

CLUB

Working Papers in Linguistics

A cura di Silvia Ballarè e Caterina Mauri

Volume 9, 2025



CLUB – CIRCOLO LINGUISTICO DELL'UNIVERSITÀ DI BOLOGNA

ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA

CLUB

Working Papers in Linguistics

A cura di Silvia Ballarè e Caterina Mauri

Volume 9, 2025

CLUB – CIRCOLO LINGUISTICO DELL'UNIVERSITÀ DI BOLOGNA

ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA

Collana
CLUB WORKING PAPERS IN LINGUISTICS
(CLUB-WPL)

Comitato di Direzione

Cristiana De Santis
Nicola Grandi
Francesca Masini
Fabio Tamburini

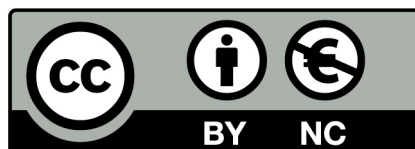
Comitato Scientifico

Marianna Bolognesi
Claudia Borghetti
Gloria Gagliardi
Chiara Gianollo
Elisabetta Magni
Yahis Martari
Caterina Mauri
Marco Mazzoleni
Emanuele Miola
Rosa Pugliese
Mario Vayra
Matteo Viale

*Il CLUB – Circolo Linguistico dell'Università di Bologna nasce nel 2015
con l'obiettivo di riunire coloro che, all'interno dell'Alma Mater,
svolgono attività di ricerca in ambito linguistico.*

*Il CLUB organizza ogni anno un ciclo di seminari e pubblica una selezione degli
interventi nella collana CLUB WORKING PAPERS IN LINGUISTICS.*

*I volumi, sottoposti a una procedura di peer-review, sono pubblicati online sulla
piattaforma AMS Acta dell'Università di Bologna e sono liberamente accessibili.*



CC BY-NC

CLUB Working Papers in Linguistics, Volume 9, 2025

ISBN: 9788854972148

ISSN: 2612-7008

DOI: <https://doi.org/10.6092/unibo/amsacta/8647>

CLUB – CIRCOLO LINGUISTICO DELL'UNIVERSITÀ DI BOLOGNA [http://
corpora. iclit.unibo.it/CLUB](http://corpora.iclit.unibo.it/CLUB)

CLUB Working Papers in Linguistics

Volume 9

a cura di *Silvia Ballarè e Caterina Mauri*

Indice

Presentazione <i>Silvia Ballarè e Caterina Mauri</i>	6
La variazione individuale nell'uso dei segnali discorsivi in spagnolo L2 <i>Valeria Baruzzo</i>	8
Il corpus IT_SEAH per la didattica della comunicazione accademico-disciplinare nel settore dell'Architettura e delle Costruzioni <i>Mariapia D'Angelo</i>	34
L'italiano delle traduzioni come varietà di contatto: un approccio tipologico a un corpus di narrativa <i>Emanuela De Vita</i>	54
Proto-ausiliari e verbi supporto nella costruzione dell'interlingua di migranti adulti con alfabetismo limitato <i>Egle Mocciaro</i>	81
Reference in teenagers' narratives between plurilingualism and restandardisation <i>Nicola Perugini ed Eleonora Zucchini</i>	99

SEZIONE CLUB PROGETTI

Per un archivio dell'italiano regionale e dei dialetti emilano-romagnoli <i>Silvia Ballarè, Isabella Colliva ed Emanuele Miola</i>	125
---	-----

Il progetto “AKiD – Acoustic and kinematic characteristics of speech in Dementia <i>Elena Martinelli, Gloria Gagliardi, Sabina Capellari, Chiara Meluzzi, Yuka Naito e Maria Ranù</i>	133
Il progetto DiverSIta. Documentare la diversità nell’italiano parlato <i>Caterina Mauri, Eugenio Gorla, Beatrice Bernasconi ed Eleonora Zucchini</i>	147
Il progetto “Racconta – Competenze narrative e sviluppo della lettoscrittura tra variazione interna e plurilinguismo” <i>Nicola Perugini e Chiara Gianollo</i>	157

Presentazione

Silvia Ballarè

Università di Bologna

silvia.ballare@unibo.it

Caterina Mauri

Università di Bologna

caterina.mauri@unibo.it

Con questo volume, la collana *CLUB Working Papers in Linguistics* prosegue il proprio percorso come spazio di confronto e di condivisione scientifica all'interno della comunità dei linguisti e delle linguiste dell'Università di Bologna, raccogliendo una selezione di contributi che riflettono la varietà di temi, dati e approcci emersi nel corso delle attività del CLUB.

Il CLUB – Circolo Linguistico dell'Università di Bologna, attivo dal 2015 – riunisce studiose e studiosi che, all'interno dell'Ateneo, svolgono attività di ricerca in ambito linguistico. Nel corso dell'anno accademico 2023-2024, il programma delle iniziative del CLUB è stato piuttosto ricco (il programma è disponibile [qui](#)). Sono stati organizzati seminari a cadenza regolare a cui hanno preso parte colleghi e colleghe dell'Università di Bologna e di diversi altri atenei, italiani ed esteri. Come da tradizione, hanno poi avuto luogo il *Club Day*, intitolato “La lingua che cambia: uno sguardo dinamico a variazione e mutamento” e il *Club Tesi* in cui, tra i diversi elaborati magistrali di ambito linguistico presentati, i membri del CLUB hanno selezionato come più meritevole quello ad opera di Emanuela De Vita. Inoltre, nella programmazione, hanno trovato spazio due novità: una giornata dedicata ai progetti di ricerca in corso nei diversi Dipartimenti e alcuni incontri organizzati dai dottorandi di linguistica dell'Università di Bologna.

Questo volume, che si inserisce nella collana *CLUB Working Papers in Linguistics*, raccoglie alcuni dei contributi presentati durante questo anno accademico, con l'obiettivo di dare visibilità editoriale a questo scambio continuo di idee, prospettive e pratiche di ricerca.

Il presente volume raccoglie contributi che affrontano questioni centrali per la linguistica contemporanea, con particolare attenzione alla variazione, all'uso e all'acquisizione delle lingue, alla linguistica dei corpora e allo studio del parlato. I saggi di ricerca inclusi in questa prima parte del volume sono firmati da Valeria Baruzzo, Mariapia D'Angelo, Emanuela De Vita, Egle Mocciaro, Nicola Perugini ed Eleonora Zucchini.

Nel dettaglio, Valeria Baruzzo analizza la variazione individuale nell'uso dei segnali discorsivi in spagnolo L2 da parte di apprendenti italofoeni in contesto di immersione. Mariapia D'Angelo presenta il corpus IT_SEAH, illustrandone la progettazione e le potenzialità per la didattica della comunicazione accademico-disciplinare nel settore dell'Architettura e delle Costruzioni. Emanuela De Vita propone un'analisi tipologica

dell'italiano delle traduzioni come varietà di contatto, basata su un corpus di narrativa. Egle Mocciaro si concentra sul ruolo dei proto-ausiliari e dei verbi supporto nella costruzione dell'interlingua di migranti adulti con alfabetismo limitato. Nicola Perugini ed Eleonora Zucchini esplorano infine le strategie di riferimento nelle narrazioni di adolescenti, mettendo in relazione plurilinguismo e processi di ristandardizzazione.

Accanto ai saggi di ricerca, il volume ospita una sezione *CLUB Progetti*, pensata come spazio di presentazione e valorizzazione di iniziative di ricerca collettive che coinvolgono membri del CLUB. In questa sezione trovano spazio diversi contributi. Silvia Ballarè, Isabella Colliva ed Emanuele Miola descrivono la progettazione di un archivio dell'italiano regionale e dei dialetti emiliano-romagnoli. Elena Martinelli, Gloria Gagliardi, Sabina Capellari, Chiara Meluzzi, Yuka Naito e Maria Ranù presentano il progetto *AKiD – Acoustic and kinematic characteristics of speech in Dementia*. Caterina Mauri, Eugenio Gorla, Beatrice Bernasconi ed Eleonora Zucchini illustrano il progetto *DiverSIta*, volto alla documentazione della diversità nell'italiano parlato. Infine, Nicola Perugini e Chiara Gianollo descrivono il progetto *Racconta*, dedicato allo sviluppo delle competenze narrative e della lettoscrittura in contesti di variazione interna e plurilinguismo.

Tutti i contributi inclusi nella prima sezione del volume sono stati sottoposti a revisione anonima ai fini della pubblicazione. Ringraziamo sentitamente le colleghe e i colleghi che hanno contribuito al processo di referaggio e tutti coloro che, nel corso delle attività del CLUB, hanno partecipato all'organizzazione degli incontri e dei progetti di ricerca. Il CLUB è gestito in modo collegiale e si sostiene grazie al contributo scientifico e organizzativo dei suoi membri, che condividono con noi la soddisfazione di vedere raccolti in questo volume i risultati delle proprie iniziative.

Ci auguriamo che questo nono volume dei *CLUB Working Papers in Linguistics* possa offrire spunti di riflessione utili e stimolanti, e continuare a rappresentare un luogo di dialogo aperto tra approcci, dati e tradizioni di ricerca diverse.

Bologna e Milano, 22 dicembre 2025

La variazione individuale nell'uso dei segnali discorsivi in spagnolo L2

Valeria Baruzzo

Università degli Studi di Padova

valeria.baruzzo2@gmail.com

Abstract

Questo studio esplora l'uso dei segnali discorsivi (SD) in spagnolo L2 da parte di apprendenti italofoni in contesto di immersione linguistica. In particolare, si concentra sull'effetto della variazione individuale nella conformazione dell'inventario di SD e nella scelta delle forme con cui realizzare le funzioni discorsive di tipo interazionale. Lo studio adotta un approccio qualitativo e si concentra sulle produzioni individuali di tre informanti selezionate da un corpus più ampio, che raccoglie dati conversazionali a partire da interviste semi-strutturate. I risultati dell'analisi evidenziano che la variabilità individuale ricopre un ruolo importante nella conformazione dei repertori di segnali discorsivi. Tuttavia, nella realizzazione di certe funzioni, si verifica una convergenza nelle strategie discorsive selezionate.

1. Introduzione

Questo articolo, adottando una prospettiva qualitativa, punta a esplorare la dimensione della variazione individuale¹ nello studio dell'acquisizione dei segnali discorsivi (d'ora in poi, SD). Di fatto, è ampiamente registrato che i SD sono fortemente sensibili alla variabilità individuale dei parlanti (Hasselgren 2002; Bazzanella 2015), al punto che, ricorrendo alle parole di Magliacane (2019: 228): "l'utilizzo dei marcatori discorsivi può essere considerato come parte dell'idioletto di ciascun parlante e il loro uso può variare da una persona all'altra". Questa peculiarità, tra le molte che contraddistinguono i SD, rappresenta una sfida importante per gli studi di stampo acquisizionale che si interessano dell'utilizzo di questi elementi da parte di non nativi nella lingua target (LT). Infatti, come riportano Koch & Thörle (2021), gli inventari di SD nella L2 spesso contengono forme insolite, risultato dell'interferenza della L1 o di altre lingue presenti nel repertorio individuale dei parlanti, oppure frutto di creazioni originali tipiche del singolo parlante. Pertanto, come suggerito da studi precedenti (Gilquin 2016; Magliacane 2019), condurre un'analisi individuale che consideri le eventuali differenze tra i singoli soggetti del campione permette di cogliere in modo più approfondito il comportamento dei SD nell'interlingua di apprendenti.

¹ In questo contributo si fa riferimento al concetto di "variazione individuale" così come descritto da Bülow & Pfenniger (2021: 1), ossia: "the observable variation within individual behavior".

Per cercare di soddisfare questa necessità di approfondimento, vengono analizzati dati conversazionali elicitati tramite intervista semi-strutturata a tre persone italofone residenti in Spagna. L'approccio all'analisi dei SD è di tipo onomasiologico² e si concentra sulle strategie discorsive impiegate per realizzare funzioni di tipo interazionale (presa di turno, manifestazione di accordo/disaccordo, controllo della ricezione, mitigazione, ecc.). L'obiettivo dell'analisi è osservare se i partecipanti ricorrono a strategie individuali per realizzare questo tipo di funzione o se, al contrario, possono essere rilevati dei pattern comuni.

L'articolo è organizzato come segue: nella prima sezione (§2, 2.1, 2.1.1) vengono brevemente descritti i segnali discorsivi, soffermandosi in particolare sulla presentazione di una tassonomia funzionale per la loro classificazione e ponendo particolare attenzione alla descrizione della funzione interazionale; a seguire (§3) viene riportato il quadro generale dell'acquisizione dei SD in L2, anche in riferimento alla questione della variazione individuale (§3.1) e al caso specifico dell'acquisizione dei SD in spagnolo L2 da parte di italofoni (§3.2); successivamente, in §4 e §5 viene presentata la metodologia impiegata in questo studio e l'analisi; in §6 vengono discussi i risultati ottenuti; infine, in §7 vengono presentate le riflessioni conclusive.

2. I segnali discorsivi

Proporre una definizione condivisa e cross-linguisticamente accettata di cosa si intenda per “segnali discorsivi” è impresa tutt'altro che semplice. Di fatto, non solo abbondano le definizioni (cfr. ad esempio la revisione critica di Pons Bordería 1998) ma persino la varietà di etichette terminologiche con cui gli studiosi scelgono di riferirsi a questa classe di elementi linguistici. Di conseguenza, risulta estremamente difficile orientarsi in questa “giungla” (Fischer 2006: 1). Senza addentrarsi in questo dibattito, una definizione valida e pertinente per questo contributo è quella proposta da Bazzanella (2010: 1139), secondo la quale:

si definiscono *segnali discorsivi* quegli elementi che, svuotandosi in parte del loro significato originario, assumono dei valori che servono a sottolineare la strutturazione del testo, a connettere elementi frasali, interfrasali, extrafrasali, a esplicitare la collocazione dell'enunciato in una dimensione interpersonale, ad evidenziare processi cognitivi. I segnali discorsivi non costituiscono una classe morfologica o lessicale, ma funzionale. Dal punto di vista grammaticale, appartengono a diverse categorie: congiunzioni, avverbi, sintagmi verbali, espressioni frasali.

Come si può evincere dalla definizione, i segnali discorsivi apportano dei “valori” che contengono un contributo semantico particolare. Questo contributo semantico viene

² A partire dalla funzione comunicativa specifica, si osservano le forme con cui viene realizzata tale funzione.

definito “procedurale”³ e consiste nella capacità di fornire le istruzioni per interpretare correttamente i significati e le relazioni tra un elemento e l’altro all’interno delle sequenze discorsive (Murillo Ornat 2010: 243). In merito a questo, è necessario specificare altri due aspetti che contribuiscono a rendere molto peculiare la classe dei segnali discorsivi. In primo luogo, va evidenziato il loro carattere polifunzionale: ogni SD può svolgere più funzioni nello stesso contesto (polifunzionalità sintagmatica) o essere associato a funzioni diverse in base ai contesti in cui compare (polifunzionalità paradigmatica). In secondo luogo, va evidenziata la sensibilità al contesto che, sempre Bazzanella (2015: 38) distingue in “globale” e “locale”. Le variabili costitutive del contesto globale, riporta Bazzanella (2015), corrisponderebbero a quelle di tipo sociolinguistico, quindi variabili di tipo diamesico, diafasico, diatopico, diastratico, diacronico a cui si aggiungono, come suggerisce Sansò (2020: 81), anche le situazioni di contatto linguistico. Per quanto riguarda invece il contesto locale, i SD possono variare in base a (i) contesto linguistico, ossia tutti i parametri linguistici che co-occorrono con il SD, quindi elementi come le pause, l’intonazione e la posizione sintattica e (ii) contesto enunciativo (Bazzanella 2015: 39), ossia la situazione concreta dell’interazione, che incide sulla forma linguistica selezionata e sull’attivazione di determinate funzioni del SD. L’approccio teorico proposto da Aijmer (2013) riprendendo Norén & Linell (2007) cerca di cogliere proprio questo aspetto. Nella sua proposta, infatti, i segnali discorsivi non avrebbero un significato fisso, ma: “a meaning potential which is available as a communicative resource to language users” (Aijmer 2013: 30). Secondo questa visione, quindi, il contesto sociale e interazionale in cui si svolge la comunicazione sarebbe responsabile dell’attivazione di uno tra i vari potenziali significati che il SD può veicolare.

2.1 Una tassonomia funzionale

Gli approcci adottati per descrivere le funzioni dei SD sono svariati e hanno dato esito a numerose classificazioni (cfr. ad esempio, Martín Zorraquino & Portolés Lázaro 1999; Pons Bordería 2006; Ghezzi & Molinelli 2014; Fuentes-Rodríguez 2023). In questo contributo è stata adottata la tassonomia di Borreguero Zuloaga (2015) che consiste in un’elaborazione di quella di López Serena & Borreguero Zuloaga (2010), formulata a partire da Bazzanella (1995; 2006; 2011). La tassonomia individua tre macro-funzioni discorsive, a cui sono associate varie sub-funzioni. Nello specifico, questa tripartizione prevede: (i) la funzione interazionale, che guida i rapporti tra gli interlocutori nell’interazione ed è regolata da due parametri, ossia il ruolo ricoperto dai partecipanti nella conversazione (emittente o ascoltatore) e l’intenzione che hanno i partecipanti di cambiare suddetto ruolo; (ii) la funzione meta-discorsiva, invece, raggruppa le sub-funzioni svolte dai SD che creano la struttura del discorso e ne facilitano la costruzione durante lo svolgimento; (iii) la funzione cognitiva, infine, è svolta dai SD che, da un lato, indicano le relazioni tra le varie componenti del discorso; dall’altro, suggeriscono l’atteggiamento del parlante di fronte a quanto sta emettendo o viene emesso nella conversazione (Borreguero Zuloaga 2015: 162-163).

³ Il significato procedurale si distingue da quello concettuale – veicolato, ad esempio, da elementi come *casa*, *andare*, *bello* – che forniscono informazioni lessicali sul contenuto proposizionale degli enunciati (Nadal et al. 2016: 59).

2.1.1 La funzione interazionale

All'interno della macro-funzione interazionale possono essere identificate varie tipologie di sub-funzioni. La tipologia di sub-funzioni dipende dal ruolo ricoperto nella conversazione dalla persona che emette il segnale discorsivo. La prima tipologia è composta dalle funzioni a cui ricorre la persona che ha il ruolo di emittente, e include sub-funzioni come: presa e cessione del turno, controllo della ricezione e comprensione del messaggio; mitigazione (Borreguero Zuloaga 2015: 161). La seconda tipologia di funzioni invece riguarda la persona che ha il ruolo di ascoltatore e vuole mantenere quel ruolo. In altre parole, vuole continuare a partecipare nella conversazione ma senza intervenire. Questa classe di funzioni include sub-funzioni come: espressione di emotività nei confronti di quanto dice l'emittente (sorpresa, rabbia, ecc.); conferma della comprensione e della ricezione del messaggio (Borreguero Zuloaga 2015: 161). Infine, la terza tipologia è composta dalle funzioni a cui ricorre la persona che ha il ruolo di ascoltatore e desidera invece intervenire nella conversazione. Comprende i SD che servono a introdurre: risposta collaborativa (o manifestazione di accordo); risposta reattiva (o manifestazione di disaccordo); richiesta di chiarimento (Borreguero Zuloaga 2015: 161-162).

3. Acquisizione dei SD in L2

Ad oggi non esiste una teoria condivisa sull'acquisizione dei SD in L2 (De Cristofaro & Badan 2021; De Marco 2023). Tuttavia, gli esperti concordano nell'affermare che l'acquisizione dei SD non avviene in modo lineare (Bello Viruega et al. 2019) e nemmeno seguendo fasi ben definite (Borreguero Zuloaga 2017; 2019). Gli studi condotti finora riportano che i SD sono già presenti nelle prime fasi dell'interlingua, ma il repertorio di forme è inizialmente molto limitato, soprattutto nei livelli più bassi di competenza (Andorno 2007; Bini & Pernas 2011; Solís García & León Gómez 2017; Borreguero Zuloaga 2019). In questi stadi, gli apprendenti dispongono di un insieme ristretto di segnali discorsivi che sono per lo più foneticamente semplici e formalmente e/o funzionalmente vicini ai segnali discorsivi presenti nella L1 (Andorno 2007; Koch & Thörle 2019). Nelle fasi successive, il repertorio di SD si amplia, anche se non in proporzione all'aumento di altre categorie di parole (Gillani & Pernas 2021) e comunque mai in maniera stabile, tanto che in alcuni casi si osserva un decremento di occorrenze nella produzione di SD anche tra parlanti di livello avanzato (Borreguero Zuloaga 2017). In alcuni casi, la presenza di SD formalmente e pragmaticamente simili nella L1 facilita l'acquisizione degli elementi nella L2 (Nogueira Da Silva 2010). In altri casi (Álvarez-Ejzenberg 2021; Di Nunno 2024) la somiglianza tra SD nella L1 e nella L2 porta gli apprendenti a formulare delle predizioni non adeguate sul funzionamento dei SD nella L2. Questo accade soprattutto nei casi in cui esiste un equivalente formale tra un elemento nella L1 e nella L2, ma tra le due forme non sussiste un rapporto di equivalenza pragmatica o esiste solo in parte. In questi casi, gli apprendenti tendono a incorrere in "prestiti pragmatici" funzionali di varia entità (Andersen 2014). In alcuni casi, l'interferenza della L1 può manifestarsi anche in termini di pattern di frequenza (Koch & Thörle 2019). In altre parole, se un SD viene usato molto frequentemente nella L1, allora è possibile che l'equivalente di quella forma nella L2 venga usato altrettanto frequentemente quando la persona si esprime nella lingua seconda in questione. Infine,

un elemento che sembra avere un effetto positivo nell'ampliamento dell'inventario formale e funzionale dei SD nella L2 è la possibilità di trascorrere un periodo di immersione nella comunità linguistica target (Hellerman & Vergun 2007; Diskin & Regan 2015; Liu 2016; De Cristofaro & Badan 2021; Thomas & Mitchell 2022), anche se gli studi non rivelano quanto stabile e duraturo sia il beneficio per gli apprendenti.

3.1 La variazione individuale nell'acquisizione dei SD in L2

Come anticipato nell'introduzione di questo contributo, l'uso dei SD da parte dei parlanti risponde a fattori caratterizzati da un valore fortemente idiosincratico. Questo valore non riguarda solo il contesto L2, ma riguarda anche l'uso nella L1. Di fatto, in riferimento al carattere idiosincratico dei segnali discorsivi in L1, Bazzanella (2015: 39) lo descrive come l'impiego "preferenziale, ripetuto e inconsapevole" da parte di un interlocutore di uno o più SD nel corso della conversazione. In contesto L2, questo comportamento linguistico viene spesso associato al concetto di *lexical teddy bear* coniato da Hasselgren (1994). Con questo termine, la studiosa si riferisce a "core words [...] learnt early, widely usable, and above all safe - because they do not show up as errors" (Hasselgren 1994: 250) che, tendenzialmente, vengono anche sovra-utilizzate da parte degli apprendenti. In questo senso, due fattori possono essere influenti. Da un lato l'"effetto ripetizione", ossia la "reoccurrence of a marker that frequently appeared in the preceding discourse" (Kerr-Barnes 1998: 205). Dall'altro, le strategie di evitamento (Schachter 1974: 212), cioè gli apprendenti tendono a evitare forme che non garantiscono loro "correttezza grammaticale" né "valutazione positiva" (de Santiago-Guervós 2020: 15). Questa caratteristica nell'uso dei SD da parte dei parlanti rappresenta un limite da tenere in considerazione. Infatti, molti studi evidenziano come questo costituisca un problema per la generalizzazione delle osservazioni. Ad esempio, Gilquin (2016) sottolinea come la variazione individuale degli usi dei SD nel suo campione abbia portato a dei risultati "sometimes messy". Götz & Mukherjee (2018), nella loro ricerca sull'effetto del soggiorno di studio all'estero da parte di studenti universitari germanofoni sull'uso dei SD in inglese, formulano considerazioni simili: il periodo all'estero e il contatto con i membri della comunità target hanno avuto un impatto positivo poiché hanno permesso di aumentare la frequenza d'uso dei SD nell'interlingua, ma la grande variabilità individuale nelle diverse produzioni non permette di tracciare delle proiezioni forti.

3.2 L'acquisizione dei SD in spagnolo L2 da parte di italofoni

L'acquisizione delle abilità comunicative e discorsive in spagnolo L2 da parte di italofoni non è stata ancora ben approfondita. Infatti, come riporta Bailini (2016: 64), la maggior parte degli studi che si sono interessati di questo target specifico di apprendenti si è concentrata su elementi di natura morfo-sintattica. I pochi studi dedicati all'acquisizione dei SD in spagnolo L2 da parte di italofoni (cfr. ad esempio, Pascual Escagedo 2013; 2015; Solís García & León Gómez 2017; Di Nunno 2024) suggeriscono che la profonda affinità linguistica tra italiano e spagnolo, se da un lato è molto positiva e vantaggiosa per la comprensione (Ainciburu 2008), non sembra esserlo così tanto nel caso dei segnali discorsivi. Di fatto, come sottolinea Della Putta (2015: 28), nonostante la forte convergenza delle strutture sintattiche che conferisce a italiano e spagnolo "the apparent

impression of two perfectly overlapping systems”, i punti in cui i due sistemi differiscono sono estremamente complessi da acquisire.⁴

Per quanto riguarda l’acquisizione delle funzioni, Pascual Escagedo (2015) fornisce indicazioni sull’incrementalità delle funzioni dei SD in rapporto al livello di competenza linguistica. In particolare, nota che le funzioni più impiegate in assoluto sono quelle ‘interazionali’, indipendentemente dal livello degli apprendenti. Tuttavia, nota che il livello degli apprendenti influisce sulla tipologia di funzioni attivate: quelle interazionali sono usate con più frequenza nei livelli più avanzati, mentre l’impiego di quelle cognitive decresce. Le funzioni meta-discorsive sono generalmente poco sfruttate, e tendono a dominare nei livelli iniziali. Sempre all’interno dell’ambito delle funzioni interazionali, focus di questo contributo, Solís García & León Gómez (2017), nel loro studio sull’acquisizione delle repliche affermative (*sí, vale, ya, de acuerdo*) da parte di studenti italofofoni, evidenziano un grado di fossilizzazione nell’uso e nella varietà di queste particelle anche nei livelli più avanzati di competenza. Infatti, le Autrici notano: “una modulación del sistema de L1 que intenta acercarse al de la L2 pero que no lo consigue” (Solís García & León Gómez 2017: 20). Pascual Escagedo (2013) invece confronta le produzioni di *bueno* in un gruppo di studenti italofofoni e in uno di nativi, e nota un generale infra-utilizzo di questo SD negli apprendenti italofofoni rispetto ai nativi. Per quanto riguarda le funzioni, nota che i parlanti nativi tendono a utilizzare *bueno* principalmente in posizione mediale, mentre gli apprendenti italiani lo collocano prevalentemente all’inizio dell’enunciato. Questi dati suggeriscono che, per gli apprendenti italiani, il SD *bueno* viene usato soprattutto come segnale per prendere la parola, mentre per i parlanti nativi spagnoli rappresenta più che altro una strategia collaborativa per esprimere accordo con l’interlocutore.

4. Obiettivi e metodologia

L’obiettivo di questo contributo è descrivere l’uso dei SD da parte di apprendenti italofofoni nel corso di conversazioni semi-strutturate in spagnolo. Per raggiungere questo obiettivo, vengono analizzati i repertori di SD impiegati da tre informanti, tenendo in considerazione le forme e le funzioni con cui intervengono nel discorso. L’analisi è focalizzata sui profili individuali, per cui si cercherà di riflettere sulle simmetrie o asimmetrie degli usi dei SD nelle singole conversazioni. In particolare, si cercherà di individuare se esistono delle strategie discorsive individuali o se l’uso dei SD è piuttosto da ricondurre a pattern più generali.

4.1 Corpus generale

L’analisi di questo contributo si basa su un corpus più ampio costituito da 20 conversazioni simmetriche semi-strutturate condotte in spagnolo da parte di italofofoni. Oltre alle interviste, ai partecipanti è stato chiesto di compilare un questionario socio-biografico formulato a partire dalle indicazioni di Dörnyei (2003) e dai modelli elaborati da Marian et al. (2007), Li et al. (2006; 2014), Anderson et al. (2018). Le risposte al

⁴ Naturalmente questa osservazione non è limitata al solo caso dei segnali discorsivi.

questionario hanno permesso di raccogliere informazioni dettagliate sui profili bio-linguistici dei partecipanti e, in particolare, sulle loro abitudini linguistiche.⁵

Le conversazioni si sono svolte online tramite piattaforme di videoconferenza e la maggior parte è stata videoregistrata. Le conversazioni hanno avuto una durata che variava da 20 a 50 minuti. Tuttavia, per l'analisi è stato selezionato un estratto di 10 minuti di conversazione. L'estratto selezionato corrisponde al momento in cui gli informanti raccontavano dell'arrivo a Madrid, in modo da avere conversazioni omogenee dal punto di vista tematico. Le conversazioni sono state trascritte adottando una versione semplificata delle convenzioni di Jefferson (2014). In particolare, sono stati annotati le pause, gli allungamenti vocalici e le sovrapposizioni dei turni. Le osservazioni in merito alle produzioni dei nativi, invece, provengono dalla selezione (n = 3) di interviste provenienti dal corpus PRESEEA-Madrid.⁶

4.2 Annotazione dei SD

Il sistema di annotazione si è basato sulla tassonomia che prevede la tripartizione tra funzione interazionale, meta-discorsiva e cognitiva (cfr. paragrafo 2.1). La procedura di annotazione ha previsto due fasi: (i) l'individuazione della sub-funzione svolta dall'elemento linguistico con valore procedurale;⁷ (ii) la riconduzione alla corrispondente macro-funzione. In questa fase dell'analisi è stata annotata solo la sub-funzione prominente, lasciando ad una seconda fase di analisi l'esplorazione di casi di polifunzionalità sintagmatica.

5. Analisi

La selezione delle conversazioni è stata operata sulla base dei diversi contesti di uso della L1 e della L2 e delle diverse gradualità delle abitudini linguistiche.⁸ In particolare, è stata selezionata una persona con abitudini linguistiche prevalentemente ispanofone (Inf_01), una persona con abitudini tendenzialmente italofone (Inf_17) e una che si colloca in una situazione intermedia (Inf_02). L'analisi individuale è presentata così come segue: in primo luogo si esplicita il profilo bio-linguistico dell'informante, per poi riportare il repertorio individuale dei SD elicitati nel corso della conversazione. Successivamente, ci si concentra sulle strategie discorsive con cui vengono realizzate le diverse funzioni di tipo interazionale.

⁵ Nel senso di usi prevalenti della L1, della L2 o di entrambe.

⁶ Corpus creato per il *Proyecto para el Estudio Sociolingüístico del español de España y de América* (PRESEEA). Comprende interviste semi-strutturate a parlanti ispanofoni, tenendo conto della variazione diatopica e diastratica.

⁷ Sono state consapevolmente escluse dall'annotazione le interiezioni (ah, eh, oh) e le occorrenze di y.

⁸ Il campione da cui proviene questa selezione, infatti, è stato distribuito in un continuum sulla base delle diverse situazioni di uso della L2 e delle abitudini linguistiche. Gli estremi di questo continuum sono, da un lato, un contesto in cui le interazioni avvengono quasi totalmente in spagnolo e, dall'altro, uno in cui si svolgono quasi interamente in italiano.

Inf_01

Al momento della raccolta dati viveva a Madrid da 9 anni, ma aveva già trascorso un periodo risiedendo in un'altra città spagnola. Ha una formazione universitaria in ambito linguistico. La sua L1 è l'italiano, ed è stata esposta al dialetto durante l'infanzia, l'adolescenza e la post-adolescenza. Nel suo repertorio linguistico rientra anche l'inglese a livello avanzato.⁹ Per quanto riguarda lo spagnolo, la sua competenza è avanzata. La prima esposizione è avvenuta a 19 anni in contesto universitario ed è provenuta da insegnanti e prodotti culturali.

Le sue abitudini linguistiche propendono verso un contesto prevalentemente ispanofono. Di fatto, quasi tutte le sue interazioni (con amici, conviventi, partner, colleghi e clienti) avviene totalmente in spagnolo o in parte (con i parenti). L'unico contesto in cui le interazioni sono mantenute in italiano è con i genitori.

Il repertorio di SD

La tabella 1 illustra i dettagli relativi al repertorio di SD utilizzati da Inf_01.

Repertorio	<i>Sí/sí</i> (n = 16); <i>pues</i> (n = 11); <i>claro</i> (n = 8); <i>también</i> (n = 8); <i>entonces</i> (n = 6); <i>pero</i> (n = 6); <i>digamos</i> (n = 4); <i>de hecho</i> (n = 3); <i>porque</i> (n = 3); <i>o sea</i> (n = 3); <i>bueno</i> (n = 3); <i>pero bueno</i> (n = 2); <i>además</i> (2); <i>vale</i> (n = 1); <i>pero claro</i> (n = 1); <i>justo</i> (n = 1); <i>entonces claro</i> (n = 1); <i>pues sí</i> (n = 1); <i>porque también</i> (n = 1); <i>¿ves?</i> (n = 1); <i>al final</i> (n = 1); <i>te digo</i> (n = 1).
Tot SD	85
Tot parole/10 min	1159
% SD	7,3%
RTT	26%

Tabella 1. Descrizione repertorio Inf_01

Nelle parti di conversazione in cui Inf_01 ricopriva il ruolo di emittente, la presenza di SD e forme con significato procedurale nel discorso è abbondante (n = 85) e corrisponde al 7,3% degli elementi linguistici elicitati. L'inventario delle forme è piuttosto vario, come indica il rapporto tra *tokens* e *types* (RTT = 26%) e contiene i SD più frequenti nel corpus di riferimento dei nativi (*sí*, *pues*, *pero*, *bueno*, *también*, *porque*, *o sea*) ad eccezione di *¿no?* Di queste forme, 13 occorrono più di una volta durante l'eloquio. Le forme utilizzate con più frequenza sono *sí* e *pues*. *Sí* corrisponde anche alla forma più frequente impiegata dal corpus generale di L2 e di L1, mentre *pues* costituisce il secondo SD più frequente impiegato nel corpus di nativi di riferimento. Tra le combinazioni di SD, l'informante ricorre a *pero bueno*, *pero claro*, *entonces claro* e *porque también*. Solo

⁹ Tutte le informazioni relative alla competenza linguistica è frutto dell'auto-valutazione da parte degli informanti.

pero bueno viene considerata¹⁰ come “formula conversazionale” e compare anche tra gli usi dei nativi. Le altre combinazioni, invece, non compaiono tra gli usi dei nativi del corpus di riferimento.¹¹

Le strategie discorsive interazionali

La tabella 2 illustra i SD usati da Inf_01 per realizzare funzioni discorsive di tipo interazionale.

Funzione interazionale	SD
Presa di turno	<i>Vale, sí</i>
Accordo	<i>Claro, sí, pues sí, bueno</i>
Intensificazione	<i>¿ves?</i>
Mitigazione	<i>Digamos (que)</i>
Fatica	<i>Sí</i>

Tabella 2. I SD per la funzione interazionale in Inf_01

Per quanto riguarda la presa di turno, Inf_01 la segnala solo in due occasioni attraverso un SD: in un caso, con *vale*, e nell’altro caso, con *sí*. Va segnalato, tuttavia, che in entrambe le situazioni risulta difficile separare la funzione di presa di turno da quella di accordo, in particolare per quanto riguarda *vale* (1):

(1)

02 (.) *y: pue- lo lo que*

‘aver accettato di partecipare e beh quello che’

03 *quieras*

‘vuoi’

04 (.) *contarme como me interesa sobre todo saber por*

‘raccontarmi mi interessa soprattutto sapere’

05 *qué decidiste mudarte a Madri:d cuándo llegaste y:*

‘perché hai deciso di trasferirti a Madrid quando sei arrivata e’

06 *y nada un poquito un poquito sobre tu historia*

‘niente un po’ un po’ della tua storia’

07 *persona:l*

‘personale’

08 *e:h (.) por lo que ataÑe Madrid*

‘per quanto riguarda Madrid’

¹⁰ Dal DPDE, <http://www.dpde.es/#/entry/bueno1>

¹¹ Tuttavia, questo potrebbe essere dovuto alla ristrettezza del sub-corpus considerato. Di fatto, sono presenti nel corpus più ampio (<https://preseca.uah.es/corpus-preseca>).

09 I1 *vale con la con la ciudad (.) pues yo en realidad*
'ok con la città beh io in realtà'

10 *llegué a a España en como en dos mil doce de Erasmus pero*
'sono arrivata in Spagna nel tipo nel duemiladodici in Erasmus però'

11 *no a Madrid fui a [nombre ciudad] entonces como un año*
'non a Madrid sono stata a [nome della città] quindi tipo un anno'

12 *después*
'dopo'

Di fatto, in (1) *vale* (it. ok) esprime sia la funzione di segnalare il fatto che Inf_01 prende il turno e inizia la sua narrazione, ma anche esprime a E di aver capito la richiesta e di accoglierla positivamente.

Altre funzioni esplicitate da Inf_01 quando detiene il turno sono quella di intensificazione, che segnala con l'utilizzo di *¿ves?* (it. vedi) e di mitigazione, che invece è introdotta da *digamos que* (it. diciamo che).

Nei casi in cui Inf_01 ricopre il ruolo di ascoltatrice e non desidera prendere il turno, ricorre alla forma *sí/sì* per svolgere funzioni di tipo fático. In altre parole, non interviene nella conversazione ma segnala esplicitamente di essere attiva nell'ascolto. Di fatto, come si nota nell'estratto (2), il *sí/sì* di Inf_01 viene emesso in sovrapposizione al turno di E.

(2)

258 I1 *mientras que La Latina es como más de domingo de*
'Invece La Latina è tipo più da domenica da'

259 *tardeo de te pones en una terraza como la gente de*
'pomeriggio di metti fuori tipo le persone dai'

260 *los treinta para adelante le gusta más La Latina o:*
'trenta in su gli piace di più La Latina'

261 *(.) ay (.)Lavapiés también*
'ah e anche Lavapiés'

262 E *sí como estar ahí un poquito sentados [tranquilos]*
'sí tipo stare lì un po' seduti'

263 I1 [*Sí*]
'sì'

264 E = *echando el rato simplemente*
'semplicemente passando il tempo'

Per quanto riguarda le funzioni dell'ascoltatore che desidera cambiare il suo ruolo e intervenire nella conversazione, Inf_01 ne realizza di due tipologie: espressione

dell'accordo (o risposta collaborativa) ed espressione di disaccordo (o risposta reattiva). Per quanto riguarda l'accordo, le forme più utilizzate sono *sí/sì*, *claro* e, in un caso, *pues sí*. Anche *bueno* introduce una manifestazione di accordo ma, a differenza degli altri casi, si tratta di un accordo parziale (estratto 3).

(3)

223 *la ciudad e:h de recursos que al final la gente que*
'la città eh le risorse che alla fine la gente che'

224 *viene de fuera luego se queda más días en Madrid*
'viene da fuori poi rimane più giorni a Madrid'

225 *entonces haces [como]*
'quindi fai tipo'

226 E *[gastan muchísimos]*
'spendono tantissimo'

227 I1 = *más turismo claro*
'più turismo certo'

228 E *Incluso los bares y:*
'Anche i bar e'

229 I1 *sí sí*
'sì sì'

230 E *Así de simple sí sí sí*
'Così semplicemente sì sì'

231 I1 ***bueno*** *no es que Madrid tenga problemas en llenar*
'beh non è che Madrid abbia problemi a riempire'

232 *los bares también te digo eh (.) porque siempre*
'i bar anche ti dico eh perché sempre'

233 *está así aquí se sale (.) todo el rato*
'sei così qui si esce tutto il tempo'

Nell'estratto (3), si può osservare innanzitutto l'impiego di *claro*, che esprime l'accordo con l'affermazione di E rispetto al fatto che "le persone che vengono da fuori" spendono molto quando si recano a Madrid. Nel turno di E, poi, viene aggiunta l'osservazione che persino i bar beneficiano del turismo e in questo punto si nota una mossa conversazionale interessante. Infatti, la prima reazione di Inf_01 è di esprimere il suo accordo totale con quanto espresso da E (*sí sí*). Tuttavia, nel turno successivo, Inf_01 dimostra di accettare parzialmente quanto suggerito da E, aggiungendo che in realtà i bar non hanno bisogno

di turisti per essere frequentati. Questo passaggio viene marcato con *bueno* e precede una mitigazione di quanto detto anteriormente.

Inf_02

Al momento della raccolta dati viveva a Madrid da 7 anni, ma aveva già trascorso un periodo risiedendo in un'altra città spagnola. Ha una formazione universitaria in ambito linguistico. La sua L1 è l'italiano e nel suo repertorio linguistico rientrano anche inglese (avanzato), francese (principiante-intermedio) e tedesco (principiante). Per quanto riguarda lo spagnolo, la sua competenza è avanzata. La prima esposizione è avvenuta a 18 anni durante una vacanza ed è provenuta da parlanti nativi. Successivamente, ha proseguito con lo studio dello spagnolo all'Università, anche dopo il trasferimento a Madrid.

Le sue abitudini linguistiche si caratterizzano da un contesto misto di uso di italiano (con genitori e parenti), spagnolo (con amici in Spagna, partner e conviventi) e inglese (con amici in generale, clienti e colleghi).

Il repertorio di SD

La tabella 3 illustra i dettagli relativi al repertorio di SD utilizzati da Inf_02.

Repertorio	<i>¿sabes?</i> (n = 13); <i>bueno</i> (n = 12); <i>pero</i> (n = 9); <i>sí/sì</i> (n = 6); <i>ya</i> (n = 6); <i>entonces</i> (n = 5); <i>también</i> (n = 5); <i>porque</i> (n = 4); <i>por ejemplo</i> (n = 3); <i>o sea</i> (n = 3); <i>pues</i> (n = 3); <i>sí que</i> (n = 3); <i>vale</i> (n = 2); <i>y tal</i> (n = 2); <i>y nada</i> (n = 2); <i>eso</i> (n = 2); <i>y todo eso</i> (n = 1); <i>entonces bueno</i> (n = 1); <i>nada</i> (n = 1); <i>pero bueno</i> (n = 1); <i>claro</i> (n = 1); <i>no digo que</i> (n = 1); <i>o sea en plan</i> (n = 1); <i>tal cual</i> (n = 1).
Tot SD	88
Tot parole/10 min	994
% SD su tot parole	8,8%
RTT	27%

Tabella 3. Descrizione repertorio Inf_02

Anche nel caso di Inf_02, i SD e altre forme con significato procedurale sono abbondanti nel corso della conversazione (n = 88) e corrispondono all'8,8% degli elementi linguistici pronunciati. L'inventario è molto ampio, dato che l'informante ricorre a 24 forme diverse (RTT = 27%), di cui 17 ripetute più di una volta. Tra queste, sono presenti i SD più frequenti nel corpus di riferimento dei nativi (*sí*, *pues*, *pero*, *bueno*, *también*, *porque*, *o sea*) ad eccezione di *¿no?* Nel suo caso, le forme discorsive più frequenti sono *¿sabes?* e *bueno*. La forma *bueno* rientra tra i SD usati con più frequenza anche dai nativi, mentre *¿sabes?* non compare come forma più frequentemente utilizzata né da parte dei nativi, né dal gruppo più ampio di informanti italofofoni. In questo senso, si potrebbe ipotizzare una certa influenza dall'inglese, che l'informante usa abitualmente (Borba & Jaeger 2011).

Per quanto riguarda le combinazioni di SD, compare la formula conversazionale *pero bueno* e anche la costruzione *entonces bueno*. In quest'ultimo caso, la costruzione non viene considerata come formula rutinaria, ma è presente nel corpus dei nativi di riferimento. Il caso di *o sea en plan*, probabilmente, è da considerare come una sequenza casuale di SD (Crible & Degand 2021). Di fatto, non sembra essere utilizzato né da altri informanti, né dai nativi del corpus di riferimento.¹² Le due forme coinvolte svolgono tipicamente funzioni di riformulazione o pianificazione discorsiva (Pons 2016; Rojas 2023). L'uso di *en plan* è particolarmente interessante, dato che sembrerebbe essere associato a un uso tipicamente giovanile e particolarmente diffuso nelle varietà peninsulari dello spagnolo (Rojas 2023).

Le strategie discorsive interazionali

La tabella 4 illustra i SD usati da Inf_02 per realizzare funzioni discorsive di tipo interazionale.

Funzione interazionale	SD
Presa di turno	<i>Vale, ya</i>
Accordo	<i>Sí, eso, claro, ya, tal cual, bueno</i>
Mitigazione	<i>No digo que, bueno</i>
Fatica	<i>Ya</i>
Controllo della ricezione	<i>¿sabes?</i>
Espressione attitudinale	<i>(Ah) vale</i>

Tabella 4. I SD per la funzione interazionale in Inf_02

Le funzioni discorsive espresse da Inf_02 nelle parti della conversazione in cui ricopriva il ruolo di emittente sono varie. Anche nel suo caso la presa di turno viene segnalata due volte: in un caso viene marcata con *vale* e, in un altro, con *ya*. Va precisato che, analogamente al caso analizzato precedentemente, *vale* viene utilizzato sia per segnalare la presa di turno, sia per dimostrare all'interlocutrice di aver compreso la richiesta di raccontare la propria storia personale relativa all'arrivo in Spagna. Un'altra funzione realizzata da Inf_02 quando svolge il ruolo di emittente è la mitigazione. Per marcare questa funzione ricorre a forme come *no digo que, bueno*. Oltre a questo, Inf_02 elicitava la funzione discorsiva di "controllo della ricezione". Questa funzione è volta ad assicurarsi che il messaggio sia ricevuto (e/o compreso) da parte dell'interlocutore, anche se non sempre ci si aspetta una conferma verbalizzata da parte dell'altro partecipante nella conversazione. Nel caso di Inf_02, questa funzione è elicitata attraverso la forma *¿sabes?*¹³ (4) che, tra l'altro, è anche il segnale discorsivo emesso più frequentemente da questa informante.

(4)

59 I2 = *es el estilo de vida que permiti:te soltarte un poco más*
 'è lo stile di vita che ti permette di lasciarti andare un po' di più'

¹² Anche considerando quello più esteso (<https://preseaa.uah.es/corpus-preseaa>).

¹³ Forme alternative "specializzate" per questo tipo di funzione discorsiva sono anche *¿entiendes?*, *¿no?*, *¿verdad?* (Borreguero Zuloaga 2015: 161).

- 60 (.) *yo creo sí eso (0.2) e:h te permite soltarte en todos*
 ‘io credo sí questo eh ti permette di lasciarti andare in tutti’
- 61 *los sentidos ¿sabes? sociales y: también e:h no sé*
 ‘i sensi sai? sociali e anche eh non so’
- 62 *personales también como íntimos ¿sabes? (.) no sé (.) hay*
 ‘personali anche tipo intimi sai? non so ci sono’
- 63 *más ocasiones para hacerlo*
 ‘più occasioni per farlo’

Le situazioni in cui Inf_02 interviene per prendere la parola sono marcate con forme come *sí, eso, claro, ya, tal cual* quando l’informante desidera evidenziare il suo accordo con quanto emesso dall’interlocutrice. Tra queste forme, quella a cui Inf_02 ricorre più frequentemente è *sí/sì*. In un caso, Inf_02 interviene per esprimere l’accordo parziale con quanto emesso dall’interlocutrice. Anche in questo caso, la forma che viene impiegata per segnalare l’accordo parziale è *bueno* (5).

Estratto (5)

- 170 I2 *sí porque bueno hubo como ¿sabes? cosas con la pandemia y*
 ‘sì perché beh ci sono stati tipo sai? cose con pandemia e’
- 171 *ta:l que cerra:ron y luego abrie:ron y luego:(.) bueno*
 ‘cose del genere che hanno chiuso poi hanno aperto e poi vabbè’
- 172 *cambios varios*
 ‘cambiamenti vari’
- 173 E *Un lío sí sí siempre (.)siempre el riesgo m:*
 ‘Un casino sì sì il rischio è sempre mmh’
- 174 I2 **Bueno** *sobre todo durante una pandemia ¿sabes?*
 ‘Beh soprattutto durante una pandemia sai?’

Nei casi in cui Inf_02 ricopre il ruolo di ascoltatrice, ma non vuole intervenire nella conversazione, esplicita due tipologie di funzioni discorsive: fatica e di espressione dell’emotività veicolata da quanto riportato dall’altra persona coinvolta nella conversazione. La funzione fatica è svolta da *ya*, mentre l’espressione emotiva è marcata con la combinazione dell’interiezione *ah* con il segnale discorsivo *vale* e, probabilmente, può essere interpretata come manifestazione di sorpresa rispetto all’elemento inaspettato introdotto da E (6).

(6)

- 67 E *Como que (.) pilla el aspecto más abstracto de la ciudad*
 ‘Tipo che coglie l’aspetto più astratto della città’

- 68 *(.) y: va un poquito más allá de decir vale que hay buen*
 ‘e va un po’ oltre del dire ok che c’è una buona’
- 69 *rollo la gente es muy maja y tal pero eso que has dicho me*
 ‘vibe la gente è bella e tutto però questo che hai detto mi’
- 70 *ha impactado mucho de hecho cuando vivía allí (.) pues yo*
 ‘ha colpito molto di fatto quando vivevo lì beh io’
- 71 *no viví en Madrid sino en [nombre de la ciudad] (.)*
 ‘non ho vissuto a Madrid ma a [nome della città]’
- 72 I2 *¡Ah vale!*
 ‘Ah ok!’

Inf_17

Al momento della raccolta dati viveva a Madrid da più di 30 anni, ma aveva già trascorso un periodo risiedendo in un altro paese ispanofono. Ha una formazione universitaria in ambito linguistico. La sua L1 è l’italiano, ed è stata esposta al dialetto durante l’infanzia, l’adolescenza e la post-adolescenza. Nel suo repertorio linguistico rientrano anche inglese (intermedio) e francese (principiante). Per quanto riguarda lo spagnolo, la sua competenza è avanzata. La prima esposizione è avvenuta a 25 anni durante un soggiorno all’estero tramite amici ispanofoni. Pur dichiarando di conoscere lo spagnolo prima del suo arrivo a Madrid, non ha mai seguito un percorso di apprendimento formale. Le sue abitudini linguistiche propendono verso un contesto prevalentemente italofono. Di fatto, quasi tutte le sue interazioni (parenti, partner, colleghi e clienti) avvengono totalmente in italiano o in parte (con i figli e con gli amici).

Il repertorio di SD

La tabella 5 illustra i dettagli relativi al repertorio di SD utilizzati da Inf_17.

Repertorio	<i>¿no?</i> (n = 15); <i>porque</i> (n = 14); <i>sí/sì</i> (n = 9); <i>entonces</i> (n = 9); <i>también</i> (n = 8); <i>digamos</i> (n = 7); <i>però</i> (n = 6); <i>pero</i> (n = 4); <i>no/no</i> (n = 3); <i>esatto</i> (n = 3); <i>entonces claro</i> (n = 2); <i>claro</i> (n = 2); <i>ma</i> (n = 2); <i>¿sabes?</i> (n = 2); <i>o sea</i> (n = 1); <i>incluso</i> (n = 1); <i>por ejemplo</i> (n = 1); <i>además</i> (n = 1); <i>esattamente</i> (n = 1); <i>ma però</i> (n = 1); <i>ma por esto</i> (n = 1); <i>de verdad</i> (n = 1); <i>esto es</i> (n = 1).
Tot SD	95
Tot parole/10 min	1118
% SD su tot parole	8,4%
RTT	26%

Tabella 5. Descrizione repertorio Inf_17

Inf_17 ricorre abbondantemente a SD e altre forme con valore procedurale (n = 95), che rappresentano l'8,4% degli elementi pronunciati. Anche nel suo caso l'inventario è diversificato, come indica il rapporto tra *tokens* e *types* (RTT = 26%). Tuttavia, la varietà di forme è minore rispetto a quella impiegata dalle altre informanti: su 22 elementi linguistici, solo 13 sono ripetuti più di una volta. Nel suo caso, non sono presenti tutte le forme più frequenti nel corpus di riferimento dei nativi. Di fatto, nell'inventario di Inf_17 – per quanto riguarda la porzione di conversazione analizzata – non compaiono occorrenze di *pues* e *bueno*. Tuttavia, a differenza delle altre due informanti, in questo caso compare la forma *¿no?*. Di fatto, *¿no?* risulta essere anche la forma più ricorrente, seguita da *porque*. Per quanto riguarda le combinazioni di SD, Inf_17 ricorre a *entonces claro*, che non è stata riscontrata nel corpus di nativi di riferimento, ma compare anche nella produzione di Inf_01, e a *ma por esto*. Questo caso è particolarmente interessante, in quanto la costruzione è composta da due elementi adiacenti, di cui uno in italiano. In un altro caso, invece, le forme adiacenti sono entrambe in italiano (*ma però*). Di fatto, nel repertorio di Inf_17 sono presenti molti elementi discorsivi in italiano: *esatto*, *ma*, *esattamente* e *però* – che alterna con la forma spagnola *pero*.¹⁴

Le strategie discorsive interazionali

La tabella 6 illustra i SD usati da Inf_17 per realizzare funzioni discorsive di tipo interazionale.

Funzione interazionale	SD
Accordo	<i>Sí, esatto, claro, esto es, sí no</i>
Disaccordo	<i>Ma però</i>
Mitigazione	<i>Digamos (que)</i>
Controllo/conferma della ricezione	<i>¿sabes?, ¿no?</i>

Tabella 6. I SD per la funzione interazionale in Inf_17

Quando Inf_17 ricopre il ruolo di emittente, le funzioni discorsive che vengono marcate con dei segnali discorsivi sono quella di mitigazione e di controllo/conferma della ricezione del messaggio. Per quanto riguarda la funzione di mitigazione, la forma utilizzata da Inf_17 è *digamos (que)* (7). Tuttavia, va specificato che non tutte le occorrenze di *digamos*, emesse da Inf_17, svolgono una funzione di mitigazione. Infatti, in alcuni casi (n = 3) questa forma viene impiegata nel suo valore di pianificazione discorsiva, ossia, come “riempitivo” per permettere a chi sta parlando di prendere del tempo per proseguire la formulazione del discorso. Quest'uso, come suggerisce Di Nunno (2024) può essere considerato un transfer pragmatico dalla forma italiana equivalente (*diciamo*) che svolge sia funzioni di mitigazione, sia di pianificazione discorsiva.

¹⁴ Nel caso degli italofofoni, è difficile stabilire se *sí* e *no* appartengono al repertorio spagnolo o italiano, dato che le forme coincidono nella forma e nella pronuncia. In questo contributo, non vengono considerati come *tag-switches* (a differenza di *però* al posto di *pero*) dato che non ci sono elementi prosodici, sintattici o pragmatici che permetterebbero di associare le forme all'inventario italiano o a quello spagnolo. Tuttavia, va precisato che sono considerati SD quando appaiono all'inizio di un enunciato o più raramente alla fine come particelle autonome, non integrate in una frase.

(7)

202 I17 *porque se integra después (.) entonces ha vivido*
'perché si integra dopo allora ha vissuto'

203 *sí (.) una una (.) situación política bastante:*
'sì una una situazione politica abbastanza'

204 *eh digamos: que ha penalizado España*
'eh diciamo che ha penalizzato la Spagna'

205 E *sí*
'sì'

Per quanto riguarda la funzione di controllo/conferma della ricezione del messaggio (8), Inf_17 alterna tra le forme *¿sabes?* e *¿no?*. Tuttavia, la forma *¿no?* è decisamente quella di preferenza, dato che non solo occorre più frequentemente (n = 17) rispetto a *¿sabes?* (n = 2), ma è anche in assoluto il segnale discorsivo più utilizzato da questa informante. In generale, questa funzione interazionale è quella più rappresentata nel discorso di Inf_17.

(8)

194 I17 *ha llegado tarde respeto a los demás de: otros*
'è arrivato tardi rispetto al resto di altri'

195 *países europeos ¿no? però los cambios han sido*
'paesi europei no? però i cambiamenti sono stati'

196 *muy muy e:h muy eficientes y rápidos sobre todo*
'molto molto eh molto efficienti e rapidi soprattutto'

Non sono stati registrati casi in cui Inf_17 marca in modo esplicito le funzioni discorsive tipiche del ruolo di ascoltatore che non desidera intervenire.¹⁵ Tuttavia, nei casi in cui Inf_17 desidera intervenire nella conversazione, esplicita strategie discorsive di accordo e disaccordo. Per esprimere accordo, l'informante alterna tra le forme *sí*, *esatto*, *claro*, *esto es*, *sí no*. In questo caso, la forma di preferenza per esprimere l'accordo è *sí/sì* (n = 8), ma è interessante notare la presenza anche della forma italiana *esatto* (n = 3) nell'estratto (9).

(9)

153 E *Claro y: todo el mundo se viene a España porque*
'Certo e tutti vengono in Spagna perché'

¹⁵ È possibile che, ad esempio, queste funzioni vengano esplicitate con forme non considerate in questa analisi oppure con elementi di natura prossemica.

- 154 *piensa que: el español es fácil*
 ‘pensano che lo spagnolo sia facile’
- 155 I17 **Esatto** (.) *y después hubo un cambio social*
 ‘Esatto e poi c’è stato un cambio sociale’
- 156 *también muy importante* (.)
 ‘anche questo molto importante’

Infine, l’unico caso in cui Inf_17 interviene per esprimere un disaccordo viene esplicitato con la formula italiana *ma però*. In questo caso, l’informante smentisce un’inferenza di E riguardo al fatto che i suoi figli volessero intraprendere una carriera formativa e lavorativa in Italia (10).

(10)

- 176 I17 **Ma però** *mis hijos no querían ir a Italia ninguno*
 ‘Ma però i miei figli non volevano andare in Italia nessuno’
- 177 *de nosotros quiere regresar a Italia*
 ‘di noi vuole tornare in Italia’

6. Discussione

Questo contributo si era posto come obiettivo osservare i repertori individuali di italofoeni in contesto di immersione, soffermandosi su eventuali casi di variazione intra e interindividuale nell’uso dei segnali discorsivi in spagnolo L2.

A livello quantitativo (cfr. Appendice I) si notano dei pattern simili. Infatti, tutte e tre le informanti hanno elicitato una quantità simile di occorrenze di SD (n = 85 per Inf_01; n = 88 per Inf_02; n = 95 per Inf_17). Anche considerando il totale di SD prodotti in rapporto al numero di parole pronunciate nei 10 minuti di conversazione analizzati si osservano risultati simili (per Inf_01, i SD corrispondono al 7,3% delle unità linguistiche pronunciate; per Inf_02 corrispondono all’8,8%; per Inf_17 corrispondono all’8,4%). Per quanto riguarda il rapporto tra *tokens* (numero complessivo di SD elicitati) e *types* (forme diverse di SD elicitati), anche in questo caso i riscontri sono analoghi nelle tre produzioni. Infatti, la diversità delle forme utilizzate dalle informanti, rispetto al totale di segnali discorsivi prodotti, è piuttosto simile (26% per Inf_01; 27% per Inf_02; 25,6% per Inf_17).

In generale, i SD più frequenti nelle produzioni delle informanti sono *sí/sì* (n = 31); *pero* (n = 25); *entonces* (n = 21); *porque* (n = 21); *también* (n = 21). Questo riscontro è in parte supportato da studi precedenti (Solís García & León Gomez, 2017) in cui viene registrato l’uso abbondante di *sí/sì* da parte degli informanti italofoeni durante conversazioni in spagnolo L2. Inoltre, questi riscontri si allineano anche con le considerazioni per cui i non nativi tendono ad usare più frequentemente SD formalmente o funzionalmente simili alla L1 (Koch & Thörle, 2019). Di fatto *sí/sì*, *pero* (it. però) e *porque* (it. perché) possono essere considerati formalmente simili alle forme nella L1,

mentre *entonces* (it. quindi, allora) e *también* (it. anche) possono essere considerati (parzialmente) simili dal punto di vista funzionale.

Se a livello quantitativo non si notano asimmetrie nei vari comportamenti linguistici, l'analisi qualitativa delle forme che compaiono nei singoli repertori illustra delle situazioni di divergenza. Di fatto, solo 7 forme sono condivise in tutti e tre i repertori (*claro, entonces, o sea, pero, porque, sí/sì, también*). Oltre a questo, ogni informante presenta delle forme “preferite” che utilizza con più frequenza rispetto al resto dei SD nel repertorio. Per Inf_01 si tratta di *sí/sì* (n = 16) e *pues* (n = 11); per Inf_02 corrispondono a *¿sabes?* (n = 13) e *bueno* (n = 12); per Inf_17 a *¿no?* (n = 15) e *porque* (n = 14). Pertanto, riprendendo il concetto di Hasselgren (1994), queste forme possono essere considerate come *lexical teddy bears*. Tuttavia, come puntualizzano anche Koch & Thörle (2021: 182) rispetto a riscontri simili a quelli evidenziati in questo contributo, solo per alcune forme si può parlare di “usi sicuri”. Di fatto, per le osservazioni di questo contributo, questo vale solo per le forme *sí/sì, ¿sabes?, ¿no?* e *porque*, che sono formalmente e/o funzionalmente simili ai corrispondenti italiani. Questi criteri non si applicano a *bueno* e *pues*, per i quali le equivalenze formali e funzionali in italiano sono molto complesse (per una rassegna contrastiva su *bueno*, cfr. Calvi & Mapelli 2004; Medina Montero 2013; su *pues*, cfr. Sainz 2003; Flores 2004; Esposito 2021).

Un'ultima considerazione riguarda la presenza di casi di *tag-switching* (Poplack 1980) che, per questo contributo, riguardano l'inserzione di segnali discorsivi in italiano (L1) nel discorso in spagnolo (L2). Queste istanze si manifestano solo nel caso di una persona (Inf_17). Interessante è anche notare che si tratta dell'informante con il profilo più tendente all'italofonia, ossia, le cui abitudini linguistiche si svolgono principalmente in italiano pur trovandosi in Spagna da molti anni. Le forme italiane a cui ricorre sono *però, esatto, esattamente, ma, ma però* e la combinazione ibrida *ma + por esto*. Il caso più peculiare consiste nell'alternanza tra *ma, pero* e *però*, che condividono tutti lo stesso valore di contrasto. La sovrapposizione di *pero* e *però* è riscontrata anche nello studio di Pascual Escagedo (2015).

