di 10 punti.

Figura 6.9: Esempio di lettura e completamento della tabella

Testo n.4:
La scrittura deve servire innanzitutto a se stessi, per riflettere e trovare la coscienza di sé, come la meditazione nelle filosofie orientali. A proposito del fatto che la scrittura dev'essere per prima cosa un piacere in sé, vorrei ricordare cosa mi hanno detto Natalia Aspesi: "Penso che scrivere sia un grande piacere personale che ripaga di ogni delusione".
http://guide.supereva.it/case_editorici/interventi/2005/07/218030.shtml

Quando è usato il congiuntivo? Scrivi gli esempi che hai trovato nei testi:

Con un verbo di opinione	Con una forma impersonale	Con una congiunzione subordinante	Nella frase interrogativa indiretta

Figura 6.10: Esempio attività di completamento

5. dopo alcuni connettivi

Connettivi temporali:

- Andiamo a fare la spesa **prima che** chiudano i negozi
- Rimarrò con mia madre **fino a che** non stia bene.

1. La domanda va spedita alla segreteria una settimana prima che (scadere) _____ i termini della consegna.
2. Tagliate le foglie e fatele asciugare al sole, finché non (diventare) _____ da verde scuro a gialle.

Riflessione

Nella pagina di riflessione sulle strutture grammaticali viene presentato un brano tratto da un blog con la recensione di un romanzo⁷. Nel testo vengono messe in evidenza due frasi, nelle quali il verbo principale è un verbo di opinione, seguito in un caso dal verbo coniugato al congiuntivo e nell'altro caso del verbo coniugato all'indicativo. In un'attività di completamento della regola grammaticale, a cui corrisponde un totale di 6 punti, viene chiesto agli apprendenti di fare ipotesi sulle

⁷ Adattamento brano tratto da: www.cineblog.it/post/20702/tutta-da-rifare-il-romanzo-di-giorgia-wurth-il-book-trailer/1.

subordinata interrogativa indiretta sono presenti verbi all'indicativo e successivamente viene chiesto di trasformarle per adattare ad una situazione formale, ossia di coniugare i verbi della subordinata al congiuntivo e di apportare qualunque altra modifica ritenuta necessaria (uso del *Lei* e di espressioni di cortesia, eccetera). Nell'attività successiva viene chiesto, invece, di adattare altre frasi ad un contesto informale.

Figura 6.12: Esempio di attività sull'uso del congiuntivo vs indicativo nell'interrogativa indiretta

Leggi le seguenti frasi:

- 1) Cominciamo ad andare in giro per la città. Già nei primi momenti mi chiede se io sono la giornalista straniera che è venuta a visitare la città.
- 2) Se si intende lasciare l'auto in una città o in un posto diverso da quello dove si noleggia, bisogna dirlo subito, e chiedere se sono previsti sovrapprezzi in questo caso.
- 3) La ragazza chiede se c'è qualcuno che proviene da fuori e Clem solleva la sua mano.
- 4) Mi chiede se per caso io ho letto "I Libri di Natale" di Dickens.

Ora trasforma le frasi precedenti per adattare ad una situazione formale: stai scrivendo un articolo sulla tua esperienza come studente/ssa in scambio all'estero che sarà pubblicato sul sito della tua università.

In entrambe le attività, in contrasto con l'approccio induttivo su cui si basa il percorso delle unità prodotte, si chiede esplicitamente di coniugare, a seconda dei casi, i verbi al congiuntivo o all'indicativo. Questa scelta di un approccio chiaramente deduttivo ("coniuga al...") a prima vista potrebbe destare delle perplessità. Come evidenziato da Balboni (2002, p. 139) le tecniche di manipolazione tipiche dell'approccio formalistico⁸, in cui ci si serve di consegne del tipo "Trasforma al...", mirano ad applicare meccanicamente le regole, piuttosto che a introdurre una riflessione. Tuttavia, lo stesso Balboni suggerisce che sia possibile inserire tali attività in un approccio di tipo comunicativo e induttivo attribuendo una giustificazione comunicativa alle tecniche di manipolazione: ad esempio, chiedere nella consegna di un esercizio di raccontare un

⁸ Approccio glottodidattico che ha caratterizzato l'insegnamento delle lingue straniere dal XVIII secolo agli anni '70 del XX secolo, basato su un percorso di tipo deduttivo, in cui il docente, fonte di informazioni e modello da seguire, fornisce delle regole da cui gli studenti deducono i comportamenti linguistici. Le tecniche didattiche utilizzate sono, oltre ai già citati esercizi di manipolazione ("Trasforma al..."), il dettato e la traduzione tra la lingua materna e la lingua straniera (metodo grammaticale-traduttivo).

film giustifica la trasformazione dei verbi dal presente al passato. Analogamente, nel caso specifico dell'unità didattica sul congiuntivo, lo stesso risultato può essere ottenuto modificando le frasi per adattarle ad una diversa situazione comunicativa che preveda un diverso livello di formalità (esempio: racconta le tue esperienze come studente *Erasmus* ai tuoi amici vs ora modifica quanto hai scritto per riferire le stesse esperienze e opinioni personali in un articolo del portale web dell'università).

Questa svolta verso l'approccio deduttivo in queste pagine non interrompe la progressione crescente di difficoltà dell'unità, poiché nell'attività finale, come verrà illustrato nella sezione successiva, spetta agli apprendenti stabilire se utilizzare il congiuntivo o l'indicativo rispetto al grado di formalità del contesto.

Verifica

Infine, sono state progettate due attività finalizzate a verificare il raggiungimento degli obiettivi generali dell'unità: conoscere la coniugazione del congiuntivo presente e passato, conoscere i casi in cui nella varietà standard è prescritto l'uso del congiuntivo e i casi in cui nella varietà neostandard l'uso dell'indicativo al posto del congiuntivo corrisponde ad un grado inferiore di formalità, e, infine, saper scegliere in base al registro formale o informale della comunicazione tra l'uso del congiuntivo o dell'indicativo. Alle attività finali è stato attribuito un totale di 20 punti.

Vengono proposte due attività: nella prima vengono fornite delle frasi complete con verbo coniugato al congiuntivo e per ciascuna frase viene chiesto agli apprendenti di selezionare da una lista a scelta multipla la risposta che spiega perché in quel caso specifico è obbligatorio il congiuntivo (ossia, se è presente un verbo di opinione/sentimento o una forma impersonale nella frase principale, oppure se è presente una congiunzione subordinante o una frase interrogativa indiretta). Segue una seconda attività di completamento: agli apprendenti è richiesto di completare delle frasi coniugando il verbo della frase secondaria all'indicativo o al congiuntivo a seconda del contesto formale informale in cui è inserita la frase.

Figura 6.13: Esempio della prima attività

Scegli la risposta che spiega perché, secondo lo standard, è obbligatorio il congiuntivo nelle seguenti frasi:

1) Quando facevo in moto quel percorso, mi sembrava una cosa lunga. Adesso mi sembra che non siano passati due minuti.

- ☐ a C'è un verbo di opinione/sentimento nella frase principale
- ☐ b C'è una forma impersonale nella frase principale
- ☐ c C'è una congiunzione subordinante
- ☐ d È una frase interrogativa indiretta

Figura 6.14: Esempio della seconda attività

Indicativo o congiuntivo? Scegli in base al contesto informale o formale della frase:

1) L'insegnante dice: "Mi chiedo se per caso tu (avere) _____ letto "I Libri di Natale" di Dickens."

2) Un giornalista scrive: "Liz Taylor, rimasta evidentemente affascinata da Depardieu, gli chiede se (essere) _____ sposato. L'attore risponde di sì e la Taylor commenta "Peccato!"

3) Hai fatto un incidente stradale e hai danneggiato l'auto di un'altra persona:
 "Basta, mi sono stancato! Se vuoi che io ti (pagare) _____ il danno, prendi il modulo per la denuncia!"

6.3. Esiti della sperimentazione

In questa sezione verrà svolta un'analisi quantitativa e qualitativa dei dati raccolti tramite le sperimentazioni delle unità didattiche. Le unità sono state sperimentate nel periodo compreso tra maggio 2019 e gennaio 2020 nell'ambito dei corsi di italiano di livello intermedio attivati presso gli atenei di seguito elencati: Università di Breslavia, Università Complutense di Madrid, Università di Murcia e Università di Malaga. L'attività ha coinvolto, complessivamente, un collettivo di 53 partecipanti.

Al fine di adattare le unità al livello dei diversi gruppi-classe coinvolti, lo svolgimento delle attività ha subito alcune variazioni nelle diverse sedi di sperimentazione. Pertanto, per una più rapida confrontabilità dei dati, in questa sede si analizzeranno solo gli esiti della verifica finale. Per l'analisi completa dei dati di tutte le attività che compongono entrambe le unità si rinvia a Zingaro (2021, capitolo 6).

L'analisi degli esiti delle verifiche finali è stata integrata dalle risposte ottenute tramite un questionario con cui sono state raccolte informazioni sulla popolazione coinvolta nella sperimentazione e le

opinioni sulle attività svolte (Nunan, 1989; Coonan, 2000; Pozzo, 2008).

Nella prima parte il questionario presenta domande aperte di tipo conoscitivo: L1 e altre LS conosciute oltre all'italiano; da quanto tempo è in corso l'apprendimento dell'italiano; eventuali conoscenze pregresse sul tema linguistico-grammaticale dell'unità; eventuale presenza di fenomeni linguistici analoghi nella lingua madre e/o in altre LS conosciute.

Nella seconda parte si chiede ai partecipanti di valutare su una scala numerica da 0 a 3 alcune affermazioni che descrivono le attività svolte al fine di esprimere il loro parere rispetto alla loro percezione di utilità, di acquisizione del tema grammaticale (esempio: *dopo queste attività ho un'idea più chiara dell'uso dei riflessivi*) e di soddisfazione rispetto alla piattaforma e-learning in cui è stata collocata l'unità. Seguono spazi per aggiungere apporti personali.

Prima di procedere all'analisi dei dati raccolti nell'arco di queste sperimentazioni, è necessario premettere quanto segue. Avendo a disposizione un collettivo estremamente ridotto, è bene precisare che lo scopo della sperimentazione e del successivo monitoraggio non è di ricavare conclusioni di carattere generale, valide su vasta scala, per le quali sarebbe stato necessario un collettivo più ampio, bensì di svolgere un primo test dei materiali didattici e valutarne l'efficacia in termini di utilità e contributo motivazionale nell'acquisizione linguistica.

Inoltre, quando si progetta una sperimentazione, è necessario che l'ambiente sperimentale venga tenuto in condizioni costanti, ad eccezione della variabile sperimentale, che invece viene volutamente cambiata. Nel caso della presente ricerca, sono state svolte delle lezioni di lingua italiana di livello B1, aventi tra gli obiettivi linguistici un ripasso dell'uso standard di due tratti grammaticali generalmente affrontati nel livello intermedio, all'interno del quale si inserisce, come elemento di novità, un approfondimento dell'uso neostandard di tali tratti grammaticali. L'obiettivo della sperimentazione è valutare se e in che misura i partecipanti al termine delle attività siano in grado di formulare delle scelte tra uso standard e neostandard rispetto al grado di formalità/informalità del contesto e di riconoscere l'uso neostandard da parte di parlanti nativi italofoeni come il frutto di una scelta e non di una scarsa padronanza delle regole grammaticali.

Una sperimentazione controllata prevede che i partecipanti vengano divisi in due gruppi: un gruppo sperimentale sottoposto ai cambiamenti della variabile e un gruppo di controllo che non sia esposto a quei cambiamenti, la cui presenza garantisce che i dati prodotti dal gruppo sperimentale siano realmente dovuti alla variabile che si sta verificando e non a influenze esterne sconosciute. Nel caso in questione, la prassi per ottenere dei dati sull'efficacia delle attività svolte consisterebbe nel dividere la popolazione in due gruppi: un gruppo che svolge tutte le attività della sperimentazione, incluso l'approfondimento sulla scelta tra uso standard e neostandard rispetto al grado di formalità/informalità del contesto, e un gruppo di controllo a cui non venga fornito tale approfondimento.

Tuttavia, negli accordi siglati con gli atenei esteri coinvolti nella sperimentazione è stato concesso un numero di ore strettamente necessario allo svolgimento delle due unità didattiche, ossia 1 ora e 30 per ciascuna unità, fino ad un massimo di 3 ore totali, che non consentiva di dividere le classi ed effettuare incontri differenziati con il gruppo di sperimentazione e il gruppo di controllo. Di conseguenza, è stato necessario fornire a tutti i partecipanti gli stessi contenuti. Se ne deduce che i dati raccolti non forniscono indicazioni sull'efficacia delle attività di verifica finale in assenza dello svolgimento dell'intera unità didattica.

Inoltre, è necessario specificare che il totale dei partecipanti alla sperimentazione in taluni casi varia in base alle diverse attività svolte. I fattori che hanno determinato la mancata consegna di alcune attività sono principalmente di natura logistica (ad esempio, l'arrivo in ritardo o, viceversa, l'uscita anticipata dall'aula di alcune persone rispetto all'orario della lezione), mentre solo in pochi casi isolati sono state riscontrate delle difficoltà nel salvataggio delle risposte sulla piattaforma, causati dall'instabilità della connessione internet o da incomprensioni circa la modalità di salvataggio.

Infine, è necessario precisare che prima di ciascuna sperimentazione si è svolto un confronto con i docenti referenti presso i dipartimenti ospiti per valutare se i materiali progettati fossero adeguati al gruppo-classe in cui sarebbero stati sperimentati ed eventualmente selezionare alcune attività a cui dare maggiore rilevanza o, invece, da escludere, in base alle competenze linguistiche degli apprendenti. Ad esempio, la

sperimentazione dell'unità sulla scelta tra indicativo e congiuntivo non è stata svolta a Madrid, né nella classe del I anno a Breslavia, dato che il tema del congiuntivo non era ancora stato trattato in entrambi i corsi di lingua italiana, dunque si è ritenuto che l'unità non fosse adeguata al livello di competenza degli apprendenti.

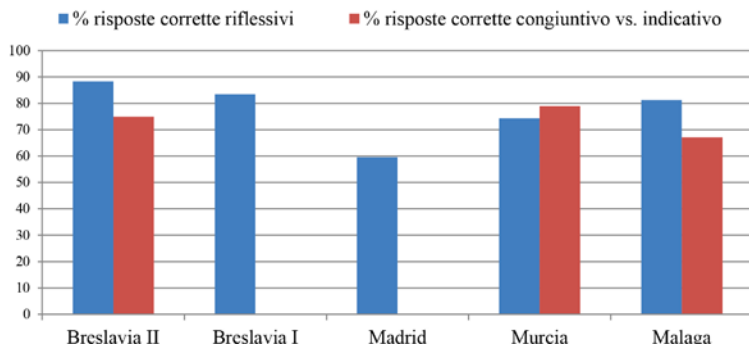
6.3.1. Esiti delle attività finali sui verbi riflessivi e sull'uso del congiuntivo vs indicativo

In questa sezione verrà svolto un confronto degli esiti delle attività di verifica svolte al termine delle due unità sperimentate. Prima di mostrare i dati raccolti e al fine di contestualizzarli, si ricorda che nelle attività finali di verifica (rispettivamente *Applicazione pratica 3* e *Uso del congiuntivo: formalità vs informalità*) viene chiesto agli studenti di riutilizzare le strutture grammaticali su cui è incentrata l'unità e fare delle riflessioni di tipo socio-pragmatico (ossia, scegliere di utilizzare o meno una determinata struttura linguistica in base allo specifico contesto comunicativo e al registro formale o informale ad esso correlato). Allo scopo di confrontare i dati ottenuti, per ciascuna classe di sperimentazione è stata calcolata la percentuale delle risposte corrette totalizzate nelle attività finali di verifica, che permette di fare un confronto degli esiti sia a livello globale tra le diverse sedi coinvolte, sia all'interno di ciascuna sede tra le due attività svolte. L'analisi comparativa escluderà sia la classe del I anno di Breslavia sia la classe di Madrid, dove – come già anticipato – l'unità didattica sul congiuntivo non è stata svolta, poiché l'argomento grammaticale non era ancora stato affrontato nel loro percorso curricolare.

Innanzitutto, per ciascuna classe di sperimentazione è stato calcolato il rapporto tra le risposte corrette date da ciascun partecipante e il totale dei quesiti dell'attività. Ad esempio, nel caso dell'attività finale sui verbi riflessivi, un partecipante ha totalizzato 13 risposte corrette su 15; il rapporto tra risposte corrette e quesiti totali è pari a $13/15=0,8666667$. Dopo aver svolto questa operazione per i punteggi di tutti i partecipanti, è stata calcolata la media dei rapporti ottenuti e, infine, da questa è stata ricavata la percentuale. Le percentuali ottenute sono state utilizzate per realizzare il grafico 6.1 e sono state inserite nella seguente tabella riassuntiva:

Tabella 6.2: Confronto delle percentuali delle risposte corrette nelle due attività finali di verifica

	Percentuale risposte corrette	
	Verifica sui verbi riflessivi “affettivi”	Verifica su congiuntivo vs indicativo
Breslavia I	83,5%	Unità didattica non sperimentata
Breslavia II	88,3%	75%
Madrid	59,6%	Unità didattica non sperimentata
Murcia	74,4%	79%
Malaga	81,3%	67,1%

Grafico 6.1: Confronto delle percentuali delle risposte corrette nelle due attività finali di verifica

L'analisi di tali dati verrà integrata in una prima fase dalle risposte date dai partecipanti al quesito *Da quanto tempo studi l'italiano?*, contenuto nel questionario sulle opinioni degli studenti (Appendice 3), che forniscono informazioni generali sulla durata media dello studio pregresso per ciascuna classe.

Prima di commentare i dati, si precisa che il gruppo-classe del II anno dell'università di Breslavia è l'unico in cui le stesse persone hanno partecipato ad entrambe le sperimentazioni, mentre i gruppi-classe di Malaga hanno conservato solo parzialmente la stessa composizione. Di conseguenza, solo nel caso di Breslavia può essere svolta un'effettiva analisi di tipo contrastivo su come lo stesso gruppo di persone abbia svolto l'una e l'altra unità didattica.

Osservando i dati dell'unità sui verbi riflessivi, emerge che la percen-

tuale più alta, pari all'88,3%, è stata totalizzata dalla classe del II anno di Breslavia, segue la classe di Breslavia del I anno con l'83,5%, Malaga con l'81,3, Murcia con il 74,4% e, infine, Madrid con il 59,6%.

Per quanto riguarda invece il congiuntivo il punteggio più alto è stato ottenuto dall'università di Murcia con il 79% di risposte corrette, segue il II anno di Breslavia con il 75% e, infine, Malaga con il 67,1%. Dalla correzione emerge che gli obiettivi delle unità sono stati globalmente raggiunti: in entrambe le attività di verifica i punteggi medi ottenuti in ciascuna classe sono superiori alla sufficienza; inoltre, i partecipanti sono in grado di reimpiegare le strutture grammaticali sulle quali è incentrata l'unità e fare delle riflessioni di tipo socio-pragmatico (ossia, scegliere di utilizzare o meno una determinata struttura linguistica in base allo specifico contesto comunicativo e al registro formale/informale ad esso correlato). Salvo poche eccezioni, gli errori rilevati sono principalmente di tipo morfologico (ad esempio, dubbi nella coniugazione dei verbi, nel sistema pronominale ecc.) o relativi a strutture linguistiche che non rappresentavano il *focus* dell'unità.

Entrando nel dettaglio, nel caso del II anno dell'università di **Breslavia** si nota che la percentuale di risposta più alta è stata ottenuta nell'unità didattica dei verbi riflessivi (88,3%) rispetto all'unità sull'uso dell'indicativo vs congiuntivo (75%). Questa differenza può essere spiegata sia in base al fatto che l'uso standard dei verbi riflessivi era già stato svolto al primo anno di corso, sia in base alle risposte fornite nel questionario di valutazione, dove tutti i partecipanti affermano di aver già trovato in precedenza esempi d'uso dei verbi riflessivi in senso "affettivo". Dunque per quanto riguarda l'uso standard dei verbi riflessivi, si tratta di un contenuto già consolidato nella competenza linguistica degli apprendenti del secondo anno, a cui la sperimentazione ha aggiunto ulteriori informazioni e ha fornito un inquadramento teorico circa l'uso riflessivo "affettivo" del neostandard. Al contrario, il tema del congiuntivo è stato trattato in tempi più recenti rispetto alla sperimentazione e, data anche la sua complessità, è comprensibile che permangano ancora dei dubbi (come emergerà dalle risposte date nel questionario nella sezione 6.4).

Una riflessione analoga può essere fatta nel caso dell'università di **Malaga**, dove ad esempio nel caso dell'unità sui verbi riflessivi la percentuale delle risposte corrette è molto elevata, pari all'81,3%. Ciò è

avvenuto nell'ambito di una classe molto disomogenea al suo interno dal punto di vista dello studio pregresso: su un totale di 5 persone la durata media dello studio della lingua italiana è di 2,4 anni, con un valore minimo di 1 anno e massimo di 6 anni.

Tuttavia, come per la classe di Breslavia, poiché l'argomento dei verbi riflessivi viene normalmente trattato tra il livello A2 e il B1, anche le persone aventi un solo anno alle spalle di studio dell'italiano sono riuscite a svolgere con successo l'attività, proprio perché l'argomento era probabilmente già noto.

Nel caso di **Murcia**, la percentuale più elevata delle risposte corrette è quella dell'attività finale sull'uso del congiuntivo vs indicativo (79%), caso unico tra le sperimentazioni condotte. Una possibile spiegazione può essere ricercata nei dati sullo studio pregresso della lingua italiana. Infatti, la classe che ha partecipato alla sperimentazione dell'unità sull'uso del congiuntivo vs indicativo era composta da 5 persone con una lunga durata dello studio pregresso della lingua italiana, nello specifico la durata media dello studio della lingua italiana è di 3,2 anni, compresa tra un valore minimo di 2 anni⁹ e un valore massimo di 4 anni e anche la durata più breve tra quelle indicate presuppone che il tema dell'uso standard del congiuntivo sia già stato affrontato e che le attività possano essere svolte agevolmente da tutti.

Nella classe che ha partecipato alla sperimentazione dell'unità sui verbi riflessivi, dei 6 partecipanti del III anno 4 hanno studiato l'italiano per 2 anni e 2 l'hanno studiato per 3 anni. In questo caso, invece, non si nota una differenza sostanziale in termini di rapporto tra la durata dello studio pregresso e i risultati ottenuti (per i dettagli sui singoli punteggi più alti e più bassi totalizzati in ciascuna classe di sperimentazione, si rinvia a Zingaro, 2021, capitolo 6).

Per quanto riguarda, invece, le classi in cui è stata svolta la sola sperimentazione dell'unità sui riflessivi, merita attenzione la percentuale delle risposte corrette pari all'83,5% totalizzata dalla classe del **I anno di Breslavia**, separata di soli 5 punti dagli esiti prodotti dalla classe

⁹ Nelle risposte date al questionario non è stato chiarito come mai una persona iscritta al III anno abbia dichiarato 2 anni e non 3 di conoscenza pregressa dell'italiano. Si ipotizza un cambiamento nel piano di studi nell'arco del quadriennio o un errore nel conteggio del periodo di studio.

del II anno dello stesso ateneo (88,3%). Un risultato così elevato potrebbe essere stato determinato da due fattori: innanzitutto il fatto che l'argomento dell'uso standard della forma riflessiva era stato trattato recentemente nell'ambito dell'insegnamento di lingua italiana curriculare, trattandosi del primo anno di studio della laurea triennale; in secondo luogo, all'interno della classe sono presenti alcune persone con una lunga durata di studi pregressi e un caso di bilinguismo familiare, nello specifico la durata media dello studio della lingua italiana è di 27,6 mesi¹⁰, con un valore minimo di 7 mesi e un massimo di 120 mesi (ossia, 10 anni).

A questo proposito è interessante notare che anche le persone con una durata dello studio pregresso inferiore ad un anno, compresa tra i 7 e 9 mesi, hanno condotto una sperimentazione di successo, con degli esiti assolutamente più che sufficienti: si citano i casi di A.K. e H.M. che, pur avendo soli 8 mesi pregressi, hanno totalizzato un punteggio pari a 13/15.

Nel caso di **Madrid**, la classe ha totalizzato il punteggio più basso tra le sperimentazioni svolte, pari al 59,6%. Per quanto riguarda la durata degli studi pregressi dell'italiano, su 15 persone la durata media dello studio della lingua italiana è di 14,8 mesi, con un valore minimo di 0, nel caso di un principiante assoluto, e un valore massimo di 2 anni. Nella fase preliminare alla sperimentazione, il docente referente a Madrid aveva messo in luce un livello di competenza pari ad A2 raggiunto al termine dell'anno accademico precedente, precisando inoltre che i tre mesi di pausa estiva intercorsa tra la fine del primo anno e lo svolgimento della sperimentazione (settembre 2019) avrebbero potuto incidere negativamente sulle prestazioni dei componenti della classe. Per quanto riguarda il punteggio più basso, pari a 4/15 e ottenuto da 3 persone, solo in un caso è noto che si tratti di un principiante assoluto. Non è stato possibile, invece, ricavare informazioni sulla durata dello studio pregresso né delle altre 2 persone che hanno totalizzato lo stesso punteggio, né della persona che ha totalizzato il punteggio più alto pari a 14/15, poiché alcuni partecipanti non hanno utilizzato lo stesso

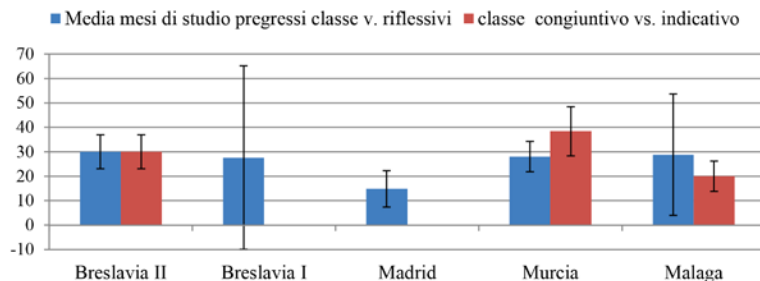
¹⁰ Come verrà di seguito più diffusamente illustrato, è stata effettuata la conversione in mesi della durata dello studio pregresso, data l'eterogeneità delle risposte raccolte.

username per accedere alla sperimentazione e al questionario finale. I dati più volte citati circa la durata dello studio pregresso, utilizzati per integrare l'analisi dei risultati ottenuti nelle attività di verifica finale, sono tratti dalle risposte dei partecipanti al quesito *Da quanto tempo studi l'italiano?*, contenuto nel questionario sulle opinioni degli studenti, come già anticipato all'inizio della sezione. Data l'eterogeneità delle risposte raccolte, che variano tra pochi mesi a molti anni di studio, al fine di permettere il confronto dei dati, è stata effettuata una conversione degli anni in mesi e, successivamente, è stata calcolata la media dei mesi di studio pregresso dell'italiano, come illustrato nella tabella 6.3 e nel grafico 6.2, che forniscono una visione d'insieme funzionale al confronto tra le classi di sperimentazione di entrambe le unità didattiche.