Per quanto riguarda invece l'effetto della variazione individuale nella selezione delle forme per realizzare funzioni discorsive di tipo interazionale (cfr. Appendice II), si nota innanzitutto che non tutte le funzioni sono marcate dalle tre informanti. Ad esempio, Inf_01 non utilizza SD per marcare la funzione di controllo/conferma della ricezione del messaggio, che invece è fortemente esplicitata sia da Inf_02 (con *¿sabes?*) sia da Inf_17 (con *¿sabes?* e *¿no?*).¹⁶ Un altro caso è quello della presa di turno, che non viene mai segnalato da Inf_17 con un segnale discorsivo, mentre invece Inf_01 e Inf_02 lo fanno in due casi utilizzando le forme *vale* (sia Inf_01 sia Inf_02), *sí/sì* (Inf_01), *ya* (Inf_02). Altri casi riguardano la realizzazione dell'intensificazione (solo Inf_01 con *¿ves?*); della funzione fatica (Inf_01 con *sí/sì*; Inf_02 con *ya*); l'espressione attitudinale (solo Inf_02, con *¡ah vale!*); il disaccordo (solo Inf_17 con *ma però*). La spiegazione di questa condizione può essere ricondotta al fatto che, come sottolinea anche Magliacane (2019) commentando risultati analoghi, siano stati i bisogni comunicativi specifici di ciascun parlante ad aver determinato le differenze e le preferenze delle funzioni elicitate. In altre parole, è possibile che l'assenza della marcatura di una certa funzione dipenda non tanto

¹⁶ Anche Pascual Escagedo (2015) nota che la funzione di controllo/conferma della ricezione è marcata poche volte da parte degli apprendenti italofofoni considerati per il suo studio. Tuttavia, nei pochi casi in cui viene marcata, anche nel suo caso avviene principalmente con la forma *¿sabes?*

dall'incapacità del parlante di realizzarla¹⁷, bensì dalla sua necessità o meno di veicolare quel tipo di istruzione. Le uniche due funzioni marcate con dei SD da tutte le informanti sono quella di mitigazione e di manifestazione di accordo (totale e parziale). Per quanto riguarda la mitigazione, si notano sia degli usi condivisi (*digamos, digamos que, no digo que*) e degli altri personali (*pero bueno* per Inf_01; *bueno* per Inf_02). Anche nel caso della manifestazione di accordo si notano degli usi condivisi (per l'accordo totale *claro, sí/sí*; per l'accordo parziale, *bueno*) e altri usi più individuali (*pues sí* per Inf_01; *eso, ya, tal cual* per Inf_02; *esatto, esto es* per Inf_17). A questo proposito, è interessante notare che l'uso di *bueno* con valore di espressione dell'accordo (parziale) si discosta da quanto riscontrato nello studio di Pascual Escagedo (2013), focalizzato proprio sugli usi di questo SD da parte di apprendenti italofofoni. Nel suo studio, infatti, gli italofofoni usano *bueno* principalmente come strategia per prendere il turno, mentre i nativi ispanofoni lo usavano come strategia collaborativa per esprimere accordo con l'interlocutore. Nel caso di questo contributo, le funzioni di *bueno* si allineano piuttosto con quest'ultimo utilizzo.

Infine, l'ultima osservazione riguarda la polifunzionalità. Di fatto, nelle produzioni di Inf_01 e di Inf_02 si nota l'uso polifunzionale di un elemento (*sí/sí* per Inf_01; *ya* per Inf_02) per esprimere più funzioni (presa di turno, fatica, accordo). Invece, Inf_17 ha utilizzato forme diverse in base alla funzione discorsiva da realizzare. Questa tendenza, probabilmente, va interpretata tenendo in considerazione l'elevata complessità dello sviluppo della polifunzionalità dei SD, che vale tanto per la L1 (Salameh 2022) quanto per la L2 (cfr. Andorno 2007; Ivanova & Bello Viruega 2019 e, più in generale, Ellis & Wulff 2015).

7. Conclusioni

Questo contributo ha cercato di esplorare la manifestazione della variazione intra e interindividuale nell'uso dei segnali discorsivi in L2. In particolare, ha voluto osservare eventuali simmetrie o asimmetrie nella conformazione degli inventari di segnali discorsivi a disposizione dei non nativi e identificare eventuali strategie discorsive comuni. Per questo, il lavoro è stato impostato seguendo un approccio qualitativo e si è focalizzato sui singoli profili individuali. Dall'analisi è emerso che le tre parlanti selezionate per questo studio non mostrano particolari divergenze per quanto riguarda: (i) le occorrenze di SD nella conversazione; (ii) il rapporto tra SD e parole elicitate nella conversazione; (iii) la diversità di forme utilizzate. Tuttavia, un'ampia varietà individuale emerge osservando le forme "preferite" dalle parlanti e le creazioni individuali, in particolare per quanto riguarda le combinazioni di SD e le forme di SD adiacenti. In questo senso, la variazione si mostra sul piano intra e interindividuale.

Per quanto riguarda invece le strategie discorsive per trasmettere funzioni di tipo interazionale, le informanti selezionate per questo studio mostrano sia pattern convergenti, manifestati dalla scelta delle stesse forme per realizzare una determinata funzione, sia divergenti. In quest'ultimo caso si nota anche un'asimmetria delle funzioni interazionali realizzate dovuta, probabilmente, alle diverse necessità comunicative sorte nel corso delle varie conversazioni. In conclusione, si ritiene che lo studio rappresenti un

¹⁷ Su questo va ricordato che probabilmente la funzione è stata realizzata ma non lessicalizzata. Ad esempio, può essere stata realizzata con interiezioni o con elementi prossemici non considerati in questa fase di analisi.

valido esempio per cogliere al meglio il forte carattere idiosincratico che caratterizza i segnali discorsivi che, spesso, rappresenta un elemento di complessità nell'interpretazione generale dei comportamenti linguistici per gli studi che si interessano dei processi acquisizionali di questi elementi.

Naturalmente, lo studio non è privo di limiti. Infatti, va segnalata soprattutto la numerosità degli esempi considerati: trattandosi di uno studio esplorativo che puntava sull'analisi approfondita delle produzioni individuali, non è stato possibile includere un numero maggiore di interviste. Sviluppi futuri della ricerca, pertanto, prevedono l'estensione dell'analisi a un maggior numero di produzioni. Inoltre, potrebbe senz'altro essere interessante non limitarsi all'analisi della funzione interazionale, bensì includere anche quella relativa alle funzioni metadiscorsiva e cognitiva. In questo senso, potrebbe risultare rilevante osservare tutte le strategie discorsive, oltre all'impiego dei segnali discorsivi, attivate per realizzare tali funzioni.

Appendice I

Repertorio SD	Inf_01	Inf_02	Inf_17
además	2	-	1
al final	1	-	-
bueno	3	12	-
claro	8	1	2
de hecho	3	-	-
de verdad	-	-	1
digamos	4	-	7
entonces	7	5	9
entonces bueno	-	1	2
entonces claro	1	-	-
eso	-	2	-
esattamente	-	-	1
esatto	-	-	3
esto es	-	-	1
incluso	-	-	1
justo	1	-	-
ma	-	-	2
ma però	-	-	1
ma por esto	-	-	1
nada	-	1	-
no	-	-	3
¿no?	-	-	15
no digo que	-	1	-
o sea	3	3	1
o sea en plan	-	1	-
pero	6	9	10

Repertorio SD	Inf_01	Inf_02	Inf_17
pero bueno	2	1	-
pero claro	1	-	-
por ejemplo	-	3	1
porque	3	4	14
porque también	1	-	-
pues	11	3	-
pues sí	1	-	-
¿sabes?	-	13	2
sí	16	6	9
sí que	-	3	-
tal cual	-	1	-
también	8	5	8
te digo	1	-	-
vale	1	2	-
¿ves?	1	-	-
y nada	-	2	-
y tal	-	2	-
y todo eso	-	1	-
ya	-	6	-

Appendice II

Funzione	Inf_01	Inf_02	Inf_17
Mitigazione	Digamos que, pero bueno	No digo que, bueno	Digamos
Controllo/conferma della ricezione	-	¿sabes?	¿sabes?, ¿no?
Presa di turno	Vale, sí	Vale, ya	-
Intensificazione	¿ves?	-	-
Fatica	Sí	Ya	-
Espressione attitudinale	-	¡Ah vale!	-
Accordo	Claro, sí, pues sí, bueno	Sí, eso, claro, ya, tal cual, bueno	Sí, esatto, claro, esto es, sí no
Disaccordo	-		Ma però

Bibliografía

- Aijmer, Karin. 2013. *Understanding pragmatic markers: a variational pragmatic approach*. Edinburgh University Press.
- Ainciburu, María Cecilia. 2008. *Aspectos del aprendizaje del vocabulario. Tipo de palabra, método, contexto y grado de competencia en las lenguas afines*. Frankfurt am Main: Peter Lang.
- Álvarez-Ejzenberg, Fabiana. 2021. Análisis contrastivo de las partículas pues/pois/puis: empleos discursivos de la forma pues en el español como L2. In Ballester De Celis, Carmen & Sampedro Mella, Maria (a cura di), *Aportes del CAES a la enseñanza del español como lengua extranjera: Estudios aplicados de gramática y pragmática*, 235-256. Universidad de Santiago de Compostela: Servicio de Publicaciones e Intercambio Científico.
- Anderson, A.E. John & Mak, Lorinda & Keyvani Chahi, Aram & Bialystok, Ellen. 2018. The Language and Social Background Questionnaire - Assessing Degree of Bilingualism in a Diverse Population (LSBQ). *Behav Res* 50. 250–263.
- Andorno, Cecilia. 2007. Apprendere il lessico: elaborazione di segnali discorsivi (sì, no, così). In Chini, Marina & Desideri, Paola & Favilla, Maria Elena & Pallotti, Gabriele (a cura di.), *Imparare una lingua: recenti sviluppi teorici e proposte applicative, Atti del VI Congresso di Studi Aitla*, 95-122. Perugia: Guerra.
- Bailini, Sonia. 2016. *La interlengua de lenguas afines. El español de los italianos, el italiano de los españoles*. Milano: LED.
- Bazzanella, Carla. 2015. Segnali discorsivi a confronto. Dati e teoria, un percorso integrato. In Borreguero Zuloaga, Margarita & Gómez-Jordana Ferary, Sonia (a cura di), *Les Marqueurs du discours dans les langues romanes: une approche contrastive*, 35-46. Lambert-Lacas.
- Bazzanella, Carla. 1995. I segnali discorsivi. In Renzi, Lorenzo & Salvi, Giampaolo & Cardinaletti, Anna (a cura di), *Grande grammatica italiana di consultazione vol. III*, 225-257. Bologna: Il Mulino.
- Bazzanella, Carla. 2001. Segnali discorsivi e contesto. In Heinrich, Wilma & Heiss, Christine (a cura di), *Modalità e Substandard*, 41-64: Bologna: CLUEB.
- Bazzanella, Carla. 2006. Discourse markers in Italian: towards a “compositional” meaning. In Fisher, Kerstin (a cura di), *Approaches to discourse markers*, 449–464. Amsterdam: Elsevier.
- Bello Viruega, Iria & Bernales, Carolina & Calvi, Maria Vittoria & Landone, Elena. 2019. *Cognitive Insights into Discourse Markers and Second Language Acquisition*. Peter Lang.
- Bini, Milena & Pernas, Almudena. 2008. Marcadores discursivos en los primeros estadios de adquisición del italiano L2. In Monroy Casas, Rafael & Sánchez Pérez, Aquilino (a cura di), *25 años de lingüística aplicada en España: hitos y retos*, 25-36. Universidad de Murcia.
- Borba, Rodrigo & Jaeger, Aline. 2011. They never realized that, you know: linguistic collocations and interactional functions of *you know* in contemporary academic spoken English. *ESpecialist* 32(2). 195-215.
- Borreguero Zuloaga, Margarita. 2017. Topic-shift markers in L2 Italian: Paths of acquisition by Spanish speakers. *Language, Interaction, Acquisition* 8 (2). 173-203.
- Borreguero Zuloaga, Margarita. 2019. Expressing agreement in L2 Italian Strategies and discourse markers in Spanish learners. In Bello Viruega, Iria & Bernales, Carolina &

- Calvi, Maria Vittoria & Landone, Elena (a cura di), *Cognitive Insights into Discourse Markers and Second Language Acquisition*, 124-143. Bern: Peter Lang.
- Borreguero Zuloaga, Margarita. 2015. A vueltas con los marcadores del discurso: de nuevo sobre su definición y sus funciones. In Ferrari, Angela & Lala, Letizia, *Testualità. Fondamenti, unità, relazioni*, 151-170. Firenze: Franco Cesati.
- Bülow, Lars & Pfenninger, E. Simone. 2021. Introduction: Reconciling approaches to intra- individual variation in psycholinguistics and variationist sociolinguistics. *Linguistics Vanguard* 7(2). 1-10.
- Calvi, Maria Vittoria & Mapelli, Giovanna. 2004. Los marcadores bueno, pues, en fin, en los diccionarios de español e italiano. *Artifara* 4.
- Crible, Ludivine & Degand, Liesbeth. 2021. Co-occurrence and ordering of discourse markers in sequences: a multifactorial study in spoken French. *Journal of Pragmatics* 177. 18–28.
- De Cristofaro, Elisa & Badan, Linda. 2021. The Acquisition of Italian Discourse Markers as a Function of Studying Abroad. *Corpus Pragmatics* 5(1). 95–120.
- De Marco, Anna. 2023. L'insegnamento esplicito nello sviluppo della competenza pragmatica in L2/LS. Il caso dei segnali discorsivi. In Balboni, Paolo & Caon, Fabio & Menegale, Marcella & Serragiotto, Graziano (a cura di), *La linguistica educativa tra ricerca e sperimentazione. Scritti in onore di Carmel Mary Coonan*, 83-92. Edizioni Ca' Foscari.
- Della Putta, Paolo. 2015. Discouraging constructional negative transfer: Theoretical aspects and classroom activities for Spanish-speaking students of L2 Italian. In Masuda, Kyoko (a cura di), *Cognitive Linguistics and Sociocultural Theory: Applications for Second and Foreign Language Teaching*, 25-49. De Gruyter.
- Di Nunno, Valentina. 2024. La enseñanza de los marcadores del discurso en la clase de ELE: análisis de los mecanismos de transferencia pragmalingüística en el aprendizaje de lenguas afines. In Fuentes Rodríguez, Catalina & Brenes Peña, Ester (a cura di), *Macrosintaxis y pragmática del discurso persuasivo*, 323-346. Pamplona: EUNSA ediciones, Universidad de Navarra.
- Diskin, Chloé & Regan, Vera. 2015. Migratory Experience and Second Language Acquisition Among Polish and Chinese Migrants in Dublin, Ireland. In Forsberg Lundell, Fanny & Bartning, Inge (a cura di), *Cultural Migrants and Optimal Language Acquisition*, 137–177. Multilingual Matters.
- Dörnyei, Zoltán. 2003. *Questionnaires in Second Language Research – Construction, Administration and Processing*. Lawrence Erlbaum Associates Inc.
- Ellis, C. Nick & Wulff, Stephanie. 2015. Usage-based approaches to SLA. In VanPatten, Bill & Williams, Jessica (a cura di), *Second language acquisition research series: Theories in second language acquisition*, 75-94. New York: Routledge.
- Ellis, C. Nick & Wulff, Stephanie. 2020. Usage-Based Approaches to L2 Acquisition. In VanPatten, Bill & Keating, D. Gregory & Wulff, Stephanie (a cura di), *Theories in Second Language Acquisition*, 63–82. New York: Routledge.
- Esposito, Giorgia. 2021. Partículas discursivas y contrastividad: un estudio sobre pues. *Orillas* 10. 347-377.
- Fischer, Kirsten. 2006. *Approaches to discourse particles*. Amsterdam: Elsevier.
- Flores, María José. 2004. Los marcadores del discurso II: a propósito de los comentadores pues y pues bien. Aproximación a un análisis contrastive español-italiano. *Cultura Latinoamericana Annali* (6), 137-168.

- Fuentes-Rodríguez, Catalina & Brenes Peña, Ester. 2023. *La enseñanza-aprendizaje de conectores y operadores discursivos en ELE*. Arco Libros- La muralla.
- Ghezzi, Chiara & Molinelli, Piera. 2014. *Discourse and Pragmatic markers from Latin to the Romance Languages*. Oxford: Oxford University Press.
- Gillani, Eugenio & Pernas, Paloma. 2021. Didattica della competenza conversazionale in italiano L2/LS: lo strumento A.Ma.Dis, una risorsa basata sulla ricerca acquisizionale. In Borreguero Zuloaga, Margarita (a cura di), *Acquisizione e didattica dell'italiano: riflessioni linguistiche, nuovi apprendenti e uno sguardo al passato*, 615-630. Berlin: Peter Lang.
- Gilquin, Gaëtanelle. 2016. Discourse markers in L2 English - From classroom to naturalistic input. In Timofeeva, Olga & Gardner, Anne-Christine & Honkapohja, Alpo & Chevalier, Sarah (a cura di), *New Approaches to English Linguistics: Building Bridges*, 213-249. Studies in Language Companion Series, Amsterdam: John Benjamins.
- Götz, Sandra & Mukherjee, Joybrato. 2018. Investigating the Effect of the Study Abroad Variable on Learner Output: A Pseudo-Longitudinal Study on Spoken German Learner English. In Brezina, Vaclav & Flowerdew, Lynne (a cura di), *Learner Corpus Research*, 47-65. Londra: Bloomsbury.
- Hasselgren, Anita. 1994. Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4(2), 237-260.
- Hasselgren, Anita. 2002. Learner corpora and language testing: smallwords as markers of learner fluency. In Granger, Sylviane & Hung, Joseph & Petch-Tyson, Stephanie (a cura di), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, 143-173. Philadelphia: John Benjamins.
- Hellermann, John & Vergun, Andrea. 2007. Language which is not taught: the discourse marker use of beginning adult learners of English. *Journal of Pragmatics*, 39. 157-179.
- Ivanova, Olga & Bello Viruega, Iria. 2019. Pragmatic processing in second language: what can focus operators tell us about cognitive performance in L2?. In Bello Viruega, Iria & Bernales, Carolina & Calvi, Maria Vittoria & Landone, Elena, *Cognitive Insights into Discourse Markers and Second Language Acquisition*, 45-60, Bern: Peter Lang.
- Jefferson, Gail. 2004. Glossary of Transcript Symbols with an Introduction. In Lerner, Gene H. (a cura di), *Conversation Analysis: Studies from the first generation. Pragmatics & Beyond*, 13-31. New Series, John Benjamins Publishing Company.
- Kerr-Barnes, Betsy. 1998. The acquisition of connectors in French L2 narrative discourse. *French Language Studies* 8. 189-208.
- Koch, Christian & Thörle, Britta. 2019. The discourse markers sí, claro and vale in Spanish as a Foreign Language. In Bello Viruega, Iria & Bernales, Carolina & Calvi, Maria Vittoria & Landone, Elena, *Cognitive Insights into Discourse Markers and Second Language Acquisition*, 76-95. Bern: Peter Lang.
- Koch, Christian & Thörle, Britta. 2021. Metadiscursive Activities in Oral Discourse Production in L2 French: A Study on Learner Profiles. *Corpus Pragmatics* 5. 153-186.
- Li, Ping & Zhang, Fan & Tsai, Erlfang & Puls, Brendan. 2014. Language History Questionnaire (LHQ 2.0): A New Dynamic Web-Based Research Tool. *Bilingualism: Language and Cognition* 17(3). 673-680.
- Liu, Binmei. 2016. Effect of L2 exposure: From a perspective of discourse markers. *Applied Linguistics Review* 7(1). 73-98.

- Magliacane, Annarita. 2019. L'utilizzo di marcatori discorsivi nella lingua seconda: Uno studio longitudinale sull'utilizzo di 'you know' da parte di studenti Erasmus. *Studi AiTla*. 215-230.
- Martín Zorraquino, María Antonia & Portolés Lázaro, José. 1999. Los marcadores del discurso. *Gramática descriptiva de la lengua española*, 4051-4213. Madrid: Espasa.
- Medina Montero, Jose Francisco. 2013. El marcador del discurso "bueno": análisis y propuestas de traducción hacia el italiano. *Studi Italiani di Linguistica Teorica e Applicata* 42 (2). 381-400.
- Murillo Ornat, Silvia. 2010. Los marcadores del discurso y su semántica. In Loureda Lamas, Óscar & Acín Villa, Esperanza (a cura di), *Los estudios sobre marcadores del discurso en español, hoy*, 241-270. Madrid: Arco Libros.
- Nogueira da Silva, Antonio Messias. 2010. La enseñanza de los marcadores del discurso del español en relación con los géneros y secuencias textuales. *Revista Nebrija de Lingüística Aplicada* 5(9). 59-85.
- Norén, Kerstin & Linell, Per. 2007. Meaning potentials and the interaction between lexis and contexts: an empirical substantiation. *International Pragmatics Association, Pragmatics* 17 (3). 387-416.
- Pascual Escagedo, Consuelo. 2013. Análisis del marcador discursivo bueno en las conversaciones de estudiantes italianos y españoles hablando en español. In Ainciburu María Cecilia (a cura di), *Actas del I Congreso Internacional Nebrija en Lingüística Aplicada a la Enseñanza de Lenguas, Madrid, 28-30 settembre 2012*, 196-223. Universidad Antonio de Nebrija.
- Pascual Escagedo, Consuelo. 2015. Análisis de las funciones de los marcadores discursivos en las conversaciones de estudiantes italianos de ELE. *Lingue e linguaggi*, 13. 119-161.
- Pons Bordería, Salvador. 1998. *Conexión y conectores: estudio de su relación en el registro informal de la lengua*. Anejo XXVII de la Revista Cuadernos de filología.
- Pons Bordería, Salvador. 2006. A functional approach to the study of discourse markers. In Kerstin Fischer, *Approaches to discourse particles*, 77-100. Elsevier.
- Pons Bordería, Salvador. 2016. Evolución diacrónica de o sea. In *Boletín de la Real Academia Española (RAE)*, Tomo XCVI, quaderno CCCXIII, 291-350.
- Poplack, Shana. 1980. Sometimes I'll start a sentence in English and termino en Español; toward a typology of code-switching. *Linguistics* 18(7/8). 581-618.
- Rojas, Cristian. 2023. Marcadores del discurso e indexicalidad. Una lectura del comportamiento pragmático de *en plan*, *onda* y *tipo* desde la sociolingüística de la identidad. *SOPRAG* 11(1). 23-41.
- Sainz, Eugenia. 2003. Marcador discursivo e interferencia en la enseñanza del ELE. *Rassegna Iberistica* 78. 91-96.
- Salameh Jiménez, Shima. 2022. Measuring the polyfunctionality of discourse markers experimentally: Eye-tracking and visual attention as cognitive-processing indicators. Peninsular Spanish o sea. *Studii de lingvistică* 12 (2). 143-178.
- Sansò, Andrea. 2020. *I segnali discorsivi*. Roma: Carocci.
- Solís García, Inmaculada & León Gómez, Magdalena. 2017. La adquisición de las respuestas afirmativas en ELE por parte de aprendices italianos. *Marco ELE* 24. 1-27.
- Thomas, Anita & Mitchell, Rosamond. 2022. Can variation in input explain variation in typical spoken targetlanguage features during study abroad?. *Journal of the European Second Language Association* 6 (1). 60-77.

Il corpus IT_SEAH per la didattica della comunicazione accademico-disciplinare nel settore dell'Architettura e delle Costruzioni

Mariapia D'Angelo

Università di Chieti-Pescara

mariapia.dangelo@unich.it

Abstract

Al fine di rispondere ai bisogni specifici degli studenti internazionali nell'ambito del progetto SEAH (*Sharing European Architectural Heritage*)¹ è stata sviluppata una metodologia per la creazione di materiali didattici innovativi indirizzati all'insegnamento della comunicazione accademico-disciplinare nel macrosettore dell'Architettura e delle Costruzioni, fruibili online e in modalità di autoapprendimento. Tali risorse si basano su corpora specialistici compilati ad hoc, in ciascuna delle cinque lingue del progetto: francese, italiano, russo, spagnolo e tedesco. Alla presentazione delle premesse teorico-metodologiche di SEAH seguiranno qui esemplificazioni dell'uso diretto e indiretto del corpus IT_SEAH per la didattica di alcuni fenomeni lessico-grammaticali del linguaggio dell'Architettura e delle Costruzioni in lingua italiana.

1. Le risorse del progetto SEAH: premesse teorico-metodologiche

Nel contesto italiano la comunicazione accademica è stata dapprima esplorata in riferimento alle produzioni dei nativi² e in seguito a quella degli studenti internazionali, denominazione questa che raggruppa sia i corsisti stranieri, iscritti per un lungo o breve

¹ Il progetto *Sharing European Architectural Heritage: Innovative language teaching tools for academic and professional mobility in Architecture and Construction* (SEAH) è stato cofinanziato nel quadro del programma Erasmus+ K203 (2020-2022) ed è stato coordinato dalla scrivente in collaborazione con M. C. Ferro e S. Piccioni dell'Università di Chieti-Pescara, ateneo capofila. Le risorse linguistiche sviluppate sono liberamente accessibili tramite una piattaforma di risorse educative aperte sul sito di SEAH (<https://www.seahproject.eu/>), con il duplice intento di soddisfare le necessità formative degli studenti in mobilità e delle categorie professionali coinvolte con committenti stranieri, pubblici e privati, nel campo dell'Architettura e delle Costruzioni.

² Si vedano il lavoro monografico di Lavinio (2004), unitamente ai lavori collettanei a cura di Sobrero & Lavinio (1991), di Ciliberti & Anderson (1999) e il recente volume a cura di Grandi (2025), che prende in esame anche le produzioni testuali di studenti stranieri iscritti a percorsi di studio in Italia.

periodo negli atenei del nostro Paese, sia quelli che studiano la nostra lingua presso le istituzioni universitarie estere (De Mauro *et al.* 2002). In ottica interdisciplinare, i testi orali e scritti prodotti da docenti, esperti e studenti sono stati indagati in ordine ai tratti caratteristici del livello lessicale e fraseologico (Spina 2010; Gallina 2018), delle tipologie testuali, dei registri e dell'organizzazione retorica (Desideri 2011; Fiorentino 2015; Balboni 2018). Si è riflettuto, inoltre, sull'opportunità di compilare i sillabi con una maggiore flessibilità rispetto ai descrittori del QCER (Lo Duca 2006; Bonvino & Rastelli 2011; Mezzadri 2016) e di abolire la distinzione tra i corsi di italiano L2 per scopi comunicativi e quelli per scopi accademici (Vedovelli 2010; Gallina 2019) in funzione delle reali esigenze degli studenti universitari alloglotti, i quali giungono nelle sedi ospitanti con una competenza della lingua italiana spesso inferiore all'auspicato livello B2 e devono subito confrontarsi con testi specialistici del proprio settore di studio.

In linea con le ricerche da tempo condotte nel versante dell'*English for Specific Academic Purposes*, anche la riflessione glottodidattica italiana ha rilevato la necessità di un "potenziamento dell'input disciplinare in lingua" (Lubello 2019: 186) mediante attività basate sugli aspetti linguistico-retorici delle singole pratiche discorsive disciplinari, spesso non esplicitate, ma determinate dal contesto epistemico di riferimento (D'Angelo 2021). È stato infatti dimostrato che il significato di alcune unità lessicali, la loro frequenza d'uso e i pattern collocazionali possono variare a seconda del settore scientifico (Hyland & Tse 2004; Durrant 2014; Rees 2021); altrettante variazioni si sono riscontrate nell'uso delle citazioni e del discorso indiretto, nonché nella selezione delle strategie retorico-pragmatiche e dei generi testuali (Hyland 2004; Nesi & Gardner 2012).

Le suddette indagini – inizialmente applicate alla lingua inglese e poi ad altre lingue di cultura europee – sono state supportate dagli strumenti di analisi della *corpus linguistics*, in modo particolare dall'interrogazione di corpora specialistici, ovvero raccolte di testi in formato elettronico ascrivibili ad un determinato registro, ambito di discorso (es. medico, economico, ecc.) o settore disciplinare (Biber, Conrad & Reppen 1998; de Beaugrande 2001). La consultazione di tali raccolte riveste un ruolo sempre più significativo nella didattica della comunicazione accademica, poiché consente l'identificazione dei tratti costitutivi dei linguaggi in oggetto in relazione ai testi orali (Thompson & Nesi 2001) e scritti (Römer & Cortes & Friginal 2020; Flowerdew 2022; Thompson 2022), permettendo inoltre analisi contrastive con le varietà della stessa lingua (cf. Flinz & Perkuhn 2018) e con le diverse lingue che compongono il repertorio degli studenti universitari (Salzmann 2021).

Sulla scorta di queste ricerche il progetto SEAH ha previsto, in una prima fase, la compilazione di cinque corpora che includono testi della comunicazione accademico-disciplinare e del linguaggio professionale nel macrosettore dell'Architettura e delle Costruzioni (d'ora in avanti AC) in lingua francese, italiana, russa, spagnola e tedesca (Piccioni *et al.* 2021). Nella fase successiva, tali risorse sono state impiegate nell'elaborazione dei percorsi glottodidattici, ora usufruibili tramite una Open Educational Platform appositamente sviluppata e liberamente accessibile dal sito del progetto (D'Angelo *et al.* 2022).

La raccolta dei testi confluiti nei corpora SEAH è stata preceduta da una *needs analysis* (Long 2005) condotta in consultazione con i docenti dei Dipartimenti di AC del partenariato e con i coordinatori Erasmus delle istituzioni educative coinvolte (Káňa & Hradílková 2022), al fine di identificare le principali difficoltà incontrate dagli

studenti in mobilità e gli insegnamenti maggiormente frequentati durante il soggiorno all'estero. Le informazioni rinvenute hanno determinato la selezione degli ambiti tematici e dei generi testuali da inserire nei corpora, per la cui compilazione sono stati utilizzati molti documenti messi a disposizione dai docenti stessi (es. sillabi, dispense e lezioni registrate), nonché testi reperiti direttamente nel web effettuando ricerche per parole chiave (tesi di dottorato o articoli scientifici), oltre che testi afferenti al dominio professionale individuati anche grazie alle indicazioni degli *associate partners* (report di progetto, materiale pubblicitario, ecc.).

Nella redazione dei materiali didattici, i corpora SEAH sono stati utilizzati in modo “indiretto” (Leech 1997; Römer 2008), vale a dire per la selezione delle caratteristiche lessico-grammaticali e retorico-pragmatiche da includere nelle attività, come pure per il reperimento dei testi e delle frasi impiegate nei vari esercizi che si basano, dunque, su produzioni linguistiche autentiche e attinenti all'ambito disciplinare specifico, nell'intento di offrire attività didattiche percepite dai discenti come rilevanti per il proprio percorso accademico (Hyland 2016).

Inoltre, secondo l'approccio del *data-driven learning* (Johns 1990) le attività sviluppate prevedono la consultazione “diretta” del corpus da parte degli studenti per la risoluzione di alcuni esercizi, tramite le funzionalità del software gratuito NoSketch Engine (Rychlý 2007). Numerose teorie sull'acquisizione linguistica supportano infatti l'impiego dei corpora nella classe di lingua³, da quelle di matrice costruttivista e socioculturale (Flowerdew 2015), alla teoria *usage-based* di Tomasello (2003; cfr. anche Bybee 2006). Nel quadro di quest'ultima, l'acquisizione di una lingua avverrebbe “attraverso l'intervento determinante di meccanismi cognitivi, capaci di astrarre le regolarità delle coppie forma-significato presenti nell'input” (Spina *et. al.* 2020: 78) che si attiverebbero nel risolvere un compito mediante l'osservazione diretta dei dati linguistici forniti dal corpus. Focalizzare l'attenzione dei discenti su determinati fenomeni linguistici concorrerebbe inoltre a migliorare l'acquisizione degli input, secondo la nota ipotesi del *noticing* elaborata da Schmidt (2001).

L'innovativa integrazione di usi diretti e indiretti dei corpora SEAH – non riscontrabile nelle risorse glottodidattiche per il settore dell'AC ad oggi disponibili – mira al conseguimento di molteplici obiettivi. Tra questi figurano l'esplorazione attiva dei dati linguistici per la scoperta autonoma dei significati (Aston 2001), delle regole e degli schemi ricorrenti (Richards & Rodgers 2014), unitamente allo sviluppo delle competenze necessarie per un uso indipendente del corpus per l'osservazione autonoma di fenomeni linguistici di interesse e la risoluzione di dubbi specifici. E ancora, i materiali perseguono l'obiettivo di potenziare la comprensione dei testi specialistici e la trasmissione di contenuti tecnici connessi ai saperi scientifici, nonché alle pratiche comunicative dell'AC nei vari contesti discorsivi, geografici e culturali.

Le unità realizzate si prestano, infine, all'elaborazione di attività per il potenziamento della competenza metalinguistica in chiave contrastiva nelle cinque lingue del progetto e sono utilizzabili nei corsi di lingua e traduzione o con i docenti in formazione⁴.

Nei paragrafi a seguire, dopo una breve introduzione sui tratti salienti del linguaggio italiano dell'AC, verranno illustrati alcuni usi del corpus IT_SEAH (§ 2.1) nei materiali didattici già sviluppati (§ 2.2) e ulteriori applicazioni di questa risorsa per la didattica

³ Un'utile e puntuale rassegna in Forti (2023: 21-25).

⁴ cfr. D'Angelo & Flinz (2025) per applicazioni nella didattica della diatesi passiva in tedesco e italiano.

dei derivati, dei marchionimi e degli acronimi (§ 2.3), come pure delle collocazioni e delle polirematiche del settore (§ 2.4).

2. Le risorse SEAH per l'italiano dell'AC

Il linguaggio dell'architettura è stato oggetto di numerose indagini condotte da diverse prospettive di studio, quali quella storica, sociolinguistica e semiologica⁵. Per ragioni di spazio, ci limitiamo a ricordare che è stato uno dei primi linguaggi settoriali ad affermarsi durante il Rinascimento (Nencioni 1995 [2000]: 24; Biffi 2003), “così maturo alla fine del Cinquecento da estendersi anche in molte lingue d'Europa” (Biffi 2010: 106). A partire dal XVI secolo iniziano a diffondersi in tutto il continente non solo i tecnicismi degli artigiani, ma anche i manuali redatti dagli architetti italiani (Motolese 2012; Bruni 2013), quali la *Regola delli cinque ordini d'architettura* di Barozzi da Vignola tradotta in ben nove lingue, oppure il trattato di Vincenzo Scamozzi intitolato *L'idea della Architettura Universale* divenuto un testo fondamentale nelle scienze delle costruzioni.

Il lessico dell'AC si presenta come un sistema notevolmente composito e intessuto sin dalle origini di varianti diatopiche, diamesiche e diastratiche. In relazione a quest'ultima dimensione, come individuato nel noto lavoro di Nencioni (1995 [2000]), la specificità della lingua architettonica risiede nel suo doppio registro. Da un lato, essa trae origine dalla trattatistica latina (fondamentale il *De architectura* di Vitruvio) e dalle relative volgarizzazioni che hanno contribuito a stabilire una base concettuale condivisa. Dall'altro lato, deriva da un sapere custodito e trasmesso oralmente all'interno delle botteghe artistiche e artigiane, essendo l'architettura una *ars mechanica*. Tale aspetto determina la marcata variabilità diatopica, peculiare di tutte le lingue tecniche dei mestieri, poiché corrispondente alla diversità geografica delle sue origini, attestata dai molteplici geosinonimi rinvenuti (Biffi 2019). A ciò si affianca la duplice varietà diamesica di un sapere tramandato sia oralmente dalle maestranze, sia per iscritto nei trattati di architettura. Questa duplice natura si manifesta anche nella lingua architettonica contemporanea (Bellucci 1997; Bellucci & Biffi 2012) che si configura sia come lingua di cultura, ricca e mutevole per dinamiche interne, sia come strumento di comunicazione essenziale per le relazioni estese, profondamente inserita nei processi di innovazione culturale, connessa alle trasformazioni ambientali, politiche, tecnologiche ed economiche della società. Ne costituiscono una traccia i forestierismi di diversa origine attestati nel lessico dell'AC, fino a qualche decennio fa essenzialmente di provenienza greca o latina e con la presenza di “pochissimi” anglicismi (Bellucci &

⁵ Il tema sulle origini del lessico architettonico, già introdotto nei lavori di Migliorini (1960) e Marazzini (1993: 78; 1994: 252), diviene oggetto di indagine soprattutto a partire dallo studio fondante di Nencioni 1995 [2000], seguito da una nutrita serie di ricerche (cfr. almeno Biffi 1999, 2006), anche in ordine alla diffusione del lessico architettonico di origine italiana nelle varie lingue europee (Motolese 2012). Tra gli studi condotti in prospettiva sincronica menzioniamo Bellucci (1997; 2002), a cui in questo contributo ci siamo più volte riferiti, oltre che Bellucci & Biffi (2012) e Biffi (2003), mentre prediligono una prospettiva semio-linguistica Eco (1968: 191 ff.), Garroni (1970), Lenza (1975) e Koenig (1964). Per una rassegna sugli studi a partire dagli anni Novanta del secolo scorso rinviamo al puntuale lavoro di Biffi (2022), contenuto nella recente miscellanea di taglio interdisciplinare a cura di Armando *et al.* (2022) con contributi che esplorano ulteriori intersezioni tra i principali versanti della ricerca linguistica e di quella dell'AC.

Biffi 2012: 316), che invece oggi prevalgono, quanto meno in ambiti quali la Bioarchitettura o le Scienze dei materiali e delle Costruzioni (Gay 2015).

Dal punto di vista sociolinguistico emerge inoltre una chiara variabilità diastratica, poiché da sempre gli architetti interagiscono con un'ampia gamma di attori sociali: dai committenti fino ai professionisti sul campo come geologi, capimastri e manovali, oltre alle maestranze oggi molto spesso straniere (Biffi 2022: 24, 32). Anche nello specifico della comunicazione scientifico-disciplinare dell'AC, dal punto di vista della dimensione verticale (Berruto 1987; Cortelazzo 1990) esistono diversi livelli di specializzazione del discorso che vanno da quello altamente specialistico della comunicazione tra esperti a quello molto più vicino alla lingua comune nei testi della comunicazione divulgativa, con ulteriori variazioni anche in considerazione della dimensione diamesica (Gualdo & Telve 2015: 238-253).

2.1 Il corpus IT_SEAH

Allo stato attuale, le dimensioni del corpus IT_SEAH ammontano a 1.124.034 token, 970.521 parole e 293 testi ascrivibili a dodici generi testuali diversi. Come per gli altri corpora, la struttura generale tiene conto del canale di ricezione, orale o scritto, del dominio e del genere testuale di ciascun testo. Il 52% dei documenti collezionati appartiene al dominio accademico (monografie, articoli scientifici, tesi), il 22% a quello didattico (lezioni, dispense, manuali), il 23% è costituito da generi del dominio professionale (report di progetto, descrizione di progetto, descrizione di materiali), mentre solamente il restante 3% afferisce al dominio divulgativo (podcast, lezioni pubbliche, documentari). Nei prossimi anni è in programma un ampliamento del corpus IT_SEAH nella direzione di un maggiore bilanciamento tra i testi orali e scritti (che costituiscono ora, rispettivamente, il 21% e il 79% dei documenti raccolti), come pure di un'estensione degli ambiti da includere, data la molteplicità dei settori scientifico-disciplinari coinvolti nella macroarea dell'AC una delle "più eterogenee e trasversali" (Bellucci & Biffi 2012: 300). Al momento, per le ragioni connesse alle tempistiche del progetto, sono state individuate quattro aree tematiche attorno alle quali sono state sviluppate successivamente le attività dei moduli di lingua: storia dell'architettura, urbanistica, restauro architettonico e architettura sostenibile (cfr. § 1; Piccioni *et al.* 2021).

Il corpus composto da testi raccolti, ripuliti manualmente⁶, lemmatizzati e annotati utilizzando le funzionalità di Sketch Engine (Kilgarriff *et al.* 2014), è stato poi reso liberamente accessibile nel sito per una consultazione autonoma attraverso NoSketch Engine (§1), oppure per lo svolgimento dei compiti di ricerca guidata previsti nella sezione dei percorsi didattici. Questi contengono diverse tipologie di attività a vari livelli dell'analisi linguistica (cfr. D'Angelo *et al.* 2022: 15-16), unitamente ad esercizi in cui agli apprendenti viene richiesto di osservare le occorrenze di un determinato costrutto grammaticale o di una parola, al fine di sviluppare la loro capacità di

⁶ I testi orali, inizialmente trascritti automaticamente, sono stati sottoposti a post-editing manuale e uniformati in ordine alla trascrizione delle pause, dei frammenti non chiari, degli errori di pronuncia, delle parole troncate ecc. I file sono poi stati etichettati in base a un set di dieci classi di metadati che permettono il recupero automatico dei testi e la creazione di subcorpora per ricerche mirate in domini o generi specifici (Piccioni *et al.* 2021). Per la lemmatizzazione e l'annotazione morfosintattica dei testi di IT_SEAH è stato usato il software FreeLing (Padró & Stanilovsky 2012).

formulare e verificare ipotesi attraverso alcune abilità fondamentali, quali notare tratti linguistici o validare descrizioni. Risultano utili al tal proposito gli esercizi basati sull'uso della funzione *Concordance* di NoSketch Engine che presenta l'immediata visualizzazione dei co-testi in cui occorre un determinato fenomeno linguistico secondo il KWIC (*KeyWord In Context*). Tali compiti di ricerca, descritti nei paragrafi a seguire, mirano a promuovere la competenza metalinguistica dei discenti e, allo stesso tempo, l'apprendimento di un "life-long L2 learning tool" (Vyatkina 2016: 172)⁷.

2.2 Usi diretti del corpus nelle attività didattiche SEAH

Nella terza e settima sezione di ogni unità dei moduli di lingua SEAH sono previste attività che implicano la consultazione diretta dei corpora, come nel caso di un esercizio dedicato al tema della Bioarchitettura (unità 1, sezione 3). Agli studenti viene dapprima chiesto di visualizzare le occorrenze del termine *pannello* mediante la funzione *concordance*, per poter poi individuare le preposizioni usate in italiano per esprimere il complemento di materia, ovvero sia *di* che *in* (es. pannello *in/di* vetro). Nella sezione sette di ogni unità, il compito consiste nel redigere un mini-glossario formato da una tabella che raccoglie, nella prima colonna, alcune unità lessicali incontrate nelle attività precedenti e affiancate da una traduzione in inglese da noi fornita nella seconda colonna, mentre nella terza lo studente stesso dovrà inserire la resa di tali unità nella propria lingua madre (v. Fig. 1). In linea con altri studi sulla didattica dell'italiano accademico (Rosi & Berardini 2020), si è ritenuto opportuno inserire esercizi che coinvolgessero le altre lingue conosciute dagli apprendenti, seguendo il "cumulative-enhancement model" (Flynn & Foley & Vinnitskaya 2004) secondo cui tutte le lingue già presenti nel repertorio individuale concorrono a favorire l'acquisizione di nuove. Nel confronto con le unità lessicali della prima e della seconda colonna, gli studenti possono notare la somiglianza fra le parole italiane e quelle inglesi (in questo caso *capturing panels*/pannelli captanti e *connective movements*/moti connettivi) attivando utili meccanismi di transfer, di cui potrebbero beneficiare non solo sul piano della comprensione, ma anche su quello della produzione.

⁷ Nell'anno accademico 2022 il modulo di lingua italiana è stato sperimentato da 36 studenti in mobilità, partecipanti a corsi di Italiano L2 di livello B1 erogati dall'Ateneo di Chieti-Pescara. Attraverso un questionario anonimo di Google Forms i partecipanti hanno espresso il proprio giudizio in merito, tra l'altro, alla validità degli esercizi che prevedono l'impiego del corpus IT_SEAH per l'apprendimento dell'italiano dell'AC e alla possibilità in futuro di usare tale risorsa linguistica in maniera autonoma (D'Angelo 2022: 15). In maniera anonima, gli apprendenti hanno espresso il loro grado di accordo/disaccordo secondo una scala Likert con valori da 1 (=massimo disaccordo) a 5 (=massimo accordo). Pur nella consapevolezza di dover raccogliere un numero maggiore di riscontri, si evidenzia l'esito positivo di questa primissima indagine, dalla quale risulta che ben il 42% è favorevole alla consultazione del corpus per lo studio del linguaggio italiano dell'AC, il 33% si è dichiarato addirittura molto favorevole, mentre il 14% neutrale, l'8% non favorevole e il 3% in netto disaccordo. Altrettanto positivo il giudizio sulla possibilità di usare in futuro di IT_SEAH: il 31% degli intervistati si esprime con un giudizio di massimo accordo, il 28% di accordo, il 36% di neutralità e il restante 5% di disaccordo.



Riproduci audio

Italiano:	English:	Come si dice nella tua lingua?
Nelle scienze dell'atmosfera, l' effetto serra è un particolare fenomeno di regolazione della temperatura di un pianeta provvisto di atmosfera.	<i>Greenhouse effect</i>	La tua risposta...
La serra è un ambiente artificiale costruito per coltivare fiori e piante con temperature simili a quelle del loro habitat naturale.	<i>Greenhouse</i>	La tua risposta...
I pannelli captanti , sono superfici in grado di assorbire e trattenere energia nelle sue diverse forme.	<i>Capturing (the sun power) panels</i>	La tua risposta...
L' edilizia è l'insieme di attività riferite alla costruzione di edifici.	<i>Construction</i>	La tua risposta...
L' architettura bioclimatica usa elementi naturali (sole, vento, acqua...) per realizzare edifici termicamente efficienti.	<i>Bioclimatic (architecture)</i>	La tua risposta...
I Moti convettivi nei fluidi e gas, sono forme di movimento dell'energia termica che portano la temperatura ad una maggiore omogeneità.	<i>Convective movement</i>	La tua risposta...
Dal punto di vista energetico è previsto l'impiego di energia rinnovabile geotermica per il riscaldamento e	<i>Geothermics</i>	La tua risposta...

Figura1. Compilazione del mini-glossario (unità 1, sezione 7.1)

Il mini-glossario autonomamente compilato, potrà quindi essere salvato in formato pdf ed eventualmente ampliato in futuro. Seguono quindi esercizi di consultazione guidata del corpus volti al consolidamento del bagaglio lessicale (Fig. 2), per la cui esecuzione non viene richiesto agli apprendenti di padroneggiare la sintassi delle *query*, ma semplicemente di modificarla (in questo caso relativamente al lemma) dovendo fornire ulteriori aggettivi che occorrono assieme ai sostantivi 'edilizia', 'pannello' ed 'energia', incontrati precedentemente.

Mediante compiti guidati e di graduale complessità l'apprendente familiarizza con le diverse funzionalità di NoSketch Engine (come l'ordinamento delle concordanze, l'elaborazione di liste di frequenza in base a diversi criteri e l'analisi dei collocati) verso lo sviluppo di una prima *corpus literacy* (Mukherjee 2002), funzionale ad accrescere la capacità di analizzare testi specialistici a un livello profondo. Per i docenti, come illustreremo nei paragrafi a seguire, la consultazione di IT_SEAH potrebbe compensare le limitazioni dei dizionari e delle risorse tradizionali disponibili in rete per l'elaborazione di esercitazioni specifiche in ordine ai marchionimi (o nomi commerciali), ai termini formati da processi di derivazione o per mutamento semantico, come pure alle collocazioni tipiche del linguaggio disciplinare dell'AC per cui può rivelarsi utile un impiego diretto del corpus.



es. moderna, scolastica, continua _____

Per completare l'esercizio accedi al corpus SEAH (<https://www.seahproject.eu/resources.php>) e:

1) Seleziona Concordance > Advanced
 2) Seleziona Query type: CQL
 3) Copia e incolla questa "query": [lemma="edilizia"] [tag="A.*"]

The screenshot shows the Concordance software interface. On the left is a dark blue sidebar with icons for home, search, and other functions. The main area has tabs for BASIC, ADVANCED, and ABOUT. The ADVANCED tab is selected, showing a 'Query type' dropdown set to 'CQL'. Below this is a 'CQL BUILDER' section with a text input field containing the query '[lemma="edilizia"] [tag="A.*"]'. There are buttons for 'Insert', 'TAGS', and 'CQL BUILDER'. Below the query field is a 'Default attribute' dropdown set to 'lemma'. At the bottom, there is a 'SORT' section with a description: 'Sorts the concordance lines alphabetically by the KWIC or by the token to the left or right of KWIC. The lines can also be sorted by text types.' Below this is a 'Sort by:' section with buttons for 'left context', 'KWIC', and 'right context'. The 'right context' button is highlighted with a red arrow. Below the 'Sort by:' section is a 'Sort attribute' dropdown set to 'word'.

CONCORDANCE

IT_SEAH

BASIC ADVANCED ABOUT

Query type [?]

simple
lemma
phrase
word
character
CQL

CQL

[lemma="edilizia"] [tag="A.*"]

Insert [] { } < > " " & \

| ~ # TAGS

CQL BUILDER

Default attribute [?]

lemma

Or puoi visualizzare le concordanze del sostantivo edilizia, seguito da un aggettivo

Per velocizzare la tua ricerca ti consigliamo di ordinare la visualizzazione in ordine alfabetico premendo il comando **Sort** (indicato dalla freccia) e selezionando la prima parola alla destra della KWIC:

≡ SORT

Sorts the concordance lines alphabetically by the **KWIC** or by the token to the left or right of KWIC. The lines can also be sorted by **text types**.

Sort by:

left context [?] 3 2 1 KWIC **1** 2 3 right context [?]

Sort attribute [?]

word

Figura 2. Esercizi di consultazione diretta del corpus IT SEAH (unità 1, sezione 7.1)

2.3 Applicazioni per la didattica di derivati, marchionimi e acronimi

Nel lessico dei linguaggi specialistici il processo di formazione dei termini si avvale di diversi meccanismi morfologici, quali la derivazione per affissazione, ovvero mediante prefissi e suffissi (Grossmann & Reiner 2004: §§ 3-4-5; Dardano 2009: §§2-3), con esiti che vanno dall'accumulo di questi ultimi (es. rinvenuto in IT_SEAH: *riutilizzabilità*) fino a nomi deverbali a suffisso zero (es. *collaudo* da collaudare, v. Zolli 1989). Di quest'ultima tipologia Bellucci (1997: 168) annovera i sostantivi *affaccio*, *arriccio*, *concio*, *getto*, *intonaco*, *rinfianco*, *rincasso* e *rinzafo* che, eccezion fatta per gli ultimi due, sono attestati anche nel nostro corpus. Interrogando IT_SEAH con la funzione *concordance*, gli insegnanti possono agevolmente reperire numerosi esempi contestualizzati dei suddetti derivati per l'elaborazione di esercizi ad hoc, potendo scegliere di visualizzare i risultati della ricerca sia secondo l'indice KWIC sia frase per

frase secondo l'indice GDEX (Kilgarrieff *et al.* 2008). Scegliendo quest'ultima opzione, l'ordine di visualizzazione delle frasi è determinato da criteri automatici di 'buon esempio' definiti dall'algoritmo GDEX, ovvero proposizioni informative e facilmente comprensibili, alle quali viene assegnato automaticamente un punteggio di qualità testuale. Sebbene sia sempre auspicabile un controllo manuale, l'elenco risultante delle frasi può agevolare la selezione degli esempi da utilizzare per i materiali didattici (v. Fig. 3).

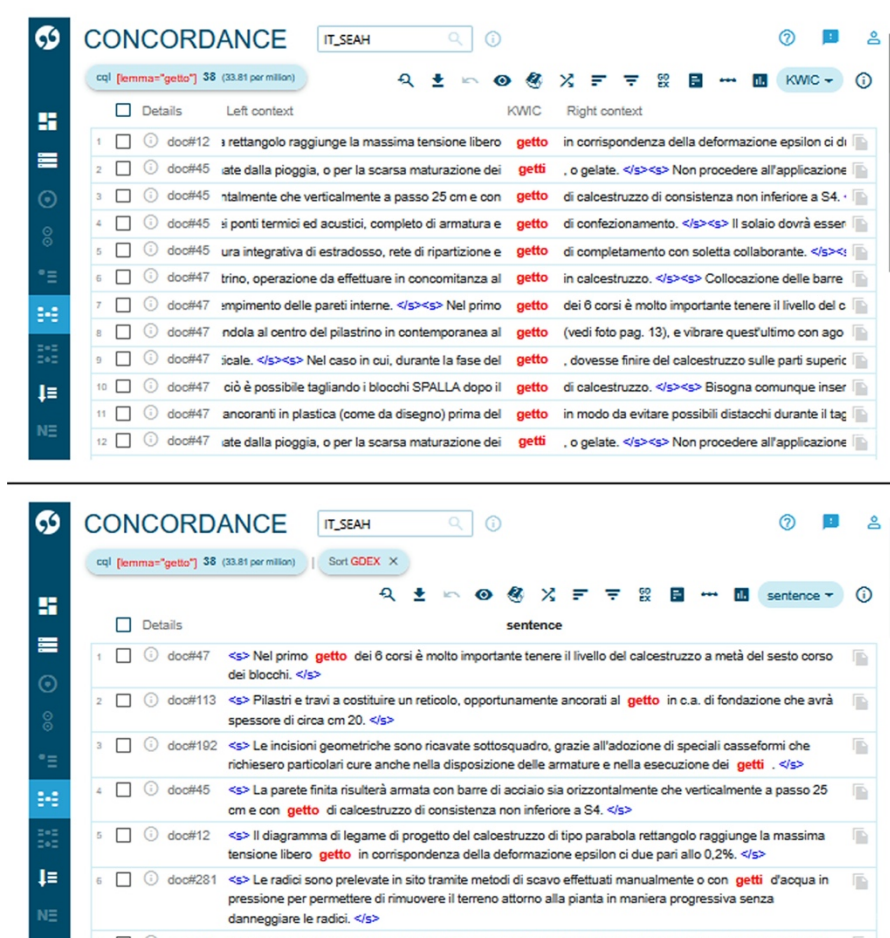


Figura 3. Visualizzazione secondo l'indice KWIC e secondo l'indice GDEX del deverbale *affaccio*

Allo stesso modo, seppur con un necessario spoglio manuale, la funzione Wordlist di NoSketch Engine consente in maniera altrettanto agevole il reperimento di sostantivi derivati per suffissazione, come per il tipo in *-tura* (v. Fig. 4), che Bellucci (1997: 169) ricorda essere molto produttivo nel linguaggio dell'architettura (esempi da IT_SEAH: *muratura, sigillatura, intelaiatura, rasatura*), o ancora di alterati (es. *casetta, villetta*), per i quali una volta individuati, sarà possibile reperire esempi contestualizzati interrogando il corpus con la funzione *Concordance*⁸.

⁸ Nell'interrogare il corpus per reperire esempi di alterati, sono stati individuati anche falsi alterati del lessico architettonico (*sanpietrino, cavalletto, lunetta, rosone, e frontone*) che abbiamo deciso di inserire in una breve scheda nella sezione 5 dell'unità 4, corredati di immagini esplicative (e divertenti) per far riflettere gli apprendenti su tale categoria.

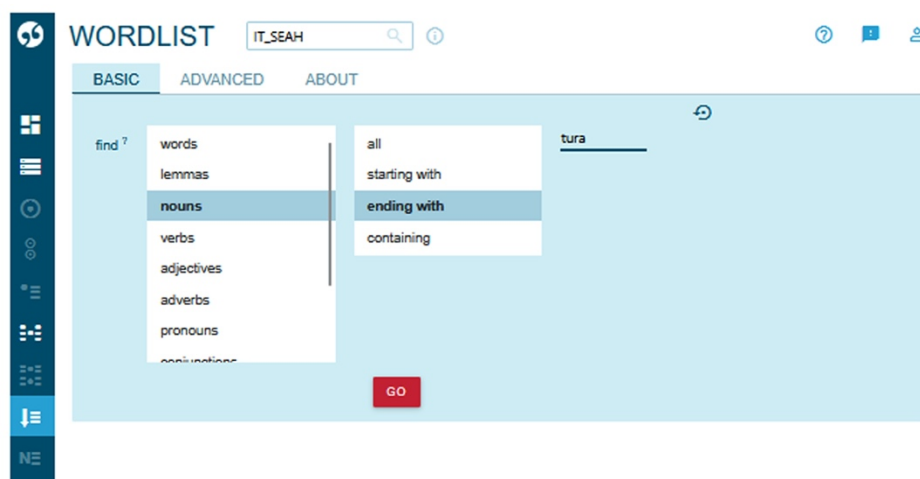


Figura 4. Ricerca di derivati tramite la funzione Wordlist nel corpus IT_SEAH

L'apparato terminologico dell'AC comprende inoltre i marchionimi che in genere derivano da nomi propri, come pure da nomi di fantasia, voci del lessico, sigle e acronimi. Bellucci ne rinviene alcuni nel corpus da lei consultato, quali ad esempio *eternit*, da tempo lessicalizzato e usato al posto del sostantivo *fibrocemento*, al quale aggiungiamo Velux, dal nome del marchio per finestre da tetto che nell'uso settoriale può funzionare da nome generico (es. *tipo velux* in IT_SEAH). Frequenti poi anche gli eponimi, derivati da nomi propri, come *mansarda* (dal nome dell'architetto François Mansart), *palladiano* e *palladianesimo* (aggettivo e sostantivo derivati dal nome dell'architetto Andrea Palladio) o ancora *tubi Innocenti* (dal nome dell'inventore Luigi Innocenti). Questi termini risultano in genere di facile comprensione per gli apprendenti, come pure quelli formatisi con l'aggiunta di prefissi e suffissi di origine greca o latina, detti anche prefissoidi e suffissoidi particolarmente diffusi nelle terminologie tecnico-scientifiche (es. dal nostro corpus: *mono-* in *monolocale* o *-metria* in *planimetria*). Tali elementi formativi, infatti, pur non essendo parole autonome, possiedono un valore semantico intrinseco, dal significato spesso riconoscibile nelle varie discipline scientifiche nelle diverse lingue. Questa trasparenza facilita la denominazione concettuale all'interno di un dominio e la successiva diffusione dei termini nella lingua comune (es. *monolocale*)⁹.

Risultano invece poco comprensibili le sigle e gli acronimi, assai diffusi nel settore dell'AC e per i quali abbiamo previsto una scheda esplicativa nei moduli SEAH, dietro specifica richiesta dei docenti dei Dipartimenti consultati. Nella sezione 6 dell'unità 3 dedicata all'urbanistica incontriamo, ad esempio, PRG (Piano Regolatore Regionale), NTA (Norme Tecniche di Attuazione), ZTO (Zone Territoriali Omogenee) e PTCP (Piano Territoriale Coordinamento Provinciale), attestati nei testi del dominio sia professionale che accademico¹⁰. Di contro, non sono state trattate le sigle e gli acronimi di origine internazionale presenti nel corpus, in quanto trasparenti per i discendenti, quali CAD (*Computer-Aided Design*), HVAC (*Heating Ventilation and Air Conditioning*) e

⁹ Rimandiamo a Iacobini (2011a:1145-1147; 2011b: 1428-1431) per una rassegna generale e Iacobini (2004: 69-95) per gli affissi di origine neoclassica, con i riferimenti alla vasta letteratura sul tema; cfr. inoltre Cortelazzo (1990).

¹⁰ IT_SEAH attesta 46 occorrenze per PRG, 6 per NTA, 2 per ZTO e una per PTCP.

RAL (*Reichs-Ausschuß für Lieferbedingungen*, usato per definire una scala di colori normalizzata nell'ambito delle vernici e dei rivestimenti).

L'utilizzo del corpus si rivela altresì utile per l'elaborazione di attività didattiche incentrate sui termini che si sono formati per “mutamento semantico” o “rideterminazione semantica” di parole della lingua quotidiana attraverso la specializzazione, ad esempio con l'aggiunta di un aggettivo, o tramite processi metaforici (cfr. Dardano 2009; Gualdo & Telve 2015: 81-90). Tra questi segnaliamo il sostantivo *cappotto* che nei principali dizionari gratuitamente accessibili online¹¹ è menzionato nell'unica accezione di indumento invernale, mentre in alcuni dizionari cartacei e glossari del settore è da tempo consolidato con l'accezione di sistema di isolamento a parete. È possibile risalire alla suddetta definizione tecnica soltanto effettuando una ricerca della specializzazione *cappotto termico* nella sezione riservata ai neologismi della Treccani¹². Non è detto però che gli apprendenti incontrino l'intero sintagma, nei testi specialistici, come nel caso di quelli raccolti nel nostro corpus e nei quali il lemma *cappotto* è presente con 63 occorrenze, di cui ben 46 senza il suddetto aggettivo.

Infine, ci pare utile sottolineare che in IT_SEAH il sostantivo sia attestato in un'unica occorrenza con il significato di capo di abbigliamento, a riprova di quanto sia più agevole per i docenti reperire esempi d'uso pertinenti rispetto a quelli che si otterrebbero interrogando valide risorse gratuite, che però non si basano su corpora specialistici, quali il corpus itSkELL (Baysa & Sucomelh 2014)¹³. Lo stesso è riscontrabile per l'eponimo *sandwich* (Caffarelli 2019: 13) che nel corpus SEAH occorre una sola volta con il significato di ‘tramezzino’, con cui è entrato dall'inglese nella lingua comune italiana, mentre è attestato per ben 21 volte nell'accezione specialistica di ‘pannello coibentato a tre strati’ (es. pannello *sandwich*, sistema isolante *sandwich* ecc.).

2.4 Applicazioni per la didattica delle collocazioni e delle polirematiche

Le collocazioni, ovvero le combinazioni di due o più unità lessicali co-occorrenti in modo frequente¹⁴, rivestono un ruolo primario nell'apprendimento linguistico. Una loro corretta comprensione e produzione è indice di un livello avanzato di competenza comunicativa del parlante (Howart 1996) e della sua eventuale appartenenza ad un certo gruppo sociale o professionale (Wray 2002). Inoltre, un uso estensivo e corretto delle collocazioni è una delle evidenze più significative nel distinguere le produzioni dei

¹¹ Sono state consultate le versioni gratuite online dei dizionari sui portali della Treccani e della Garzanti Linguistica, oltre che la versione gratuita del Sabatini Coletti accessibile dal sito del *Corriere della Sera*. I primi due, almeno nella versione gratuita, indicano unicamente l'accezione di *cappotto* come indumento, mentre il Sabatini Coletti riporta anche il significato di *cappotto* nel linguaggio marinairesco, detto cioè di un'imbarcazione che si capovolge, e della locuzione *fare cappotto* in ambito sportivo, dal significato di vincere senza lasciar fare punti all'avversario (https://dizionari.corriere.it/dizionario_italiano/C/cappotto_2.shtml) [1.7.2025].

¹² (https://www.treccani.it/vocabolario/neo-cappotto-termico_Neologismi) [1.7.2025].

¹³ Tale risorsa gratuita raccoglie testi tratti dal corpus itTenTen 2016, consultabili tramite l'interfaccia web SKELL (acronimo di *Sketch Engine for Language Learning*). I risultati delle visualizzazioni vengono mostrati sulla base di un punteggio assegnato alle frasi dal già menzionato algoritmo GDEX.

¹⁴ In riferimento alle collocazioni della lingua italiana si vedano i lavori di Gualdo & Telve (2015), Marelli (1996), Ježek (2005), Tiberii (2012).

nativi da quelle dei non-nativi, i quali tendono a sottoutilizzarle o utilizzarle in maniera impropria a fronte dell'impredicibilità delle combinazioni lessicali (Ellis *et al.* 2015). La padronanza nell'uso risulterebbe quindi difficoltosa per gli apprendenti di una L2, indipendentemente dagli anni di istruzione ricevuta nella lingua target, dalla lingua d'origine o dal tipo di attività svolta (Laufer & Waldman 2011). D'altro canto, una volta apprese, le collocazioni costituiscono una sorta di “prefabricated units” su cui gli studenti possono fare affidamento in produzione e in ricezione, a beneficio anche della loro *fluency* (Nesselhauf 2005: 2).

I corpora della suite di SEAH possono essere d'aiuto ai docenti per l'individuazione di tali fenomeni lessico-grammaticali nel linguaggio settoriale in questione e utilizzare gli esempi reperiti per l'elaborazione di esercizi di consultazione del corpus finalizzati ad attivare processi di *noticing* e sviluppare una competenza metalinguistica in merito a tale fenomeno. Ad esempio, nell'interrogazione del corpus alla ricerca delle collocazioni del tipo verbo+articolo+nome, quali *gettare il cemento* vs. versare il cemento o *assorbire una spinta* vs. resistere a una spinta. E ancora prendere consapevolezza dell'eventuale non corrispondenza dei collocati verbali tra le lingue del repertorio degli studenti, utilizzando i corpora SEAH per una riflessione in chiave contrastiva. A tal riguardo, un tipo specifico di collocazione che merita particolare attenzione è rappresentato dalle costruzioni con verbi leggeri (Jespersen 1942) o verbi supportanti (Elia & D'Agostino & Martinelli 1985) che designano quelle sequenze verbo-nominali in cui il significato è veicolato quasi interamente dal sostantivo, mentre il verbo perde parte del suo valore semantico, assumendo una funzione più grammaticale o ‘leggera’, come nel caso di *fare una domanda*, *dare un esame*, ecc.

La decodifica e soprattutto la produzione delle costruzioni a verbo supporto (CVS) può comportare difficoltà anche per apprendenti con livelli avanzati di competenza linguistica, come attestano gli esiti delle indagini applicate alla didattica della lingua inglese (Altenberg & Granger 2001), versante della ricerca glottodidattica nel quale è altresì sostenuta la validità dell'approccio DDL per l'insegnamento delle CVS mediante l'osservazione di dati contestualizzati per giungere alla scoperta induttiva delle combinazioni più frequenti e appropriate¹⁵.

In relazione alle CVS, il corpus IT_SEAH può essere usato per l'ideazione di una vasta gamma di compiti di ricerca guidata, da modulare sulla base del livello di competenza degli apprendenti. A questi si potrebbe richiedere di individuare le collocazioni più frequenti tra verbi e nomi, riconoscere le varianti sintattico-lessicali ammesse, analizzare il contesto d'uso per cogliere le sfumature semantico-aspettuali, o ancora osservare varietà diastratiche, diafasiche e diamesiche, mediante un'interrogazione filtrata per il dominio (accademico, didattico, professionale o divulgativo) e per il canale (scritto o orale).

Inoltre, nei corsi di formazione per futuri docenti di italiano L2 le esplorazioni si possono estendere a considerazioni sulla classificazione delle CVS individuate talora ascrivibili alla categoria delle collocazioni o a quella delle polirematiche verbali (Voghera 2004). Queste ultime rappresentano una difficoltà per gli studenti alloglotti poiché oltre alla fissità lessicale e sintattica interna, si caratterizzano per la non

¹⁵ In Koesters Gensini *et al.* (2023) si descrivono le premesse teorico-metodologiche e i primi risultati di un progetto avente per obiettivo la realizzazione di una risorsa multilingue, innovativa e collaborativa, per la didattica delle CVS destinata a studenti universitari di italiano L3. Lo studio contiene inoltre una preziosa rassegna delle principali indagini linguistiche e glottodidattiche condotte in ordine alle CVS in albanese, italiano, inglese, francese, olandese, rumeno, russo e tedesco.

composizionalità del significato. Tuttavia, qualora il sostantivo della CVS sia polisemico e, in ambito settoriale, usato nell'accezione specialistica, può presentarsi il caso il cui una polirematica appartenente alla lingua comune e pertanto più volte incontrata dagli apprendenti, risulti di più facile comprensione. In IT_SEAH ad esempio, tra le occorrenze ottenute per il verbo *dare*¹⁶ si rinvencono utili esempi che presentano diversi gradi di opacità semantica, come nelle frasi che contengono:

- (1) una polirematica verbale (Masini 2010: 1111), però frequente nella lingua comune: 'il più forte *dia una mano* al più debole'
- (2) una CVS di facile disambiguazione: 'proviamo a *dare delle misure* a questo ambiente' vs. *misurare*
- (3) una CVS con il sostantivo polisemico *mano*: '*dare una mano di fondo* con ...' vs. applicare uno strato (o più) di un prodotto su una superficie'.

Sarebbe pertanto utile focalizzare l'attenzione degli apprendenti sull'accezione dell'esempio (3), meno consueta nel linguaggio comune, con l'osservazione di altre frasi autentiche presenti nel nostro corpus¹⁷. Allo stesso modo, la risorsa agevola la ricerca delle polirematiche nominali, assai frequenti nei "linguaggi tecnico-specialistici" (De Mauro 2005: 89), per le quali abbiamo rinvenuto attestazioni del tipo N+Prep+N come *lana di roccia*, *bocca di lupo*, *latte di calce*, *corpo di fabbrica*. La non composizionalità di tali unità lessicali ne determina la difficile comprensione per gli studenti alloglotti, ai quali è opportuno proporre attività osservative e preferibilmente plurilingui che mirino ad una consapevolezza del fenomeno in sé, oltre che all'ampiamiento del vocabolario specialistico.

3. Conclusioni e prospettive di sviluppo

Il progetto SEAH ha inteso elaborare risorse linguistiche il più possibile aderenti alle esigenze formative degli studenti internazionali, con un focus sui partecipanti ai programmi di scambio di breve durata, la cui esperienza formativa all'estero risulta spesso ostacolata da insufficienti competenze linguistiche nel settore disciplinare di afferenza. Per ragioni organizzative, come pure per mancanza di docenti esperti nei vari linguaggi specialistici, l'offerta didattica delle sedi ospitanti è in genere limitata a corsi di italiano per scopi accademici generali, destinati agli *incoming students* dei livelli più avanzati. Le attività didattiche SEAH, invece, sono state ideate per apprendenti di livello B1 e possono essere svolte in autonomia online oppure in corsi in presenza. Esse ambiscono all'acquisizione di contenuti linguistici e disciplinari, come pure al potenziamento della competenza metalinguistica sulle peculiarità del linguaggio dell'AC e allo sviluppo di una prima *corpus literacy*.

¹⁶ Utilizzando le funzionalità del *Concordance* tool di NoSketch Engine si sono ottenute 172 occorrenze *dare* + nome (con la query [lemma="dare"][tag="N.*"]), 171 occorrenze del tipo *dare* + preposizione + nome ([lemma="dare"][tag="S.*"][tag="N.*"]), 136 del tipo *dare* + articolo + nome ([lemma="dare"][tag="S.*"][tag="N.*"]). Si è resa tuttavia necessaria una ulteriore ripulitura manuale degli esempi, per errori derivanti ad esempio dall'etichettatura errata, come nel caso di *dato/data* a volte considerati verbo anziché sostantivo.

¹⁷ Tra gli esempi ottenuti figurano *mano di adesivo*, *due mani di primer*, *una prima mano*, *numerose mani di ridipintura*.

In aggiunta, i corpora compilati ad hoc e liberamente consultabili dal sito del progetto, possono rappresentare per gli insegnanti una risorsa multilingue *user-friendly* che riduce potenzialmente il tempo e lo sforzo necessari per creare materiali didattici *corpus-informed* nel macrosettore in oggetto, o in ambiti affini, anche in prospettiva comparativa. Come illustrato qui per alcuni tratti caratteristici dell'apparato lessicale dell'AC, quali i termini formati per derivazione o per rideterminazione semantica, le collocazioni e le polirematiche (§ 2.3, 2.4), interrogando il corpus IT_SEAH si riduce notevolmente il cosiddetto rumore dovuto alla polisemia della parola nodo rispetto ai dati ottenuti da altre risorse gratuite che, seppur molto valide e intuitive nell'uso, sono basate sulla compilazione di corpora generali.

Nei prossimi anni le unità contenute nel modulo SEAH per la lingua italiana saranno ampliate in corrispondenza di un'estensione del corpus IT_SEAH, in particolar modo nella direzione sia di un maggiore bilanciamento tra i testi scritti e orali sia dell'inclusione di ulteriori aree tematiche, quali il Design e le Scienze delle Costruzioni. Al termine di questa fase, verrà sviluppato un test di competenza per verificare l'efficacia delle attività implementate in chiave DDL, un ambito questo ancora poco esplorato dalla ricerca empirica per l'italiano L2 (cfr. Forti 2023). In parallelo si utilizzeranno le risorse già disponibili nei corsi per futuri insegnanti, in risposta all'esigenza da più parte espressa di offrire percorsi formativi ai docenti sugli usi diretti e indiretti dei corpora (Römer 2008), finalizzati alla creazione autonoma di materiali adattabili ai vari contesti educativi non soltanto a livello di istruzione terziaria, ma anche nei contesti scolastici, quali i licei scientifici di vario indirizzo o gli istituti tecnici Costruzioni, Ambiente e Territorio.

Bibliografia

- Armando, Alessandro & Durbiano, Giovanni & Lucarini, Costanza & Scarpa, Raffaella (a cura di). 2022. *Tra le righe dell'Architettura: lingua, stile, testo*. Milano/Udine: Mimesis edizioni.
- Altenberg, Bengt & Granger, Sylviane. 2001. *Applied linguistics* 22 (2). 173–195.
- Aston, Guy (a cura di). 2001. *Learning with corpora*. Bologna: CLUEB.
- Balboni, Paolo E. 2018. Competenze morfosintattiche e testuali necessarie per lo studio accademico. In Jafrancesco Elisabetta (a cura di), *Le competenze trasversali dello studente universitario di Italiano L2*, 93–106. Siena: Becarelli.
- Baysa, Vít & Suchomel, Vít. 2014. SkELL: Web interface for English language learning. In Horák, Aleš & Rychlý, Pavel (a cura di), *Eighth workshop on recent advances in Slavonic natural language processing*, 63–70. RASLAN.
- Beaugrande, Robert de. 2001. Large corpora, small corpora, and the learning of “language”. In Ghadessy, Mohsen & Henry, Alex & Roseberry, Robert L. (a cura di), *Small Corpus Studies and ELT: Theory and practice*, 3–28. Amsterdam: John Benjamins Publishing Company.
- Bellucci, Patrizia. 1997. Gli usi speciali della lingua. Il linguaggio contemporaneo dell'architettura, con particolare riferimento al lessico. *Quaderni del Dipartimento di Linguistica* 8. 153–212.
- Bellucci, Patrizia. 2002. Barriere linguistiche e barriere architettoniche. Ostacoli nello spazio urbano e nello spazio linguistico *Quaderni del Dipartimento di Linguistica* 12. 47–77.

- Bellucci, Patrizia & Biffi, Marco. 2012. Note sulla lingua dell'architettura del nuovo millennio. In Nesi, Annalisa & De Martino, Domenico (a cura di) *Atti del Convegno Internazionale "Lingua italiana e scienza"* (Firenze 6–8 febbraio), 299–330. Firenze: Accademia della Crusca.
- Berruto, Gaetano. 1987. *Sociolinguistica dell'italiano contemporaneo*. Firenze: La Nuova Italia Scientifica.
- Biber, Douglas & Conrad, Susan & Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*, vol. 1. Cambridge: Cambridge University Press.
- Biffi, Marco. 1999. Sul lessico architettonico: alcuni casi controversi dalle traduzioni vitruviane. In *Studi di Lessicografia italiana* 26. 31–161.
- Biffi, Marco. 2003. Aspetti del lessico architettonico italiano. In Maraschio, Nicoletta & Poggi Salani, Teresa (a cura di), *Italia linguistica anno Mille. Italia linguistica anno Duemila*, Atti del XXXIV Congresso Internazionale di Studi della Società di Linguistica Italiana (SLI), (Firenze, 19-21 ottobre 2000), 303–316. Roma: Bulzoni.
- Biffi, Marco. 2006. Il lessico dell'architettura nella storia della lingua italiana. In Gudelj, Jasenka & Nicolin, Paola (a cura di), *Costruire il dispositivo storico. Tra fonti e strumenti*, 75–132. Milano: Mondadori.
- Biffi, Marco. 2010. Arte e critica d'arte, lingua dell'. In Simone, Raffaele (a cura di), *Enciclopedia dell'italiano*, 106–108. Roma: Treccani.
- Biffi, Marco. 2019. All'alba di un lessico intellettuale europeo dell'architettura. In Aresti, Alessandro (a cura di), *Lingua delle arti e lingua di artisti in Italia fra Medioevo e Rinascimento*, 37–60. Firenze: Cesati.
- Biffi, Marco. 2022. La lingua di architettura: alcune riflessioni a margine di venticinque anni di studi. In: Armando, Alessandro & Durbiano, Giovanni & Lucarini, Costanza & Scarpa, Raffaella (a cura di), *Tra le righe dell'Architettura: lingua, stile, testo*, 13–40, Milano/Udine: Mimesis edizioni.
- Bonvino, Elisabetta & Rastelli, Stefano (a cura di). 2011. *La didattica dell'italiano a studenti cinesi e il progetto Marco Polo*. Atti del XV seminario AICLU. Pavia: Pavia University Press.
- Bruni, Francesco. 2013. *L'italiano fuori d'Italia*. Firenze: Cesati.
- Bybee, Joan L. 2006. From usage to grammar: The mind's response to repetition. *Language*, 82(4). 711–733.
- Caffarelli, Enzo. 2019. *Parole comuni da nomi propri*. Milano: RCS.
- Ciliberti, Anna & Anderson, Laurie (a cura di) 1999. *Le forme della comunicazione accademica. Ricerche linguistiche sulla didattica universitaria in ambito umanistico*. Milano: Franco Angeli.
- Cortelazzo, Michele. 1990. *Lingue speciali. La dimensione verticale*. Padova: Unipress.
- D'Angelo, Mariapia. 2021. Discorso accademico-disciplinare in prospettiva glottodidattica: una proposta multilingue *corpus based* per il linguaggio dell'architettura e delle costruzioni. In *Lingue, testi e discorsi. Studi in onore di Paola Desideri*, 347–361. Firenze: Cesati.
- D'Angelo, Mariapia. 2022. Risorse linguistiche *corpus based* per l'italiano dell'architettura e delle costruzioni. *Italiano a Stranieri* 23. 11–15.
- D'Angelo, Mariapia & Ferro Maria C. & Piccioni, Sara. 2022. Introduzione alla sezione monografica "Insegnare il linguaggio dell'Architettura e delle Costruzioni: corpora e moduli e-learning SEAH". *Rassegna Italiana di Linguistica Applicata* 54 (1/2). 9–22.

- D'Angelo, Mariapia & Flinz, Carolina. 2025. Corpora di lingua parlata per la didattica dell'italiano e del tedesco per scopi accademici. In Roccaforte, Maria (a cura di), *La comunicazione parlata / Spoken Communication. Roma 2023 - I venti anni del GSCP*, 563–583. Roma: Aracne.
- Dardano, Maurizio. 2009. *Costruire le parole: la morfologia derivativa dell'italiano*. Bologna: il Mulino.
- De Mauro, Tullio. 2005. *La fabbrica delle parole. Il lessico e problemi di lessicologia*. Torino: UTET.
- De Mauro, Tullio & Vedovelli, Massimo & Barni, Monica & Miraglia, Lorenzo. 2002. *Italiano 2000. I pubblici e le motivazioni dell'italiano diffuso fra stranieri*. Roma: Bulzoni.
- Desideri, Paola. 2011. *Pragmatica, argomentazione e glottodidattica del discorso accademico*. In Desideri, Paola & Tessuto, Girolamo (a cura di), *Il discorso accademico. Lingue e pratiche disciplinari*, 47–72. Urbino: Quattro Venti.
- Durrant, Philip. 2014. Discipline- and level-specificity in university students' written vocabulary. *Applied Linguistics* 35(3). 328–356.
- Eco, Umberto. 1968. *La struttura assente*. Milano: Bompiani.
- Elia, Annibale & D'Agostino, Emilio & Martinelli, Maurizio. 1985. Tre componenti della sintassi italiana: frasi semplici, frasi a verbo supporto e frasi idiomatiche. In Franchi De Bellis, Annalisa & Savoia, Leonardo M. (a cura di), *Sintassi e morfologia della lingua italiana d'uso. Teorie e applicazioni descrittive*. Atti del XVII congresso internazionale della Società di Linguistica Italiana (Urbino, 11-13 settembre 1983), 311–325. Roma: Bulzoni.
- Ellis, Nick C. & Simpson-Vlach, Rita & Römer, Ute & O'Donnell, Matthew B. & Wulff, Stefanie. 2015. Learner corpora and formulaic language in SLA. In Granger, Sylviane & Gilquin, Gaëtanelle & Meunier, Fanny (a cura di), *Cambridge Handbook of Learner Corpus Research*, 357–378. Cambridge: Cambridge University Press.
- Fiorentino, Giuliana. 2015. Aspetti problematici del discorso accademico. *Cuadernos de Filología Italiana* 22. 263–284.
- Flinz, Carolina & Perkuhn, Rainer. 2018. Wortschatz und Kollokationen in „Allgemeine Reisebedingungen“. Eine intralinguale und interlinguale Studie. In Čibej, Jaka & Gorjanc, Vojko & Kosem, Iztok & Krek, Simon (a cura di), *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Context*, 959–67. Ljubljana: Ljubljana University Press, Faculty of Arts.
- Flowerdew, Lynne. 2015. Data-driven learning and language learning theories: Whither the twain shall meet. In Lenko-Szymanska, Agnieszka & Boulton, Alex (a cura di), *Multiple Affordances of Language Corpora for Data-driven Learning*, 15–36. Amsterdam. John Benjamins.
- Flowerdew, Lynne. 2022. Using corpora for writing instruction. In *The Routledge handbook of corpus linguistics*, 534–546. Oxon/New York: Routledge.
- Flynn, Suzanne & Foley, Claire & Vinnitskaya, Inna. 2004. The cumulative-enhancement model for language acquisition: Comparing adults and children's patterns of development in first, second and third language acquisition of relative clauses. *International Journal of Multilingualism* 1(1), 3–16.
- Forti, Luciana. 2023. *Corpus Use in Italian Language Pedagogy. Exploring the Effects of Data-driven learning*. London/New York: Routledge.

- Gallina, Francesca. 2018. Studenti internazionali in mobilità: la questione del lessico della conoscenza in italiano L2. In Coonan, Mary C. & Bier, Ada & Ballarin, Elena (a cura di), *La didattica delle lingue nel nuovo millennio: le sfide dell'internazionalizzazione*, Atti del IV Convegno Internazionale della Società di Didattica delle Lingue e Linguistica Educativa (DILLE), Venezia 2-4 febbraio 2017, 323–339. Venezia: Edizioni Ca' Foscari.
- Gallina, Francesca. 2019. Condizioni di possibilità per un piano di alfabetizzazione in italiano L2 per studenti internazionali. *Italiano Lingua Due* (1), 29–44.
- Garroni, Emilio. 1970. Semiotica e architettura. Alcuni problemi teorico-applicativi. «*Op. Cit.*» 18. 5–33.
- Gay, Debora. 2015. Architettura (eco)sostenibile, bioedilizia, Green Building: osservazioni sul lessico architettonico italiano in prospettiva sincronica e diacronica. In Pirvu, Elena (a cura di), *La lingua e la letteratura italiana in prospettiva sincronica e diacronica*, Atti del Convegno Internazionale di Studi di Craiova, 19-20 settembre 2014, 155–172. Firenze: Cesati.
- Grandi, Nicola (a cura di). 2025. *L'italiano scritto degli studenti universitari. Quadro sociolinguistico, tendenze tipologiche, implicazioni didattiche*. Milano: FrancoAngeli.
- Grossmann, Maria & Rainer, Franz (a cura di). 2004. *La formazione delle parole in italiano*. Tübingen: Niemeyer.
- Gualdo, Riccardo & Telve, Stefano. 2015. *Linguaggi specialistici dell'italiano*. Roma: Carocci. (1ª edizione 2011).
- Howarth, Peter. 1996. *Phraseology in English academic writing: some implications for language learning and dictionary making*. Berlin/New York: de Gruyter.
- Hyland, Ken & Tse, Polly. 2004. "I would like to thank my supervisor". Acknowledgements in graduate dissertations. *International Journal of Applied Linguistics* 14(2). 259–275.
- Hyland, Ken. 2004. *Disciplinary discourses: Social interactions in academic writing*. Ann Arbor: University of Michigan Press.
- Hyland, Ken. 2016. General and specific EAP. In Hyland, Ken & Shaw, Philip (a cura di), *The Routledge handbook of English for Academic Purposes*, 17–29. London: Routledge.
- Iacobini, Claudio. 2004. Composizione con elementi neoclassici. In Grossmann, Maria & Rainer, Franz (a cura di), *La formazione delle parole in italiano*, 69-95. Tübingen: Niemeyer.
- Iacobini, Claudio. 2011a. *Prefissi*. In Simone, Raffaele (a cura di), *Enciclopedia dell'italiano*, 1145–1147. Roma: Treccani.
- Iacobini, Claudio. 2011b. *Suffissi*. In Simone, Raffaele (a cura di), *Enciclopedia dell'italiano*, 1428–1431. Roma: Treccani.
- Jespersen, Otto. 1942. *A Modern English Grammar on Historical Principles*, vol. 4: Morphology, London: George Allen and Unwin Ltd.
- Ježek, Elisabetta (2005). *Lessico. Classi di parole, strutture, combinazioni*. Bologna: Il Mulino.
- Johns, Tim. 1990. From printout to handout: Grammar and vocabulary teaching in the context of Data-Driven Learning. *CALL Austria* 10. 14–34.
- Káňa, Tomas & Hradílková, Helena. 2022. Unterschiede zwischen allgemeinsprachlichen und fachsprachlichen Texten als Grundlage für Data-Driven-

- Learning: ein Beitrag zur Förderung korpuslinguistischer Kompetenz. *Rassegna Italiana di Linguistica Applicata* 54 (1/2). 83–99.
- Kilgarrieff, Adam & Baisa, Vít & Bušta, Jan & Jakubiček, Miloš & Kovář, Vojtěch & Michelfeit, Jan & Rychlý, Pavel & Suchomel, Vít. 2014. The Sketch Engine: Ten Years On. *Lexicography – Journal of ASIALEX* 1(1). 7–36.
- Kilgarrieff, Adam & Husák, Miloš & McAdam, Katy & Rundell, Michael & Rychlý, Pavel. 2008. GDEX: Automatically finding good dictionary examples in a corpus. In Bernal, Elisenda & De Cesaris, Janet (a cura di), *Proceedings of the 13th EURALEX International Congress*, 425–432. https://euralex.org/elx_proceedings/Euralex2008/00printedbook.pdf [1.6.2025].
- Koenig, Giovanni. 1964. *Analisi del linguaggio architettonico*. Firenze: LEF.
- Koesters Gensini, Sabine E. & Verkade, Suze A. & Berardini, Andrea & Escoubas, M. Pierre & Dashi, Brunilda & De Salazar, Danilo & Forti, Luciana & Nikolaeva, Joulia & Roccaforte, Maria & Vaccari, Debora. 2023. Dalle costruzioni a verbo supporto italiane alle lingue terze: un percorso di studio universitario. In Sostero Henrot, Geneviève (a cura di), *Alle radici della fraseologia europea*, 507–554. Lausanne/Berlin, ecc.: Peter Lang.
- Laufer, Batia & Waldman, Tina. 2011. Verb-noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning* 61 (2). 647–672.
- Lavinio, Cristina. 2004. *Comunicazione e linguaggi disciplinari. Per un'educazione linguistica trasversale*. Roma: Carocci.
- Leech, Geoffrey. 1997. Teaching and language corpora: A convergence. In Wichmann, Anne & Fligelstone, Steven & McNery, Tony & Knowles, Gerry (a cura di), *Teaching and language corpora*, 1–23. London: Longman.
- Lenza, Corrado. 1975. La teoria di Hjelmslev e l'architettura: alcuni problemi, in «*Op. Cit.*» 34. 37–77.
- Lo Duca, Maria G. 2006. *Sillabo di italiano L2*. Carocci: Roma.
- Long, Mike H. 2005. Methodological issues in learner needs analysis. In Long, Mike H. (a cura di), *Second language needs analysis*, 19–76. Cambridge: Cambridge University Press.
- Lubello, Sergio. 2019. L'italiano scritto accademico all'università tra L1 e L2: riflessioni e proposte per un curriculum. *TESTI E LINGUAGGI* 13.178–187.
- Marazzini, Claudio. 1993. *Storia della lingua italiana. Il secondo Cinquecento e il Seicento*. Bologna: il Mulino.
- Marazzini, Claudio. 1994. *La lingua italiana. Profilo storico*. Bologna: il Mulino.
- Marengo, Carla. 1996. *Le parole dell'italiano. Lessico e dizionari*. Bologna: Zanichelli.
- Masini, Francesca. 2010. Polirematiche, parole. In Simone, Raffaele (a cura di), *Enciclopedia dell'italiano*, 1109–1112. Roma: Treccani.
- Mezzadri, Marco. 2016. *Studiare in italiano all'università. Prospettive e strumenti*. Torino: Bonacci.
- Migliorini, Bruno. 1960. *Storia della letteratura italiana*. Firenze: Sansoni.
- Motolese, Matteo. 2012. *Italiano lingua delle arti. Un'avventura europea (1250-1650)*. Bologna: Il Mulino.
- Mukherjee, Joybrato. 2002. *Korpuslinguistik und Englischunterricht: Eine Einführung*. Frankfurt/M. ecc.: Peter Lang.
- Nencioni, Giovanni. 1995. Sulla formazione di un lessico nazionale dell'architettura. *Bollettino d'informazioni del Centro di Ricerche Informatiche per i Beni Culturali*

- (5) 2, 7–33 (poi in Id., *Saggi e memorie*. Pisa: Scuola Normale Superiore, 2000, 51–74).
- Nesi, Hilary & Gardner, Sheena. 2012. *Genres across the Disciplines: Student Writing in Higher Education*. Cambridge: Cambridge University Press.
- Nesselhauf, Nadja. 2005. *Collocations in a learner corpus*. Amsterdam/Philadelphia: Benjamins.
- Padró, Lluís & Stanilovsky, Evgeny. 2012. FreeLing 3.0: Towards Wider Multilinguality. In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)* ELRA. <https://nlp.lsi.upc.edu/publications/papers/padro12.pdf>. [1.6.2025].
- Piccioni, Sara & D'Angelo, Mariapia & Ferro Maria C. 2021. I Corpora SEAH di comunicazione specializzata nel settore dell'Architettura e delle Costruzioni. Struttura, compilazione e usi. *Linguistica* 61 (2). 97–122.
- Rees, Geraint P. 2021. Discipline-specific academic phraseology: Corpus evidence and potential applications. In Charles, Maggie & Frankenberg-Garcia, Ana (a cura di), *Corpora in ESP/EAP writing instruction: Preparation, Exploitation, Analysis*, 32–54. New York/London: Routledge.
- Richards, Jack C. & Rodgers, Theodore S. 2014. *Approaches and methods in language teaching*. Cambridge: Cambridge University Press.
- Römer, Ute & Cortes, Viviana & Friginal Eric (a cura di). 2020. *Advances in Corpus-Based Research on Academic Writing: Effects of Discipline, Register, and Writing Expertise*. Amsterdam/Philadelphia: John Benjamins.
- Römer, Ute. 2008. Corpora and language teaching. In Lüdeling, Anke & Kytö Merja (a cura di), *Corpus Linguistics. An International Handbook*, Vol. 1, 112–130. Berlin: Mouton de Gruyter.
- Rosi, Fabiana & Berardini, Fabiana. 2020. L'acquisizione dell'italiano in Italia e in Germania: strategie lessicali per costruire la referenza, *Italiano LinguaDue* 2. 189–200.
- Rychlý, Pavel. 2007. Manatee/Bonito – A Modular Corpus Manager. In *1st Workshop on Recent Advances in Slavonic Natural Language Processing*, 65–70. Brno: Masaryk University.
- Salzmann, Katharina. 2021. Zum Einsatz des GeWiss-Korpus im Rahmen einer mehrsprachigkeitsdidaktischen Vermittlung der gesprochenen Wissenschaftssprache. *Korpora Deutsch als Fremdsprache* 1(1). 79–96.
- Schmidt, Richard. 2001. Attention. In Robinson, Peter (a cura di), *Cognition and second language instruction*, 3–32. Cambridge: Cambridge University Press.
- Sinclair, John M. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sobrero, Alberto A. & Lavinio, Cristina (a cura di). 1991. *La lingua degli studenti universitari*. Firenze: La nuova Italia.
- Spina, Stefania & Forti, Luciana & Grego Bolli, Giuliana. 2020. Input, frequenza e modelli usage-based: il Data-driven learning come interfaccia tra teorie acquisizionali e pratica didattica. In Rastelli, Stefano & Bagna, Claudia (a cura di), *Manifesto per l'insegnamento della lingua italiana agli studenti internazionali. Otto commenti*, 73–98. Pisa: Pacini.
- Spina, Stefania. 2010. AIWL: una lista di frequenza dell'italiano accademico. In Bolasco, Sergio & Chiari, Isabella & Giuliano, Luca (a cura di), *Statistical Analysis*

- of Textual Data*, Proceedings of the 10th Conference JADT (Rome, 9-11 June 2010), 1317–25. Milano: Editrice Universitaria LED.
- Thompson, Paul & Nesi, Hilary. 2001. The British Academic Spoken English (BASE) Corpus Project. *Language Teaching Research* 5(3). 263–264.
- Thompson, Paul. 2022. Corpus analysis of disciplinary variation and the teaching of ESP/EAP. In Jablonkai, Reka R. & Csomay, Eniko (a cura di), *The Routledge handbook of corpora and English language teaching and learning*, 177–192. London: Routledge.
- Tiberii, Paola. 2012. *Il dizionario delle collocazioni. Le combinazioni delle parole in italiano*. Bologna: Zanichelli.
- Tomasello, Michael. 2003. *Constructing A Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Vedovelli, Massimo. 2010. *Guida all'italiano per stranieri*. Roma: Carocci.
- Voghera, Miriam. 2004. Polirematiche. In Grossmann, Maria & Rainer, Franz (a cura di). *La formazione delle parole in italiano*, 56–69. Tübingen: Niemeyer.
- Vyatkina, Nina. 2016. Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology* 20(3). 159–179.
- Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Zolli, Paolo. 1989. *Come nascono le parole italiane*. Milano: Rizzoli.

L'italiano delle traduzioni come varietà di contatto: un approccio tipologico a un corpus di narrativa

Emanuela De Vita

Università di Bologna

emanuela.devita3@unibo.it

Abstract

Si presentano i principali risultati di una ricerca¹ volta a esplorare le caratteristiche tipologiche dell'italiano delle traduzioni dall'inglese nell'ambito della narrativa commerciale. L'analisi, quantitativa e qualitativa, si basa sull'idea ampiamente dimostrata in letteratura che la traduzione costituisca un luogo insolito di contatto linguistico, capace di generare fenomeni di interferenza. In particolare, un corpus di romanzi italiani originali è stato messo a confronto con un corpus di romanzi tradotti dall'inglese relativamente ai seguenti tratti sintattici: esplicitazione del pronome soggetto, posizione del soggetto con verbi inaccusativi, posizione dell'aggettivo qualificativo rispetto al nome e uso della perifrasi '*stare* + gerundio' per l'espressione dell'aspetto progressivo. La ricerca conferma il ruolo cruciale dell'interferenza che, pur non essendo l'unica forza in gioco nel processo traduttivo, è in grado di generare anomalie nel profilo tipologico dell'italiano tradotto dall'inglese rispetto all'italiano originale.

1. Introduzione

In un'epoca più che mai dominata da connessioni globali, produzione e consumo su larga scala, accade sempre più spesso che anche prodotti culturali come libri e film vengano "preconfezionati" in luoghi del mondo diversi rispetto a quelli dove avviene la loro ricezione. Oltre a generare timori a causa della tendenza alla delocalizzazione e all'omologazione dei contenuti, la questione ha una conseguenza prettamente linguistica: la maggior parte delle opere (letterarie, cinematografiche, audiovisive...) viene creata in una lingua – spesso l'inglese – e fruita in un'altra, il che presuppone che essa debba essere sottoposta a un processo di traduzione.

¹ Il presente contributo rappresenta la sintesi di parte di una più ampia ricerca confluita nella tesi magistrale *L'italiano delle traduzioni come varietà di contatto. Un approccio tipologico a un corpus di narrativa* (De Vita 2024). Si ringraziano il prof. Nicola Grandi e la prof.ssa Silvia Ballarè per i loro preziosi consigli.

Le ricerche condotte negli ultimi cinquant'anni dimostrano che tale processo non è un processo neutro: il fatto stesso che un testo provenga da una traduzione implica che la sua lingua si discosti dall'uso spontaneo dei parlanti nativi, e al contempo assomigli, in qualche misura, a quella degli altri testi tradotti.

Più nello specifico, a partire dalla svolta descrittivista che ha interessato i *Translation Studies* negli anni '70-'80 del secolo scorso, l'attenzione dei ricercatori ha iniziato a rivolgersi sempre di più alle similitudini identificabili in testi tradotti differenti per lingua d'origine e lingua di destinazione. Esse sarebbero da ricondursi a una certa regolarità nel comportamento del traduttore che, indipendentemente dalle sue caratteristiche individuali, tende a mettere in atto strategie volte a rendere il testo più facilmente fruibile per il nuovo pubblico di lettori. Questa considerazione ha portato gli studiosi a ipotizzare l'esistenza di "universal features of translation" (Baker 1993: 243) o "laws of translational behaviour" (Toury 1995: 259), in seguito detti semplicemente *translation universals* ("universali della traduzione" o "universali traduttivi" nella letteratura italiana). Sotto questa etichetta si fanno comunemente rientrare diversi fenomeni, spesso suddivisi nelle macro-categorie di esplicitazione, semplificazione, normalizzazione e convergenza stilistica (Baker 1993, 1996; Salsnik 2007; Ondelli 2020). Non essendo gli universali traduttivi il *focus* di questa ricerca, non si entrerà nello specifico della loro descrizione. Basti in questa sede ricordare che l'etichetta "esplicitazione" vuole descrivere la propensione del traduttore a rendere esplicite informazioni implicite nel testo originale; si parla invece di "semplificazione" per indicare la tendenza a rendere il testo di più immediata comprensione, attraverso scelte lessicali, sintattiche e stilistiche più semplici e dirette; la "normalizzazione" riguarda l'adattamento del testo tradotto alle norme o convenzioni della lingua d'arrivo, con conseguente possibile perdita di tratti originali o innovativi; e infine, la "convergenza stilistica" ("levelling out" in Baker 1996) evidenzia come i testi tradotti tendano ad assomigliarsi tra loro, mostrando una ridotta variabilità linguistica rispetto a quanto non facciano corpora comparabili di testi originali. Insieme, questi fenomeni delineano una tendenza generale alla semplificazione – intesa questa volta in senso lato –, la cui natura, tuttavia, è ancora dibattuta. Lo stesso termine "universali" è stato più volte contestato, poiché il suo rimando diretto agli universali linguistici può suggerire l'esistenza di una comune matrice cognitiva che, nel caso degli universali traduttivi, non è affatto accertata. Non a caso, ad oggi, molti studiosi preferiscono termini come "leggi" o "regolarità" (Salsnik 2007).

Tuttavia, la lingua della traduzione non è determinata unicamente da questa "forza centripeta" che porta alla comparsa di tratti comuni in traduzioni differenti. Esiste, infatti, una seconda forza, generata dall'interazione delle due specifiche lingue coinvolte nel processo. Essa riguarda la propensione del traduttore a modellare le sue scelte linguistiche sulla base della lingua in cui è scritto il testo originale. È in questo senso che la traduzione può essere concepita come luogo insolito di contatto linguistico, un contatto che non coinvolge intere comunità di parlanti protraendosi a lungo nel tempo, ma si concentra nella singola competenza del traduttore², e ciononostante genera un fenomeno ben noto: l'interferenza linguistica.

Come precisa Toury, formulando la cosiddetta "law of interference" (1995: 275), l'interferenza è una sorta di condizione di *default* nelle traduzioni: sarebbe infatti la creazione di un ipotetico testo privo di qualsiasi effetto di interferenza a necessitare di

² Sebbene, soprattutto in alcuni contesti di largo consumo, gli esiti dei fenomeni di contatto siano fortemente condivisi dalla comunità dei traduttori e talvolta finiscano per essere adottati dalla comunità parlante (Garzone, 2004).

uno sforzo fuori dal comune da parte del traduttore. In condizioni normali, al contrario, tratti linguistici relativi alla composizione del testo di partenza, e dunque tipici di quella lingua, tenderanno a essere trasferiti nella lingua del testo di arrivo.

Si noti che l'interferenza non opera in maniera totalmente casuale; è possibile, infatti, analizzarla e delimitarne i confini. Cardinaletti (2005), ad esempio, osserva che il fenomeno che ha luogo nelle traduzioni è particolarmente affine, per certi versi, a un tipo particolare di interferenza che si verifica in precise condizioni: l'attrito linguistico³. In particolare, l'interferenza, così come l'attrito, non è obbligatoria, ossia non si verifica sempre: si può verificare laddove il traduttore abbia la possibilità di compiere una scelta. Tipicamente, una delle alternative rappresenta l'opzione più "naturale" per la lingua d'arrivo, la seconda – quella suggerita dalla lingua da cui si traduce – risulta invece "meno naturale" o meno adeguata al contesto, senza tuttavia essere totalmente agrammaticale. Ciò non significa che il traduttore compia coscientemente una scelta ogniqualevolta esista questa doppia possibilità: così come nel caso dell'attrito, conseguenza dell'interferenza cognitiva tra due o più sistemi linguistici, l'interferenza nella traduzione non si seleziona a piacere, bensì viene attuata in maniera inconsapevole, seppur non sistematica (Cardinaletti, 2005).

Dalla volontà, dunque, di esplorare le peculiarità della lingua della traduzione, e nello specifico l'italiano della traduzione, nasce l'idea alla base della ricerca che qui si presenta. Si è scelto di focalizzare l'attenzione sulla narrativa commerciale, ambito nel quale la traduzione ha un peso sempre più consistente sia in termini di numero assoluto di opere immesse sul mercato che in termini di successo editoriale e, dunque, di contatto con il pubblico di italofoni (Ondelli 2020). I dati sul numero di traduzioni hanno portato, inoltre, alla decisione di circoscrivere il campo d'indagine alle traduzioni in italiano a partire dall'inglese, lingua che primeggia, con grande distacco rispetto agli altri, tra gli idiomi da cui provengono le opere tradotte.

2. Stato dell'arte

Per quanto riguarda la ricerca sulla lingua delle traduzioni in ambito italiano, gli studi sono ancora poco numerosi. Si citano qui alcuni dei principali.

Una delle prime serie di studi significativi sull'argomento è stata condotta nell'ambito del progetto *Osservatorio dell'italiano contemporaneo. Analisi linguistica e applicazioni didattiche e traduttive*, a cura di Anna Cardinaletti e Giuliana Garzone, i cui risultati sono stati pubblicati nei volumi collettivi *Lingua, mediazione linguistica e interferenza* (2004) e *L'italiano delle traduzioni* (2005). Tale progetto ha portato all'individuazione di numerose caratteristiche tipiche dell'italiano delle traduzioni, soprattutto a livello di morfologia pronominale e verbale e sintassi della frase e del periodo.

³ L'attrito è un fenomeno che interessa i parlanti "quasi nativi", cioè persone che hanno vissuto a lungo in un paese straniero, raggiungendo un grado molto elevato di competenza nella nuova lingua, seppur questa non sia per loro nativa. In questo tipo di parlanti, il contatto esteso nel tempo con la nuova lingua provoca "la modificazione (parziale) della grammatica della loro lingua nativa" (Cardinaletti 2005: 59).

Si ricorda in questa sede il contributo di Garzone (2005), che si basa sull'analisi di testi inglesi di vario genere (narrativa commerciale, cronaca, articoli scientifico-divulgativi...) relativi a un arco temporale di circa mezzo secolo. I risultati mostrano alcune caratteristiche ricorrenti in aree della lingua definite. Per esempio, le traduzioni risultano innaturalmente ricche di pronomi personali e, al contempo, si registrano ripetizioni ingiustificate di nomi propri. Inoltre, si rileva una tendenza significativa a porre il soggetto in posizione preverbale, seguendo l'ordine sintattico obbligatorio in inglese, anche in casi in cui l'italiano avrebbe preferito una posizione posposta. Interessante, sempre in ambito pronominale, anche l'incertezza tra la serie della tradizione prescrittiva (*egli, ella, essi, esse*) e quella del cosiddetto italiano neo-standard (*lui, lei, loro*),⁴ con oscillazioni anche all'interno del medesimo testo o, addirittura, paragrafo.

Risultati simili ma altrettanto significativi provengono da una ricerca condotta dall'*École de Traduction et d'Interprétation* dell'Università di Ginevra su un corpus di testi tradotti da inglese, francese e tedesco (Salsnik 2007). Limitatamente alla tipologia testuale dei publiredazionali,⁵ sono emersi numerosi tratti che differenzierebbero i testi originali dall'italiano tradotto: tra questi, i già citati esplicitazione / ripetizione del soggetto e la sua collocazione in posizione preverbale, una maggior aderenza all'ordine canonico soggetto-verbo-oggetto (SVO), la posizione marcata (prenominale) dell'aggettivo qualificativo, problematiche legate alla struttura informativa della frase, nonché improprietà lessicali, ripetizioni ed errori grammaticali.

Sulla lingua dei giornali si concentrano anche Ondelli e Viale (2010; si veda anche Ondelli 2020), raccogliendo dagli archivi del *Corriere della Sera*, *la Repubblica* e *l'Unità* circa un migliaio di articoli tradotti da lingue straniere,⁶ per un totale di circa un milione di parole. Confrontando tramite analisi computazionale questo corpus con un corpus corrispondente creato *ad hoc* a partire da articoli scritti originariamente in italiano, è emerso che le traduzioni sono caratterizzate da una minore ricchezza lessicale e, di contro, da una maggiore presenza di termini appartenenti al cosiddetto vocabolario di base,⁷ cosa che sembrerebbe confermare una tendenza alla semplificazione. Uno spoglio "quali-quantitativo" (Ondelli 2020: 71) – che quindi combina osservazioni descrittive e interpretative con conteggi e misurazioni statistiche – ha fornito altri dati interessanti: per esempio, pare che i traduttori usino meno forestierismi dei giornalisti italiani (e, nel caso, preferiscano prestiti di necessità), tendano maggiormente alla scelta di varianti più ricercate (come *debbono* anziché *devono*) e al sovrasfruttamento del congiuntivo. Per

⁴ Per approfondire il trattamento dei pronomi personali in italiano standard *versus* italiano neo-standard, si vedano Sabatini 1985, Grandi 2019, Ballarè 2020.

⁵ Vengono così definiti da Salsnik (2007) testi sponsorizzati da enti pubblici e aziende private che vengono redatti presso società estere al fine di garantire la coerenza dell'immagine aziendale, dopodiché tradotti, spesso in maniera poco sorvegliata, e inseriti nella stampa italiana.

⁶ Le lingue di partenza sono varie; tra queste spiccano, per numero di articoli, il francese, lo spagnolo, il tedesco e soprattutto l'inglese.

⁷ Il *Vocabolario di base della lingua italiana* (VdB) è frutto di un'impresa condotta da Tullio De Mauro e collaboratori una prima volta alla fine degli anni '70 al fine di stilare la lista dei termini più usati dagli italofoni nella lingua quotidiana. Nel 2016 è stata pubblicata la sua versione aggiornata: il *Nuovo vocabolario di base della lingua italiana* (NVdB) comprende circa 2.000 parole 'fondamentali', 3.000 di 'alto uso' e 2.500 di 'alta disponibilità', per un totale di circa 7.500 parole. Il NVdB è consultabile al seguente link: <https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana>.

quanto riguarda la morfologia pronominale, di nuovo la frequenza assoluta dei pronomi soggetto esplicitati è maggiore nelle traduzioni.

Tra gli altri studi, si cita infine un'indagine esplorativa condotta da Gallitelli (2016). La studiosa ha messo a confronto romanzi scritti direttamente in lingua italiana con romanzi tradotti dall'inglese per tre diverse fasi cronologiche (1870-1910, 1930-1970, 1989-2011), con l'obiettivo di stabilire se tra la lingua dei romanzi originali italiani e quella delle traduzioni vi sia una differenza quantificabile e se questa abbia subito cambiamenti nel tempo. Se tra i due corpora Gallitelli non rileva significative differenze a livello di lunghezza media dei periodi e densità lessicale, l'interferenza dall'inglese viene invece confermata soprattutto per quanto riguarda l'aumento di perifrasi progressive nelle traduzioni. In generale, Gallitelli nota una certa coerenza interna al sottosistema delle traduzioni, che descrive come più conservatore rispetto ai testi autoctoni.

In sintesi, le ricerche condotte negli ultimi anni sembrano confermare la tesi di una lingua della traduzione con caratteristiche proprie: nello specifico un italiano delle traduzioni che si distingue per vari tratti dall'italiano scritto dell'uso spontaneo dei parlanti nativi. Oltre alla presenza di fenomeni riconducibili a tendenze universali, come la semplificazione o la maggiore esplicitatezza, l'italiano delle traduzioni sarebbe caratterizzato da una serie di tratti anomali, che acquistano però un senso se ricondotti al profilo tipologico della lingua di partenza del processo traduttivo, ossia, nella grande maggioranza dei casi, l'inglese.

3. Presentazione dello studio

L'obiettivo dello studio è quello di fornire un contributo nel campo circoscritto della letteratura commerciale tradotta dall'inglese, attraverso un'analisi che si sviluppi al contempo sui due piani possibili della comparazione:

- (i) confronto tra italiano originale e italiano delle traduzioni (analisi quantitativa);
- (ii) confronto tra la lingua di partenza, l'inglese, e la lingua di arrivo, l'italiano (analisi qualitativa).

Con questo scopo, si è costruito un corpus di narrativa diviso al suo interno in tre subcorpora. Ciascuno di essi contiene a sua volta dieci romanzi: il primo contiene dieci romanzi di autori italiani, scritti dunque fin dal principio in italiano; il secondo dieci romanzi scritti originariamente in inglese e in seguito tradotti in italiano; il terzo dieci romanzi tradotti in italiano a partire da lingue diverse dall'inglese. Quest'ultimo subcorpus è stato introdotto con funzione di controllo, al fine di stabilire se i fenomeni rilevati fossero l'effetto specifico dell'interferenza della lingua inglese o, al contrario, dovessero essere ricondotti a tendenze che prescindono dalle singole lingue coinvolte, come gli universali traduttivi. Si rimanda all'Appendice per un prospetto riassuntivo della costituzione del corpus con i relativi subcorpora.

La totalità dei romanzi analizzati appartiene agli anni '80-'90 del Novecento. Un ristretto gruppo di romanzi inglesi e di altre lingue (secondo e terzo subcorpus) risale nella sua prima stesura ad anni precedenti rispetto a questa fascia cronologica, e tuttavia le rispettive traduzioni in italiano rientrano nel periodo preso in considerazione. La scelta di questi due specifici decenni è da ricondursi alla volontà di indagare la (quasi) contemporaneità dell'italiano tradotto: le traduzioni realizzate in questo periodo riflettono verosimilmente uno stile traduttivo moderno. Inoltre, questa fascia cronologica si colloca

a una distanza temporale tale per cui l'italiano delle traduzioni potrebbe aver in qualche modo già influenzato l'italiano di oggi.

I tre subcorpora sono stati tra loro bilanciati in primo luogo a livello di numero complessivo di parole (circa un milione di token rispettivamente per il corpus dell'italiano originale e quello dell'italiano tradotto dall'inglese, poco meno per il corpus di controllo), per un totale complessivo di 2.841.998 parole. Quasi tutti i romanzi, inoltre, rientrano in una fascia di lunghezza compresa tra le 40.000 e le 200.000 parole (le eccezioni, solo due, sono dovute alla necessità di tener conto di altre esigenze di bilanciamento). In secondo luogo, si è considerato un ventaglio di generi narrativi, scegliendo di rappresentare in ciascun subcorpus romanzi d'avventura, psicologici, storici, thriller, gialli e fantastici.

Per quanto riguarda il terzo subcorpus, che funziona da corpus di controllo, si è cercato di riprodurre al suo interno una certa varietà genealogica (prevalentemente nell'ambito della famiglia indoeuropea, ma non solo) e soprattutto tipologica. A livello di tipologia morfologica, sono rappresentate perlopiù lingue fusive; a livello di tipologia sintattica, l'ordine dei costituenti della frase dichiarativa di gran lunga preponderante è SVO, con l'eccezione di due lingue classificate dal WALS (Dryer & Haspelmath 2013)⁸ come 'no dominant order' e una, il giapponese, SOV. La posizione reciproca tra aggettivo e testa nominale (rilevante per l'indagine riportata al par. 4.3) è varia: in tre casi l'aggettivo viene di norma posposto al nome, mentre in sei è tendenzialmente prenominal. Lo stesso dicasi per il carattere pro-drop (rilevante per l'indagine riportata al par. 4.1), che appartiene a cinque lingue su nove. Si rimanda alla tabella I in Appendice per le caratteristiche tipologiche specifiche delle lingue presenti nel subcorpus.

Nonostante questi sforzi, si è consapevoli che un corpus di soli trenta romanzi non sarà probabilmente esente da distorsioni, che si è tuttavia cercato di ridurre al minimo; per la stessa ragione, non si pretende che i risultati ottenuti vengano considerati esaustivi o definitivi, quanto piuttosto che possano fornire un apporto nella conferma o smentita di alcune questioni dibattute nella letteratura o costituire un punto di partenza per nuove analisi più approfondite.

Quanto all'organizzazione interna, lo studio si compone di due parti, le quali rispecchiano i due piani di comparazione cui si accennava sopra. Il fine principale rimane, in ambo le parti, quello di verificare l'entità dell'interferenza linguistica dall'inglese.

La prima parte (par. 4) è dedicata al confronto tra l'italiano originale e l'italiano della traduzione, e si basa su un'analisi principalmente quantitativa. Sono stati infatti estratti e confrontati dati numerici riguardanti i tre subcorpora, mettendo in luce eventuali peculiarità della lingua della traduzione. Dove non indicato diversamente, si è utilizzato il software di consultazione e analisi di corpora AntConc a cui si sono affiancate, laddove necessario, verifiche manuali.

La seconda parte (par. 5), dedicata al confronto tra l'originale inglese e la sua traduzione italiana, si compone di una serie di analisi qualitative di alcuni brani estratti dai romanzi. Esse permettono di indagare sul piano delle parole e delle costruzioni linguistiche, e non solo su quello dei numeri, il fenomeno in esame, consentendo inoltre di ipotizzare le motivazioni alla base delle singole scelte del traduttore.

Il fenomeno dell'interferenza è stato indagato, in entrambe le parti, principalmente attraverso l'analisi di quattro tratti linguistici per i quali il profilo tipologico dell'italiano si discosta da quello dell'inglese, come riassunto in Tab. 1:

⁸ The World Atlas of Language Structures: <https://wals.info/>.

Tratto	Inglese	Italiano
Esplicitazione del pronome soggetto	Obbligatoria	Pro-drop
Posizione dell'aggettivo qualificativo rispetto al nome	AN es. <i>a black cat</i>	NA (/AN) es. <i>un gatto nero</i>
Posizione del soggetto rispetto al verbo inaccusativo	SV es. <i>something happened</i>	VS (/SV) es. <i>è successo qualcosa</i>
Uso della perifrasi progressiva per l'espressione dell'aspetto progressivo	Obbligatorio es. <i>I was reading</i>	Opzionale es. <i>leggevo / stavo leggendo</i>

Tabella 1. Tratti tipologici oggetto dello studio.

4. Risultati dell'analisi quantitativa

4.1 Esplicitazione del pronome soggetto

Una delle caratteristiche tipiche dell'italiano tradotto dall'inglese concerne il sistema pronominale: si è da più parti notato (Cardinaletti 2004, 2005; Giusti 2004; Ondelli & Viale 2010), infatti, che i traduttori, a fronte del modello inglese, tendono a esplicitare i pronomi personali in posizione di soggetto anche laddove l'italiano nativo li ometterebbe (tratto *pro-drop*, Dryer 2013a).

La Tab. 2 riporta i risultati della ricerca delle occorrenze dei pronomi soggetto all'interno dei tre subcorpora. Si è reso necessario uno spoglio manuale dell'esito dell'estrazione automatica dei dati al fine di eliminare i casi in cui il pronome rappresentava un costituente sintattico diverso dal soggetto. I risultati dell'indagine vengono presentati anche sotto forma di grafico (Fig. 1), al fine di visualizzare in maniera più immediata gli andamenti dei dati per ciascun corpus.

Appare anzitutto evidente la netta separazione tra le due serie di pronomi: i pronomi della tradizione prescrittiva e letteraria (*egli, ella, esso*, ecc.) compaiono, infatti, in ciascuno dei corpora, in numero molto inferiore rispetto alla serie neo-standard (*lui, lei, loro*) (per il passaggio da una serie all'altra come tratto dell'italiano neo-standard: Sabatini 1985, Ballarè 2020). *Egli* è, tra i pronomi standard, quello che “sopravvive” in misura leggermente maggiore, e tuttavia non regge comunque il confronto con il corrispondente neo-standard (*lui*), presente nei diversi corpora in misura da cinque a sei volte maggiore.

Un'altra differenza tra la serie standard e quella neo-standard, tuttavia, risulta più interessante ai fini di questo contributo. Se, infatti, tra i pronomi standard non si registrano scarti rilevanti tra i diversi subcorpora (così come in Gallitelli 2016, che tuttavia prende in analisi solo il pronome *egli*), i pronomi neo-standard, al contrario, mostrano un comportamento che sembra avvalorare l'ipotesi dell'interferenza della lingua fonte nella direzione di una maggiore esplicitazione dei pronomi soggetto. Nello specifico, il corpus di testi italiani mostra un numero notevolmente inferiore di occorrenze per ciascun pronome neo-standard rispetto agli altri due corpora. Il corpus di traduzioni dall'inglese,

di contro, registra le più alte frequenze sia di *lui* (+39,7% rispetto all'italiano originale) che di *lei* (+87,24% rispetto all'italiano originale).

	Originali italiani	Traduzioni dall'inglese	Traduzioni da lingue varie
<i>egli</i>	2,67 (268)	2,68 (277)	2,46 (199)
<i>ella</i>	0,07 (7)	0,12 (12)	0,26 (21)
<i>esso</i>	0,46 (46)	0,28 (29)	0,32 (26)
<i>essa</i>	0,73 (73)	0,30 (31)	0,63 (51)
<i>essi</i>	1,02 (102)	0,90 (93)	0,83 (67)
<i>esse</i>	0,19 (19)	0,09 (10)	0,22 (18)
<i>lui</i>	12,60 (1263)	17,60 (1813)	13,36 (1081)
<i>lei</i>	11,60 (1163)	21,72 (2240)	20,14 (1629)
<i>loro</i>	7,64 (765)	8,25 (851)	11,3 (914)

Tabella 2. Occorrenze dei pronomi soggetto in un corpus di romanzi italiani originali e romanzi tradotti. I valori sono normalizzati e forniti su 10.000 parole; tra parentesi si riportano i valori assoluti delle occorrenze.

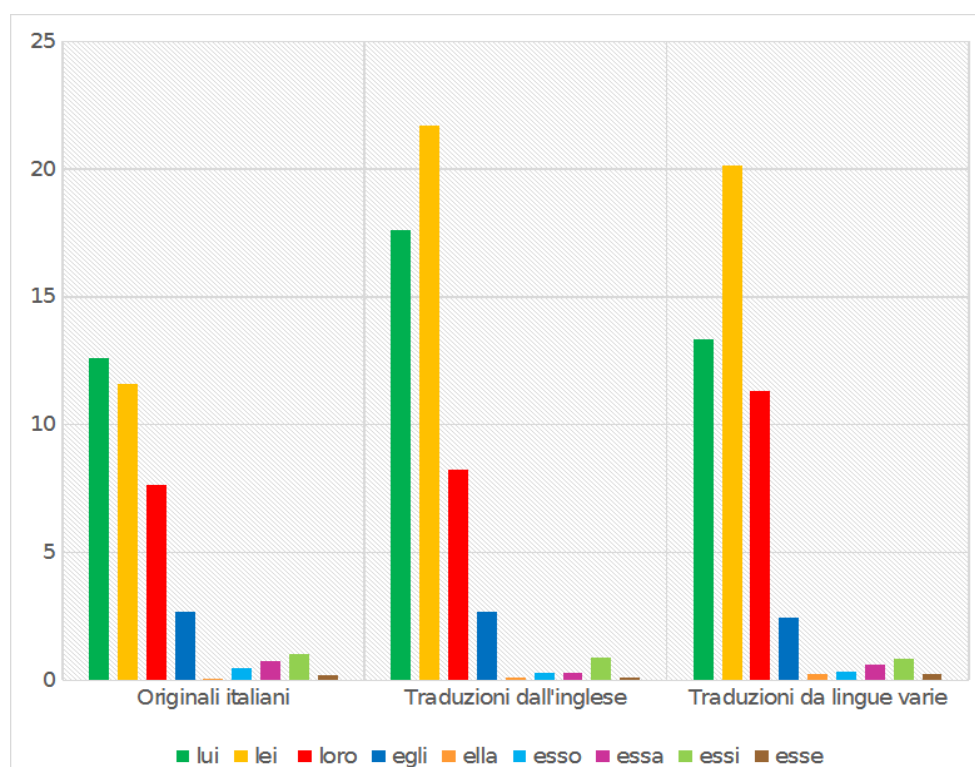


Figura 1. Occorrenze dei pronomi soggetto in un corpus di romanzi italiani originali e romanzi italiani tradotti. I valori sono normalizzati e forniti su 10.000 parole.

Relativamente alla serie di pronomi neo-standard, per i quali, come si è appena detto, i valori divergono tra i corpora in maniera visibilmente considerevole, si è eseguito il test del chi-quadrato di indipendenza per verificare la correlazione tra le variabili ed escludere che la distribuzione dei valori ottenuta sia casuale. Dall'esito del test si evince che le variabili sono dipendenti, ossia vi è evidenza statistica della dipendenza tra l'uso dei pronomi e il fatto che il romanzo sia stato scritto originariamente in italiano o, al contrario, provenga da una traduzione dall'inglese, χ^2 (2, N = 8095) = 85.09, $p < .001$ (livello di significatività: $\alpha = 0.05$).

4.2 Posizione del soggetto con verbi inaccusativi

I verbi intransitivi inaccusativi in italiano tendono a preferire il soggetto posposto (es. *È arrivato il treno; È sparito un biscotto*; Bentley 2006; Ježek 2003, 2010); in inglese, al contrario, l'ordine SV è rispettato – quasi senza eccezioni – con ogni tipo di verbo, compresi gli inaccusativi. Tra gli altri, Garzone (2005) e Giusti (2004) hanno mostrato come in alcuni casi il traduttore si lasci influenzare dalla struttura sintattica della frase di partenza e riproduca l'ordine inglese, collocando il soggetto prima del verbo inaccusativo.

Con l'obiettivo di esplorare tale fenomeno, si è scelto di verificare nei subcorpora la posizione del soggetto relativamente a un insieme limitato dei principali verbi inaccusativi in italiano. La selezione tiene conto della classificazione proposta da Ježek (2003; 2010), secondo cui si possono riconoscere sette tipi di verbi inaccusativi rispetto a due parametri, quali il grado di inaccusatività e la presenza del clitico *-si*. Si è scelto nella fattispecie di verificare il comportamento del soggetto con alcuni verbi del primo gruppo, gli inaccusativi semplici: si è ritenuto che questo gruppo costituisse il “prototipo” dei verbi intransitivi inaccusativi. La Tab. 3 riassume i risultati dello spoglio automatico.