Tabella 6.3: Media dei mesi di studio pregresso della lingua italiana

	Classe di sperimentazione verbi riflessivi "affettivi"		Classe di sperimentazione congiuntivo vs indicativo	
	Media mesi	Deviazione standard	Media mesi	Deviazione standard
Breslavia I	27,6	37,5	La sperimentazione non è stata svolta	
Breslavia II	30	6,9	30	6,9
Madrid	14,8	7,4	La sperimentazione non è stata svolta	
Murcia	28	6,1	38,4	10
Malaga	28,8	24,8	20	6,1

Grafico 6.2: Confronto della media dei mesi di studio



Dal confronto dei dati emerge quanto segue. Innanzitutto, il valore della deviazione standard ottenuto per ciascuna classe di sperimentazione rispecchia un'elevata eterogeneità dei dati sulla durata dello studio pregresso, compresi tra i seguenti valori minimi e massimi: classi di sperimentazione della forma riflessiva: Breslavia I anno: 7-120 mesi (10 anni), più un caso di bilinguismo familiare non conteggiato; Murcia: 24-36 mesi; Malaga 12-72 mesi (6 anni); Madrid: 0-24 mesi; classi di sperimentazione del congiuntivo: Murcia 24-48 mesi; Malaga: 12-24 mesi. Infine, Breslavia II anno, in cui lo stesso gruppo ha svolto entrambe le sperimentazioni: 24-72 mesi.

A questo punto è possibile fare una prima riflessione sugli esiti delle attività di verifica. Per quanto riguarda le classi che hanno partecipato alla sperimentazione delle unità didattiche sui verbi riflessivi "affettivi", si può notare che tra le classi di Breslavia (I e II anno), Malaga e Murcia non vi sono molte differenze, sia in termini di durata media del periodo di studio pregresso (27,6 mesi, 30 mesi, 28,8 mesi e 28 mesi), sia in termini di profitto ottenuto (88,3% Breslavia II anno, 83,5% Breslavia I anno, 81,3 Malaga, 74,4% Murcia). L'unico caso che si discosta da questi è la classe di Madrid, dove in effetti ad una durata media dello studio pregresso della lingua italiana inferiore rispetto a quella delle altre classi, pari a 14,8 mesi, corrisponde anche un profitto inferiore, con la percentuale delle risposte corrette pari al 59,6%.

Nel caso della sperimentazione dell'unità didattica dedicata al congiuntivo, la graduatoria delle classi che hanno totalizzato il profitto migliore (Murcia 79%, Breslavia II anno 75% e Malaga 59,6%) rispecchia perfettamente l'ordine decrescente della durata media del periodo di studio pregresso, con Murcia che si attesta a 38,4 mesi, Breslavia II anno a 30 mesi e Malaga a 20.

Confrontando in particolare il grafico 6.1, che rappresenta la percentuale di risposte corrette totalizzate nelle due attività conclusive dalle diverse classi, e il grafico 6.2 sulla durata dello studio pregresso della lingua italiana, emerge che la disposizione dei dati sia pressoché speculare: le classi che hanno svolto con migliore profitto le attività conclusive presentano anche una più lunga durata dello studio pregresso, il che, però, rende necessaria una precisazione per una corretta interpretazione dei dati. La lunga durata dello studio dell'italiano non implica la conoscenza pregressa del tema della forma riflessiva "affet-

tiva” o dell’uso dell’indicativo al posto del congiuntivo: i partecipanti hanno infatti dichiarato di averne già incontrato alcuni esempi nella vita quotidiana, senza, tuttavia, avere alcuna cognizione teorica che ne spiegasse il funzionamento. Ciò lascia intendere che l’uso standard di entrambi i temi grammaticali trattati fosse un contenuto già consolidato nella competenza linguistica degli apprendenti, a cui la sperimentazione ha aggiunto ulteriori informazioni e ha fornito un inquadramento teorico circa l’uso nel neostandard.

Grafico 6.1: Confronto delle percentuali delle risposte corrette nelle due attività finali di verifica

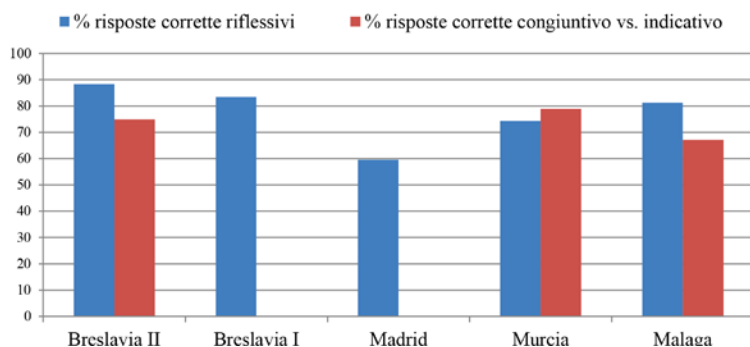
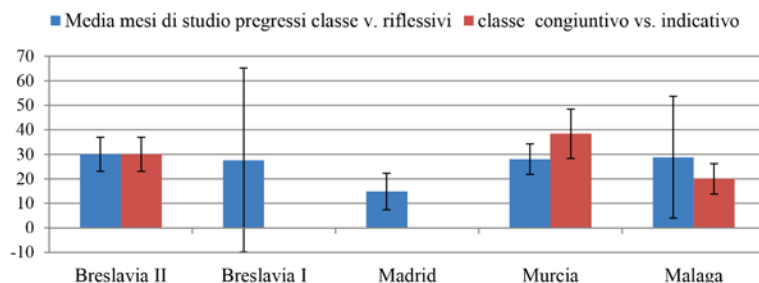


Grafico 6.2: Confronto della media dei mesi di studio



Queste riflessioni saranno arricchite da informazioni aggiuntive nella sezione successiva, dedicata alla descrizione degli esiti del questionario per la raccolta delle opinioni dei partecipanti sulle attività svolte. I risultati ottenuti nelle verifiche finali e la durata dello studio pregresso della lingua verranno incrociati con i dati relativi alle

L1 e alle altre eventuali lingue straniere (LS), alla presenza nella L1 e/o nelle LS di argomenti grammaticali simili o paragonabili alla forma riflessiva “affettiva” e all’alternanza dell’indicativo vs congiuntivo. Lo scopo è mettere in luce quanto questi fattori possano aver avuto un impatto sugli esiti delle sperimentazioni condotte in aula.

Verranno, infine, commentati i dati sulla percezione di utilità delle attività svolte e della piattaforma *Wizer.me* che le ospita.

6.3.2. Riflessioni finali sugli esiti delle sperimentazioni

In questa sezione le percentuali di risposte corrette ottenute in ciascuna classe verranno messe in relazione non solo alla durata dello studio pregresso della lingua, ma a tutti i dati emersi dai primi 7 quesiti del questionario, che hanno permesso di tracciare un profilo dei partecipanti: la lingua materna degli apprendenti (L1), la conoscenza di altre lingue straniere (LS), la familiarità dei partecipanti con gli argomenti trattati, sia dal punto di vista dell’uso standard che dell’uso neostandard e, infine, la presenza nella L1 e nelle altre LS conosciute di argomenti grammaticali simili o paragonabili alla forma riflessiva “affettiva” e all’alternanza dell’indicativo vs congiuntivo. I dati di seguito commentati verranno poi riassunti nelle tabelle conclusive (6.4 e 6.5).

Per quanto riguarda il possibile rapporto tra la L1 e le LS conosciute dai partecipanti e i risultati da loro ottenuti, in tutte le classi delle università spagnole coinvolte i partecipanti hanno dichiarato che la lingua madre e le altre lingue conosciute sono state utili a capire meglio gli argomenti grammaticali trattati, motivando la risposta sulla base delle similarità presenti tra italiano e spagnolo come L1 o come LS, nel caso di studenti *Erasmus*, (esempio di una risposta data a Malaga nel questionario: *sì, perché lo spagnolo anche ha il congiuntivo come tempo verbale*). In alcuni casi sono state evidenziate anche le analogie esistenti tra italiano e francese come LS, essendo quest’ultima la lingua straniera più diffusa tra i partecipanti delle classi di Murcia (e la lingua straniera romanza più diffusa tra le classi di Breslavia, come verrà di seguito illustrato).

Per le classi di Murcia e Malaga, dove sono state ottenute delle alte

percentuali di risposte corrette nelle attività finali¹¹, si può affermare che ad una lunga durata media dello studio progresso della lingua italiana si accompagna anche la percezione del vantaggio costituito dalla L1 romanza e dalla conoscenza di altre lingue straniere romanze che presentano analogie con i temi trattati nella sperimentazione. Al contrario, questa riflessione non si applica al caso della classe di Madrid.

La classe di Madrid si caratterizza per la presenza al suo interno della più alta varietà di lingue romanze oltre allo spagnolo (L1 o LS) e all'italiano (LS): una persona ha indicato il rumeno come L1, mentre tra le lingue straniere conosciute sono state indicate il francese, il catalano, il galiziano, il portoghese e il valenciano. A ciò si accompagna una certa consapevolezza circa l'esistenza di un fenomeno grammaticale analogo nella L1. Alla domanda *Esiste un fenomeno simile nella tua lingua?* (in riferimento alla forma riflessiva "affettiva") 12 persone hanno risposto sì, 4 persone non ne sono del tutto sicure (esempio: *Non ci avevo mai pensato. Diciamo in teoria sì, ma per me si manifesta in modo diverso*). Tuttavia, in assenza di una lunga durata pregressa dello studio dell'italiano, la sola presenza di numerose lingue romanze all'interno della classe non sembra aver costituito un fattore determinante e sufficiente per il raggiungimento di risultati equiparabili a quelli delle altre classi, dato che, come già affermato, nella classe è stata totalizzata la percentuale più bassa di risposte corrette tra tutte le classi di sperimentazione (59,6%).

Queste riflessioni sembrano trovare conferma osservando i dati delle classi di Breslavia, dove si è verificata una situazione analoga a quella delle classi di Murcia e Malaga, con il raggiungimento di percentuali elevate di risposte corrette¹². A prima vista il profilo dei partecipanti polacchi alla sperimentazione presenta alcune differenze rispetto a quello dei partecipanti spagnoli, ossia la L1 slava e la presenza molto ristretta di parlanti di lingue romanze come LS diverse dall'italiano: in entrambe le classi di Breslavia la lingua straniera più frequente tra

¹¹ Forma riflessiva "affettiva": Murcia 74,4%; Malaga 81,3%; congiuntivo: Murcia 79%; Malaga: 67,1%.

¹² Forma riflessiva "affettiva": II anno 88,3%; I anno 83,5%; congiuntivo: II anno 75%.

i partecipanti è l'inglese, segue una ristretta minoranza che ha indicato il francese o lo spagnolo e altre lingue non romanze.

Tuttavia, nonostante queste differenze nel profilo linguistico, un fattore che ha contribuito al raggiungimento di alte percentuali di risposte corrette nel caso dell'unità dedicata ai verbi riflessivi "affettivi" è probabilmente la presenza in polacco di una struttura grammaticale simile, utilizzata nel registro informale, come dichiarato dai partecipanti nelle risposte al questionario. A titolo di esempio, si riportano qui alcune risposte, per le quali è stata rispettata la grafia originale, senza apportare correzioni e la stessa scelta verrà applicata da qui in avanti: *Si, in polacco esiste l'argomento uguale; Si, la mia lingua madre ha la stessa cosa in grammatica*; alla domanda *La tua lingua madre o altre lingue che conosci ti hanno aiutato/a a imparare/capire meglio l'argomento di grammatica di oggi?* 12 persone hanno risposto di sì: *Sì, polacco mi ha aiutato; La conoscenza del francese o anche dell'inglese mi fa diventare piu saggio in questo caso.*

Nel caso dell'unità sul congiuntivo, invece, alla domanda *Esiste un fenomeno simile nella tua lingua?* tutti nella classe del II anno di Bre-slavia hanno risposto di no¹³ e hanno affermato che le lingue straniere conosciute non sono state utili a familiarizzare con il tema della sperimentazione, incluse le persone che hanno una lingua romanza nel proprio curriculum, poiché la loro conoscenza di quelle lingue era troppo limitata per poter essere d'aiuto. Ad esempio: *Inglese non aiuta. Spagnolo potrebbe aiutarmi ma ho un livello della lingua troppo bassa.*

Sulla base di quanto affermato, emerge che le classi di sperimentazione che hanno svolto con migliore profitto le attività conclusive sull'uso neostandard dei due tratti grammaticali sono quelle che presentano, in generale, i seguenti fattori: innanzitutto, una più lunga durata pregressa dello studio della lingua italiana, a cui si aggiunge la presenza nella L1 di un fenomeno grammaticale simile a quello affrontato nell'unità. Quest'ultimo fattore si è verificato sia nel polacco che nelle varie lingue romanze dei partecipanti coinvolti: tuttavia, come mostrato dai dati precedentemente presentati, specie nel caso

¹³ "Il polacco non possiede la categoria del congiuntivo" (Sosnowski, 2010, p. 144).

della classe di Madrid, in assenza di una lunga durata dello studio pregresso, la sola analogia tra l'italiano e la L1 e/o le diverse lingue romanze conosciute non ha costituito un fattore determinante per lo svolgimento positivo delle attività.

Oltre a ciò, la maggior parte dei partecipanti delle classi di sperimentazione, ad eccezione di pochi, ha dichiarato di aver già incontrato nella vita quotidiana alcuni esempi dell'uso neostandard degli argomenti trattati nelle unità, senza, tuttavia, averne alcuna consapevolezza o cognizione teorica che ne spiegasse il funzionamento. Si citano a titolo di esempio alcune risposte raccolte per la forma riflessiva "affettiva": *sì, ma non sapevo che venga usato dai Italiani così spesso; un po', ma nessuno me li ha mai spiegati* (Breslavia I anno); *Lo avevo sentito ma non comprendevo il senso esatto di quest'uso* (Madrid). Se ne deduce che l'uso standard di entrambi i temi grammaticali trattati fosse un contenuto già presente nella competenza linguistica degli apprendenti e che l'impatto delle unità si sia concretizzato non solo nell'aver aggiunto ulteriori informazioni circa l'uso standard, ma soprattutto nell'aver fornito un inquadramento teorico riguardo all'uso neostandard, che sino ad allora era percepito dagli studenti come errato o inspiegabile, in quanto fuori dalla norma linguistica che era stata loro trasmessa. Ad ulteriore conferma, come già anticipato in precedenza, che una lunga durata dello studio pregresso non implica la conoscenza dell'uso neostandard dei fenomeni grammaticali oggetto di sperimentazione.

Attraverso l'uso delle unità i partecipanti hanno, quindi, dimostrato di saper riutilizzare le strutture grammaticali su cui erano incentrate le unità e di saper scegliere di utilizzare o meno una determinata struttura linguistica in base allo specifico contesto comunicativo e al registro formale o informale ad esso correlato.

I dati finora commentati sono riportati nelle seguenti tabelle 6.4 e 6.5, che riassumono le risposte date ai primi 7 quesiti a risposta aperta del questionario del questionario, per alcuni dei quali è stato riportato tra parentesi il numero di persone che hanno dato una determinata risposta.

Tabella 6.4: Dati relativi alle classi di sperimentazione dell'unità sui verbi riflessivi "affettivi"

	Breslavia II	Breslavia I	Madrid	Murcia	Malaga
Media mesi (studio pregresso)	30	27,6	14,8	28	28,8
Deviazione standard	6,9	37,5	7,4	6,1	24,8
% risposte corrette verifica finale	88,3%	83,5%	59,6%	74,4%	81,3%
L1 e LS	L1: polacco; LS: inglese (4), francese (3), hindi (1), spagnolo (1), tedesco (1)	L1: polacco (tutti), polacco e italiano (1) LS: inglese (13), francese (3), russo (3), tedesco (3), spagnolo (2), hindi (1), svedese (1)	L1: spagnolo (18), rumeno (1), finlandese (1), tedesco (1). LS: inglese (17), tedesco (9), francese (6), spagnolo LS (3), catalano (2), galiziano (2), giapponese (1), olandese (1), portoghese (1), russo (1), svedese (1), valenciano (1)	L1: spagnolo; LS: francese (6), inglese (4)	L1: spagnolo (3), francese (2); LS: inglese (5), francese LS (2), spagnolo LS (2), portoghese (1)
Utilità della L1/LS nella comprensione del tema?	Sì (2) la L1, perché ha un argomento grammaticale simile; un po' (1); no (1)	Sì (12) la L1, ha un argomento grammaticale simile; no (2)	Sì (tutti): similarità presenti tra l'italiano e lo spagnolo	Sì (tutti): similarità presenti tra l'italiano e lo spagnolo	La L1: sì, (tutti); similarità presenti tra l'italiano e il francese (1)
Hai già incontrato esempi dell'uso neostandard?	Sì (tutti)	Sì (11) No (3)	Sì (13) No (4)	Sì (3); 2 (no); 1 (non so)	Sì (tutti)

Tabella 6.5: Dati relativi alle classi di sperimentazione dell'unità sul congiuntivo vs indicativo

	Breslavia II	Murcia	Malaga
Media mesi (studio pregresso)	30	38,4	20
Deviazione standard	6,9	10	6,1
% risposte corrette verifica finale	75%	79%	67,1%
L1 e LS	L1: polacco; LS: inglese (4), francese (3), hindi (1), spagnolo (1), tedesco (1)	L1: spagnolo; LS: francese (5), inglese (3), portoghese (1)	L1: spagnolo (5) francese (1); LS: inglese (6), francese (3), spagnolo (1)
Utilità della L1/LS nella comprensione del tema?	Tutti (no): né la L1, né le LS, di cui alcuni dichiarano di avere un livello troppo basso	La L1: sì, (tutti); LS: similarità presenti tra italiano, spagnolo e francese (1)	La L1: sì, (5), per le similarità presenti tra italiano, spagnolo; no (1)
Hai già incontrato esempi dell'uso neostandard?	Sì (3), No (1)	Sì (2), No (3)	Sì (3), No (3)

Infine, nelle sezioni successive verranno commentate le opinioni dei partecipanti sulle attività svolte e sull'uso della piattaforma che ospita le unità didattiche.

6.4 Opinioni sulle attività svolte e sulla piattaforma

Dopo aver illustrato gli esiti delle unità didattiche, nelle sezioni successive saranno descritti i dati relativi alle opinioni dei partecipanti sulle attività svolte, raccolti tramite la piattaforma *Socrative* (cfr. capitolo 4). Nell'analisi delle risposte date, la trattazione sarà suddivisa in base alle 2 parti del questionario finora ancora non commentate: opinioni sulle attività e opinioni sull'uso della piattaforma *Wizer.me*. I dati raccolti nei quesiti a scelta multipla saranno rappresentati graficamente.

6.4.1. Unità didattica sui verbi riflessivi “affettivi”: opinioni sulle attività
Dopo aver tracciato un profilo dei partecipanti alle sperimentazioni tramite i dati raccolti nei quesiti 1-7 e commentati precedentemente, nei quesiti 8-25 viene chiesto ai partecipanti di esprimere la propria opinione sulle attività svolte sulla scala di valori (0) *no*, (1) *poco*, (2) *abbastanza*, (3) *sì* o di selezionare la voce *attività non svolta*, nel caso in cui quell’attività non sia stata proposta in classe. Nello specifico, i quesiti riguardano la percezione di utilità delle attività svolte (8-17), la percezione personale di acquisizione dei contenuti (18-21) e l’utilizzo della piattaforma (22-25). Le percentuali delle risposte verranno presentate nei grafici 6.3-6.8.

Di seguito la trattazione si focalizzerà sui dati riguardanti la percezione di utilità delle attività svolte (quesiti 8-17): *È stata utile questa attività?*
8) Leggere il testo “L’appuntamento”; 9) Esercizi di collegamento (parole del testo-significato); 10) Rispondere alle domande sul tema del testo; 11) Cercare nel testo i verbi che descrivono la cura del corpo; 12) Completare la coniugazione dei verbi riflessivi; 13) Distinguere le diverse azioni espresse dai verbi riflessivi (*azione del soggetto su se stesso/ su parti del corpo o vestiti/accessori che appartengono al soggetto; azione reciproca tra più soggetti; riflessivo “affettivo”*); 14) Scegliere tra la forma riflessiva-affettiva e la forma non riflessiva; 15) Contenuto extra: i riflessivi nei modi di dire; 16) Produzione scritta e orale; 17) Fare esercizi basati su frasi vere della vita quotidiana (tratte da blog e altre pagine web) che non sono state inventate per “fare grammatica”.

Grafico 6.3: Percentuali delle risposte date nella classe di Breslavia II ai quesiti sulla percezione di utilità delle attività svolte

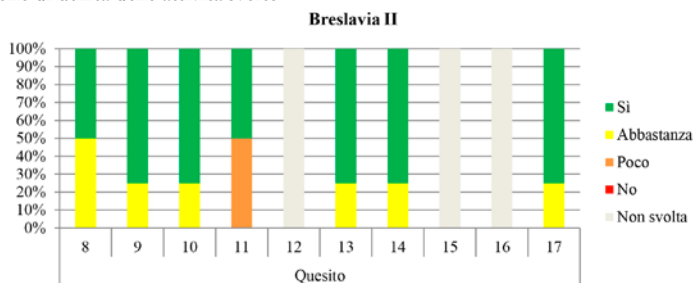


Grafico 6.4: Percentuali delle risposte date nella classe di Breslavia I ai quesiti sulla percezione di utilità delle attività svolte

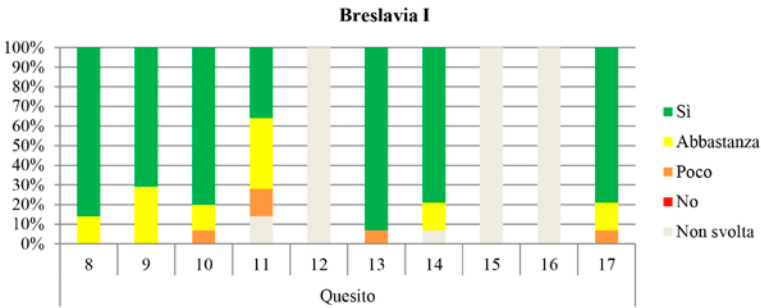


Grafico 6.5: Percentuali delle risposte date nella classe di Madrid ai quesiti sulla percezione di utilità delle attività svolte

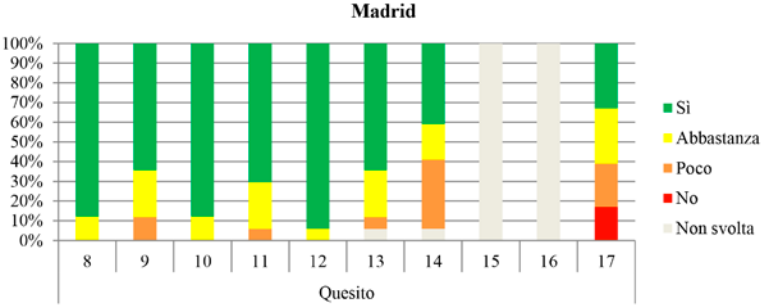


Grafico 6.6: Percentuali delle risposte date nella classe di Murcia ai quesiti sulla percezione di utilità delle attività svolte

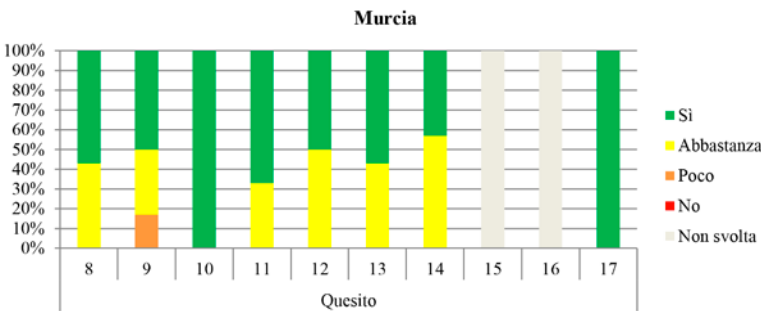


Grafico 6.7: Percentuali delle risposte date nella classe di Malaga ai quesiti sulla percezione di utilità delle attività svolte

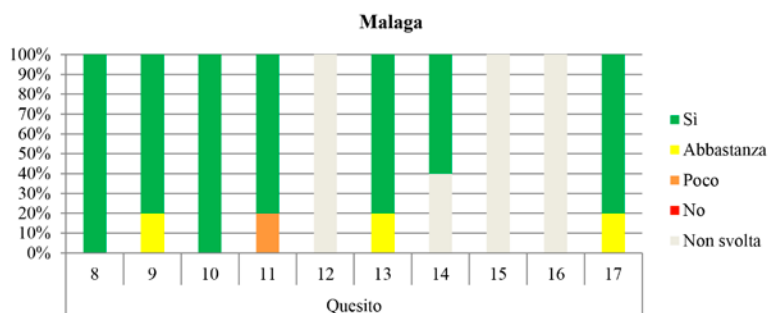
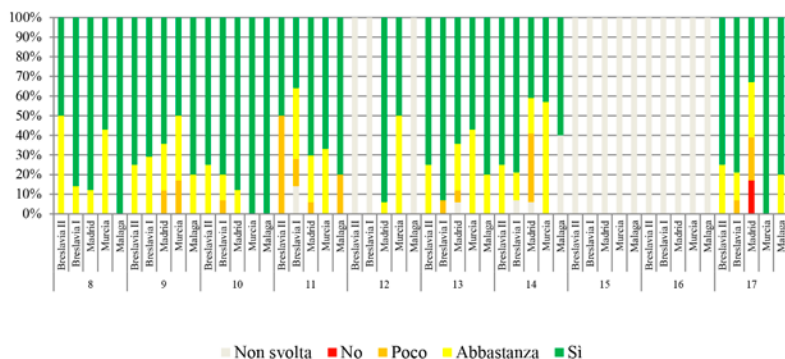


Grafico 6.8: Panoramica delle percentuali delle risposte date in tutte le classi di sperimentazione ai quesiti sulla percezione di utilità delle attività svolte



Osservando le percentuali delle risposte date¹⁴, emerge a livello globale una generale percezione di elevata utilità delle attività svolte. Di seguito verranno commentati alcuni casi particolari, che hanno trovato spiegazione nei commenti liberi lasciati dai partecipanti al termine delle attività. Al punto 11. *Cercare nel testo i verbi che descrivono la cura del corpo* il fatto che il 50% della classe del II anno di Breslavia e il 20% della classe di Malaga abbiano definito l'attività come *poco utile* è giustificabile sulla base del livello di competenza linguistica molto alto dei partecipanti, per i quali l'at-

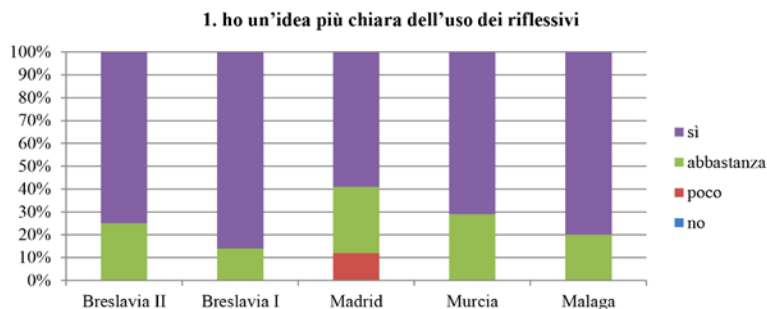
¹⁴ Si precisa che le percentuali del questionario sono state arrotondate all'intero dal sistema di calcolo presente nella piattaforma *Socrative*.

tività è stata percepita come banale. Viceversa, nel caso del punto 12. *Completare la coniugazione dei verbi riflessivi*, l'attività è stata percepita come *utile* dal 94% della classe di Madrid, dove il livello di competenza della classe era inferiore rispetto a quello delle altre classi partecipanti, pari ad un A2 pienamente raggiunto, e dunque l'attività di ripasso della coniugazione è stata ritenuta adeguata. Nel caso dei punti 13. *Distinguere le diverse azioni espresse dai verbi riflessivi* e 14. *Scegliere tra la forma riflessiva-affettiva e la forma non riflessiva* si ipotizza che la scelta dell'opzione *attività non svolta* (cfr. punto 13: Madrid 6%; punto 14: Breslavia I: 7%; Madrid: 6%; Malaga: 40%) derivi da una errata interpretazione della domanda da parte dei partecipanti, dato che le attività sono state proposte in tutte le sedi di sperimentazione e, di fatto, sono state svolte da tutti i partecipanti. Diverso, è invece, il caso dei quesiti 15 e 16, che si riferiscono ad attività pensate come supplementari (ossia: 15. *I riflessivi nei modi di dire*; 16. *Produzione scritta e orale*) e che di fatto non sono state svolte in nessuna sede e, dunque, la scelta del descrittore non svolta corrisponde alla reale prassi didattica.