	Originali italiano		Traduzioni dall'inglese		Traduzioni da lingue varie	
	SV	VS	SV	VS	SV	VS
<i>arrivare</i>	1,05 (105)	0,72 (72)	0,78 (81)	0,58 (60)	0,73 (59)	0,96 (78)
<i>scendere</i>	0,51 (51)	0,13 (13)	0,46 (48)	0,13 (14)	0,39 (32)	0,12 (10)
<i>cadere</i>	0,50 (50)	0,17 (17)	0,53 (55)	0,16 (17)	0,68 (55)	0,16 (13)
<i>morire</i>	1,04 (104)	0,22 (22)	1,05 (109)	0,15 (16)	0,89 (72)	0,22 (18)
<i>scoppiare</i>	0,08 (8)	0,15 (15)	0,18 (19)	0,15 (16)	0,14 (11)	0,14 (11)
<i>sparire</i>	0,14 (14)	0,03 (3)	0,25 (26)	0,02 (2)	0,28 (23)	0,09 (7)

Tabella 3. Posizione del soggetto con una selezione di verbi inaccusativi in un corpus di romanzi italiani originali e romanzi tradotti. I valori sono normalizzati e forniti su 10.000 parole; tra parentesi si riportano i valori assoluti delle occorrenze.

Anzitutto, si sarà subito notato che i sintagmi SV sono, per quasi ciascun verbo e per tutti e tre i subcorpora, più frequenti dei corrispondenti VS. Ciò si può spiegare con il fatto che tipicamente un soggetto con un alto grado di agentività e animatezza, così come un

soggetto precedentemente introdotto nel discorso, tenderà con più probabilità a occupare la posizione preverbale anche con verbi inaccusativi. Inoltre, la prima posizione viene preferita anche quando il soggetto, all'interno di un periodo complesso, è soggetto anche di una proposizione coordinata o subordinata alla principale. Per esempio, numerosissimi in tutti i corpora sono i casi come quelli in (1):

- (1) a. *Marianne è arrivata con il pane tagliato a fette in un vassoio di vimini chiaro.*
 b. *Bartleboom scese dal letto e senza una parola prese il ragazzino per un braccio.*
 c. *Enrico VIII è morto lasciando dietro di sé uno stuolo di figli che si contendono la corona.*

(a. *Uto*, A. De Carlo, 1995; b. *Oceano mare*, A. Baricco, 1993; c. *Q*, L. Blisset, 1999)

In casi come questi il soggetto preverbale è non solo accettato, ma preferibile e più naturale. Ciò spiega il fatto che, nel complesso, i sintagmi SV sono comunque in numero maggiore rispetto ai sintagmi VS.

Chiarito questo punto, si osserva che i dati risultano piuttosto uniformi tra i tre subcorpora. Per quanto riguarda nello specifico l'ipotesi dell'interferenza della struttura inglese, i risultati non mostrano scarti particolarmente significativi tra l'occorrenza di sintagmi SV con verbi inaccusativi nelle traduzioni dall'inglese rispetto all'occorrenza degli stessi nei testi originali italiani. Solo *scoppiare* registra un numero più che doppio di sintagmi SV nelle traduzioni dall'inglese rispetto agli originali italiani, e *sparire* un buon +78,5%. Con *cadere* e *morire* lo scarto è nettamente inferiore.

È probabile, dunque, che un'indagine puramente quantitativa non riesca a cogliere appieno l'entità del fenomeno in indagine, sia a causa della scarsa presenza degli specifici verbi nel corpus, sia – soprattutto – della naturale variabilità della posizione del soggetto in italiano.

4.3 Posizione dell'aggettivo qualificativo rispetto al nome

Notoriamente, l'ordine non marcato della lingua inglese prevede che all'interno dei sintagmi nominali l'aggettivo preceda sempre il nome (Dryer 2013b). La situazione dell'italiano è più complessa, in quanto, nonostante l'ordine di base sia NA (nome-aggettivo), la collocazione preominale non è sempre preclusa all'aggettivo.

A questo proposito, si prende qui a riferimento la classificazione proposta da Cerruti (2015), che si basa sull'intersezione di due proprietà, una legata al comportamento sintattico dell'aggettivo (\pm PREDICATIVO) e l'altra alla sua natura semantica (\pm INTERSETTIVO)⁹. Combinando le due proprietà si ottengono quattro classi, per ciascuna delle quali è possibile riconoscere una correlazione tendenziale con la posizione pre- o postnominale. Si veda la seguente tabella:

⁹ La prima proprietà (\pm PREDICATIVO) riguarda l'accettabilità dell'aggettivo in posizione postcopulare, cioè dopo il verbo *essere* e altri verbi copulativi; la seconda proprietà (\pm INTERSETTIVO) concerne il tipo di significato dell'aggettivo: un aggettivo ha significato intersettivo "quando la denotazione del sintagma nominale di cui fa parte è l'intersezione dell'insieme denotato dal nome e dell'insieme denotato dall'aggettivo" (Cerruti 2015: 402).

Classe	Proprietà	Posizione
I classe	[+PREDICATIVO] [+INTERSETTIVO]	NA es. <i>pittore francese</i>
II classe	[+PREDICATIVO] [−INTERSETTIVO]	NA o AN (alcuni solo NA) es. <i>formica enorme / enorme formica</i>
III classe	[−PREDICATIVO] [+INTERSETTIVO]	NA es. <i>giornalista sportivo</i>
IV classe	[−PREDICATIVO] [−INTERSETTIVO]	NA o AN (alcuni solo AN) es. <i>motivo principale / principale motivo</i>

Tabella 4. Classi di aggettivi qualificativi in italiano (adattata da Cerruti 2015).

Con l'obiettivo di indagare la frequenza dei sintagmi NA e AN all'interno del corpus in esame, si è sottoposto l'insieme dei testi a una procedura automatica di POS-tagging tramite il software TagAnt. I risultati della ricerca sono riportati in Tab. 5:

	Originali italiani	Traduzioni dall'inglese	Traduzioni da lingue varie
NA	112,17	117,99	123,46
AN	75,76	78,80	83,95

Tabella 5. Posizione dell'aggettivo qualificativo rispetto al nome in un corpus di romanzi italiani originali e romanzi tradotti. I valori sono normalizzati e forniti su 10.000 parole.

Come si può osservare, i risultati indicano, inaspettatamente, una maggiore presenza complessiva di aggettivi nelle traduzioni, sia dall'inglese, che, ancor di più, da lingue varie.

Una possibile spiegazione di questo dato inatteso potrebbe fare riferimento all'universale dell'esplicitazione: tra le sue manifestazioni, infatti, rientra anche l'aggiunta di aggettivi, insieme ad avverbi e congiunzioni (Blum-Kulka, 1986). Anche la volontà di rendere il testo in qualche modo più semplice e fruibile potrebbe in alcuni casi portare il traduttore a sostituire un iponimo con un suo iperonimo, magari reso più specifico tramite un aggettivo. O ancora, la maggiore varietà lessicale del corpus degli originali italiani (si veda il par. 4.5.2) potrebbe significare un uso più vario di nomi, laddove la traduzione potrebbe tendere all'impiego di termini più comuni seguiti da un aggettivo, sia per interferenza lessicale dalla lingua di partenza, sia per effetto

dell'universale della semplificazione. Si è provato, per esempio, a verificare le occorrenze del sintagma *bambino/i piccolo/i*. Esso compare solo sette volte nel corpus dei romanzi italiani e ben undici nel corpus di traduzioni dall'inglese. Nel corpus di romanzi italiani, tuttavia, compaiono anche *neonato/i* (quindici volte), *infante/i* (dieci volte), *bebè* (cinque volte), *moccioso/i* (cinque volte), *pargolo/i* (quattro volte), *marmocchio/i* (tre volte), *piccino/i* (due volte), *lattanti* (una volta), *pupo* (una volta). Di questi, nel corpus di traduzioni dall'inglese compaiono solo *neonato/i* (quattordici volte), *bebè* (tre volte) e *marmocchi* (una volta).

Tornando all'ipotesi dell'interferenza, è vero che le traduzioni risultano quantitativamente più ricche di sintagmi del tipo AN rispetto all'italiano, e tuttavia, come si è appena osservato, questo incremento rientra in un più generale incremento di sintagmi contenenti aggettivi. Non vi sono le basi, dunque, per affermare che i dati raccolti supportino l'ipotesi di partenza. Come già osservato con il caso della posizione del soggetto rispetto a un verbo inaccusativo, d'altronde, il comportamento "naturale" dell'italiano (in questo caso nella collocazione pre- o postnominale dell'aggettivo) è talmente vario che non stupisce che un'indagine quantitativa come quella appena condotta non riesca a individuare con certezza un'influenza della lingua da cui si traduce.

Nondimeno, si propongono di seguito i risultati di una seconda ricerca, limitata a tre specifiche coppie di nomi e aggettivi: *voce alta / alta voce*, *capelli lunghi / lunghi capelli*, e *occhi grandi / grandi occhi*. Come si vede, gli aggettivi selezionati risultano accettabili in italiano sia in posizione prenominali che postnominale: non a caso, essi rientrano nella II classe dello schema di Cerruti (2015) riportato sopra (Tab. 4).

Di seguito (Tab. 6 e Fig. 2, 3 e 4) si riportano i risultati di questo secondo spoglio.

	Originali italiani		Traduzioni dall'inglese		Traduzioni da lingue varie	
	NA	AN	NA	AN	NA	AN
<i>voce – alta</i>	0,15 (15)	0,19 (19)	0,19 (20)	0,36 (37)	0,24 (20)	0,27 (22)
<i>capelli – lunghi</i>	0,15 (15)	0,05 (5)	0,07 (7)	0,28 (29)	0,07 (6)	0,09 (7)
<i>occhi – grandi</i>	0,06 (6)	0,04 (4)	0,00 (0)	0,13 (13)	0,01 (1)	0,17 (14)

Tabella 6. Posizione dell'aggettivo qualificativo rispetto al nome in un corpus di romanzi italiani originali e romanzi tradotti: casi specifici. I valori sono normalizzati e forniti su 10.000 parole; tra parentesi si riportano i valori assoluti delle occorrenze.

A partire da questi dati si nota che il corpus di romanzi italiani privilegia, in due casi su tre, la sequenza NA; per l'unico caso in cui ad essere preferita è la sequenza opposta (*alta voce*)¹⁰, lo scarto è del +26 % rispetto a *voce alta*. Se si guarda al corpus di traduzioni dall'inglese, invece, la sequenza AN è nettamente preferita per ciascuna delle diadi (*alta voce* compare quasi il doppio delle volte rispetto a *voce alta*; *lunghi capelli* compare il

¹⁰ Si potrebbe notare a questo proposito che le costruzioni AN qui citate sono forse più frequenti in espressioni polirematiche. Soprattutto nel caso di *ad alta voce*, potrebbe entrare in gioco una componente di fissità (si pensi al test di modificabilità: *a voce molto alta* contro **a molto alta voce*) che rende l'espressione probabilmente più ricorrente di un semplice costrutto AN. Nonostante ciò, si ritiene che il valore del dato comparativo tra il corpus di romanzi originali e i corpora di traduzioni rimanga significativo.

quadruplo delle volte rispetto a *capelli lunghi*; e addirittura *occhi grandi* non compare mai). Il corpus di traduzioni varie registra anch'esso una predominanza della sequenza AN, anche se con scarto minore rispetto a NA.

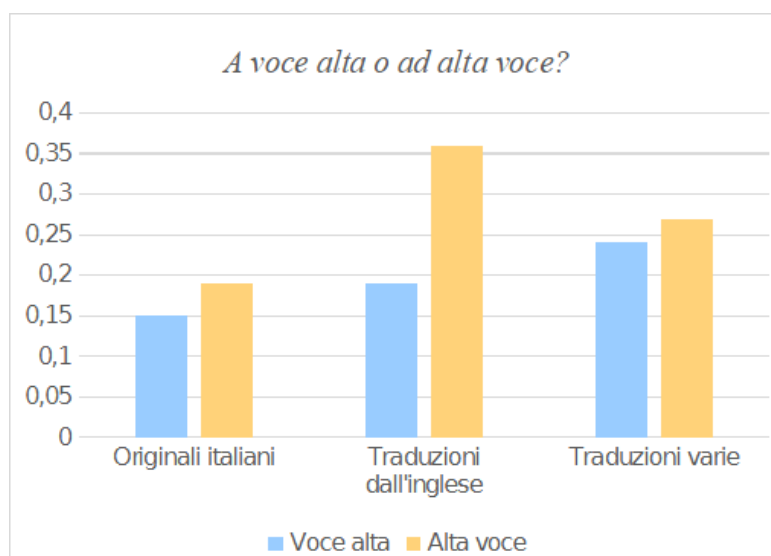


Figura 2. Posizione dell'aggettivo qualificativo rispetto al nome in un corpus di romanzi italiani originali e romanzi tradotti. *A voce alta o ad alta voce?*

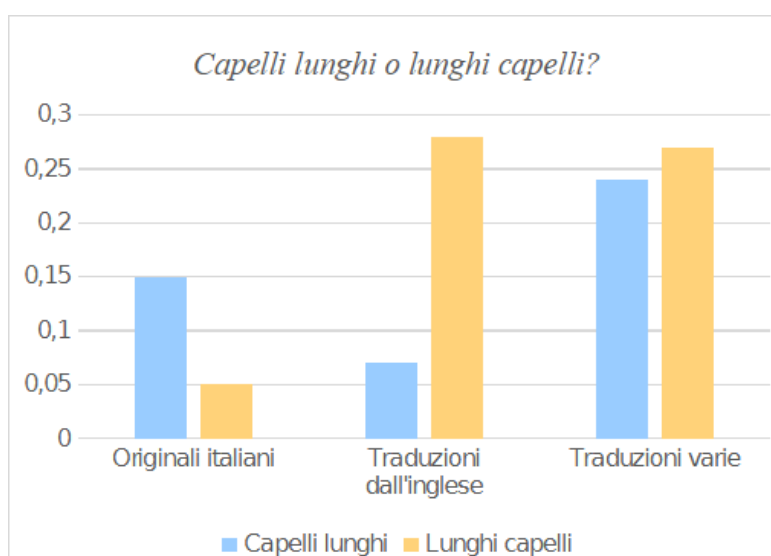


Figura 3. Posizione dell'aggettivo qualificativo rispetto al nome in un corpus di romanzi italiani originali e romanzi tradotti. *Capelli lunghi o lunghi capelli?*

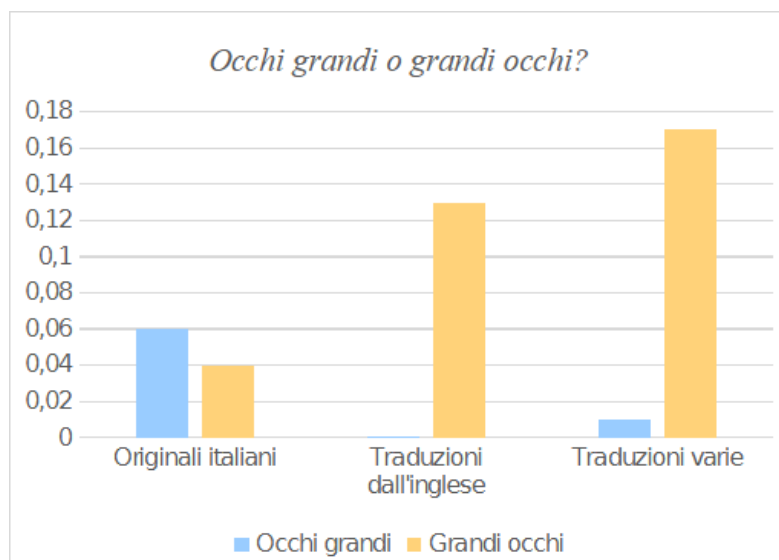


Figura 4. Posizione dell'aggettivo qualificativo rispetto al nome in un corpus di romanzi italiani originali e romanzi tradotti. *Occhi grandi o grandi occhi?*

Considerando queste coppie lessicali specifiche, dunque, è evidente come nel complesso le traduzioni dall'inglese tendano a selezionare l'ordine AN molto più di quanto non faccia l'italiano originale. Sommando le occorrenze relative alle tre coppie per ciascuno dei due corpora, si osserva che i romanzi originali italiani usano costruzioni NA per 36 volte e costruzioni AN per 28 volte, mentre le traduzioni dall'inglese usano costruzioni NA per 27 volte e costruzioni AN per 79 volte. Il test esatto di Fisher conferma che, nonostante la differenza in termini di numero assoluto di occorrenze, la distribuzione non è casuale, bensì statisticamente significativa: il valore che si ottiene applicando la formula del test di Fisher è infatti inferiore allo 0.001 (livello di significatività: $\alpha = 0.05$).

Diversamente dai dati complessivi (Tab. 5), questa ricerca specifica sembra insomma dare credito all'ipotesi di una effettiva interferenza del modello inglese (o di altre lingue AN) laddove il traduttore si trova davanti a una scelta più o meno equivalente.

4.4 Perifrasi progressiva: stare + gerundio

Per quanto riguarda la perifrasi progressiva, la letteratura (per esempio Durante 1981; Bertinetto *et al.* 2000; Degano 2005) è concorde nell'affermare che il modello costituito dalla forma inglese (del tipo *I am reading*) abbia avuto un ruolo chiave, a partire dalla seconda metà del Novecento, nell'aumentare la diffusione della perifrasi progressiva italiana (del tipo *sto leggendo*) e accelerare la sua naturale evoluzione in termini di accentuazione del carattere aspettuale imperfettivo (si veda in particolare la Tab. 3 in Bertinetto *et al.* 2000: 540).

Una conferma di ciò è data dalla presenza tendenzialmente maggiore di perifrasi progressive nei testi tradotti, dovuta alla predisposizione del traduttore a riprodurre il carattere perifrastico della struttura inglese, piuttosto che affidarsi a tempi dell'indicativo come il presente o l'imperfetto, che pure in italiano possono esprimere il medesimo significato aspettuale (es. *Quando sono arrivata, Lorenzo dormiva*).

Nella Tab. 7 si riportano i risultati della ricerca di perifrasi progressive nei subcorpora.

	Originali italiani	Traduzioni dall'inglese	Traduzioni da lingue varie
Perifrasi progressive (<i>stare</i> + gerundio)	13,76 (1379)	22,44 (2315)	14,04 (1163)

Tabella 7. Incidenza della perifrasi progressiva in un corpus di romanzi italiani originali e romanzi tradotti. I valori sono normalizzati e forniti su 10.000 parole; tra parentesi si riportano i valori assoluti delle occorrenze.

La presenza nettamente maggiore di perifrasi progressive nel corpus delle traduzioni dall'inglese richiama subito l'attenzione. Se si considera, inoltre, che il corpus di traduzioni da lingue varie registra un dato pressoché simile a quello degli originali italiani, il fatto che i romanzi tradotti dall'inglese presentino un numero di perifrasi progressive più alto del 63% rispetto a quello dell'italiano sembra del tutto avvalorare la tesi dell'interferenza dalla lingua inglese.

Di seguito (Tab. 8) si forniscono anche le collocazioni più frequenti per ciascun subcorpus (la frequenza è data nella colonna *Freq*; *Rank* indica la posizione in 'classifica', e *Range* il numero di romanzi, su dieci, in cui la collocazione è presente).

Originali italiani					Traduzioni dall'inglese					Traduzioni da lingue varie				
	Cluster	Rank	Freq	Range		Cluster	Rank	Freq	Range		Cluster	Rank	Freq	Range
1	sta facendo	1	17	5	1	stava dicendo	1	49	8	1	stava succedendo	1	17	4
2	sta succedendo	2	14	5	2	stava facendo	2	32	9	2	sta succedendo	2	16	5
3	sto facendo	2	14	6	3	sta facendo	3	28	7	3	stava facendo	2	16	8
4	sto parlando	2	14	5	4	stava parlando	4	27	6	4	stava accadendo	4	15	8
5	stava facendo	5	11	6	5	stava cercando	5	25	8	5	stai facendo	5	10	5
6	sto cercando	5	11	5	6	stava guardando	5	25	9	6	stava parlando	5	10	7
7	sta cercando	7	10	4	7	sta succedendo	7	24	8	7	stava andando	7	9	5
8	stava diventando	8	9	5	8	stai parlando	7	24	7	8	stava dicendo	7	9	5
9	stava cercando	9	8	4	9	stai facendo	9	21	5	9	stava diventando	9	8	4
10	stavo cercando	9	8	4	10	sta cercando	10	18	6	10	stava ascoltando	10	7	2
11	sta aspettando	11	7	3	11	stava aspettando	10	18	4	11	stava cercando	10	7	3
12	sta prendendo	11	7	4	12	sto cercando	10	18	6	12	stesse facendo	10	7	5
13	stai dicendo	13	6	4	13	stai dicendo	13	16	4	13	sta facendo	13	6	3
14	stai parlando	13	6	3	14	stava osservando	13	16	6	14	stava aspettando	13	6	4
15	stanno facendo	13	6	5	15	stava pensando	13	16	6	15	stava morendo	13	6	3

Tabella 8. Perifrasi progressiva. Collocazioni più frequenti in un corpus di romanzi italiani originali e romanzi tradotti.

Interessante, per esempio, notare che *stava dicendo*, al primo posto nel corpus di traduzioni dall'inglese con ben 49 occorrenze, è molto meno frequente negli altri due corpora. Nel corpus di romanzi originali italiani, addirittura, non rientra nelle prime quindici collocazioni qui mostrate, registrando solo 4 occorrenze; di contro, una veloce

ricerca mostra invece che *diceva* compare 347 volte. Sebbene da queste ultime sarebbero da eliminare, a rigore, tutte le occorrenze in cui *diceva* non ha valore aspettuale progressivo, rimane un dato interessante se si pensa che il corpus di traduzioni dall'inglese – che pure è leggermente più consistente a livello di token – contiene 220 occorrenze della stessa voce verbale. Dunque, semplificando, si potrebbe dire che nei romanzi italiani si usa di più *diceva* e meno *stava dicendo*, mentre la seconda opzione è molto più selezionata nelle traduzioni dall'inglese. Il corpus di controllo si colloca in posizione intermedia tra i due, con 9 occorrenze di *stava dicendo* e 228 di *diceva*.

Si può dunque ragionevolmente affermare, sulla base dei dati ottenuti, che l'influenza del modello inglese nel caso della perifrasi progressiva operi in maniera piuttosto consistente.

L'esito dell'indagine qui condotta è coerente con quello ottenuto da Gallitelli (2016), la quale rileva nel suo corpus un uso molto più frequente della perifrasi nelle traduzioni dall'inglese rispetto ai romanzi originali italiani, oltre che un suo notevole aumento nel tempo (con un dato di occorrenze nel periodo 1989-2011 di ben sedici volte maggiore rispetto a quello relativo agli anni 1870-1910!).

Una delle ragioni del successo della perifrasi progressiva nelle traduzioni (e poi, di conseguenza, nella lingua d'uso) può essere individuata proprio nel fatto che, come ricorda Cortelazzo (2007), la perifrasi progressiva è *già* un costrutto ampiamente diffuso e in diffusione in italiano, a prescindere da influenze esterne. Dunque, diversamente dagli altri tratti qui indagati, in questo caso l'interferenza dell'inglese non va in una direzione “anomala” per l'italiano, scontrandosi con quella che è la sua convenzione d'uso, bensì finisce per accelerare una tendenza già presente nella lingua.

4.5 Altre misurazioni

4.5.1 Lunghezza media e articolazione interna dei periodi

Si propone, inoltre, sulla scia di Gallitelli (2016), un insieme di dati riguardanti in generale la composizione sintattica dei periodi e la loro lunghezza.

Le medie sono state ottenute analizzando tramite il software Read-It (Dell'Orletta *et al.* 2011)¹¹ cinque brani composti ciascuno da dieci periodi estratti casualmente da ogni romanzo per ciascun subcorpus. La Tab. 9 riporta una sintesi dei risultati ottenuti (sono indicati i valori in media).

I risultati mostrano anzitutto uno scarto interessante tra il corpus di traduzioni dall'inglese e gli altri due corpora relativamente al numero di parole per periodo: in particolare, i periodi dei romanzi italiani (22,93 parole in media) sono più lunghi del 23,2% rispetto a quelli delle traduzioni dall'inglese (18,61 parole in media). Il corpus di controllo, invece, registra un dato affine (minore del 3,4%) a quello degli originali italiani. Inoltre, si nota che il numero medio di parole per proposizione è più basso nei testi italiani originali (anche se solo dell'1,9% rispetto alle traduzioni dall'inglese), e tuttavia l'italiano originale esibisce il valore più alto relativamente alla media di proposizioni per periodo (+29% rispetto a traduzioni dall'inglese, +10,45% rispetto a traduzioni da lingue varie).

¹¹ Read-It è uno strumento di valutazione della leggibilità di testi italiani sviluppato dall'ItaliaNLP Lab dell'Istituto di Linguistica Computazionale ‘Antonio Zampolli’ (ILC-CNR). Sito ufficiale dell'Italian Natural Processing Lab (ItaliaNPL Lab): <http://www.italianlp.it/>.

	Originali italiani	Traduzioni dall'inglese	Traduzioni da lingue varie
Parole per periodo	22,93	18,61	22,18
Parole per proposizione	6,90	7,03	7,41
Proposizioni per periodo	3,38	2,62	3,06
Proposizioni principali o coordinate	64,32%	66,87%	63,19%
Proposizioni subordinate	35,68%	33,13%	36,81%
Indice Gulpease	56,36	59,40	55,52

Tabella 9. Lunghezza media e articolazione interna dei periodi in un corpus di romanzi italiani originali e romanzi tradotti. Per 'periodo' si intende l'enunciato compreso tra due punti fermi; per 'proposizione' l'unità frasale minore di senso compiuto costruita intorno ad un predicato (comprendente anche un soggetto ed eventuali complementi).

Provando a tradurre i dati numerici in parole, si può supporre che l'autore italiano tenda a comporre in media periodi più lunghi, articolati al suo interno in un numero maggiore di proposizioni; di contro, la traduzione dall'inglese sarebbe caratterizzata da periodi complessivamente più corti, composti da meno proposizioni seppur queste contino in media leggermente più parole. In altri termini, sembrerebbe che il corpus di traduzioni dall'inglese lasci trasparire quantitativamente un'influenza della lingua fonte a livello di composizione del periodo, nella direzione di periodi più corti e meno articolati.

Le penultime due righe della tabella portano ulteriore conferma a questa interpretazione. In media, infatti, le traduzioni dall'inglese registrano la minore percentuale di frasi subordinate, mentre il dato relativo alle frasi principali o coordinate è il più alto. Il corpus di originali italiani e quello di traduzioni da lingue varie mostrano invece risultati più simili tra loro.

L'ultima riga della Tab. 9 riporta l'indice di leggibilità Gulpease, elaborato alla fine degli anni '80 dal Gruppo universitario linguistico e pedagogico (GULP) dell'Istituto di Filosofia dell'Università degli studi di Roma La Sapienza. Al di là della sua funzione specifica, ossia quella di valutare la leggibilità di un testo in riferimento al livello di scolarizzazione del lettore,¹² si ritiene che i valori ottenuti siano rilevanti soprattutto se considerati in relazione l'uno con l'altro, più che nei loro valori assoluti. Coerentemente con i risultati precedentemente esposti, il corpus di traduzioni dall'inglese registra l'indice Gulpease più alto e dunque una leggibilità maggiore. Nello specifico, è più alto di più di tre punti rispetto agli indici degli altri due corpora, che sono tra loro più simili.

¹² Se l'indice risulta superiore a 80 significa che la lettura è facile per un lettore con licenza elementare, se superiore a 60 lo è per un lettore con licenza media, se superiore a 40 lo è per un lettore con diploma di scuola superiore.

4.5.2 Indice di varietà lessicale

Il rapporto type/token (*type/token ratio*, TTR) è un indice che misura la varietà delle parole contenute in un testo, dividendo il numero di parole diverse utilizzate in un testo (type) per il numero totale di parole (token). Un valore TTR elevato indica una maggiore varietà di vocaboli, mentre un valore basso suggerisce ripetizioni frequenti. I valori possono variare da 0, che indica la ripetizione continua di un singolo type, a 1, che rappresenta un testo composto esclusivamente da *hapax legomena*, ovvero parole che ricorrono una volta soltanto (Lenci *et al.* 2005). Spesso il TTR viene espresso tramite una percentuale, che si ottiene moltiplicando il valore per 100, come si è scelto di fare in questo caso (il valore minimo sarà dunque di 0%, il massimo di 100%).

La Tab. 10 riporta gli indici TTR dei subcorpora qui in analisi. Si presti attenzione al fatto che non si stanno mettendo a confronto due lingue diverse, inglese e italiano, la cui differenza in termini di TTR potrebbe essere dovuta a diversità strutturali; al contrario, la comparazione avviene tra due corpora di una stessa lingua (italiano), con l'unica differenza che uno dei due è composto da testi tradotti.

	Originali italiani	Traduzioni dall'inglese	Traduzioni da lingue varie
Numero di type	61.403	55.224	48.563
Numero di token	1.001.816	1.031.365	808.817
TTR	6,13%	5,35%	6,00%

Tabella 10. Rapporto type / token in un corpus di romanzi italiani originali e romanzi tradotti.

Lo scarto percentuale che viene rilevato diventa, dunque, interessante: si potrebbe ipotizzare che esso derivi, oltre che da tendenze universali della traduzione (come suggeriscono Ondelli e Viale a fronte di un risultato simile nel loro studio del 2010), da un processo traduttivo che tende a riprodurre in maniera abbastanza fedele la struttura sintattica della frase inglese (per esempio esplicitando i pronomi soggetto, utilizzando più frasi coordinate e dunque congiunzioni di un certo tipo, ecc.) o alcune scelte lessicali tipiche o molto frequenti in inglese in determinati registri (per esempio, al par. 4.3 si cita il fatto che una veloce ricerca ha rilevato una presenza massiccia dell'espressione *bambino piccolo* nelle traduzioni dall'inglese, a fronte di un uso di termini molto più vario nei romanzi originali italiani). Resta tuttavia una supposizione che andrebbe corroborata con altri studi, magari specifici sulla misurazione della densità e varietà lessicale in italiano, inglese e nelle traduzioni.

5. Risultati dell'analisi qualitativa

Di seguito sono riportati, per ciascun tratto preso in esame, alcuni brani che nella ricerca sono stati analizzati da un punto di vista qualitativo, permettendo di confermare e completare i risultati dell'analisi quantitativa. Per questioni di spazio, si riporta in questa sede un numero molto limitato di brani a titolo esemplificativo, selezionati tra quelli che

offrono spunti di riflessione più interessanti, oltre che più paradigmatici del fenomeno analizzato.

Per ogni brano tratto da un romanzo, si fornisce in (a) il testo di partenza in inglese e in (b) la traduzione presente nella versione italiana del romanzo stesso. In (c) – ed eventualmente in (d) – si propone una traduzione che si ritiene, sulla base della mia intuizione e competenza di italoфона nativa, più “naturale” per l’italiano.

5.1 Esplicitazione del pronome soggetto

Un esempio di esplicitazione superflua di pronome soggetto presumibilmente dovuta a interferenza dall’inglese è offerto dalla traduzione di questo brano, tratto dal celebre romanzo *Il signore delle mosche*:

(2)

- a. *Ralph counted. “I’m chief, then”. The circle of boys broke into applause. Even the choir applauded; and the freckles on Jack’s face disappeared under a blush of mortification. **He** started up, then changed his mind and sat down again while the air rang.*
- b. *Ralph contò le mani. “Allora il capo sono io”. Un applauso scoppiò tra i ragazzi in cerchio. Anche il coro applaudiva, e per la mortificazione Jack diventò così rosso che le lentiggini non si vedevano più. **Egli** si alzò, poi cambiò idea e si sedette di nuovo tra il frastuono.*
- c. *Ralph contò le mani. “Allora il capo sono io”. Tra i ragazzi in cerchio scoppiò un applauso. Anche il coro applaudiva, e per la mortificazione Jack diventò così rosso che le lentiggini non si vedevano più. [Ø] Si alzò, poi cambiò idea e si sedette di nuovo tra il frastuono.*

(a. Romanzo originale: *Lord of the Flies*, William Golding, 1954; b. Traduzione italiana: *Il signore delle mosche*, a cura di Filippo Donini, 1986; grassetto mio)

Come si vede in (2b), nonostante il personaggio di nome Jack fosse stato appena citato (in italiano addirittura come soggetto dell’ultima frase, mentre in inglese il soggetto è *the freckles*), il traduttore rende il pronome *he* (in grassetto) con un pronome esplicito – tra l’altro, un pronome dello standard letterario oggi ormai in declino (*egli*). Si tratta di una scelta assai poco giustificabile, che si ritiene dovuta soprattutto all’interferenza del testo di partenza. In (2c) si propone una versione del brano a mio giudizio più naturale. Si noti anche la posizione reciproca tra il verbo inaccusativo *scoppiò* e il relativo soggetto *un applauso*, che nella traduzione (2b) segue l’ordine inglese (cfr. par. 4.2 e 5.2), mentre in (2c) viene riportata all’ordine tipico dell’italiano.

5.2 Posizione del soggetto con verbi inaccusativi

Nel seguente brano si può osservare come la presenza di un verbo inaccusativo come *arrivare* con soggetto rematico non sia sufficiente a orientare il traduttore verso una struttura VN (come sarebbe più naturale), a fronte di un testo di partenza inglese con struttura NV.

(3)

- a. "I can assure you, you won't have to do that. And as for the sketches, I showed them to Mr. Boothby and he was immensely interested. **Nothing so important has come on the market for years.**"
- b. "Posso assicurarle che non sarà così. E quanto agli schizzi, li ho fatti vedere a Mister Boothby, che ha mostrato un estremo interesse. **Nulla di così importante è arrivato sul mercato da anni.**"
- c. "Posso assicurarle che non sarà così. E per quanto riguarda gli schizzi, li ho fatti vedere al signor Boothby, che ha mostrato un estremo interesse. Da anni **non arriva nulla di così importante sul mercato.**"
- d. "[...] Sono anni che sul mercato **non arriva qualcosa di così importante.**"

(a. Romanzo originale: *The Shell Seekers*, Rosamunde Pilcher, 1988; b. Traduzione italiana: *I cercatori di conchiglie*, a cura di Amina Pandolfi, 1996; grassetto mio)

In questo estratto, infatti, la struttura inglese *nothing so important has come* (3a) viene perfettamente riprodotta nella traduzione italiana (3b. *nulla di così importante è arrivato*), nonostante sia presente un verbo inaccusativo (*arrivare*), nonché un quantificatore negativo (*nulla*). In effetti, il caso in esame è del tutto simile a quello individuato da Giusti (2004: 159) in *Harry Potter e la camera dei segreti* (...*niente di male potrà accadervi...*, calco di ...*no harm can befall you...*).

Una traduzione più naturale è quella che viene proposta in (3c), che prevede non solo lo spostamento del soggetto in posizione posposta al verbo, ma anche una riorganizzazione complessiva della frase e un cambiamento di tempo verbale: con la locuzione *da anni* (e simili: *da giorni*, *da mesi*, ecc.), difatti, la mia competenza di italoфона mi suggerisce un presente semplice, piuttosto che un passato prossimo, forse in quanto inconsciamente quest'ultimo è legato all'aspetto perfettivo. Una versione ancora alternativa è proposta in (3d) ed è quella che maggiormente si discosta dal modello inglese, ma che oggi verrebbe probabilmente più spontanea a un italofono: si tratta infatti di una frase scissa (è X che Y), costruito tipico del neo-standard (Sabatini, 1985), che permette tra l'altro di volgere il soggetto indefinito da negativo a positivo ed eliminare la doppia negazione.

Dal punto di vista lessicale, si noti incidentalmente che nella traduzione ufficiale (3b) la sigla *Mr.* viene sciolta in *Mister* (universale dell'esplicitazione), al quale, tuttavia, non si sceglie di sostituire l'italiano *signor* (che si propone in 3c, seppur come alternativa possibile e non necessariamente auspicabile). Entrambe le opzioni sono valide: *Mister* è coerente con l'ambientazione inglese del romanzo, *signor* risulta sicuramente più familiare a un lettore italofono.

5.3 Posizione dell'aggettivo qualificativo rispetto al nome

Per quanto riguarda la collocazione prenominale di aggettivi qualificativi, si è scelto di riportare due esempi che mostrino come, in virtù della natura dell'aggettivo (Tab. 4), tale collocazione può risultare più o meno accettabile.

Nel primo caso, seppur si possa ipotizzare l'interferenza dell'inglese nella scelta di anteporre sistematicamente una serie di aggettivi alla propria testa nominale,¹³ il risultato è comunque accettabile, in quanto gli aggettivi appartengono alla II classe dello schema di Cerruti (2015):

(4)

- a. *It wasn't the normal universe he knew. It was a universe of densely enfolded worlds, of **wild topographies, towering mountain peaks, heart stopping ravines**, of moons shattering off into sea horses [...] and bottomless hurtling hooping funts.*
- b. *Non era il normale universo a lui noto. Era formato da mondi tutti avviluppati tra loro, da **inconcepibili topografie, torreggianti vette, vertiginosi burroni**, lune che si disgregavano in cavallucci marini [...] e insondabili, rimbombanti, avvolgenti blob.*
- c. *Non era il normale universo a lui noto. Era formato da mondi tutti avviluppati tra loro, **topografie inconcepibili, vette torreggianti, burroni vertiginosi**, lune che si disgregavano in cavallucci marini [...] e insondabili, rimbombanti, avvolgenti blob.*

(a. Romanzo originale: *Mostly Harmless*, Douglas Adams, 1992; b. Traduzione italiana: *Praticamente innocuo*, a cura di Laura Serra, 1993; grassetto mio)

Nella traduzione seguente, al contrario, si può osservare un caso di interferenza che tende a comparire meno di frequente, proprio perché l'aggettivo *unfamiliar* (sconosciuta) appartiene alla I classe di aggettivi, che di norma si associa a una collocazione post-nominale (Tab. 4).

(5)

- a. *She saw Noel's old Jaguar parked by the garage, an **unfamiliar bicycle** leaning against the wall of the house, and a forlorn group of unburnable objects that had obviously been dumped, pending disposal.*
- b. *Vide la vecchia Jaguar di Noel parcheggiata presso il garage, una **sconosciuta bicicletta** appoggiata al muro della casa e un mucchio di strani oggetti abbandonati che evidentemente non si potevano bruciare e che aspettavano la loro sorte.*
- c. *Vide la vecchia Jaguar di Noel parcheggiata presso il garage, una **bicicletta sconosciuta** appoggiata al muro della casa e un mucchio di oggetti abbandonati che non si potevano bruciare ed evidentemente erano stati ammassati in attesa dello smaltimento.*

(a. Romanzo originale: *The Shell Seekers*, Rosamunde Pilcher, 1988; b. Traduzione italiana: *I cercatori di conchiglie*, a cura di Amina Pandolfi, 1996; grassetto mio)

¹³ Oltre a ciò, in casi come questo, si potrebbe mettere in associazione l'ordine AN con la volontà di innalzare il tono del racconto, rendendo il brano in qualche modo più "poetico". Non è da escludere, infatti, che dietro alla scelta del traduttore si possano collocare molteplici fattori, di diversa natura, i quali potrebbero aver agito in questo caso in maniera concorde, rafforzandosi a vicenda.

5.4 Perifrasi progressiva

Se in generale non sorprende l'ampio utilizzo della perifrasi progressiva in italiano (cfr. par. 4.4; per approfondire: Cortelazzo 2007), può risultare interessante la sua estensione all'ambito delle descrizioni, unità narrative dove tradizionalmente predomina il tempo imperfetto. Si veda, per esempio, il seguente estratto:

(6)

- a. *The woman's head was bowed and her lips were moving. **Belt radios were crackling** across the room. The alarm would be up any second now. Dead cop.*
- b. *La donnetta teneva la testa bassa e muoveva le labbra. **Radio ricetrasmittenti stavano crepitando** al lato opposto della stanza. L'allarme sarebbe stato dato da un momento all'altro. Poliziotto ucciso.*
- c. *La donnetta teneva la testa bassa e muoveva le labbra. **Radio ricetrasmittenti crepitavano** dal lato opposto della stanza. L'allarme sarebbe stato dato da un momento all'altro. Poliziotto ucciso.*

(a. Romanzo originale: *Black Sunday*, Thomas Harris, 1975; b. Traduzione italiana: *Black Sunday*, a cura di Bruno Oddera, 1989; grassetto mio)

In questo brano viene descritta una scena poliziesca in cui sta per essere dato l'allarme per l'uccisione di un agente. Interessante notare che il traduttore (6b) usa l'imperfetto nella prima frase – e non rende, per esempio, *her lips were moving* con *le sue labbra si stavano muovendo* –, mentre in quella seguente abbandona l'imperfetto a favore della perifrasi progressiva. Probabilmente nella prima frase l'eventuale perifrasi è stata “bloccata” dal verbo *tenere*, usato qui in accezione stativa (*teneva la testa bassa* = stava a testa bassa, stava piegata): con i verbi stativi, infatti, in italiano resta valida (per adesso) una restrizione per cui di norma la perifrasi progressiva non viene utilizzata (Degano, 2005). Una frase come *la donnetta stava tenendo la testa bassa e muovendo le labbra* suona ancora piuttosto “strana”. Al contrario, delle radio ricetrasmittenti, si dice tranquillamente che *stavano crepitando*, calco del modello inglese. Ovviamente, anche qui, l'imperfetto sarebbe stato non solo accettabile, ma avrebbe creato anche una certa continuità di tempi verbali in rapporto alla prima frase (vd. la traduzione proposta in (6c)).

6. Conclusioni

Nel complesso, l'indagine condotta conferma l'ipotesi di partenza, ossia l'esistenza di differenze quantitativamente e qualitativamente misurabili tra italiano originale e italiano tradotto dall'inglese per quanto riguarda la lingua della narrativa.

In particolare, i risultati avvalorano la tesi di una precisa interferenza dell'inglese sull'italiano, che porta il traduttore ad allentare alcune regole della propria grammatica mentale, optando talvolta per un'alternativa possibile ma meno “naturale” o adeguata al contesto. L'interferenza dall'inglese è confermata dal fatto che in molti casi il corpus di controllo (traduzioni da lingue varie) fa registrare valori simili al corpus di romanzi originali italiani. Nello specifico dei tratti tipologici analizzati, la ricerca conferma le intuizioni di Cardinaletti (2005) e Giusti (2004), ma anche Ondelli & Viale (2010),

sull'aumento dell'uso di pronomi nelle traduzioni, soprattutto per quanto riguarda i cosiddetti pronomi neo-standard.

Particolarmente evidente è anche l'influsso dell'inglese nell'uso della perifrasi progressiva, che, in accordo con quanto già rilevato da Gallitelli (2016), è nettamente più frequente nei romanzi tradotti rispetto ai romanzi originali italiani (+ 63% nel mio corpus). Diversamente da quanto rilevato da Gallitelli (2016), invece, le analisi qui condotte sembrano individuare anche una certa influenza della lingua inglese nella direzione di periodi più corti e minore ricorso all'ipotassi, oltre che un leggero abbassamento del valore TTR.

Più complesso è stato rilevare numericamente l'interferenza dell'inglese sulla posizione del soggetto rispetto ai verbi inaccusativi, seppur da un punto di vista quantitativo siano emersi numerosi casi simili a quelli individuati da Giusti (2004) nel suo studio sulle traduzioni della saga di *Harry Potter*.

Per quanto riguarda infine la posizione dell'aggettivo rispetto al nome, risulta determinante la natura dell'aggettivo: sulla base dei dati ottenuti, infatti, sembra che l'interferenza dell'inglese sia maggiore laddove il sistema linguistico italiano contempla già la posizione prenominali (aggettivi della II e IV classe della classificazione di Cerruti, 2015).

Questo porta direttamente a una considerazione che è bene rimarcare, nella scia di Cardinaletti (2005) e Giusti (2004): l'interferenza che ha luogo nelle traduzioni non produce (quasi) mai frasi totalmente agrammaticali. Non si tratta, infatti, di usi del tutto scorretti, ma piuttosto di un uso più vasto di alternative insolite, meno naturali per il profilo tipologico della lingua d'arrivo, o "non perfettamente accettabili nel contesto" (Giusti, 2004: 154), ma comunque previste dal sistema. Non a caso, l'interferenza interessa maggiormente i cosiddetti "fenomeni di interfaccia" (Cardinaletti 2004, 2005), aree instabili del sistema o che stanno vivendo processi di risistemazione (come il caso dei pronomi o della perifrasi progressiva). Ciò rende, in ultima analisi, particolarmente rilevante la ricerca in questo ambito, soprattutto in relazione a una lingua come l'inglese, con cui l'italiano si trova sempre più virtualmente a contatto.

Appendice

Composizione del corpus

I. Originali italiani

1. *Anima mundi*, Susanna Tamaro, 1997;
2. *Ascolta il mio cuore*, Bianca Pitzorno, 1991;
3. *Il conto dell'ultima cena*, Andrea G. Pinketts, 1988;
4. *Il nome della rosa*, Umberto Eco, 1980;
5. *L'amore è un bar sempre aperto*, Grazia Varesani, 1999;
6. *La sindrome*, Dario Argento, 1996;
7. *Oceano mare*, Alessandro Baricco, 1993;
8. *Q*, Luther Blisset, 1999;
9. *Sostiene Pereira*, Antonio Tabucchi, 1994;
10. *Uto*, Andrea De Carlo, 1995;

II. Traduzioni dall'inglese

11. *Black Sunday* (*Black Sunday*), Thomas Harris, 1975; trad. it.: Bruno Oddera, 1989;
12. *I cercatori di conchiglie* (*The Shell Seekers*), Rosamunde Pilcher, 1987; trad. it.: Amina Pandolfi, 1996;
13. *Il martello dell'Eden* (*The Hammer of Eden*), Ken Follett, 1998; trad. it.: Annamaria Raffo, 1998;
14. *Il signore delle mosche* (*Lord of the Flies*), William Golding, 1954; trad. it.: Filippo Donini, 1986;
15. *In Patagonia* (*In Patagonia*), Charles Bruce Chatwin, 1977; trad. it.: Marina Marchesi, 1981;
16. *La famiglia Winshaw* (*What a Carve Up!*), Jonathan Coe, 1994; trad. it.: Alberto Rollo, 1995;
17. *Le parole della notte* (*Reading in the Dark*), Seamus Deane, 1996; trad. it.: Vincenzo Mantovani, 1997;
18. *Morte innaturale* (*Unnatural Exposure*), Patricia Cornwell, 1997; trad. it.: Renato Pera, 1990;
19. *Nemesis* (*Nemesis*), Isaac Asimov, 1989; trad. it.: Piero Anselmi, 1990;
20. *Praticamente innocuo* (*Mostly Harmless*), Douglas Adams, 1992; trad. it.: Laura Serra, 1993;

III. Traduzioni da lingue varie

21. *Il figlio della luce* (*Le fils de la lumière*), Christian Jacq, 1995, lingua originale: **francese**; trad. it.: Francesco Saba Sardi, 1997;
22. *Il maestro e Margherita* (*Мастер и Маргарита*), Michail Bulgakov, 1967, lingua originale: **russo**; trad. it.: Claudia Zonghetti, 1995;
23. *Il paradiso degli orchi* (*Au bonheur des ogres*), Daniel Pennac, 1985, lingua originale: **francese**; trad. it.: Yasmina Mélaouah, 1991;
24. *L'insostenibile leggerezza dell'essere* (*Nesnesitelná lehkost bytí*), Milan Kundera, 1984, lingua originale: **ceco**; trad. it.: Giuseppe Dierna, 1995;
25. *La casa degli spiriti* (*La casa de los espíritus*), Isabel Allende, 1982, lingua originale: **spagnolo**; trad. it.: Angelo Morino e Sonia Piloto Di Castri, 1982;
26. *La donna e la scimmia* (*Kvinden og aben*), Peter Høeg, 1996, lingua originale: **danese**; trad. it.: Bruno Berni, 1997;
27. *Le braci* (*A gyertyák csonkig égnek*), Sándor Márai, 1942, lingua originale: **ungherese**; trad. it.: Marinella D'Alessandro, 1998;
28. *Monte cinque* (*O Monte Cinco*), Paulo Coelho, 1996, lingua originale: **portoghese**; trad. it.: Rita Desti, 1998;
29. *Recita estiva* (*Sommerstück*), Christa Wolf, 1989, lingua originale: **tedesco**; trad. it.: Anita Raja, 1989;
30. *Tsugumi* (*Tugumi*), Banana Yoshimoto, 1989, lingua originale: **giapponese**; trad. it.: Alessandro G. Gerevini, 1994;

Lingua	Famiglia linguistica	Tipo morfologico	Casi	Ordine dei costituenti	Sintagma nominale	Pro-drop
Francese	Lingue indoeuropee (romanze)	Flessivo-fusiva	No	SVO	NA / (AN)	No (pronome obbligatorio in posizione di sogg.)
Russo	Lingue indoeuropee (slave orientali)	Flessivo-fusiva	Si (6)	SVO	AN	No (pronome obbligatorio in posizione di sogg.)
Francese	Lingue indoeuropee (romanze)	Flessivo-fusiva	No	SVO	NA / (AN)	No (pronome obbligatorio in posizione di sogg.)
Ceco	Lingue indoeuropee (slave occidentali)	Flessivo-fusiva	Si (7)	SVO	AN	Si (marca di accordo sul verbo)
Spagnolo	Lingue indoeuropee (romanze)	Flessivo-fusiva	No	SVO	NA / (AN)	Si (marca di accordo sul verbo)
Danese	Lingue indoeuropee (germaniche settentrionali)	Flessivo-fusiva	No	SVO	AN	No (pronome obbligatorio in posizione di sogg.)
Ungherese	Lingue uraliche (ugrofinniche)	Agglutinante	Si (18)	Nessun ordine dominante	AN	Si (marca di accordo sul verbo)
Portoghese	Lingue indoeuropee (romanze)	Flessivo-fusiva	No	SVO	NA	Si (marca di accordo sul verbo)
Tedesco	Lingue indoeuropee (germaniche occidentali)	Flessivo-fusiva	Si (4)	Nessun ordine dominante	AN	No (pronome obbligatorio in posizione di sogg.)
Giapponese	Lingue nipponiche	Agglutinante	Si (8-9)	SOV	AN	Si (pronome opzionale in posizione di sogg.)

Tabella I. Caratteristiche genealogiche e tipologiche delle lingue che compongono il terzo subcorpus (romanzi tradotti da lingue varie).

Bibliografia

- Baker, Mona. 1993. Corpus Linguistics and Translation Studies: Implication and Applications. In Baker, Mona & Francis, Gill & Tognini-Bonelli, Elena (eds), *Text and Technology: in honour of John Sinclair*, 233-250. John Benjamins Publishing Company, Philadelphia/Amsterdam.
- Baker, Mona. 1996. Corpus-based Translation Studies: the Challenges that Lie Ahead. *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*, 18, 175-188. John Benjamins Publishing Company, Philadelphia/Amsterdam.
- Ballarè, Silvia. 2020. L'italiano neo-standard oggi: stato dell'arte. In *Italiano LinguaDue*, 12 (2): 469-492.
- Bentley, Delia. 2006. *Split intransitivity in Italian*. De Gruyter Mouton, Berlin/New York.
- Bertinetto, Pier Marco & Ebert, Karen H. & De Groot, Casper. 2000. The progressive in Europe. In Dahl, Östen (ed.) *Tense and aspect in the languages of Europe*, 517-558. De Gruyter Mouton, Berlin/New York.

- Blum-Kulka, Shoshana. 1986. Shift of Cohesion and Coherence in Translation. In House, Juliane & Blum-Kulka, Shoshana (eds.), *Interlingual and Intercultural Communication*, 17-35. Gunter Narr, Tübingen.
- Cardinaletti, Anna. 2004. La traduzione dei pronomi: interferenza sintattica e cambiamento linguistico. In Garzone, Giuliana & Cardinaletti, Anna (a cura di), *Lingua, mediazione linguistica e interferenza*, 129-150. FrancoAngeli, Milano.
- Cardinaletti, Anna & Garzone, Giuliana (a cura di). 2005. *L'italiano delle traduzioni*. FrancoAngeli, Milano.
- Cardinaletti, Anna. 2005. La traduzione: un caso di attrito linguistico. In Cardinaletti, Anna & Garzone, Giuliana (a cura di), *L'italiano delle traduzioni*, 59-83. FrancoAngeli, Milano.
- Cerruti, Massimo. 2015. La collocazione prenominal di sintagmi aggettivali complessi nell'italiano contemporaneo. Il contatto linguistico come 'rinforzo' di una possibilità del sistema. In Consani, Carlo (a cura di), *Contatto interlinguistico fra presente e passato*, 397-420. LED Edizioni, Milano.
- Cortelazzo, Michele. 2007. La perifrasi progressiva in italiano è un anglicismo sintattico? In *Studi in onore di Pier Vincenzo Mengaldo per i suoi settant'anni, a cura degli allievi padovani*, 1753-1764. SISMEL Edizioni del Galluzzo, Firenze.
- De Vita, Emanuela. 2024. *L'italiano delle traduzioni come varietà di contatto. Un approccio tipologico a un corpus di narrativa*. Alma Mater Studiorum, università di Bologna. (Tesi di laurea magistrale).
- Degano, Chiara. 2005. Influssi inglesi sulla sintassi italiana: uno studio preliminare sul caso della perifrasi progressiva. In Cardinaletti, Anna & Garzone, Giuliana (a cura di), *L'italiano delle traduzioni*, 85-105. FrancoAngeli, Milano.
- Dell'Orletta, Felice & Montemagni, Simonetta & Venturi, Giulia. 2011. READ-IT: Assessing Readability of Italian Texts with a View to Text Simplification. In *Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies*, 73-83. Association for Computational Linguistics, Edinburgh, Scotland, UK.
- Dryer, Matthew S. & Haspelmath, Martin (eds.). 2013. WALS Online (v2020.4) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13950591>. (Available online at <https://wals.info>, Accessed on 2025-02-12.)
- Dryer, Matthew S. 2013a. *Expression of Pronominal Subjects*. In Dryer, Matthew S. & Haspelmath, Martin (eds.) WALS Online (v2020.4) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13950591>. (Available online at <http://wals.info/chapter/101>, Accessed on 2025-02-12).
- Dryer, Matthew S. 2013b. *Order of Adjective and Noun*. In Dryer, Matthew S. & Haspelmath, Martin (eds.) WALS Online (v2020.4) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13950591>. (Available online at <http://wals.info/chapter/87>, Accessed on 2025-02-12.).
- Durante, Marcello. 1981. *Dal latino all'italiano moderno: saggio di storia linguistica e culturale*. Zanichelli, Bologna.
- Gallitelli, Eleonora. 2016. *Il ruolo delle traduzioni in Italia. Dall'unità alla globalizzazione*. Aracne editrice, Ariccia (RM).
- Garzone, Giuliana & Cardinaletti, Anna (a cura di). 2004. *Lingua, mediazione linguistica e interferenza*. FrancoAngeli, Milano.

- Garzone, Giuliana. 2004. Traduzione e interferenza linguistica: il punto di vista della traduttologia. In Garzone, Giuliana & Cardinaletti, Anna (a cura di), *Lingua, mediazione linguistica e interferenza*, 105-127. FrancoAngeli, Milano.
- Garzone, Giuliana. 2005. Osservazioni sull'assetto del testo italiano tradotto dall'inglese. In Cardinaletti, Anna & Garzone, Giuliana (a cura di), *L'italiano delle traduzioni*, 35-57. FrancoAngeli, Milano.
- Giusti, Giuliana. 2004. Interferenza dell'inglese sulla posizione del soggetto in italiano: alcune considerazioni sulle traduzioni italiane di Harry Potter. In Garzone, Giuliana & Cardinaletti, Anna (a cura di), *Lingua, mediazione linguistica e interferenza*, 151-166. FrancoAngeli, Milano.
- Grandi, Nicola. 2019. Che tipo, l'italiano neostandard!. In Moretti B., Kunz A., Natale S. & Krakenberger E. (a cura di), *Le tendenze dell'italiano contemporaneo rivisitate. Atti del LII Congresso Internazionale di Studi della Società di Linguistica Italiana (Berna, 6-8 settembre 2018)*, 59-74.
- Ježek, Elisabetta. 2003. Classi di verbi italiani tra semantica e sintassi. In Giacomo-Marcellesi, Mathée & Rocchetti, Alvaro (a cura di), *Il verbo italiano: studi diacronici, sincronici, contrastivi, didattici. Atti del XXXV Congresso internazionale di studi della Società di Linguistica italiana, (Parigi 20-22 settembre 2001)*, 251-263.
- Ježek, Elisabetta. 2010. Inaccusativi, verbi. In Raffaele, Simone (a cura di), *Enciclopedia dell'Italiano Treccani*. (https://www.treccani.it/enciclopedia/verbi-inaccusativi_%28Enciclopedia-dell%27Italiano%29/, ultima consultazione: 22 novembre 2024).
- Lenci, Alessandro & Montemagni, Simonetta & Pirrelli, Vito. 2005. *Testo e computer. Elementi di linguistica computazionale*. Carocci editore, Roma.
- Ondelli, Stefano. 2020. *L'italiano delle traduzioni*. Carocci editore, Roma.
- Ondelli, Stefano & Viale, Matteo. 2010. L'assetto dell'italiano delle traduzioni in un corpus giornalistico. Aspetti qualitativi e quantitativi. In *Rivista internazionale di tecnica della traduzione*, 12, 1-62.
- Sabatini, Francesco. 1985. L'italiano dell'uso medio': una realtà tra le varietà linguistiche italiane. In Holtus, Günter & Radtke, Edgar (Hrsg.), *Gesprochenes Italienisch in Geschichte und Gegenwart*, 154-184. Gunter Narr Verlag, Tübingen.
- Salsnik, Erika. 2007. Dagli universali traduttivi all'italiano delle traduzioni. In Montella, Clara & Marchesini, Giancarlo (a cura di), *I saperi del tradurre. Analogie, affinità, confronti*, 101-131. FrancoAngeli, Milano.
- Toury, Gideon. 1995. *Descriptive Translation Studies and beyond*. John Benjamins Publishing Company, Philadelphia/Amsterdam.

Proto-ausiliari e verbi supporto nella costruzione dell'interlingua di migranti adulti con alfabetismo limitato.

Egle Mocciaro

Masaryk University, Brno

egle.mocciaro@mail.muni.cz

Abstract

Questo contributo analizza l'emergere e la persistenza di costruzioni verbali analitiche non target nell'italiano L2 di migranti adulti con alfabetismo limitato, con particolare attenzione ai proto-ausiliari (*essere, avere*) e alle costruzioni a verbo supporto con *fare*. I dati provengono da uno studio longitudinale condotto su 20 apprendenti maschi, di diversa provenienza linguistica, seguiti per 13 mesi in contesto migratorio siciliano. Inquadrata nella prospettiva della Basic Variety, l'analisi qualitativa mostra che tali costruzioni emergono sistematicamente nella fase di transizione tra varietà basica e post-basica iniziale. Pur non essendo esclusive degli apprendenti poco alfabetizzati, esse risultano significativamente più frequenti e stabili nelle loro interlingue rispetto a quelle degli apprendenti alfabetizzati. L'articolo propone che questa asimmetria sia spiegabile in termini di accesso all'input: l'assenza o la forte limitazione dell'alfabetismo riduce l'esposizione alla lingua scritta e favorisce il riuso di forme funzionali fonologicamente salienti presenti nell'input orale. Le costruzioni analitiche sono interpretate come soluzioni grammaticali alternative, discorsivamente motivate, che possono stabilizzarsi in condizioni di input scarso, contribuendo a una riflessione più ampia sul ruolo dell'alfabetismo nello sviluppo delle interlingue in contesto migratorio

1. Campioni sconvenienti: da dove nasce la ricerca¹

La ricerca sullo sviluppo delle lingue seconde (L2) o, più in generale, aggiuntive (LA) da parte di immigrati adulti rappresenta oggi un'area ancora piuttosto marginale in ambito acquisizionale. Questa lacuna non è costitutiva della disciplina, ma si è prodotta dopo una fondamentale stagione di studi che, tra gli anni Settanta e i primi Novanta, aveva

¹ Il lavoro qui presentato ha ricevuto il supporto del progetto "A lifetime with language: the nature and ontogeny of linguistic communication ([LangInLife](#))", reg. no.: CZ.02.01.01/00/23_025/0008726, finanziato dal Fondo Europeo di Sviluppo Regionale (FESR).

permesso di delineare i contorni delle grammatiche delle interlingue proprio a partire da dati raccolti in contesto migratorio (Becker et al. 1977; Bernini & Giacalone Ramat 1990; Clahsen et al. 1983; Giacalone Ramat 2003; Klein & Perdue 1992, 1997; Perdue 1993 inter al.). Negli anni successivi, e soprattutto a partire dagli anni Duemila, l'attenzione si è spostata in modo sempre più deciso verso altri tipi di apprendenti, per lo più studenti, spesso universitari, che volontariamente intraprendono un percorso di apprendimento guidato nelle classi di lingua.² Si tratta, quindi, di campioni di convenienza, che offrono senz'altro innumerevoli vantaggi dal punto di vista dell'organizzazione della ricerca (anzitutto in termini di costruzione del campione e di controllo delle variabili, come l'età, le lingue madri, la scolarizzazione), ma che rispecchiano dinamiche di apprendimento statisticamente poco rappresentative (Arnett 2008; Plonsky 2023).³

In effetti, la stragrande maggioranza di coloro che sviluppano nuove lingue da adulti lo fa nell'ambito di un'esperienza migratoria, quindi attraverso l'interazione spontanea (o *naturalistica*) con i locali,⁴ piuttosto che nelle aule universitarie. Tuttavia, lavorare al di fuori delle aule significa muoversi su un terreno molto accidentato dal punto di vista della stabilità del campione, del controllo delle variabili, delle tecniche di raccolta dei dati, con importanti implicazioni anche sul piano dell'etica della ricerca.

Le migrazioni più recenti verso l'Europa coinvolgono in larga misura individui provenienti da aree del mondo, come l'Asia meridionale o l'Africa sub-sahariana, che offrono modelli sociolinguistici molto distanti da quelli europei, per esempio tassi di scolarizzazione e di alfabetizzazione molto bassi (UIS 2017a, 2017b), modelli educativi diversi (si pensi alle scuole coraniche; cfr. Saleem 2015, 2018), ma anche forme di plurilinguismo "fluido" che, specialmente in molte realtà africane, si sviluppano nell'interazione spontanea piuttosto che nelle classi (Lüpke 2017) e continuano in fasi diverse della vita, incluso il viaggio migratorio (D'Agostino 2021). Né vanno sottovalutate, per l'impatto sulle dinamiche linguistiche in gioco, le condizioni drammatiche e spesso traumatiche che precedono, accompagnano e seguono la migrazione, incluso l'alto grado di emarginazione sociale nei paesi di arrivo (D'Agostino 2021; D'Agostino & Mocciaro 2021; Mocciaro 2020; Monsen & Steien 2022), che ha conseguenze negative anche sul piano dell'esposizione alla lingua locale e, dal punto di vista del ricercatore, su quello dell'accessibilità al dato e alle modalità di raccolta.

Se queste difficoltà spiegano la persistente preferenza degli acquisizionalisti per i campioni di convenienza, la focalizzazione pressoché esclusiva su quella che è stata efficacemente definita popolazione "WEIRD (Western, Educated, Industrialised, Rich, Democratic)" (Henrich et al. 2010) pone più di un problema in termini di rappresentatività e generalizzabilità dei risultati della ricerca (Andringa & Godfroid 2020; Bigelow & Tarone 2004; Godfroid & Andringa 2023; Plonsky 2023; Tarone & Bigelow 2005; Tarone 2010). E se è vero che, negli ultimi anni, alcune delle lettere dell'acronimo hanno finalmente perso nitidezza, grazie a una crescente attenzione verso apprendenti

² Una scorsa agli indici delle principali riviste di linguistica acquisizionale supporta questa affermazione (cfr. osservazioni simili in Arnett 2008; Bigelow & Tarone 2004 inter al.).

³ O, altrimenti, proiettate per default su altri tipi di apprendenti. Come recentemente osservato da Forsberg Lundell & Beaulieu (2023), gli acquisizionalisti hanno talvolta lavorato su dati raccolti in contesto migratorio, senza però indagarne le specificità. L'analisi sistematica di queste specificità costituirebbe, d'altra parte, un importante passo avanti nella direzione di quel *social turn* da tempo invocato nella ricerca acquisizionale, ma rimasto per lo più confinato nella dimensione del manifesto (Block 2003; The Douglas Fir Group 2016; Ortega 2019).

⁴ I cosiddetti "nativi", ma portatori, in effetti, di diverse varietà di una data lingua. Per una critica del concetto di "nativo", cfr. inter al. Berruto (2003); Coulmas (1981); Davies (2013).

provenienti da paesi non occidentali, né necessariamente industrializzati, ricchi o democratici, lo status della E (Educated) rimane ancora piuttosto stabile.

1.1 Il contesto della ricerca: tra vecchie e nuove domande

Le riflessioni qui proposte si situano in una prospettiva di ricerca relativamente nuova, che cerca di colmare le lacune nel modo in cui il dato acquisizionale è costruito attraverso l'assunzione di un'esplicita prospettiva migratoria. Lo sfondo più immediato è quello degli studi condotti all'Università di Palermo sulle pratiche plurilingui dei nuovi migranti che raggiungono le coste siciliane attraverso le rotte del Mediterraneo. Il profilo sociolinguistico di questa popolazione di parlanti, già tratteggiato in 1, è stato ben descritto da Mari D'Agostino (2021), che spiega come l'aggettivo "nuovo" vada riferito non solo alla natura relativamente recente di queste esperienze migratorie, ma anche alle loro nuove specificità. Si tratta di individui con competenze alfabetiche spesso limitate⁵ e che non necessariamente vengono inclusi, dopo l'arrivo, nel sistema scolastico locale.

Molti provengono da territori massicciamente multilingui, come sono i paesi dell'Africa subsahariana: un multilinguismo che si riverbera in un alto grado di plurilinguismo individuale, che a sua volta tende ad aumentare durante la mobilità migratoria, attraverso il contatto con altri migranti provenienti da altri paesi o, eventualmente, con gli autoctoni durante alcune fasi del viaggio. Mobilità, ma anche immobilità, fatta di continue interruzioni del viaggio, non di rado forzate, durante le quali i migranti vivono isolati in centri di accoglienza, hub, o prigioni: una condizione che, a livello linguistico, significa bassissima interazione con i locali e, quindi, bassissima esposizione alle lingue del territorio, ma anche stretta interazione con altri migranti e con altri repertori.

In questo contesto, ho condotto negli anni scorsi uno studio sulla morfosintassi del verbo in apprendenti con alfabetismo limitato, per verificare se e, eventualmente, in che modo questa condizione influisca sullo sviluppo della morfosintassi della LA. I partecipanti allo studio erano 20 adulti maschi dai 18 ai 30 anni, reclutati durante i test di alfabetizzazione somministrati all'Università di Palermo nel 2017-2018 (cfr. nota 5). Arrivavano dall'Africa occidentale (Gambia, Senegal, Guinea, Costa d'Avorio, Mali, Burkina Faso, Nigeria) e dal Bangladesh; i loro repertori (dichiarati) includevano, quindi, Pulaar, varie lingue Mande, Benue-Congo, Gur, e Bangla. Nessuno di loro era stato esposto all'italiano prima dell'arrivo in Sicilia, tra i 10 e i 21 mesi prima, ma molti avevano seguito, sebbene in modo assai frammentario, un corso di italiano subito dopo l'arrivo, in contesti di volontariato, che, offrendo un'esperienza di interazione con parlanti di italiano aveva, in alcuni casi, almeno garantito l'emersione di un'interlingua italiana, che però non andava oltre le primissime fasi.

Metà del campione risultava priva di competenze alfabetiche (non in grado di leggere o scrivere parole isolate in nessun sistema di scrittura) o solo parzialmente alfabetizzata (in grado di riconoscere le lettere dell'alfabeto latino o i caratteri di altri alfabeti; di decifrare solo con lentezza alcune parole isolate; di scrivere il proprio nome etc.); per gli altri 10 il test aveva rilevato buoni livelli di alfabetizzazione, derivante, secondo i casi,

⁵ Un sondaggio condotto alla Scuola di Lingua italiana per Stranieri (ItaStra) dell'Università di Palermo tra il 2017 e il 2018 su un campione di quasi 800 migranti (nuovi arrivati, ma anche residenti di medio-lungo periodo), ha portato alla luce un tasso di analfabetismo o di alfabetismo limitato di circa il 30% in qualsiasi lingua e sistema di scrittura (D'Agostino 2018).

dall'esperienza scolastica nel paese d'origine o, altrimenti, da percorsi informali (per es., l'insegnamento di un fratello maggiore o di un altro migrante durante l'esperienza migratoria).

Nessuno dei 20 partecipanti era stato inserito nelle classi di lingua e alfabetizzazione dopo la somministrazione del test. Lo sviluppo delle interlingue italiane è stato tracciato attraverso una raccolta di dati longitudinale di 13 mesi, durante cui i partecipanti sono stati testati cinque volte in sessioni individuali svolte nel centro di accoglienza in cui vivevano. Le sessioni comprendevano una versione ridotta dell'iniziale test di alfabetizzazione, interviste semiguide e compiti narrativi (racconto di video senza parole, ricostruzione di eventi attraverso l'uso di immagini etc.)⁶. L'analisi dei dati ha seguito criteri di tipo qualitativo e si è ispirata ai principi della *Basic Variety* di Klein & Perdue (1997) e delle sue applicazioni alle interlingue italiane (Banfi & Bernini 2003)⁷.

Non mi soffermerò qui sui risultati complessivi dello studio (che sono già stati pubblicati in Mocciaro 2019, 2020), ma su alcuni aspetti specifici, che sembrano correlarsi con il grado di alfabetismo dei partecipanti e, forse, con le specificità delle pratiche comunicative che li coinvolgono.

2. Di cosa parliamo quando parliamo di alfabetismo: alcune premesse teoriche

Quella di alfabetismo è una nozione tutt'altro che univoca, e non è certo questo il contesto appropriato per una disamina (cfr. Alves 2019; Wiley 2005). Vanno però almeno menzionati per grandi linee i diversi approcci che possono condizionare l'interpretazione di come questa fondamentale abilità interferisca con lo sviluppo di nuove lingue da parte degli adulti.

In senso stretto, l'alfabetismo è la condizione di chi sa usare un alfabeto (o, più in generale, un sistema di scrittura), che risulta da un percorso di alfabetizzazione, spesso ma non necessariamente nell'ambito della scolarizzazione primaria. Ma, al di là di questa nozione tecnica (consistente nel conoscere e saper tracciare segni grafici, riconnetterli a suoni etc.), esiste una nozione più ampia di alfabetismo (o, per usare un'etichetta divenuta di uso comune anche in Italia, *literacy*), secondo cui è alfabeto una persona che non solo sa decodificare e codificare testi scritti, ma è anche in grado, attraverso questa abilità, di comprendere e interpretare la realtà in cui si muove (è questa la definizione UNESCO 1958). In ambito antropologico, si è in effetti affermata l'idea che l'alfabetismo sia essenzialmente una pratica sociale che riflette funzioni specifiche della lingua, diverse da quelle svolte dalla comunicazione orale e culturalmente specifiche (Cardona 2009; Street 1984). Si tratta cioè non solo di possedere un codice, ma di sapersi orientare in un universo semiotico caratterizzato dalla scrittura e, quindi, di beneficiare di una esposizione articolata e continua ai testi scritti.

⁶ Tutti i compiti orali erano di natura fondamentalmente interazionale e i partecipanti potevano quindi negoziare il racconto, la descrizione etc. nello scambio dialogico con gli intervistatori (chi scrive e alcuni collaboratori della Scuola di Lingua italiana per Stranieri di Palermo). Le sessioni, di durata variabile anche in ragione dei diversi livelli linguistici e della diversa motivazione dei partecipanti, sono state registrate in formato .wav e quindi trascritte con ELAN.

⁷ In particolare, è stato ricostruito il percorso di sviluppo del verbo nelle interlingue dei venti apprendenti rispetto alla sequenza di acquisizione proposta in Banfi & Bernini (2003: 93). Le costruzioni che sono oggetto di questo lavoro sono emerse non elicitate dalle produzioni degli apprendenti.

Tuttavia, i pochi studi acquisizionali che hanno tenuto conto delle competenze alfabetiche di chi sviluppa una lingua seconda hanno considerato l'alfabetismo essenzialmente nel suo senso basilico di competenza tecnica, di codifica e decodifica di un dato sistema di scrittura – e, in particolare, delle scritture alfabetiche (cfr. Bigelow & Tarone 2004; Tarone & Bigelow 2005; Tarone et al. 2006; Tarone et al. 2009; Vainikka & Young-Scholten 2007; Vainikka et al. 2017 inter al.). Questa preferenza si deve all'interazione di questa fondamentale abilità con certe aree della cognizione (come la consapevolezza metalinguistica e, in particolare, fonemica, la capacità di individuare i confini di parola etc.), che a loro volta sono state da alcuni considerate responsabili dello scarso sviluppo morfologico delle interlingue di analfabeti (cfr. in particolare Bigelow & Tarone 2004; per una discussione sui diversi approcci, cfr. Mocciaro 2020: 50-59 e Mocciaro & Young-Scholten 2026). Per questo motivo, cioè per ragioni di comparabilità dei dati, ho impiegato un test di alfabetizzazione per misurare esplicitamente le competenze alfabetiche degli apprendenti nel mio corpus (cfr. 1). Tuttavia, i miei dati non hanno mostrato alcuna divaricazione nello sviluppo delle interlingue riconducibile a questa variabile. In ciò che segue, si vedrà come sia soprattutto l'altra interpretazione dell'alfabetismo, come capacità di partecipare a un universo scritto, a giocare un ruolo nella costruzione delle interlingue. In breve, considererò l'alfabetismo come una chiave d'accesso aggiuntiva all'input in LA.

3. Protoausiliari e costruzioni dell'interlingua

La presenza in italiano L2/LA di formazioni analitiche non target, cioè non riconducibili in quanto tali all'input, è stata per la prima volta rilevata da Bernini (1989: 201),⁸ che le interpretava come una “strategia transitoria per far fronte a problemi di espressione di categorie, temporali o anche di persona, per cui manca la forma appropriata (o ritenuta tale) nel paradigma incompleto dell'interlingua”. La strategia consiste nel codificare separatamente il significato lessicale e quello grammaticale, quest'ultimo affidato a forme sovraestese di *essere*, *avere* o altre parole funzionali combinate a forme inanalizzate (“basiche”, in termini di Klein & Perdue 1997) di un verbo lessicale, per esempio: *miei amici ero si chiama* ‘si chiamava’, *sono lavora* ‘(io) lavoro’, *siamo partenza* ‘partiamo’, *non avevo credo nel questo cosa* ‘non credevo in questa cosa’. In questi esempi (tratti da Bernini 1989: 201 e Banfi & Bernini 2003: 106), alla forma di *essere* o (meno frequentemente) *avere* è assegnato il compito di esprimere informazioni di natura temporale, aspettuale (passato imperfettivo o evento in atto nel presente) o di persona

⁸ Gli apprendenti erano quelli del corpus di Pavia (Bernini & Giacalone Ramat 1990; Giacalone Ramat 2003), quindi individui di età, sesso, provenienza, background linguistico e sociolinguistico diversi, che sviluppavano la grammatica della L2 in modo (prevalentemente) naturalistico. Tuttavia, il profilo sociolinguistico degli apprendenti non è stato tracciato in modo sistematico nell'ambito del progetto di Pavia e alcune informazioni possono essere parzialmente ricavate solo per via inferenziale (per esempio, il grado di istruzione dall'occupazione). Banfi & Bernini (2003: 106) osservano che le formazioni analitiche in esame si osservavano in misura diversa nei diversi apprendenti, ma specialmente in PE, MK e AB: il primo (PE) ha 25 anni, viene da Singapore, parla cantonese, malese e inglese, è ingegnere (quindi, inferiamo, ha un grado di istruzione alto ed è certamente alfabetizzato anche in alfabeto latino), era in Italia da 25 mesi ai tempi della prima intervista; il secondo (MK) ha 20 anni, è eritreo e parla tigrino, fa l'elettricista ed era in Italia da un mese alla prima intervista; la terza (AB) ha 21 anni, è eritrea e parla tigrino, fa la collaboratrice domestica ed era in Italia già da un anno ai tempi della prima intervista.

(cfr. anche Bernini 2003; Giacalone Ramat 1992: 307), mentre il secondo elemento⁹ esprime solo significato lessicale. In altre parole, il primo elemento della costruzione si comporta come un ausiliare (o “protoausiliare”, secondo Starren 2001), che esprime a livello analitico quello che nella lingua target è espresso per mezzo della flessione su un verbo lessicale.

È essenziale osservare che l’emergere di queste formazioni è strettamente connesso a una precisa fase di sviluppo dell’interlingua. Nella fase in questione, l’apprendente identifica nell’input la presenza di parole funzionali (come la copula), di cui evidentemente riconosce il potenziale grammaticale se può riusarle per codificare – sebbene in modo ancora non target – le categorie del verbo. In altri termini, i protoausiliari introducono la categoria della finitezza nel sistema verbale degli apprendenti prima che questa si propaghi ai verbi lessicali. Nei dati da me esaminati, in effetti, formazioni di questo tipo si situano invariabilmente nella fase di transizione tra ciò che nella formulazione di Klein & Perdue (1997) è detta “varietà basica” (in cui il verbo esiste come forma autonoma ed è dotato di struttura argomentale, ma è ancora privo di flessione, es. *io mangia*¹⁰) e quella postbasica iniziale (segnata dalla comparsa del participio passato per l’espressione della perfettività e, insieme, da quella delle prime parole funzionali, anzitutto la copula). Banfi & Bernini (2003: 93) schematizzano questa progressione in italiano L2 come riportato nella Tab. 1.

FASI	VALORI ASPETTUALI E TEMPORALI				
	PRESENTE	PASSATO PERFETTIVO	PASSATO IMPERFETTIVO	FUTURO	NON FATTUALE
1	Presente/Inf.	Presente/Inf.	Presente/Inf.	Presente/Inf.	Presente/Inf.
2	Presente	Part. pass.	Presente	Presente	Presente
3	Presente	(Aus.) Part. pass.	Imperfetto	Presente	Presente
4	Presente	(Aus.) Part. pass.	Imperfetto	Futuro	Futuro

Tabella 1. La complessificazione del sistema verbale nelle varietà postbasiche (Banfi & Bernini 2003: 93, adattato)

In particolare, le fasi che ci interessano sono la 2 e la 3: superata la fase basica (1), in cui il verbo appare solo in una forma inanalizzata, ed emerso il participio passato (2), che introduce un’opposizione di natura aspettuale, cominciano a comparire, nella fase 3, le prime parole funzionali, che però, diversamente da quanto la Tabella potrebbe suggerire (cfr. “Aus.(iliare)”), non si accompagnano necessariamente al participio passato per formare il passato prossimo (anzi, nei miei dati questo passaggio è raro e tutt’altro che sistematico), ma si prestano a creare formazioni non target come quelle appena descritte, combinandosi dunque con forme basiche. Gli esempi (1)-(3) illustrano alcuni casi di

⁹ Tipicamente ma non necessariamente un verbo, cfr. il già citato *sono partenza* (Banfi & Bernini 2003: 106).

¹⁰ O meglio, è pseudoflesso: diversamente da quanto avviene in altre interlingue in cui la forma basica è realizzata attraverso l’assenza del morfema grammaticale (es. ingl. *he walk-Ø*), in italiano L2 la forma basica presenta perlopiù una terminazione che richiama ora una terza persona singolare (o un tema, es. *io mangia*) ora altre forme, ma che è sostanzialmente inanalizzata, priva cioè di qualsiasi significato grammaticale. Nella fase basica, le informazioni di natura temporale, aspettuale o di persona sono espresse, eventualmente, per via lessicale (es. attraverso avverbi o pronomi).

protoausiliare *essere* nelle interlingue post-basiche iniziali di tre apprendenti, rispettivamente senegalese, maliano e burkinabe:

- (1) *adesto sono estudiato CPIA* (YS_3¹¹)
'Adesso studio al CPIA.'
- (2) *signorə è scendre* (MF_5)
'L'uomo scende.'
- (3) *allora lui era dormo* (MLG_5)
'Allora lui dormiva.'

In (1)-(3), le forme di *essere* codificano informazioni tempo-aspettuali e di persona: in (1) e (2), *sono* ed *è* indicano che la situazione codificata dalle forme (inanalizzate) del verbo lessicale, *estudiato* e *scendre*, è in atto nel presente; in (3), *era* marca l'aspetto imperfettivo nel passato (Bernini 1990, 2003); in tutti e tre i casi, il protoausiliare codifica anche la persona.

3.1 Costruzioni non target a verbo supporto

I miei dati mostrano anche un altro tipo di costruzione, che coinvolge il verbo *fare* costruito con un elemento lessicale inanalizzato, come in (4) e (5):

- (4) *fare dentifricia* (MJ_2)
ACT [dentifricio]
'Mi sono lavato i denti.'
- (5) *fare | lavare dentis e poi rasare | barba fare dolce (...)*
ACT [lavare denti] [rasare barba] ACT doccia
fare cuzina (MF_2)
ACT [cucinare]
'Mi sono lavato i denti, poi mi sono rasato, ho fatto la doccia e poi ho cucinato.'

In italiano, *fare* è usato come verbo supporto (o *light*) con nomi d'azione, come in *fare sport*, *fare ricerca* etc. In queste costruzioni, il nucleo semantico è costituito dal nome, che impone la struttura argomentale e assegna i ruoli semantici agli argomenti, mentre il verbo esprime la componente semantica "attività/azione", nonché le informazioni grammaticali di tipo tempo-aspettuale e di persona (cfr. inter al. Butt 2003; Ježek 2004, 2023). Gli apprendenti possono impiegare la costruzione con *fare* per colmare un gap lessicale (cfr. anche esempi come *fare danza* 'danzare' riportati in Giacalone Ramat 1997: 262), non necessariamente con nomi che esprimono in sé un'azione, come nell'esempio in (4), in cui l'apprendente, semplicemente, sembra non conoscere il verbo che codifica la situazione evocata da *dentifricio* e seleziona *fare* per renderne esplicito il carattere predicativo; in particolare, *fare* sembra indicare, non diversamente da quanto avviene nella lingua target, che la situazione descritta è una *attività* (segnalata come ACT nelle glosse¹²). Come mostra l'esempio in (5), *fare* può svolgere la stessa funzione con verbi

¹¹ Il numero accanto alle iniziali dell'apprendente si riferisce alla sessione di raccolta.

¹² Le glosse qui usate hanno valore funzionale, piuttosto che morfologico, e sono concepite come uno spazio di riflessione preliminare sulle categorie d'analisi da impiegare per interpretare le forme dell'interlingua. Nel caso specifico, ACT rimanda a una semantica azionale che include azioni e attività, mentre le parole

formalmente sottospecificati, rendendone esplicita proprio la “verbità” e, in particolare, indicando che si tratta di attività o azioni (e solo raramente, e in subordine, veicolando informazioni tempo-aspettuali e di persona). È interessante osservare che ACT è anche la caratterizzazione azionale predominante del secondo elemento della costruzione, che invariabilmente indica eventi come ‘lavare’, ‘cucinare’, ‘studiare’ etc.

La costruzione con *fare* è, come quella con *essere*, legata alla fase dell’interlingua: in particolare, emerge al passaggio tra la varietà basica e quella post-basica, prima ancora di quella che impiega la copula, almeno nelle interlingue in cui entrambe sono documentate.

3.2 Costruzioni dell’interlingua e alfabetismo degli apprendenti

Bernini (1989: 201; cfr. anche Banfi & Bernini 2003: 106) ha interpretato le costruzioni con protoausiliare descritte in 3 come strategie occasionali e temporanee, usate “per ovviare a deficit momentanei di espressione” e destinate a scomparire via via che la finitezza si stabilizza sui verbi lessicali. Questa descrizione potrebbe probabilmente essere estesa ai miei dati, includendo in essi anche le costruzioni con verbo supporto osservate in 3.1, se le interlingue in questione esibissero una tale stabilizzazione morfologica. Ma così non è, e solo pochissimi tra gli apprendenti nel corpus raggiungono davvero la fase 3: nelle loro interlingue, continuano ad alternarsi, secondo i casi, forme basiche, participi passati, forme di presente eventualmente flesso per persona, costruzioni analitiche anche target (per esempio, quella progressiva), rare forme di imperfetto. In Mocciano (2020), ho sostenuto che questa scarsa progressione riflette la pochezza quantitativa e qualitativa dell’input in italiano; non mi soffermerò oltre su questo in questa sede. Ciò su cui vorrei invece mettere qui l’accento è la correlazione tra la presenza e la persistenza delle costruzioni non target e il livello di alfabetismo degli apprendenti.

In generale, le costruzioni non target non sono certo esclusive degli apprendenti con alfabetismo limitato: si ritrovano, per quanto occasionalmente, in apprendenti certamente scolarizzati del Corpus di Pavia (cfr. 3, in particolare n. 8) e anche in altre interlingue, su cui tornerò tra poco (cfr. 4). Ma la loro incidenza cambia significativamente. Nel mio corpus, 11 apprendenti su 20 producono costruzioni non target, a cavallo tra fase basica e post-basica iniziale: 7 su 11 hanno competenze alfabetiche da nulle a molto limitate per l’intera durata della raccolta dati. Soprattutto, queste costruzioni persistono nel discorso dei poco alfabetizzati fino all’ultima sessione della raccolta, mentre negli apprendenti alfabetizzati esse sono non solo del tutto episodiche ma anche meno stabili nella microdiacronia delle loro interlingue.

Questa distribuzione è riportata nella Tab. 2, che mostra come le prime formazioni analitiche non target emergano solo a partire dalla varietà basica, quando il verbo si precisa come categoria autonoma (cfr. 3); si tratta per lo più di formazioni con *fare*, che, come si è visto in 3.2, serve a esprimere il valore predicativo della costruzione in cui

racchiuse tra parentesi quadre indicano che le forme in questione veicolano una mera semantica lessicale e sono, invece, sottospecificate a livello morfologico. Quella dell’etichettatura delle forme dell’interlingua è una questione piuttosto spinosa, specialmente se si adotta una prospettiva “interna” che non misuri le produzioni degli apprendenti nei termini della lingua target. Come acutamente osservato da Andorno & Rastelli (2009: 13-14), se lo scopo della ricerca acquisizionale è quello di ricostruire specifiche categorie dell’interlingua, non è chiaro quali categorie vadano impiegate per etichettarne le forme. In altri termini, le etichette dovrebbero essere il risultato e non la premessa dell’analisi.

occorre piuttosto che informazioni grammaticali come tempo, aspetto o persona; non stupisce, pertanto, che compaia prima del protoausiliare *essere*. Questo compare invece, accanto a *fare*, solo con l'avvenuta transizione alla varietà postbasica e, più in particolare, con l'emersione della copula. I numeri tra parentesi indicano le occorrenze, che sono con tutta evidenza assai più numerose e più stabili negli apprendenti con alfabetismo limitato (una cinquantina) rispetto a quelli alfabetizzati (meno di una decina). In 5, proporrò un'interpretazione di questi dati.

SESSIONE	GRUPPO CON ALFABETIZZAZIONE LIMITATA			GRUPPO CON ALFABETIZZAZIONE PIENA		
	STADIO DELL'INTERLINGUA			STADIO DELL'INTERLINGUA		
	prebasico	basico	postbasico	prebasico	basico	postbasico
1-2	-	fare (2)	fare (7) essere (2)	-	fare (1)	essere (2)
3	-	-	fare (2) essere (6)	-	-	-
4	-	-	fare (9) essere (4)	-	fare (1)	fare (1)
5	-	avere (3)	fare (9) essere (6)	-	-	essere (1) avere (1)

Tabella 2. Distribuzione delle costruzioni non target nel corpus Mocciaro (2020: 13, adattato).

4. Le costruzioni non target in altre varietà

Costruzioni analitiche come quelle descritte nei paragrafi precedenti sono state osservate in L2 diverse, il che suggerisce che si tratti di una fase più generale nel percorso di costruzione dell'interlingua piuttosto che di un fenomeno specifico di quelle italiane (cfr. inter al. Blom et al. 2013; Fleta 2003; Huebner et al. 1992; Jordens e Dimroth 2006; Julien et al. 2016; Starren 2001; von Stutterheim 1986, che hanno osservato fenomeni simili in diverse interlingue europee, tra cui francese, olandese, tedesco e inglese). Gli esempi in (6) e (7) illustrano due casi, discussi in Benazzo & Starren (2007: 142), nell'olandese L2 di apprendenti con L1 arabo marocchino; in entrambi i casi, una forma di *zijn* 'essere' (*was*, *is*) è combinata con altre due forme verbali, cioè un secondo ausiliare (*is*, *heft*) e la forma inanalizzata di un verbo lessicale (*gestaan*, formalmente un participio passato, e *aanvragen*, formalmente un infinito); il secondo ausiliare è una forma di 'essere' in (6) e di 'avere' in (7); il primo ausiliare esprime valore temporale (passato in 6, presente in 7), mentre il secondo valore aspettuale (imperfettivo in 6, perfettivo in 7); insomma, tempo e aspetto sono codificati separatamente:

- (6) *die was bij Charlie is gestaan*
 lei essere:PST.IPFV accanto Charlie essere:PRS.3SG stare:PST.PTCP
 'Lei era in piedi accanto a Charlie.'
- (7) *dan is hij heft werk aanvragen*
 poi essere:PRS.3SG lui avere:PRS.3SG lavoro domandare:INF
 'Poi ha chiesto lavoro.'

Altri esempi interessanti sono quelli analizzati da Pfaff (1992: 292) nel tedesco L2 di bambini turcofoni, che frequentavano un asilo nido bilingue in una *speech community*

multilingue di Berlino. In (8), troviamo una forma di ‘essere’ combinata con un verbo lessicale di forma non finita per esprimere una situazione in corso nel presente (simile, chiosa Pfaff, al progressivo inglese, impiegato infatti nella sua traduzione e qui riproposto in italiano); si tratta di una soluzione che non è attestata né nella L1 né nella lingua target.

- (8) *hier das is kommen*
 qui questo essere:PRS.3SG venire:INF
 ‘Sta arrivando qui.’

Ancora Pfaff (1992: 292) riporta alcuni esempi, come quello in (9), in cui una forma non target di ‘fare’, *mach*, si combina con la forma inanalizzata di un verbo lessicale; anche questo caso è interpretato come una struttura di valore progressivo:

- (9) *Elefant komm. die mach hauen*
 Elefante [venire]¹³ il [fare] battere:INF
 ‘L’elefante viene, sta combattendo (battendo).’

Pfaff (1992: 291-292) descrive formazioni come quelle in (8) e (9) come “compound verb constructions” che illustrano processi di grammaticalizzazione attivi nelle prime fasi dell’acquisizione della lingua seconda, forse, almeno per il caso di (9), per influsso della L1, in cui storicamente costruzioni di questo tipo sono state impiegate per integrare prestiti stranieri. Pfaff rimane tuttavia cauta circa la possibilità di considerare queste costruzioni come una prova del sovrapporsi delle grammatiche della L1 e della L2, sia perché si tratta comunque di esempi occasionali, sia perché i bambini che le producono in tedesco L2 non usano, d’altra parte, la parallela costruzione turca verbo + *yapmak* ‘fare’ e viceversa. Resta comunque interessante il parallelo tra dati acquisizionali e dati storici, ricordato anche da Bernini (1989: 202), che cita proprio l’esempio turco *tefrik etmek* ‘(letter. ‘distinzione fare’) distinguere’, dove *tefrik* è un prestito dall’ar. *tafriiqu* integrato per mezzo della formazione analitica con *etmek* ‘fare’. Oltre al caso dei prestiti da una lingua con maggiore prestigio culturale, Bernini (1989: 202) cita anche altre situazioni di contatto interlinguistico assimilabili in una certa misura a quelle dell’acquisizione della seconda lingua. Tra queste, richiama il caso del dialetto walser di Gressonay studiato da Giacalone Ramat (1989: 45), in cui formazioni analitiche con ‘essere’, ‘avere’ e ‘fare’ sono impiegate in diversi settori del paradigma, es. wenn *hit tet-es schnue* ‘(letter. ‘se oggi facesse esso nevicare’) se oggi nevicasse’. Più in generale, osserva Bernini, le formazioni analitiche servono in molte lingue a complessificare i paradigmi introducendo, eventualmente, specifiche sfumature aspettuali.

Il parallelismo con diverse situazioni di contatto suggerisce che i dati acquisizionali come quelli qui in discussione riflettano strategie più generali di accesso alle categorie della lingua e di costruzione della grammatica, piuttosto che un’influenza diretta della L1 sulla L2, tanto più che tanto le costruzioni analitiche con protoausiliari quanto quelle con *light verbs* si ritrovano anche in apprendenti con L1 diverse dal turco e che non necessariamente possiedono strutture simili. Per quanto riguarda le seconde, oltre a quelle emerse nei miei dati, costruzioni con *fare* verbo supporto si osservano occasionalmente anche in altre interlingue italiane, per esempio nell’italiano LA di adulti con L1 cinese

¹³ La forma *komm* non esiste nella lingua target, per questo motivo si trova qui tra parentesi quadre, che non si trovano tuttavia nell’originale: questa scelta riflette il tentativo di presentare tutti gli esempi in modo formalmente coerente.

studiato da Di Pinto (2020). Qui gli esempi sembrano concentrarsi nelle produzioni di Gioia, una donna di 36 anni, cresciuta in campagna, con alle spalle un regolare percorso di studi conclusosi con la licenza media (Di Pinto 2020: 66; tuttavia, non abbiamo informazioni certe sul livello di alfabetizzazione¹⁴). L'esempio in (10) mostra una situazione del tutto analoga ai casi discussi in 3.1, in cui una forma non flessa di *fare*, impiegata per veicolare valore predicativo, si combina con una forma non flessa di un verbo lessicale (nei dati in Di Pinto 2020: 175-178, entrambe le forme sono sempre all'infinito):

- (10) *bam 'bini? || bam 'bini qua cinesi | tutti 'fale: | an 'dale scuola*
 'I bambini cinesi vanno tutti a scuola.' (Di Pinto 2020: 176)

Sebbene in modo solo occasionale, *fare* + V si ritrova nell'interlingua degli apprendenti universitari di italiano L2 studiati da Lorenzová (2023 e in prep.); in (11), è riportato un esempio prodotto da un'apprendente russofona nella sua fase post-basica iniziale:

- (11) *RE: c'è qualche attività durante cui ti rilassi che è un po' antistress per te*
UD1: nel prima è fumare (...)
e fare passeggiare (.) comprare
comprare: abbigliamento o cibo (.) basta

La breve rassegna di casi qui proposta lascia intravedere alcune direzioni di studio che andrebbero percorse per comprendere meglio le costruzioni in esame. Come si è visto, i dati discussi in 3 e 3.1 suggeriscono che possa esistere una correlazione tra il grado di alfabetismo degli apprendenti e la frequenza e la persistenza delle costruzioni analitiche non target nelle loro interlingue. Gli esempi ricavati da altri studi e discussi in questa sezione non permettono di confermare questa ipotesi, sia perché non sempre abbiamo informazioni chiare sul grado e il tipo di alfabetizzazione degli apprendenti (in genere solo inferibile dalla scolarizzazione dichiarata), sia perché a volte essi sono prodotti da apprendenti certamente alfabetizzati (come PE nel corpus di Pavia o l'apprendente russofona dell'esempio 11). Non confermano l'ipotesi, ma nemmeno la sconfessano: che le costruzioni non target siano un fenomeno più generale (anche se non generalizzato) delle interlingue (e forse di altre situazioni di contatto) sembra un fatto acclarato – e osservabile, del resto, anche nei dati in Mocciaro (2020). Ciò che, piuttosto, va ancora spiegato è quali siano le circostanze che ne favoriscono non tanto l'emersione quanto la maggiore frequenza e la relativa stabilità. In 5, continuerò dunque a percorrere l'ipotesi dell'alfabetismo limitato.

5. Questione di input

La mia ipotesi riprende la spiegazione fonologica proposta da Vainikka & Young-Scholten (1998), secondo cui, soprattutto in contesti di bassa esposizione alle lingue

¹⁴ Dal momento che ha completato le scuole secondarie di primo grado, possiamo inferire che abbia appreso il *pīnyīn*, il sistema di trascrizione in alfabeto latino dei suoni cinesi; tuttavia, non abbiamo informazioni dirette sulle sue reali competenze alfabetiche, perché questa variabile non fa parte di quelle indagate da Di Pinto (2020).

target, gli apprendenti faticano ad acquisire costituenti di dimensioni inferiori a un piede fonologico, come sono i morfemi legati. Nel parlato, oltre a essere poco salienti dal punto di vista fonologico, i morfemi lo sono anche sostanzialmente (a causa dei frequenti fenomeni di riduzione che caratterizzano la lingua parlata, cfr. Savy 1999) e contestualmente, perché la frequente compresenza di altre forme più salienti, come gli avverbi, li rende ridondanti (per es. *domani partirò*; cfr. Cintrón-Valentín & Ellis 2016). Gli apprendenti che non hanno accesso (o hanno solo un accesso limitato) alla lingua scritta, perché non conoscono ancora la lingua target o non conoscono la sua forma scritta o non conoscono affatto la scrittura, ricevono l'input solo nella sua forma orale e non possono giovare di quella "permanent visual substance" (Cintrón-Valentín & Ellis 2016) che rende le forme grammaticali più facilmente identificabili e, quindi, più facilmente processabili. Questa "greater reliance on auditory memory than visual memory" (Vainikka et al. 2017: 248) fa sì che gli apprendenti si affidino a forme più "pesanti" (cioè fonologicamente e percettivamente più salienti), che possono essere più facilmente identificate nell'input orale: parole frequenti come sono quelle funzionali, che possono essere riusate per creare specifici pattern morfosintattici che compensano la mancanza di morfologia.

Due osservazioni sono a questo punto indispensabili. Anzitutto, sulla base dell'interpretazione appena esposta, l'alfabetismo – o la sua assenza – gioca un ruolo cruciale semplicemente perché determina la quantità e la qualità dell'input cui gli apprendenti sono esposti. In secondo luogo, questo input va situato: non è, ovviamente, *la* lingua target (in questo caso, l'italiano) nel suo senso più vasto e astratto¹⁵, ma gli specifici contesti discorsivi in cui gli apprendenti si trovano impegnati con italofofoni, il che – in contesto migratorio (in particolare, in quello da me studiato) – si riduce a una manciata di casi¹⁶ e, tra questi, proprio quelli innescati dai task per la raccolta dei dati. In questi contesti, gli apprendenti sono esposti con una certa frequenza a un certo tipo di strutture, per esempio a quelle impiegate nelle domande che facilitano le conversazioni guidate o i compiti narrativi.

Consideriamo per esempio l'uso di *fare* nelle domande dell'intervistatore (es. "Che cosa hai fatto/fai? Che cosa fanno?"), in situazioni in cui ai partecipanti viene chiesto di descrivere o raccontare qualcosa, come nel già discusso esempio (5), che riproduco qui insieme, appunto, alla domanda:

- (12) *dimmi, che cosa hai fatto ieri?*
fare | *lavare dentis* *e poi* *rasare* | *barba* *fare dolce (...)*
 ACT [lavare denti] [rasare barba] ACT doccia
fare *cuzina*
 ACT [cucinare] (MF_2)
 'Mi sono lavato i denti, poi mi sono rasato, ho fatto la doccia e poi ho cucinato.'

¹⁵ Come giustamente osserva un revisore anonimo, ciò è in effetti vero per qualsiasi situazione di apprendimento.

¹⁶ Sebbene l'analisi dell'input non sia stata tra gli obiettivi del mio studio, l'isolamento e la scarsa interazione con i locali emergono come dato incontrovertibile dai racconti dei partecipanti, come anche da quelli di altri partecipanti con un simile profilo incontrati nel contesto di lavori diversi (cfr. D'Agostino & Mocchiari 2022). Non c'è dubbio che studiare da vicino l'input cui chi apprende una lingua si trova via via esposto sia un compito tanto cruciale quanto arduo, specialmente in contesti poco controllati come quello di cui mi occupo qui; ringrazio il revisore anonimo che ha sollevato la questione.

In casi come questi, è lecito chiedersi se non sia la presenza (ricorrente) di *fare* nella domanda a innescare il suo riuso nella produzione dell'apprendente – ciò che è in genere definito *priming*. In altri termini, questo specifico contesto di discorso (quello dell'intervista o del task monolingue) si configura come il *locus* del cambiamento: uno spazio di reinterpretazione contestuale, in cui una data forma può sviluppare una nuova sfumatura semantica, che può eventualmente stabilizzarsi (che è il processo di grammaticalizzazione nella descrizione, per esempio, di Heine et al. 1991). A volte, la presenza di pause (|) o di esitazioni tra *fare* e il verbo lessicale suggerisce più chiaramente questa diretta replica dello stimolo contenuto nella domanda, il che ci potrebbe indurre a considerare questi casi come meri fatti di *performance* o, per usare gli efficaci termini di Bernini (1989), temporanee “strategie di discorso” destinate a dissiparsi rapidamente, piuttosto che come costruzioni dell'interlingua degli apprendenti. Questo è senz'altro vero in molti casi: nei casi cioè degli apprendenti che rapidamente superano la fase analitica dell'interlingua per approdare alla finitezza nei verbi lessicali. Va però tenuto conto del fatto che le pause, le esitazioni e, in generale, le disfluenze sono inerenti alla modalità parlata in generale (cfr. Bonvino 2005; Voghera 2017). Inoltre, nel caso specifico qui in esame, troviamo pause o esitazioni anche in altre situazioni sintattiche, come in *rasare* | *barba*, in cui non è possibile ravvisare alcun effetto di *priming*. Soprattutto, nel mio corpus in generale e anche specificamente nell'esempio in (5), la costruzione occorre altrettanto spesso con un profilo intonazionale continuo (*fare dolce* ‘ho fatto la doccia’, *fare cucina* ‘ho cucinato’). E che questo processo di grammaticalizzazione sia quanto meno in atto, se non compiuto, è suggerito dal fatto che *fare* + elemento lessicale inanalizzato può emanciparsi dal contesto che lo ha stimolato e apparire anche in stadi più avanzati dell'interlingua, quando la morfologia, se non completamente stabilizzata, è comunque ormai chiaramente emersa nei verbi lessicali, come nell'esempio in (13):

- (13) *ritorneresti nel tuo paese?*
il mio città (...) ci sono violenza (...) e ci fanno tagliano gola alle genti. (MLG_5)

6. Verso una conclusione

I costrutti analitici descritti in questo lavoro vanno interpretati come soluzioni grammaticali alternative a quelle morfologiche, che derivano da un uso specifico di forme frequenti nell'input (come la copula o i verbi “leggeri”), dotate di un significato generale e di un potenziale grammaticale, quindi di una forte e naturale predisposizione alla grammaticalizzazione. Questo materiale viene riutilizzato dagli apprendenti, ricombinato in nuove coppie forma-funzione, che hanno in genere carattere provvisorio e che, di norma, scompaiono non appena la finitezza emerge nei verbi lessicali (Bernini 1989; Julien et al. 2016): a meno che l'input (compreso quello scritto) non rimanga troppo scarno perché l'apprendente sia in grado di processare adeguatamente i costituenti morfologici e, in questo caso, le costruzioni mostrano una maggiore persistenza. Una condizione tipica di input ridotto è quello provocato dall'alfabetismo limitato, che va quindi considerato come una condizione che favorisce la stabilizzazione delle soluzioni analitiche per l'espressione delle categorie del verbo.

Un altro aspetto cruciale di cui tenere conto è il carattere inerentemente discorsivo di queste costruzioni, che emergono nell'interazione con gli italofoeni inizialmente “per

ovviare a deficit momentanei di espressione” (Bernini 1989), ma che in certi casi, come si è detto, possono stabilizzarsi. La natura di queste interazioni andrà, dunque, approfondita nel prossimo futuro, per esempio tenendo conto del carattere rigidamente monolingue imposto (necessariamente ma un po’ artificialmente), nei contesti di raccolta dati, ad apprendenti plurilingui (D’Agostino 2021; D’Agostino & Mocciaro 2026), avvezzi a pratiche interazionali “fluide”, tipicamente mistilingui, in cui l’italiano occorre solo come lingua franca.

Bibliografia

- Alves, Rui. 2019. The early steps in becoming a writer: Enabling participation in a literate world. In Horst, Jessica S. & von Koss Torkildsen, Jenna (a cura di), *International handbook of language acquisition*, 567–590. London: Routledge.
- Andorno, Cecilia & Rastelli, Stefano (a cura di). 2009. *Corpora di italiano L2: tecnologie, metodi, spunti teorici*. Perugia: Guerra.
- Andringa, Sible & Godfroid, Alina. 2020. Sampling bias and the problem of generalizability in applied linguistics. *Annual Review of Applied Linguistics* 40. 134–142.
- Arnett, Jeffrey Jensen. 2008. The neglected 95%: Why American psychology needs to become less American. *American Psychologist* 63. 602–614.
- Banfi, Emanuele & Bernini, Giuliano. 2003. Il verbo. In Giacalone Ramat, Anna (a cura di), *Verso l’italiano. Percorsi e strategie di acquisizione*, 70–115. Roma: Carocci.
- Becker, Angelika & Dittmar, Norbert & Gutmann, Margit & Klein, Wolfgang & Rieck, Bo & Senft, Gunter & Senft, Ingeborg & Steckner Wolfram & Thielicke, Elisabeth. 1977. *Heidelberger Forschungsprojekt ‘Pidgin-Deutsch Spanischer Und Italienischer Arbeiter in Der Bundesrepublik’: Die Ungesteuerte Erlernung Des Deutschen Durch Spanische Und Italienische Arbeiter; Eine Soziolinguistische Untersuchung*. Osnabrück: Universität Osnabrück.
- Benazzo, Sandra & Starren, Marianne. 2007. L’émergence de moyens grammaticaux pour exprimer les relations temporelles en L2. *Aile* 25. 129–157.
- Bernini, Giuliano. 1989. Strategie di costruzione dei paradigmi verbali in lingua seconda. *Quaderni del Dipartimento di linguistica e letterature comparate – Istituto Universitario di Bergamo* 5. 195–207.
- Bernini, Giuliano. 1990. Lo sviluppo di paradigmi verbali nelle varietà elementari di apprendimento dell’italiano lingua seconda. In Bernini, Giuliano & Giacalone Ramat, Anna (a cura di), *La temporalità nell’acquisizione di lingue seconde*, 81–101. Milano: FrancoAngeli.
- Bernini, Giuliano. 2003. The copula in learner Italian. Finiteness and verbal inflection. In Dimroth, Christine & Starren, Marianne (a cura di), *Information structure, linguistic structure and the dynamics of language acquisition*, 159–185. Amsterdam: John Benjamins.
- Bernini, Giuliano & Giacalone Ramat, Anna (a cura di). 1990. *La temporalità nell’acquisizione di lingue seconde*. Milano: FrancoAngeli.
- Berruto, Gaetano. 2003. Sul parlante nativo (di italiano). In Radaz, Hans-Ingo & Schlösser, Reiner (a cura di), *Donum grammaticorum. Festschrift für Harro Stammerjohann*, 1–14. Berlin: Walter De Gruyter.

- Bigelow, Martha & Tarone, Elaine. 2004. The role of literacy level in second language acquisition: Doesn't who we study determine what we know? *TESOL Quarterly* 38(4). 689–700.
- Block, David. 2003. *The social turn in second language acquisition*. Edinburgh: Edinburgh University Press.
- Blom, Elma, van de Craats, Ineke & Verhagen, Josje. 2013. *Dummy auxiliaries in first and second language acquisition*. Berlin: De Gruyter Mouton.
- Bonvino, Elisabetta. 2005. *Le sujet postverbale. Un étude sur l'italien parlé*. Paris: Ophrys / Maison des Sciences de l'Homme.
- Butt, Miriam. 2003. The light verb jungle. In Aygen, Gülsat & Bower, Claire & Quinn, Conor (a cura di), *Harvard working papers in linguistics. Papers from the GSAS/Dudley House workshop on light verbs*. Vol. 9, 1–49. Harvard: Harvard Working Papers in Linguistics.
- Cardona, Giorgio Raimondo. 2009. *Antropologia della scrittura*. Torino: UTET.
- Cintrón-Valentín, Myrna C. & Ellis, Nick C. 2016. Salience in second language acquisition: Physical form, learner attention, and instructional focus. *Frontiers in Psychology* 7(1284).
- Clahsen, Harald & Meisel, Jürgen & Pienemann, Manfred. 1983. *Deutsch als Zweitsprache. Der Spracherwerb ausländischer Arbeiter*. Tübingen: Narr.
- Coulmas, Florian. 1996. *The Blackwell encyclopedia of writing systems*. Oxford: Blackwell.
- D'Agostino, Mari (a cura di). 2018. *La forza delle lingue, nella migrazione e nella inclusione*. Palermo: Scuola di Lingua italiana per Stranieri.
- D'Agostino, Mari. 2021. 'Noi che siamo passati dalla Libia'. *Giovani in viaggio fra alfabeti e multilinguismo*. Bologna: Il Mulino.
- D'Agostino, Mari & Mocciaro, Egle. 2021. Literacy and literacy practices: Plurilingual connected migrants and emerging literacy. *Journal of Second Language Writing* 51. 100792.
- D'Agostino, Mari & Mocciaro, Egle. 2022. Palermo 2000–2020: Sicilian in old and new migrations. In Goglia, Francesco & Wolny, Matthias (Eds), *Italo-Romance dialects in the linguistic repertoires of immigrants in Italy*, 19–46. Cham: Palgrave Macmillan.
- D'Agostino, Mari & Mocciaro, Egle. 2026 (in stampa). Multilinguismi fluidi, in presenza e a distanza, e mediazioni 'conviviali'. In D'Agostino, Mari (a cura di), *Camminare attraverso tante lingue. Il plurilinguismo in contesti multilingui tra didattica, ricerca e formazione*. AitLA.
- Davies, Alan. 2013. *Native speakers and native users. Loss and gain*. Cambridge: Cambridge University Press.
- Di Pinto, Jessica. 2020. *Aspetti linguistici e psico-sociali del contatto. Uno studio sugli immigrati cinesi residenti nella città di Napoli*. Tesi di dottorato non pubblicata, Università di Napoli «Federico II».
- Douglas Fir Group. 2016. A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal* 100. 19–47.
- Fleta, M. Teresa. 2003. Is-insertion in L2 grammars of English: A step forward between developmental stages? In Liceras, Juana M. & Zobl, Helmut & Goodluck, Helen (a cura di), *Proceedings of the 6th generative approaches to second language acquisition conference (GASLA 2002)*, 85–96. Somerville: Cascadia Proceedings Project.
- Forsberg Lundell, Fanny & Beaulieu, Suzie. 2023. Second Language Acquisition in different migration contexts. *Languages* 8(1).

- Giacalone Ramat, Anna. 1989. Per una caratterizzazione linguistica e sociolinguistica dell'area Walser. In Rizzi, Enrico (a cura di), *Lingua e comunicazione simbolica nella cultura Walser. Atti del VI Convegno Internazionale di Studi Walser*, Anzola d'Ossola, 1989, 37–66. Fondazione Arch. Enrico Monti.
- Giacalone Ramat, Anna. 1992. Grammaticalization processes in the area of temporal and modal relations. *Studies in Second Language Acquisition* 14(3). 297–322.
- Giacalone Ramat, Anna. 1997. Progressive periphrases, markedness, and second-language data. In Eliasson, Stig & Jahr, Ernst Häkon (a cura di), *Language and its ecology. Essays in memory of Einar Haugen*, 261–285. Berlin: Mouton de Gruyter.
- Giacalone Ramat, Anna (a cura di). 2003. *Verso l'italiano. Percorsi e strategie di acquisizione*. Roma: Carocci.
- Godfroid, Aline, & Andringa, Sible. 2023. Uncovering sampling biases, advancing inclusivity, and rethinking theoretical accounts in second language acquisition: Introduction to the special issue SLA for All? *Language Learning* 73(4): 981–1002.
- Heine, Bernd & Claudi, Ulrike & Hünemeyer, Friederike. 1991. *Grammaticalization: A conceptual framework*. Chicago: The University of Chicago Press.
- Henrich, Joseph & Heine, Steven J. & Norenzayan, Ara. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33(2-3). 61–83.
- Huebner, Thom & Carroll, Mary & Perdue, Clive. 1992. The acquisition of English. In Klein, Wolfgang & Perdue, Clive (a cura di), *Utterance structure: Developing grammars again*, 1–121. Amsterdam: John Benjamins.
- Ježek, Elisabetta. 2004. Types et degrés de verbes supports en italien. *Linguisticae investigations*. 27(2): 185–201.
- Ježek, Elisabetta. 2023. Semantic co-composition in light verb constructions. In Pompei, Anna & Mereu, Lunella & Piunno, Valentina (a cura di), *Light verb constructions as complex verbs*, 221–238. Berlin: De Gruyter.
- Jordens, Peter & Dimroth, Christine. 2006. Finiteness in children and adults learning Dutch. In Gagarina, Natalia & Güllow, Insa (a cura di), *Discovering the world of verbs*, 167–195. Dordrecht: Kluwer.
- Julien, Manuela & van Hout, Roeland & van de Craats, Ineke. 2016. Meaning and function of dummy auxiliaries in adult acquisition of Dutch as an additional language. *Second Language Research* 32(1). 49–73.
- Klein, Wolfgang & Perdue, Clive. 1992. *Utterance structure. Developing grammar again*. Amsterdam: John Benjamins.
- Klein, Wolfgang & Perdue, Clive. 1997. The Basic Variety (or: Couldn't natural languages be much simpler?). *Second Language Research* 13(4). 301–347.
- Lorenzová, Kristýna. In prep. *The development of discourse markers in the interlanguages of Czech and Slovak learners of Italian as a non-mother tongue language: a longitudinal study*. Tesi di dottorato, Università Masaryk, Brno.
- Lorenzová, Kristýna. 2023. Lo sviluppo dei marcatori discorsivi nelle interlingue di apprendenti cechi e slovacchi di italiano come lingua non materna. *Italiano LinguaDue* 15(2). 65–83.
- Lüpke, Friederike. 2017. African(ist) perspectives on vitality: Fluidity, small speaker numbers, and adaptive multilingualism make vibrant ecologies (Response to Mufwene). *Language* 93(4). e275–e279.
- Mocciaro, Egle. 2020. *The development of L2 Italian morphosyntax in adult learners with limited literacy*. Palermo: Palermo University Press.

- Mocciaro, Egle. 2019. Emerging constructions in the L2 Italian spoken by low literate migrants. *Languages* 4(4). 86.
- Mocciaro, Egle & Young-Scholten, Martha. 2026 (in pubblicazione). Literacy and adults' acquisition of morphosyntax. In Herschensohn, Julia et al. (a cura di), *The handbook of second language acquisition* (2nd ed.). Cambridge: Cambridge University Press.
- Monsen, Marte & Bordal Stein, Guri. 2022. *Language learning and forced migration*. Bristol: Multilingual Matters.
- Ortega, Lourdes. 2019. SLA and the study of equitable multilingualism. *The Modern Language Journal* 103. 23–38.
- Perdue, Clive. 1993. *Adult language acquisition: Cross-linguistic perspectives*, 2 voll. Cambridge: Cambridge University Press.
- Pfaff, Carol W. 1992. The issue of grammaticalization in early German second language. *Studies in Second Language Acquisition* 14. 273–296.
- Plonsky, Luke. 2023. Sampling and generalizability in Lx research: A second-order synthesis. *Languages* 8(1). 75.
- Saleem, Amjad. 2015. Does memorization without comprehension result in language learning? Cardiff: Cardiff University PhD dissertation.
- Saleem, Amjad. 2018. Memorization without comprehension: A window onto the “extremities” of the capability of human brain. *Al-Idah* 36(1). 25–40.
- Savy, Renata. 1999. Riduzioni foniche nella morfologia del sintagma nominale nel parlato spontaneo: indagini quantitative e aspetti strutturali. In Benincà, Paola & Mioni, Alberto & Vanelli, Laura (a cura di), *Fonologia e morfologia dell'italiano e dei dialetti d'Italia. XXXI Congresso della Società di Linguistica Italiana, Padova 25-27 settembre 1997* (SLI 41), 201–221. Roma: Bulzoni.
- Starren, Marianne. 2001. *The second time: The acquisition of temporality in Dutch and French as a second language*. Utrecht: LOT.
- Street, Brian V. 1984. *Literacy in theory and in practice*. Cambridge: Cambridge University Press.
- Tarone, Elaine. 2010. Second language acquisition by low-literate learners: An understudied population. *Language Teaching* 43(1). 75–83.
- Tarone, Elaine & Bigelow, Martha. 2005. Impact of literacy on oral language processing: Implication for SLA research. *Annual Review of Applied Linguistics* 25. 77–97.
- Tarone, Elaine & Bigelow, Martha, & Hansen, Kit. 2009. *Literacy and second language oracy*. Oxford: Oxford University Press.
- Tarone, Elaine & Swierzbins, Bonnie, & Bigelow, Martha. 2006. The impact of literacy level on features of interlanguage in oral narratives. In Baldwin, Thomas D. & Selinker, Larry (a cura di), *Interlanguage: Current thought and practices. Rivista di Psicolinguistica Applicata* 6. 65–77.
- UIS. 2017a = UNESCO Institute for Statistics. 2017a. [More than one-half of children and adolescents are not learning worldwide](#). Fact sheet n. 46, September 2017.
- UIS. 2017b = [Literacy rates continue to rise from one generation to the next](#). Fact sheet n. 45, September 2017.
- UNESCO. 1958. Recommendation concerning the international standardization of educational statistics. In records of the General Conference, Tenth Session, Paris, 1958: Resolutions, Paris, UNESCO.
- Vainnika, Anne & Young-Scholten, Martha. 1998. Morphosyntactic triggers in adult SLA. In Beck, Marie-Louise (a cura di), *Morphology and its interfaces*, 89–113. Amsterdam: John Benjamins.

- Vainikka, Anne & Young-Scholten, Martha. 2007. The role of literacy in the development of L2 morpho-syntax from an Organic Grammar perspective. In Faux, Nancy (a cura di), *Low-educated adult second language and literacy acquisition (LESLLA): Research, policy, and practice. Proceedings of the second annual forum*, 123–148. Richmond: The Literacy Institute at Virginia Commonwealth University.
- Vainikka, Anne & Young-Scholten, Martha & Ijuin, Colleen & Jarad, Samawal. 2017. Literacy in the development of L2 English morphosyntax. In Sosiński, Marcin (a cura di), *Literacy education and second language learning for adults: research, policy and practice*, 239–250. Granada: Universidad de Granada.
- Voghera, Miriam. 2017. *Dal parlato alla grammatica*. Roma: Carocci.
- von Stutterheim, Christiane. 1986. *Temporalität in der Zweitsprache*. Berlin: Mouton de Gruyter.
- Wiley, Terrence G. 2005. Second language literacy and biliteracy. In Hinkel, Eli (a cura di), *Handbook of research in second language teaching and learning*, 529–544. Mahwah, NJ: Erlbaum.

Reference in teenagers' narratives between plurilingualism and restandardisation

Nicola Perugini

Università di Bologna

nicola.perugini6@unibo.it

Eleonora Zucchini

Masaryk University, Brno

zucchini@phil.muni.cz

Abstract

This study examines referential choices in narrative texts in Italian, written by monolingual and bilingual adolescents; it adopts a double perspective, investigating both strategies for reference tracking and sociolinguistic variation in pronoun choice. Results show that referential expressions (REs) are used adequately across all groups, with monolinguals and simultaneous bilinguals displaying similar patterns, while sequential bilinguals using pronouns less frequently. In terms of pronoun choice, adolescents overwhelmingly favor neostandard forms for both subject and indirect-object pronouns, while standard variants have nearly disappeared. Notably, the study provides evidence of an interesting linguistic tendency among Italian teenagers: the use of *li*, lit. 'them', as indirect object for both genders and numbers.

1. Introduction

This study investigates the referential choices of monolingual and plurilingual Italian adolescents in narrative written productions. The development of cohesive strategies for the construction of extended texts seems to continue throughout school career and to be a linguistic domain of distinction between speakers with monolingual family background and those with multilingual repertoires (Hendriks et al. 2014; Andreou et al. 2020 among others). Furthermore, not much is known about the age range from the end of primary school to adulthood, since only few studies on Italian have involved groups of adolescents. We examine how adolescents encode referents, focusing on the use of pronouns, zero anaphora, and full noun phrases. Given the variability of personal pronoun forms in Italian, we also analyse whether the different groups involved in the study behave differently regarding pronoun choice in terms of standard and non-standard variants.

2. Theoretical framework

2.1 Reference in narrative texts

To construct a well-formed narrative text, that can be easily followed by the readers, the narrator must guide them in the construction of a coherent discourse model and in the identification of the essential elements that shape the text's progression. In narratives, these reference points include characters, temporal and spatial settings, and events (Chini 2005).

By anchoring these elements within the text, the narrator ensures cohesion and enhances comprehension, allowing the reader to follow the story's development with ease.

Referential expressions (REs, henceforth) play a crucial role in this regard, by structuring the narrative world, maintaining continuity, and managing information flow throughout the discourse. When selecting an appropriate RE from a range of linguistic options (e.g., "a cat," "the cat," "that," "it"), the speaker must consider the referent's accessibility to the listener, which involves assessing the assumed cognitive status of a referent within the addressee's mental representation of the discourse (e.g., Ariel 2001; Arnold 2010; Chafe 1976; Gundel & Hedberg & Zacharski 1993; Prince 1981). In turn, addressees identify referents by integrating verbal cues from the speaker while drawing on their presumed shared knowledge (Clark 1996). This collaborative process is evident not only in synchronous communication, such as face-to-face conversation (Peeters et al. 2020), but also in asynchronous contexts, such as written text construction.

With regard to characters, which constitute the focus of this study, the narrator must introduce them in a way that establishes sufficient common ground with the interlocutor. As the text unfolds, the accessibility status of referents may shift based on both the interlocutor's knowledge (i.e., whether the character is already given or identifiable) and their attentional relevance at a given point in the discourse (i.e., whether the referent is in focus). For example, during transitions between narrative episodes, the addressee's attention must be redirected, either shifting from one character to another, re-establishing reference to a previously mentioned character, or introducing a new character interacting with those already present in the discourse. These transitions give rise to what are known as referential movements, which play a crucial role in maintaining discourse cohesion (Chini 2005). Within the domain of person reference, three primary referential movements are traditionally identified: the introduction of one or more characters, the maintenance of their reference throughout the narrative, and their reintroduction (Chini 2005; Colozzo & Whitley 2014). Each of these movements corresponds to varying degrees of referent accessibility, and require speakers to choose the adequate referential expression accordingly (Ariel 1991, p. 443).

In Italian, the introduction of new characters, which are minimally accessible, generally requires full forms, typically indefinite noun phrases (NPs with an indefinite article such as *un gatto* 'a cat', partitive articles, *dei gatti* 'some cats', numerals, *due gatti* 'two cats', or indefinite quantifiers *alcuni gatti* 'a few cats'). In contrast, the maintenance of already known characters tends to rely on leaner forms, such as pronouns or zero anaphora: verbal agreement and clitic pronouns are preferably used for highly accessible referents, particularly those present in the immediate textual context, whereas strong pronouns are primarily employed for contrastive or emphatic anaphora (Chini 2005). When referring to previously mentioned characters with low discourse

continuity, Italian uses definite descriptions or NPs with demonstratives, especially when disambiguation is needed among multiple characters with similar degrees of accessibility. However, in cases where multiple referents compete, lighter forms such as pronouns may still be employed, depending on the narrative salience of the characters: the main character is often the most likely referent to be identified at specific points in the narration (Schimcke et al. 2020).