Di seguito la trattazione si focalizzerà sui dati riguardanti la percezione personale di efficacia delle attività svolte nell'acquisizione dei contenuti dell'unità (quesiti 18-21): *Dopo queste attività...*: 18) ho un'idea più chiara dell'uso dei riflessivi; 19) riesco a distinguere meglio le situazioni formali da quelle informali; 20) ho capito che in alcune situazioni informali è accettabile usare i riflessivi "affettivi"; 21) ho ancora dei dubbi sull'uso dei riflessivi.

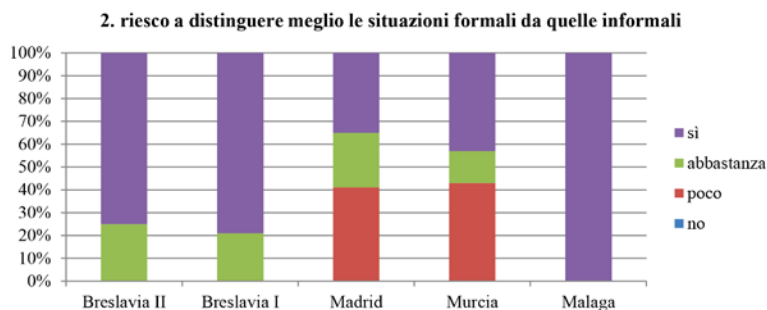
Osservando il grafico 6.9 emerge che quasi la totalità dei partecipanti afferma di avere un'idea più chiara sull'uso dei riflessivi, con l'unica eccezione della classe di Madrid, dove vi è una scarsa percentuale (12%) che afferma di avere ancora un'idea poco chiara. A questo proposito va tenuto conto, come già affermato in precedenza, del livello di competenza inferiore rispetto ai componenti delle altre classi.

Grafico 6.9: Percentuali delle risposte date in tutte le classi di sperimentazione al primo quesito sulla percezione personale di acquisizione dei contenuti



Al secondo quesito (grafico 6.10) *riesco a distinguere meglio le situazioni formali da quelle informali*, la totalità dei partecipanti di Murcia e la maggioranza dei partecipanti di Breslavia (79% per il I anno e 75% per il II anno) afferma di *sì*, nel caso di Madrid il 41% ha risposto *poco*, seguito dal 35% *sì* e dal 24% *abbastanza*. Questo dato va considerato alla luce delle considerazioni già fatte sul livello di competenza più basso degli apprendenti del gruppo classe di Madrid. Nel caso di Murcia vi è una parità di risposte (43%) per le voci *sì* e *poco* e la percentuale rimanente ha risposto *abbastanza*.

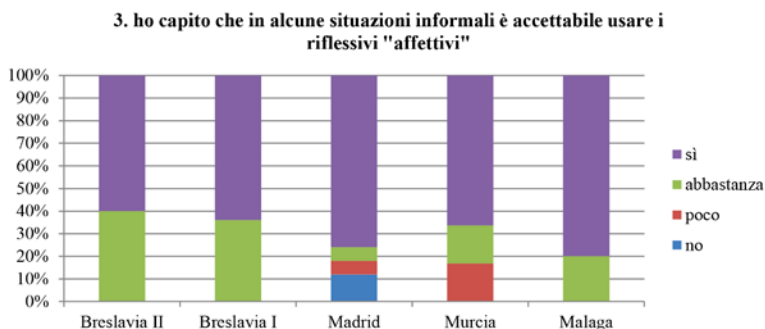
Grafico 6.10: Percentuali delle risposte date in tutte le classi di sperimentazione al secondo quesito sulla percezione personale di acquisizione dei contenuti



Al terzo quesito (grafico 6.11) *ho capito che in alcune situazioni informali è accettabile usare i riflessivi "affettivi"*, in ciascuna classe la mag-

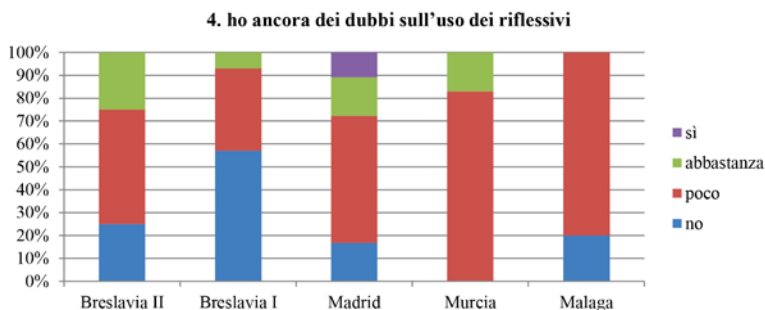
gioranza delle risposte è stata attribuita alla voce *sì*, permangono dei dubbi ancora nelle classi di Madrid (12% *no*, 6% *poco* e *abbastanza*) e Murcia (17% *poco* e *abbastanza*).

Grafico 6.11: Percentuali delle risposte date in tutte le classi di sperimentazione al terzo quesito sulla percezione personale di acquisizione dei contenuti



Al quarto quesito (grafico 6.12) *Ho ancora dei dubbi sull'uso dei riflessivi "affettivi"*, nelle classi del II anno di Breslavia, Madrid, Murcia e Malaga prevale la risposta *poco* (rispettivamente pari al 50, 56, 83 e 80%), mentre nella classe del I anno di Breslavia prevale la risposta *no* (57%).

Grafico 6.12: Percentuali delle risposte date in tutte le classi di sperimentazione al quarto quesito sulla percezione personale di acquisizione dei contenuti



I dati ottenuti dai quesiti sulla percezione di utilità delle attività svolte e sulla percezione personale di acquisizione dei contenuti trovano conferma nella sezione del questionario dedicata ai commenti liberi, che vengono riportati qui di seguito. Nella trascrizione è stata rispettata la grafia originale, senza apportare correzioni, indicando tra parentesi la classe in cui sono stati formulati.

Alcuni di questi commenti esprimono un apprezzamento per l'operato della docente-ricercatrice o per la piacevolezza e utilità dell'attività (*L'attività è stata molto piacevole :*) (Breslavia II anno); *Era molto brave* (Breslavia I anno); *Tutto bene. Mi è piaciuta la lezione; Mi è piaciuta l'attività e ho imparato; Per me, l'esercizio è molto bene per imparare l'italiano* (Madrid); *Penso che siano state delle attività utile e, anzi, ci ha imparato qualcosa di nuovo; Seconde me questi esercizi sono molto utili per aiutarci a comprendere meglio questo tema* (Murcia).

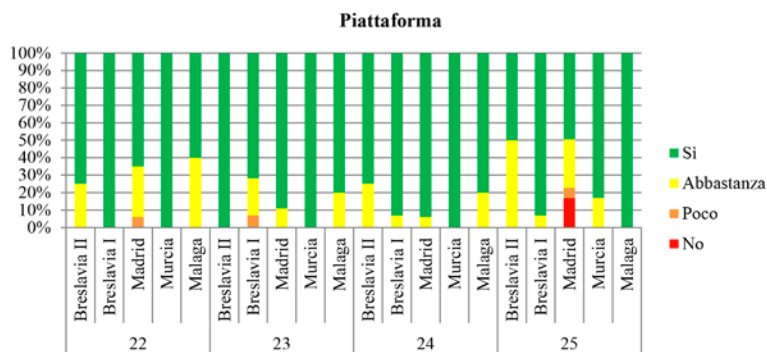
In altri commenti vengono dati suggerimenti sullo svolgimento della lezione (*Penso che Lei possa aggiungere un po' piu' di conversazione, abbiamo abbastanza grammatica all'universita'. Vorremmo avere piu' contatto con il parlato* (Breslavia I anno); *piu tempo per capire la grammatica!* (Madrid)), sulle preferenze personali circa le strategie di apprendimento (*È chiara per me la lezione di oggi, ma mi piace fare i esercizi dopo una spiegazione con un libro perche è più chiaro per me la teoria con appunti; la actividad esta muy bien hecha, pero me gustaria mas ejemplos y/o mas definiciones para poder comprenderlo mejor* (Madrid)) e, infine, sono stati inseriti dei commenti sulla grafica e la struttura delle unità, che verranno citati al termine di questa sezione.

6.4.2. Unità didattica sui verbi riflessivi "affettivi": opinioni sulla piattaforma *Wizer.me*

Per quanto riguarda i quesiti sulla fruizione della piattaforma (22-25), dal grafico 6.13 emerge la generale soddisfazione degli utenti, le cui percentuali di risposta si dividono tra i descrittori *sì* e *abbastanza*¹⁵. Di seguito verranno commentate alcune particolarità.

¹⁵ Si ricorda che le percentuali del questionario sono state arrotondate all'intero dal sistema di calcolo presente nella piattaforma *Socrative*.

Grafico 6.13: Panoramica delle percentuali delle risposte date in tutte le classi di sperimentazione ai quesiti sull'utilizzo della piattaforma



Circa il punto 22 *Il login e l'accesso alle diverse parti dell'unità è stato facile*, nella classe di Madrid il 6% dei partecipanti ha scelto la risposta *poco*, che presumibilmente si giustifica alla luce dello svolgimento della sperimentazione su dispositivi elettronici appartenenti ai partecipanti (in alcuni casi persino su *smartphone*) e del conseguente rallentamento dei tempi di connessione e apertura delle pagine rispetto all'utilizzo di un'aula computer già predisposta alle attività dall'insegnante.

Al punto 23 *Le istruzioni delle attività erano comprensibili* nella classe del I anno di Breslavia il 7% ha risposto *poco*: tale scelta rispecchia le richieste da parte di alcuni partecipanti di spiegazione delle consegne degli esercizi o di singole parole e trova spiegazione un commento lasciato da un partecipante nella sezione del questionario dedicata ai commenti liberi: *Secondo me, alcune consegne non erano chiare e poco precise, in generale gli esercizi mi sono piaciuti ma le consegne appunto, potrebbero essere più precise.*

Nel caso, invece, del punto 25 *È stato utile avere esercizi con correzione automatica*, solo nella classe di Madrid sono stati scelti i descrittori *no* (17%) e *poco* (6%), tuttavia queste risposte non trovano spiegazione nella sezione del questionario dedicata ai commenti liberi, dove invece sono stati formulati vari suggerimenti per il miglioramento delle attività sulla piattaforma, per i quali è stata rispettata la grafia originale, senza apportare correzioni.

Alcuni suggerimenti della classe di Madrid riguardano la progettazione strutturale e grafica delle attività: *Per me, può essere meglio di rafforzare*

il layout della pagina; Dovresti avere due opzioni in alcuni esercizi per quando ci sono due possibili risposte; Suggestisco collocare tre opzioni di risposta nell'esercizio dove sono possibili più risposte. Questo suggerimento è stato formulato in riferimento ad alcune attività in cui erano forniti nella consegna vari verbi all'infinito, chiedendo ai partecipanti di utilizzarne la forma coniugata per completare alcune frasi e tenendo conto del fatto che, per alcune frasi, fossero possibili più soluzioni: per questi casi specifici i partecipanti suggeriscono di fornire una scelta multipla, invece di scegliere le risposte da un unico elenco di opzioni valido per tutto l'esercizio. Un'alternativa alla scelta multipla viene fornita nella classe di Murcia, dove si suggerisce di progettare l'esercizio facendo in modo che i verbi utilizzati per completare le frasi non siano poi più selezionabili dall'elenco: *In generale penso che sarebbe meglio di potere togliere la risposta quando usi la parola nell'esercizio e potere mettere le diversi risposte.*

6.4.3. Unità didattica sul congiuntivo: opinioni sulle attività

Nei quesiti 8-20 viene chiesto ai partecipanti di esprimere la propria opinione sulla seguente scala di valori: (0) *no*; (1) *poco*; (2) *abbastanza*; (3) *sì* o di selezionare la voce *attività non svolta*, nel caso in cui quell'attività non sia stata proposta in classe. Le percentuali delle risposte date ai quesiti sulla percezione di utilità delle attività svolte (8-11), sulla percezione personale di acquisizione dei contenuti (12-15) e sull'utilizzo della piattaforma *Wizer.me* (16-20) verranno presentate nei grafici 6.14-6.21. Di seguito la trattazione si focalizzerà sui dati riguardanti la percezione di utilità delle attività svolte (quesiti 8-11) *È stata utile questa attività?*: 8) Cercare nei 4 brevi testi gli esempi d'uso del congiuntivo; 9) Ripassare i casi in cui il congiuntivo è obbligatorio; 10) Fare ipotesi sulla scelta tra congiuntivo e indicativo in base al contesto formale/informale; 11) Fare esercizi basati su frasi vere della vita quotidiana (tratte da blog e altre pagine web) che non sono state inventate per "fare grammatica". Osservando le percentuali delle risposte date, emerge a livello globale una percezione di elevata utilità delle attività svolte.

Grafico 6.14: Percentuali delle risposte date nella classe di Breslavia II ai quesiti sulla percezione di utilità delle attività svolte

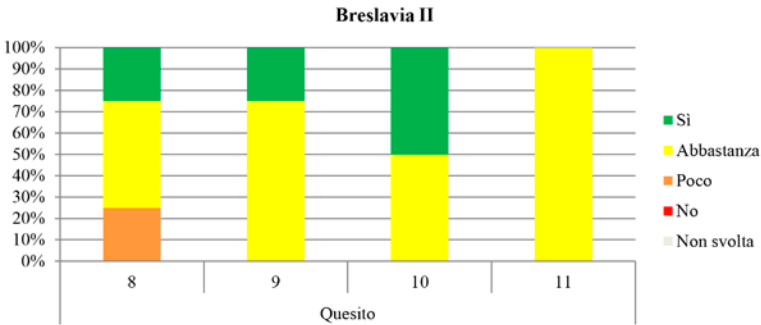


Grafico 6.15: Percentuali delle risposte date nella classe di Murcia ai quesiti sulla percezione di utilità delle attività svolte

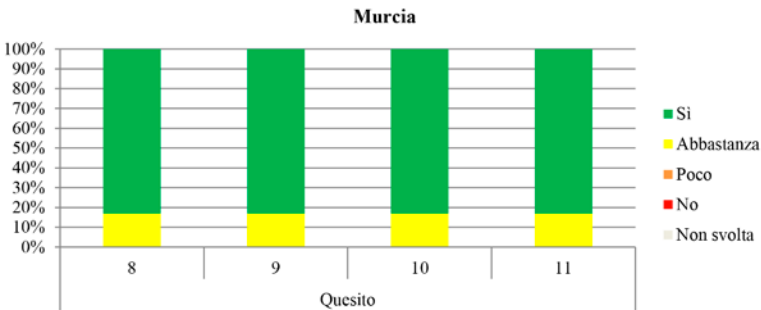


Grafico 6.16: Percentuali delle risposte date nella classe di Malaga ai quesiti sulla percezione di utilità delle attività svolte

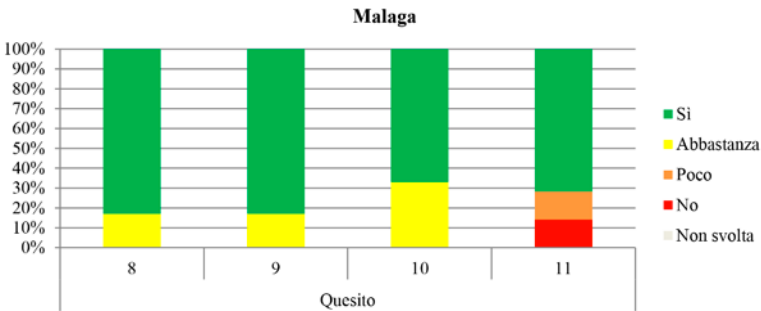
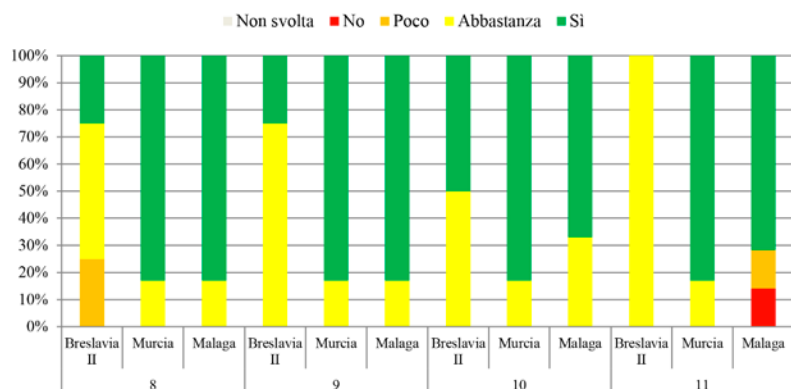


Grafico 6.17: Panoramica delle percentuali delle risposte date in tutte le classi di sperimentazione ai quesiti sulla percezione di utilità delle attività svolte



Di seguito verranno commentati solo due casi in cui le risposte comprendono anche i descrittori *poco* e *no*. Nel primo caso, rispetto al punto 8. *Cercare nei 4 brevi testi gli esempi d'uso del congiuntivo* nella classe di Breslavia le risposte date si suddividono tra *poco utile* (25%), *abbastanza utile* (50%) e *sì* (25%): data l'assenza di spiegazioni nella sezione dedicata ai commenti personali, si ipotizza che il 25% delle preferenze date alla voce *poco utile* possa essere dovuto alla durata pregressa dello studio della lingua dei componenti di questa classe (cfr. sezioni 6.3.1 e 6.3.2), per cui la semplice individuazione nel testo delle forme verbali sia un'attività banale, non percepita come sfidante e come utile allo sviluppo delle competenze.

Nel secondo caso, rispetto al punto 11. *fare esercizi basati su frasi vere della vita quotidiana (tratte da blog e altre pagine web) che non sono state inventate per "fare grammatica"*, nella classe di Malaga le risposte date si suddividono tra *no* (14%), *poco utile* (14%) e *sì* (71%)¹⁶. Data anche in questo caso l'assenza di spiegazioni nella sezione dedicata ai commenti personali, si ipotizza che il 14% delle preferenze date alla voce *no* dipenda dal fatto che, semplicemente, dal punto di vista dell'apprendente l'uso di testi autentici non influisca nella percezione di utilità dell'attività, a differenza di al-

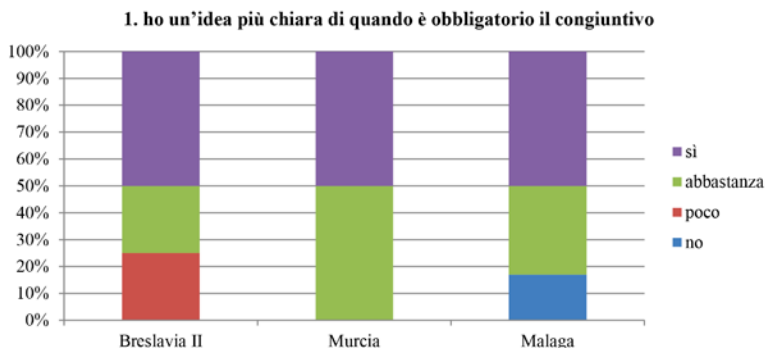
¹⁶ Si ricorda, come già specificato in precedenza, che le percentuali del questionario sono state arrotondate all'intero dal sistema di calcolo presente nella piattaforma *Socrative*.

tri fattori menzionati nel questionario (esempio, la chiarezza delle tracce degli esercizi).

Di seguito la trattazione si focalizzerà sui dati riguardanti la percezione personale di efficacia delle attività svolte nell'acquisizione dei contenuti dell'unità (quesiti 12-15), ossia: *Dopo queste attività...*: 12) Ho un'idea più chiara di quando è obbligatorio il congiuntivo; 13) Riesco a distinguere meglio le situazioni formali da quelle informali; 14) Ho capito che in alcune situazioni informali è accettabile non usare il congiuntivo; 15) Ho ancora dei dubbi sulla scelta tra congiuntivo e indicativo.

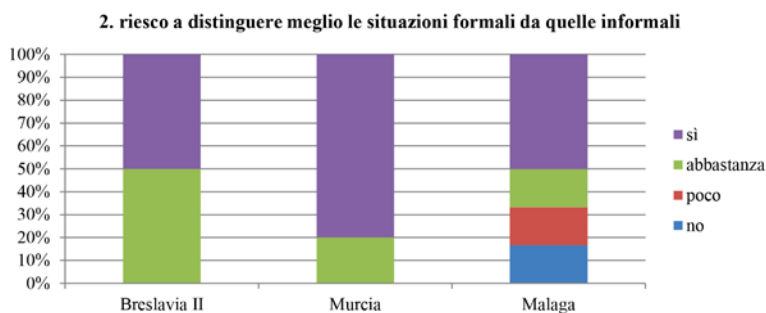
Al primo quesito (grafico 6.18) *ho un'idea più chiara di quando è obbligatorio il congiuntivo*, in tutte le classi di sperimentazione il 50% dei partecipanti ha risposto *sì*; nel caso di Murcia vi è parità tra le voci *sì* e *abbastanza*, mentre solo nella classe di Malaga il 17% ha risposto *no*.

Grafico 6.18: Percentuali delle risposte date in tutte le classi di sperimentazione al primo quesito sulla percezione personale di acquisizione dei contenuti



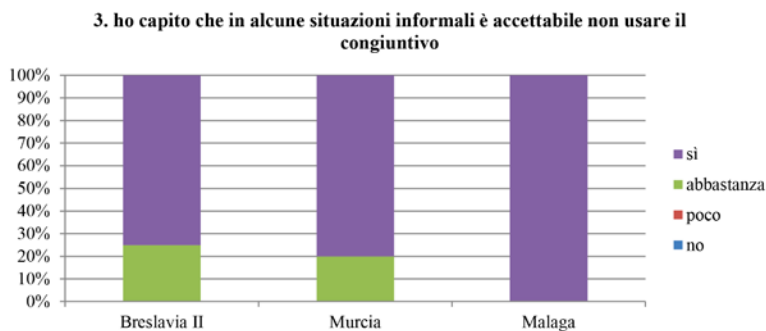
Al secondo quesito (grafico 6.19) *riesco a distinguere meglio le situazioni formali da quelle informali* nella classe di Murcia prevale il *sì* (80%), nella classe di Breslavia vi è parità tra le voci *sì* e *abbastanza*; restano alcuni dubbi nella classe di Malaga (le voci *no*, *poco* e *abbastanza* hanno ricevuto il 16,6% delle preferenze, mentre la maggioranza si è espressa per il *sì*).

Grafico 6.19: Percentuali delle risposte date in tutte le classi di sperimentazione al secondo quesito sulla percezione personale di acquisizione dei contenuti



Osservando il grafico 6.20, emerge che, al termine della verifica, è chiaro a tutti gli apprendenti coinvolti che in alcune situazioni informali è accettabile non usare il congiuntivo. La percentuale del *sì* è pari al 100% a Malaga, all'80% a Murcia e al 75% a Breslavia e in questi ultimi due casi le percentuali rimanenti corrispondono alla voce *abbastanza*.