2.2 Personal pronouns in standard and neostandard Italian

Among referring expressions in Italian, the paradigm of third person pronouns is particularly interesting from a (socio)linguistic point of view. In fact, it is extremely rich: it includes tonic and clitic forms, but also polysemic forms and forms competing for the same function or in the process of acquiring new ones. Moreover, pronoun choice is subject to factors of different nature: it responds to case, number and gender agreement, to sentence syntax and information structure (i.e. post-verbal position, presence of coordinated clauses and focalization) (Cardinaletti 2021).

The pronominal system of standard Italian is summarised in Tables 1 and 2 (Serianni 1988: 169; 175):

Tonic pronouns	Subject		Direct object and obliques	
	[+Hum]	[-Hum]	[+Hum]	[-Hum]
M.SG.	<i>egli, lui</i>	<i>esso</i>	<i>lui</i>	<i>esso</i>
F.SG.	<i>ella, lei</i>	<i>essa</i>	<i>lei</i>	<i>essa</i>
M.PL.	<i>essi, loro</i>	<i>essi</i>	<i>loro</i>	<i>essi</i>
F.PL.	<i>esse, loro</i>	<i>esse</i>	<i>loro</i>	<i>esse</i>

Table 1. Tonic pronouns in standard Italian

Clitic pronouns	Direct object	Indirect object
M.SG.	<i>lo</i>	<i>gli</i>
F.SG.	<i>la</i>	<i>le</i>
M.PL.	<i>li</i>	<i>loro</i>
F.PL.	<i>le</i>	<i>loro</i>

Table 2. Clitic pronouns in standard Italian

According to previous studies, such paradigm is going through a process of restandardisation, along with several other morphosyntactic structures (Sabatini 1985, 1990; Berruto 2012[1987]). Specifically, subject pronouns *egli* and *ella* have been relegated to very formal written contexts (see Ballarè in press), whereas in less formal styles the object and oblique pronoun *lui* and *lei* are preferred also in subject position; moreover, *esso/a/i/e* are used more rarely and are being replaced by *lui*, *lei* and *loro*, across all functions. Clitic pronouns are included in the spectrum of variation as well, since the masculine singular dative *gli* is expanding and is employed quite regularly also

to refer to plural and feminine referents in informal and spoken varieties; studies have also detected, though more rarely, cases of the opposite phenomenon, namely *le* used to refer to masculine referents (Benincà & Penello 2009).

It is possible to say that, thanks to the restandardisation process, the factors driving the set of options of REs in Italian is further expanded, since extralinguistic factors, such as medium (spoken vs written) and register, play a role alongside linguistic ones.

2.3. Referential choices in multilingual speakers: a complex interaction of linguistic and cognitive factors

Selecting adequate REs based on a referent's informational status is a complex task in narrative discourse and requires speakers to access a full range of linguistic forms and map them onto their adequate discourse functions (e.g., introducing a new referent, maintaining reference to previously mentioned entities). This challenge becomes even more pronounced for multilingual speakers, who must handle this process across two or more languages. However, research on referential choices in multilinguals presents a nuanced and sometimes contradictory picture. While some studies find no significant differences between monolingual and multilingual speakers, suggesting that referential choices are governed by universal principles (Andreou et al. 2015; Topaj 2010), others report a tendency among multilingual children to be over-informative when compared to their monolingual peers. This happens when a speaker or writer provides more information than necessary for the addressee to correctly identify a referent, for example when the referent is already highly accessible or uniquely identifiable within the discourse context, making additional specification redundant, if not used for emphasis or stylistic effect. Over-informativeness may manifest through the use of full noun phrases instead of pronouns, excessive modifiers, or explicit descriptions when a simpler reference would suffice. This pattern has been attributed to a range of interrelated linguistic and cognitive factors.

One key factor is the dominance in each language of the individual repertoire, which encompasses both exposure and proficiency (Torregrossa et al. 2021). In a study on bilingual children (aged 7-13) with diverse language combinations, Torregrossa and colleagues (2021) found that greater dominance in the non-target language correlated with a stronger tendency toward over-specification. Conversely, more balanced bilinguals exhibit a greater tendency toward under-specification (Torregrossa et al. 2019; Fiorentini & Gianollo 2021).

Less balanced bilinguals may exhibit slower and less efficient processing of referential options in their non-dominant language, leading them to favour structurally simpler and more easily retrievable forms, potentially at the expense of pragmatic-discursive appropriateness (Torregrossa et al. 2018). This holds true especially for sequential bilingual speakers (individuals exposed to a second language later than age 6) who might have more limited morphosyntactic resources (i.e. a less wide range of linguistic options)¹. Together with morphosyntactic competence, as Arnold (2010, Arnold & Zerkle 2019) discusses, cognitive load also plays an important role in reference production, with speakers under higher cognitive demands showing a

¹ As shown by Schulz and Grimm (2018), differences between simultaneous and sequential bilinguals do not necessarily persist into adulthood. Their study suggests that sequential bilinguals can catch up with simultaneous bilinguals over time, especially in domains acquired early in development.

preference for more explicit referring expressions. Due to their linguistic limitations, during production, early L2 speakers might choose the cognitively lighter path, often leading to faster but less structurally complex processing strategies: the choice of over-informative forms would appear the lightest from a cognitive point of view, since it optimizes the underlying processing mechanisms of morphosyntactic form selection (De Lange et al. 2009).

This kind of explanation points to a general effect of bilingualism on information processing speed. Research suggests that bilingualism may influence referential strategies through both cognitive and communicative effects: the need to inhibit options from the non-target language, the lower exposure to and proficiency in the target language may slow down referential processing (Torregrossa et al. 2021; Perugini 2025). Additionally, some studies (Sorace 2011; 2016) suggest that bilinguals exhibit a heightened sensitivity to the interlocutor's needs, leading to a stronger preference for avoiding ambiguity and, consequently, a higher tolerance for over-informativeness compared to monolinguals (Sorace 2016).

While the specific language pair in a bilingual's repertoire may shape certain strategies, (Lindgren & Bohnacker 2024) studies show that bilinguals, regardless of whether their languages are structurally similar (e.g., Greek-Spanish: Georgopoulos 2017; Italian-Spanish: Leonetti-Escandell 2024) or different (e.g., Greek-English, Italian-English: Tsimpli et al. 2004; Russian-Greek: Tsimpli & Sorace 2006), tend to over-specify referents. This tendency appears to stem more from processing demands and linguistic experience than from grammatical transfer, suggesting that cross-linguistic influence does not significantly affect the degree of informativeness in referential choices.

Overall, the intricate interplay of linguistic dominance, cognitive processing, morphosyntactic competence, and language pair effects highlights the complexity of referential choices in multilingual speakers.

2.4 Sociolinguistic variation and teenagers' language

The topic of language learning has been addressed also in a sociolinguistic perspective, particularly to study the acquisition of patterns of variation.²

According to initial work by Labov (1964; 2001) young children only learn basic grammar and begin adopting local vernacular through contact with peers in school; only from adolescence onwards, they gradually become aware of the social meaning of language and grow progressively capable of modulating speech according to social norms (*Sequential view*, see also De Vogelaer et al. 2017). On the contrary, successive studies describe the acquisition of variability as setting off immediately after birth (*Parallel view*; Eckert 1997; Johnson 2016; Roberts 1997 among others).

Up to now, the evidence in favour of the two views is not sufficient to give a definite answer to the question (Johnson & White 2019): children have been proven to be sensitive to language-internal variability from a very young age, but studies focus mostly on phonological variables; moreover, the acquisition of socially-dictated variation, which is surely a slower process, still needs further investigation and also

² Unfortunately, there is no space here to summarize all the relevant theories and approaches; see Nardy et al. (2013) and Johnson & White (2019) for thorough reviews.

requires a deeper understanding of children's comprehension of social matters (Barbu et al. 2013; Foulkes et al. 2005; Johnson & White 2019; Smith et al. 2007).

Most research on the acquisition of variation patterns has involved children of a very young age; the language of adolescents has been investigated in comparison to adult's rather to unveil the trajectories of linguistic innovations (Eckert 2004; Labov 1964, Labov 2001; Tagliamonte & D'Arcy 2009; Tagliamonte 2016 among others), since teenagers have an important role in language change: as a matter of fact, they have been proven to adopt vernacular variants very frequently and push change forward at an increasingly quicker pace (*adolescence peak*), without full awareness of social factors. The phase of change in which the variable is found, and factors such as gender and social background of the speaker also play a role (Kohn 2014; Tagliamonte & D'Arcy 2009).³

3. Data and methodology

3.1 Data

The data analysed in this paper consist of 48 narrative texts written in school settings by lower secondary school students. They were extracted from a corpus of 345 texts (126,095 tokens) collected between 2020 and 2023 in schools located in the cities of Bologna and Modena in Northern Italy. The texts can be considered examples of authentic school writing, since they were written during a school test and not elicited on purpose for this study.

The texts discuss different topics and are of various genres,⁴ among which creative narratives (114 texts); in addition, the corpus includes 12 book reviews, always containing a summary of the plot which can be considered a short narrative text.

3.2 Participants

The students involved in the data collection filled out a questionnaire aimed to collect metadata regarding basic biographical information, linguistic repertoire and practices, and migration background. All the students involved, at the moment of data collection, were attending the last year of lower secondary school and were between 13 and 14 years of age.

Table 3 summarizes the information regarding the country of origin of the students and their families. Thanks to the information obtained through the questionnaires, it was possible to find out that 28 narrative texts were written by students with a migration background, of which 8 were born abroad, and moved to Italy between 6 and 10 years of age, and 20 were born in Italy or moved there very early (before 4 years old), and have parents of foreign origins (from Albania, Morocco and Bangladesh, among others).

³ For methodological aspects, see Tagliamonte & D'Arcy (2009, p. 61ff.) about the apparent-time approach, and Kohn (2014, p. 5ff.) about the limitations of such approach and the relation between individual change and community-led variation.

⁴ For more details on the data collection and on the corpus, see Zucchini (2025).

	no. students	%
Students and parents born in Italy	221	64%
Students born in Italy and parents born abroad	63	18,3%
Students born abroad and parents born abroad	24	7%
One parent born abroad	23	6,7%
No information	13	3,7%
Tot.	345	100%

Table 3. Participants according to family background (entire corpus).

The students of the first group will be considered sequential bilinguals, as their acquisition of Italian as an additional language, mainly in school settings, began after the age of six. The students in the second group will be considered simultaneous bilinguals, since they all report speaking a language other than Italian at home but have presumably been exposed to Italian through educations since an early age. Students born in Italy to Italian parents will be considered monolinguals, as none of them reported speaking any other language at home or from an early age.

Two samples of the corpora were therefore compiled: the first contains 40 narrative texts, 20 written by students born in Italy with migration background (considered simultaneous bilinguals) and 20 by students born in Italy from Italian parents, selected randomly (monolinguals); the second sample contains 24 texts, 8 written by simultaneous bilinguals, 8 by sequential bilinguals and 8 by monolingual Italian speakers. Tables 4 and 5 summarize the size of the two samples.

First sample	n. texts	n. tokens
Monolingual	20	9.400
Simultaneous bil.	20	8.292
Tot.	40	17.692

Table 4. Size of the first sample

Second sample	n. texts	n. tokens
Monolingual	8	3.657
Simultaneous bil.	8	2.525
Sequential bil.	8	2.925
Tot.	24	9.107

Table 5. Size of the second sample

3.3 Coding

To investigate the management of reference in the texts of the sample, a coding protocol was developed based on the literature (see 2.1), with three levels:

1. Function: introduction or maintenance;
2. Type of form: full NP, personal pronoun, zero anaphora;
3. Informativeness of RE: optimal, overinformative, underinformative.

Only for pronouns, also number, gender, person and function (subject, dative or other function) was coded since it was necessary for the sociolinguistic section of this paper (see section 5).

The first level concerns the function performed by the RE, which can either be that of introducing a referent for the first time, as in the case of *due ragazzi*, ‘two boys’ in example (1), or that of mentioning again a referent that has already been introduced (maintenance function), as in the case of the zero anaphora *erano* ‘were’ and *sembravano* ‘seemed’.

- 1) *Escono **due ragazzi** [introduction] che sembrano più o meno della mia età, o anche un po' più grandi, insieme **hai loro genitori** [introduction,], **erano** [maintenance] molto eleganti e **sembravano** [maintenance] anche molto amichevoli.* (MCA18)
‘**Two boys** appear from the house, that look like they are my age or maybe a bit older, **with their parents**, (they) **were** very elegant and **seemed** very friendly’

Annotating the function performed by REs is essential to later verify whether the form chosen by the author is adequate at the discourse-pragmatic level. It should be noted that only two values (introduction or maintenance) were assigned at this level, since the texts considered are fairly short (400 words per text, on average) and the characters involved in the narration are not very numerous: therefore, the function of reintroduction (considered in other studies, see Pallotti, Borghetti & Rosi 2020) was not taken into account, assuming that characters introduced at the beginning of the text would remain active in the reader’s memory until the end.

Levels 2 concerns the form; some examples of how the REs were annotated can be found in (2):

- (2) *Il giorno dopo **io** [pronoun] e **mio padre** [NP] siamo andati a fare la spesa e dopo essere tornati **ho** visto [zero anaphora] **Marco** [NP] e **Nicholas** [NP] portare dei scatoloni dentro la casa, allora **io** [pronoun] **gli** [pronome,] chiesi se **potevo** [zero anaphora] **aiutarli** [pronoun] **loro** [pronoun] dissero [...]* (MCA18)
‘The next day **my father** and **I** went shopping and after coming home (I) **saw** Marco and Nicholas take some boxes in the house so I asked **them** if (I) **could** help them **they** said [...]

As mentioned before, observing the type of form is fundamental for assessing whether the RE is adequate; as a matter of fact, the annotation of the second level, if considered in relation to the first, prepares the ground for the third level, namely informativeness, which will reveal difficulties in managing reference. Thanks to the annotation of the

type of form, it will also be possible to observe whether the three populations involved in the study use the same range of linguistic strategies or if there are differences.

Finally, the presence of underinformative or overinformative referential choices was annotated (level 3). A RE was considered underinformative if the form and determiner chosen were not sufficient to allow the reader to retrieve the referent from memory, that is, when the form chosen would be more adequate for a more accessible referent: for example, in (3a) a zero anaphora is used to retrieve a referent that has not yet been introduced⁵.

On the contrary, an expression was considered overinformative when the chosen form corresponds to a lower level of accessibility, thus requiring a less informative and less marked form (example 3b). It is quite evident that in example (3b) repeating the characters' first names in full is not necessary, since the last mention occurs in the previous sentence and performs the same syntactic role; thus, zero anaphora would have been sufficient.

- (3) a. *La prima missione se invece veniva fallita non se ne potevano fare altre e **venivano** trasferiti sul pianeta dei buoni a nulla.* (MDA13)
 'If the first mission failed, no more could be done and (they) **were** transferred to the planet of the good for nothing'.
- b. ***Sofia** e il **GFG** volevano fermare e catturare i giganti che mangiano gli uomini. **Sofia** e il **GFG** volevano convincere la regina di Inghilterra a trovare un rimedio* (MVL4).
 '**Sofia** and the **BFG** wanted to stop and captures the giants who ate humans. **Sofia** and the **BFG** wanted to persuade the queen of England to find a solution'.

This level of annotation will make it possible to verify whether the three populations investigated have reached the same level of competence given the same number of years of schooling, and whether there are particularly problematic contexts for all or for only one of the three groups.

Finally, a short section (section 5) is dedicated to restandardisation trends in pronoun choice, where the methodology of variable rules (Labov 1973) will be adopted for the data analysis. More specifically, third person pronouns, already extracted and coded, will be grouped by syntactic function and gender and number of the referent; then, the percentage of the pronouns performing the same function will be calculated, to verify which one is the preferred choice in general and for each group.

Before proceeding with the analysis, it should be added that, due to the extremely high number of inanimate referents mentioned in the texts, only human ones were considered, following a methodology already adopted in other studies (Chini 2005; Pallotti et al. 2020 among others).

⁵ (Under/over) informativeness was evaluated following established annotation guidelines, with doubtful cases resolved through discussion among annotators, in line with previous descriptions of referential expression use in Italian (e.g., Carminati, 2002; Andreou et al. 2020).

4. Results

4.1. Distribution of REs in the first sample

This section considers the first subsample of texts extracted from the entire corpus (see section 3.2), consisting of 20 texts produced by simultaneous bilingual and 20 by monolingual students. The analysis presented here aims to investigate the distribution of REs used by students in their narrative texts for character introduction and maintenance. Table 6 provides raw figures and percentages of REs by group and by discourse function.

Table 6 shows also that for referent introduction, both bilingual and monolingual students predominantly use full NPs (71% and 69%, respectively), while null pronouns (22% and 17%) and overt pronouns (14% and 7%) are used less frequently. This suggests a similar distributional pattern across groups, although monolinguals employ slightly more pronouns overall. For referent maintenance, both groups strongly prefer zero anaphora (54%, 56%), which is not surprising: due to the brevity of the texts and the limited number of characters, referents generally exhibit high discourse continuity. Instances of character shifts or referential competition requiring disambiguation are rare. The use of pronouns and full NPs in this function is also comparable between groups (monolinguals: 25%, 21%; bilinguals: 24%, 20%).

	Form	Introduction	Maintenance	Total
Simultaneous bilingual	Pro	13 (7%)	263 (24%)	276
	NP	129 (71%)	219 (20%)	348
	Zero	39 (22%)	601 (56%)	640
	Tot.	181	1083	1264
Monolinguals	Pro	23 (14%)	282 (25%)	305
	NP	117 (69%)	245 (21%)	362
	Zero	29 (17%)	625 (54%)	654
	Tot.	169	1152	1321

Table 6. Distribution of REs in the first sample

A Chi-square test was conducted to compare the distribution of referential forms (pronouns, noun phrases, zero anaphora) across functions (introduction and maintenance) between simultaneous bilingual and monolingual students. The differences were not statistically significant, neither for the introductions ($\chi^2 = 4.42$; $df = 2$; $p > 0.1$), nor in the maintenance function ($\chi^2 = 0.45$; $df = 2$; $p > 0.7$), indicating similar referential strategies across the two groups in Italian written narrative texts.

4.1.1. Informativeness of REs in the first sample

Turning to analyse the informativeness of the referential choices made by the two groups of participants, the raw figures and percentages of the forms coded as optimal, underinformative and overinformative are reported below in Table 7.

From the table, it emerges that both groups produce more overinformative forms than underinformative ones, but in relatively low percentages (around 4% in both groups). No significant differences are observed between the two groups for any of the levels of informativeness considered, and in fact, such differences were not found to be significant in the Chi-Square test ($\chi^2 = 2.04$; $df = 2$; $p > 0.3$).

	Optimal REs	Underinformative REs	Overinformative REs
Simultaneous bilingual	1186/1264 (94%)	27/1264 (2%)	51/1264 (4%)
Monolinguals	1229/1321 (93%)	40/1321 (3%)	52 /1321 (4%)

Table 7. Informativeness in the first sample

Some instances of non-optimal REs across the two groups are illustrated below. As for underinformative REs, they occur in the texts at very low frequency and the few instances are distributed across all forms, with a higher frequency of ambiguous uses of zero anaphora, as shown in example (4), produced by a simultaneous-bilingual student.

- (4) *Quindi abbiamo deciso di andare da mio padre che era uno che sapeva di queste cose, e ci ha detto che **era** un super hacker di nome Anonymen.* (MVI23)
‘So we decided to go to my father who was someone who knew about these things, and (he) told us that (he) **was** a super hacker named Anonymen’.

In example (4), the use of zero anaphora *era* (‘was’, in bold) indicates high referential accessibility. The most accessible referent, based on recency of mention, is the father. However, this interpretation conflicts with the broader discourse coherence, as it becomes clear that the referent of the zero subject (the “super hacker”) is not the father. The referential choice made here momentarily misguides the reader, creating a mismatch between local accessibility cues and global coherence.

Example (5) taken from a text produced by a monolingual student, features an interesting use of the clitic pronoun *gli* (‘to him/her/them’; see sections 2.2 and 5). In this case, the ambiguity does not stem from the choice of a pronoun over a full NP or zero anaphora, but rather from the selection of a pronominal form underspecified in terms of gender and number, which does not facilitate the identification of the referent.

- (5) *Scoppiai a piangere: era **mia mamma** corsi giù e la abbracciai forte, dalla macchina uscì anche **mio papà**, dopo gli abbracci **gli** chiesi come avessero fatto a trovarmi **loro** non risposero dissero solo di salutare.* (MCA10)
‘I burst into tears: it was **my mom**, I ran down and hugged her tightly, **my dad** also got out of the car, after the hugs I asked **him/them** how they had managed to find me **they** did not answer only said to say goodbye’.

In example (5), referential ambiguity arises from the use of the Italian clitic pronoun *gli*, which can refer to both singular and plural masculine indirect objects (see 2.2). In this case, the referent of *gli* could plausibly be interpreted as the father, who is mentioned immediately prior in the discourse, but also as “the parents,” since both the mother and

father are present in the scene. The reader is not provided with sufficient linguistic cues to unambiguously resolve the reference at this point in the text. Only in the following sentence (*loro non risposero* “they did not answer”) where the subject is explicitly referred to as *loro* (“they”), does the intended referent become clear.

This use highlights the need to consider processes of restandardization within the Italian pronominal system when analysing referential choices, an issue that will be addressed in section 5. Moreover, in this example, the referential ambiguity reflects a broader pattern observed in the data, whereby most referential issues stem from the presence of multiple potential referents, particularly in plural forms such as *loro* (“they”) or *noi* (“we”), where it is often unclear which characters present in the scene are included in the referential set in addition to the speaker. These cases reveal a challenge in the management of plural reference, especially when the discourse does not provide explicit cues to delimit the intended group membership.

4.2 Distribution of REs in the second sample

This section considers the narratives written by the students in the second subsample, which consists of 24 texts, equally distributed across three groups: 8 texts produced by simultaneous bilinguals (SimBil), 8 by monolinguals (Mono) and 8 written by sequential bilinguals (SeqBil) (see section 3.2). The analysis presented below aims to investigate the distribution of REs used by students in their narrative texts for character introduction and maintenance. Table 8 provides raw figures and percentages organised by group and by discourse function.

	Form	Introduction	Maintenance	Total
Simultaneous bilingual	Pro	4 (6%)	67 (23%)	71
	NP	46 (71%)	62 (21%)	108
	Zero	15 (23%)	165 (56%)	180
	Tot.	65	294	359
Monolinguals	Pro	8 (11%)	118 (26%)	126
	NP	49 (65%)	97 (21%)	146
	Zero	18 (24%)	238 (53%)	256
	Tot.	75	453	528
Sequential bilinguals	Pro	5 (10%)	48 (13%)	53
	NP	33 (65%)	53 (14%)	86
	Zero	13 (25%)	278 (73%)	291
	Tot.	51	379	430

Table 8. REs in the second sample

Table 8 shows that all groups exhibit the same distributional pattern, preferring full NPs for the introduction of new referents and zero anaphora for maintenance, albeit with differing percentages. While all groups show similar percentages in using full noun phrases to introduce referents (71%, 65%, 65%), for the maintenance function,

sequential bilingual speakers show a stronger preference for zero anaphora (73%) compared to the other groups (56%, 53%).

The use of pronouns shows more variability across groups, with the monolingual group employing them more frequently overall. Pronouns are primarily employed by all three groups for referent maintenance, with similar percentages among simultaneous bilinguals (23%) and monolinguals (26%), and much less frequently by sequential bilinguals (13%). Overall, while simultaneous bilinguals and monolinguals appear to follow comparable referential strategies, sequential bilingual speakers tend to differ in terms of frequency of use rather than in the type of pattern employed.

In order to verify whether the differences in distribution of REs across the three groups were significant, a Chi-square test was conducted in pairwise comparison for each function. For the introduction function, the results indicate that there is no significant difference across any of the comparisons (Mono vs SimBil: $\chi^2 = 0.99$, $df = 2$, $p > 0.5$; Mono vs SeqBil: $\chi^2 = 0.05$, $df = 2$, $p > 0.9$; SimBil vs SeqBil: $\chi^2 = 0.71$, $df = 2$, $p > 0.7$). As for maintenance, differences in the referential strategies of simultaneous bilinguals and monolinguals do not reach the level of significance (Mono vs SimBil: $\chi^2 = 1.20$, $df = 2$, $p = 0.5$). However, when compared to monolingual and simultaneous peers in the maintenance function, sequential bilinguals always show highly significant differences in their referential patterns (Mono vs SeqBil: $\chi^2 = 39.25$, $df = 2$, $p < 0.001$; SimBil vs SeqBil: $\chi^2 = 22.29$, $df = 2$, $p < 0.001$), confirming that the referential behaviour of this group is substantially different from the others.

To better understand the source of the significant group differences, we examined standardized residuals (SRs) from the Chi-square analyses, focusing only on the maintenance function and on comparisons involving sequential bilinguals. These residuals indicate which specific response types (in terms of selected forms) are over- or under-represented in each group relative to chance expectations, shedding light on the distinct referential strategies employed (values greater than ± 2 are typically considered statistically noteworthy, highlighting which cells contribute most to the overall chi-square effect).

In the comparison between monolinguals and sequential bilinguals, the residuals revealed clear contrasts: monolinguals significantly overuse pronouns ($SR = +2.91$) and underuse zero anaphora ($SR = -2.56$), sequential bilinguals, conversely, significantly underuse pronouns ($SR = -3.18$) and overused zero anaphora ($SR = +2.80$). A similar pattern emerged in the comparison between simultaneous and sequential bilinguals: simultaneous bilinguals overused pronouns ($SR = +2.36$) and underused zero anaphora ($SR = -2.05$), while sequential bilinguals again underuse pronouns ($SR = -2.08$) and showed a marginal overuse of zero anaphora ($SR = +1.81$). Taken together, these results confirm that the use of pronouns is the main point of divergence across groups, with sequential bilinguals adopting a referential strategy that favors zero anaphora and disfavors pronouns, setting them apart from both monolinguals and simultaneous bilinguals.

4.3 Informativeness of REs in the second sample

We now turn to analysing the informativeness of the referential choices made by the three groups of participants of the second subsample. The absolute occurrences and percentages of the forms coded as optimal, underinformative and overinformative are reported in Table 9.

	Optimal REs	Underinformative REs	Overinformative REs
Simultaneous bilinguals	338/359 (94%)	8/359 (2%)	13/359 (4%)
Monolinguals	500/528 (95%)	16/528 (3%)	12/528 (2%)
Sequential bilinguals	426/430 (99%)	0/430 (0%)	4/430 (1%)

Table 9: informativeness in the second sample

All groups exhibit a strong tendency to produce optimal REs, with proportions exceeding 95%. Sequential bilinguals show the highest rate, reaching 99%. Underinformative REs were absent among sequential bilinguals (0%), while they appear at low frequencies (2–3%) in the other two groups. Overinformative REs are slightly more frequent among simultaneous bilinguals (3%) compared to monolinguals and sequential bilinguals.

Three pairwise Chi-square tests were conducted to examine whether there were significant differences in the level of informativeness of REs across the three groups. The test between the simultaneous bilinguals and monolinguals indicates that there is no significant difference in the distribution of forms according to their informativeness between the two groups ($\chi^2 = 1.89$; $df = 2$; $p > 0.3$). On the other hand, when compared to monolingual and simultaneous bilingual peers, sequential bilinguals always show highly significant differences in the informativeness of their referential choices (Mono vs SeqBil: $\chi^2 = 16.06$; $df = 2$; $p < 0.001$; SimBil vs SeqBil: $\chi^2 = 16.65$, $df = 2$; $p < 0.001$).

The Chi-square results confirm the descriptive trends observed in the distribution of REs. Specifically, the lack of significant difference between simultaneous bilinguals and monolinguals suggests a broadly comparable pattern in informativeness levels, with both groups producing a similarly high proportion of optimal REs and a low, but visible, use of under and overinformative forms. In contrast, the sequential bilingual group display a significantly distinct and unique pattern.

Some instances of overinformative referential expressions produced by sequential bilingual students are provided below.

It is worth mentioning that three out of the four occurrences among sequential bilinguals were produced by the same student (MP6), see example (6):

- (6) *Una ragazza mi salutò e si sedette vicino a me, non sapevo chi chi era, **la ragazza** mi diede una bottiglia mi disse che dovevo berla. Mi alzai dicendogli che dovevo tornare a casa, **la ragazza** mi prese il braccio e mi fermò [...].* (MP6)
 ‘A girl greeted me and sat down next to me, I didn’t know who she was. **The girl** gave me a bottle and told me I had to drink it. I stood up and told her I had to go home; **the girl** grabbed my arm and stopped me.’

This repeated mention of the girl through a full NP (*la ragazza*, ‘the girl’), despite her referent being highly accessible, reveals a pattern of referential overspecification in this speaker’s narrative.

It should be noted that this type of overspecification can also be found in texts from the other groups. Among monolinguals the occurrences of overinformative REs come from 4 students, while among simultaneous bilinguals, 10 (out of 13 instances) of

overinformative REs come from a single student (MV18), who consistently shows a pervasive strategy throughout the text: the overt reintroduction of the subject pronoun immediately after direct speech (e.g., “*Tu no!*” *rispose lei*, ‘Not you!’ **she** answered’), also when the speaker is contextually clear and unambiguous (even with 1st person sg.). Notably, this tendency can also be found in the other groups, although less frequently.

5. Restandardisation trends in pronoun choice

As mentioned in section 2.2, the paradigm of 3rd person pronouns in Italian has undergone a shift in formality, especially in the case of subject pronouns and clitics expressing indirect feminine and plural objects. Whereas the forms *lui*, *lei* and *loro* (lit. ‘him’, ‘her’ and ‘them’), devoted to object marking according to standard grammar, have already spread to spoken and written varieties of medium formality, the use of *gli* (‘to him’) for plural and, especially, feminine indirect object are more resistant to change and are still considered very informal (Serianni 1988; see also Zucchini 2025 regarding school grammars).

This section is dedicated to investigating how teenagers behave in this respect when writing in school settings.

As the main aim of this study is analysing reference in written narratives, which required a very fine-grained annotation, the dataset is quite small, smaller than those usually adopted for sociolinguistic inquiries. This is why the analysis presented in this section must be considered very exploratory: no strong conclusion will be drawn from the discussion of the results, which will only open new paths for future studies.

Table 10 shows the occurrences for each pronominal form in the three groups of participants involved in this study (monolingual, simultaneous and sequential bilinguals). It is necessary to specify that the whole sample (48 texts, of which 20 written by monolinguals, 20 by simultaneous bilinguals and 8 by sequential bilinguals) was considered, despite not being balanced in terms of number of texts for each group.

The figures contained in the table refer to the number of occurrences of each form, whereas the percentage within parenthesis indicates the rate of selection of that specific form over the total cases of pronouns performing that function; we did not consider in the calculation other strategies such as zero anaphora or NP.

As far as subject pronouns are concerned, almost no variation emerges, whereas a higher degree of variability can be found in the case of indirect object pronouns.

A comparison between the choice of subject pronouns in the monolingual and in the bilingual group show that their behaviour is the same: the authors of the text always choose the forms *lui*, *lei* and *loro*, and only one case of *esso* (literally ‘it’) can be found, in reference to “my father”, in a text by a monolingual student (ex. 7). Such example, albeit episodic, is interesting because, according to Serianni (1988), the pronoun *esso* should be used only for inanimate referents:

- (7) *A questo punto scendiamo ed **esso** mi chiede [...].* (MCB13)
 ‘At that point we get out and **he** asks me [...]

Function	Form	Monolinguals	Simultaneous bilinguals	Sequential bilinguals	Total
Subject M. SG.	<i>lui</i>	15 (94%)	19 (100%)	4 (100%)	38 (97%)
	<i>esso</i>	1 (4%)	/	/	1 (3%)
	Tot.	16	19	4	39
Subject F. SG.	<i>lei</i>	6 (100%)	24 (100%)	1 (100%)	31
Subject PL.	<i>loro</i>	7 (100%)	5 (100%)	/	12
Indirect Object M. SG.	<i>gli</i>	14 (93%)	19 (100%)	4 (100%)	37 (97%)
	<i>li</i>	1 (7%)	/	/	1 (3%)
	Tot.	15	19	4	38
Indirect Object F. SG.	<i>le</i>	4 (30%)	5 (25%)	/	7 (22%)
	<i>gli</i>	1 (7%)	14 (70%)	2	15 (47%)
	<i>li</i>	9 (64%)	1 (5%)	/	10 (31%)
	Tot.	14	20	2	32
Indirect Object PL.	<i>loro</i>	1 (10%)	1 (25%)	/	2 (14%)
	<i>gli</i>	8 (80%)	3 (75%)	/	11 (79%)
	<i>li</i>	1 (10%)	/	/	1 (7%)
	Tot.	10	4	/	14

Table 10. Pronoun selection in the sample.

The picture is further complicated by the presence of *li* (lit. ‘them’), the clitic pronoun devoted to masculine plural direct object, as in *li ho visti*, ‘I saw them’ or *ho smesso di vederli* ‘I stopped seeing them’. As shown by Table 10, this form is used in the data also to express indirect object, across all genders and numbers, replacing *gli* also where its use is considered non-standard. Ex. (8a) shows *li* used for masculine singular, (8b) for feminine singular and (8c) for plural:

- (8) a. *Il figlio però implorò il padre di prometterli che dopo la spedizione si sarebbero rivisti.* (MDA15)
‘But the son begged the father to promise **him** they would meet again after the expedition’
- b. *Alessio quindi li fa sedere sul tappeto volante e li dice.* (MDA15)
‘So Alessio makes them sit on the flying carpet and tells them’
- c. *ovviamente io decisi di scriverli perche non sarei mai riuscito a parlarli.* (MCD1)
‘Obviously I decided I would write to **her** because I would never be able to speak to **her**’.

This use is not considered neostandard by studies, but rather a feature of *Italiano popolare* (Berruto 1983), the variety spoken by non-educated speakers with an italo-romance dialect as native language.

As the number of occurrences in the data is not very high, the whole corpus (described in section 3.1, see also Zucchini 2025) was searched for similar uses of *li*, to rule out the possibility of it being an idiosyncrasy of a few writers. If we consider the

corpus as a whole, the number of occurrences of *li* used by monolinguals increases, whereas the cases found in bilingual students' texts remain the same.⁶

Table 11 summarizes the occurrences of *li* in the larger corpus, organized by gender and number of the referent and subcorpus. Since the texts written by monolingual and bilingual students are not the same size (monolingual 92.731 tokens; simultaneous bilingual 22.797, sequential bilinguals 6820), the table also includes the frequency of *li* (multiplied by 1 million for better legibility).

	Occurrences			
	Monolinguals	Sim. Bilinguals	Seq. bilinguals	Tot.
M. SG.	6	/	/	6
M. PL.	4	/	/	4
F. SG.	10	1	1	12
Tot.	20	1	1	22
Frequency	215,68	43,87	146,62	174,47

Table 11. Use of *li* as indirect object pronoun (entire corpus).

The data from the larger corpus confirms that monolingual students are mainly responsible for the introduction of *li* in the paradigm of indirect object pronouns, since the frequency is higher than the general rate. The table also shows that *li* is adopted more often for feminine referents (see also Table 10).

As far as individual choices are concerned, the data analysis shows that *li* is used as a complete replacement for *gli* by eight students, who never use the latter in their texts, whereas two students use both; in examples (9a) and (9b), the two pronouns perform the same function (feminine singular):

(9) MCB21

- a. *Si era fatta sera e neanche me ne ero accorto caspita con lei passa infretta il tempo, perciò mi alzai e gli dissi: "io dovrei andare anche perche si è fatto tardi".*
 'It was already night time and I didn't even realise dang time flies with her, so I got up and told her "I should go because it's late"'.
- b. *Dopodiché uscì anche una bellissima ragazza [...] ad un tratto mentre giocavo a pallone i nostri sguardi si incrociarono e mi sorrise, quindi decisi di andarli a parlare.*
 'Then a gorgeous girl got out [...] at some point while I was playing football our eyes met and she smiled, so I decided to go speak to her'.

Tables 12-14 give more details about the context of occurrence, more specifically by which verb the pronoun *li* is controlled (tab. 13) and its argument structure (tab. 15), which may be relevant to understand the phenomenon in question; the verb may be ditransitive and take two objects (one direct and one indirect, ex. 9a), or intransitive and can only take indirect objects (ex. 9b).

⁶ A small number of occurrences (5 in total) of *li* replacing *gli* could be found also in a corpus of essays of high school students (98.771 tokens; see Zucchini 2023).

As can be seen from tables 12-14, the occurrences of *li* are found in equal proportions in pre and postverbal position; moreover, the majority of verbs (13/22) are ditransitive, 3 are intransitive unaccusative and only *parlare* ‘talk’ is intransitive unergative, but is adopted quite frequently in combination with *li* (6 occurrences).

Syntactic position	Occurrences
Postverbal	11
Preverbal	11
Tot.	22

Table 12. Syntactic position of pronoun *li*

Lemma	Occurrences
<i>parlare</i> , ‘talk’	6
<i>dire</i> , ‘say’	4
<i>scrivere</i> , ‘write’	4
<i>chiedere</i> , ‘ask’	2
<i>piacere</i> , ‘like’	2
<i>fare qualcosa a qualcuno</i> ‘do something to someone’	1
<i>mandare</i> ‘send’	1
<i>promettere</i> ‘promise’	1
<i>venire un’idea</i> ‘have an idea’	1
Tot.	22

Table 13. Lemma of controlling verb

Verb argument structure	Occurrences
Ditransitive	13
Intransitive unergative	6
Intransitive unaccusative	3
Tot.	22

Table 14. Argument structure of controlling verb

6. Discussion

This study analyzes narrative texts by Italian monolingual and bilingual adolescents. Our findings show that what may seem like two distinct domains, referential structure (distribution of NPs, pronouns, and null subjects) and sociolinguistic variation (neostandard pronominal forms), are in fact meaningfully interconnected. Clitic pronouns, in particular, emerge as a key site of convergence between referential strategy and sociolinguistic innovation.

As for referential strategies, we conducted two comparisons using two different subsamples, in which the groups were selected based on the age of first exposure to Italian.

The first comparison involved monolinguals and simultaneous bilinguals (i.e., individuals exposed to Italian before the age of 4). We analyzed both the distribution of referential forms used in two discourse functions (introduction and maintenance) and the informativeness level of these expressions (over- or under-informative). No significant differences emerged from this comparison: the only noteworthy observation is a higher overall use of pronominal forms among monolinguals.

The second comparison involved a smaller sample and aimed to examine the distribution of referential forms and their informativeness levels across three groups: monolinguals, simultaneous bilinguals, and sequential bilinguals (i.e. exposed to Italian after the age of 6). While the first two groups (monolinguals and simultaneous bilinguals) continued to show no significant differences, the third group, sequential bilinguals, differed from the other two both in the distribution of referential forms and in the informativeness level.

Sequential bilinguals show a higher use of zero anaphora, which nevertheless does not result in underinformativeness. This pattern may reflect a different organization of discourse in this group. Notably, they produce no underinformative expressions, which can be interpreted as the outcome of a compensatory strategy whereby more explicit referential forms are employed to secure referential clarity (with only a negligible number of overinformative choices). Interestingly, while all groups performed well in terms of producing optimal referential expressions (monolinguals: 95%; simultaneous bilinguals: 94%; sequential bilinguals: 99%), only sequential bilinguals show a markedly reduced reliance on pronominal forms. This suggests that their preference for more explicit referential forms may not be driven by increased referential precision, but rather by a reduced confidence in managing the full range of referential options available within the language, especially in the case of pronouns, which are cognitively demanding and require fine-grained discourse tracking (Arnold & Zerkle 2019).

This distribution suggests a strategy oriented toward maximizing referential clarity while minimizing cognitive load in pronoun production, since it optimizes the underlying processing mechanisms of morphosyntactic form selection. Interestingly, even simultaneous bilinguals use pronominal forms less frequently than monolinguals. This subtle difference may point to the fact that the Italian pronoun system takes longer to be fully mastered and is cognitively more demanding to process (Arnold & Zerkle 2019), especially in written tasks (Bel & Albert 2016).

Such findings are consistent with prior research indicating that pronoun acquisition and mastery can be protracted both in bilingual and monolingual speakers. Sorace et al. (2009), for instance, notes that pronouns pose persistent challenges in bilingual development. Similarly, studies in other null-subject languages such as Spanish (Bel & Albert 2016) show that the acquisition of adult-like pronominal usage tends to occur later than the mastery of null subject forms, often stabilizing around early adolescence, coinciding with the age of participants in our sample.

Overall, all participant groups produced cohesive and structurally coherent texts. The differences observed between groups appear to be primarily quantitative, especially concerning the frequency of pronoun use, rather than qualitative: no group exhibited fundamentally different referential patterns; rather, the variation lies in the extent to which particular forms are selected or avoided. Notably, only the use of pronouns

emerges as the key variable differentiating the groups, with monolinguals using them most frequently, followed by simultaneous bilinguals, and then by sequential bilinguals. It is also worth noting that instances of inappropriate use tend to be attributable to individual participants rather than representative of group-wide tendencies. However, given the relatively small sample size, these group-level differences should be interpreted with caution.

The management of the Italian pronominal system is further complicated by sociolinguistic factors, which must be taken into account when selecting referential forms in order to avoid ambiguity (recall example 5), which is why an analysis of pronoun selection was included in our investigation.

As far as subject pronouns are concerned, very little variation emerged and students always adopt neostandard solutions; the overall results are not surprising, since a previous study on the corpus the sample was extracted from has found very few occurrences of old-standard pronouns (22 cases in 345 texts; Zucchini 2025), whereas older students have been shown to use them much more frequently (see Zucchini 2023, on high school students, and Ballarè in press, on university students).

As for dative clitics, the degree of variation is higher both for feminine and plural referents and the neostandard pronoun choice is generally predominant, with the standard option being chosen only in 22% and 14% of times. However, the use of non-standard forms (among which we include *li*) is more frequent in the case of feminine than in the case of plural referents, despite the latter being more accepted in formal speech and written texts according to the literature.

The comparison between the three groups shows that all students strongly prefer *gli* for the expression of plural, with a small difference (5%). The picture regarding feminine referents is complicated by the presence of 9 occurrences of *li* in texts by monolingual students, which, therefore, is their preferred choice (64%), followed by *le*, the standard Italian option (30%); simultaneous bilinguals, on the contrary, strongly prefer *gli* also for feminine referents (around 70%) and only choose *li* once (5%).

The most interesting element of the whole picture is the presence of *li*, which can be considered the most innovative choice. Thanks to a comparison with the whole corpus, it was possible to increase the number of contexts where *li* is used in an innovative way, which has enabled to further confirm our findings and observe the distribution of the phenomenon on a larger dataset.

Knowing that we cannot entirely rule out factors related to performance errors, given the similarity between the two forms, we will attempt an interpretation of the phenomenon.

We believe the emergence of *li* to express indirect object may be the outcome of a two-phase change, especially because it is adopted across genders and numbers:

- Phase 1: *gli*, *le*, *loro* (indirect-object pronouns differentiated for gender and number) > *gli* (syncretic form);
- Phase 2: *gli* > *li*.

In phase 1, already described by studies on restandardisation, the system of forms used for indirect object in Italian simplifies into one form *gli* which can be used for masculine and feminine referent, singular and plural. In phase 2, still very incipient, the form *gli* alternates with *li*, clitic pronoun devoted to masculine singular object marking.

The second phase may be favoured, in our opinion, by two factors: the phonetic similarity between the two forms (*gli* and *li*) or the argument structure of the controlling verb, since ditransitive verbs allow the construction with both pronouns.

We cannot confirm whether the phenomenon is caused by the first, the second factor or a combination of the two, due to the low number of total occurrences. Moreover, since the occurrences of *li* found in the corpus are controlled by ditransitive verbs only in a slight majority of cases (13/22), the role of this construction can only be confirmed by investigating larger corpora.

If we consider individual texts (and individual students), we can see that the majority settles on one form, either *gli* or *li*, and uses it categorically; only two students use both forms in the same way, therefore we cannot say that any of the two pronouns is specializing for a specific function. Moreover, it is important to bear in mind that the language sample for each student is very limited (around 400 tokens) and does not allow to draw solid conclusions regarding individual patterns of linguistic behavior.

As far as a comparison between the two groups is concerned, we see that it is monolingual students that use *li* the most; bilingual students show, in this respect, more conservative behavior than monolinguals, since it is the latter that adopt, while writing, a form which is considered typical of a low variety of Italian, the *Italiano popolare* (Berruto 1983). However, due to the scarcity of metadata, we are not able to understand in details which is the real factor causing this difference, which may be due to motivations going beyond bilingualism, such as domestic and individual language practices, home literacy and socio-economic status, among many.

7. Concluding remarks

The study aims to investigate REs in narrative texts written by monolingual and bilingual Italian adolescents in school settings, both regarding the effective management of reference tracing and pronoun choice with respect to competing standards in contemporary Italian.

The first part is dedicated to a review of the literature about the use of REs in narrations and their relation with referent accessibility, the differences between monolingual and bilingual speakers in the acquisition of reference, the restandardisation process that involves the paradigm of personal pronouns in Italian and the role of children and adolescents in language change.

Studies reveal that the use of REs in narrations is affected by bilingualism, especially due to the effect of a greater cognitive load and to the proficiency and exposure to the target language.

REs in Italian, in particular the paradigm of third personal pronouns, is undergoing change, but the behaviour of teenagers in this respect has not been investigated yet; however, according to sociolinguistic studies, the younger generations play a significant role in propagating language change and, therefore, we believe they should be included in the picture, all the more considering the high percentage of bilingual students in Italian schools.

The investigation described in this paper is carried out on a sample of narrative texts, extracted from a larger corpus that include texts written by monolinguals, simultaneous and sequential bilinguals.

The comparison between monolinguals and simultaneous bilinguals reveals similar distributional patterns of REs in both groups, who are also equally competent in reference tracing; the only differences that emerged regards the use of pronouns in sequential bilinguals, which are used less frequently in comparison to both simultaneous bilinguals and monolinguals. Once again, the majority of REs are used appropriately across groups.

As far as pronoun choice is concerned, unsurprisingly all students prefer neostandard solutions both regarding subject pronouns and indirect-object pronouns, with old standard pronouns that have basically disappeared. More worth of interest is the use of *li* as indirect object pronoun across genders and numbers, phenomenon observed only by Berruto (1983) in the language of dialect speakers with low levels of education.

Despite the dataset size, too small to draw any robust conclusion, we believe it is a phenomenon worth monitoring, especially if we consider the findings of the literature on the role of teenagers in leading linguistic change; the presence of this phenomenon also calls for more research aimed to compare learners' and low varieties. Finally, we believe that adopting a sociolinguistic perspective on the study of acquisition would add interesting insights on the topic, especially if it regards linguistics phenomena undergoing a shift.

Abbreviations

F = feminine
M = masculine
HUM = human
PL = plural
RE = referential expression
SG = singular

References

- Andreou, Maria & Knopp, Eva & Bongartz, Christiane Maria & Tsimpli, Ianthi M. 2015. Character Reference in Greek-German Bilingual Children's Narratives. In L. Roberts, K. McManus, N. Vanek, D. Trenkic (Eds), *EUROSLA Yearbook 15*, 1-40. Amsterdam, John Benjamins.
- Andreou, Maria & Torregrossa, Jacopo & Bongartz, Christiane Maria. 2020. The Sharing of Reference Strategies across Two Languages: The Production and Comprehension of Referring Expressions by Greek-Italian Bilingual Children. *Discours* [online], 26 | 2020, published online November 2, 2020.
- Ariel, Mira. 1991. The function of accessibility in a theory of grammar, *Journal of Pragmatics*, 16(5), 443–464.
- Ariel, Mira. 2001. Accessibility theory: An overview. In Sanders, Ted & Schliperoord, Joost & Spooren, Wilbert (eds), *Text Representation*, 29–87, Amsterdam: John Benjamins.
- Arnold, Jennifer E. 2010. How speakers refer: The role of accessibility, *Language and Linguistics Compass* 4(4), 187–203.

- Arnold, Jennifer E. & Zerkle Sandra A. 2019. Why do people produce pronouns? Pragmatic selection vs. rational models. *Language, Cognition and Neuroscience* 34.9, 1152-1175.
- Ballarè, Silvia (in press). Competing standards in university students' formal production: a corpus-based study. In Elspaß, Stephan & Mendoza, Imke & Pöll, Bernhard & Schleef, Erik (eds.), (1600-2020) – Historical Sociolinguistics series. Frankfurt am Main, Peter Lang.
- Barbu, Stéphanie & Nardy, Aurélie & Chevrot, Jean-Pierre & Juhel, Jacques. 2013. Language evaluation and use during early childhood: Adhesion to social norms or integration of environmental regularities? *Linguistics* 51(2). 381–411.
- Bel, Aurora & Albert, Milagros. 2016. The development of referential choice in Spanish narratives among school-age children and adolescents. In Perera, Joan & Aparici, Melina, & Rosado, Elisa & Salas, Naymé (ed), *Written and Spoken Language Development across the Lifespan*, 251-269. Berlin: Springer.
- Benincà, Paola & Penello, Nicoletta 2009. L'uso di *le* al di là dei suoi confini. In Cardinaletti, Anna & Munaro, Nicola (a cura di), *Italiano, italiani regionali e dialetti*. Milano, Franco Angeli, pp. 13-27.
- Berruto, Gaetano (2012[1987]), *Sociolinguistica dell'italiano contemporaneo*. Roma: Carocci.
- Cardinaletti, Anna. 2021. I pronomi deboli nel sistema pronominale dell'italiano. *Storie e linguaggi* (7)2.
- Carminati, Maria Nella. 2002. *The Processing of Italian Subject Pronouns*. PhD thesis. University of Massachusetts at Amherst.
- Chafe, Wallace L. 1976. *Givenness, contrastiveness, definiteness, subjects, topics and point of view*. In Charles N. Li (ed.): *Subject and topic*, 27–55. Academic Press, New York.
- Chini, Marina. 2005. *Reference to person in learner discourse*. In Hendriks, Henriëtte, *The structure of learner varieties*. Berlin/Boston: De Gruyter, 65–110.
- Colozzo, Paola & Whitely, Cristy. 2014. Keeping track of characters: Factors affecting referential adequacy in children's narratives. *First Language* 34(2), 155–177.
- De Lange, Joke & Vasic Nada & Avrutin, Sergey. 2009. Reading between the (head) lines: A processing account of article omissions in newspaper headlines and child speech. *Lingua* 119(10), 1523–1540.
- De Vogelaer, Gunther & Chevrot, Jean-Pierre & Katerbow, Matthias & Nardy, Aurélie. 2017. *Bridging the gap between language acquisition and sociolinguistics: Introduction to an interdisciplinary topic*. In De Vogelaer, Gunther & Katerbow, Matthias (eds.), *Studies in Language Variation*, 1–41. Amsterdam: John Benjamins.
- Eckert, Penelope. 1997. Age as Sociolinguistic Variable. In Coulmas, Florian (ed.), *The handbook of sociolinguistics*, 151–167. Oxford: Blackwell.
- Eckert, Penelope. 2004. Adolescent language. In Finegan, Edward & Rickford, John R. (eds.), *Language in the USA*, 361–374. Cambridge: Cambridge University Press.
- Foulkes, Paul & Docherty, Gerard & Watt, Dominic. 2005. Phonological Variation in Child-Directed Speech. *Language*, 81(1), 177–206.
- Fiorentini, Ilaria & Gianollo, Chiara. 2021. L'alfabetizzazione nella classe plurilingue. Un'indagine a Bologna. *Lingue e linguaggi* 41, 215-232.
- Georgopoulos, Athanasios. 2017. Anaphora resolution in the interlanguage of Greek and English learners of Spanish: A corpus study. *Studies in Greek Linguistics* 37, 239-252.

- Gundel, Jeanette K. & Hedberg, Nancy & Zacharaski, Ron. 1993. Cognitive status and the form of referring expressions. *Language* 69, 274–307.
- Hendriks, Petra & Charlotte Koster & John C.J. Hoeks. 2013. Referential Choice across the Lifespan: Why Children and Elderly Adults Produce Ambiguous Pronouns. *Language, Cognition and Neuroscience* 29 (4), 391–407.
- Johnson, Elizabeth K. 2016. Constructing a Proto-Lexicon: An Integrative View of Infant Language Development. *Annual Review of Linguistics*, 2(1), 391–412.
- Johnson, Elizabeth K. & White, Katherine S. 2019. Developmental sociolinguistics: Children’s acquisition of language variation. *WIREs Cognitive Science* 11(1), e1515.
- Kohn, Mary Elizabeth. 2014. *The way I communicate changes but how I speak don’t: A longitudinal perspective on adolescent language variation and change*. Duke: Duke University Press.
- Labov, William. 1964. Stages in the acquisition of Standard English. In Shuy Roger W. (ed.), *Social dialects and language learning*, 77–104. National Council of Teachers of English.
- Labov, William. 1973. *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press.
- Labov, William. 2001. *Principles of linguistic change. Vol. 2: Social factors*. Oxford: Blackwell.
- Leonetti Escandell, Vicky. 2024. Reference management in written narrative production by Spanish-Italian bilingual children. *Frontiers in Communication* 9, 1400984.
- Lindgren, Josefin, & Bohnacker, Ute. 2024. (In) definites, pronouns and bare nouns: How Turkish/Swedish bilingual 4-to-7-year-olds introduce characters in narratives in Swedish. *Nordic Journal of Linguistics* 2024, 1–30.
- Nardy, Aurélie & Chevrot, Jean-Pierre & Barbu, Stéphanie. 2013. The acquisition of sociolinguistic variation: Looking back and thinking ahead. *Linguistics* 51(2), 255–284.
- Pallotti, Gabriele & Borghetti, Claudia & Rosi, Fabiana. 2020. *Una didattica della scrittura efficace e inclusiva. Il progetto Osservare l’interlingua*, Bologna/Cesena: Caissa.
- Perugini, Nicola. 2025. Specific patterns of referential use in Mandarin-Italian bilingual children. *ZAS Papers in Linguistics* 66 (2025). 141–161
- Prince, Ellen. 1981. Toward a Taxonomy of Given-New Information. In Cole, Peter (ed.), *Radical pragmatics*, 223–255. New York: Academic Press.
- Roberts, John. 1997. Acquisition of variable rules: A study of (-t, d) deletion in preschool children. *Journal of Child Language* 24(2), 351–372.
- Sabatini, Francesco. 1985. L’“italiano dell’uso medio”: una realtà tra le varietà linguistiche italiane. In Holtus, Günter & Radtke, Edgar (eds.), *Gesprochenes Italienisch in Geschichte und Gegenwart*. Tübingen: Narr, 154–184.
- Sabatini, Francesco. 1990. Una lingua ritrovata: l’italiano parlato. *Studi latini e italiani* 4, 215–234.
- Schimke, Sarah & Colonna, Saveria & Hickmann, Maya. 2020. Reference in French and German, in Gagarina, Natalia & Musan, Renate (eds.), *Referential and Relational Discourse Coherence in Adults and Children*, Berlin/Boston, De Gruyter, 139–170.
- Schulz, Petra & Grimm, Angela. 2019. The Age Factor Revisited: Timing in Acquisition Interacts With Age of Onset in Bilingual Acquisition. *Frontiers in Psychology*. 9:2732

- Serianni, Luca. 1988. *Grammatica italiana: italiano comune e lingua letteraria. Suoni, forme, costrutti*. Torino: UTET.
- Smith, Jennifer & Durham, Mercedes & Fortune, Liane. 2007. “Mam, my trousers is fa’in doon!”: Community, caregiver, and child in the acquisition of variation in a Scottish dialect. *Language Variation and Change* 19(01), 63–99.
- Sorace, Antonella. 2011. Pinning down the concept of “interface” in bilingualism. *Linguistic approaches to bilingualism* 1(1), 1–33.
- Sorace, Antonella. 2016. Referring expressions and executive functions in bilingualism. *Linguistic Approaches to Bilingualism* 6(5), 669–684.
- Sorace, Antonella & Serratrice, Ludovica & Filiaci, Francesca & Baldo, Michela. 2009. Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua* 119(3), 460–477.
- Tagliamonte, Sali A. & D’Arcy, Alexandra. 2009. Peaks Beyond Phonology: Adolescence, Incrementation, and Language Change. *Language* 85(1), 58–108.
- Tagliamonte, Sali A. 2006. *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.
- Tagliamonte, Sali A. 2016. *Teen talk: The language of adolescents*. Cambridge: Cambridge University Press.
- Topaj Nathalie. 2010. Topical referential expressions in narratives of Russian-German bilingual children. In Chini, Marina (ed). *Topic, information structure and language acquisition*, Milano: Franco Angeli, 59–72.
- Torregrossa, Jacopo & Andreou, Maria & Bongartz, Christiane & Tsimpli Ianthi M. 2021. Bilingual acquisition of reference: The role of language experience, executive functions and crosslinguistic effects. *Bilingualism: Language and Cognition* 24(4), 694–706.
- Torregrossa Jacopo & Bongartz, Christiane. 2018. Teasing Apart the Effects of Dominance, Transfer, and Processing. Reference Production by German–Italian Bilingual Adolescents. *Languages* 3(36).
- Torregrossa, Jacopo & Bongartz, Christiane, & Tsimpli, Ianthi M. 2019. Bilingual reference production: A cognitive-computational account. *Linguistic Approaches to Bilingualism* 9(4-5), 569-599.
- Tsimpli, Ianthi M. & Sorace, Antonella & Heycock, Caroline & Filiaci, Francesca. 2004. First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of bilingualism* 8(3), 257–277.
- Tsimpli, Ianthi & Sorace, Antonella. 2006. Differentiating interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena. In Bamman, David & Magnitskaia, Tatiana & Zaller, Colleen, *BUCLD 30: 30th Annual Boston-University Conference on Language Development*, Somerville MA: Cascadilla Press, 653–664.
- Zucchini, Eleonora. 2023. *L’italiano neostandard nella lingua a scuola: il caso dell’alternanza fra indicativo e congiuntivo*. Doctoral dissertation, University of Bologna. DOI: 10.48676/unibo/amsdottorato/10505.
- Zucchini, Eleonora. 2025. *L’italiano neostandard a scuola. Con un caso di studio sull’alternanza modale*. Bologna, Pátron.

CLUB Working Papers in Linguistics

SEZIONE CLUB PROGETTI

Per un archivio dell'italiano regionale e dei dialetti emilano-romagnoli

Silvia Ballarè

Università di Bologna

silvia.ballare@unibo.it

Isabella Colliva

Università di Bologna

isabella.colliva@unibo.it

Emanuele Miola

Università di Bologna

emanuele.miola@unibo.it

1. Introduzione

L'unità bolognese del progetto PNRR PE5 CHANGES (*Cultural Heritage Active Innovation for Next-Gen Sustainable Society*) – Spoke 3, WP 5¹ si è occupata principalmente di acquisire materiali linguistici prodotti in Emilia-Romagna. Più specificamente, per dare conto del repertorio della regione, si è cercato di considerare dati dei dialetti parlati sul territorio e dati di italiano, con particolare riferimento all'area metropolitana di Bologna. I dati dialettali, strettamente monolingui, hanno natura più eterogenea e comprendono primariamente testi scritti (§ 2). I dati di italiano sono costituiti da registrazioni (e trascrizioni) di parlato e comprendono, per quanto sporadicamente, episodi di commutazione di codice con dialetti italo-romanzi.

2. Dati dialettali

Al fine di creare un archivio dei dialetti dell'Emilia-Romagna, si è proceduto battendo due strade. La prima ha portato alla digitalizzazione, l'archiviazione e la trascrizione di schede lessicali redatte da Fabio Foresti (v. § 2.1); la seconda, invece, aveva come obiettivo la digitalizzazione di materiali parlati audio e video ma, a causa della natura del dato, si è rivelata più impervia (v. § 2.2).

¹ PNRR PE5: CHANGES – Cultural Heritage Active Innovation for Next-Gen Sustainable Society PE 0000020 CHANGES – CUP F53C22000700006, D53C22002530006

2.1 Il fondo Foresti

All'avvio del progetto si è ricevuta, da parte di Fabio Foresti, la disponibilità a fornire al Dipartimento FICLIT del nostro Ateneo un grande numero di schede preparate in passato per la costituzione di un dizionario storico del dialetto bolognese, che, però, non riuscì mai a vedere la luce.

Il Fondo Foresti è costituito da circa 15.000 schede lessicali compilate a supporto del progetto di dizionario dialettale appena ricordato. Ogni scheda (un esempio è visibile nella fig. 1) contiene una parola dialettale e vari elementi che indicano l'opera letteraria da cui la parola è tratta: troviamo in alto a destra della scheda una sigla che designa il testo di riferimento, accanto al lemma un numero che indica la pagina e in basso la citazione del contesto in cui compare il termine. In alcune schede è presente un commento del redattore, che fornisce indicazioni sulla traduzione della parola o precisazioni riguardo al contesto, spesso utili a comprenderne meglio il significato.

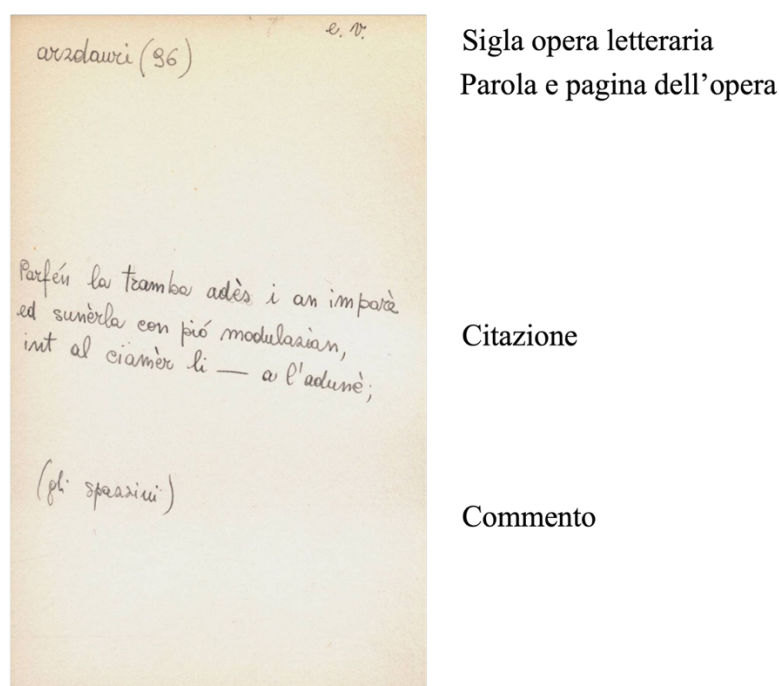


Figura 1. Scheda del Fondo Foresti (voce del lemma *arzdaur*).

Le opere letterarie dialettali utilizzate per la compilazione sono più di 60 e vanno dalla fine del 1500 (è il caso, ad esempio, delle opere minori di Giulio Cesare Croce) fino alla metà del 1900. I testi considerati nella redazione delle schede sono sia in prosa che in poesia e hanno argomenti molto vari: sono presenti, ad esempio, raccolte favolistiche, traduzioni di opere letterarie, storie di vita quotidiana e copioni di commedie teatrali. Tale differenza di generi fa emergere alcune peculiarità, che rendono il materiale interessante in prospettiva storica: una su tutte, le diverse grafie che nei secoli il dialetto bolognese ha assunto e il loro cambiamento, talora dovuto anche ai mutamenti di pronuncia, visibile in molti casi nelle varie attestazioni di una stessa parola. Interessante è anche la presenza di termini desueti (e dunque difficili da tradurre) che dimostrano le evoluzioni lessicali avvenute nel tempo.

Nell’ambito del progetto, i materiali sono stati in primo luogo organizzati e digitalizzati grazie all’aiuto dei tirocinanti dell’ADLab (Laboratorio Analogico Digitale del FICLIT). La digitalizzazione consente di evitare l’usura delle schede fisiche, che, essendo scritte a mano, sono molto delicate; inoltre, siccome sono state annotate a matita, alle volte risultano difficili da leggere e in questi casi l’immagine digitale è d’aiuto. Contestualmente alla digitalizzazione, si sta procedendo alla trascrizione dei dati presenti sulle schede. In questa fase del lavoro le schede sono state raggruppate per lemmi, affiancando alla voce della scheda la forma di riferimento (si è scelto, a titolo esemplificativo, l’infinito per i verbi coniugati). L’obiettivo della trascrizione è immettere i dati sulla piattaforma CHANGES, dove sono consultabili per lemma. Sulla piattaforma il lemma è indicato con “documento” ed è presente una schermata per la sua descrizione denominata “dati del documento” (come riportato nella fig. 2).

Le attestazioni della parola invece sono raggruppate sotto la voce “Dati dei Lemmi” e appaiono come mostrato nella fig. 3.

Come possiamo vedere nella figura 3, in ognuna delle voci riprodotta nei Dati dei Lemmi troviamo il “lemma” che corrisponde alla parola che si trova sulla scheda fisica e la “frase contesto” che riproduce la citazione; nel “nome Informatore” è riportata l’opera letteraria corrispondente alla sigla presente sulle schede e nella “descrizione” troviamo riportato l’eventuale commento del redattore.

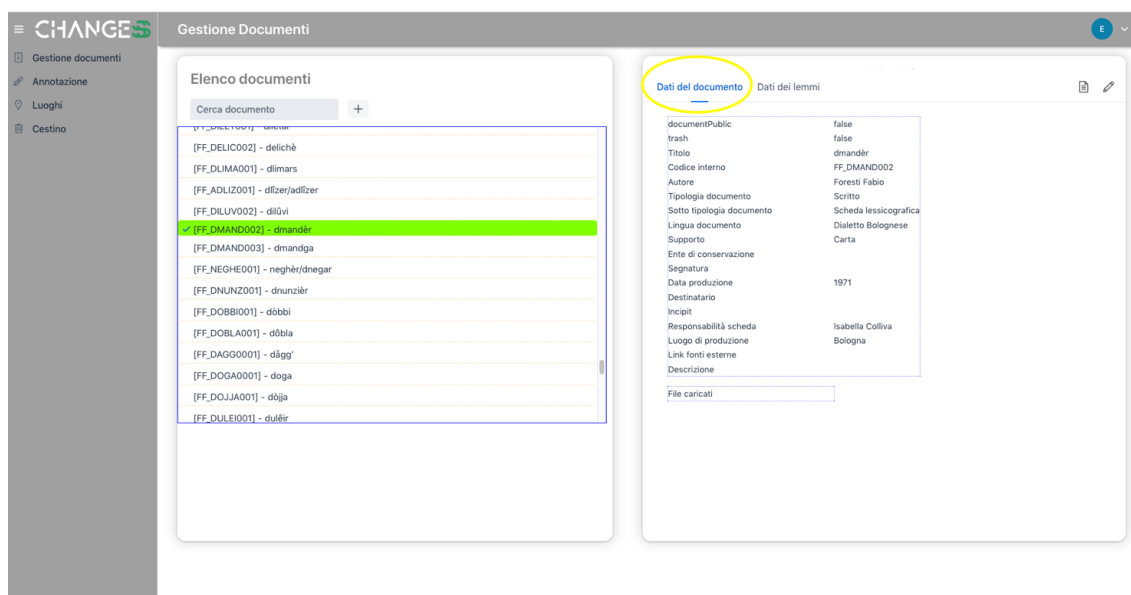


Figura 2. Schermata della sezione Gestione Documenti di CHANGES (lemma *dmandèr*).

X

Dati dei lemmi	
Lemma ->	dmandà
Lingua del lemma originaria	Dialetto Bolognese
Frase Contesto	es m'è sta __ da una mezza duzina d' spurach
Lemma trascritto o tradotto in italiano corrente	domandato, chiesto
Parti del discorso	v
Descrizione	[la figlia è stata chiesta in moglie da...]
nomeInformatore	B.P.III: Fabrizio NANNI-La finta verità del medico per amore. Comedia di Fabritio Nanni-Bologna: Stamperia Longhi-1703
Non compilati:	Professione informatore; Genere infomatore; Grado cultura informatore; Anno nascita informatore; Luogo citato; Area Lessicale; Connotazione; Trascrizione Ipa; Nota Etimologica; Link a fonti esterne; Nomenclatura;
Lemma ->	dmandari
visibility	false
trashL	false
Lingua del lemma originaria	Dialetto Bolognese
Frase Contesto	...s'i accusò un puvrett a __ la limosna,....
Lemma trascritto o tradotto in italiano corrente	domandare, chiedere
Parti del discorso	v
nomeInformatore	ME.SS: Eraclito MANFREDI-Sett sturielli in dialett bulgnèis, fattura d'Eraclit Manfred, mai vgnò a la lus-Bologna: Compositori-1884
Non compilati:	Professione informatore; Genere infomatore; Grado cultura informatore; Anno nascita informatore; Luogo citato; Area Lessicale; Connotazione; Trascrizione Ipa; Nota Etimologica; Link a fonti esterne; Nomenclatura; Descrizione;
Lemma ->	dmandà
visibility	false
trashL	false
Lingua del lemma originaria	Dialetto Bolognese
Frase Contesto	à vuoi dir cuell à Minghina mia fiola, azzò sa fuss __ ch'la saper dund à son.
Lemma trascritto o tradotto in italiano corrente	chiesto, domandato
Parti del discorso	v
Descrizione	azzo sa fuss dmandà-->[se si chiedesse di me]
nomeInformatore	Banch.M: Adriano BANCHIERI-La Minghina da Barbiàn-Venezia-1620
Non compilati:	Professione informatore; Genere infomatore; Grado cultura informatore; Anno nascita informatore; Luogo citato; Area Lessicale; Connotazione; Trascrizione Ipa; Nota Etimologica; Link a fonti esterne; Nomenclatura;

Figura 3. Visualizzazione dati dei lemmi su CHANGES (voci del lemma *dmandèr*).

2.2 I materiali audio e video

Come già accennato, l'individuazione e l'acquisizione di materiali dialettali audio e video, per diverse ragioni, si è rivelata piuttosto difficoltosa. Innanzitutto, com'è noto, la voce di chi parla è un dato sensibile protetto dalle leggi sulla privacy e non può essere acquisita se non dietro la firma di una liberatoria da parte del diretto interessato o di chi ne fa le veci, per esempio in caso di decesso. Quest'ultima eventualità non è infrequente nel caso di registrazioni non approntate per il progetto che stiamo discutendo, ma già effettuate, nel passato recente o meno recente, da esperti o da altri raccoglitori: in genere, infatti, i testimoni interpellati sono parlanti anziani, bilingui simultanei in italiano ed emiliano-romagnolo o anche L1 emiliano-romagnoli. Non vanno nemmeno trascurati i casi in cui chi ha raccolto i materiali non ricorda (o non conoscesse fin dall'inizio) le generalità di coloro che si sono prestati all'inchiesta: occorre, a norma di legge, procedere a divulgare l'intenzione dell'Ateneo di usare gli audio e i video registrati tramite pubblici proclami su tutta l'area in cui si sono svolte le indagini dialettali.

In secondo luogo, non è facile definire, per le registrazioni come quelle in questione, i contorni del diritto d'autore. Se, infatti, possono sicuramente accampare diritti sulla registrazione il ricercatore/raccoglitore e il parlante, è meno semplice definire il ruolo di

un eventuale compilatore dei questionari o del canovaccio di inchiesta sottoposta agli informanti.

Al momento della stesura tutte le liberatorie e i documenti relativi all'acquisizione dei materiali sono stati predisposti, dopo un lungo scambio con gli Uffici preposti. Si è iniziata a sondare l'esistenza di associazioni dialettali e privati che abbiano materiali di nostro interesse, e la loro disponibilità a donarli all'Ateneo.

La possibilità di fare registrazioni *ad hoc* per il progetto CHANGES di materiale dialettale è stata presa in considerazione ma, almeno al momento della stesura di queste righe, è stata accantonata. La principale ragione sta nella difficoltà di reperire persone, anche – a causa della ben nota contrazione della dialettologia in tutte le aree gallo-italiche – tra i laureati magistrali in materie letterarie, con le giuste competenze sociolinguistiche e linguistico-teoriche e, nel contempo, capaci di condurre un'intervista semi-direttiva praticando il dialetto locale abbastanza bene da seguire il discorso dell'informante interlocutore e da sostenere la conversazione senza dover cambiare il codice verso l'italiano, inducendo così l'intervistato a smettere di parlare dialetto.

3. L'italiano regionale

Nell'ambito del progetto qui presentato, ci si è anche occupati della raccolta e della pubblicazione di dati di parlato (primariamente in italiano) raccolti nella città di Bologna.

ParlaBO è il quarto modulo del corpus KIParla (www.kiparla.it) ed è stato pubblicato nel 2024 grazie ai finanziamenti dei progetti PRIN PNRR *DiverSIta* (per cui v. Mauri et al. in questo volume) e PNRR PE5 CHANGES. Il modulo raccoglie oltre 65 ore di registrazioni di interviste semi-strutturate raccolte nella città metropolitana di Bologna. I parlanti coinvolti risiedono in varie zone del territorio e sono diversificati in base alla loro caratterizzazione sociale (quindi, ad esempio, per titolo di studio, tipo di attività lavorativa, etc.). Il modulo è bilanciato in relazione alla fascia d'età degli intervistati, come illustrato nella tabella (1).

Fascia d'età	Ore di registrazione
Giovani (dai 16 ai 29 anni)	18:46:50
Adulti (dai 30 ai 59 anni)	24:21:43
Anziani (oltre i 60 anni)	22:43:25
Totale:	65:43:25

Tabella 1. Bilanciamento ParlaBO

Le interviste sono state raccolte dagli studenti, debitamente formati e affiancati, che hanno preso parte al tirocinio KIParla tra il 2021 e il 2024. Gli argomenti affrontati durante le conversazioni, pur mantenendo un margine di variabilità, hanno riguardato principalmente il rapporto degli intervistati con il territorio e con le tradizioni locali. Di seguito, con finalità meramente esemplificativa, si riportano alcune delle domande contenute nel canovaccio dell'intervista elaborato per la raccolta dati.

- a) Ti piace la tua zona?
 - i. Perché?
 - ii. Ci sono dei problemi? Come miglioreresti la tua zona?

- iii. Trovi tutto nel tuo quartiere o devi spostarti spesso?
- iv. Come ti muovi in città?
- b) Qual è il cibo bolognese che più ti piace? Perché?
 - i. Come si fa il ragù a casa tua?
 - ii. Qual è un piatto tradizionale che ti ricorda casa?
 - iii. Come si prepara?

Numerosi, poi, sono i racconti relativi al passato e legati alla città di Bologna. Nelle righe successive, in (1), si riporta un estratto di un'intervista a una donna di età compresa tra i 76 e gli 80 anni, in cui è narrato un episodio riconducibile alla Seconda Guerra Mondiale.

1) ParlaBO, PBC005

- BOI014 [>no mi è mi è] n~ mi è venu~ mi viene in mente subito perchè mia madre<
ha sempre raccontato che
- BOI014 e:h (.) ogni volta che sentivano
- BOI014 pippo
- BOI014 che era
- BOI014 eh (.) <l'aereo>
- BOI014 che (.) che:
- BOI014 dava l'allarme no?
- BOR002 m:hmh
- BOI014 E
- BOI014 e:h non so se era pippo che dava l'allarme >comunque< c'era sto pippo (.)
non lo so [xx]
- BOR002 [ma] era una persona o era [un nome per qualcosa?]
- BOI014 [no no un aereo un ae]re[o]
- BOR002 [ah] era un [ae]reo
- BOI014 [ve]ro credo che fosse l'aereo (.) comunque
- BOI014 e:h
- BOI014 quando suonava l'allarme tutti andavano (.) a nei rifugi
- BOI014 uno è in via: dei bregoli
- BOR002 a:h
- BOI014 non l'hai mai visto? quando hai [fatto la strada (.) ecco c'è adesso] l'hanno
chiuso però
- BOR002 [<non c'ho mai fatto caso>]
- BOR002 [a:h]
- BOI014 [proprio vic]ino alla chiesa
- BOI014 su poco dopo c'era
- BOI014 <sotto nel mont[e avevano] (.) fatto i rifugi no?>
- BOR002 [okay]
- BOI014 [i rifu]gi (.) oppure si andava in cantina insomma SI SCAPPAVA

Il modulo ParlaBO, dunque, costituisce di fatto un serbatoio di racconti riguardanti la città di Bologna, la sua storia e le sue tradizioni. Esso, inoltre, risulta essere di grande rilievo per interrogativi di ricerca di natura (socio)linguistica. La metodologia impiegata per la raccolta dati, infatti, è stata selezionata in modo tale da elicitare produzioni linguistiche poco controllate. Laddove possibile, sono stati coinvolti contemporaneamente più parlanti che avessero un alto grado di conoscenza (ad esempio, amici, coppie, membri della stessa famiglia, etc.), nel tentativo di favorire l'interazione tra loro e ridurre il peso specifico dell'intervistatore all'interno dello scambio comunicativo. Inoltre, si è scelto di svolgere le interviste in luoghi familiari all'intervistato (come, ad esempio, abitazioni private ma anche locali pubblici e luoghi di lavoro), anche a scapito della qualità audio della registrazione. Infine, si è preferito impiegare per la registrazione semplici smartphone (e non registratori professionali), in modo tale che non costituissero una distrazione rispetto al contesto abituale. Queste scelte hanno fatto sì che nei dati comparissero tratti linguistici propri dell'italiano colloquiale e regionale, oltre che non sporadici episodi di commutazione di codice con il dialetto locale. In (2) si riporta un esempio di impiego del dialetto locale all'interno di un racconto portato avanti da due informanti, ovvero BOI033 (donna, 61-65 anni, licenza media) e BOI034 (uomo, over85, licenza elementare).

2) ParlaBO, PBC012

BOI034 dove dormivamo noi, han dormit[o e: le truppe]
BOI033 [i tedeschi]
BOI034 dove dormi~ ((ride))
BOI033 c'erano i par~ c'erano i partigiani ((ride))
BOI034 allora un giorno
BOI033 erano [tutti là]
BOI034 [un gio~]
BOI034 un giorno dice:: i: il co~ l'ufficiale delle cose le quell~
BOI034 il responsabile diciamo del carrarmato li
BOI034 #al dis (a)scolta,
BOI034 #al di~ perché non si poteva tener fer~ al chiuso tutti [a:]: un po' de eh,
BOI033 [eh no]
BOI034 #al dis (a)scolta vieni qua:
BOI034 #al dis chi eser quello là
BOI033 chi sono ((ride))
BOI034 e di~ e dico mio cugino
BOI034 #al fe
BOI034 apri l'occhio
BOI034 eh aveva capito tutto (.) tutto
BOI034 per eh però
BOI033 però son stati [zitti]
BOI034 [avevo] tutto [(l'interesse) di]: da
BOI033 [tanto ormai]

BOI034 sì eh [tant'è vero]

BOI033 [era finita la guerra] ((ride))

Le fasi successive del progetto prevedono l'annotazione semi-automatizzata e manuale di tratti linguistici di natura lessicale e morfosintattica.

Il progetto “AKiD – Acoustic and Kinematic characteristics of speech in Dementia”

Elena Martinelli

Università di Bologna

elena.martinelli12@unibo.it

Gloria Gagliardi

Università di Bologna

gloria.gagliardi@unibo.it

Sabina Capellari

Università di Bologna,

IRCSS Istituto delle Scienze

Neurologiche di Bologna

sabina.cappellari@unibo.it

Chiara Meluzzi

Università degli Studi di Milano

“La Statale”

chiara.meluzzi@unimi.it

Yuka Naito

Università degli Studi di Milano

“La Statale”

yuka.naito@unimi.it

Maria Ranù

Università di Bologna

maria.ranu@unibo.it

1. Stato dell’arte

La demenza è stata riconosciuta come una priorità globale di salute pubblica nel Rapporto dell'Organizzazione Mondiale della Sanità e dell'Alzheimer Disease International (Fleming, Zeisel & Bennet 2020). Con l'incremento della popolazione anziana, anche il numero di persone affette da patologie neurodegenerative è infatti destinato ad aumentare: i dati degli studi epidemiologici sono preoccupanti, poiché si prevede che il numero di diagnosi triplicherà nei prossimi trent'anni, passando dai 55 milioni di casi identificati nel 2019 ai 139 milioni stimati per il 2050, con 7,7 milioni di nuovi casi ogni anno – ovvero uno ogni 4 secondi – e una sopravvivenza media post-diagnosi di 4-8 anni (Long & Weidner 2023; World Health Organization 2023). Questa situazione rappresenta una vera e propria emergenza per i sistemi sanitari e assistenziali pubblici: una popolazione più longeva richiede una redistribuzione più efficiente delle risorse sanitarie, nonché l'adozione di strategie mirate per la prevenzione e la diagnosi precoce. In questo contesto, cruciale è la condizione clinica nota come *Mild Cognitive*

Impairment (MCI; Winblad et al. 2004), che si caratterizza per un lieve deterioramento di uno o più domini cognitivi, autonomia completa nelle ADL-IADL (Lawton & Brody, 1969), e da un rischio maggiore di evoluzione verso la demenza rispetto ad una condizione di invecchiamento neurotipico. Per ottimizzare gli interventi assistenziali, consentire ai pazienti di pianificare la loro vita mentre le capacità decisionali sono ancora integre e facilitare la ricerca di terapie in grado di modificare il decorso della patologia (Sabbagh et al. 2020), l'identificazione tempestiva del declino cognitivo diventa un obiettivo scientifico fondamentale (ISS 2024). La comunità scientifica è da tempo alla ricerca di biomarcatori di malattia che possano supportare la diagnosi precoce della demenza e che, allo stesso tempo, siano di agile ed economico impiego nelle pratiche di *screening* su vasta scala (es. caratteristiche demografiche, indici di rischio cardiovascolare e biomarcatori liquorali e ematici; Alzheimer's Disease International, 2021). Anche le caratteristiche linguistiche, da questo punto di vista, possono essere di prezioso ausilio: dagli anni '70 del Novecento, studi condotti a livello internazionale hanno permesso di evidenziare e descrivere il deterioramento delle competenze linguistiche causato dall'insorgenza di una patologia dementigena, tra cui alto numero di disfluenze, alterazione delle caratteristiche spettro-acustiche e ritmiche della voce, impoverimento morfo-sintattico e lessicale, difficoltà di *word-retrieval*, deficit di pianificazione ed ecolalie a livello pragmatico e testuale (Boschi et al. 2017). Sebbene all'*onset* le abilità in comprensione e la correttezza morfo-sintattica e fonologica delle produzioni verbali non risultino solitamente compromesse, micro-alterazioni linguistiche sono comunque rilevabili diversi anni prima dell'esordio clinico della patologia, nelle fasi prodromiche del disturbo cognitivo (Taler & Phillips 2008). Tali alterazioni sono difficilmente identificabili mediante i tradizionali test neuropsicologici "carta e matita", ma possono essere estratte dall'eloquio dei pazienti sfruttando tecniche proprie del *Natural Language Processing* (NLP, o Trattamento Automatico del Linguaggio), e utilizzate come veri e propri biomarcatori per lo screening e la diagnosi precoce della malattia (Gagliardi 2024).

Sebbene le caratteristiche acustiche/ritmiche della voce siano soggette a naturali cambiamenti con il progredire dell'età, in virtù di eventi fisiologici dell'età anziana quali – ad esempio – l'indebolimento dei muscoli respiratori, l'ossificazione delle cartilagini laringali e l'atrofizzazione dei muscoli faringali e facciali, necessariamente coinvolti in mutamenti nei meccanismi alla base del parlato (Pessin et al. 2017; Rojas, Kefalianos & Vogel 2020), studi precedenti hanno evidenziato come alcuni indici acustici risultino statisticamente significativi nel distinguere i pazienti con demenza dai controlli, e quanto siano altamente dominanti in tutti gli esperimenti di previsione che sfruttano i Decision Tree (Calzà et al. 2021; Gagliardi & Tamburini 2021). D'altra parte, in tale ambito la ricerca si è principalmente focalizzata sull'indagine del dominio soprasegmentale e temporale (es. durata e incidenza delle pause nell'eloquio, *speech rate*; cfr. Canning et al. 2004; König et al. 2015; Carozza, Cantor & Bell-Berti 2012): non è chiaro, dunque, in che misura il declino cognitivo possa comportare variazioni fonetiche a livello segmentale, né se vi siano differenze a livello articolatorio in grado di chiarire le alterazioni acustiche descritte in letteratura. A tal proposito, vari metodi (con diversi gradi di invasività) sono stati sviluppati e testati per l'analisi articolatoria del parlato tipico e atipico (Calzà et al. 2021), ma gli studi in quest'ottica sono stati condotti principalmente sulla lingua inglese e su persone con disturbo della fluenza/balbuzie (Maffia et al. 2021). Sia nella ricerca fonetica che sociofonetica, gli sviluppi recenti nell'interfaccia tra fonetica articolatoria e fonologia hanno evidenziato come caratteristiche importanti per l'analisi della coesione vocale-consonante nel dominio

articolatorio-cinematico aspetti quali l'anticipazione dei gesti apicali sulla vocale precedente e la velocità di picco dei gesti della punta della lingua (Zeroual 2008; Strycharczuk & Scobbie 2015; Celata et al. 2022): tali misure potrebbero dunque essere applicate per quantificare eventuali differenze circa il grado di coesione gestuale tra la produzione di parlanti anziani neurotipici e quelli con declino cognitivo.

Attualmente non vi è dunque una conoscenza esaustiva delle caratteristiche fonetiche associabili al declino cognitivo, poiché – nonostante la mole di ricerche al riguardo (Gosztolya et al. 2019; Hajjar et al. 2023; Luz et al. 2021), non è stata mai condotta un'analisi fonetica “fine”: solo pochi studi pilota hanno considerato il livello segmentale, evidenziando una potenziale riduzione dello spazio vocalico nei soggetti affetti da una patologia dementigena (Nishikawa et al. 2022; Gagliardi & Meluzzi 2025), e non vi sono studi circa il substrato articolatorio delle alterazioni acustiche evidenziate. Mentre i disturbi lessicali, sintattici e pragmatici possono essere associati alla tipica localizzazione delle atrofie cerebrali (come il lobo temporale mediale e l'ippocampo nella malattia di Alzheimer) e correlati ad altre alterazioni cognitive (come la memoria episodica/semantica e le funzioni esecutive), i fondamenti neurobiologici di tali anomalie rimangono ancora in gran parte poco chiari da una prospettiva clinica.

2. Obiettivi

Il progetto “Acoustic and Kinematic Characteristics of Speech in Dementia – AKiD” (finanziato dall'Unione europea- Next Generation EU, Missione 4 Componente 2, codice progetto P2022MMH7R – CUP J53D23017290001) si propone di esplorare il profilo fonetico dell'eloquio di pazienti con declino cognitivo lieve/moderato dovuto a patologia neurodegenerativa / *misfolding* proteico (es. Malattia di Alzheimer, demenza fronto-temporale, malattia a corpi di Lewy; cfr. Abu-Rumeileh et al. 2018; Gorno-Tempini et al. 2011; McKhann et al. 2011; McKeith et al. 2017; Rascovsky, et al. 2011), mediante la combinazione dell'analisi acustica e articolatoria delle produzioni linguistiche dei partecipanti reclutati. AKiD prevede un metodo sperimentale innovativo, poiché l'analisi articolatoria sarà condotta mediante rilevazione dei movimenti linguali tramite *Ultrasound Tongue Imaging* – UTI, una sonda ecografica convessa che – posta sotto al mento del parlante e mantenuta in posizione da un casco – permette di rilevare i movimenti linguali realizzati durante il parlato (cfr. Figura 1).

In particolare, tra le tecnologie articolatorie, l'Ultrasound Tongue Imaging (UTI) offre alcuni vantaggi, tra cui l'essere meno invasiva e/o pericolosa rispetto ad altre tecniche (es. MRI o fMRI) e il consentire al parlante di assumere una posizione normale mentre parla. Attualmente non esistono corpora UTI per lingue diverse dall'inglese (e quelli disponibili sono comunque di dimensioni limitate; cfr. Hu et al. 2022), e il numero di studi sull'eloquio dementigeno che utilizzano l'UTI è estremamente esiguo: a tal proposito, si può citare un unico case-study su una paziente con complesso parkinson-demenza (Hardin 2005; Keller 1987). Lo studio articolatorio condotto su parlanti affetti da demenza rappresenta dunque un elemento fondamentale e innovativo del progetto: sulla base di ricerche precedenti in ambito di linguaggio patologico (sebbene non dementigeno; si veda ad esempio lo studio di Lenoci & Ricci, 2018 su soggetti con balbuzie), si può ipotizzare che le caratteristiche articolatorie possano rivelare in modo più preciso e affidabile l'emergere di un declino cognitivo, anche nelle sue fasi iniziali, rispetto ai soli dati acustici. Uno degli obiettivi principali è pertanto

determinare se caratteristiche acustiche simili sottendano una diversa organizzazione dei gesti linguistici nella produzione del parlato. In particolare, grazie alle immagini ecografiche fornite dall'Ultrasound Tongue Imaging (UTI), verrà indagata la coesione dei gesti articolatori in due modi: qualitativamente, attraverso l'analisi dei movimenti della lingua nella cavità orale, e quantitativamente, valutando il grado di anticipazione del gesto consonantico durante la produzione della vocale accentata in vari contesti fonologici (cfr. Celata et al. 2022).

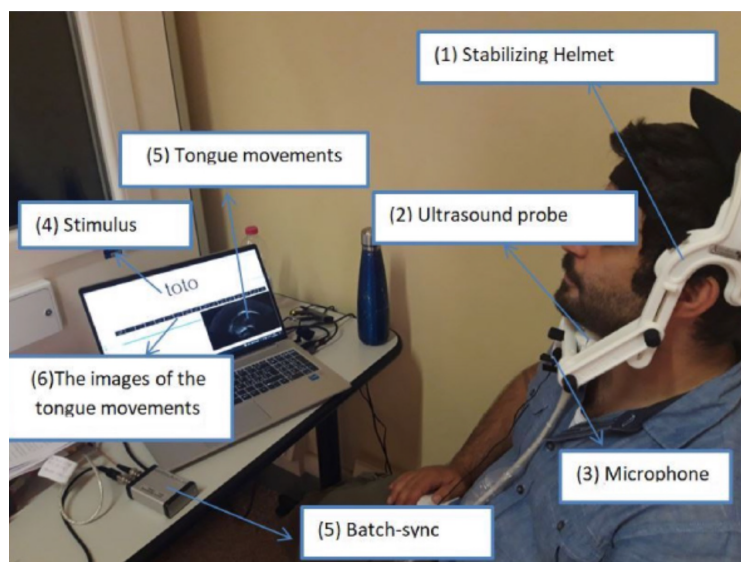


Figura 1. Ultrasound Tongue Imaging (UTI).

3. Il campione

3.1 Il campione

Al fine di raggiungere gli obiettivi prefissati, il progetto prevede il reclutamento di partecipanti adulti/anziani, maschi e femmine, residenti in Emilia-Romagna, divisi in tre coorti bilanciate:

- DemGroup: pazienti di età compresa tra 65 e 85 anni, con diagnosi di declino cognitivo preclinico e clinico dovuto a malattia neurodegenerativa/*misfolding* proteico (es. Malattia di Alzheimer, Demenza Fronto-Temporale, Demenza a Corpi di Lewy), afferenti al Centro Disturbi Cognitivi e Demenze (UOC Clinica Neurologica) dell'IRCCS Istituto delle Scienze Neurologiche di Bologna;
- HeGroup: soggetti anziani neurotipici, bilanciati per sesso, età ed istruzione con i soggetti del DemGroup;
- HAGroup: soggetti adulti neurotipici, bilanciati per sesso ed istruzione con i soggetti del DemGroup.

Il design del campione è finalizzato a contrastare due fenomeni che potrebbero influenzare i risultati delle analisi, ovvero:

- La variazione diatopica, che influisce in modo significativo sulle caratteristiche segmentali e prosodiche del parlato ed è particolarmente rilevante nel caso della

lingua italiana, che presenta sotto questo aspetto una notevole eterogeneità. Per limitare l'influenza della variabilità regionale, l'arruolamento di pazienti e controlli sarà quindi circoscritto a parlanti residenti in Emilia-Romagna;

- La presbifonia: come evidenziato da Pessin et al. (2017) e Rojas et al. (2020), le caratteristiche acustiche e ritmiche della voce subiscono cambiamenti fisiologici, considerati del tutto "naturali", durante il processo di invecchiamento (ad esempio, modifiche nei meccanismi di produzione del parlato che coinvolgono i sistemi respiratorio, fonatorio e sopraglottico). Queste alterazioni, conosciute come "presbifonia", sono così evidenti che gli ascoltatori riescono a stimare l'età di un parlante semplicemente ascoltando la sua voce (Pettorino & Giannini 2011). Per queste ragioni, è previsto il reclutamento di un gruppo di soggetti adulti: il confronto tra le caratteristiche acustico-articolatorie dell'eloquio di soggetti anziani (sia normotipici che affetti da demenza) e quelle di soggetti più giovani permetterà di isolare al meglio le anomalie vocaliche associate al declino cognitivo da quelle legate invece al naturale processo di invecchiamento.

Per il calcolo della numerosità campionaria, sono stati presi come riferimento alcuni studi osservazionali precedenti (Calzà et al. 2021; Gagliardi & Tamburini 2021; Gagliardi & Meluzzi 2025): sulla base di una Power Analysis condotta con il software G*Power, utilizzando un t-test per il confronto tra medie a due code, con un effect size (cioè, Cohen's d) di 0.8 ($\alpha = 0.05$, $1 - \beta = 0.95$, Allocation Ratio $N2/N1 = 1$, actual power = 0.951825), verranno reclutati 42 pazienti (DemGroup) e 42 soggetti di controllo anziani (HEGroup). La dimensione campionaria per l'HAGroup (42 soggetti) è stata invece stimata sulla base della numerosità degli altri due gruppi, seguendo i principi metodologici della Corpus Linguistics, in particolare il principio del "bilanciamento". In totale, dunque, è previsto il reclutamento di 126 partecipanti.

Per ogni coorte sono previsti specifici criteri di inclusione/esclusione. Nello specifico, per il DemGroup i criteri di inclusione sono: l'età (65-85); la presenza di declino cognitivo dubbio, accertato in base ad un punteggio pari a 0.5 (per il MCI) ottenuto alla *Clinical Dementia Rating Scale* (CDR; Morris, 1997), o lieve/moderato, accertato in base ad un punteggio pari a 1/2 ottenuto alla CDR (Morris, 1997); adeguata comprensione verbale. Criteri di esclusione sono invece: cecità, sordità, altri deficit sensoriali; disturbi neurologici acuti, storia di malattia cerebrovascolare; alcolismo; patologie internistiche severe associate a deficit cognitivo; altre cause del disturbo cognitivo (es. ematoma subdurale, neoplasia etc.); malattie psichiatriche severe, trattamento cronico con farmaci psicotropi, utilizzo di farmaci che possano interferire con le funzioni attentive/cognitive a giudizio del medico reclutante.

Per i soggetti di controllo, i criteri di inclusione sono l'età (nello specifico: 65-85 per l'HEGroup; 30-50 per l'HAGroup) e l'assenza di declino cognitivo (accertata da un punteggio alla *Clinical Dementia Rating Scale* pari a 0). I criteri di esclusione sono i medesimi per entrambi i sottogruppi di soggetti neurotipici, ovvero: cecità, sordità, altri deficit sensoriali; disturbi neurologici acuti, storia di malattia cerebrovascolare; alcolismo; patologie internistiche severe associate a deficit cognitivo; altre cause del disturbo cognitivo (es. ematoma subdurale, neoplasia etc.); malattie psichiatriche severe, trattamento cronico con farmaci psicotropi, utilizzo di farmaci che possano interferire con le funzioni attentive/cognitive a giudizio del Medico reclutante; familiarità per demenza ad esordio precoce.

Ai partecipanti è stato somministrato un esteso protocollo neuropsicologico della

durata complessiva di circa 90 minuti, costituito da scale cliniche finalizzate a misurare il grado di autonomia strumentale e della vita quotidiana (ADL-IADL; Lawton & Brody 1969) e la presenza di disturbi neuropsichiatrici (NPI - *Neuropsychiatric Inventory*; Binetti et al. 1995) nonché test neuropsicologici volti ad indagare i principali domini cognitivi (memoria, attenzione, linguaggio, abilità esecutive e visuo-spaziali). In particolare, il protocollo neuropsicologico si compone delle seguenti prove:

- MMSE - Mini-Mental State Examination (Measso et al. 1993; Magni et al. 1996);
- MoCA - Montreal Cognitive Assessment (Aiello et al. 2022);
- Phonemic and semantic fluency test (Novelli et al. 1986);
- Episodic Memory Score (Marra et al. 2016);
- FCSRT - Free and Cued Selective Reminding Test (Frasson et al. 2011)
- FAB - Frontal Assessment Battery (Appollonio et al. 2005);
- TMT - Trail Making Test A and B (Giovagnoli et al. 1996);
- RAVLT- Rey Auditory Verbal Learning test (Caltagirone et al. 1995; Carlesimo et al. 1995a; Carlesimo et al. 1995b; Carlesimo et al. 1996);
- Stroop (Caffarra et al. 2002a);
- Rey-Osterrieth Complex Figure B Test (Caffarra et al. 2002b);
- SAND Screening for Aphasia NeuroDegeneration (Catricalà et al. 2016);
- STAI The State-Trait Anxiety Inventory (Pedrabissi & Santinello 1989);
- BECK BDI- II Beck Depression Inventory - II (Montano & Flebus 2006).

3.2 Materiali

Per raccogliere campioni di linguaggio orale controllato e semi-spontaneo è stata progettata una batteria composta da quattro task linguistici. Nello specifico, per acquisire campioni di eloquio controllato sono previsti i seguenti task:

- a. *Picture naming*, ovvero la denominazione di item lessicali mostrati su una mappa;
- b. *Reading Task*, ovvero lettura di frasi-stimolo (nello specifico: 92). Durante l'esecuzione del task, a un sotto-campione di soggetti sarà richiesto di indossare l'Ultrasound Tongue Imaging per raccogliere informazioni relative all'articolazione delle variabili fonetiche oggetto di analisi.

Il materiale utilizzato per il *picture naming task* e per la lettura di frasi è parzialmente sovrapponibile. In primo luogo, infatti, è stata creata una lista di item costituita da parole target che includono le 7 vocali italiane e le occlusive sorde e sonore dell'italiano. Si tratta di parole esclusivamente bisillabiche, diverse per vocale accentata, qualità vocalica e contesto fonologico della seconda consonante, con un'opposizione tra contesti scempi (es. /'CV.CV/) e geminati (es. /'CVC.CV/).

Successivamente sono stati creati materiali sperimentali specifici per ognuno dei due task. Per il *picture naming task*, sono state selezionate parole piane, appartenenti al lessico italiano e di alta frequenza, quando possibile. In alcuni casi, si è preferito mantenere il contesto fonologico anche con parole di bassa frequenza (come ad esempio "bisso"). In totale sono stati dunque creati 25 stimoli. Sono state sviluppate poi due versioni di questo compito: nella versione "facile" sono presenti 5 stimoli in ogni mappa per un totale di 5 mappe differenti; nella versione "difficile", invece, vengono proposti 8

o 9 stimoli per mappa per un totale di 3 mappe (in particolare: 2 mappe con 8 stimoli ciascuna e 1 mappa con 9 stimoli). Gli stimoli sono stati assegnati in modo casuale a ciascuna mappa, in modo da evitare la frequente ripetizione di uno stesso contesto fonologico. La versione “facile” è stata pensata per agevolare l’elicitazione linguistica nei pazienti affetti da declino cognitivo severo, mentre la versione “difficile” è stata progettata per la somministrazione a tutti i soggetti neurotipici e ai pazienti con declino cognitivo lieve (i.e., MCI). Il task di *picture naming* segue il modello del *map-task*: ogni mappa è infatti fornita in due versioni, una con un percorso tracciato che collega gli elementi sulla mappa, e l’altra senza il percorso, ma con gli stessi elementi rappresentati. Il partecipante (in termini tecnici: *giver of instruction*, o più semplicemente *giver*) riceve la mappa con il percorso e fornisce istruzioni al ricercatore (che tecnicamente assume il ruolo di *follower*) su come collegare le immagini, ricostruendo così il percorso e pronunciando le parole target. Il ricercatore, a sua volta, chiede più volte di ripetere singoli passaggi, cercando di simulare una conversazione naturale, per indurre la pronuncia ripetuta degli stimoli.

Per il task di lettura di frasi, invece, sono state costruite delle frasi stimolo cercando di creare dei contesti frasali più realistici possibile. In particolare, le frasi hanno tutte lunghezza sillabica simile e un contorno prosodico-intonativo uguale; le parole target non sono mai inserite all’inizio o alla fine dell’enunciato, in modo da evitare l’influenza di un’eccessiva energia articolatoria o, al contrario, di un’eccessiva diminuzione del pitch. Si è poi cercato di far precedere - ma soprattutto seguire - all’item target delle parole con consonante occlusiva o fricativa in posizionale iniziale, in modo da riuscire a segmentare correttamente le singole parole target e, contemporaneamente, evitare effetti co-articolatori (es. parola target “fata”, frase-stimolo: “Poi appare la *fata* turchina sul carro”; parola target “scafo”, frase-stimolo: “C’è un buco nello *scafo* della nave”). A partire dalla lista principale, sono state generate altre cinque liste in cui le frasi-stimolo sono presentate in ordine casuale, per evitare che l’inizio o la fine della lettura possano influenzare sempre le stesse verbalizzazioni. Per ogni lista di frasi è stato creato un *power point* da sottoporre al soggetto, in cui ad ogni frase corrisponde una singola slide.

Per l’acquisizione dei campioni di eloquio semi-controllato, invece, è prevista la somministrazione dei seguenti task:

- c. *Picture description*: descrizione orale di due figure stimolo;
- d. *Narrazione biografica*: racconto di una tipica giornata (lavorativa).

Per quanto riguarda il *picture description task*, i materiali impiegati sono tratti rispettivamente da Ciurli et al. (1996) e Berube et al. (2019):



Figura 2. Figura stimolo - Esame del Linguaggio II (Ciurli et al. 1996).



Figura 3. Figura stimolo - Versione a colori della figura “Cookie Theft” (Goodglass, Kaplan & Barresi 2001; Berube et al. 2019).

Infine, nel task di narrazione biografica al soggetto verrà richiesto il racconto di una tipica giornata (lavorativa o non). Le domande stimolo previste per l’elicitazione linguistica nei due task appena descritti sono rispettivamente: “Potresti per favore descrivermi questa figura?” e “Potresti per favore descrivermi la tua tipica giornata (di lavoro)?”.

4. Piano di analisi

I campioni di linguaggio, acquisiti secondo le modalità descritte precedentemente, saranno inizialmente annotati dal punto di vista acustico utilizzando il software open source Praat (Boersma & Weenink 2024). A tal fine, è stato già sviluppato un protocollo di annotazione articolato su cinque livelli, denominati rispettivamente "sentence", "word", "phone", "stop" e "fricative". I primi due livelli sono dedicati alla trascrizione ortografica – rispettivamente – di frasi e parole, accompagnata dalla registrazione di eventuali fenomeni tipici della lingua parlata (ad esempio, sovrapposizioni tra i parlanti). Il terzo livello riguarda la trascrizione fonetica dei singoli suoni, con particolare attenzione all’annotazione delle vocali (inclusa la tonicità e la presenza di eventuale nasalizzazione o rotacizzazione), nonché delle consonanti occlusive e fricative. Per ciascuna classe di consonante è previsto un livello specifico di annotazione e analisi. Il quarto livello, relativo alle occlusive (ovvero il *tier* “stop”), prevede la separazione delle fasi di silenzio da quelle di *Voice Onset Time* (VOT). Infine, il quinto livello, dedicato alle fricative, si concentra sull’annotazione delle vocali che le precedono e/o seguono, poiché una delle variabili da analizzare riguarda proprio la potenziale persistenza o anticipazione della frizione nel contesto vocalico circostante.

Parallelamente, le immagini ecografiche medio-sagittali della forma e della posizione della lingua acquisite tramite l'utilizzo di UTI da un sotto-campione di partecipanti (pazienti con demenza, soggetti di controllo anziani e adulti) saranno annotate in modo semi-automatico per identificare i movimenti e la posizione della lingua durante la fonazione delle vocali e delle consonanti target.

Una volta che il campione sarà annotato a livello acustico-articolatorio, il progetto prevede:

- l'analisi spettrografica contrastiva dei campioni di parlato acquisiti dal gruppo dei pazienti (DemGroup) e dal gruppo bilanciato di coetanei con invecchiamento "tipico" (HEGroup);
- l'analisi dinamica contrastiva dei dati UTI acquisiti da un campione di pazienti con demenza (DemGroup) ed un gruppo bilanciato di coetanei con invecchiamento «tipico» (HEGroup);
- l'analisi contrastiva dei dati ottenuti dal gruppo dei pazienti (DemGroup) e dal gruppo di controlli anziani (HEGroup) rispetto ai dati ottenuti dal gruppo di adulti neurotipici (HAGroup).

Nello specifico, le analisi - supportate dall'impiego di adeguati test statistici applicati tramite il software R (R Core Team 2014) – si baseranno sull'estrazione di parametri considerati rilevanti ai fini dell'evidenziazione di una compromissione acustico-articolatoria. Nello specifico, per il livello acustico saranno oggetto di esame i valori formantici F1, F2 e F3 (Hz), la Frequenza fondamentale - F0 (Hz), lo spazio vocalico/distanza vocalica (Hz), il VOT (ms), la durata del fono, della parola e dei silenzi (ms), il Centro di gravità, e la percentuale di intervalli vocalici e consonantici (%). Al contrario, per il livello articolatorio saranno esaminati i profili linguali (avanzamento/arretramento, innalzamento/abbassamento della lingua), la velocity (mm/s; Scobbie et al. 2013), il grado di anticipazione gesto (Celata et al. 2022), la Pixel difference (McMillan & Corley 2019) e la Scanline-based Pixel Difference (SBPD; Palo, 2019).

Credit author statement

SC: Writing - Review & Editing, Conceptualization, Methodology; GG: Writing - Review & Editing, Conceptualization, Methodology, Project administration, Funding acquisition; CM: Writing - Review & Editing, Conceptualization, Methodology; EM: Writing - Original Draft, Data Curation, Formal analysis; YN: Writing - Review & Editing, Data Curation, Formal analysis; MR: Writing - Review & Editing, Data Curation.

Bibliografia

- Abu-Rumeileh, Samir & Mometto, Nicola & Bartoletti-Stella, Anna et al. 2018. Cerebrospinal Fluid Biomarkers in Patients with Frontotemporal Dementia Spectrum: A Single-Center Study. *Journal of Alzheimer's disease*, 66(2). 551–563. <https://doi.org/10.3233/JAD-180409>
- Aiello, Edoardo Nicolò & Gramegna, Chiara & Esposito, Antonella et al. 2022. The Montreal Cognitive Assessment (MoCA): Updated norms and psychometric

- insights into adaptive testing from healthy individuals in Northern Italy. *Aging Clinical and Experimental Research*. 1- 8. <https://doi.org/10.1007/s40520-021-01943-7>
- Alzheimer's Disease International, Gauthier, Serge & Rosa-Neto, Pedro & Morais, José A., & Webster, Claire 2021. *World Alzheimer Report 2021: Journey through the diagnosis of dementia*. London, England: Alzheimer's Disease International.
- Appollonio, Ildebrando & Leone, Maria Ernesta & Isella, Valeria et al. 2005. The frontal assessment battery (FAB): normative values in an Italian population sample. *Neurological sciences: official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, 26(2). 108–116. <https://doi.org/10.1007/s10072-005-0443-4>
- Bertinetto, Pier Marco & Burani, Cristina & Laudanna, Alessandro & Marconi, Lucia & Ratti, Daniela & Rolando, Claudia, & Thornton, Anna Maria. 2005. *Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS)*. Istituto di Scienze e Tecnologie della Cognizione, CNR.
- Berube, Shauna & Nonnemacher, Jodi & Demsky, Cornelia et al. 2019. Stealing Cookies in the Twenty-First Century: Measures of Spoken Narrative in Healthy Versus Speakers With Aphasia. *American journal of speech-language pathology*, 28(1S). 321–329. https://doi.org/10.1044/2018_AJSLP-17-0131.
- Binetti, Giuliano & Magni, Eugenio & Rozzini, Luca et al. 1995. Neuropsychiatric inventory: validazione italiana di una scala per la valutazione psicopatologica nella demenza. *Giornale di Gerontologia*, 43. 864-865.
- Boersma, Paul & Weenink, David J. M. 2024. Praat: doing phonetics by computer [Computer program]. Version 6.4.12, retrieved 2 May 2024 from <http://www.praat.org/>
- Boschi, Veronica & Catricalà, Eleonora & Consonni, Monica et al. 2017. Connected Speech in Neurodegenerative Language Disorders: A Review. *Frontiers in psychology*, 8, 269. <https://doi.org/10.3389/fpsyg.2017.00269>
- Caffarra, Paola & Vezzadini, Giuliana & Dieci, Francesca et al. 2002a. Una versione abbreviata del test di Stroop: Dati normativi nella popolazione Italiana. *Nuova Rivista di Neurologia*, 12(4). 111-115.
- Caffarra Paola & Vezzadini Giuliana & Dieci Francesca et al. 2002b. Rey-Osterrieth Complex Figure: Normative values in an italian population sample. *Neurological Sciences*, 22. 443-447.
- Caltagirone, Carlo & Gainotti, Guido & Carlesimo, Giovanni Augusto et al. .1995. Batteria per la valutazione del Deterioramento Mentale (parte I): descrizione di uno strumento di diagnosi neuropsicologica. *Archivio di Psicologia, neurologia e psichiatria*, 56(4). 461-470.
- Calzà, Laura & Gagliardi, Gloria & Favretti, Rema Rossini & Tamburini, Fabio. 2021. Linguistic features and automatic classifiers for identifying Mild Cognitive Impairment and Dementia. *Computer, Speech & Language*, 65: 101113. <https://doi.org/10.1016/j.csl.2020.101113>
- Canning, Susan J & Leach, Linda & Stuss, Donald T. et al. 2004. Diagnostic utility of abbreviated fluency measures in Alzheimer disease and vascular dementia. *Neurology*, 62(4). 556–562. <https://doi.org/10.1212/wnl.62.4.556>
- Carlesimo, Giovanni Augusto & Caltagirone, Carlo & Fadda, Luigi et al. 1995. Batteria per la valutazione del deterioramento mentale (parte III): analisi dei profili

- qualitativi di compromissione cognitiva. *Archivio di Psicologia, neurologia e psichiatria*, 56(4). 489-502.
- Carlesimo, Giovanni Augusto & Caltagirone, Carlo & Gainotti, Guido et al. 1995. Batteria per la valutazione del Deterioramento Mentale (parte II): standardizzazione e affidabilità diagnostica nell'identificazione di pazienti affetti da sindrome demenziale. *Archivio di Psicologia, neurologia e psichiatria*, 56(4). 471-488.
- Carlesimo, Giovanni Augusto & Caltagirone, Carlo & Gainotti Guido et al. 1996. The Mental Deterioration Battery: Normative Data, Diagnostic Reliability and Qualitative Analyses of Cognitive Impairment. *European Neurology*, 36(6). 378-384.
- Carozza, Linda & Cantor, Pamela & Bell-Berti, Fredericka. 2012. Temporal structure in the speech of persons with Dementia of the Alzheimer's Type (DAT). *The Journal of the Acoustical Society of America*, 132(3_Supplement). 2089-2089. <https://doi.org/10.1121/1.4755720>
- Catricalà, Eleonora & Gobbi, Elena & Battista, Petronilla et al. 2017. SAND: a Screening for Aphasia in NeuroDegeneration. Development and normative data. *Neurological sciences: official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, 38(8). 1469–1483. <https://doi.org/10.1007/s10072-017-3001-y>
- Celata, Chiara & Meluzzi, Chiara & Bertini, Chiara. 2022. Acoustic and kinematic correlates of heterosyllabicity in different phonological contexts. *Language and Speech*, 65(3). 755-780.
- Ciurli, Paola & Marangolo, Paola & Basso, Anna. 1996. *Esame del Linguaggio II. Manuale e materiale d'esame*. Firenze: Giunti.
- Fleming, Richiard & Zeisel, John & Bennet Kristy. 2020. *World Alzheimer Report 2020: Design Dignity Dementia: dementia-related design and the built environment*, Volume 1. London, England: Alzheimer's Disease International.
- Frasson, Paola & Ghiretti, Roberta & Catricalà, Eleonora et al. 2011. Free and Cued Selective Reminding Test: an Italian normative study. *Neurological sciences: official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, 32(6). 1057–1062. <https://doi.org/10.1007/s10072-011-0607-3>
- Gagliardi Gloria. 2024. Natural language processing techniques for studying language in pathological ageing: A scoping review. *International journal of language & communication disorders*, 59(1). 110–122. <https://doi.org/10.1111/1460-6984.12870>
- Gagliardi, Gloria & Tamburini, Fabio. 2021. Linguistic biomarkers for the detection of Mild Cognitive Impairment. *Lingue & Linguaggio* XX(1). 3-31. <https://dx.doi.org/10.1418/101111>
- Gagliardi, Gloria & Meluzzi, Chiara. 2025. Vowel formant variability in the speech of patients with dementia: a preliminary analysis. *Italiano LinguaDue*, 16(2). 635–651. <https://doi.org/10.54103/2037-3597/27869>
- Giovagnoli, Anna Rita & Del Pesce, M. & Mascheroni S., et al. 1996. Trail Making Test: nor-mative values from 287 normal adult controls. *The Italian Journal of Neurology/ Sciences*, 17(4). 305-309.
- Goodglass, Harold & Kaplan, Edith & Barresi, Barbara. 2001. *The Assessment of Aphasia and Related Disorders*. Austin: Pro-ed.

- Gorno-Tempini, Maria Luisa & Hillis, Argye E. & Weintraub, Sandra et al. 2011. Classification of primary progressive aphasia and its variants. *Neurology*, 76(11). 1006–1014. <https://doi.org/10.1212/WNL.0b013e31821103e6>
- Gosztolya, Gábor & Vincze, Veronika & Tóth, László et al. 2019. Identifying mild cognitive impairment and mild Alzheimer's disease based on spontaneous speech using ASR and linguistic features. *Computer Speech & Language* 53. 181–197.
- Hajjar, Ihab & Okafor, Maureen & Choi, Jinho D. et al. 2023. Development of digital voice biomarkers and associations with cognition, cerebrospinal biomarkers, and neural representation in early Alzheimer's disease. *Alzheimer's & dementia*, 15(1), e12393. <https://doi.org/10.1002/dad2.12393>
- Hardin, Sarah A. 2005. Reliability of hand measures of ultrasound analysis. (Master's thesis). University of South Florida.
- Hu, Shujie & Xie, Xurong & Geng, Mengzhe & Cui, Mingyu & Deng, Jiajun & Li, Guinan & Wang, Tianzi & Liu, Xunying, & Meng, Helen. 2022. Exploiting cross-domain and cross-lingual ultrasound tongue imaging features for elderly and dysarthric speech recognition. *arXiv preprint arXiv:2206.07327*
- Istituto Superiore di Sanità. 2024. *Diagnosi e trattamento di demenza e Mild Cognitive Impairment* (Linee Guida). https://www.iss.it/documents/20126/9140509/LG+Demenza+e+MCI_v3.0.pdf/45961ff0-aa16-5017-1244-8426403600ec?t=1707121203312
- Keller, Eric. 1987. Ultrasound measurements of tongue dorsum movements in articulatory speech impairments. *Folia Phoniatrica et Logopaedica*, 39(2). 93–112.
- König, Alexandra, Satt, Aharon, Sorin, Alexander et al. 2015. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. *Alzheimer's & dementia*, 1(1). 112–124. <https://doi.org/10.1016/j.dadm.2014.11.012>
- Lawton, M. Powell & Brody, Elaine M. 1969. Assessment of older people: self-maintaining and instrumental activities of daily living. *The gerontologist*, 9(3_Part_1). 179–186.
- Lenoci, Giovanna, & Ricci, Irene. 2018. An ultrasound investigation of the speech motor skills of stuttering Italian children. *Clinical Linguistics & Phonetics*, 32(12). 1126–1144
- Long, Simon & Benoist, Chloé & Weidner, Wendy. 2023. *World Alzheimer Report 2023: Reducing dementia risk: never too early, never too late*. London, England: Alzheimer's Disease International.
- Luz, Saturnino & Haider, Fasih & de la Fuente Garcia, Sofia et al. 2021. Editorial: Alzheimer's Dementia Recognition through Spontaneous Speech. *Frontiers in computer science*, 3, 780169. <https://doi.org/10.3389/fcomp.2021.780169>
- Maffia, Marta & De Micco, Rosa & Pettorino, Massimo et al. 2021. Speech Rhythm Variation in Early-Stage Parkinson's Disease: A Study on Different Speaking Tasks. *Frontiers in psychology*, 12, 668291. <https://doi.org/10.3389/fpsyg.2021.668291>
- Magni, Eugenio & Binetti, Giuliano & Bianchetti, Angelo et al. 1996. Mini-Mental State Examination: a normative study in Italian elderly population. *European journal of neurology*, 3(3). 198–202. <https://doi.org/10.1111/j.1468-1331.1996.tb00423.x>

- Marra, Camillo & Gainotti, Gudo & Fadda, Lucia et al. 2016. Usefulness of an Integrated Analysis of Different Memory Tasks to Predict the Progression from Mild Cognitive Impairment to Alzheimer's Disease: The Episodic Memory Score (EMS). *Journal of Alzheimer's disease: JAD*, 50(1). 61–70. <https://doi.org/10.3233/JAD-150613>
- McKeith, Ian G. & Boeve, Bradley F. & Dickson, Dennis W. et al. 2017. Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium. *Neurology*, 89(1). 88–100. <https://doi.org/10.1212/WNL.0000000000004058>
- McKhann, Guy M. & Knopman, David S. & Chertkow, Howard et al. 2011. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia*, 7(3). 263-269. <https://doi.org/10.1016/j.jalz.2011.03.005>
- McMillan, Corey T. & Corley, Martin. 2010. Cascading influences on the production of speech: Evidence from articulation. *Cognition*, 117(3). 243 – 260.
- Measso, Giovanni & Cavarzeran, Fabiano & Zappalà, Giuseppe et al. 1993. The Mini-Mental State Examination: Normative study of an Italian random sample. *Developmental Neuropsychology*, 9(2). 77–85.
- Montano, Antonella & Flebus, Giovanni Battista. 2006. Presentazione del Beck Depression Inventory- Seconda Edizione (BDI-II): conferma della struttura bifattoriale in un campione di popolazione italiana. *Psicoterapia cognitiva e comportamentale*, 12(1). 67-82.
- Morris, John C. 1997. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *International psychogeriatrics*, 9 Suppl 1. 173–178. <https://doi.org/10.1017/s1041610297004870>
- Nishikawa, Kazu & Hirakawa, Rin & Kawano, Hideaki et al. 2022. Analysis of prosodic features and formant of dementia speech for Machine Learning. In *Proceedings of 5th International Conference on Information and Computer Technologies (ICICT)*, New York, USA. 173-176.
- Novelli, Giovanni & Papagno, Costanza & Capitani, Erminio et al. 1986. Tre test clinici di ricerca e produzione les-sicale. Taratura su soggetti normali. *Archivio di Psicologia Neurologia e Psichiatria*, 4(47). 477-506.
- Palo, Pertti. 2019. Measuring pre-speech articulation, PhD Thesis, Queen Margaret University.
- Pedrabissi, Luigi & Santinello, Massimo. 1989. *STAI State-Trait Anxiety Inventory Forma Y Manuale*. Firenze : Giunti Organizzazioni Speciali.
- Pessin, Adriana Bueno Benito & Tavares, Elaine Lara Mendes & Gramuglia, Andrea Cristina Joia et al. 2017. Voice and ageing: clinical, endoscopic and acoustic investigation. *Clinical otolaryngology: official journal of ENT-UK; official journal of Netherlands Society for Oto-Rhino-Laryngology & Cervico-Facial Surgery*, 42(2). 330–335. <https://doi.org/10.1111/coa.12725>
- Pettorino, Massimo & Giannini, Antonella. 2011. The speaker's age: a perceptual study. In LEE, W. S., & ZEE, E. (Eds.). *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII): August 17-21, 2011*. City University of Hong Kong, 1582-1585.

- R Core Team. 2014. *R: A language and environment for statistical computing*. R Foundation for Statistical.
- Rascovsky, Katya & Hodges, John R. & Knopman, David et al. 2011. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain: a journal of neurology*, 134 (Pt 9). 2456–2477. <https://doi.org/10.1093/brain/awr179>
- Rojas, Sandra & Kefalianos, Elaina & Vogel, Adam. 2020. How Does Our Voice Change as We Age? A Systematic Review and Meta-Analysis of Acoustic and Perceptual Voice Data From Healthy Adults Over 50 Years of Age. *Journal of speech, language, and hearing research: JSLHR*, 63(2). 533–551. https://doi.org/10.1044/2019_JSLHR-19-00099
- Sabbagh, Marwan N. & Boada, Mercè & Borson, Soo et al. 2020. Rationale for Early Diagnosis of Mild Cognitive Impairment (MCI) Supported by Emerging Digital Technologies. *The Journal of Prevention of Alzheimer's Disease*, 7(3). 158–164. <https://doi.org/10.14283/jpad.2020.19>
- Scobbie, James M. & Punnoose, Reenu & Khattab, Ghada. 2013. *Articulating five liquids: A single speaker ultrasound study of Malayalam*, in Spreafico, L. & Vietti, A. (eds.) *Rhotics. New Data and Perspectives*. Bozen-Bolzano, Italy: Bozen-Bolzano University Press, 99-124.
- Strycharczuk, Patrycja & Scobbie, James M. 2015. Velocity measures in ultrasound data: Gestural timing of post-vocalic /l/ in English. In *Proceedings of the 18th International Congress of Phonetic Sciences*. International Phonetic Association, 133-136.
- Taler, Vanessa & Phillips, Natalie A. 2008. Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review. *Journal of clinical and experimental neuropsychology*, 30(5). 501–556. <https://doi.org/10.1080/13803390701550128>
- Winblad, Bengt & Palmer, Katie & Kivipelto, Mila et al. 2004. Mild cognitive impairment--beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of internal medicine*, 256(3). 240–246. <https://doi.org/10.1111/j.1365-2796.2004.01380.x>
- World Health Organization. 2023. Dementia. <https://www.who.int/news-room/fact-sheets/detail/dementia>
- Zeroual, Chakir & Hoole, Philip, & Gafos, Adamantios I. 2008. Spatio-temporal and kinematic study of Moroccan Arabic coronal geminate plosives. In *Proceedings of the 8th International Seminar on Speech Production*. 133-136.

Il progetto DiverSIta. Documentare la diversità nell'italiano parlato

Caterina Mauri

Università di Bologna

caterina.mauri@unibo.it

Eugenio Gorla

Università di Torino

eugenio.gorla@unito.it

Beatrice Bernasconi

Università di Torino

beatrice.bernasconi@unito.it

Eleonora Zucchini

Università Masaryk, Brno

zucchini@phil.muni.cz

1. Introduzione: la diversità come tratto costitutivo dell'italiano parlato

Il progetto “DiverSIta – *Diversity in Spoken Italian*”¹ nasce con l’obiettivo di documentare, analizzare e rendere osservabile la diversità dell’italiano parlato contemporaneo, assumendola non come deviazione marginale rispetto a una norma astratta, ma come dimensione strutturale e costitutiva della lingua. La diversità non è dunque concepita come un problema da affrontare, bensì come una risorsa conoscitiva essenziale per comprendere il funzionamento reale dell’italiano.