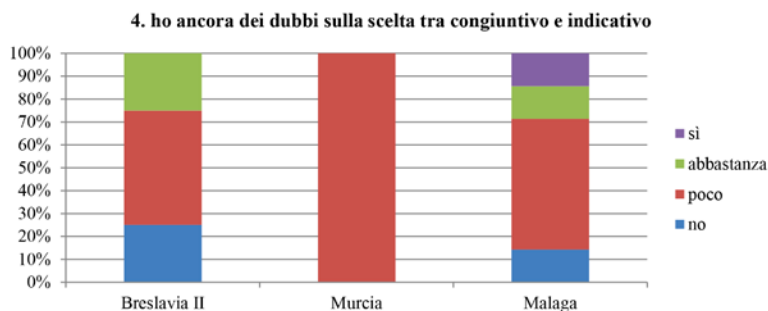
Grafico 6.20: Percentuali delle risposte date in tutte le classi di sperimentazione al terzo quesito sulla percezione personale di acquisizione dei contenuti



Infine dal grafico 6.21 emerge che la maggioranza dei partecipanti in tutte le classi ha ancora dubbi nella scelta tra congiuntivo e indicativo: Breslavia: *no* 25%, *poco* 50% e *abbastanza* 25%; Murcia 100% *poco* e

Malaga 57% *poco*, mentre le restanti voci hanno totalizzato ciascuna il 14,3%.

Grafico 6.21: Percentuali delle risposte date in tutte le classi di sperimentazione al quarto quesito sulla percezione personale di acquisizione dei contenuti



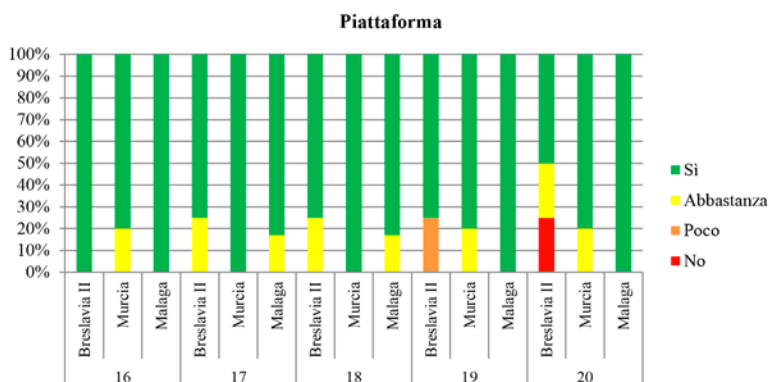
Pertanto, sono ancora presenti dei dubbi sui casi in cui la grammatica prescrive nello standard l'uso del congiuntivo, sulla scelta tra indicativo e congiuntivo e sulla distinzione tra situazioni formali e informali, mentre invece appare chiaro che in alcuni contesti informali è accettabile non usare il congiuntivo. Dunque, le attività svolte sono state generalmente apprezzate e il tema generale dell'unità è stato compreso, sebbene necessiti ancora di ulteriore pratica e riflessione, come confermano le risposte libere date nella sezione dedicata ai commenti personali, per i quali si è conservata la grafia originaria nella trascrizione: *Ho bisogno di praticare di più; Bisogno di imparare dei verbi irregolari; Tutto è bene* (Murcia); *È più facile per lo studio; Tutto ha stato bene molto facile a comprendere* (Malaga).

6.4.4. Unità didattica sul congiuntivo: opinioni sulla piattaforma Wizer.me

Segue la trattazione dei dati relativi ai punti 16-20 del questionario sulle opinioni dei partecipanti sull'uso della piattaforma: 16) Il login e l'accesso alle diverse parti dell'unità è stato facile; 17) Le istruzioni delle attività erano comprensibili; 18) Le attività si visualizzano bene sullo schermo; 19) È stato utile avere esercizi con correzione automatica; 20) Mi è piaciuta l'idea delle attività online. Dal grafico 6.22 emerge la generale soddisfazione degli utenti, le cui percentuali di risposta si

dividono, nella quasi totalità dei casi, tra i descrittori *sì* e *abbastanza*. A conferma di ciò nella classe di Malaga un utente ha lasciato il seguente commento: *Mi piacciono le attività online perché sono più dinamiche. Solo nel caso del punto 19. È stato utile avere esercizi con correzione automatica* e del punto 20. *Mi è piaciuta l'idea delle attività online* nella classe di Breslavia una percentuale ridotta, pari in entrambi i casi al 25%, è stata attribuita rispettivamente ai descrittori *poco* e *no*, tuttavia nella sezione dei commenti liberi non sono state date spiegazioni in proposito.

Grafico 6.22: Panoramica delle percentuali delle risposte date in tutte le classi di sperimentazione ai quesiti sull'utilizzo della piattaforma



Dall'analisi di questi dati emerge che i partecipanti alla sperimentazione hanno comunicato il proprio apprezzamento per l'operato della docente-ricercatrice e hanno fornito anche un riscontro sulle attività svolte, dando dei suggerimenti per il miglioramento dello svolgimento della lezione, della fruizione della piattaforma dal punto di vista della grafica e della struttura delle attività.

CONCLUSIONI

Il tema di questo lavoro si intreccia strettamente con la questione della ricerca di uno standard linguistico, una questione ricorrente nella storia dell'italiano e che perdura almeno dall'Unità d'Italia. Come esposto nei capitoli 1 e 2, la ristrutturazione del sistema linguistico italiano ha fatto emergere a partire dalla seconda metà del Novecento una nuova varietà che "si candida ad occupare, dopo secoli di ostracismo, il baricentro dell'intero sistema linguistico italiano o per lo meno a condividere con lo standard ufficiale il crisma della norma" (Sabatini, 1985, p. 175). Alla descrizione dei tratti distintivi di tale varietà linguistica si affianca anche la necessità di definire quali di questi fenomeni grammaticali siano da considerarsi effettivamente costitutivi di un nuovo standard linguistico e, di conseguenza, quali di essi vadano trasmessi nell'insegnamento dell'italiano a stranieri.

È proprio dal divario tra l'italiano codificato nelle grammatiche e l'italiano parlato che nasce la necessità per l'insegnante di italiano di fare riferimento ad una norma e, allo stesso tempo, la difficoltà di come

stabilirla, dati i continui cambiamenti nella lingua e le diverse posizioni adottate dagli studiosi circa quali fenomeni siano da considerarsi parte del nuovo standard.

Questa ricerca si inserisce nell'ambito di questa riflessione e, sulla base dei principali studi condotti sul neostandard e citati nel corso della trattazione, propone un possibile approccio per l'insegnamento dell'italiano, che consiste nel presentare i fenomeni grammaticali del neostandard come segnali di un cambiamento tendenziale dell'italiano (in accordo con quanto proposto da Sobrero, 1992), avendo cura però di fare una distinzione basata principalmente sulla funzionalità e, in seconda battuta, sull'accettabilità sociale.

Per alcuni fenomeni il distacco dalla norma grammaticale tradizionale produce un effettivo apporto funzionale alla lingua, poiché senza di essi non sarebbe possibile esprimere un certo concetto o raggiungere un determinato scopo pragmatico; altri, invece, abbassano il registro linguistico senza apportare di fatto alcun vantaggio sul piano funzionale, poiché sono solo un'alternativa arbitraria a forme standard perfettamente funzionanti (in accordo con quanto proposto da Prandi, 2020). I fenomeni che rientrano nella prima categoria andrebbero trasmessi nell'insegnamento come parte integrante della nuova varietà standard; tutti gli altri possono essere menzionati come varianti sub-standard, seppur diffuse (cfr. capitolo 2).

Questa proposta non nega l'importanza dell'elaborazione di concetti come "italiano neostandard" (cfr. Berruto, 1987) o "italiano dell'uso medio" (cfr. Sabatini, 1985), ossia la revisione dello standard ereditato dai secoli passati, un'esigenza più volte affiorata nel corso della storia. Tuttavia, se da un lato le descrizioni di Francesco Sabatini e di Gaetano Berruto hanno avuto il pregio di documentare anche dal punto di vista storico i cambiamenti avvenuti nell'italiano, dall'altro includere nelle liste tutto ciò che si distaccava dalle norme grammaticali ha portato di fatto all'accostamento di risorse funzionali essenziali a forme che invece non apportano alcun vantaggio funzionale e che producono solo una caduta di registro. Ne deriva che persistono ancora dei dubbi dal punto di vista della messa a punto di uno standard dell'uso comune che dia sicurezza linguistica ai parlanti e un modello definito a chi ne valuta gli usi.

Rispetto alle pubblicazioni già citate sul tema del neostandard, che

principalmente si focalizzano solo sulla frequenza d'uso o sulla percezione di correttezza da parte dei parlanti, questa ricerca propone dei criteri la cui applicazione permette di non cadere né nell'errore di rifiutare in blocco tutto ciò che differisce dalla norma prescrittiva, né nell'errore opposto, ossia di ritenere che tutti gli errori di oggi saranno le regole di domani in base ai soli dati sulla frequenza o sulla preferenza dei parlanti.

L'applicazione dei criteri della funzionalità e dell'accettabilità sociale rappresenta il punto di partenza da cui è possibile determinare se le forme linguistiche dell'uso medio appartengano a uno standard condiviso e insegnabile. Questo schema può essere applicato non solo ai fenomeni elencati nel 1990 da Sabatini, ma anche a qualunque innovazione linguistica che entri nell'uso e si candidi a entrare nello standard sulla base della sua diffusione tra i parlanti.

La diffusione statistica di una struttura linguistica – sia essa già presente nel sistema o un'innovazione – ci dice solo che essa dev'essere presa in considerazione, il che non implica affatto che il suo uso sia socialmente accettabile e consigliabile. Infatti, basarsi unicamente sul criterio della diffusione statistica porterebbe a considerare come parte integrante dello standard qualunque struttura linguistica ampiamente diffusa e, nel fare ciò, ad osservare la questione solo dal punto di vista dei parlanti che la utilizzano, tralasciando il punto di vista del linguista e dell'insegnante di lingue, ossia: queste strutture linguistiche danno un effettivo apporto funzionale alla lingua? Il loro uso evita un involontario abbassamento del registro linguistico e l'impressione di una scarsa conoscenza della lingua? E quindi sarebbe davvero consigliabile insegnarle?

Di conseguenza, un possibile approccio nell'insegnamento della lingua potrebbe essere quello di affrontare sia le strutture grammaticali "ritrovate" – per usare la scelta lessicale di Sabatini – sia le più recenti innovazioni linguistiche presentandole come diffuse tra i parlanti nativi e come segnali di un possibile cambiamento tendenziale dell'italiano (in accordo con quanto già proposto da Sobrero, 1992b). Nel fare ciò, sarà necessario distinguere le strutture che deviano dalla norma grammaticale tradizionale dando un effettivo apporto funzionale alla lingua e quelle che, per quanto siano diffuse, di fatto non apportano alcun vantaggio funzionale e rappresentano una deviazione del tutto arbitra-

ria rispetto alla norma, il cui uso produce solo un calo di registro ed espone il parlante alla censura sociale (Prandi, 2020).

Solo in una fase più avanzata del processo di acquisizione della lingua l'apprendente potrà, eventualmente, scegliere consapevolmente di deviare dalla norma linguistica sulla base di una valutazione della situazione comunicativa in cui si trova, del suo livello di (in)formalità, dei suoi interlocutori e del registro linguistico. Fino ad allora, è necessario fornire un modello linguistico sicuro anche attraverso i materiali didattici in uso.

Da questo punto di vista, un aspetto innovativo della progettazione delle unità didattiche è l'aver fornito materiali utili ad un ripasso dell'uso standard di due tratti grammaticali, all'interno del quale si inserisce, come elemento di novità, un approfondimento dell'uso neostandard di tali tratti grammaticali.

Dall'analisi incrociata degli esiti delle attività di verifica di ciascuna unità con alcuni dati personali raccolti tramite il questionario sulle opinioni dei partecipanti è emerso quanto segue. Le classi di sperimentazione che hanno svolto con migliore profitto le attività conclusive delle due unità sull'uso neostandard sono, in generale, quelle che presentano una più lunga durata pregressa dello studio della lingua italiana, a cui si aggiunge la presenza nella L1 o nelle LS conosciute di un fenomeno grammaticale simile a quello affrontato nell'unità. Quest'ultimo fattore si è verificato nel caso dell'unità sui riflessivi, sia per il polacco che per le varie L1 romanze dei partecipanti coinvolti (principalmente spagnolo e francese). Ciò consente di ipotizzare un possibile rapporto tra la L1 e/o le LS conosciute dai partecipanti e i risultati da loro ottenuti.

Sebbene dai dati raccolti emerga che i profitti migliori si accompagnino sempre ad una lunga durata dello studio pregresso della lingua italiana, anche nel caso in cui il fenomeno grammaticale oggetto di studio non ha un corrispettivo effettivo nella L1 (ad esempio, il congiuntivo non esiste in polacco, ma la verifica finale dedicata a questo argomento è stata svolta con successo dalla classe del secondo anno di Breslavia), non è possibile affermare il contrario: in assenza di una lunga durata pregressa dello studio dell'italiano la sola analogia tra l'italiano e la L1 e/o le diverse lingue romanze conosciute sembra non costituire un fattore determinante per lo svolgimento positivo delle attività. Ciò è

emerso in particolare nella classe di Madrid, dove a fronte della più breve durata dello studio pregresso e della più ampia varietà di L1/LS romanze in aula (L1: spagnolo e rumeno; LS: francese, catalano, galiziano, portoghese e valenzano) è stata totalizzata la percentuale più bassa di risposte corrette tra tutte le classi di sperimentazione.

Per quanto riguarda, invece, le conoscenze pregresse specifiche sul tema, la maggior parte dei partecipanti delle classi di sperimentazione ha dichiarato di aver già incontrato nella vita quotidiana alcuni esempi dell'uso neostandard degli argomenti trattati nelle unità, senza tuttavia avere alcuna cognizione teorica che ne spiegasse il funzionamento.

Tutto ciò permette di ipotizzare un possibile rapporto tra la durata pregressa dello studio, la L1 e le LS conosciute dai partecipanti e i risultati positivi da loro ottenuti, ove il primo fattore sembra avere un peso maggiore sulla riuscita delle attività, anche in assenza di analogie tra la L1 e/o le diverse lingue romanze conosciute.

Sulla base di quanto esposto finora, è possibile trarre le seguenti conclusioni. Il contributo delle unità didattiche progettate si è concretizzato nell'aver fornito un inquadramento teorico non solo circa l'uso standard di alcuni fenomeni grammaticali – in taluni casi come primo approccio, in altri come integrazione e potenziamento di un argomento già parzialmente noto – ma anche riguardo al loro uso neostandard, che sino ad allora era percepito dai partecipanti come errato o inspiegabile, essendo in contrasto con la norma linguistica che era stata loro trasmessa.

Pertanto una teorizzazione linguistica e una pratica didattica che tenga conto delle riflessioni condivise in questa ricerca permetterà a chi insegna e a chi apprende la lingua italiana innanzitutto di disporre di un modello linguistico sicuro e di criteri con cui valutare l'appartenenza al neostandard sia delle forme linguistiche già attestate, sia di tutte le innovazioni linguistiche che entreranno nell'uso. In secondo luogo, una progettazione di materiali didattici che tenga conto anche dei fenomeni linguistici del neostandard può supportare efficacemente la formazione degli apprendenti di lingua italiana L2/LS (ma anche degli italofoni nativi), favorendo una percezione realistica delle stratificazioni sociolinguistiche della lingua in uso.

BIBLIOGRAFIA

- Acquaviva P. (2000), *La grammatica italiana: il lavoro comincia adesso*, «Lingua e stile», 2, pp. 249-272.
- Aijmer K. (2009, a c. di), *Corpora and Language Teaching*, John Benjamins, Amsterdam-Philadelphia.
- AIL: Accademia Italiana di Lingua (2015), *AIL presentazione ed introduzione. 6 esami di italiano, da A1 a C2, e 2 esami di italiano commerciale per tutti i livelli di conoscenza*, Firenze. Disponibile online: <http://www.acad.it/downloads/sillabo.html>.
- Alhujaylan H. (2014), *Teaching collocations and lexical phrases: A data driven learning approach*, contributo presentato alla 11a conferenza internazionale su *Teaching and Language Corpora* (TaLC11), Lancaster University, 20-23 luglio 2014. Disponibile online: <http://ucrel.lancs.ac.uk/talc2014/doc/TALC2014-abstract-book.pdf>.
- Alisova T. (1972), *Strutture semantiche e sintattiche della proposizione semplice in italiano*, Sansoni, Firenze.

- Ammon U. (1986), *Explikation der Begriffe 'Standardvarietät' und 'Standardsprache' auf normtheoretischer Grundlage*, in *Sprachlicher Substandard. Vol. I*, a c. di G. Holtus, E. Radtke (1986-1990), Niemeyer, Tübingen, pp. 1-62.
- Ammon U. (2003), *On the social forces that determine what is standard in a language and on conditions of successful implementation*, «Sociolinguistica», 17, pp. 1-10.
- Amor S. (2002), *Authenticity and Authentication in Language Learning: Distinctions, Orientations, Implications*, Peter Lang, Frankfurt.
- Andorno C. (2003), *La grammatica italiana*, Mondadori, Milano.
- Andorno C., Bernini G., Giacalone Ramat A., Valentini A. (2003), *Sintassi*, in *Verso l'italiano*, a c. di A. Giacalone Ramat, Carocci, Roma, pp. 116-127.
- Antonelli G. (2007), *L'italiano nella società della comunicazione*, Il Mulino, Bologna.
- Antonelli G. (2011), *Lingua*, in *Modernità italiana. Cultura, lingua e letteratura dagli anni Settanta a oggi*, a c. di A. Acribo, E. Zinato, Carocci, Roma, pp. 15-52.
- Antonelli G. (2016), *Un italiano vero. La lingua in cui viviamo*, Milano, Rizzoli.
- Arcangeli M., Giugni S., Cardillo G., D'Angelo K., Ferrari S., Greco N., Lucatorto A. (2014), *Attestato ADA. Piano dei corsi*, Alma & Società Dante Alighieri, Firenze.
- Asor Rosa A., Serianni L., Trifone P. (1994, a c. di), *Storia della lingua italiana. Vol. I: I luoghi della codificazione; vol. II: Scritto e parlato; vol. III: Le altre lingue*, Einaudi, Torino.
- Aston G., Bernardini S., Stewart D. (2004, a c. di), *Corpora and language learners*, John Benjamins, Amsterdam-Philadelphia.
- Babych S. (2012), *Exploring Patterns of Textual Cohesion in Multilingual Corpora and their Application for Teaching Language and Translation*, contributo presentato alla 10a conferenza internazionale su *Teaching and Language Corpora* (TaLC10), Uniwersytet Warszawski, 11-14 luglio 2012. Disponibile online: <http://talc10.ils.uw.edu.pl/files/2011/09/BABYCHSvitlana1.pdf>.
- Baker M. (1993), *Corpus linguistics and translation studies: implications and applications*, in *Text and Technology: In Honour of John Sinclair*,

- a c. di M. Baker, G. Francis e E. Tognini-Bonelli, John Benjamins, Amsterdam, pp. 233-352.
- Baker M. (1995), *Corpora in translation studies: an overview and some suggestions for future research*, «Target», 7/2, pp. 223-243.
- Baker M. (1999), *The role of corpora in investigating the linguistic behaviour of professional translators*, «International Journal of Corpus Linguistics», 4/2, pp. 281-298.
- Baker P., Hardie A., McEnery T. (2006), *A Glossary of Corpus Linguistics*, Edinburgh University Press, Edinburgh.
- Balboni P.E. (1994), *Didattica dell'italiano a stranieri*, Bonacci, Roma.
- Balboni P.E. (1999), *Dizionario di glottodidattica*, Guerra Edizioni, Perugia.
- Balboni P.E. (2002), *Le sfide di Babele. Insegnare le lingue nelle società complesse*, UTET, Torino.
- Balboni P.E. (2008), *Imparare le lingue straniere*, Marsilio, Venezia.
- Bambini V., Trevisan M. (2012), *EsploraCoLFIS: Un'interfaccia web per le ricerche sul Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS)*, Quaderni del Laboratorio di Linguistica, 11, Scuola Normale Superiore, Pisa, pp. 1-16.
- Banfi E. (2003 a c. di), *Italiano di cinesi. Percorsi acquisizionali*, Franco Angeli, Milano.
- Banfi E., Bernini G. (2003), *Il verbo*, in *Verso l'italiano. Percorsi e strategie di acquisizione*, a c. di A. Giacalone Ramat, Carocci, Roma, pp. 70-115.
- Barbera M. (2007), *Per la storia di un gruppo di ricerca. Tra bmanuel.org e corpora.unito.it*, in *Corpora e linguistica in Rete*, a c. di M. Barbera, E. Corino, C. Onesti, Guerra Edizioni, Perugia, pp. 3-20.
- Barbera M. (2013, a c. di), *Una introduzione ai NUNC: storia della creazione di un corpus*, in *Molti occhi sono meglio di uno: saggi di linguistica generale 2008-12*, Qu.A.S.A.R., Milano, pp. 97-114.
- Barbera M. (2015, a c. di), *La linguistica dei corpora in Italia all'alba del terzo millennio*, in *Quanto più la relazione è bella: saggi di storia della lingua italiana 1999-2014*, bmanuel.org. glottologica et philologica-series maior, Youcanprint, Torino, pp. 117-138. Disponibile online: https://iris.unito.it/retrieve/handle/2318/1670619/420144/Barbera_Relazione%3dSaggiStoriaLinguaItaliana_2015.pdf.

- Barbera M., Marellò C. (2008), *Tra scritto-parlato, Umgangssprache e comunicazione in rete: i corpora NUNC*, «Studi di Grammatica Italiana», 27 (2008, recte 2011), a c. di A. Antonini, S. Stefanelli. Atti del convegno *Per Giovanni Nencioni. Convegno Internazionale di Studi* (Pisa-Firenze, 4-5 maggio 2009), Le Lettere, Firenze, pp. 157-185.
- Barni M., Bandini A., Sprugnoli L., Lucarelli S., Scaglioso A.M., Strambi B., Fusi C., Arruffoli A.M. (2009). *Linee guida CILS. Certificazione di Italiano come Lingua Straniera* (presentazione di M. Vedovelli), Guerra Edizioni, Perugia. Disponibile online: https://cils.unistrasr.it/public/articoli/52/Files/linee_guida_cils_pdf.pdf.
- Baroni M. (2010), *Corpora di italiano*, in *Enciclopedia dell'italiano*, a c. di R. Simone, G. Berruto, P. D'Achille (2010-2011), Istituto della Enciclopedia Italiana, Roma, pp. 300-303. Disponibile online: http://www.treccani.it/enciclopedia/corpora-di-italiano_%28Enciclopedia-dell%27Italiano%29/.
- Baroni M., Bernardini S. (2004), *BootCaT: Bootstrapping corpora and terms from the web*, in *Proceedings of Fourth International Conference on Language Resources and Evaluation (LREC 2004)*. Atti del convegno (Lisboa, 26-28 maggio 2004), ELRA - European Language Resources Association, Paris, pp. 1313-1316.
- Baroni M., Bernardini S. (2006, a c. di), *Wacky! Working papers on the web as corpus*, Gedit, Bologna. Disponibile online: <http://wacky-book.sslmit.unibo.it/wackycontents.html>.
- Baroni M., Bernardini S., Comastri F., Piccioni L., Volpi A., Aston G., Mazzoleni M. (2004), *Introducing the La Repubblica corpus: A large, annotated, TEI(XML)-compliant corpus of newspaper Italian*, in *Proceedings of Fourth International Conference on Language Resources and Evaluation (LREC 2004)*. Atti del convegno (Lisboa, 26-28 maggio 2004), ELRA - European Language Resources Association, Paris, pp. 1771-1774.
- Baroni M., Bernardini S., Ferraresi A., Zanchetta E. (2009), *The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-crawled Corpora*, «Journal of Language Resources and Evaluation», 43/3, pp. 209-226.
- Baroni M., Ueyama M. (2006), *Building general- and special-purpose corpora by Web crawling*, in *Proceedings of the NIJL International*

- Workshop on Language Corpora* (NIJL, 2006). Atti del convegno (Tokyo, marzo 2006), pp. 31-40.
- Bazzanella C. (2014), *Linguistica cognitiva. Un'introduzione*, Laterza, Roma-Bari.
- Begotti P. (2011), *Didattizzazione di materiali autentici e analisi di manuali di italiano per stranieri*, FILIM - Formazione degli Insegnanti di Lingua Italiana nel Mondo. Disponibile online: https://www.italy.it/sites/default/files/Filim_didattizzazione_analisi_teoria.pdf.
- Begotti P., (2006), *Didattizzazione di materiali autentici e analisi dei manuali di italiano per stranieri*, modulo telematico in ambito del Laboratorio Itals, Venezia. Disponibile online: https://www.italy.it/sites/default/files/Filim_didattizzazione_analisi_teoria.pdf.
- Bembo P. (1989[1966]), *Prose della volgar lingua. Gli Asolani*. Rime, a c. di C. Dionisotti, Tea, Torino.
- Benincà P., Salvi G., Frison L. (1988), *L'ordine degli elementi della frase e le costruzioni marcate*, in *Grande grammatica italiana di consultazione. Vol. I: La frase. I sintagmi nominale e preposizionale*, a c. di L. Renzi, Salvi G., Cardinaletti A., Il Mulino, Bologna, pp. 129-239.
- Benko V. (2016), *Two years of Aranea: Increasing counts and tuning the pipeline*, in *Proceedings of the 10th LREC Conference* (LREC 2016). Atti del convegno (Portorož, 23-28 maggio 2016), ELRA - European Language Resources Association, Paris, pp. 4245-4248.
- Benko V. (2014a), *Compatible Sketch Grammars for Comparable Corpora*, in *Proceedings of the XVI EURALEX International Congress: The User In Focus*, a c. di A. Abel, C. Vettori, N. Ralli. Atti del convegno (Bolzano/Bozen, 15-19 luglio 2014), Eurac Research, Bolzano-Bozen, pp. 417-430.
- Benko V. (2014b), *Aranea: Yet Another Family of (Comparable) Web Corpora*, in *Text, Speech and Dialogue. Proceedings of the 17th International Conference* (TSD 2014), a c. di P. Sojka, A. Horák, I. Kopeček, K. Pala. Atti del convegno (Brno, 8-12 settembre 2014), Springer International Publishing, Basel, pp. 257-264.
- Benucci A. (2007, a c. di), *Sillabo di Italiano per Stranieri. Una proposta del Centro Linguistico dell'Università per Stranieri di Siena*, Guerra Edizioni, Perugia.
- Berisso M. (2000), *Livelli linguistici e soluzioni stilistiche. Sondaggi sulla nuova narrativa italiana 1991-1998*, «Lingua e Stile», 3, pp. 471-504.

- Berretta M. (1986), *Riprese anaforiche e tipi di testo: il monologo espositivo*, in *Parallela2. Aspetti della sintassi dell'italiano contemporaneo*, a c. di K. Lichem, E. Mara, S. Knaller, Narr, Tübingen, pp. 47-59.
- Berretta M. (1987), *Bricolage di parole in lingua straniera*, «IeO», 5, pp. 222-226.
- Berretta M. (1992), *Deissi e anafora nella conversazione*, in *Grammatica del parlare e dell'ascoltare a scuola*, a c. di L. Brasca, M.L. Zambelli, La Nuova Italia, Firenze, pp. 13-31.
- Berretta M. (1993), *Morfologia*, in *Introduzione all'italiano contemporaneo; vol. I: Le strutture; vol. II: La variazione e gli usi* a c. di A.A. Sobrero, Laterza, Roma-Bari, pp. 193-245.
- Berretta M. (1995), *Ordini marcati dei costituenti maggiori di frase: una rassegna*, «Linguistica e filologia», 1, pp. 125-170.
- Berretta M. (2002), *Quello che voglio dire è che: le scisse da struttura topicalizzanti a connettivi testuali*, in *La parola al testo. Scritti per Bice Mortara Garavelli*, a c. di G.L. Beccaria, C. Marelli, Edizioni dell'Orso, Alessandria, pp. 15-31.
- Berruto G. (1983), *L'italiano popolare e la semplificazione linguistica*, «Vox Romanica», 42, pp. 38-79.
- Berruto G. (1985a), *Per una caratterizzazione del parlato: l'italiano parlato ha un'altra grammatica?*, in *Geschprochenes Italienisch in Geschichte und Gegenwart*, a c. di G. Holtus, E. Radtke, Narr, Tübingen, pp. 120-53.
- Berruto G. (1985b), *'Dislocazioni a sinistra' e 'grammatica' dell'italiano parlato*, in *Sintassi e morfologia della lingua italiana d'uso. Teorie ed applicazioni descrittive*, a c. di A. Franchi De Bellis, L.M. Savoia, Bulzoni, Roma, pp. 59-82.
- Berruto G. (1986), *Le dislocazioni a destra in italiano*, in *Tema-Rema in Italiano*, a c. di H. Stammmerjohann, Narr, Tübingen, pp. 55-69.
- Berruto G. (1987), *Sociolinguistica dell'italiano contemporaneo*, Carocci, Roma.
- Berruto G. (1990), *Semplificazione linguistica e varietà sub-standard*, in *Sprachlicher Substandard. Vol. III*, a c. di G. Holtus, E. Radtke (1986-1990), Niemeyer, Tübingen, pp. 17-43.
- Berruto G. (1993a), *Le varietà del repertorio*, in *Introduzione all'italiano contemporaneo. Vol. II*, a c. di A.A. Sobrero, Laterza, Roma-Bari, pp. 3-36.

- Berruto G. (1993b), *Varietà diamesiche, diastratiche, diafasiche*, in *Introduzione all'italiano contemporaneo. Vol. II*, a c. di A.A. Sobrero, Laterza, Roma-Bari, pp. 37-92.
- Berruto G. (2010), *Italiano standard*, in *Enciclopedia dell'italiano*, a c. di R. Simone, G. Berruto, P. D'Achille (2010-2011). Disponibile online: <https://www.treccani.it/enciclopedia/ricerca/italiano-standard/>.
- Berruto G. (2011a), *Substandard*, in *Enciclopedia dell'italiano*, a c. di R. Simone, G. Berruto, P. D'Achille (2010-2011). Disponibile online: https://www.treccani.it/enciclopedia/substandard_%28Enciclopedia-dell%27Italiano%29/.
- Berruto G. (2011b), *Varietà*, in *Enciclopedia dell'italiano*, in Simone R., Berruto G., D'Achille P., (2010-1, a c. di). Disponibile online: https://www.treccani.it/enciclopedia/variet%C3%A0_%28Enciclopedia-dell%27Italiano%29/.
- Berruto G. (2012 [1987]), *Sociolinguistica dell'italiano contemporaneo*, La Nuova Italia Scientifica, Roma, 1987, Carocci, Roma, 2012.
- Berruto G. (2012), *Sul parlante nativo (di italiano)*, in Idem, *Saggi di sociolinguistica e linguistica*, a c. di G. Bernini, B. Moretti, S. Schmid, T. Telmon, Edizioni dell'Orso, Alessandria, pp. 87-99.
- Berruto G., Cerruti M. (2014), *Manuale di sociolinguistica*, UTET, Torino.
- Bettoni C., Di Biase B. (2005), *Sviluppo obbligatorio e progresso morfologico-sintattico. Un caso di Processabilità in italiano L2*, «ITALS: Didattica e linguistica dell'italiano come lingua straniera», a c. di P. E. Balboni, M. Santipolo, 3/7, pp. 27-48.
- Bettoni C., Di Biase, B. (2015), *Processability theory: theoretical bases and universal schedules*, in *Grammatical Development in Second Languages: Exploring the Boundaries of Processability Theory*, a c. di C. Bettoni, B. Di Biase, The European Second Language Association, Amsterdam, pp. 19-80.
- Biber D. (2012), *Corpus-Based and Corpus-driven Analyses of Language Variation and Use*, in *The Oxford Handbook of Linguistic Analysis (1 ed.)*, a c. di B. Heine e H. Narrog. Disponibile online: <https://eclass.uoa.gr/modules/document/file.php/MEDIA139/Readings%20//%20%CE%91%CE%BD%CE%B1%CE%B3%CE%B-D%CF%8E%CF%83%CE%BC%CE%B1%CF%84%CE%B1/>

- Biber%2C%20Corpus-Based%20and%20Corpus-Driven%20Analyses%20of%20Language%20Variation%20and%20Use.pdf*.
- Biber D., Conrad S., Reppen R. (1998), *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge University Press, Cambridge.
- Bolasco S. (1999), *Analisi multidimensionale dei dati. Metodi, strategie e criteri d'interpretazione*, Carocci, Roma.
- Bolwler L., Pearson J. (2002), *Working with Specialized Language. A practical Guide to Using Corpora*, Routledge, London-New York.
- Bonomi I. (1996), *La narrativa e l'italiano dell'uso medio*, «Studi di Grammatica Italiana», 16, pp. 321-338.
- Bonomi I. (1997), *La grammatica del parlato in alcuni scrittori contemporanei*, in *Norma e lingua in Italia: alcune riflessioni fra passato e presente*, Istituto Lombardo di Scienze e Lettere, Milano, pp. 167-183.
- Bonomi I. (2002), *L'italiano giornalistico. Dall'inizio del '900 ai quotidiani on line*, Cesati, Firenze.
- Bonomi I., Mauroni E. (2003), *L'innovazione grammaticale in testi scritti di fine millennio*, in *Italia linguistica anno Mille. Italia linguistica anno Duemila*, a c. di N. Maraschio, T. Poggi Salani. Atti del XXXIV Congresso Internazionale di Studi della SLI (Firenze, 19-21 ottobre 2000), Bulzoni, Roma, pp. 491-519.
- Bonvino E. (2004), *I materiali didattici per l'insegnamento dell'italiano L2: criteri di selezione*, ICON. Disponibile online: <https://languageadvisor.net/wp-content/uploads/2022/03/Criteri-di-selezione-materiali-didattici.pdf>.
- Borghetti C., Castagnoli S., Brunello M. (2011), *I generi del web tra tradizione e innovazione: un'analisi linguistica sulla base del corpus PAISÀ*, in *Scritto e parlato, formale e informale: La comunicazione mediata dalla rete*, a c. di M. Cerruti, E. Corino, C. Onesti, Carocci, Roma, pp. 147-170.
- Bortolini U., Pizzuto E. (1997, a c. di), *Il progetto CHILDES-Italia, Contributi di ricerca sulla lingua italiana*, Edizioni del Cerro, Tirrenia.
- Bortolini U., Zampolli A. (1971), *Lessico di frequenza della lingua italiana contemporanea: prospettive metodologiche*, in *L'insegnamento dell'italiano in Italia e all'Estero*, a c. di M. Medici, R. Simone.

- Atti del IV Convegno internazionale di Studi (Roma, 1-2 giugno 1970), II, Bulzoni, Roma, pp. 639-648.
- Breen M.P. (1985), *Authenticity in the language classroom*, «Applied linguistics», 6, pp. 60-70.
- Brezina V., Gablasova D., Marín Cervantes I. (2018), *Using corpora to teach sociolinguistics: A practical workshop*, workshop tenuto alla 13a conferenza internazionale su Teaching and Language Corpora (TaLC13), Cambridge University, 18-21 luglio 2018. Disponibile online: https://www.researchgate.net/publication/326548400_13th_Teaching_and_Language_Corpora_Conference_TaLC_2018_Book_of_Abstacts.
- Bresnan J. (2001), *Lexical-functional syntax*, Blackwell, Malden (MA).
- Bruni F. (1984), *L'italiano. Elementi di storia della lingua e della cultura*, UTET, Torino.
- Buchholz S., Marsi E. (2006), *CoNLL-X Shared Task on Multilingual Dependency Parsing*, in *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X '06)*. Atti del convegno (New York 8-9 giugno 2006), Association for Computational Linguistics, Stroudsburg, pp. 149-164.
- Budassi M. (2017), *Estrazione del 'che' polivalente da un corpus POS-tagato: Limiti e possibilità*, «Analisi Linguistica e Letteraria», 25, pp. 147-172.
- Budinich L. (2016), *La variazione semasiologica di sottocodice: un esempio di analisi lessicale corpus-based*, «CHIMERA. Romance Corpora and Linguistic Studies», 3/1, pp. 23-56.
- Calleri D., Chini M., Cordin P., Ferraris P. (2003), *Confronti tra l'acquisizione di italiano L1 e l'acquisizione di italiano L2*, in *Verso l'italiano*, a c. di A. Giacalone Ramat, Carocci, Roma, pp. 220-253.
- Cardillo G., Vecchio P. (2015a, a c. di), *Nuovo sillabo della Certificazione PLIDA. Livelli A1-C2*, Società Dante Alighieri, Roma. Disponibile online: https://www.almaedizioni.it/media/upload/pdf_vari/novosillaboplida2015.pdf.
- Cardillo G., Vecchio P. (2015b a c. di), *Livello PLIDA B1 - Quaderno delle specifiche*, Società Dante Alighieri, Roma.
- Cardillo G., Vecchio P. (2016 a c. di), *Livello PLIDA B2 - Quaderno delle specifiche*, Società Dante Alighieri, Roma.
- Cardillo G., Vecchio P. (2017 a c. di), *Livello PLIDA C1 - Quaderno delle specifiche*, Società Dante Alighieri, Roma.

- Cardillo G., Vecchio P. (2018 a c. di), *Livello PLIDA A2 - Quaderno delle specifiche*, Società Dante Alighieri, Roma.
- Cardinale U. (2011), *Manuale di scrittura giornalistica*, UTET Università, Torino.
- Castellani A. (1991), *Italiano dell'uso medio o italiano senz'aggettivi?*, «Studi Linguistici Italiani», 17, pp. 233-256.
- Charkova K., Charkova D., *A linguistically informed, systematic, and in-depth approach to developing data driven language learning materials*, contributo presentato alla 11a conferenza internazionale su *Teaching and Language Corpora* (TaLC11), Lancaster University, 20-23 luglio 2014. Disponibile online: <http://ucrel.lancs.ac.uk/talc2014/doc/TALC2014-abstract-book.pdf>.
- Chiappini L., De Filippo N. (2014), *Congiuntivo che passione!*, Bonacci Editore, Formello.
- Disponibile online: https://cdn3.scuolabook.it/Uploaded/loescher_b4442e_preview/b4442e_preview.pdf.
- Chini M. (2012b), *Linguistica educativa e linguistica acquisizionale: qualche suggerimento dalla ricerca su L2*, in *Linguistica educativa*, a c. di S. Ferreri, Bulzoni, Roma, pp. 123-140.
- Chini M. (2005), *Che cos'è la linguistica acquisizionale*, Carocci, Roma.
- Chini M. (2011), *Qualche riflessione sulla didattica di L2 ispirata alla recente ricerca acquisizionale*, in «Italiano LinguaDue», 3/2, pp. 1-22. Disponibile online: <http://riviste.unimi.it/index.php/promoitals/article/view/1912>.
- Chini M. (2012a), *Alcune ricadute didattiche degli studi acquisizionali sull'italiano L2*, in «Rassegna Italiana di Linguistica Applicata. Numero monografico L'italiano L2 nella scuola secondaria di secondo grado», a cura di C. Bosio, B. Cambiaghi, M. Zanola, 44/1, pp. 69-91.
- Chini M. (2012c), *Spunti per la didattica linguistica a partire da alcuni recenti filoni di studio di L2*, in (Far) *apprendere, usare e certificare una lingua straniera. Studi in onore di Serena Ambroso*, a c. di E. Bonvino, E. Luzi, A.R. Tamponi, Bonacci, Roma, pp. 61-74.
- Chini M. (2016), *Elementi utili per una didattica dell'italiano L2 alla luce della ricerca acquisizionale*, «Italiano LinguaDue», 8/2, pp. 1-18. Disponibile online: <https://riviste.unimi.it/index.php/promoitals/article/view/8172>.

- Christ O. (1994a), *The IMS Corpus Workbench Technical Manual*, Institut für maschinelle Sprachverarbeitung, Universität Stuttgart.
- Christ O. (1994b), *The XKwic User Manual*, Institut für maschinelle Sprachverarbeitung, Universität Stuttgart.
- Cignetti L. (2006), *L'ordine delle parole nello scritto e nel parlato (con alcune osservazioni sul fenomeno della "doppia dislocazione"*, in *Parole frasi testi tra scritto e parlato*, a c. di A. Ferrari, «Cenobio», 55/3, pp. 207-214.
- Cignetti L. (2010), *Dativo etico*, in *Enciclopedia dell'italiano*, a c. di R. Simone, G. Berruto, P. D'Achille (2010-2011), Disponibile online: <https://www.treccani.it/enciclopedia/ricerca/Dativo-etico/>.
- Ciliberti A. (1994), *Manuale di glottodidattica*, La Nuova Italia, Firenze.
- Cimaglia R. (2011), *Verbi di percezione*, in *Enciclopedia dell'italiano*, a c. di R. Simone, G. Berruto, P. D'Achille (2010-2011), Disponibile online: <https://www.treccani.it/enciclopedia/ricerca/Verbi-di-percezione/>.
- Comodi A. (1995), *Materiali autentici: selezione e uso nella didattica dell'italiano come lingua straniera*, Guerra Edizioni, Perugia.
- Consiglio d'Europa (2002), *Quadro comune europeo di riferimento per le lingue: apprendimento insegnamento, valutazione*, traduzione di F. Quartapelle e D. Bertocchi, La Nuova Italia, Firenze.
- Consiglio d'Europa (2018), *Common European framework of reference for languages: learning, teaching, assessment. Companion volume with new descriptors*.
- Conte M.E. (2019), *Deissi testuale e anafora*, in *Condizioni di coerenza. Ricerche di linguistica testuale*, a c. di B. Mortara Garavellim, Edizioni dell'Orso, Alessandria [riedizione ampliata del volume edito nel 1988 a Firenze per i tipi La Nuova Italia], pp. 11-27.
- Coonan C.M. (2000), *La ricerca azione*, Università Ca' Foscari, Venezia.
- Cordin P. (2001), *La frase: i pronomi personali*, in *Grande grammatica italiana di consultazione. Vol. I: La frase. I sintagmi nominale e preposizionale*, a c. di L. Renzi, G. Salvi, A. Cardinaletti, Il Mulino, Bologna, pp. 549-606.
- Corino E. (2014), *Didattica delle lingue corpus based*, «EL.LE», 3/2, pp. 231-256.

- Corino E., Marellò C. (2009), *Didattica con i corpora di italiano per stranieri*, «Italiano LinguaDue», pp. 279-285.
- Cortelazzo M. (1972), *Avviamento critico allo studio della dialettologia italiana. Vol. III: Lineamenti di italiano popolare*, Pacini, Pisa.
- Cortelazzo M. (1976 [1972]), *Avviamento critico allo studio della dialettologia italiana. Vol. III: Lineamenti di italiano popolare*, Pacini, Pisa.
- Cortelazzo M. (1984), *Perché “a mi me gusta” sì e “a me mi piace” no?*, in *Umgangssprache im Iberoromanischen. Festschrift für Heinz Kröll*, a c. di G. Holtus, E. Radtke, Narr, Tübingen, pp. 25-28.
- Cortelazzo M.A. (2000a), *La lingua italiana di fine millennio*, in Idem, *Italiano d'oggi*, Esedra, Padova, pp. 9-24.
- Cortelazzo M.A. (2000b), *Per la storia dell'italiano scolastico*, in Idem, *Italiano d'oggi*, Esedra, Padova, pp. 91-109.
- Cortelazzo M.A. (2012), *Evoluzione della lingua, percezione del cambiamento, staticità della norma*, in *I sentieri della lingua. Saggi sugli usi dell'italiano tra passato e presente*, a c. di C. Di Benedetto, S. Ondelli, A. Pezzin, S. Tonello, V. Ujcich, M. Viale, Esedra, Padova, pp. 15-20.
- Coseriu E. (1973), *Lezioni di linguistica generale*, Boringhieri, Torino.
- Cosme C., Gilquin G. (2008), *Free and bound prepositions in a contrastive perspective. The case of with and avec*, in *Phraseology: An Interdisciplinary Perspective*, a c. di S. Granger, F. Meunier, John Benjamins, Amsterdam, pp. 259-274.
- Coxhead A. (2008), *Phraseology and English for academic purposes: Challenges and opportunities*, in *Phraseology Foreign Language Learning and Teaching*, a c. di F. Meunier, S. Granger, John Benjamins, Amsterdam-Philadelphia, pp. 149-161.
- Cresti E. (2000), *Corpus di italiano parlato*, Accademia della Crusca, Firenze.
- Cresti E., Panunzi A. (2013), *Introduzione ai corpora dell'italiano*, Il Mulino, Bologna.
- Curado-Fuentes A. (2012), *Keyword Decoding Reliability for News B1-level Reading Comprehension*, contributo presentato alla 10a conferenza internazionale su *Teaching and Language Corpora (TaLC10)*, Uniwersytet Warszawski, 11-14 luglio 2012. Disponibile online: <http://talc10.ils.uw.edu.pl/files/2011/09/CURADO-FUENTES1.pdf>.

- Curry N., Goodman O. (2018), *Using the Cambridge learner corpus to develop learning materials*, workshop tenuto alla 13a conferenza internazionale su *Teaching and Language Corpora* (TaLC13), Cambridge University, 18-21 luglio 2018. Disponibile online: https://www.researchgate.net/publication/326548400_13th_Teaching_and_Language_Corpora_Conference_TaLC_2018_Book_of_Abstracts.
- D'Achille P. (1990), *Sintassi del parlato e tradizione scritta della lingua italiana. Analisi di testi dalle origini al secolo XVIII*, Bonacci, Roma.
- D'Achille P. (1994), *L'italiano dei semicolti*, in: *Storia della lingua italiana. Vol. II Scritto e parlato*, a c. di A. Asor Rosa, L. Serianni, P. Trifone, Einaudi, Torino, pp. 41-79.
- D'Achille P. (2003), *L'italiano contemporaneo*, Il Mulino, Bologna.
- D'Achille P. (2010), *Lingua d'oggi*, in *Enciclopedia dell'italiano, Vol. I*, Istituto dell'Enciclopedia Italiana, Roma, pp. 793-799.
- D'Achille P. (2016), *Architettura dell'italiano di oggi e linee di tendenza*, in *Manuale di linguistica italiana*, a c. di S. Lubello, De Gruyter, Berlino-Boston, pp. 165-189.
- D'Achille P., Proietti D., Viviani A. (2005), *La frase scissa in italiano: aspetti e problemi*, in *Tipologia linguistica e società. Considerazioni inter- e intralinguistiche. Due Giornate italo-danesi di studi linguistici (Roma, 27-28 novembre 2003)*, a c. di P. D'Achille, I. Korzen, Cesati, Firenze, pp. 249-279.
- Daloiso M. (2009), *I fondamenti neuropsicologici dell'educazione linguistica*, Cafoscarina, Venezia.
- Daloiso M. (2015), *Scienze del linguaggio e educazione linguistica*, Bonacci-Loescher, Torino.
- Danesi M. (1988), *Neurolinguistica e glottodidattica*, Liviana, Padova.
- Danesi M. (1998), *Il cervello in aula!*, Guerra Edizioni, Perugia.
- Dardano M. (1994), *Profilo dell'italiano contemporaneo*, in *Storia della lingua italiana. Vol. II: Scritto e parlato* a c. di L. Serianni, P. Trifone, Einaudi, Torino, pp. 343-432.
- Dardano M. (2020, a c. di), *Sintassi dell'italiano antico II. La prosa del Duecento e del Trecento. La frase semplice*, Carocci, Roma.
- De Cesare A.M. (2006), *'C'è la tua bambina con i fiammiferi'. Funzioni del costrutto presentativo c'è... che...*, in *Parole frasi testi tra scritto e parlato*, a c. di A. Ferrari, «Cenobio», 55/3, pp. 215-221.
- de Haan P. (1992), *The optimum corpus sample size?*, in *New Directions*

- in English Language Corpora: Methodology, Results, Software Developments*, a c. di G. Leitner, Mouton de Gruyter, Berlin, pp. 3-19.
- De Mauro T. (1970), *Per lo studio dell'italiano popolare unitario*, in *Lettere da una tarantata*, a c. di A. Rossi, De Donato, Bari, pp. 43-75.
- De Mauro T. (1997 [1980]), *Guida all'uso delle parole*, Editori Riuniti, Roma.
- De Mauro T. (2011 [1983]), *Storia della lingua Italiana*, Laterza, Bari.
- De Mauro T. (2014), *Storia linguistica dell'Italia repubblicana dal 1946 ai nostri giorni*, Laterza Roma-Bari.
- De Mauro T., Mancini F., Vedovelli M., Voghera M. (1993), *Lessico di frequenza dell'italiano parlato*, Etaslibri, Milano.
- De Roberto E. (2010), *Strutture assolute*, in *Enciclopedia dell'italiano*, a c. di R. Simone, G. Berruto, P. D'Achille (2010-2011). Disponibile online: <https://www.treccani.it/enciclopedia/ricerca/Strutture-assolute/>.
- De Santis C. (2003), *La parola alle narratrici (e ai narratori)*, in *Quaderni dell'Osservatorio Linguistico, vol. I*, a c. di F. Frasnè et al., Franco Angeli, Milano, pp. 54-96.
- Della Putta P., Visigalli M. (2010), *Come aiutare l'acquisizione della morfosintassi del sintagma nominale italiano: confronto fra tre percorsi glottodidattici rivolti a studenti anglofoni*, «Italiano LinguaDue», 2/2, pp. 23-52. Disponibile online: <http://riviste.unimi.it/index.php/promoitals/article/view/821/1062>.
- Diadori P., Palermo M., Troncarelli D. (2015), *Insegnare italiano come seconda lingua*, Carocci, Roma.
- Diadori P., Semplici S., Troncarelli D. (2020, a c. di), *Didattica di base dell'italiano L2*, Carocci, Roma.
- Doughty C. (2003), *Instructed SLA: constraints, compensation, and enhancement*, in *The handbook of second language acquisition*, a c. di C.J. Doughty, M.H. Long, Blackwell, Oxford, pp. 256-310.
- Egerland V. (1996), *The syntax of past participles. A generative study of nonfinite constructions in ancient and modern Italian*, Lund University Press, Lund.
- Ellis R. (1997), *SLA research and language teaching*, Oxford University Press, Oxford.
- Evans V. (2007), *A Glossary of Cognitive Linguistics*, Edinburgh University Press, Edinburgh.

- Evert S. (2009), *The CQP Query Language Tutorial*. Institute for Natural Language Processing, University of Stuttgart. Disponibile online: <https://pdfs.semanticscholar.org/6a0b/ed8affc49daf0f0b7ccfb-35da5abc8f732fa.pdf>.
- Ferrari A. (1999), *L'extra-posizione a destra in italiano, con osservazioni sul francese*, in *Linguistica testuale comparativa*, a c. di G. Skytte, F. Sabatini, Museum Tusculanum Press, København, pp. 111-140.
- Ferrari A. (2003), *Le ragioni del testo. Aspetti sintattici e interpuntivi dell'italiano contemporaneo*, Accademia della Crusca, Firenze.
- Ferrari A. (2008, a c. di), *L'interfaccia lingua testo. Natura e funzioni dell'articolazione informativa dell'enunciato*, Edizioni dell'Orso, Alessandria.
- Ferraris S. (1999), *Imparare la sintassi*, Mercurio, Bergamo.
- Florentino G. (2010), *Che polivalente*, in *Enciclopedia dell'italiano*, in Simone R., Berruto G., D'Achille P., (2010-2011, a c. di). Disponibile online: <https://www.treccani.it/enciclopedia/ricerca/Che-polivalentel>.
- Firth J.R. (1957), *Papers in Linguistics 1934-1951*, Oxford University Press, London.
- Freddi G. (1970), *Metodologia e didattica delle lingue straniere*, Minerva Italica, Bergamo.
- Freddi G. (1994), *Glottodidattica. Fondamenti, metodi, tecniche*, UTET, Torino.
- Freddi G. (1999), *Psicolinguistica, sociolinguistica, glottodidattica*, UTET, Torino.
- Frosini G. (2020, a c. di), *Storia dell'italiano. La lingua, i testi*, Salerno, Roma.
- Galli de' Paratesi N. (1984), *Lingua toscana in bocca ambrosiana. Tendenze verso l'italiano standard: un'inchiesta sociolinguistica*, Il Mulino, Bologna.
- Gatta F. (2014), *Giornalismo*, in *Storia dell'italiano scritto, Vol III: Italiano dell'uso*, a c. di G. Antonelli, M. Motolese e L. Tomasin, Carocci, Roma, pp. 293-347.
- Geeraerts D. (2006), *Methodology in cognitive linguistics*, in *Cognitive linguistics: current applications and future perspectives*, a c. di G. Kristiansen, M. Achard, R. Dirven, F.J.R. Ruiz de Mendoza Ibáñez, Mouton de Gruyter, Berlin-New York.

- Gheno V. (2017), *Social-linguistica. Italiano e italiani dei social network*, Firenze, Franco Cesati Editore.
- Gheno V. (2019), *Potere alle parole. Perché usarle meglio*, Torino, Einaudi.
- Giacalone Ramat A. (2003, a c. di), *Verso l'italiano. Percorsi e strategie di acquisizione*, Carocci, Roma.
- Giacalone Ramat A., Chini M., Andorno C. (2013), *Italiano come L2*, in *La linguistica italiana all'alba del terzo millennio (1997-2010)*, a c. di G. Iannàcaro, Bulzoni, Roma, pp. 149-206.
- Giorgi A., Pianesi F. (1997), *Tense and Aspect. From Semantics to Morphosyntax*, Oxford University Press, New York-Oxford.
- Giuliano L., La Rocca G. (2008), *L'analisi automatica e semi-automatica dei dati testuali. Software e istruzioni per l'uso*, LED, Milano.
- Glisan E., Drescher V. (1993), *Textbook grammar: Does it reflect native speaker speech?*, «The Modern Language Journal», 77/1, pp. 23-33.
- Gossen C.T. (1954), *Studien zur syntaktischen und stilistischen Hervorhebung im modernen Italienisch*, Akademie Verlag, Berlin.
- Grandi N. (2011), *Costruzione passiva*, in *Enciclopedia dell'italiano*, a c. di R. Simone, G. Berruto, P. D'Achille (2010-2011). Disponibile online: <https://www.treccani.it/enciclopedia/ricerca/Costruzione-passiva/>.
- Grandi N. (2018), *Sulla penetrazione di tratti neo-standard nell'italiano degli studenti universitari. Primi risultati di un'indagine empirica*, Griseldaonline (sez. Dibattiti), Disponibile online: <https://griseldaonline.unibo.it/article/view/9021>.
- Grandi N. (2019), *Che tipo, l'italiano neostandard!*, in *Le tendenze dell'italiano contemporaneo rivisitate. Atti del LII Congresso Internazionale di Studi della Società di Linguistica Italiana* (Berna, 6-8 settembre 2018), Officinaventuno, Milano, pp. 59-74.
- Granger S., Meunier F. (2008, a c. di), *Phraseology: An Interdisciplinary Perspective*, John Benjamins, Amsterdam.
- Grassi R., Bozzone Costa R., Ghezzi C. (2008, a c. di), *Dagli studi sulle sequenze di acquisizione alla classe di italiano L2*, Guerra Edizioni, Perugia.
- Gries S.T., Stefanowitsch A. (2006, a c. di), *Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis*, Mouton de Gruyter, Berlin-New York.

- Gries S.T., Wulff, S. (2005), *Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora*, in «Annual Review of Cognitive Linguistics», 3, pp.182-200.
- Gries S.T., Yoon J. (2016, a c. di), *Corpus-based Approaches to Construction Grammar (Constructional Approaches to Language)*, John Benjamins, Amsterdam-Philadelphia.
- Gualdo R. (2007), *L'italiano dei giornali*, Carocci, Roma.
- Guidetti M.G., Lenzi G., Storch, S. (2012), *Potenzialità e limiti dell'uso dei corpora linguistici per la didattica dell'italiano LS*, «Bollettino ITALS», 44. Disponibile online: <https://www.italis.it/potenzialit%C3%A0-e-limiti-dell%E2%80%99uso-dei-corpora-linguistici-la-didattica-dell%E2%80%99italiano-ls>.
- Hadley G., Hadley H. (2018), *Investigating data-driven learning with receptive skills: a contemporary approach*, contributo presentato alla 13a conferenza internazionale su *Teaching and Language Corpora* (TaLC13), Cambridge University, 18-21 luglio 2018. Disponibile online: https://www.researchgate.net/publication/326548400_13th_Teaching_and_Language_Corpora_Conference_TaLC_2018_Book_of_Abstracts.
- Harrich K. (2011), *Il notiziario ovvero l'importanza dell'uso di materiali autentici nell'insegnamento – apprendimento linguistico*, «Italiano LinguaDue», 2/2, pp. 321-335. Disponibile online: <http://riviste.unimi.it/index.php/promoitals/article/view/835>.
- Hilpert M. (2013), *Corpus-based approaches to constructional change*, in *The Oxford Handbook of Construction Grammar*, a c. di G.Trousdale, T. Hoffmann, Oxford University Press, Oxford, pp. 458-477.
- Hinkel E. (2005, a c. di), *Handbook of research in second language teaching and learning*, Erlbaum, Mahwah NJ.
- Hinkel E. (2011, a c. di), *Handbook of Research in Second Language Teaching and Learning: Volume 2*, Routledge, New York-London.
- Hoffmann T., Trousdale G. (2013, a c. di), *The Oxford Handbook of Construction Grammar*, Oxford University Press, Oxford-New York.
- Holtus G. (1984), *Codice parlato e codice scritto*, in *Il dialetto dall'oralità alla scrittura. Atti del XIII Convegno per gli Studi Dialettali Italiani* (Catania-Nicosia, 28 settembre 1981), Pacini, Pisa, pp. 1-12.

- Holtus G., Radtke E. (1986-1990, a c. di), *Sprachlicher Substandard. Vol I: Sprachlicher, Substandard; vol II: Standard und Substandard in der Sprachgeschichte und in der Grammatik; vol. III: Standard, Substandard und Varietätenlinguistik*, Niemeyer, Tübingen.
- Housen A., Pierrard M. (2005, a c. di), *Investigations in instructed second language acquisition*, Mouton de Gruyter, Berlin.
- Hunston S. (2002), *Corpora in Applied Linguistics*, Cambridge University Press, Cambridge.
- Hunston S., Francis G. (2000), *Pattern grammar. A corpus-driven approach to the lexical grammar of English*, John Benjamins, Amsterdam-Philadelphia.
- Ivanovska-Naskova R. (2014), *Il corpus parallelo italiano-macedone come strumento nella didattica dell'Italiano LS*, in *Dal manoscritto al web: canali e modalità di trasmissione dell'italiano. Tecniche, materiali e usi nella storia della lingua*, a c. di E. Garavelli, E. Suome-la-Härmä. Atti del XII Congresso SILFI - Società Internazionale di Linguistica e Filologia Italiana (Helsinki, 18-20 giugno 2012), Franco Cesati Editore, Firenze, pp. 471-479.
- Ivanovska-Naskova R. (2015), *I corpora e l'insegnamento dell'italiano LS*, in *Incroci. Studi sulla letteratura, la traduzione e la glottodidattica*, a c. di V. Zaccaro, R. Ivanovska-Naskova, Università degli Studi di Bari "Aldo Moro" e Università "Ss. Cirillo e Metodio" di Skopje, Bari-Skopje, pp. 105-123.
- Ivanovska-Naskova R. (2018), *L'insegnamento della grammatica dell'italiano LS attraverso corpora*, «Italica Wratislaviensia», 2018, 9/1, pp. 71-87.
- Jakubíček M., Kilgariff A., Kovář V., Rychlý P., Suchomel V. (2013), *The tenten corpus family*, in *7th International Corpus Linguistics Conference (CL2013)*. Atti del Convegno (Lancaster University, 23-26 luglio 2013), UCREL – Lancaster University, Lancaster, pp. 125-127.
- Ježek E. (2011), *Verbi pronominali*, in *Enciclopedia dell'italiano*, a c. di R. Simone, G. Berruto, P. D'Achille (2010-2011). Disponibile online: <https://www.treccani.it/enciclopedia/ricerca/verbi-pronominali/>.
- Johns T.F. (1991), *Should you be persuaded? - Two Samples of Data-Driven Learning Materials*, «English Language Research Journal», 4, pp. 1-16.

- Jones M., Haywood S. (2004), *Facilitating the acquisition of formulaic sequences: An exploratory study in an EAP context*, in *Formulaic sequences*, a c. di N. Schmitt, John Benjamins Amsterdam, pp. 269-300.
- Kennedy C., Miceli, T. (2010), *Corpus-Assisted Creative Writing: Introducing Intermediate Italian Learners to a Corpus as a Reference Resource*, «Language Learning & Technology», 14/1, pp. 28-44.
- Kilgarrieff A., Grefenstette G. (2003), *Introduction to the special issue on the Web as a Corpus*, «Computational linguistics», 29/3, pp. 333-347.
- Koch P. (2014), *La scelta manzoniana tra selezione e ristandardizzazione*, in *Dall'architettura della lingua italiana all'architettura linguistica dell'Italia. Saggi in omaggio a Heidi Siller-Runggaldier*, a c. di P. Danler, C. Konecny, Peter Lang, Frankfurt am Main *et al.*, pp. 75-102.
- Koffka K. (1935), *Principles of Gestalt-Psychology*, Harcourt, Brace & Company, New York.
- Köhler W. (1929), *Gestalt Psychology*, Liveright, New York.
- Krashen S.D. (1982), *Principle and Practice in Second Language Acquisition*, Pergamon, Oxford.
- Krashen S.D. (1985), *The Input Hypothesis: Issues and Implications*, Longman, London.
- Lambrecht K. (1994), *Information Structure and Sentence Form. Topic, Focus, and the Mental Representations of Discourse Referents*, Cambridge University Press, Cambridge.
- Laudanna A., Thornton A.M., Brown G., Burani C., Marconi L. (1995), *Un corpus dell'italiano scritto contemporaneo dalla parte del ricevente*, in *JADT 1995. III Giornate internazionali di Analisi Statistica dei Dati Testuali*, a c. di S. Bolasco, L. Lebart, A. Salem, I. Cisu, Roma, pp. 103-109.
- Lavinio C. (1986), *Tipologia dei testi parlati e scritti*, «Linguaggi», 3/1-2, pp. 14-22.
- Leech G. (1997), *Teaching and language corpora: A convergence*, in *Teaching and language corpora*, a c. di A. Wichmann, S. Fligelstone, T. McEnery e G. Knowles, Longman, London, pp. 1-23.
- Lepschy A.L., Lepschy G.C. (1988), *La lingua italiana: storia, varietà dell'uso, grammatica*, Bompiani, Milano.
- Levelt W.J.M. (1989), *Speaking. From intention to articulation*, MIT Press, Cambridge (MA).

- Lewandowska-Tomaszczyk B., Dziwirek K. (2009 a c. di), *Studies in cognitive corpus linguistics*, Peter Lang, Frankfurt am Main.
- Liscia R. (2004, a c. di), *E-learning. Stato dell'arte e prospettive di sviluppo*, Apogeo, Milano.
- Lo Duca M.G. (1990), *Creatività e regole*, Il Mulino, Bologna.
- Lo Duca M.G., Solarino R. (2006), *Lingua italiana. Una grammatica ragionevole*, Unipress, Padova.
- Lo Duca M.G. (2006), *Sillabo di italiano L2*, Carocci, Roma.
- Loporcaro M., Seiler S. (2009), *Evoluzione diacronica delle subordinate partecipiali in italiano*, in *Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione e giustapposizione, vol. I*, a cura di A. Ferrari. Atti del X congresso della Società Internazionale di Linguistica e Filologia italiana (Basilea, 30 giugno-3 luglio 2008), Cesati, Firenze, pp. 485-512.
- Lorenzetti L. (2002), *L'italiano contemporaneo*, Carocci, Roma.
- Lüdeling A., Kytö M. (2008-2009, a c. di), *Corpus linguistics. An international handbook. Vol. I e II*, Mouton de Gruyter, Berlin.
- Lyding V., Borghetti C., Dittmann H., Nicolas L., Stemle E. (2013), *Open corpus interface for Italian language learning*, in *Proceeding of ICT for Language Learning*, VI ed. Atti del convegno (Firenze 14-15 novembre 2013). Disponibile online: https://www.researchgate.net/publication/257043915_Open_Corpus_Interface_for_Italian_Language_Learning.
- Lyding V., Chiocchetti E., Sérasset G., Brunet-Manquat F. (2006), *The LexALP Information System: Term Bank and Corpus for Multilingual Legal Terminology Consolidated*, in *Proceedings of the Workshop on Multilingual Language Resources and Interoperability (MLRI '06)*. Atti del convegno (Sydney, luglio 2006), Association for Computational Linguistics Stroudsburg, PA, pp. 25-31.
- Lyding V., Stemle E., Borghetti C., Brunello M., Castagnoli S., Dell'Orletta F., Dittmann H., Lenci A., Pirrelli V. (2014), *The PAISÀ Corpus of Italian Web Texts*, in *Proceedings of the 9th Web as Corpus Workshop (WaC-9)* (EACL 2014). Atti del Convegno (Göthenburg, 26 aprile 2014), Association for Computational Linguistics, Stroudsburg, pp. 36-43.
- Maiden M. (2005), *Il problema del congiuntivo nell'insegnamento dell'italiano: verso una soluzione?*, in *Prospettive sull'italiano come Lingua*

- Straniera*, a c. di A.L. Lepschy, A.R. Tamponi, Guerra Edizioni, Perugia, pp. 109-117.
- Malinowska M. (1996), *Contenuti semantici del congiuntivo - ruolo della modalità e dei fattori extralinguistici nell'acquisizione della competenza comunicativa in italiano come lingua 2*, Viridis, Kraków.
- Mandelli M. (2006), *Quella 'e' a inizio frase... Scarti prosodici e interpretivi tra elementi coordinati*, in *Parole frasi testi tra scritto e parlato*, a c. di A. Ferrari, «Cenobio», 55/3, pp. 231-240.
- Manili P. (2015), *Il modo congiuntivo e l'italiano L2*, «Gentes», 2, pp. 154-167.
- Marani T. (2016), *Didattica del congiuntivo e (neo)purismo linguistico*. Aggiornamenti, «Rivista dell'Associazione dei docenti di italiano in Germania», 9, pp. 28-40.
- Marello C. (1996), *Le parole dell'italiano. Lessico e dizionari*, Zanichelli, Bologna.
- Mauranen A. (2004), *Speech corpora in the classroom*, in *Corpora and language learners*, a c. di G. Aston, S. Bernardini e D. Stewart, John Benjamins, Amsterdam-Philadelphia, pp. 195-211.
- Mazzoleni M. (1992), *"Se lo sapevo non ci venivo": l'imperfetto indicativo ipotetico nell'italiano contemporaneo*, in *Linee di tendenza dell'italiano contemporaneo*, a c. di B. Moretti, D. Petrini, S. Bianconi, Bulzoni, Roma, pp. 171-190.
- McEnery A., Xiao R. (2005), *Help or help to: What do corpora have to say*, «English Studies», 86/2, pp. 161-187.
- McEnery T. (2003), *Corpus Linguistics*, in *The Oxford Handbook of Computational Linguistics*, a c. di R. Mitkov, Oxford University Press, Oxford, pp. 448-463.
- McEnery T., Hardie A. (2012), *Corpus Linguistics: Method, Theory and Practice*, Cambridge University Press, Cambridge.
- McEnery T., Xiao R. (2011), *What corpora can offer in language teaching and learning*, in *Handbook of research in second language teaching and learning*, a c. di E. Hinkel, Routledge, New York, pp. 364-380.
- McEnery T., Wilson A. (2001[1996]), *Corpus Linguistics*, Edinburgh University Press, Edinburgh.
- MacWhinney B. (1997), *Il progetto CHILDES: Strumenti per l'analisi del linguaggio parlato*, ed. italiana a c. di E. Pizzuto, U. Bortolini

- (*The CHILDES project: Tools for analyzing talk* (2nd ed.). Lawrence Erlbaum Associates, Inc.). Pisa, Del Cerro.
- Mengaldo P.V. (1994), *Storia della lingua italiana*, Il Novecento, Bologna.
- Meunier F., Granger S. (2008, a c. di), *Phraseology in Language Learning and Teaching*, John Benjamins, Amsterdam.
- Minciarelli F., Comodi A. (2005), *Sillabo per i cinque gradi del corso di lingua e cultura italiana per stranieri*, Guerra Edizioni, Perugia.
- Mioni A.M. (1983), *Italiano tendenziale: osservazioni su alcuni aspetti della standardizzazione*, in *Scritti linguistici in onore di Giovan Battista Pellegrini. Vol. I*, a c. di P. Benincà et al., Pacini, Pisa, pp. 495-517.
- Moneti A. (1988), *Sull'insegnamento del congiuntivo in italiano lingua seconda: proposta didattica*, «Italice», 65/1, pp. 19-30.
- Moretti B., Petrini D., Bianconi S. (1992, a c. di), *Linee di tendenza dell'italiano contemporaneo*, Bulzoni, Roma.
- Moro A. (1998), «Esserci» e «averci»: les clitiques sujets et une analyse en «petite proposition» pour «avoir», in «Etre» et «avoir»: syntaxe, statique, typologie, a c. di A. Rouveret, Presses Universitaires de Vincennes, Saint-Denis, pp. 155-170.
- Mortara Garavelli B. (2001), *Le parole e la giustizia*, Einaudi, Torino.
- Mudraya O., Piao S.S.L., Rayson P., Sharoff S., Babych B., Löffberg L. (2008), *Automatic extraction of translation equivalents of phrasal and light verbs in English and Russian*, in *Phraseology: An Interdisciplinary Perspective*, a c. di S. Granger, F. Meunier, John Benjamins, Amsterdam, pp. 293-309.
- Mukherjee J. (2002), *Korpuslinguistik und Englischunterricht: Eine Einführung*, Peter Lang, Frankfurt am Main.
- Nencioni G. (1976), *Parlato-parlato, parlato-scritto, parlato recitato*, «Strumenti critici», 10, pp. 1-56, ora in G. Nencioni, *Di scritto e di parlato. Discorsi linguistici*, Zanichelli, Bologna, 1983, pp. 126-179.
- Nencioni G. (1987), *Costanza dell'antico nel parlato moderno*, in *Gli Italiani parlati. Sondaggi nella lingua di oggi*, AA.VV., Accademia della Crusca, Firenze, pp. 7-25.
- Nesi H., Thompson P. (2014), *Using Sketch Engine with BAWE*. Disponibile online: <https://www.coventry.ac.uk/research/research-directories/current-projects/2015/british-academic-written-english-corpus-bawe/search-the-bawe-corpus/>.

- Newman J. (2011), *Corpora and cognitive linguistics*, «Brazilian Journal of Applied Linguistics», 11/2, pp. 521-559.
- Nunan D. (1989), *Understanding Language Classrooms. A Guide for Teacher-Initiated Research*, Prentice Hall, New York.
- Ondelli S., Pontrandolfo G. (2014), *La negazione multipla nei testi giuridici: veramente non si può negare che sia un tratto caratteristico?*, «Rivista internazionale di tecnica della traduzione», 16, pp. 153-176.
- Ondelli S., Romanini F. (2018), *Norma interiorizzata e uso: un'indagine preliminare su parlanti italiani*, «Italica Wratislaviensia», 9/1, pp. 185-207.
- Ondelli S., Viale M. (2010), *L'assetto dell'italiano delle traduzioni in un corpus giornalistico. Aspetti qualitativi e quantitativi*, «Rivista internazionale di tecnica della traduzione», 12, pp. 1-62.
- Orenha-Ottaiano A. (2012), *Collocations and the Design of Teaching Materials for Language Learners*, contributo presentato alla 10a conferenza internazionale su *Teaching and Language Corpora* (TaLC10), Uniwersytet Warszawski, 11-14 luglio 2012. Disponibile online: <http://talc10.ils.uw.edu.pl/files/2011/09/ORENHA-OTTAIANO.pdf>.
- Palermo M. (1997), *L'espressione del pronome personale soggetto nella storia dell'italiano*, Bulzoni, Roma.
- Palermo M. (2011), *Soggetto*, in *Enciclopedia dell'italiano*, a c. di R. Simone, G. Berruto, P. D'Achille (2010-2011). Disponibile online: https://www.treccani.it/enciclopedia/soggetto_%28Enciclopedia-dell%27Italiano%29/
- Palermo M. (2020), *Linguistica italiana*, Il Mulino, Bologna.
- Panunzi A., Cresti E., Gregori L. (2014), *RIDIRE. Corpus and Tools for the Acquisition of Italian L2*, in *Proceedings of the XVI EURALEX International Congress: The User in Focus*, a c. di A. Abel, C. Vettori, N. Ralli. Atti del convegno (Bolzano, 15-19 luglio 2014), EURAC research, Bolzano, 447-462.
- Patota G. (1990), *Sintassi e storia della lingua italiana: tipologia delle frasi interrogative*, Bulzoni, Roma.
- Pecman M. (2008), *Compilation, formalisation and presentation of bilingual phraseology: problems and possible solutions*, in *Phraseology in Language Learning and Teaching*, a c. di F. Meunier, S. Granger, John Benjamins, Amsterdam-Philadelphia, pp. 203-222.

- Pellegrini G.B. (1960), *Tra lingua e dialetto in Italia*, «Studi mediolatini e volgari», 8, pp. 137-153.
- Pesce A. (2003), *Una lingua per l'infanzia*, in *Quaderni dell'Osservatorio Linguistico, vol. I*, a c. di F. Frasnè et al., Franco Angeli, Milano, pp. 293-315.
- Picoral A., Staples S., Shin J., Swatek A. (2018), *Exploring variation and intertextuality in L2 undergraduate writing in English: Using the Corpus and Repository of Writing Online Platform for research and teaching*, workshop tenuto alla 13a conferenza internazionale su *Teaching and Language Corpora* (TaLC13), Cambridge University, 18-21 luglio 2018. Disponibile online: https://www.researchgate.net/publication/326548400_13th_Teaching_and_Language_Corpora_Conference_TaLC_2018_Book_of_Abstracts.
- Pieneman M. (1986), *L'effetto dell'insegnamento sugli orientamenti degli apprendenti nell'acquisizione di L2*, in *L'apprendimento spontaneo di una seconda lingua*, a c. di A. Giacalone Ramat, Il Mulino, Bologna, pp. 307-326.
- Pieneman M. (1998), *Language Processing and Second Language Development. Processability Theory*, John Benjamins, Amsterdam-Philadelphia.
- Pienemann M. (2005), *Cross-Linguistic Aspects of Processability Theory*, John Benjamins, Amsterdam.
- Pistoletti E. (2015), *Diamesia: la nascita di una dimensione*, in *Parole, gesti, interpretazioni. Studi linguistici per Carla Bazzanella*, a c. di E. Pistoletti, R. Pugliese, B. Gili Fivela, Aracne, Roma, pp. 27-56.
- Poggi Salani T. (2010), *Italiano regionale*, in *Enciclopedia dell'italiano*, a c. di R. Simone, G. Berruto, P. D'Achille (2010-2011). Disponibile online: https://www.treccani.it/enciclopedia/italiano-regionale_%28Enciclopedia-dell%27Italiano%29/.
- Poggi Salani T. (2012 [1986]), *Per lo studio dell'italiano. Avviamento storico descrittivo*, Liviana, Padova.
- Pozzo G. (2008), *L'osservazione: uno strumento per conoscere cosa succede in classe*, «Lend – Lingua e nuova Didattica», 37/1, pp. 15-28.
- Prandi M. (2002), *C'è un valore per il congiuntivo?*, in *Intorno al congiuntivo*, a c. di L. Schena, M. Prandi, M. Mazzoleni, CLUEB, Bologna, pp. 29-44.

- Prandi M. (2020), *Le regole e le scelte. Grammatica italiana*, UTET, Torino.
- Radtke E. (1992), *Varietà dell'italiano*, in *La linguistica italiana degli anni 1976-1986*, a c. di A. A. Mioni, M.A. Cortelazzo, Bulzoni, Roma, pp. 59-74.
- Ramat P. (1993), *L'italiano lingua d'Europa*, in *Introduzione all'italiano contemporaneo. Vol. I*, a c. di A.A. Sobrero, pp. 3-39.
- Rastelli S. (2009), *Che cos'è la didattica acquisizionale*, Carocci, Roma.
- Rastelli S. (2010 a c. di), *Italiano di Cinesi, Italiano per Cinesi. Dalla prospettiva della didattica acquisizionale*, Guerra Edizioni, Perugia.
- Rehm G., Santini M., Mehler A., Braslavski P., Gleim R., Stubbe A., Symonenko S., Tivosanis M., Vidulin V. (2008), *Towards a reference corpus of web genres for the evaluation of genre identification systems*, in *Proceedings of the the sixth international conference on Language Resources and Evaluation (LREC 2008)*. Atti del convegno (Marrakech, 28-30 maggio 2008), ELRA - European Language Resources Association, Paris, pp. 351-358.
- Renzi L. (2012), *Come cambia la lingua. L'italiano in movimento*, Il Mulino, Bologna.
- Roggia C.E. (2006), *Costruzioni marcate tra scritto e parlato: la frase scissa*, in *Parole frasi testi tra scritto e parlato*, a c. di A. Ferrari, «Cenobio», 55/3, pp. 222-230.
- Roggia C.E. (2009), *Le frasi scisse in italiano. Struttura informativa e funzioni discorsive*, Éditions Slatkine, Genève.
- Römer U. (2004), *Comparing real and ideal language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching*, in *Corpora and language learners*, a c. di G. Aston, S. Bernardini e D. Stewart, John Benjamins, Amsterdam-Philadelphia, pp. 151-168.
- Römer U. (2008), *Corpora and language teaching*, in *Corpus Linguistics. An International Handbook Vol. 1*, a c. di A. Lüdeling e M. Kytö, Mouton de Gruyter, Berlin-New York, pp. 112-131.
- Rossi F. (2016), *Dislocazione a destra/la sinistra in due corpora di italiano scritto: tra grammaticalizzazione, ammiccamento e coesione*, in *La lingua variabile nei testi letterari, artistici e funzionali contemporanei*, a c. di G. Ruffino, M. Castiglione, Cesati, Firenze, pp. 847-859.
- Rossini Favretti R., Tamburini F., De Santis C. (2002), *CORIS/CODIS*:

- A corpus of written Italian based on a defined and a dynamic model*, in *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*, a c. di A. Wilson, P. Rayson, T. McEnery, Lincom-Europa, München, pp. 27-38.
- Rychlý P. (2007), *Manatee/Bonito-A Modular Corpus Manager*, in *Proceedings of Recent Advances in Slavonic Natural Language Processing* (RASLAN 2007), a c. di P. Sojka, A. Horák. Atti del convegno (Brno, 14-16 dicembre 2007), Masaryk University, Brno, pp. 65-70.
- Sabatini F. (1982), *La comunicazione orale, scritta e trasmessa: la diversità del mezzo, della lingua e delle funzioni*, in *Educazione linguistica nella scuola superiore. Sei argomenti per un curriculum*, a c. di A.M. Boccafurni, S. Serromani, Provincia di Roma e Consiglio Nazionale delle Ricerche, Roma, pp. 105-127.
- Sabatini F. (1985), *L'italiano dell'uso medio: una realtà tra le varietà linguistiche italiane*, in *Geschprochenes Italienisch in Geschichte und Gegenwart*, a c. di G. Holtus, E. Radtke, Narr, Tübingen, pp. 154-184.
- Sabatini F. (1990), *Una lingua ritrovata: l'italiano parlato*, «Studi latini e italiani», 4, pp. 215-234.
- Sabatini F. (1997), *Pause e congiunzioni nel testo. Quel "ma" all'inizio di frase...*, in *Norma e lingua in Italia: alcune riflessioni tra passato e presente*, a c. di I. Bonomi, Istituto Lombardo di Scienze e Lettere, Milano, pp. 113-146.
- Sabatini F. (1997), *Prove per l'italiano 'trasmesso' (e auspici di un parlato serio semplice)*, in *Gli italiani trasmessi: la radio*. Atti del Convegno (Firenze, Villa Medicea di Castello, 13-14 maggio 1994), Accademia della Crusca, Firenze, pp. 11-30.
- Sabatini F. (2011), *L'italiano nel mondo moderno, Saggi scelti dal 1968 al 2009, vol. II*, a c. di V. Coletti, R. Coluccia, P. D'Achille, N. De Blasi, D. Proietti, bibliografia degli scritti a c. di R. Cimaglia, Liguori, Napoli.
- Sabatini F. (2004), *Risposta al quesito di Marina Hubacech sull'uso del verbo venire come ausiliare solo nei tempi semplici*, in «La Crusca per voi», 29, pp. 8-9. Disponibile online: www.accademiadellacrusca.it/it/lingua-italiana/consulenza-linguistica/domande-risposte/motivo-per-verbo-venire-utilizzato-ausiliare.

- Sabatini F. (2016), *Lezione di italiano*, Mondadori, Milano.
- Salvi G. (1988), *La frase semplice*, in *Grande grammatica italiana di consultazione. Vol. I: La frase. I sintagmi nominale e preposizionale*, a c. di L. Renzi, G. Salvi, A. Cardinaletti, Il Mulino, Bologna, pp. 37-127.
- Sansò A. (2003), *Degrees of event elaboration. Passive construction in Italian and Spanish*, Franco Angeli, Milano.
- Santini M. (2005), *Genres in formation? An exploratory study of web pages using cluster analysis*, in *Proceedings of the 8th Annual Colloquium for the UK Special Interest Group for Computational Linguistics (CLUK05)*. Atti del convegno (Manchester, 11 gennaio 2005). Disponibile online: https://www.researchgate.net/publication/28550769_Genres_in_formation_An_exploratory_study_of_Web_pages_using_cluster_analysis.
- Santini M., Mehler A., Sharoff S. (2010), *Riding the Rough Waves of Genre on the Web. Concepts and Research Questions*, in *Genres on the Web: Computational Models and Empirical Studies*, a c. di A. Mehler, S. Sharoff, M. Santini, Springer, Dordrecht, pp. 3-33.
- Santipolo M. (2014), *L'impiego dei materiali autentici per lo sviluppo della competenza sociolinguistica: riflessioni teoriche e spunti operativi*, «Cultura & Comunicazione», 5, pp. 15-22.
- Schena L., Prandi M., Mazzoleni M. (2002, a c. di), *Intorno al congiuntivo*, CLUEB, Bologna.
- Scherer K.R. (1984), *On the nature and function of emotion: a component process approach*, in *Approaches to emotion*, a c. di K.R. Scherer, P. Ekman, Erlbaum, Hillsdale, pp. 293-317.
- Scherer K.R. (2009), *The dynamic architecture of emotion: Evidence for the component process model*, «Cognition and Emotion», 23/7, pp. 1307-1351.
- Scherer K.R., Schorr A., Johnstone T. (2001, a c. di), *Appraisal processes in emotion: theory, methods, research*, Oxford University Press, New York.
- Schneider S. (1999), *Il congiuntivo tra modalità e subordinazione: uno studio sull'italiano parlato*, Carocci, Roma.
- Schumann J.H. (1999), *The Neurobiology of Affect in Language*, Blackwell, Oxford.
- Schumann J.H., Crowell S.E., Jones N.E., Lee N., Schuchert S.A.,

- Wood L.A. (2004), *The neurobiology of learning. Perspectives from second language acquisition*, Lawrence Erlbaum, Mahwah (NJ).
- Selinker L. (1972), *Interlanguage*, «International Review of Applied Linguistics in Language Teaching», 10, pp. 209-241.
- Seracini F. (2018), *Phraseology in specialised translation learning: a corpus-based study*, contributo presentato alla 13a conferenza internazionale su *Teaching and Language Corpora* (TaLC13), Cambridge University, 18-21 luglio 2018. Disponibile online: https://www.researchgate.net/publication/326548400_13th_Teaching_and_Language_Corpora_Conference-TaLC_2018_Book_of_Abstracts.
- Serianni L. (1986), *Il problema della norma linguistica dell'italiano*, «Gli Annali della Università per Stranieri [di Perugia]», 7, pp. 47-69.
- Serianni L. (1988), *Grammatica italiana. Italiano comune e lingua letteraria. Suoni, forme, costrutti*, con la collaborazione di A. Castelvocchi, UTET, Torino.
- Serianni L. (1989a), *Le varianti fonomorfologiche dei "Promessi Sposi" 1840 nel quadro dell'italiano ottocentesco*, in *Saggi di storia linguistica italiana*, L. Serianni, Morano, Napoli, pp. 141-213.
- Serianni L. (2006), *Prima lezione di grammatica*, Laterza, Roma-Bari.
- Serianni L., Trifone P. (1994, a c. di), *Storia della lingua italiana. Vol. II: Scritto e parlato*, Einaudi, Torino.
- Sgroi S.C. (2013), *Dove va il congiuntivo? Ovvero il congiuntivo da nove punti di vista*, UTET, Torino.
- Sharoff S. (2006), *Creating General-Purpose Corpora Using Automated Search Engine Queries*, in *Wacky! Working Papers on the Web as Corpus*, a c. di M. Baroni, S. Bernardini, Gedit, Bologna, pp. 63-98.
- Sharoff S. (2010), *Analysing similarities and differences between corpora*, in *7th Language Technologies Conference*, Jožef Stefan Institute, Ljubljana, pp. 5-11.
- Simone R. (1991), *Why linguists need variation. Reflections on Italian in "troubled contexts"*, «Rivista di linguistica», 3/2, pp. 407-421.
- Simone R. (1993), *Stabilità e instabilità nei caratteri originali dell'italiano*, in *Introduzione all'italiano contemporaneo. Vol. 1: Struttura e variazioni*, a c. di A.A. Sobrero, Laterza, Roma-Bari, pp. 41-100.
- Simone R., Berruto G., D'Achille P. (2010-2011, a c. di), *Enciclopedia dell'italiano*, Treccani, Roma (<http://www.treccani.it/>).
- Sinclair J. McH. (1996), *Preliminary recommendations on Corpus Ty-*

- pology, *EAGLES Document EAG-TCWG-CTYP/R, Version of May, 1996*. Disponibile online: <http://www.ilc.cnr.it/EAGLES96/corpus-yp/corpusyp.html>.
- Sinclair J. McH. (1997), *Corpus evidence in language description*, in *Teaching and language Corpora*, a c. di A. Wichmann, S. Fligelstone, T. McEnery e G. Knowles, Longman, London, pp. 27-39.
- Sinclair J. McH. (2004, a c. di), *How to Use Corpora in Language Teaching*, John Benjamins, Amsterdam.
- Sinclair J. McH. (2005), *Corpus and Text — Basic Principles*, in *Developing Linguistic Corpora: a Guide to Good Practice*, a c. di M. Wynne, AHDS (Arts and Humanities Data Service), versione online priva di numero di pagina. Disponibile online: <https://users.ox.ac.uk/~martinw/dlc/chapter1.htm>.
- Sinclair J. McH. (1991), *Corpus, concordance, collocation*, Oxford University Press, Oxford.
- Słapek D. (2016, aprile 14-16), *Argomenti grammaticali nei manuali d'italiano L2 al livello C1/C2* [presentazione a convegno], Gli orizzonti dell'italianistica: tradizione, attualità e sfide di ricerca, Uniwersytet Pedagogiczny, Kraków, Polska. <https://in.up.krakow.pl/2016/03/10/gli-orizzonti-dellitalianistica-tradizione-attualita-e-sfide-di-ricerca-ii-incontri-cracoviensi-di-italianistica/>.
- Słapek D. (2017), *Argomenti grammaticali nei manuali e nei certificati di italiano LS*, «Rassegna Italiana di Linguistica Applicata», 49/1, pp. 109-127.
- Sobrero A.A. (1992a), *L'italiano di oggi*, Istituto della Enciclopedia Italiana Treccani, Roma.
- Sobrero A.A. (1992b), *Forme e norme*, «Italiano e oltre», 7, p. 153.
- Sobrero A.A. (1993, a c. di), *Introduzione all'italiano contemporaneo. Vol. I: le strutture; Vol. II: la variazione e gli usi*, Laterza, Roma-Bari.
- Sornicola R. (1981), *Sul parlato*, Il Mulino, Bologna.
- Sosnowski R. (2010), *Errori dovuti a interferenze negli scritti degli apprendenti polacchi (livello intermedio e avanzato). Ricerca corpus based*, «Romanica Cracoviensia», 10/1, pp. 135-156.
- Spina S. (2010a), *The dictionary of Italian Collocations: Design and intergration in an Online Learning Environment*, in *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. Atti del convegno (Malta, 17-23 maggio

- 2010), ELRA - European Language Resources Association, Paris, pp. 3202-3208.
- Spina S. (2010b), *AIWL: una lista di frequenza dell'italiano accademico*, in *Statistical Analysis of Textual Data, Proceedings of the 10th JADT Conference*, a c. di S. Bolasco, I. Chiari, L. Giuliano. Atti del convegno (Roma, Università La Sapienza, 9-11 giugno 2010), Editrice universitaria LED, Roma, pp. 1317-1325.
- Spina S. (2014), *Il Perugia Corpus: una risorsa di riferimento per l'italiano. Composizione, annotazione e valutazione*, in *Proceedings of the First Italian Conference on Computational Linguistics (CLiC-it 2014)*, a c. di R. Basili, A. Lenci, B. Magnini, 1, Pisa University Press, Pisa, pp. 354-359.
- Spinelli B., Parizzi F. (2010 a c. di). *Profilo della lingua italiana. Livelli di riferimento del QCER A1, A2, B1, B2*, La Nuova Italia, Firenze.
- Stefanowitsch A., Gries S.T. (2006, a c. di), *Corpus-based approaches to metaphor and metonymy*, Mouton de Gruyter, Berlin-New York.
- Stubbs M. (2001), *Texts, corpora, and problems of interpretation: A response to Widdowson*, «Applied linguistics», 22/2, pp. 149-172.
- Svartvik, J. (1992). *Corpus linguistics comes of age*, in *Directions in corpus linguistics. Proceedings of Nobel Symposium 82* (Stockholm, 4-8 agosto 1991), a c. di J. Svartvik, Mouton de Gruyter, Berlin-New York.
- Tamburini F. (2000), *Annotazione grammaticale e lemmatizzazione di corpora in italiano*, in *Linguistica e informatica: multimedialità, corpora e percorsi di apprendimento*, a c. di R. Rossini, Bulzoni, Roma, pp. 57-73.
- Tavoni M. (2002), *Caratteristiche dell'italiano contemporaneo e insegnamento della scrittura*, in *Manuale dell'italiano professionale. Teoria e didattica*, a c. di F. Bruni, T. Raso, Zanichelli, Bologna, pp. 139-152.
- Taylor D. (1994), *Inauthentic authenticity or authentic in authenticity?*, «TESL-EJ», 1/2, pp. 1-12. Disponibile online: <https://www.tesl-ej.org/wordpress/issues/volume1/ej02/ej02a1/>.
- Tesi R. (2005), *Storia dell'italiano. La lingua moderna e contemporanea*, Zanichelli, Bologna.
- Titone R. (1976), *Psicodidattica*, La Scuola, Brescia.
- Tognini-Bonelli E. (2001), *Corpus Linguistics at Work*, John Benjamins, Amsterdam.

- Tomlin R.S. (1986), *Basic Word Order. Functional Principles*, Croom Helm, London-Sydney-Wolfeboro.
- Tono Y. (2018), *CEFR-Jx27: Developing corpus- and CEFR-based pedagogical resources and e-learning systems for 27 languages*, contributo presentato alla 13a conferenza internazionale su *Teaching and Language Corpora* (TaLC13), Cambridge University, 18-21 luglio 2018. Disponibile online: https://www.researchgate.net/publication/326548400_13th_Teaching_and_Language_Corpora_Conference_TaLC_2018_Book_of_Abstracts.
- Trévisiol-Okamura P., Komur-Thillooy G. (2011, a c. di), *Discours, acquisition et didactique des langues: Les termes d'un dialogue*, Orizons, Paris.
- Vališová P. (2014), *Teaching Czech Verbs on Elementary Level: Comparing Textbook Corpus and National Corpus Data*, contributo presentato alla 11a conferenza internazionale su *Teaching and Language Corpora* (TaLC11), Lancaster University, 20-23 luglio 2014. Disponibile online: <http://ucrel.lancs.ac.uk/talc2014/doc/TALC2014-abstract-book.pdf>.
- Vanelli L. (1999), "Punti di crisi" nell'italiano contemporaneo, in *Insegnare italiano nella scuola del 2000*, a c. di U. Cardinale, Unipress, Padova, pp. 99-121.
- Vedovelli M., Villarini A. (2003), *Dalla linguistica acquisizionale alla didattica acquisizionale: le sequenze sintattiche nei materiali per l'italiano L2 destinati agli immigrati stranieri* in *Verso l'italiano. Percorsi e strategie di acquisizione*, a c. di A. Giacalone Ramat, Carocci, Roma, pp. 270-304.
- Vedovelli M. (2003), *Note sulla glottodidattica italiana oggi: problemi e prospettive*, «Studi Italiani di Linguistica Teorica e Applicata», 18/2, pp. 173-197.
- Véronique D. (2000 a c. di), *Didactique des langues étrangères et recherches sur l'acquisition*, «Etudes de Linguistique Appliquée, 120 (numero monografico)», Trévisiol-Okamura, Komur-Thillooy (2011).
- Véronique D. (2005), *Les interrelations entre la recherche sur l'acquisition du français langue étrangère et la didactique du français langue étrangère*, «Acquisition et Interaction en Langue Etrangère/AILE», 23, pp. 9-41.

- Viganò P.B. (2011), *I corpora e il loro sfruttamento in didattica*, «Italiano LinguaDue», 2, pp. 115-128.
- Voghera M. (1992), *Sintassi e intonazione nell'italiano parlato*, Il Mulino, Bologna.
- Wandruszka U. (1991), *Frasi subordinate al congiuntivo*, in *Grande grammatica italiana di consultazione. Vol. II: I sintagmi verbale, aggettivale, avverbiale. La subordinazione*, a c. di L. Renzi, G. Salvi, A. Cardinaletti, Il Mulino, Bologna, pp. 415-481.
- Wible D. (2008), *Multiword expressions and the digital turn*, in *Phraseology in Language Learning and Teaching*, a c. di F. Meunier, S. Granger, John Benjamins, Amsterdam-Philadelphia, pp. 163-181.
- Widdowson H.G. (1991), *The description and prescription of language*, in *Georgetown University round table on language and linguistics. Linguistics and language pedagogy: The state of the art*, a c. di J.E. Alatis, Georgetown University Press, Washington D.C., pp. 11-24.
- Widdowson H.G. (2000), *On the limitations of linguistics applied*, «Applied linguistics», 21/1, pp. 3-25.
- Wilkins D. (1976), *Notional syllabus*, Oxford University Press, Oxford.
- Whittle A. (2015), *Focalizzare la forma: sviluppo della competenza linguistica nella classe multilingue della scuola primaria*, «Italiano LinguaDue», 2, pp. 13-35.
- Wunderlich C. (2018), *Creating a subject-specific corpus and academic word list for business informatics*, contributo presentato alla 13a conferenza internazionale su *Teaching and Language Corpora* (TaLC13), Cambridge University, 18-21 luglio 2018. Disponibile online: https://www.researchgate.net/publication/326548400_13th_Teaching_and_Language_Corpora_Conference_TaLC_2018_Book_of_Abstracts.
- Zamora P., Alessandro A. (2011), *La didáctica del italiano coloquial: viejos recelos, nuevos alicientes*, «Revista electrónica de estudios filológicos», 21. Disponibile online: <https://www.um.es/tonosdigital/znum21/secciones/estudios-30-italiano.htm>
- Zamora P., Alessandro A., Hernández González M.B. (2010), *L'italiano, a me mi piace. Corso per l'apprendimento dell'italiano colloquiale come LS/L2*, Editum, Murcia.

- Zanchetta E., Baroni, M. (2005), *Morph-it! A free corpus-based morphological resource for the Italian language*, in *Proceedings from the Corpus Linguistics Conference 1* (1). Atti del convegno (University of Birmingham, luglio 2005), Conference e-journal disponibile online: <https://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-conf-e-journal.aspx>
- Zingaro A. (2018), *La "vita spericolata" della sintassi: il che polivalente nella canzone italiana tra gli anni '70 e '90 del Novecento*, «Italcia Wratislaviensia», 9/2, pp. 305-325.
- Zingaro A. (2021), *Progettazione e sperimentazione di unità didattiche sul neostandard per l'insegnamento dell'italiano a stranieri*, [tesi di dottorato], Alma Mater Studiorum - Università di Bologna. Dottorato di ricerca in Traduzione, interpretazione e interculturalità, 33 Ciclo. Disponibile online: <http://amsdottorato.unibo.it/9745/>. DOI 10.48676/unibo/amsdottorato/9745.

Grammatiche

- Antonelli G., Picchiorri E. (2016), *L'italiano, gli italiani*, Einaudi scuola, Torino.
- Battaglia S., Pernicone, V. (1980 [1951]), *La grammatica italiana*, Loescher, Torino.
- Dardano M., Trifone P. (1985), *La lingua italiana*, Zanichelli, Bologna.
- Fogarasi M. (1983 [1969]), *Grammatica italiana del Novecento*, Bulzoni, Roma.
- Fogliato S. (2015), *L'italiano competente*, Loescher, Torino.
- Renzi L., Salvi G., Cardinaletti A. (2001-2), *Grande grammatica italiana di consultazione, Vol. I-III*, Il Mulino, Bologna.
- Salvi G., Vanelli L. (2004), *Nuova grammatica italiana*, Il Mulino, Bologna.
- Schmitt-Jensen J. (1970), *Subjonctif et hypotaxe en italien: une esquisse de la syntaxe du subjonctif dans les propositions subordonnées en italien contemporain*, Odense University Press, Odense.
- Sensini M. (1990), *La grammatica della lingua italiana*, Mondadori, Milano.
- Serianni L. (1989b), *Grammatica italiana*, con la collaborazione di A. Castelveccchi, UTET, Torino.

-
- Serianni L., Della Valle V., Patota G. (2016), *Italiano plurale*, Bruno Mondadori, Milano.
- Sorrento L. (1951), *Sintassi romanza. Ricerche e prospettive*, Istituto Editoriale Cisalpino Varese, Milano.

SITOGRAFIA

CORPORA E PAGINE DI ISTITUTI DI RICERCA E STRUMENTI DI CONSULTAZIONE AD ESSI CORRELATI

Aranea Corpora

<http://ucts.uniba.sk/>

http://ucts.uniba.sk/aranea_about/

<https://www.korpus.cz/kontext/query?corpname=syn2020>

Araneum Italicum

http://ucts.uniba.sk/aranea_about/_italicum.html

Athenaeum Corpus

<http://www.bmanuel.org/projects/at-HOME.html>

http://www.corpora.unito.it/ateneo/ateneo_.php

<http://iris.unito.it/handle/2318/120913#.WqaOuExFx1s>

CLIPS

<http://www.clips.unina.it/it/>

COLFIS

<https://www.istc.cnr.it/grouppage/colfis>

<http://linguistica.sns.it/esploracolfis/home.htm>

<http://www.ge.ilc.cnr.it/strumenti.php>

CoLiTec

<https://corpora.dipintra.it/>

CorDIC

<http://corporadidattici.lablita.it/>

CORIS

<http://corpora.dslo.unibo.it/TCORIS/>

Corpus querying

<https://www.sketchengine.eu/documentation/corpus-querying/>

CQL Basics

<https://www.sketchengine.eu/documentation/cql-basics/>

Creative Commons

<http://creativecommons.org/licenses/>

ISST-TANL tagset

http://www.corpusitaliano.it/static/documents/POS_ISST-TANL-tagset-web.pdf

ItTenTen

<http://www.sketchengine.co.uk/itTenTen-italian-corpus/>

LABLITA

<https://www.letterefilosofia.unifi.it/vp-309-lablita.html>

la Repubblica

http://corpora.dipintra.it/public/run.cgi/first_form

LexALP

http://lexalp.eurac.edu/projects/corpus_en.htm

Laboratorio di linguistica - Scuola Normale Superiore

<http://linguistica.sns.it/CoLFIS/Testi/libri/libri.htm>

Morph-it

<https://docs.sslmit.unibo.it/doku.php?id=resources:morph-it>

NoSketch Engine

<http://nlp.fi.muni.cz/trac/noske>

NUNC

<http://www.bmanuel.org/projects/ng-HOME.html>

http://www.corpora.unito.it/index_nunc.php

PAISÀ

<http://www.corpusitaliano.it/it/>

Perugia Corpus

<https://apps.unistrapg.it/cqpweb/>

Repubblica Scuola

<http://scuola.repubblica.it/>

RIDIRE

<http://www.ridire.it/it.drwolf.ridire/home.seam>

RIDIRE-CPI

<https://github.com/lablita/ridire-cpi>

Sketch Engine

<http://www.sketchengine.co.uk/>

SpiderLing

<http://nlp.fi.muni.cz/trac/spiderling>

TenTen

<http://www.sketchengine.co.uk/documentation/tenten-corpora/>

TreeTagger

<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

WaCky

<http://wacky.sslmit.unibo.it/doku.php>

Wikipedia Extractor

http://wiki.apertium.org/wiki/Wikipedia_Extractor

PIATTAFORME DI E-LEARNING E VERIFICA DELL'APPRENDIMENTO**A-tutor**

<https://atutor.ca>

Blackboard

<https://www.blackboard.com/it-it>

Chamilo

<https://chamilo.org/it/>

Claroline

<https://www.claroline.com>

Docebo

<https://www.docebo.com/>

Dokeos

<https://www.dokeos.com/it/>

E-Front

<https://www.efrontlearning.com/>

Ilias

<https://www.ilias.de/>

Moodle

<https://moodle.org>

Sakai

<https://www.sakailms.org/>

Schoology

<https://www.schoology.com/>

Socrative

<https://socrative.com>

Wizer.me

<https://app.wizer.me/>

APPENDICE 1

Panoramica generale dei principali dati raccolti sui corpora presi in esame (cfr. capitolo 4)

Corpora	Date dei contenuti ¹	Numero di <i>token</i>	Informazioni generi testuali	Disponibilità	Licenze <i>Copyright</i>
1. Araneum Italicum Matus	Scaricati dal web nel 2014.	1,2 mld.	Dati non reperiti.	A seconda delle piattaforme, accessibile previa registrazione o abbonamento.	“The copyright status of the data is not clear”. (Benko, 2016, p. 4248)
2. Araneum Italicum Minus	Scaricati dal web nel 2014.	119 mln.	Come sopra.	Come sopra.	Come sopra.
3. Athenaeum	Dati non reperiti.	306 mila	Prosa accademica (scientifica e amministrativa).	Liberamente accessibile online.	<i>Creative Commons – Share Alike</i> .
4. CLIPS	Dati non reperiti.	Dati non reperiti.	Dati non reperiti.	Consultabile online previa registrazione gratuita.	Dati non reperiti.
5. COLFIS	Prodotti tra il 1992 e il 1994.	3,7 mln.	Quotidiani, periodici, libri.	Liberamente accessibile online.	Non sono state concesse licenze per la totalità dei testi.

¹ Ove disponibili, sono state riportate le informazioni relative alle date di *pubblicazione* dei testi. In caso contrario, sono state riportate solo le informazioni relative al periodo in cui è stata svolta la *raccolta* dei testi, che consentono perlomeno di fissare con certezza un limite oltre il quale non sono stati raccolti (e dunque prodotti) i testi che sono entrati a far parte del corpus.

Corpora	Date dei contenuti ¹	Numero di <i>token</i>	Informazioni generi testuali	Disponibilità	Licenze <i>Copyright</i>
6. CorDIC	Dati non reperiti.	500 mila	Dominio d'uso: arti, burocratico, creativo, economia e stampa.	Liberamente accessibile online.	Dati non reperiti.
7. CorIS/CoDIS	Prodotti tra gli anni '80 e il 2017.	150 mln.	Stampa, narrativa, prosa accademica, prosa giuridico-amministrativa, miscellanea ed <i>ephemera</i> .	Liberamente accessibile online.	Non sono state concesse licenze, o perlomeno non per la totalità dei testi.
8. ItTenTen	Scaricati dal web nel 2016.	5,8 mld	Dati non reperiti.	Accesso gratuito per 30 giorni. In alternativa, varie opzioni di abbonamento.	Dati non reperiti.
9. ItWaC	Scaricati dal web tra il 2005 e il 2007.	2 mld.	Dati non reperiti.	Liberamente accessibile online.	"There is no easy way of determining whether the content of a particular page is copyrighted". (Baroni et al., 2009, p. 224)

Corpora	Date dei contenuti ¹	Numero di <i>token</i>	Informazioni generi testuali	Disponibilità	Licenze <i>Copyleft</i>
10. la Repubblica	Prodotti tra il 1985 al 2000.	326 mln.	Prosa giornalistica.	Liberamente accessibile online.	Nessuna.
11. LexALP	Dati non reperiti.	3000 ²	Linguaggio giuridico.	Liberamente accessibile online.	Tattandosi di testi giuridici, non sussiste la questione del copyright.
12. LIF	Prodotti tra il 1947 e il 1968.	500.000 ³	Teatro, romanzi, cinema, periodici, sussidiari.	Non accessibile né online, né offline.	Nessuna.
13. NUNC	Prodotti tra il 2002 e il 2004.	Due sezioni: 127 mln. e 109 mln. di <i>token</i> .	Forum di cinema, cucina, fotografia e motori.	Liberamente accessibile online.	La risorsa dispone di licenza <i>Creative Commons – Share Alike</i> , ma non i singoli testi che la compongono.

² Si riferisce al numero di documenti e non di token.

³ Lemmi e non token.

Corpora	Date dei contenuti ¹	Numero di token	Informazioni generi testuali	Disponibilità	Licenze <i>Copyright</i>
14. PAISÀ	Scaricati dal web nel 2010.	250 mln.	Pagine tratte dai progetti Wikimedia Foundation, guide.supereva.it, italy.indymedia.org e vari blog.	Liberamente accessibile online.	<i>CC-Attribuzione, CC-Attribuzione-Condividi allo stesso modo, CC-Attribuzione-Non commerciale, o CC-Attribuzione-Non commerciale-Condividi allo stesso modo.</i>
15. Perugia Corpus	Prodotti tra il 1990 e il 2012.	26 mln.	Letteratura, saggistica, stampa, scritto accademico, scritto scolastico, scritto amministrativo, web.	Accessibile previa registrazione.	Nessuna.
16. RIDIRE	Scaricati dal web tra il 2009 e il 2013.	1,5 mld.	Informazione, economia e business, amministrazione e legislazione sport, architettura e design, cucina, cinema, musica, moda, arti visive, religione, letteratura e teatro.	Liberamente accessibile online per le funzioni di ricerca; è richiesto il log-in per le funzioni di gestione, monitoraggio e aggiornamento.	Dati non reperiti.

APPENDICE 2

Panoramica generale delle stringhe di ricerca prodotte (cfr. capitolo 4)

FENOMENO GRAMMATICALE	SOTTOFENOMENO	STRINGA DI RICERCA
Pronome <i>gli</i> vs <i>le</i> (<i>a lei</i>) / <i>loro</i> (<i>a loro</i>)		[word="(?)gli"] [postag="V.*"]
Il partitivo preceduto da preposizione (es.: <i>con degli amici</i> ; <i>per degli amici</i>)		[postag="E" & word="prima dopo"] [postag="EA" & lemma="di"]
Dislocazione a sinistra	Doppio partitivo (Es.: <i>di amici ne ho molti</i>)	[postag="S.* A.*"] [word="(?)di"] [postag="S.*" & word!="fatto seguito conseguenza normal solito"] [word="ne"] [postag="V.*"]
dislocazione a sinistra dell'elemento pronominale	A me mi	[word="(?)a"] [word="me"] [word="mi"]
	A te ti	[word="(?)a"] [word="te"] [word="ti"]
	A lui gli	[word="(?)a"] [word="lui"] [word="gli"]
	A lei le	[word="(?)a"] [word="lei"] [word="le"]
	A lei gli	[word="(?)a"] [word="lei"] [word="gli"]
	A noi ci	[word="(?)a"] [word="noi"] [word="ci"]
	A voi vi	[word="(?)a"] [word="voi"] [word="vi"]
	A loro gli	[word="(?)a"] [word="loro"] [word="gli"]
		[word="(?)per"] [word="cui"]
<i>Per cui</i> in funzione di congiunzione causale-consecutiva		

FENOMENO GRAMMATICALE	SOTTOFENOMENO	STRINGA DI RICERCA
Impiego di congiunzioni coordinanti e di connettivi testuali in posizione iniziale di discorso o di enunciato	<i>Ma</i> in posizione iniziale	[word="\.\, \,;\"] [word="(?)Ma"]
	<i>E</i> in posizione iniziale	[word="\.\, \,;\"] [word="(?)E"]
	<i>Allora</i> in posizione iniziale	[word="\.\, \,;\"] [word="(?)Allora"]
	<i>Comunque</i> in posizione iniziale	[word="\.\, \,;\"] [word="(?)Comunque"]
<i>Cosa</i> come interrogativo al posto di <i>che cosa</i>	Nell'interrogativa diretta	[cpostag="F"] [postag="A.* D.* R.*"]? [word="(?)cosa"] [word="\, \,;\"] {0,10} [word="\, \,;\"]
	Nell'interrogativa indiretta	[lemma="chiedere domandare"] [cpostag="F"] & word!="che"]? [word="(?)cosa"]
Soggetto in posizione rematica rispetto al predicato	con discordanza di numero rispetto al soggetto: <i>niente</i> seguito da un sostantivo plurale (Es.: <i>niente soldi</i>)	[word="niente"] [postag="S.*"] & feats=".*num=p.*"]
	sostantivo plurale seguito da <i>non ce n'è</i> (Es.: <i>soldi non ce n'è</i>)	[postag="S.*"] & feats=".*num=p.*"] [word="non"] [word="ce"] [word="n"] [lemma="essere"] & feats=".*num=s.*" & feats=".*per=3.*"]

Soggetto in posizione rematica	rispetto a verbi non composti (Es.: <i>Parla Luca</i> . Per ulteriori esempi si rinvia al capitolo 4)	[word="\." [postag="V.*" & word="[A-Z].+" & lemma="essere occorrere servire avere esistere" & deprel="ROOT" & feats=".*mod=i.*"] [postag="R.* A.* D.*"]{0,2} [postag="S.* P.*" & deprel="subj"]]
	rispetto a verbi composti con l'ausiliare <i>essere</i>	[word="\." [lemma="essere"] [postag="V.*" & feats=".*mod=p.*"] [postag="R.* A.* D.*"]{0,2} [postag="P.* S.*" & deprel="subj"]]
	rispetto a verbi composti con l'ausiliare <i>avere</i> , con oggetto diretto tra participio e soggetto	[word="\." [lemma="avere"] [postag="V.*" & feats=".*mod=p.*"] [postag="R.* A.* D.*"]{0,2} [postag="S.*" & deprel="obj"]] [postag="R.* A.* D.*"]{0,2} [postag="P.* S.*" & deprel="subj" & word!="che"]]
	rispetto a verbi composti con l'ausiliare <i>avere</i> (senza l'oggetto diretto tra participio e soggetto)	[word="\." [lemma="avere"] [postag="V.*" & feats=".*mod=p.*"] [postag="R.* A.* D.*"]{0,2} [postag="P.* S.*" & deprel="subj" & word!="che"]]
	rispetto a verbi passivi	[word="\." [lemma="essere venire andare"] [lemma="essere andare" & feats=".*mod=p.*"]? [postag="V.*" & feats=".*mod=p.*"] [postag="R.* A.* D.*"]{0,2} [postag="P.* S.*" & deprel="subj"]]
	in frase subordinata	[postag="CS" & word="[A-Z].*" [postag="V.*" & feats=".*mod=f.*"] {1,2} [postag="R.* A.* D.*"]{0,2} [postag="P.* S.*" & deprel="subj"]]

Frase scissa	verbo <i>essere</i> + gruppo nominale + che + verbo coniugato (es.: <i>è Anna che parla</i> . Per ulteriori esempi si rinvia al capitolo 4)	[postag="E.* C.*"] [lemma="essere"] [postag="R.* A.* D.*"]{0,2} [postag="P.* S.*"] [word="che"] [postag="V.*"]
	verbo <i>essere</i> + verbo all'infinito + che + verbo coniugato (es. <i>è votare che è importante</i>)	[postag="E.* C.*"] [lemma="essere"] [postag="V.*"] & feats=" .*mod=f.*" [word="che"] [postag="V.*"]
	verbo <i>essere</i> + avverbio + che + verbo coniugato (es. <i>è qui che abitiamo</i>)	[postag="E.* C.*"] [lemma="essere"] [postag="B.*"] [word="che"] [postag="V.*"]
	verbo <i>essere</i> + gruppo nominale preceduto da preposizione + che + verbo coniugato (es. <i>è a casa che dobbiamo andare</i> . Per ulteriori esempi si rinvia al capitolo 4)	[postag="E.* C.*"] [lemma="essere"] [postag="E.*"] [postag="A.* D.*"]{0,2} [postag="P.* S.*"] [word="che"] [postag="V.*"]

	verbo <i>essere</i> + gruppo nominale + a + verbo al modo infinito (es. <i>è lui a telefonare</i>)	[postag="E.* C.*"] [lemma="essere"] [postag="R.* A.* D.*"] {0,2} [postag="P.* S.*"] [word="a"] [postag="V.*"] & feats="." *mod=f.*"]
Soggetto in forma pronominale per indicare partecipazione affettiva con i verbi non composti (Es. <i>mi bevolterrò/bevvi</i> (ecc.) <i>il caffè</i>)	1a pers. sing.	[word="(?)i mi"] [lemma="mangiare leggere bere fare guardare vedere"] & feats="." *num=s.*" & feats="." *per=1.*" [postag="R.* A.* D.*"] {0,2} [postag="P.* S.*"] & deprel="obj"]
	2a pers. sing.	[word="(?)i ti"] [lemma="mangiare leggere bere fare guardare vedere"] & feats="." *num=s.*" & feats="." *per=2.*" [postag="R.* A.* D.*"] {0,2} [postag="P.* S.*"] & deprel="obj"]
	3a pers. sing.	[word="(?)i si"] [lemma="mangiare leggere bere fare guardare vedere"] & feats="." *num=s.*" & feats="." *per=3.*" [postag="R.* A.* D.*"] {0,2} [postag="P.* S.*"] & deprel="obj"]
	1a pers. plur.	[word="(?)i ci"] [lemma="mangiare leggere bere fare guardare vedere"] & feats="." *num=p.*" & feats="." *per=1.*" [postag="R.* A.* D.*"] {0,2} [postag="P.* S.*"] & deprel="obj"]
	2a pers. plur.	[word="(?)i vi"] [lemma="mangiare leggere bere fare guardare vedere"] & feats="." *num=p.*" & feats="." *per=2.*" [postag="R.* A.* D.*"] {0,2} [postag="P.* S.*"] & deprel="obj"]

	3a pers. plur.	[word="(?)si"] [lemma="mangiare leggere bere fare guardare vedere"& feats="." num=p.*" & feats="." per=3.*"] [postag="R. A. D. *"] {0,2} [postag="P.* S.*" & deprel="obj"]
Soggetto in forma pronominale per indicare partecipazione affettiva con i verbi composti (Es. <i>mi sonolero/fuì/sarei... bevuto i bevolberò/bevvi (ecc.) il caffè</i>)	1a pers. sing.	[word="(?)i mi"] [lemma="essere"& feats="." num=s.*" & feats="." per=1.*"] [lemma="mangiare leggere bere fare guardare vedere"& feats="." mod=p.*"] [postag="R. A. D. *"] {0,2} [postag="P.* S.*" & deprel="obj"]
	2a pers. sing.	[word="(?)ti"] [lemma="essere"& feats="." num=s.*" & feats="." per=2.*"] [lemma="mangiare leggere bere fare guardare vedere"& feats="." mod=p.*"] [postag="R. A. D. *"] {0,2} [postag="P.* S.*" & deprel="obj"]
	3a pers. sing.	[word="(?)si"] [lemma="essere"& feats="." num=s.*" & feats="." per=3.*"] [lemma="mangiare leggere bere fare guardare vedere"& feats="." mod=p.*"] [postag="R. A. D. *"] {0,2} [postag="P.* S.*" & deprel="obj"]
	1a pers. plur.	[word="(?)ci"] [lemma="essere"& feats="." num=p.*" & feats="." per=1.*"] [lemma="mangiare leggere bere fare guardare vedere"& feats="." mod=p.*"] [postag="R. A. D. *"] {0,2} [postag="P.* S.*" & deprel="obj"]

	2a pers. plur.	[word="(?)vi"][[lemma="essere"& feats=".*num=p.*"& feats=".*per=2.*"][[lemma="mangiare leggere bere fare guardare vedere"& feats=".*mod=p.*"]][postag="R. A. D.*"]{0,2} [postag="P.* S.*"&deprel="obj"]]
	3a pers. plur.	[word="(?)si"][[lemma="essere"& feats=".*num=p.*"& feats=".*per=3.*"][[lemma="mangiare leggere bere fare guardare vedere"& feats=".*mod=p.*"]][postag="R. A. D.*"]{0,2} [postag="P.* S.*"&deprel="obj"]]
Concordanza a senso tra il soggetto collettivo singolare seguito da un'espressione partitiva e il predicato plurale (Es.: <i>Un milione di elettori hanno votato</i>)		[postag="S.*"& feats=".*num=s.*"] [word="di"] [postag="S.*"& feats=".*num=p.*"] [postag="V.*"& feats=".*num=p.*"& feats!=".*mod=p.*"]
Ci "attualizzante"	C' + verbo <i>avere</i> + sostantivo (Es.: <i>C'ha voglia</i>)	[word="(?)c'"][[lemma="avere"]][postag="S.*"]
	Ci + verbo <i>avere</i> + sostantivo (Es.: <i>Ci ha voglia</i>)	[word="(?)ci"][[lemma="avere"]][postag="S.*"]
	C' + verbo <i>avere</i> + da + verbo (Es.: <i>c'ho da fare</i>)	[word="(?)c'"][[lemma="avere"]][word="da"] [postag="V.*"]

	Ci+ verbo <i>avere</i> + da + verbo (Es.: <i>ci ho da fare</i>)	[word="(:i)ci"] [lemma="avere"] [word="da"] [postag="V.*"]
<i>Che</i> polivalente	<i>che</i> esplicativo-consecutivo; <i>che</i> causale	[postag="V.*" & lemma="ricordare dimostrare raccontare parere sembrare sentire dire confidare convenire spere vedere volere sapere scrivere rispondere scoprire ribadire promettere mettere insistere dichiarare osservare sostenere immaginare temere ritenere premettere augurare pensare pretendere calcolare lasciare ripetere scommettere riconoscere significare vedere prevedere constatare supportare importare capire trovare decidere precisare aggiungere ammettere notare credere chiedere aspettare considerare bisognare occorrere avvenire bastare essere"] [word="che"] [postag="V.*"]
	<i>Che</i> pseudorelativo	[postag="P.*" & word="si"] [lemma="vedere sentire guardare scorgere notare osservare mirare ascoltare udire intendere percepire"] [word="che"] [postag="V.*"]
	<i>Che</i> consecutivo-presentativo con avverbi di dubbio	[word=".\ ; \ ? \ ,""] [word="(:i)che"] [word="selforse chissà eventualmente probabilmente poi magari"] [] {0,1} [postag="V.*"]

Impiego dell'indicativo al posto del congiuntivo in alcune subordinate	dopo i verbi di opinione/sentimento	[lemma="pensare credere ritenere sembrare pare supporre dubitare immaginare temere sperare dispiacere stupirsi augurarsi aspettarsi volere"] [word="che"] [word!="\."] {0,5} [postag="V.*" & feats="." *mod=i.*"]
	dopo i verbi di opinione/sentimento preceduti da negazione	[word="non"] [lemma="pensare credere ritenere sembrare pare supporre dubitare immaginare temere sperare dispiacere stupirsi augurarsi aspettarsi volere sapere"] [word="che"] [word!="\."] {0,5} [postag="V.*" & feats="." *mod=i.*"]
	dopo la costruzione “verbo <i>essere</i> + aggettivo + <i>che</i>”:	[lemma="essere"] [cpostag="A"] [word="che"] [word!="\."] {0,5} [postag="V.*" & feats="." *mod=i.*"]
	dopo la costruzione “verbo <i>avere</i> + paura/timore + <i>che</i>”:	[lemma="avere"] [word="paura timore"] [word="che"] [word!="\."] {0,5} [postag="V.*" & feats="." *mod=i.*"]
	nell'interrogativa indiretta	[lemma="chiedere domandare"] [word="se"] [word!="\."] {0,5} [postag="V.*" & feats="." *mod=i.*"]
	nelle ipotetiche	[postag="CS" & word="(?)se"] [word!="ne stress(o) i e"] {0,5} [postag="V.*" & feats="." *mod=i.*" & feats="." *ten=[i].*"] [word!="\."] {0,5} [postag="V.*" & feats="." *mod=i.*" & feats="." *ten=[i].*"]

APPENDICE 3

Valutazione dell'unità didattica sui verbi riflessivi "affettivi"

Cosa pensi delle attività di oggi? Rispondi a questo questionario.

Le tue risposte sono utili all'insegnante per programmare le future attività per gli studenti. **Grazie!** ☺

Tre domande su di te:

1. Qual è la tua lingua madre? _____
2. Quali altre lingue conosci? _____
3. Da quanto tempo studi l'italiano? _____

I contenuti di questa unità:

4. Prima delle attività di oggi, avevi già trovato degli esempi di uso dei verbi riflessivi in senso affettivo (esempio: Mi bevo un caffè)?

5. Se sì, in quale contesto? Segna con una crocetta una o più risposte:
A. Parlando con gli italiani
B. TV, giornali, libri ecc.
C. Sui social network
D. Chat/messaggi
E. Altro: _____
6. Esiste un fenomeno simile nella tua lingua? _____
7. La tua lingua madre o altre lingue che conosci ti hanno aiutato/a a imparare/capire meglio l'argomento di grammatica di oggi?

Opinioni sulle attività:

Metti una crocetta nelle tabelle qui sotto per dare la tua opinione.

Risposte possibili:

- ☹☹ 0: no
- ☹/😊 1: poco
- 😊 2: abbastanza
- 😊😊 3: sì

È stata utile questa attività?	☹☹ 0	☹/😊 1	😊 2	😊😊 3
8. Leggere il testo “L'appuntamento”				
9. Esercizi di collegamento (parole del testo – significato)				
10. Rispondere alle domande sul tema del testo				
11. Cercare nel testo i verbi che descrivono la cura del corpo				
12. Completare la coniugazione dei verbi riflessivi				
13. Ripassare le regole sulla posizione dei pronomi				
14. Distinguere le diverse azioni espresse dai verbi riflessivi (<i>azione del soggetto su se stesso/ su parti del corpo o vestiti/accessori che appartengono al soggetto; azione reciproca tra più soggetti; riflessivo “affettivo”</i>)				
15. Scegliere tra la forma riflessiva-affettiva e la forma non riflessiva				
16. Contenuto extra: i riflessivi nei modi di dire				
17. Fare esercizi basati su frasi vere della vita quotidiana (tratte da blog e altre pagine web) che non sono state inventate per “fare grammatica”				

Se vuoi, qui puoi spiegare meglio le tue risposte, aggiungere informazioni o consigli:

Dopo queste attività ...	 	 / 		 
	0	1	2	3
18. ho un'idea più chiara dell'uso dei riflessivi				
19. riesco a distinguere meglio le situazioni formali da quelle informali				
20. ho capito che in alcune situazioni informali è accettabile usare i riflessivi "affettivi"				
21. ho ancora dei dubbi sull'uso dei riflessivi				

Se vuoi, qui puoi spiegare meglio le tue risposte, aggiungere informazioni o consigli:

Opinioni sulla piattaforma:	 	 / 		 
	0	1	2	3
22. Il login e l'accesso alle diverse parti dell'unità è stato facile				
23. Le istruzioni delle attività erano comprensibili				
24. Le attività si visualizzano bene sullo schermo				
25. È stato utile avere esercizi con correzione automatica				

Se vuoi, qui puoi spiegare meglio le tue risposte, aggiungere informazioni o consigli:

Valutazione dell'unità didattica sull'uso del congiuntivo vs indicativo

Cosa pensi delle attività di oggi? Rispondi a questo questionario.

Le tue risposte sono utili all'insegnante per programmare le future attività per gli studenti. **Grazie!** 😊

Tre domande su di te:

1. Qual è la tua lingua madre? _____
2. Quali altre lingue conosci? _____
3. Da quanto tempo studi l'italiano? _____

I contenuti di questa unità:

4. Prima delle attività di oggi, avevi già trovato degli esempi di uso dell'indicativo al posto del congiuntivo in situazioni informali?

5. Se sì, in quale contesto? Segna con una crocetta una o più risposte:
 - Parlando con gli italiani
 - Nei mass media
 - Sui social network
 - Chat/messaggi
 - Altro _____
6. Esiste un fenomeno simile nella tua lingua? _____
7. La tua lingua madre o altre lingue che conosci ti hanno aiutato/a a imparare/capire meglio l'argomento di grammatica di oggi?

Opinioni sulle attività:

È stata utile questa attività?

Metti una crocetta per dare la tua opinione su quanto sono state utili le attività. Risposte possibili:

- 😞😞 0: no
- 😞/😊 1: poco
- 😊 2: abbastanza
- 😊😊 3: sì

ATTIVITÀ	  0	   1	 2	  3
8. Cercare nei 4 brevi testi gli esempi d'uso del congiuntivo				
9. Ripassare i casi in cui il congiuntivo è obbligatorio				
10. Fare ipotesi sulla scelta tra congiuntivo e indicativo in base al contesto formale/informale				
11. Fare esercizi basati su frasi vere della vita quotidiana (tratte da blog e altre pagine web) che non sono state inventate per “fare grammatica”				

Se vuoi, qui puoi spiegare meglio le tue risposte, aggiungere informazioni o consigli:

Dopo queste attività ...	SÌ	NO
12. ho un'idea più chiara di quando è obbligatorio il congiuntivo		
13. riesco a distinguere meglio le situazioni formali da quelle informali		
14. ho capito che in alcune situazioni informali è accettabile non usare il congiuntivo		
15. ho ancora dei dubbi sulla scelta tra congiuntivo e indicativo		

Se vuoi, qui puoi spiegare meglio le tue risposte, aggiungere informazioni o consigli:

Opinioni sulla piattaforma:	SÌ	NO
16. Il login e l'accesso alle diverse parti dell'unità è stato facile		
17. Le istruzioni delle attività erano comprensibili		
18. Le attività si visualizzano bene sullo schermo		
19. È stato utile avere esercizi con correzione automatica		
20. Mi è piaciuta l'idea delle attività online		

Se vuoi, qui puoi spiegare meglio le tue risposte, aggiungere informazioni o consigli:

Ultimi volumi pubblicati:

Francesca Biagini, *L'espressione della finalit  in russo. Uno studio contrastivo con l'italiano*.

Francesca Biagini e Svetlana Slavkova (a cura di), *Contributi italiani allo studio della morfosintassi delle lingue slave*.

Elio Ballardini, *Traduire devant la justice p nale. L'interpr te traducteur dans les codes de proc dure p nale italiens aux XIX  et XX  si cles*.

Ahmad Al-Addous e Sara Nanni (a cura di), *Drammi. Due opere teatrali di Sa'd Allah Wannus*.

Rachele Antonini (a cura di), *La mediazione linguistica e culturale non professionale in Italia*.

F lix San Vicente, Ana Lourdes de H riz, Mar a Enriqueta P rez V zquez (eds), *Perfiles para la historia y cr tica de la gram tica del espa ol en Italia: siglos XIX y XX. Confluencia y cruces de tradiciones gramaticogr ficas*.

Amalia Amato e Gabriele Mack, *Comunicare tramite interprete nelle indagini di polizia. Implicazioni didattiche di un'analisi linguistica*.

Gloria Bazzocchi e Raffaella Tonin (a cura di), *Mi traduci una storia? Riflessioni sulla traduzione per l'infanzia e per ragazzi*.

Francesca Gatta (a cura di), *Parlare insieme. Studi per Daniela Zorzi*.

Adele D'Arcangelo (a cura di), *Promuovere la competenza interculturale nella didattica della traduzione. L'esperienza della Scuola Interpreti e Traduttori di Forl *.

Roberta Pederzoli, Licia Reggiani, Laura Santone (dir.), *M dias et bien- tre. Discours et repr sentations*.

Elide Casali, *Il bambino e la lumaca. Rileggere Piero Camporesi (1926-1997)*.

Derek Boothman (a cura di), *La traduzione come luogo di incontro e di scontro*.

F lix San Vicente, Gloria Bazzocchi, Pilar Capanaga (eds), *Oraliter. Formas de la comunicaci n presencial y a distancia*.

Motoko Ueyama, Irena Srdanovi  (eds), *Digital resources for learning Japanese*.

Raffaella Baccolini, Roberta Pederzoli, Beatrice Spallaccia (eds), *Literature, Gender and Education for Children and Young Adults / Litt rature, genre,  ducation pour l'enfance et la jeunesse*.

Adele D'Arcangelo, Chiara Elefante, Valeria Illuminati (eds), *Translating for Children Beyond Stereotypes / Traduire pour la jeunesse au-del  des st r otypes*.

Mariachiara Russo, Emilia Iglesias Fern ndez, Mar a Jes s Gonz lez Rodr guez (eds), *Telephone interpreting*.

Anabela Cristina Costa da Silva Ferreira, *A l ngua do Direito de Fam lia*.

Greta Zanoni, *Lo sviluppo della competenza pragmatica in italiano: dalla rete alla classe*.

Patrick Leech, *Cosmopolitanism, dissent, and translation. Translating radicals in eighteenth-century Britain and France*.

Valeria Illuminati, *La traduzione dei classici per l'infanzia in una prospettiva di genere*.

Gloria Bazzocchi et al. (coords), *Nosotros somos nos y somos otros. Estudios dedicados a F lix San Vicente*.

Serena Zuccheri (ed), *Studies on Chinese Language and Linguistics in Italy*.

Finito di stampare nel mese di novembre 2023
per i tipi di Bologna University Press