L’idea di fondo che anima il progetto è che l’italiano non possa essere adeguatamente descritto come un sistema unitario e omogeneo. Al contrario, esso si manifesta quotidianamente attraverso una pluralità di varietà, registri, stili e pratiche discorsive, che emergono dall’interazione tra parlanti diversi per età, livello di istruzione, provenienza geografica, storia migratoria e repertorio linguistico. A queste dimensioni si aggiunge la variabilità legata ai contesti comunicativi, agli scopi pragmatici e alle relazioni sociali tra gli interlocutori, che incidono profondamente sulle scelte linguistiche operate anche dallo stesso individuo in situazioni diverse.

Questa consapevolezza è ben radicata nella tradizione degli studi sul parlato e nella sociolinguistica italiana, ma raramente ha trovato una rappresentazione sistematica in risorse linguistiche ampie, comparabili e accessibili, capaci di restituire la complessità del fenomeno nel suo insieme. DiverSIta nasce proprio per colmare questa lacuna: il progetto, infatti, ha come obiettivo principale la costruzione di un corpus dove le scelte

¹ Progetto finanziato dal MUR nell’ambito del PRIN 2022 PNRR (2023–2026) Progetto finanziato dall’Unione Europea - NextGenerationEU a valere sul Piano Nazionale di Ripresa e Resilienza (PNRR) – Missione 4 Istruzione e ricerca. Codice progetto MUR [P2022RFR8T] - CUP [J53D23017320001]

fatte in sede di corpus design sono informate da considerazioni di natura sociale e metodologica, con un'attenzione particolare agli aspetti di restituzione pubblica della ricerca, e con l'obiettivo ultimo di rendere visibile la diversità come tratto inerente nell'italiano parlato.

1.1 Il contesto scientifico

Lo studio del parlato ha una tradizione consolidata nella linguistica italiana ed europea, ma ha spesso dovuto confrontarsi con limiti strutturali legati alla disponibilità e alla comparabilità dei dati. In Italia, progetti pionieristici come il Lessico di frequenza dell'italiano parlato (LIP) hanno rappresentato un punto di svolta fondamentale, rendendo per la prima volta disponibile una base di dati ampia e sistematica per l'analisi empirica dell'italiano parlato contemporaneo (De Mauro et al. 1993). Il LIP ha avuto un impatto duraturo sugli studi successivi, ma era concepito come un corpus chiuso, costruito per finalità specifiche e non pensato per un'estensione incrementale nel tempo. Analogamente, altri corpora di riferimento, come C-ORAL ROM, hanno fornito dati di alta qualità per lo studio del parlato spontaneo in diverse lingue romanze, adottando standard rigorosi di trascrizione e annotazione prosodica (Cresti e Moneglia 2005), ma sono rimasti anch'essi legati a un disegno progettuale definito a monte e difficilmente estendibile oltre i confini iniziali del progetto.

Più in generale, molte descrizioni del parlato, sia in ambito italiano sia europeo, si sono basate su campioni ristretti, su dati raccolti per obiettivi specifici e non sempre facilmente integrabili tra loro, oppure su risorse che, pur estremamente ricche dal punto di vista analitico, non erano progettate per crescere nel tempo né per dialogare sistematicamente con altri insiemi di dati. In questo quadro, la diversità linguistica è stata spesso osservata in modo frammentario, privilegiando singole dimensioni della variazione – ad esempio quella diatopica o diafasica – senza la possibilità di metterne a confronto sistematico le diverse componenti, in particolare quelle legate ai profili dei parlanti e ai contesti di interazione.

Negli ultimi decenni, la linguistica dei corpora ha messo in evidenza con sempre maggiore chiarezza la necessità di infrastrutture cumulative e modulari, in grado di accogliere dati eterogenei mantenendo criteri comuni di progettazione, annotazione e accesso (Biber et al. 1998; McEnery e Hardie 2012). Per il parlato, questa esigenza è particolarmente rilevante, poiché la variabilità linguistica dipende in larga misura da fattori sociali e situazionali che devono essere documentati in modo esplicito e inclusi nelle possibilità di interrogazione del corpus. In assenza di infrastrutture di questo tipo, il rischio è quello di ridurre la complessità del parlato a pochi tratti selezionati, perdendo di vista il quadro complessivo e la possibilità di analisi comparabili.

Perciò, DiverSIIta mira non solo a produrre nuovi dati sull'italiano parlato contemporaneo, ma anche a contribuire alla costruzione di una base teorico-metodologica comune per lo studio della lingua parlata, capace di sostenere analisi comparabili e di integrare in modo coerente molteplici dimensioni della diversità linguistica.

1.2 Il corpus KIParla come infrastruttura per lo studio del parlato

Il progetto DiverSIIta si innesta sul corpus KIParla, un'infrastruttura per lo studio dell'italiano parlato la cui progettazione prende avvio nel 2016 con l'obiettivo di offrire una base empirica ampia, coerente e progressivamente estendibile, in grado di sostenere

analisi comparabili del parlato contemporaneo (Mauri et al. 2019). KIParla è concepito fin dall'inizio come un corpus modulare e incrementale, destinato a crescere nel tempo attraverso l'aggiunta di nuovi moduli progettati per obiettivi specifici, ma comparabili dal punto di vista strutturale e descrittivo. Questa scelta riflette una precisa impostazione teorico-metodologica, che considera la modularità non come una semplice soluzione tecnica, ma come una condizione necessaria per osservare la variazione linguistica mantenendo il controllo delle variabili sociali e situazionali (Ballarè et al. 2022).

L'infrastruttura KIParla si fonda su trascrizioni ortografiche e conversazionali allineate ai file audio, su un sistema di metadati dettagliato che descrive in modo sistematico sia i parlanti sia le situazioni comunicative, e su un'interfaccia di interrogazione basata su NoSketchEngine. Il sistema di trascrizione adottato tiene conto delle esigenze dell'analisi del parlato spontaneo e dell'interazione, e si ispira alle pratiche e al sistema sviluppato nell'ambito della Conversation Analysis, in particolare da Jefferson (2004). Questa architettura rende possibile l'osservazione del parlato come fenomeno dinamico, mantenendo il legame tra forma linguistica, contesto e uso, e consente di interrogare i dati a diversi livelli di granularità, in linea con le esigenze della ricerca empirica sul parlato (Mauri et al. 2019).

All'avvio di DiverSIIta, KIParla comprendeva moduli già pubblicati, come KIP, focalizzato sulla variazione di registro in contesti universitari, e ParlaTO, dedicato alla variazione sociale e territoriale nell'area torinese. Nel quadro del progetto DiverSIIta, l'infrastruttura è stata ulteriormente ampliata attraverso il completamento e il rilascio di nuovi moduli, tra cui ParlaBO e KIPasti, e attraverso la progettazione e realizzazione dei moduli StraParlaBO e StraParlaTO, dedicati all'italiano parlato in contesti plurilingui urbani. L'integrazione di questi moduli all'interno dell'architettura di KIParla è avvenuta nel rispetto delle scelte strutturali e descrittive comuni, garantendo la piena comparabilità dei dati e rafforzando il carattere cumulativo dell'infrastruttura, come discusso in Ballarè et al. (2022).

1.3 La nozione di diversità adottata in DiverSIIta

Nel progetto DiverSIIta, la nozione di diversità viene intesa in senso ampio e multidimensionale, come categoria descrittiva fondamentale per comprendere il funzionamento dell'italiano parlato contemporaneo. Il termine stesso "diversità" richiama etimologicamente l'idea di separazione e distinzione (dal latino *divertere* 'volgersi altrove, prendere direzioni diverse'), ma anche quella di molteplicità. In questa prospettiva, la diversità non è concepita come scarto rispetto a un centro stabile, bensì come condizione ordinaria dell'uso linguistico, che emerge dall'interazione tra parlanti, contesti e pratiche comunicative.

In primo luogo, diversità significa variazione interna. L'italiano parlato si differenzia sistematicamente in funzione dei contesti comunicativi, dei registri, dei generi discorsivi e delle relazioni sociali tra i parlanti. Queste differenze non rappresentano anomalie, ma costituiscono una componente strutturale della lingua stessa, come ampiamente mostrato dai primi studi di sociolinguistica (v. Labov 1972 e in Italia Berruto 1987). Lo stesso individuo può modulare le proprie scelte linguistiche in modo significativo passando da contesti informali a contesti più istituzionali, da interazioni simmetriche a interazioni asimmetriche, attivando risorse diverse sul piano lessicale, morfosintattico e pragmatico. Documentare questa variazione interna significa quindi decostruire la nozione di lingua

intesa come sistema astratto e decontestualizzato, e dare invece spazio alle pratiche in cui le diverse risorse linguistiche a disposizione degli individui sono reclutate e messe in gioco strategicamente.

In secondo luogo, diversità significa pluralità dei parlanti. L'italiano è parlato da persone con storie linguistiche, biografiche e sociali profondamente diverse, comprese persone con background migratorio e repertori plurilingui complessi, soprattutto nei contesti urbani di Torino e Bologna indagati nell'ambito del progetto. DiverSIIta, in questo senso, risponde a un'esigenza di rappresentazione di queste categorie di persone, spesso associate a contesti di marginalità economica e sociale. Da un punto di vista più strettamente linguistico, le varietà di apprendimento e le diverse strategie di avvicinamento all'italiano, che includono pratiche sia monolingui che plurilingui, sono state dunque considerate sullo stesso piano delle varietà "native" di italiano che più tipicamente sono documentate nei corpora di ambito generale.

Infine, diversità significa dinamicità. La lingua non è un oggetto statico, ma un sistema in continuo mutamento, che si modifica nel tempo attraverso l'uso ripetuto e situato delle risorse linguistiche. La diversità osservabile nel parlato non riguarda soltanto la coesistenza sincronica di varietà diverse, ma anche i processi di cambiamento che attraversano la lingua, come l'emergere di nuove costruzioni, la riorganizzazione di schemi pragmatici e la ridefinizione di categorie discorsive. In questa prospettiva, la documentazione del parlato deve essere in grado di registrare il cambiamento, offrendo strumenti per osservare processi di emergenza, stabilizzazione e trasformazione delle pratiche linguistiche nel tempo, come suggerito da approcci usage-based (Bybee 2010; Hopper 2011).

DiverSIIta assume queste tre dimensioni – variazione interna, pluralità dei parlanti e dinamicità – come assi portanti della propria impostazione teorica e metodologica. La diversità non è quindi un semplice oggetto di studio, ma una lente interpretativa che orienta le scelte di progettazione dei moduli, di raccolta dei dati e di analisi, consentendo di osservare l'italiano parlato come fenomeno complesso, stratificato e in costante evoluzione.

2. DiverSIIta: obiettivi e metodi

DiverSIIta nasce con l'obiettivo di tradurre operativamente questa concezione ampia di diversità, intervenendo su più livelli. Da un lato, il progetto mira ad ampliare la base di dati disponibile per lo studio del parlato, includendo varietà, contesti e parlanti finora meno rappresentati. Dall'altro, intende riflettere in modo esplicito sulle modalità di costruzione e documentazione delle risorse, in linea con i principi di trasparenza e riusabilità.

Gli obiettivi del progetto si articolano quindi lungo tre assi principali: l'espansione del corpus KIParla attraverso nuovi moduli, lo sviluppo di pratiche di documentazione metodologica coerenti con i principi FAIR e il coinvolgimento attivo delle comunità esterne all'accademia nella produzione e restituzione della conoscenza linguistica.

2.1 I moduli sviluppati nel progetto

Nel quadro di DiverSIIta sono stati sviluppati e completati diversi moduli che contribui-

scono in modo complementare alla documentazione della diversità dell’italiano parlato, sia sul piano della variazione sociale e territoriale sia su quello delle pratiche interazionali e dei profili dei parlanti. Questi moduli si collocano all’interno dell’infrastruttura KIParla e ne ampliano in modo significativo la rappresentatività mantenendo al tempo stesso criteri comuni di progettazione e descrizione .

Il modulo ParlaBO è dedicato all’italiano parlato nella città metropolitana di Bologna e documenta la variazione sociale attraverso interviste semi-strutturate raccolte in un contesto urbano caratterizzato da elevata eterogeneità sociolinguistica. Il modulo comprende 85 interviste realizzate tra il 2021 e il 2024, per un totale di 155 parlanti coinvolti e oltre 700.000 token trascritti. I parlanti presentano profili differenziati per età (Tab. 1), livello di istruzione e occupazione, consentendo di osservare in modo sistematico la relazione tra scelte linguistiche e variabili sociali in un contesto urbano in trasformazione. ParlaBO si affianca al modulo ParlaTO, sviluppato in precedenza nell’area torinese, rendendo possibile un confronto tra due grandi contesti urbani italiani sulla base di dati comparabili.

Fascia d’età	Ore
Giovani (16 - 29)	18:46:50
Adulti (30 - 59)	24:21:43
Anziani (> 60)	22:34:52
Totale	65:43:25

Tabella 1. Dimensioni modulo ParlaBO

Il modulo KIPasti arricchisce ulteriormente il corpus attraverso la raccolta di conversazioni spontanee in contesti domestici, registrate durante interazioni a tavola tra familiari o amici. Questo modulo, raccolto tra il 2020 e il 2024, comprende 63 conversazioni per un totale di 147 parlanti e circa 483.000 token trascritti. I dati coprono diverse aree geografiche italiane e presentano un campione socialmente bilanciato sulla base dei dati ISTAT 2020 (Tab. 2), offrendo una rappresentazione ampia del parlato naturale in situazioni non elicitate e caratterizzate da relazioni simmetriche, in cui è altamente probabile l’uso di registri informali. KIPasti consente di osservare fenomeni interazionali e pragmatici difficilmente accessibili attraverso interviste o contesti istituzionali, principalmente legati alla gestione dei turni totalmente libera, che conduce a innumerevoli sovrapposizioni, e all’assenza di topic predefiniti.

Area di raccolta	Ore
Nord	19:25:31
Centro	06:50:19
Sud e isole	16:33:29
Totale	42:49:19

Tabella 2. Dimensioni modulo KIPasti

Accanto a questi moduli, nel progetto DiverSIta sono stati realizzati i moduli StraParlaBO e StraParlaTO, dedicati all’italiano parlato all’interno di comunità con background migratorio internazionale nei contesti urbani di Bologna e Torino. Questi moduli

rappresentano un'estensione cruciale dell'infrastruttura KIParla, poiché consentono di documentare repertori plurilingui complessi e pratiche linguistiche emergenti in contesti urbani multiculturali. Il bilanciamento dei moduli prevede, per ciascuna città, circa 64 ore complessive di registrazioni, articolate in interviste semi-strutturate, conversazioni libere e interazioni di gruppo task-oriented. I metadati raccolti includono informazioni dettagliate sui parlanti, come età, genere, livello di istruzione, professione, lingua o lingue di origine, paese e regione di nascita, età di arrivo in Italia e durata della permanenza, consentendo analisi *fine-grained* della relazione tra pratiche linguistiche e traiettorie biografiche.

Nel modulo StraParlaBO sono stati coinvolti 140 partecipanti, mentre il modulo StraParlaTO include 191 parlanti, provenienti da Paesi diversi: Marocco, Cina, Ucraina e Bangladesh per Bologna, Perù, Marocco, Romania e Moldavia per Torino. I dati includono sia interazioni a tema libero, caratterizzate da un intervento minimo del ricercatore, sia interviste su temi predefiniti e attività collaborative sviluppate insieme alle comunità (cf. Tab. 3 e 4). Questa articolazione consente di osservare l'italiano parlato non solo dal punto di vista delle varietà di apprendimento, lo sguardo privilegiato ad esempio dai learner corpora, ma di considerarlo come risorsa pienamente integrata in repertori plurilingui dinamici, in cui convivono varietà dell'italiano e lingue di origine, spesso usati congiuntamente nell'ambito di pratiche plurilingui.

Comunità	Interviste	Conversazione libera	Totale
Marocco	10:42:04	5:13:49	15:55:53
Bangladesh	9:58:16	6:52:56	16:51:12
Cina	9:19:00	8:49:36	18:08:36
Ucraina	9:48:30	7:17:36	17:06:06
Totale	39:47:50	28:13:57	68:01:47

Tabella 3. Dimensioni modulo Stra-ParlaBO

Comunità	Interviste	Conversazione libera	Totale
Perù	09:38:33	09:08:10	18:46:00
Marocco	11:44:08	10:20:15	22:04:23
Romania	12:23:32	7:35:48	19:59:20
Totale	33:46:13	27:04:13	60:50:26

Tabella 4. Dimensioni modulo Stra-ParlaTO

2.2 Metodologia, documentazione e principi FAIR

Un elemento centrale del progetto DiverSIIta riguarda la documentazione sistematica delle scelte metodologiche che guidano la raccolta dei dati e la costruzione del corpus. In linea con l'impostazione generale del progetto, la metodologia non è concepita come un insieme neutro di procedure tecniche, ma come una componente integrante dell'interpretazione dei dati e della loro riusabilità nel tempo. Per questo motivo, DiverSIIta adotta una prospettiva esplicitamente autoriflessiva, che mira a rendere trasparenti le decisioni prese in ogni fase del processo di costruzione della risorsa, dalla

progettazione dei moduli alla raccolta e trascrizione dei dati, fino alla gestione dei metadati e al rilascio del corpus.

Dal punto di vista della raccolta dei dati, i diversi moduli del progetto sono costruiti a partire da disegni di campionamento differenziati, calibrati sugli obiettivi specifici di ciascun modulo, ma armonizzati attraverso criteri comuni. Le scelte relative ai tipi di interazione, al ruolo del ricercatore, al grado di strutturazione dell'evento comunicativo e alle condizioni di registrazione sono documentate in modo sistematico, al fine di rendere esplicito l'impatto che tali fattori possono avere sulle pratiche linguistiche osservate. Questa attenzione è particolarmente rilevante per il parlato, dove il confine tra "dato" e "contesto" è intrinsecamente poroso e dove la presenza stessa del ricercatore può influenzare l'andamento dell'interazione.

Un aspetto cruciale della metodologia adottata in DiverSIta riguarda la trascrizione. Le trascrizioni sono allineate ai file audio e seguono convenzioni coerenti all'interno dell'infrastruttura KIParla, ispirate alle pratiche tipicamente usate nell'ambito della Conversation Analysis (Jefferson 2004), ma adattate alle esigenze di un corpus di ampia scala e orientato alla riusabilità. Le scelte trascrittive operano un bilanciamento tra fedeltà al dato e leggibilità, evitando una rappresentazione eccessivamente *fine-grained* che renderebbe i dati difficilmente interrogabili, ma preservando al tempo stesso informazioni rilevanti per l'analisi interazionale, come sovrapposizioni, pause e fenomeni di riparazione. Questo compromesso riflette una scelta metodologica consapevole, orientata a rendere il corpus utilizzabile da una pluralità di comunità di ricerca con interessi e competenze diverse.

Particolare attenzione è dedicata alla progettazione e alla gestione dei metadati. I metadati di DiverSIta descrivono in modo dettagliato sia i parlanti sia le situazioni comunicative, includendo informazioni socio-demografiche, biografiche e linguistiche, nonché caratteristiche dell'interazione e del contesto di registrazione. Nei moduli StraParla, in particolare, i metadati consentono di ricostruire le traiettorie linguistiche dei parlanti, rendendo possibile l'analisi della relazione tra pratiche linguistiche e percorsi di vita senza ricorrere a categorie rigide o predefinite quali ad esempio quella di parlante nativo e non-nativo, o di prima e seconda generazione. La strutturazione dei metadati segue criteri che ne facilitano l'interoperabilità con altre risorse e infrastrutture di ricerca.

Un aspetto centrale del progetto consiste anche nella descrizione dei processi che hanno portato alla costruzione dei due nuovi moduli. Protocolli di raccolta, linee guida per la trascrizione, schemi di metadati e decisioni relative al rilascio dei dati sono resi espliciti e accessibili, con l'obiettivo non solo di rendere il corpus consultabile, ma anche di esplicitare le scelte che hanno portato alla sua produzione. In questo quadro, un ruolo centrale è svolto dall'infrastruttura CLARIN-IT, all'interno della quale il repository di KIParla è ospitato e reso accessibile alla comunità scientifica. L'integrazione di KIParla in CLARIN-IT garantisce un ancoraggio stabile a un'infrastruttura di ricerca europea, favorendo la visibilità, la preservazione a lungo termine e l'interoperabilità dei dati con altre risorse linguistiche.

Accanto al rilascio dei dati attraverso CLARIN-IT, DiverSIta adotta pratiche di documentazione aperta tramite piattaforme di hosting e versionamento come GitHub, utilizzate per condividere linee guida, script, schemi di annotazione e materiali di supporto. L'uso di GitHub consente di tracciare in modo trasparente l'evoluzione delle scelte metodologiche, di rendere esplicite le revisioni e di favorire il riuso e l'adattamento degli strumenti sviluppati nel progetto da parte di altre comunità di ricerca. In questo

senso, la documentazione non è concepita come un prodotto statico, ma come un processo aperto e iterativo, che accompagna l'intero ciclo di vita della risorsa.

In questo quadro si colloca l'adesione esplicita di DiverSIta ai principi FAIR (Wilkinson et al. 2016), che orientano l'intero progetto. I dati sono progettati per essere *findable*, attraverso una chiara indicizzazione e descrizione all'interno delle infrastrutture di riferimento; *accessible*, grazie al rispetto delle normative etiche e legali relative alla privacy e all'acquisizione del consenso informato; *interoperable*, grazie all'adozione di standard condivisi per la trascrizione e i metadati e all'integrazione in CLARIN-IT; e *reusable*, mediante una documentazione dettagliata e versionata che ne consenta il riuso in contesti di ricerca e di formazione diversi da quelli originari.

Nel loro insieme, le scelte metodologiche e documentali adottate in DiverSIta contribuiscono a costruire una risorsa che non si limita a raccogliere dati sul parlato, ma che rende esplicite le condizioni della loro produzione e le categorie attraverso cui vengono interpretati. In questo senso, la metodologia diventa parte integrante dell'oggetto di studio, consentendo di osservare la diversità linguistica non solo nei dati, ma anche nelle pratiche di ricerca che ne rendono possibile la documentazione.

3. Comunità, restituzione e sperimentazione sociale

Accanto alle attività di raccolta, documentazione e analisi dei dati, DiverSIta attribuisce un ruolo centrale al coinvolgimento delle comunità esterne all'accademia e alla restituzione della ricerca. Il progetto adotta una prospettiva di *community outreach* che concepisce la ricerca linguistica come un'attività dialogica, fondata sullo scambio tra ricercatori e società. In questa prospettiva, la restituzione non rappresenta una fase finale o accessoria del progetto, ma una componente strutturale che accompagna l'intero ciclo di vita della ricerca e che contribuisce a orientarne scelte, pratiche e priorità. Questa impostazione si traduce, in primo luogo, nella costruzione di relazioni continuative con associazioni, scuole, enti locali e servizi pubblici operanti nei contesti urbani coinvolti dal progetto. In particolare nei moduli StraParlaBO e StraParlaTO, il dialogo con le comunità è stato parte integrante della progettazione delle attività di raccolta dei dati, contribuendo a definire temi, formati di interazione e modalità di partecipazione coerenti con i bisogni dei contesti indagati.

Un ruolo centrale nelle attività di restituzione è svolto dall'Oral Compass, concepito come un insieme di risorse orientate a promuovere una riflessione critica sull'uso dell'italiano nel parlare di diversità culturale, identità, appartenenza e differenza. L'Oral Compass non è pensato come uno strumento normativo o prescrittivo, ma come un dispositivo di orientamento che invita a interrogare le categorie linguistiche e discorsive attraverso cui la diversità viene costruita e rappresentata nello spazio pubblico. A partire dai dati del corpus, l'Oral Compass mette in evidenza come scelte linguistiche apparentemente neutre possano veicolare stereotipi, semplificazioni o categorizzazioni rigide, offrendo al tempo stesso esempi di pratiche discorsive alternative e più inclusive, basate sull'osservazione dell'uso effettivo della lingua.

Accanto all'Oral Compass, DiverSIta ha sviluppato una serie di attività di divulgazione partecipativa che mirano a rendere accessibili i risultati della ricerca a pubblici non specialistici. Tra queste rientrano eventi pubblici e incontri di formazione rivolti a insegnanti, operatori sociali e studenti, nei quali i dati del corpus vengono utilizzati come punto di partenza per riflettere sulle pratiche linguistiche quotidiane e sulle

loro implicazioni sociali. Queste attività non si limitano a “tradurre” contenuti scientifici in forme semplificate, ma sperimentano modalità di coinvolgimento attivo che valorizzano l’esperienza linguistica dei partecipanti e favoriscono un confronto critico sui temi della diversità e del plurilinguismo.

Un aspetto particolarmente innovativo delle attività di restituzione di DiverSIta è rappresentato dall’uso di strumenti ludici, come giochi da tavolo progettati a partire dai dati del corpus. Questi strumenti danno la possibilità agli utenti non specializzati di esplorare in modo informale e collaborativo fenomeni linguistici complessi, attivando forme di riflessione condivisa, e abbassando le barriere di accesso alla ricerca linguistica.

4. Conclusioni e prospettive

Nel suo complesso, DiverSIta si configura come un progetto che integra in modo sistematico ricerca empirica sul parlato e riflessione metodologica, assumendo la diversità come chiave interpretativa trasversale. L’espansione e il rafforzamento dell’infrastruttura KIParla, la progettazione di nuovi moduli dedicati a contesti di plurilinguismo urbano, e l’attenzione costante alla documentazione delle scelte metodologiche contribuiscono a costruire una risorsa che non si limita a descrivere l’italiano parlato, ma ne rende osservabili le condizioni di uso, di variazione e di trasformazione nel tempo. In questo senso, DiverSIta propone un modello di ricerca sul parlato fondato su cumulatività, comparabilità e apertura, in continuità con una linea di lavoro che concepisce i corpora come infrastrutture vive e incrementali (Mauri et al. 2019; Ballarè et al. 2022).

A partire da questo impianto, è possibile individuare alcune prospettive di sviluppo che DiverSIta contribuisce a tracciare per la ricerca futura sul parlato. Una prima prospettiva riguarda il rapporto tra corpora di parlato spontaneo e risorse annotate strutturalmente, in particolare le treebank. I dati raccolti e documentati nel progetto mostrano con chiarezza che una descrizione sintattica del parlato non può prescindere da una base empirica ampia, sensibile al contesto e ricca di metadati. In questo quadro, iniziative come KIParla Forest (Pannitto et al. 2025) indicano una direzione possibile per l’integrazione tra corpora conversazionali e annotazione per dipendenze non come semplice applicazione di modelli pensati per lo scritto, ma come terreno di confronto critico sulle categorie stesse della rappresentazione grammaticale. La prospettiva che si apre è quella di un progressivo adattamento degli strumenti di annotazione alla complessità dei dati interazionali.

Una seconda direzione di ricerca riguarda lo studio di fenomeni di contatto nelle pratiche plurilingui. Da un punto di vista più sociologico i dati del progetto mostrano innanzitutto come l’adozione di pratiche linguistiche ibride, in cui l’italiano si alterna con le altre lingue del repertorio, sia funzionale all’autorappresentazione di individui e comunità, mentre in prospettiva acquisizionale è stato possibile osservare come la modalità plurilingue costituisca per alcune persone una strategia di avvicinamento all’italiano. Infine, in una prospettiva più analitica, i dati provenienti da comunità migrate possono contribuire notevolmente allo studio dei diversi pattern di *code-mixing* in presenza di lingue più e meno affini strutturalmente all’italiano.

Inoltre, la natura modulare e incrementale di KIParla, ulteriormente sviluppata nel quadro di DiverSIta, consente di osservare la diversità linguistica come processo, rendendo possibile lo studio di fenomeni di emergenza, stabilizzazione e cambiamento nel tempo. Questo aspetto è particolarmente rilevante nei contesti urbani e plurilingui

documentati dal progetto, dove le pratiche linguistiche riflettono trasformazioni sociali rapide e stratificate. In questa prospettiva, DiverSIta non esaurisce il proprio valore nei risultati immediati, ma pone le basi per analisi comparative di lungo periodo.

Infine, una prospettiva trasversale riguarda il rapporto tra ricerca linguistica e società. Le attività di restituzione e di sperimentazione sociale sviluppate nel progetto confermano che documentazione del parlato può diventare una risorsa condivisa, capace di alimentare consapevolezza linguistica, riflessione critica e pratiche educative. In questo senso, DiverSIta propone un modello di ricerca sul parlato che integra produzione di conoscenza, responsabilità sociale e apertura delle infrastrutture, mantenendo la diversità non come eccezione da ricondurre a norma, ma come tratto costitutivo dell'esperienza linguistica contemporanea.

Bibliografia

- Ballarè, Silvia & Gorla, Eugenio & Mauri, Caterina. 2022. *Italiano parlato e variazione linguistica. Teoria e prassi nella costruzione del corpus KIParla*. Bologna: Pátron Editore.
- Berruto, Gaetano. 1987. *Sociolinguistica dell'italiano contemporaneo*. Roma: Carocci.
- Biber, Douglas & Conrad, Susan & Reppen, Randi. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. New York: Cambridge University Press.
- Bybee, Joan. 2010. *Language usage and cognition*. Cambridge: Cambridge University Press.
- Cresti, Emanuela & Moneglia, Massimo. 2005. *C-ORAL-ROM Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam-Philadelphia: John Benjamins.
- De Mauro, Tullio & Mancini, Federico & Vedovelli, Massimo & Voghera, Miriam. 1993. *Lessico di frequenza dell'italiano parlato*. Milano: ETAS libri.
- Hopper, Paul J. 2011. Emergent Grammar and Temporality in Interactional Linguistics. In P. Auer, S. Pfänder. *Constructions: Emerging and emergent*. 22-44. Berlin-Boston: De Gruyter
- Jefferson, Gail. 2004. Glossary of transcript symbols with an introduction. In G. H. Lerner (ed.). *Conversation Analysis: Studies from the First Generation*, 13-31. Amsterdam-Philadelphia: John Benjamins,.
- Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Mauri, Caterina & Ballarè Silvia & Gorla, Eugenio & Cerruti, Massimo & Suriano Francesco. 2019. KIParla corpus: A new resource for spoken Italian. In *CEUR Workshop Proceedings* 2481, 1-7.
- McEnery Tony & Hardie, Andrew. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- Pannitto Ludovica & Zucchini, Eleonora & Ballarè, Silvia & Bosco, Cristina & Mauri, Caterina & Sanguinetti, Manuela. 2025. Introducing KIParla Forest: seeds for a UD annotation of interactional syntax. In *Proceedings of the Eighth International Conference on Dependency Linguistics* (Depling, SyntaxFest 2025), 54-73. Ljubljana: Association for Computational Linguistics.
- Wilkinson, Mark D. et alii. 2016. *The FAIR Guiding Principles for scientific data management and stewardship*. Sci Data. 2016 Mar 15;3:160018. doi: 10.1038/s41597-019-0009-6.

Il progetto “Racconta – Competenze narrative e sviluppo della lettoscrittura tra variazione interna e plurilinguismo”

Nicola Perugini

Università di Bologna

nicola.perugini6@unibo.it

Chiara Gianollo

Università di Bologna

chiara.gianollo@unibo.it

1. Le competenze narrative dall’uso spontaneo alla ‘lingua dello studio’

Il progetto “Racconta – Competenze narrative e sviluppo della lettoscrittura tra variazione interna e plurilinguismo”, finanziato dal MUR attraverso il bando PRIN 2022¹, ha lo scopo di studiare le competenze narrative durante il più generale processo di maturazione delle competenze comunicative nella scuola secondaria di primo e di secondo grado. La concentrazione sulle competenze narrative è motivata dall’osservazione che esse costituiscono un ponte naturale tra l’uso della lingua orale in situazioni comunicative spontanee e l’uso della lingua in contesti scolastici. Questa osservazione è corroborata da una lunga tradizione di studi (tra i lavori più influenti, Berman 2009; Berman e Slobin 1994; Strömquist e Verhoeven 2004; Hickmann 2004; Gagarina et al. 2016), che hanno mostrato come le competenze narrative si basino sull’abilità di utilizzare strategie linguistiche di coesione testuale e di contestualizzazione che risultano preziose anche per altri compiti di produzione e di comprensione. Nelle parole di Gagarina et al. (2016: 12), “[n]arrative skills are important for academic success, because they are an important discursive form through which academic information is disseminated and acquired”.

La concentrazione sulle competenze narrative, quindi, permette di coniugare un interesse teorico per gli aspetti macro- e micro-strutturali coinvolti nell’elaborazione testuale e un interesse applicativo per i risvolti didattici della ricerca. Il progetto “Racconta”, infatti, prevede come sua parte integrante una fase laboratoriale rivolta agli

¹ PRIN 2022 - Progetti di Rilevante Interesse Nazionale (<https://www.mur.gov.it/it/aree-tematiche/ricerca/programmi-di-finanziamento/progetti-di-rilevante-interesse-nazionale-prin>), co-finanziato con fondi dell’Unione Europea - NextGenerationEU a valere sul Piano Nazionale di Ripresa e Resilienza (PNRR) – Missione 4 Istruzione e ricerca – Componente 2 Dalla ricerca all’impresa (Investimento 1.1, Avviso Prin 2022 indetto con DD N. 104 del 2/2/2022, dal titolo Racconta - Narrative competences and literacy development between internal variation and plurilingualism, codice proposta 2022FFL8BB). Sito web del progetto: <https://site.unibo.it/racconta/it>. La stesura di questo contributo è il risultato di una collaborazione costante tra i due autori. Ai fini del sistema accademico italiano, a Chiara Gianollo va attribuita la sezione 1, a Nicola Perugini le sezioni 2 e 3.

insegnanti, in cui si sfrutterà il potenziale dei dati raccolti come strumento per l'educazione linguistica, attraverso una sensibilizzazione verso l'importanza del lavoro sulle competenze narrative a scuola, che è tra i fattori più importanti per il loro sviluppo (Snow 1983).

In questo contributo, dopo una breve presentazione generale del progetto, ci concentreremo sulla descrizione del protocollo di raccolta dei dati, che è stato elaborato durante la prima fase del progetto. La metodologia di raccolta dati è frutto della collaborazione tra le unità di ricerca dell'Università di Torino e dell'Università di Bologna. Il protocollo che verrà presentato nella Sezione 2 è stato ideato con il fine di raccogliere dati linguistici autentici attraverso una metodologia che combini i vantaggi dell'elicitazione controllata con quelli della raccolta spontanea in situazioni comunicative diverse.

Le competenze narrative sono molto studiate, per parlanti sia monolingui, sia plurilingui. Questa particolare attenzione è dovuta alla loro centralità per la più ampia analisi delle competenze linguistiche a vari livelli (lessico, morfologia, sintassi, pragmatica) e per vari fini (studio dell'acquisizione linguistica in contesti monolingui e plurilingui, valutazione clinica, confronto interlinguistico). Un'ulteriore motivazione per i numerosi studi sulle competenze narrative consiste nella relativa facilità di elicitazione in contesti sperimentali il più possibile naturalistici. Le competenze narrative, infatti, maturano in contesti spontanei in connessione con esperienze vissute in prima persona o osservate, contesti che possono essere agevolmente riprodotti anche nelle condizioni necessariamente controllate di un esperimento. Inoltre, la narrazione è un compito che già in età abbastanza precoce privilegia la componente monologica, permettendo quindi un'osservazione il più possibile scevra di interferenze da parte dei ricercatori. Per questo motivo sono a disposizione per la raccolta dati numerosi strumenti normati, schemi di annotazione e risultati quantitativi e qualitativi da studi precedenti che permettono un ampio grado di comparabilità con quelli che emergeranno dal nostro progetto biennale.

Nonostante questa situazione ampiamente favorevole, manca, nella ricerca precedente, un'attenzione specifica per la fascia di età degli adolescenti. Tenzialmente, infatti, i dati provengono da bambini o da adulti. Uno degli obiettivi del progetto è pertanto quello di ampliare la base empirica degli studi sulle competenze narrative investigando la fascia di età dai 13 ai 19 anni (si veda la Sezione 2.2), in modo da approfondire il legame tra lo sviluppo delle competenze narrative e il grado di scolarizzazione. Questo obiettivo assume particolare rilevanza in considerazione del fatto che, nell'ambito del progetto, vengono raccolti dati sia con parlanti monolingui, sia con parlanti le cui famiglie provengono da un contesto migratorio plurilingue². In questo caso, ci si può attendere che, in misura maggiore o minore a seconda degli individui e delle loro comunità di provenienza, proprio negli anni dell'adolescenza si possano verificare trasformazioni nei rapporti funzionali tra le lingue del repertorio. Per le domande di ricerca che animano il progetto, sarà interessante osservare quanto i diversi equilibri tra le lingue del repertorio influiscano sulle competenze narrative e sulla loro trasferibilità da un sistema linguistico all'altro, sia in termini di strategie testuali, sia in termini di sensibilità alla variazione diafasica.

² L'etichetta 'monolingue' è necessariamente esito di una astrazione, in un contesto, come quello italiano, ancora caratterizzato da un certo grado di dialettologia. L'etichetta è stata applicata, ai fini del nostro progetto, a partecipanti che, nel questionario sociolinguistico somministrato (su cui si veda la Sezione 2.3), dichiarano di essere esposti unicamente o in maniera preponderante all'italiano nel contesto familiare.

Un aspetto prominente della ricerca è, infatti, rappresentato dalla volontà di indagare le competenze narrative in condizioni contestuali variabili, prestando particolare attenzione a parametri quali la multimodalità, la varietà linguistica e il tipo testuale, al fine di comprendere i fattori che determinano la variazione interna. Una dimensione particolarmente rilevante per la raccolta dei dati è quella della variazione diafasica nel contesto della lingua orale. La narrazione, infatti, esercita la sua funzione di ponte tra uso linguistico spontaneo e ‘lingua della scuola’ (quindi, lingua applicata in modo consapevole in funzioni cognitive complesse legate alla lettoscrittura) quando da un contesto primariamente informale si passa a un contesto formale (come quello dell’esposizione orale in una interrogazione, o dell’esposizione scritta in un saggio). Tuttavia, è plausibile che studenti adolescenti abbiano poche occasioni di esercitare il registro formale al di fuori dei contesti scolastici. È quindi interessante osservare quanto effettivamente gli usi linguistici varino a seconda della situazione informale o formale proposta nel contesto sperimentale, e quali tratti linguistici vengano sfruttati per caratterizzare il registro usato di volta in volta. Lo studio di una popolazione adolescente in contesti variabili permetterà, inoltre, di osservare nelle produzioni eventuali cambiamenti in atto in ambiti esposti a ristandardizzazione nell’italiano contemporaneo. In particolare, l’approfondimento sulla gestione delle catene referenziali progettato dall’unità di ricerca di Bologna (Gianollo e Perugini 2025) è ideato per esaminare fenomeni, come la realizzazione dei soggetti pronominali, che sappiamo essere sottoposti a pressioni sia di natura interna (Calaresu e Dal Negro 2018), sia legate al plurilinguismo (Serratrice et al. 2004; Andorno 2013).

È evidente come l’effettiva possibilità di affrontare le questioni qui delineate dipenda crucialmente dalla natura della base empirica costruita nella fase di raccolta dei dati. L’elaborazione del protocollo sperimentale è stata, pertanto, una fase particolarmente delicata del progetto: la Sezione 2 ha lo scopo di presentare le scelte compiute e le loro motivazioni. Nel descrivere il protocollo si procederà a presentare il modello su cui ci si è basati per la progettazione (Sezione 2.1), poi i partecipanti alla raccolta dati (Sezione 2.2) e gli strumenti utilizzati (Sezione 2.3), e infine la procedura seguita per l’elicitazione (Sezione 2.4).

2. La raccolta delle produzioni narrative: il modello “Lang-Sit”

2.1 Le ragioni della scelta

Il protocollo di ricerca sviluppato per il progetto “Racconta” si ispira esplicitamente a quello elaborato nell’ambito del progetto RUEG (Research Unit “Emerging Grammars in Language-Contact Situations”), condotto tra il 2018 e il 2024 da un gruppo di università e istituzioni di ricerca coordinate dalla Università Humboldt di Berlino. L’adozione del protocollo sviluppato da RUEG è giustificata dalla convergenza di alcuni obiettivi e caratteristiche tra i due progetti. In particolare, l’interesse per la metodologia di RUEG deriva dall’impiego di una configurazione sperimentale nota come “Language Situations” (Wiese 2020). Si tratta di un approccio che consente di esplorare in modo sistematico i repertori dei parlanti e permette di raccogliere produzioni linguistiche naturali in diverse situazioni comunicative, sia informali che formali. Nel suo design, ai partecipanti viene presentato un evento fittizio e viene chiesto loro di immaginare di esserne testimoni, descrivendolo in diversi contesti

comunicativi. Ad esempio, possono raccontarlo in una telefonata o in un messaggio WhatsApp agli amici, in una conversazione con uno sconosciuto o in un verbale scritto alla polizia. Diversi studi (Wiese e Pohle 2016; Wiese et al. 2014; 2018; Vangsnes 2020 solo per citarne alcuni) hanno dimostrato sia la flessibilità del metodo, che risulta adattabile facilmente a diverse popolazioni di parlanti e a vari obiettivi di ricerca, sia la sua efficacia nel fornire dati comparabili e naturali, capaci di cogliere le scelte linguistiche legate al registro, tanto all'interno di uno stesso parlante quanto tra comunità linguistiche diverse (Wiese 2020). Poiché questo approccio prevede di riportare un evento sotto forma di testimonianza, le produzioni che risultano dalla sua adozione sono principalmente di tipo narrativo, il che lo rende particolarmente adatto al nostro progetto (per le caratteristiche propriamente narrative delle produzioni raccolte, si veda Canuto et al. 2025).

A giustificare l'adozione del protocollo RUEG sta anche il fatto che in quel progetto il metodo "Language Situations" è stato usato per raccogliere produzioni narrative in varie lingue (inglese, tedesco, greco, russo e turco) e anche da parte di parlanti bilingui, con combinazioni linguistiche diverse e appartenenti a due fasce di età: adolescenti (13-19 anni) e adulti (20-37). RUEG condivide dunque con il nostro studio, oltre all'interesse per la variazione interna tra registri formali e informali, anche quello per il fenomeno del plurilinguismo e per parlanti ancora in età scolare.

Di primario interesse per i nostri obiettivi è anche il fatto che il protocollo RUEG privilegi un approccio ecologico alla raccolta dati, assicurando che le produzioni linguistiche raccolte siano il più possibile naturali e rappresentative dell'uso autentico della lingua. Per questi motivi gli strumenti e le procedure di raccolta dati messi a punto e resi disponibili dai ricercatori RUEG sono sembrati particolarmente adatti a soddisfare le esigenze di ricerca del nostro progetto. Essi permettono infatti di indagare come parlanti adolescenti con repertori linguistici diversi modulano il proprio uso della lingua a seconda del contesto comunicativo. Una tale condivisione di strumenti e metodi inoltre garantisce la confrontabilità dei dati e dei risultati. Va sottolineato poi che il corpus di dati raccolti dal progetto RUEG con questa metodologia è già pubblicato e consultabile (Wiese et al. 2019), ma non include né l'italiano come lingua maggioritaria, né l'arabo e il cinese come lingue ereditarie, offrendo l'opportunità al nostro gruppo di ricerca di contribuire in maniera significativa ad arricchire la mole di dati a disposizione.

2.2 I partecipanti

Per il reclutamento dei partecipanti, ci siamo rivolti alle scuole secondarie di primo e secondo grado del territorio metropolitano delle città di Torino e Bologna. La scelta degli istituti scolastici è avvenuta sulla base di criteri predefiniti. Nel caso delle scuole secondarie di primo grado, si è cercato di includere alunni provenienti da una zona dell'area metropolitana considerata "media" sulla base di tre parametri principali. Il primo parametro è il reddito medio rispetto ai valori nazionali, in riferimento ai dati Istat del 2023, per garantire che il contesto economico di provenienza fosse rappresentativo di un'ampia fascia della popolazione. Il secondo criterio riguarda la percentuale di popolazione straniera residente nel quartiere, un dato rilevante dal punto di vista sociale, ma soprattutto per avere informazioni sulla diversità linguistico-culturale dell'area considerata. Sono poi stati presi in esame anche gli indici di status

socioeconomico elaborati dagli uffici di statistica dei due comuni, e infine questi criteri sono stati combinati con la rete di contatti già esistenti tra i due Dipartimenti coinvolti e gli Istituti d'istruzione del territorio. Per le scuole secondarie di secondo grado, invece, sono stati selezionati studenti provenienti da licei e da istituti tecnici e professionali, al fine di garantire un'ampia rappresentatività dei percorsi formativi in cui sono inseriti gli studenti inclusi nel campione. L'obiettivo minimo che il progetto si è posto per la raccolta dei dati in ognuna delle due città è il seguente:

- una classe terza di scuola secondaria di primo grado, composta da 20-25 studenti di età compresa tra i 13 e i 14 anni;
- una classe quarta di un istituto tecnico-professionale, composta da 20-25 studenti di età compresa tra i 17 e i 19 anni;
- una classe quarta di un liceo, composta da 20-25 studenti di età compresa tra i 17 e i 19 anni;
- un campione aggiuntivo di 10 studenti di scuola secondaria di primo grado, selezionati per il loro background linguistico ereditario.

Le fasce di età 13-14 anni e 18-19 anni sono state scelte strategicamente. La prima rappresenta l'ultimo anno di istruzione comune in Italia prima della suddivisione tra percorsi scolastici differenziati, mentre la seconda permette di coinvolgere studenti che si trovano nella parte finale del loro percorso nella scuola secondaria di secondo grado e di confrontare i dati raccolti con quelli di studi precedenti su adulti (Chini 2005 tra gli altri) e con dati di altri progetti in corso o recentemente completati presso le Università di Bologna e Torino (Univers-ITA, ParlaTo, KIParla). La differenziazione tra istituto professionale, tecnico e liceo consente inoltre di analizzare l'interazione tra parametri socioculturali, pratiche di scolarizzazione e competenze linguistiche diverse. Per il sottogruppo di 10 studenti selezionati sulla base del background linguistico ereditario, sono stati raccolti testi sia in italiano sia nella lingua di origine, con l'obiettivo di indagare più a fondo le dinamiche di contatto linguistico e le competenze narrative in condizioni di bilinguismo. Le due lingue ereditarie su cui si è basata la selezione sono il cinese mandarino per la città di Bologna e una varietà di arabo (stabilita sulla base dell'effettiva disponibilità di partecipanti) per la città di Torino. Queste lingue sono state selezionate per motivi tipologici e sociolinguistici. Dal punto di vista tipologico, il cinese è una lingua "topic-prominent" con un sistema di riferimento anaforico caratterizzato da una generalizzata omissione degli argomenti a riferimento anaforico (*radical pro-drop*) e dall'assenza di articoli. L'arabo, pur essendo una lingua "subject-prominent" come l'italiano, presenta differenze nel sistema referenziale e nelle condizioni di legittimazione del soggetto nullo. Dal punto di vista sociolinguistico, entrambe le lingue mostrano variazioni significative tra contesti formali e informali, rendendo possibile la definizione della competenza degli studenti lungo dimensioni di variazione specifiche anche nella lingua ereditaria. Infine, le possibili ricadute applicative della nostra ricerca sono rilevanti per la formazione degli insegnanti e per le attività di divulgazione, considerando la crescente presenza demografica delle comunità linguistiche di riferimento in Italia.

2.3 Gli strumenti

Per quanto riguarda gli strumenti usati per la raccolta dati, il protocollo prevede l'uso di un video stimolo per le produzioni orali e un questionario utile a tracciare il profilo sociolinguistico dei partecipanti. Per la raccolta delle produzioni narrative si è adottato come stimolo il video muto messo a punto dal gruppo RUEG, in cui viene mostrato un incidente stradale minore nel parcheggio di un supermercato³. Il questionario sociolinguistico invece è stato costruito appositamente per il progetto “Racconta”, sulla base di questionari ideati per progetti precedenti, aventi obiettivi simili (Chini e Andorno 2018; Fiorentini e Gianollo 2018). Il questionario ha un totale di 54 domande divise in 7 sezioni, ognuna delle quali mira alla raccolta di informazioni diverse. Se ne danno i titoli di seguito:

1. Dati anagrafici
2. La tua famiglia
3. Scolarizzazione
4. Esposizione linguistica
5. Competenze linguistiche
6. Tempo libero
7. Esposizione alla narrazione

Come si vede dall'elenco soprastante, la sezione 1 ha l'obiettivo di raccogliere informazioni su data e luogo di nascita, età di arrivo in Italia e quartiere di residenza. A questa fa seguito una seconda parte volta a indagare lo status socioeconomico di provenienza (Paradis 2011), in cui vengono chieste informazioni sulla famiglia (da quante persone è composta, quali membri vivono in casa con l'informante e quali eventualmente vivono in altri luoghi), sul titolo di studio e sull'occupazione dei genitori. La terza sezione riguarda invece gli anni di frequenza della scuola in Italia o in altro paese (a partire dall'asilo nido fino al grado frequentato nel momento della compilazione). La sezione successiva rappresenta il cuore del questionario e comprende domande utili a capire da quali lingue è composto il repertorio di ciascun informante, quali sono le scelte e gli usi linguistici che gli studenti compiono nei vari contesti di vita e con diversi tipi di interlocutori. La sezione quinta è invece basata su un'analoga sezione del questionario messo a punto dal progetto RUEG, volta a raccogliere informazioni sulle competenze dell'informante sulle lingue del suo repertorio, tramite l'uso di scale di autovalutazione. Poiché si è visto che le attività extrascolastiche influenzano le competenze linguistiche, è sembrato opportuno inserire domande anche sul tempo libero (sezione 6), al fine di avere informazioni sulla rete sociale dei parlanti e sui tipi di attività a cui prendono parte che possano essere linguisticamente rilevanti (Paradis 2011). Questa sezione permetterà anche di esplorare il ruolo dei social media nelle pratiche di narrazione, un fattore che si prevede sempre più rilevante data l'intensità dell'input digitale nell'adolescenza. Un focus particolare è poi stato dedicato alla esposizione alla narrazione, aspetto esplorato specificamente con la sezione 7, modellata sui risultati dell'indagine TiLeggo condotta nel 2016 dall'Istituto Treccani⁴ e

³ Il video è disponibile sul sito del progetto RUEG: <https://www.linguistik.hu-berlin.de/en/institut-en/professuren-en/rueg>

⁴ La presentazione del progetto TiLeggo e i risultati dell'indagine sono disponibili sul sito <https://tileggo.treccani.it/indagini-lettura/>

dedicata alle abitudini di lettura degli adolescenti italiani. A queste sono state aggiunte anche domande sulle abitudini di scrittura, privata e online, nonché sulle pratiche di narrazione orale in famiglia e tra pari.

2.4 La procedura

Poiché l'obiettivo con cui è stato ideato il protocollo è quello di elicitare dati linguistici autentici in situazioni di cui però si possano controllare le variabili di interesse, si è deciso di utilizzare il metodo "Language Situations" (Wiese 2020), così come è stato implementato nel progetto RUEG e descritto nella Sezione 2.1. Il metodo è infatti stato concepito per elicitare produzioni linguistiche comparabili e naturalistiche in diversi registri (formali e informali). Di seguito si illustrerà come, a partire dagli strumenti e dal protocollo messi a punto da RUEG, sia stata implementata e condotta la raccolta dati del progetto "Racconta"⁵. Essa si articola in due fasi, con un primo incontro introduttivo in classe, seguito dalla raccolta delle produzioni orali narrative in due differenti contesti comunicativi:

- **narrazione orale in contesto informale:** registrazione di un messaggio vocale WhatsApp a un amico
- **narrazione orale in contesto formale:** registrazione di una testimonianza alla polizia stradale

Nel primo incontro i ricercatori presentano il progetto all'intero gruppo classe, illustrandone gli obiettivi e le modalità operative ed evitando di fornire informazioni che possano influenzare il comportamento linguistico spontaneo degli studenti nelle successive prove di narrazione. In questo primo incontro è prevista anche la compilazione individuale del questionario sociolinguistico da parte degli studenti, sotto la costante supervisione dei ricercatori, che forniscono assistenza se necessario. Le attività di raccolta dati si svolgono nei giorni successivi e fuori dall'aula in due spazi separati, dedicati rispettivamente alla narrazione formale e informale. Le prove sono consecutive ed eseguite sempre nell'ordine informale-formale. In entrambi gli spazi ai partecipanti è chiesto di vedere il video di un incidente stradale, e raccontarlo immaginando di aver appena assistito alla scena nella vita reale. Lo spazio informale viene allestito in modo da favorire un'interazione piacevole e rilassata tra studenti e somministratori. In un primo momento lo studente viene messo a proprio agio e con una breve conversazione su temi generali si cerca di creare familiarità con il setting e con i ricercatori. In questa situazione il video viene riprodotto due volte e allo studente viene poi chiesto di registrare un audio su una chat WhatsApp fittizia e creata ad hoc, immaginando di raccontare ad un amico quanto appena visto. Nella stanza formale, invece, la conversazione iniziale è più breve e limitata alle sole istruzioni. Il somministratore (una persona diversa da quella che gestisce la stanza informale) assume un atteggiamento volto a facilitare l'adozione di un registro linguistico più formale da parte del partecipante. Dopo la visione del video, lo studente ascolta un audio

⁵ Preme sottolineare che la metodologia di ricerca qui presentata è stata approvata dal Comitato Etico dell'Università di Bologna e che, per garantire il rispetto delle normative etiche e della privacy dei partecipanti, a ogni famiglia è stato fornito il modulo di consenso informato relativo alla partecipazione e al trattamento dei dati raccolti.

appositamente creato per la ricerca, che simula la segreteria telefonica della polizia stradale dedicata alla raccolta di testimonianze sugli incidenti. Al termine dell'audio, il partecipante è invitato a rilasciare la propria dichiarazione, descrivendo l'incidente rappresentato nel video dal punto di vista di un testimone. Gli studenti bilingui italiano-cinese e italiano-arabo vengono coinvolti in prove del tutto simili, ossia nelle medesime situazioni e con lo stesso video-stimolo, anche nella lingua di famiglia. Questa impostazione consente di analizzare e confrontare l'uso della lingua italiana ed ereditaria in contesti diversi, individuando le strategie linguistiche adottate da parlanti in età scolare in base alla situazione comunicativa e al destinatario. Inoltre, un simile protocollo permette di comparare i dati linguistici dei parlanti monolingui con quelli plurilingui, e per i due gruppi di bilingui selezionati consente confronti tra le competenze nella lingua maggioritaria e quella di famiglia su diversi assi di variazione.

3. Conclusioni: prospettive per l'analisi dei dati

Nel momento in cui si conclude la redazione di questo testo (primavera 2025), il protocollo è stato applicato con successo da entrambe le unità di ricerca nelle diverse scuole che hanno partecipato al progetto. Grazie alla collaborazione delle classi che hanno ospitato i ricercatori, è stato possibile superare ampiamente l'obiettivo minimo di raccolta identificato in fase di progettazione (Sezione 2.2). Il gruppo di ricerca può pertanto contare su un'ampia base empirica per l'analisi delle produzioni, che sono in corso di trascrizione e annotazione. La prima fase di analisi si concentrerà necessariamente su una valutazione dell'efficacia del protocollo nell'elicitar testi effettivamente di natura narrativa, e nello stimolare nei partecipanti una differenziazione naturalistica tra registro formale e informale (Canuto et al. 2025). Successivamente, le due unità di ricerca si specializzeranno nello studio della macrostruttura testuale (Torino) e delle strategie di gestione delle catene referenziali (Bologna), con particolare attenzione alle dimensioni di variazione sociolinguistica (natura del repertorio, livello di scolarizzazione, fenomeni di ristandardizzazione, gestione della variazione diafasica). Parallelamente, si procederà alla fase di restituzione e di formazione degli insegnanti, che prevede l'utilizzo dei materiali raccolti durante la sperimentazione per sviluppare in modo laboratoriale strategie didattiche innovative capaci di sfruttare il potenziale della narrazione come ponte tra l'uso spontaneo delle competenze linguistiche e la loro applicazione in contesto scolastico.

Bibliografia

- Andorno, Cecilia. 2013. Usare i pronomi in un'altra lingua: *lui* contrastivo? In Geymonat, Francesca (a cura di), *Linguistica applicata con stile*, 3–20. Alessandria: Edizioni dell'Orso.
- Berman, Ruth A. 2009. Language development in narrative contexts. In Bavin, Edith L. (a cura di), *The Cambridge Handbook of Child Language*, 355–376. Cambridge: Cambridge University Press.
- Berman, Ruth A. & Slobin, Dan I. (a cura di). 1994. *Relating events in narrative: A crosslinguistic developmental study*. Londra: Routledge.

- Calaresu, Emilia & Dal Negro, Silvia (a cura di). 2018. *Attorno al soggetto. Percorsi di riflessione tra prassi didattiche, libri di testo e teoria*. Milano: Officinaventuno.
- Canuto, Claudia & Perugini, Nicola & Andorno, Cecilia & Della Putta, Paolo, & Gianollo, Chiara & Martari, Yahis. 2025. Competenze narrative e variazione di registro nell'adolescenza: un corpus di testi orali in contesto plurilingue. Presentazione al convegno *La comunicazione parlata*, 19–21 giugno 2025, Università di Foggia.
- Chini, Marina. 2005. Reference to person in learner discourse. In Hendriks, Henriëtte (a cura di), *The Structure of Learner Varieties*, 65–110. Berlino / Boston: De Gruyter Mouton.
- Chini, Marina & Andorno, Cecilia (a cura di) 2018. *Repertori e usi linguistici nell'immigrazione. Una indagine su minori alloglotti dieci anni dopo*. Milano: FrancoAngeli.
- Fiorentini, Ilaria & Gianollo, Chiara. 2018. Evviva le lingue del mondo! Questionario sociolinguistico. Progetto *La classe plurilingue. Ricerca sulla complessità linguistica per una didattica inclusiva*. Bologna: Università di Bologna e Fondazione Alsos.
- Gagarina, Natalia & Klop, Daleen & Tsimpli, Ianthi M. & Walters, Joel. 2016. Narrative abilities in bilingual children. *Applied Psycholinguistics* 37. 11–17.
- Gianollo, Chiara & Perugini, Nicola. 2025. Referring expressions in narrative texts of Italian teenagers: a guided production study. Presentazione al convegno *Referential expressions in discourse* 2025, 29–20 maggio 2025, Università di Graz.
- Hickmann, Maya. 2004. *Children's Discourse: Person, Space and Time across Languages*. Cambridge: Cambridge University Press.
- Istat. *Condizioni di vita e reddito delle famiglie. Anni 2021–2022*. Roma: Istat, 2023. <https://www.istat.it/files//2023/12/Condizioni-di-vita-e-reddito-delle-famiglie-Anni-2021-2022.pdf>.
- Paradis, Johanne. 2011. Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic approaches to bilingualism*.1(3). 213–237.
- Røyneland, Unn & Vangsnes, Øystein A. 2020. Joina du kino imårgå? Ungdomars dialektskriving på sosiale medium. *Oslo Studies in Language*. 11(2020). 357–392.
- Serratrice, Ludovica & Sorace, Antonella & Paoli, Sandra. 2004. Crosslinguistic influence at the syntax- pragmatics interface: Subjects and objects in English-Italian bilingual and monolingual acquisition. *Bilingualism: Language and Cognition* 7(3). 183–205.
- Snow, Catherine E. 1983. Literacy and language: relationships during the preschool years. *Harvard Education Review* 53(2). 165–189.
- Strömqvist, Sven & Verhoeven, Ludo (a cura di). 2004. *Relating Events in Narrative, Volume 2: Typological and Contextual Perspectives*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wiese, Heike. 2020. Language Situations: A method for capturing variation within speakers' repertoires. In Asahi, Yoshiyuki (a cura di), *Methods in Dialectology XVI*, 105–117. Berna: Peter Lang.
- Wiese, Heike & Pohle, Maria. 2016. "Ich geh Kino" oder „... ins Kino?" Gebrauchsrestriktionen nichtkanonischer Lokalangaben. *Zeitschrift für Sprachwissenschaft* 35(2). 171–216.

- Wiese, Heike & Simon, Horst J. & Zappen-Thomson, Marianne & Schumann, Kathleen. 2014. Mehrsprachiges Deutsch: Beobachtungen zu Namdeutsch und Kiezdeutsch. *Zeitschrift für Dialektologie und Linguistik*. 81(3). 247–307.
- Wiese, Heike & Bracke, Yannic & Bunk, Oliver. 2018. Register differentiation in Namibian German: informal and formal language use in a vital speech community. Comunicazione al convegno *German Abroad 18*, University of Erfurt, 8–10 marzo 2018.
- Wiese, Heike & Alexiadou, Artemis & Allen, Shanley & Bunk, Oliver & Gagarina, Natalia, & Iefremenko, Kateryna & Klotz, Martin & Krause, Thomas & Labrenz, Annika & Lüdeling, Anke & Martynova, Maria & Neuhaus, Katrin & Pashkova, Tatiana & Rizou, Vicky & Tracy, Rosemarie & Schroeder, Christoph & Szucsich, Luka & Tsehay, Wintai & Zerbian, Sabine & Zuban, Yulia. 2019. RUEG Corpus (Version 0.2.0) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.3236069>

Sitografia

<https://www.linguistik.hu-berlin.de/en/institut-en/professuren-en/rueg>
<https://site.unibo.it/racconta/it>.
<https://tileggo.treccani.it/indagini-lettura/>

CLUB Working Papers in Linguistics

A cura di Silvia Ballarè e Caterina Mauri

Volume 9, 2025

ISBN: 9788854972148

Contributi di

Valeria Baruzzo

Silvia Ballarè

Sabina Capellari

Isabella Colliva

Emanuela De Vita

Mariapia D'Angelo

Gloria Gagliardi

Chiara Gianollo

Caterina Mauri

Chiara Meluzzi

Emanuele Miola

Egle Mocciaro

Yuka Naito

Nicola Perugini

Maria Ranù

Elena Martinelli

Eleonora Zucchini

CLUB – CIRCOLO LINGUISTICO DELL'UNIVERSITÀ DI BOLOGNA

ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA