

LADDER: UN CORPUS DI SCRITTURE DIGITALI PER L'INSEGNAMENTO DELLA PRAGMATICA IN L2. UN ESEMPIO DI ANALISI DI DISDETTE IN WHATSAPP

Nicola Brocca¹

1. INTRODUZIONE

A insegnanti di una L2 è sicuramente successo di entrare in contatto con richieste, scuse e altri atti linguistici di apprendenti, che, anche se corretti da un punto di vista grammaticale, risultano troppo brevi, o troppo lunghi, troppo diretti o troppo impliciti per essere accettati da parlante un madrelingua. Nella comunicazione in un ambiente didattico tali dissonanze possono diventare strumento di riflessione e ricevere un feedback, ma in contesto naturale possono portare a fallimenti pragmatici (Thomas, 1983). È risaputo che di fronte a un errore pragmatico il nativo trasferisce una valutazione negativa sulla personalità dell'emittente o del suo gruppo talora riproducendo stereotipi etnici, mentre a un errore morfosintattico il nativo può reagire mettendo in atto strategie di adattamento al livello linguistico dell'apprendente (Nuzzo, Santoro, 2017). Nel caso della comunicazione privata in forma digitale (come l'e-mail e i servizi di *instant messaging* – d'ora in poi IM) il rischio di un fraintendimento da parte di apprendenti di una L2 è maggiore che nella comunicazione faccia a faccia per l'alta difficoltà di interazione (Pallotti, 2019): la comunicazione mediata dal computer (CMC) – senza compresenza e sincronia – non può fare affidamento su elementi prossemici e ostacola l'interazione in più turni riducendo lo spazio per la negoziazione del significato.

In questo contesto si inserisce il progetto LADDER. Lo scopo del progetto è la creazione di un *corpus* digitale di testi di e-mail e messaggi WhatsApp in L2 di germanofoni e la relativa base di riferimento nella L1 al fine di analizzare le abilità socio-pragmatiche nel processo di acquisizione della lingua italiana. Il *corpus* è stato raccolto nell'ambito di un seminario di didattica svolto durante tre semestri dall'aprile 2020 e luglio 2021 presso l'Istituto di didattica disciplinare nel dipartimento di lingue didattiche delle lingue straniere dell'Università di Innsbruck il cui docente è l'autore di questo contributo. Il *corpus* si rivolge primariamente agli studenti del corso, insegnanti di italiano lingua straniera (LS) in formazione, per ricerche effettuate durante il seminario e la scrittura di tesine (§ 2.1). In secondo luogo il *corpus* ha come destinatari altri apprendenti e docenti (§ 2.2, § 2.3) e si offre inoltre come strumento di ricerca pragmatica per l'intera comunità scientifica (§ 2.4).

Questo contributo mostra l'importanza della creazione del *corpus* LADDER per la didattica di aspetti sociopragmatici e per la ricerca in pragmatica cross-culturale (§ 2). Nel § 3 si presentano le tappe della costruzione del *corpus*. In seguito (§ 4) si mostra a titolo esemplificativo un'interrogazione del *corpus* avvenuta durante il seminario di didattica. L'interrogazione si basa su uno studio di replicazione basato su Cortés Velásquez e Nuzzo (2017) e presenta un'analisi contrastiva tra l'atto della disdetta in italiano L1 e in italiano L2 di germanofoni. Successivamente (§ 5) si discute dell'archiviazione dei risultati e delle

¹ Università di Innsbruck.

prospettive per il completamento del *corpus*. Infine si riflette come il *corpus* può implementare le ricerche nell'ambito della didattica di una L2 (§ 6).

2. ATTUALITÀ E RILEVANZA DEL CORPUS LADDER PER LA DIDATTICA DELLA PRAGMATICA

Le *digital humanities*, in particolare nella loro prima espressione, hanno mostrato i vantaggi e le sfide del lavoro scientifico sulla base di *corpora*. Con l'avanzamento delle possibilità di archiviazione e analisi di dati digitali e la progressiva automatizzazione nella collezione, nell'analisi e nell'archiviazione, è stata possibile l'analisi su *corpora* di sempre più largo formato. A trarne vantaggio è stata soprattutto la ricerca lessicografica e lessicologica, ma anche quella morfosintattica. Al contrario la pragmatica non ha finora approfittato dell'automatizzazione come le altre sottodiscipline perché l'analisi di dati pragmatici si basa per l'estrazione del loro significato su un dettagliato riferimento ad informazioni con- ed extra-testuali che richiedono valutazioni elaborate e per il momento poco automatizzabili. In questo paragrafo vengono annoverate nelle ragioni che accompagnano la creazione del *corpus* LADDER.

2.1. Il *corpus* LADDER per la formazione degli insegnanti

Riferendosi al contesto italiano Corino e Marengo (2009) sottolineano che la didattica basata sui *corpora* resta ancora patrimonio in mano alla ricerca sia linguistica che glottodidattica, ma ben poco diffusa nella pratica in classe. A dieci anni di distanza Chamber (2019: 5) conferma la discrepanza tra ricerca e pratica didattica e, riferendosi al contesto internazionale, fa notare che nonostante i progressi nella linguistica computazionale, la maggior parte della ricerca empirica sull'uso di *corpora* in didattica si basi sul paradigma della ricerca-azione e siano sempre proposti da linguisti o persone esperte di didattica. In sostanza i progressi nel *data driven learning* non hanno ancora trovato una capillare diffusione nella pratica didattica scolastica: una ragione è la mancanza di una solida base di linguistica dei *corpora* nella formazione di insegnanti di lingua straniera.

Per quanto riguarda gli aspetti sociopragmatici in una L2, nei materiali didattici l'*input* risulta talvolta inadeguato (Nuzzo, 2013; Brocca, in stampa) anche se si riscontra un'alta sensibilità per il tema tra insegnanti di L2 (Trubnikova Garofolin, 2020: 86). È da immaginare che la didattica di aspetti sociopragmatici, senza un valido riferimento, possa rappresentare una sfida per l'insegnamento soprattutto nel caso il cui insegnante non sia un parlante nativo (Cohen, 2018: 57; Glaser, 2020).

Il *corpus* LADDER fa parte di un progetto nato al fine di colmare queste lacune. Con il *corpus* LADDER gli insegnanti (non nativi) in formazione possono arrivare alla definizione di fenomeni sociopragmatici non attraverso materiali preparati *ad usum Delphini*, ma attraverso fonti primarie raccolte e interrogate in un percorso didattico basato sul paradigma del *forschendes Lernen* (Natale, Kunz, Krakenberger, 2020; Reinmann, Lubcke, Heudorfer, 2019). Secondo questo paradigma di chiara ispirazione costruttivista, il sapere non si può trasferire dal docente agli studenti, ma lo studente è autore del proprio sapere attraverso lavori autonomi, lavori di gruppo e compiti basati sulla elaborazione di materiale autentico (Barr, Tagg, 1995). Il percorso seguito nel seminario prevede la collezione di scritture digitali, la loro preparazione e infine l'analisi in base a delle domande formulate dagli studenti stessi. In questo modo gli insegnanti in formazione promuovono la loro consapevolezza linguistica (*language awareness*) (Farr, 2008) e acquisiscono

esperienza nell'uso di un *corpus* in situazioni didattiche. L'uso di un *corpus* nella classe di lingua straniera è particolarmente utile per il trattamento di domande sociopragmatiche, che si lasciano descrivere in tendenze e preferenze in un certo contesto extralinguistico piuttosto che con regole nette e indipendenti dal contesto. Pertanto il seminario in cui viene creato il *corpus* LADDER si propone come modello didattico per colmare le lacune nella formazione degli insegnanti a cui fanno riferimento Chamber (2009) e Corino, Marellò (2009) da una parte e Trubnicova, Garofolin (2020), Taguchi, Roever (2017), Stadler, Will, Eloff (in stampa) e Glaser (2020) dall'altra. Il *corpus* LADDER, se usato con un metodo adeguato, può essere utilizzato per promuovere la consapevolezza pragmatica, e attivare competenze di ricerca (Brocca, 2020).

2.2. Il corpus LADDER come contributo per materiali didattici

La CMC è stata ripetutamente oggetto di studi linguistici che ne hanno sottolineato le caratteristiche idiosincratiche (Abel, Glaznieks, 2020; Antonelli, 2018; Pistolesi, 2004, 2016). Il *Quadro comune europeo di riferimento per le lingue. Volume complementare* (Council of Europe, 2020) individua nelle comunicazioni digitali caratteristiche che richiedono competenze comunicative *sui generis* e inserisce diversi descrittori per la corrispondenza via e-mail. Anche i governi nazionali hanno recepito le nuove esigenze inserendo la CMC nei rispettivi curricula nazionali² (Brocca, Masia, Kucelman, 2020).

Nonostante la variazione diamesica richieda un'attenzione particolare, l'apprendente di una L2 difficilmente riceve un *input* adeguato per la comunicazione digitale. La scarsità dell'*input* avviene sia nell'apprendimento guidato che in quello non guidato. Per quel che riguarda la didattica dell'italiano L2 alcuni manuali (Ghezzi, Piantoni, Bozzone Costa, 2015: 85) (Ghezzi *et al.*, 2015; Orlandino, Bali, Ziglio, 2018: 26, 76, 103) trattano la scrittura per IM esclusivamente come una forma di linguaggio giovanile, mentre altri (Bali, Rizzo, 2006: 104; Ghezzi *et al.*, 2015: 109) affidano alla e-mail atti linguistici che attualmente vengono veicolati da altri media (Brocca, in stampa). Una didattica della CMC si trova di fronte a due ostacoli: a) il medium scritto come il manuale non riesce a stare al passo con la rapida obsolescenza dei media digitali; b) l'autore di libri di testo non può che fare affidamento alla propria esperienza per interpretare la comunicazione digitale privata e inevitabilmente può offrire solo un *input* parzialmente adeguato. A questi problemi si va ad aggiungere la questione più generale sull'*input* di aspetti pragmatici nei libri per italiano LS. Tale *input* è spesso meno complesso e variegato rispetto allo stesso atto in contesti autentici (Bardovi-Harlig, 1996; Nuzzo, 2013). Per quel che riguarda l'apprendimento non guidato l'apprendente di una L2 non si trova a contatto con CMC nella L2 fino a quando è lui stesso a essere coinvolto nella comunicazione. Un *silent period*, come avviene in contesti altri contesti comunicativi, è escluso.

Per questo motivo il *corpus* LADDER, che ha primariamente uno scopo didattico, diventa una risorsa per tutti gli autori di materiali didattici che vogliono usare materiali autentici in L1 o analizzare i bisogni di apprendenti germanofoni di italiano per quel che riguarda la scrittura CMC privata.

2.3. Il corpus LADDER come materiale didattico con la mediazione dell'insegnante

La glottodidattica ha espresso interesse per l'analisi di *corpora* in molti suoi approcci. Sono note le posizioni dell'approccio lessicale (Lewis, 1997; Serra Borneto, Bonvino,

² Si veda ad esempio Gazzetta ufficiale federale del codice austriaco n. 88/1985, 2018.

2003: 233) secondo il quale l'apprendente deve essere incoraggiato all'apprendimento induttivo e autonomo mentre l'*input* deve essere ampio e autentico. Queste premesse sono il via libera per l'uso di *corpora* della L2 dove l'apprendente può riflettere sull'uso di elementi lessico-grammaticali in contesto. Una simile tendenza si riscontra in molti recenti approcci di stampo costruttivista tanto che Chamber (2019) riferendosi a una letteratura ultratrentennale (Granger, 2009; Leech, 1998) parla di *corpus revolution* in didattica. Il rinnovato interesse per approcci costruttivisti, che mettono l'apprendente al centro del proprio processo di apprendimento, propongono l'uso diretto di *corpora* come strumento di consultazione, riflessione e apprendimento da parte dello studente nella pratica in classe. Inoltre il riconoscimento dell'importanza di un *input* autentico (Stadler, Will, Eloff, in stampa) ha incoraggiato il confronto tra i materiali didattici e i materiali della vita quotidiana secondo analisi computazionali. Ne traggono vantaggio materiali per apprendenti come vocabolari (Krummes, Ensslin, 2015), grammatiche (Matsumoto, 2020), libri di testo (Schmiderer, Zanasi, Konecny, Autelli, 2020), strumenti di valutazione del livello di apprendimento (Hawkins, Buttery, 2009; McCarthy, 2016). Ma numerosi e in parte ancora poco esplorati sono i benefici dell'accesso diretto ai *corpora* da parte di insegnanti e apprendenti (Boulton, 2017; Chambers, 2019; Frankenberg-Garcia, 2012) come l'aumento dell'autonomia degli studenti in corsi di microlingua (Flowerdew, 2017), o il supporto alla didattizzazione durante la lettura di storie (de la Cruz, 2019).

Il *corpus* LADDER offre anche in questo caso un materiale innovativo che è possibile consultare, con la mediazione dell'insegnante, in classi di italiano per la produzione di testi CMC. L'insegnante trova in LADDER materiale autentico da selezionare e presentare agli apprendenti per trarne, deduttivamente, delle osservazioni linguistiche. Trova anche dei testi che possono essere usati per la riflessione contrastiva (per esempio un atto di richiesta molto stringato fronte a un atto molto elaborato) e stimolare la consapevolezza linguistica per esempio attraverso la formulazione di ipotetiche risposte o la classificazione di diversi atti linguistici dal più al meno gentile (Bardovi-Harlig, 1996: 7). Trova del materiale per associare le dichiarazioni dei suoi studenti ai diversi livelli pragmlinguistici del corpus degli apprendenti e favorire l'autovalutazione. Trova anche un esempio di compito di ricerca e analisi da condurre con una classe creando ad esempio il proprio corpus di produzioni di apprendenti che potrebbe poi essere analizzato dagli apprendenti stessi con la guida dall'insegnante. Inoltre l'insegnante con qualche competenza di linguistica dei corpora (o semplicemente di gestione di un database) è in grado di condurre e far condurre ai suoi studenti delle ricerche su tutto il *corpus* LADDER e offrire la possibilità di condurre osservazioni quantitative.

2.4. Il corpus LADDER come contributo per la ricerca sociopragmatica

Negli ultimi trent'anni la ricerca pragmatica si è spesso dedicata ad aspetti cross-culturali o interlinguistici. Questo vale anche per la ricerca sulle scritture digitali private dove la letteratura offre un repertorio variegato soprattutto per quanto riguarda la scrittura per e-mail. La e-mail tra studenti universitari e i loro docenti è stata il campo di un'intensa ricerca: tra gli altri Krulatz (2014) tra il russo e l'inglese, Economidou-Kogetsidis (2011) Greco-cipriota e l'inglese, Savić (2018) norvegese e inglese, Alcón Soler (2013) per l'inglese L1 e LS. Molto minori sono le ricerche che affrontano combinazioni linguistiche che non includono l'inglese. Poco però si è fatto al di fuori dell'inglese LS. Uno dei pochi lavori che mettono a fuoco l'italiano è Pagliara (2020). L'analisi di elementi sociopragmatici condotta da Pagliara conferma che, anche per l'italiano, le produzioni dei non nativi distano da quelle dei nativi: le differenze si concentrano nel maggior numero di forme di comunicazione dirette, nella minore quantità e nella diversa qualità di

mitigatori lessicali e sintattici rispetto quelle dei nativi. Tali risultati hanno immediate ricadute didattiche perché forniscono un'analisi dei bisogni basata su dati empirici che aiutano a formulare precisi obiettivi didattici.

La ricerca sociopragmatica ha rivolto la sua attenzione anche alla scrittura negli IM negli ultimi anni (Flores-Salgado, Castineira-Benitez, 2018; König, 2019; Petitjean, Morel, 2017; Sandel, Ou, Wangchuk, Ju, Duque, 2019) ma un'ottica cross-culturale o interlinguistica non è stata ancora seriamente presa in considerazione³. Tutte le ricerche pragmatiche sulla CMC fin qui citate raggiungono i loro limiti per la difficoltà del reperimento del *corpus*. Sia la raccolta di e-mail che quella di IM si basano su materiale autentico che, oltre a porre problemi di *privacy* nella eventuale condivisione, è strutturalmente limitato e non sempre adatto ad analisi comparate per i differenti contesti in cui si sviluppa.

Il *corpus* LADDER si propone di colmare questa lacuna nella ricerca e si apre alla condivisione in formato *open access*. Inoltre il *corpus* offre la possibilità alla comunità scientifica di lavorare con una costellazione (italiano e tedesco) finora non molto osservata, ma che ha mostrato promettenti risvolti nella comparazione interlinguistica (Castagneto, Ravetto, 2015; Payne, 2013; Ravetto, 2012; Venuti, 2013).

3. LA COSTRUZIONE DEL CORPUS

Il *corpus* LADDER raccoglie e-mail e IM, due generi testuali molto diversi e che in genere veicolano la CMC privata rispettivamente in contesti formali e informali. Il *corpus* è stato raccolto attraverso un questionario online basato su *Discourse Completion Task* (DCT) e *Role Play* (Taguchi, Roever, 2017: 85, 231) premesso da alcune domande finalizzate alla raccolta di metadati con informazioni sociolinguistiche degli informanti. La raccolta è stata condotta dagli studenti del corso che avevano il compito di far completare il questionario ad alcuni informanti con le seguenti caratteristiche: essere studenti, avere un'età compresa tra i 18 e i 35 anni, non essere bilingui, essere o madrelingua italiana o di madrelingua tedesca e con competenze della lingua italiana superiori al livello A1 del QCER. Nel rispetto delle norme sulla *privacy* (*General Data Protection Regulation*) è stato raccolto il consenso all'uso e alla pubblicazione dei dati. Tutti i dati sono stati anonimizzati o pseudonimizzati. In questo paragrafo si osserva la costruzione del *corpus* nei suoi due differenti generi testuali.

La prima parte del *corpus* è composta da 232 e-mail, divise in due scenari e prodotti da 78 madrelingua e da 38 apprendenti per un totale di 18.932 *token*. Il questionario si basa un *discourse completion test* (DCT) condotto online con *Googleform* in cui veniva richiesto all'informatore di scrivere alcune e-mail secondo due scenari riportati qui di seguito:

scenario a)

Immagina di star facendo un corso con il Dr. Nicola Brocca. Domani devi fare una presentazione in classe. Non hai avuto tempo per studiare bene perché dovevi prepararti ad un esame di inglese e ti accorgi che il materiale da presentare è più di quello che avevi previsto. Scrivi una e-mail al professore: il tuo obiettivo è spostare la presentazione.

³ Si escludono da questo novero le ricerche sugli SMS a cui fa riferimento per esempio parte del *corpus* DISDIR (Cortés Velásquez, Nuzzo, 2017), perché l'IM e l'SMS hanno caratteristiche diverse. Il *corpus* multilingue *What's up, Switzerland* (Ueberwasser, Stark, 2017) non si presta all'analisi di aspetti sociopragmatici per la difficoltà di valutare aspetti extralinguistici della situazione comunicativa.

scenario b)

Hai fatto un corso con il Dr. Brocca. Hai consegnato il tuo portfolio un mese fa e non hai ancora ricevuto il voto. Ti serve il voto per registrarti per una borsa di studio (Stipendium). Mandi una e-mail al prof.: il tuo obiettivo è ricevere il voto al più presto.

Tali scenari rappresentano un compito realistico per gli informanti dato che questi sono stati scelti tra studenti universitari. I due scenari prevedono situazioni con un diverso grado di legittimità⁴: mentre nella situazione a) la richiesta è poco legittima perché lo studente viene meno ad un accordo, lo è nel caso b) in cui lo studente ha diritto all'ottenimento di un voto e l'inadempienza da parte del docente comporta dei danni allo studente. Ci si aspetta quindi una differente strategia nella formulazione della richiesta: lo scenario a) sarà probabilmente mitigato da atti giustificatori più complessi che lo scenario b). Facendo riferimento alla teoria della faccia (Goffmann, 1956) mentre nella situazione b) il richiedente (lo studente) deve evitare di minacciare la faccia della persona a cui fa la richiesta (il docente), nella situazione a) il richiedente dovrà cercare di non perdere la propria faccia e quindi mettere in atto sotto-atti giustificatori o strategie di riparazione. Per la realizzazione di questo atto è anche possibile ricorrere a una bugia dandosi per esempio per malato o non compiere del tutto questo atto (Brown, Levinson, 1987: 69).

La seconda parte del *corpus* si basa su messaggi scambiati attraverso l'applicazione per smartphone WhatsApp che funge da archetipo per il genere IM. Sono stati raccolti 1.204 messaggi di WhatsApp divisi in 6 diversi scenari e prodotti da 80 informanti nativi, 55 apprendenti germanofoni e 8 bilingui italiano tedesco per un totale di 33.966 *token*. Il design della ricerca prevede che gli studenti intervistatori mandino agli informanti prima un questionario online con *Googleform* per la raccolta di metadati e poi inviino loro per WhatsApp il DCT a cui l'informante poteva rispondere attraverso smartphone⁵. Gli stimoli vogliono elicitar tre diversi atti linguistici (rifiuto, richiesta e disdetta con poco preavviso) in due diverse situazioni comunicative che differiscono per distanza sociale: la comunicazione tra amici, quindi con distanza sociale bassa, o con il diretto superiore, quindi distanza sociale media. Il grado di imposizione (l'oggetto della richiesta o di rifiuto) restava costante, con minimali adattamenti alla diversa situazione comunicativa. Anche se le interazioni per IM avvengono tipicamente tra persone con bassa distanza sociale e interazioni in contesti formali sono meno realistiche, si è pensato di inserire la variabile della distanza sociale media per favorire la riflessione in fase di analisi.

Gli *input* sottoposti agli informanti sono stati i seguenti:

1. Richiesta con distanza sociale bassa:

Sei in Erasmus in Italia. Avete creato una chat con 10 compagni di corso. Hai perso la tua tessera della biblioteca e vuoi chiedere se qualcuno ti può aiutare perché ti serve un libro entro domani...per esempio prestandoti la sua. Cosa scrivi?

2. Rifiuto con distanza sociale bassa:

Ricevi questo messaggio da un amico/a che fa un seminario con te: "Ciao, sono a corto di tempo. Ho visto che hai preso 30 all'esame. Potresti darmi una mano e restare con me in biblioteca oggi?" Non vuoi aiutare l'amico. Come reagisci?

⁴ Il concetto di legittimità non rientra esplicitamente nel modello di Brown e Levinson (1987) dove viene incluso nella dominanza sociale. La conoscenza degli obblighi e dei diritti dei parlanti, che fornisce le chiavi per interpretare correttamente la situazione comunicativa, merita in questo caso una etichetta a sé stante.

⁵ Non si è controllato se l'informante abbia reagito agli stimoli attraverso WhatsApp per Desktop.

3. Disdetta con distanza sociale bassa:
Cinque giorni fa hai promesso ad un/a amico/a che questa sera sareste andati al cinema assieme. Però devi disdire. Cosa fai?
4. Richiesta con distanza sociale media:
Sei al lavoro e hai smarrito il documento elettronico per entrare nel parcheggio. Sei nuovo in questo gruppo di lavoro e hai solo il numero del tuo diretto superiore. Gli mandi un messaggio per chiedergli se ti può aiutare.
5. Rifiuto con distanza sociale media:
Ricevi questo messaggio dal tuo/ dalla tua superiore. “Gentile collega, domani c’è una scadenza importante. Per caso sarebbe in grado di restare oggi in ufficio oltre l’orario?” Non vuoi restare in ufficio oltre il normale. Come reagisci?
6. Disdetta con distanza sociale media:
5 giorni fa hai promesso al tuo / alla tua superiore che oggi saresti andato a una cena di lavoro. Però devi disdire. Cosa fai?

I messaggi in WhatsApp vengono esportati in formato .txt e inseriti in una tabella Excel. Il *corpus* raccolto viene preparato per l’analisi quantitativa seguendo le seguenti tappe: i messaggi corrotti o non verosimili sono stati eliminati, errori nella formattazione delle *emoticons* vengono corretti manualmente, ad ogni messaggio vengono associati manualmente i metadati dell’informante (livello di competenza linguistica secondo il QCER, L1, L2, sesso, tipo di studi, area geografica di provenienza, area di residenza, codice dell’intervistatore, etc.).

4. ESEMPIO DI INTERROGAZIONE DEL *CORPUS*

Nel seguente paragrafo si documenta un’interrogazione del *corpus* compiuta durante il seminario⁶. Lo scopo di tale documentazione è offrire un esempio dell’uso del *corpus* da parte di insegnanti di italiano L2 in formazione. Trattandosi di una ricerca compiuta da studenti del triennio lo scopo non è di arrivare a risultati originali, ma di sensibilizzare alle competenze sociopragmatiche e di acquisire delle competenze di ricerca. Il compito presentato in questo paragrafo consisteva nello studio replicativo (Block, Kuckertz, 2018) della ricerca di Cortés Velásquez, Nuzzo (2017).

La domanda di ricerca ha voluto verificare la presenza di differenze nelle funzioni pragmatiche nelle mosse di contorno di una disdetta in parlanti nativi e apprendenti. La ricerca ha percorso sette fasi:

- 1) la preparazione di un subset estratto dal *corpus* LADDER;
- 2) la scelta e preparazione dei criteri di analisi;
- 3) il *training* degli annotatori e l’annotazione del *dataset* secondo i criteri scelti;
- 4) il controllo dell’annotazione attraverso il calcolo dell’*inter-rater-agreement*;
- 5) l’analisi della distribuzione della frequenza dei criteri;
- 6) il *clustering* e la visualizzazione dei dati;
- 7) l’interpretazione dei risultati.

⁶ Si ringraziano Mirjana Ljubojevic e Raphael Eichler per aver permesso la pubblicazione della loro analisi: <https://ladder.hypotheses.org/130>.

4.1. *La preparazione del subset*

Dal *corpus* considerato si sono estratti i dati formati da disdette di apprendenti e di nativi. Sono state considerate solo le disdette con distanza sociale bassa. Si è scelto di non considerare le *emoticons*. Il *subcorpus* si compone di 77 disdette in WhatsApp divise in 906 sotto-atti.

Per la scelta dei criteri di analisi ci si è basati sulla tassonomia proposta da Cortés Velásquez, Nuzzo (2017). Tale tassonomia si discosta dalla classificazione più comunemente diffusa in ricerche di pragmatica transculturale (Blum-Kulka, House, Kasper, 2011), perché anziché individuare un nucleo attorno al quale ruotano atti di sostegno corollari, classifica il turno di conversazione in diversi sotto-atti senza attribuzione gerarchica. Questo semplifica la classificazione perché gli annotatori non devono farsi carico di ricercare l'atto testa, non sempre distinguibile dalle strategie di contorno specialmente nel caso in cui la disdetta avvenga in modo indiretto. Inoltre la tassonomia di Cortés Velásquez, Nuzzo (2017) garantisce un'analisi a grana fine (13 diversi tipi di sotto-atti: *appeal, empathy, cancellation, explanation, farewell, gratitude, irony, repair, preparation, remedial, defence, willingness, wishes*) e contemporaneamente in categorie facilmente identificabili. Tra tali categorie si è scelto di eliminare *gratitude* e *irony* perché non presenti nel *subcorpus* considerato. Le 11 categorie pragmatiche sottoposte agli annotatori sono dunque le seguenti. Accanto a ogni etichetta si è aggiunto un esempio.

1. Appeal: *Ciao, senti...*
2. Appeal to empathy: *Lo so che è tardi...*
3. Cancellation: *Sta sera non ce la faccio a venire...*
4. Explanation: *Ho conosciuto una su tinder*
5. Farewell: *A presto*
6. Offer of repair: *Possiamo fare la settimana prossima?*
7. Preparator: *ho brutte notizie..., indovina cos'è successo*
8. Remedial move: *Mi dispiace*
9. Self-defence: *Sono preso con le bombe*
10. Willingness: *Ci tenevo proprio*
11. Wishes: *Buon divertimento anche senza di me*

Dopo aver stabilito le categorie e facendo riferimento alla loro descrizione approfondita in Cortés Velásquez e Nuzzo (2017) gli annotatori, due studenti e il docente (che corrisponde all'autore dell'articolo), hanno annotato in modo congiunto 10 messaggi che non rientrano in quelli analizzati nel subset. In questa breve sessione di *training* si è voluto testare l'interpretazione dei criteri su materiale autentico e creare un *gold-standard* per un'analisi possibilmente più unitaria e meno soggetta a variazione individuale (Rapp, 2017). Successivamente gli studenti in coppia (annotatore 1) e il docente (annotatore 2) hanno intrapreso l'annotazione del *subset* indipendentemente. Si mostra qui di seguito un esempio di messaggio prodotto dall'informante 21CN e rispettivamente l'annotazione interlineare dell'annotatore 2.

21CN:

Ciao! Come va? Per sta sera, non penso che potrò venire al cinema: ho raffreddore e tosse penso sia più prudente rimanere a casa. Scusami, buona serata

Annotatore 2:

Ciao! [*appeal*] Come va? Per sta sera, non penso che potrò venire al cinema [*cancellation*]: ho raffreddore e tosse penso sia più prudente rimanere a casa [*explanation*]. Scusami [*remedial move*], buona serata [*farewell*]

4.2. *Analisi dell'accordo tra annotatori*

Nonostante la fase di *training*, il manuale dell'annotatore, il lavoro in coppia da parte degli studenti, l'annotazione di categorie pragmatiche risente di una variazione intersoggettiva (Garassino, Masia, Brocca, in stampa). Per tenere sotto controllo questa variazione si è provveduto a misurare l'indice di accordo tra gli annotatori (*inter-rater agreement*) attraverso il valore kappa di Cohen (Hoek, Scholman, 2017). Nonostante la sua diffusione in molti campi di ricerca, come ad esempio quello medico, la scelta di tale valore per le annotazioni in pragmatica linguistica è uno strumento non pienamente consolidato. Ne consegue che non esiste un accordo tra i ricercatori su quale sia il valore per stabilite che l'*inter-rater-agreement* debba superare per un livello di concordanza accettabile: Artstein, Poesio (2008) propongono il valore .80 in linguistica computazionale ma in ricerca pragmatica basata su *corpora* tale valore può arrivare a .70. A tale valore giungono anche Spooren e Degand (2010: 256) riferendosi alla coerenza discorsiva. Su quest'ultimo valore si baserà la seguente analisi. Oltre alla variazione determinata dall'intervistatore una certa variazione è dipendente dallo strumento statistico per misurare l'*inter-rated agreement* come dimostrano Garassino, Masia e Brocca (in stampa) dove gli stessi dati valutati con diversi strumenti portano a diverse conclusioni. Si è preferito usare l'indice kappa di Cohen per la sua diffusione, tenendo però conto del *bias* della frequenza (Di Eugenio, Jordan, Thomason, Moore, 2000), secondo il quale l'uso poco frequente di una certa categoria produce un indice kappa basso (Spooren, Degand, 2010: 257). Per la categoria *willingness* dove la frequenza è molto bassa, tale *bias* impedisce di ottenere un indice affidabile.

Nella tabella 1, che mostra i valori kappa di Cohen tra i due annotatori, cinque categorie (*appeal*, *cancellation*, *explanation*, *farewell*) mostrano un valore significativo e tre (*repair*, *defence*, *wishes*) completo accordo. In tre categorie (*empathy*, *preparatory*, *remedial*) l'indice kappa di Cohen mostra una concordanza non significativa.

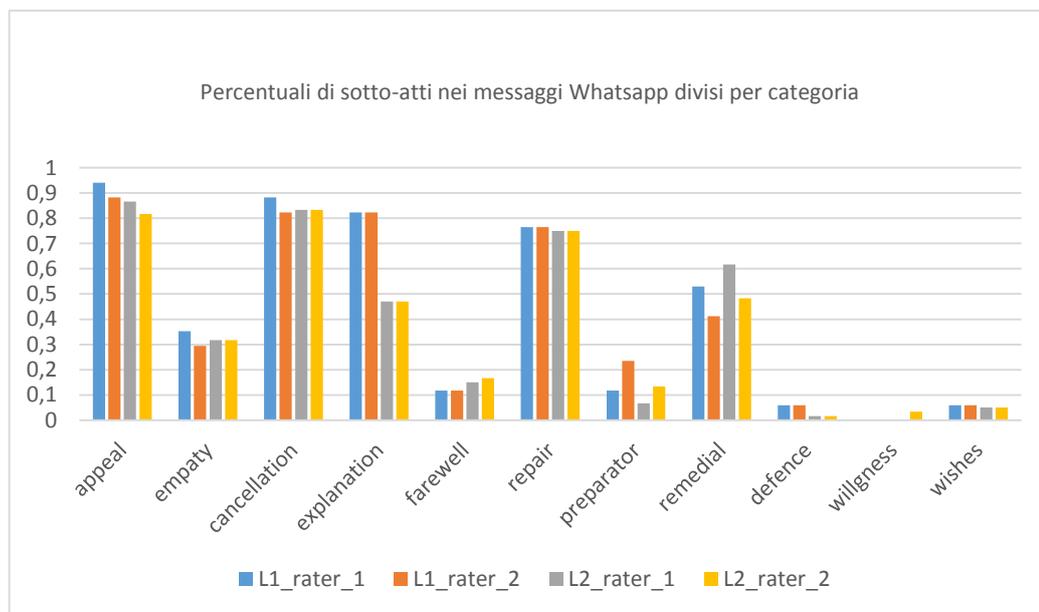
Tabella 1. *Indice kappa di Cohen per due annotatori (non ponderato), p-value: <.01*

Appeal	Empathy	Cancellation	Explanation	Farewell	Repair
.789	0.419	.761	.9	.949	1
Preparatory	Remedial	Defence	Willingness	Wishes	
.628	.692	1	.493	1	

4.3. Analisi della distribuzione della frequenza delle categorie di sotto-atti

Dopo aver tenuto conto di quale affidabilità hanno le analisi degli annotatori, si è provveduto a osservare la distribuzione della frequenza dei sotto-atti nel subset per comparare la produzione degli apprendenti con quella dei nativi, sempre tenendo in conto di entrambe le valutazioni. Nel grafico 1 le prime due barre di ogni categoria (in blu e rosso) rappresentano il gruppo di nativi valutati secondo annotazione 1 (blu) e due (rosso), mentre la terza e la quarta rappresentano i valori del gruppo di apprendenti rispettivamente nella annotazione 1 e 2. I sotto-atti di *appeal*, *cancellation*, *repair* sono tendenzialmente presenti in tutte le disdette (i valori variano da 95% a 77%) con poche differenze tra apprendenti e nativi. Mentre poco presenti in entrambi i gruppi sono le categorie *empathy*, *farewell*, *preparator*, *remedial*, *wishes* (35%-1%). Una differenza notevole nelle le produzioni di nativi e apprendenti si nota esclusivamente per il sotto-atto *explanation* presente in 82% degli atti dei nativi e 47% degli apprendenti, mentre la quantità dell'uso dei sotto-atti è nelle altre categorie tendenzialmente omogenea tra i due gruppi. Una leggera differenza si evidenzia nelle categorie *farewell* e *defence*.

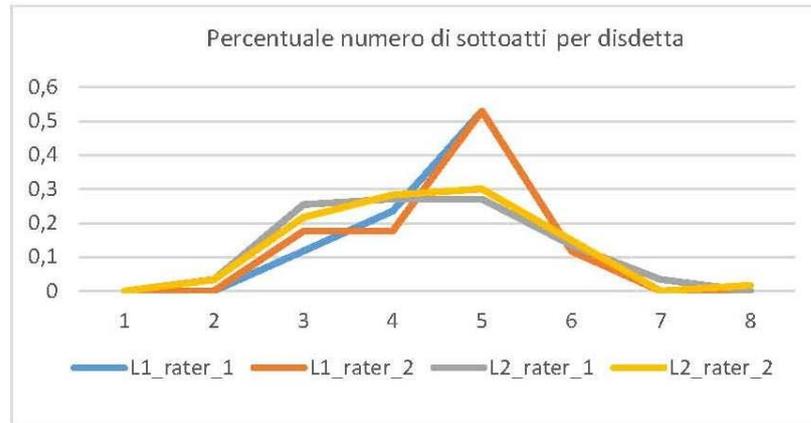
Grafico 1. Percentuali di sotto-atti divisi per categoria



Un'altra valutazione delle differenze tra madrelingua e apprendenti viene effettuata attraverso il conteggio del numero dei sotto-atti per messaggio di disdetta successivamente rappresentato in frequenza percentuale sugli atti del gruppo di parlanti (nativi o apprendenti). Anche in questo caso il grafico 2, che visualizza i risultati, ha tenuto conto delle differenze tra le due annotazioni. I valori ottenuti mostrano una scelta del numero di sotto-atti molto differenziata in particolare per gli apprendenti: le loro produzioni oscillano tra i due e gli otto sotto-atti, mentre per i nativi la scelta si concentra tra tre e sei. Per entrambi i gruppi i valori di frequenza si attestano su un picco allontanandosi dal quale le produzioni sono meno frequenti. Tuttavia mentre tale picco è accentuato per quanto riguarda le produzioni dei nativi (linee L1_rater_1 e L1_rater_2 nel Grafico 2), per gli apprendenti (linee L2_rater_1 e L2_rater_2) il picco è più disteso. In altre parole la disdette più frequente per i nativi si compone di 5 sotto-atti, mentre per gli apprendenti le disdette più frequenti sono quasi equamente distribuite tra i tre e i cinque sotto-atti.

Come si evince dal grafico 2, il 53% dei nativi costruisce la disdetta su 5 sotto-atti, fronte a 27%-30% (oscillazioni dipendono dall'annotazione 1 o 2) degli apprendenti. Gli apprendenti usano più spesso dei nativi disdette composte da tre e quattro sotto-atti: nel 25%-21% dei casi gli apprendenti costituiscono le loro disdette su tre sotto-atti mentre la stessa scelta viene compiuta dall'11%-17% dei nativi. Anche l'uso di 4 sotto-atti è scelto da 27%-28% degli apprendenti e 23%-17% dei nativi.

Grafico 2. *Percentuale numero di sotto-atti che compongono una disdetta*



4.4. *Formazione di cluster*

I risultati dell'analisi raggiungono un ulteriore valore descrittivo se i dati, una volta analizzati, permettono la formazione di cluster utili per la rappresentazione dell'atto linguistico della disdetta in nativi ed apprendenti. In questo modo oltre a ottenere una descrizione della distribuzione dei sotto-atti (grafico 1) e della quantità degli atti (grafico 2) per gruppo, è possibile analizzare la loro composizione nell'atto della disdetta.

Analizzando il nostro campione si ottengono 39 diversi tipi di set, ovvero categorie che raggruppano uguali composizioni di sotto-atti⁷. Vista l'ampia varietà delle categorie, e visto il *corpus* composto da soli 77 atti di disdetta, è evidente che molti degli atti sono composti da set che non si ripetono nel *corpus* e formano pertanto delle categorie con un solo atto di disdetta. L'ampiezza molto ridotta del *corpus* rende evidente i propri limiti e rende impossibile una visualizzazione in cluster. Tuttavia per dare un'idea dell'analisi condotta a scopi didattici si presentano alcuni risultati ottenuti ordinando i diversi tipi di set per frequenza. Il set più rappresentato nelle produzioni dei nativi (18% dei casi) è composto nel modo seguente:

appeal – appeal to empathy – cancellation – explanation – offer of repair

Un esempio di questo tipo di set è rappresentato nel seguente atto di disdetta:

22IO: Heyyy Senti.[*appeal*] Lo so che ti avevo promesso di uscire [*empathy*] ma oggi non me la sento proprio [*cancellation*] sono veramente stanca [*explanation*] possiamo fare nei prossimi giorni [*repair*]?

⁷ Per semplicità non si è tenuto in conto dell'ordine dei sotto-atti all'interno della disdetta anche se questo non è irrilevante per la forza illocutiva dell'atto.

Tale set è sottorappresentato delle produzioni degli apprendenti (3%). È interessante notare che al di fuori dal set qui descritto la frequenza d'uso della categoria *appeal* è uguale a quella dei nativi (cfr. grafico 2). Evidentemente tale categoria viene usata dagli apprendenti in un set diverso.

Il set usato dai nativi secondo per frequenza (12%) è composto da:

appeal – cancellation – explanation – offer of repair – remedial move.

Questo set è usato frequentemente anche dagli apprendenti (10%) come si osserva nel seguente esempio:

01RH: Ciao amico 😊 [*appeal*], oggi volevamo andare al cinema [*cancellation*]. Non ho tempo oggi [*explanation*]. Andiamo al cinema la prossima settimana? [*repair*] Per favore scusami [*remedial*].

Un altro set con frequenza ugualmente alta negli apprendenti (10%) è quello composto nel modo seguente:

appeal – cancellation – explanation – remedial move

Come nel seguente esempio:

13HM: Ciao [*appeal*]. Non posso andare al cinema sta sera [*cancellation*]. Mi dispiace [*remedial*], ma ho dimenticato che devo ancora fare i compiti per un corso [*explanation*].

Questo set non compare nel *corpus* dei madrelingua (0%), che invece presentano lo stesso set con l'aggiunta della categoria *offer of repair* già presentato nell'esempio 22IO.

Altri set come la combinazione *appeal – cancellation – offer of repair* o *appeal – cancellation – explanation – offer of repair – remedial move* sono invece rappresentati con valori simili sia nei nativi che negli apprendenti come si osserva nella tabella 2 che riassume i dati riportati:

Tabella 2. *Combinazioni di set e presenza percentuale in nativi e apprendenti*

Combinazioni nel set di sotto-atti	L1	L2
appeal - appeal to empathy - cancellation - explanation - offer of repair	18%	3%
appeal - cancellation - explanation - offer of repair - remedial move	12%	10%
appeal - cancellation - explanation - remedial move	0%	10%
appeal - cancellation - offer of repair	6%	5%
appeal - cancellation - offer of repair - remedial move	0%	5%

4.5. Interpretazione dei dati

Già con questa breve interrogazione condotta su un *subcorpus* di IM molto limitato si possono trarre alcune interpretazioni utili per la didattica. L'analisi ha rivelato che:

- l'atto della disdette nei nativi si compone di diversi sotto-atti, quattro dei quali sono presenti in quasi tutte le disdette: in apertura si ricorre al saluto (*appeal*),

successivamente si si fa ricorso a una disdetta esplicita dell'appuntamento preso (*cancellation*), si illustrano i motivi della disdetta (*explanation*) e si propongono degli atti per riparare alla disdetta (*repair*). Altri atti (*empathy, farewell, repair, preparatory, remedial, defence, willingness, wishes*) hanno una funzione ancillare;

- b) il gruppo degli apprendenti ricorre meno al sotto-atto *explanation* rispetto al gruppo dei nativi;
- c) le disdette degli apprendenti sono più differenziate nel numero dei sotto-atti rispetto a quelle dei nativi e tendenzialmente composte da meno sotto-atti;
- d) nonostante la forte eterogeneità, si fa strada un determinato set nei nativi (vedi 22IO) che negli apprendenti è poco rappresentato.

Certamente tali risultati sono da verificare con un *corpus* più esteso e bilanciato. La minore eterogeneità nel numero dei sotto-atti che compongono l'atto della disdetta dei nativi può dipendere dal ristretto numero di informanti madrelingua. Questo spiegherebbe la differenza con i dati raccolti in Santoro, Nuzzo (2017) che attestano a 3 il valore mediano degli atti di disdetta anche nei nativi. Inoltre il risultato dell'*inter-rater agreement* non è soddisfacente per tutte le categorie e questo getta un'ombra sulla validità dei dati. In particolare c'è da chiedersi se il ricorso ad annotatori non nativi, come gli studenti nel corso di didattica, possa influire sulla qualità dell'annotazione. Nonostante la parzialità di queste analisi preliminari l'interrogazione

- a) soddisfa lo scopo descritto in §2.1 cioè quello di innalzare la sensibilità sociopragmatica in insegnanti di italiano L2 in formazione (Brocca, in preparazione);
- b) si presta ad essere usata per la pianificazione (cf. §2.2): posto che l'obiettivo didattico sia di rendere i parlanti consapevoli delle loro scelte comunicative, una pianificazione didattica basata su questa analisi potrebbe mettere in luce (attraverso azioni di *noticing*) lo scarso uso del sotto-atto della *explanation*, la brevità dei set nelle produzioni degli apprendenti, e presentare i possibili benefici di una disdetta composta dal seguente set "*appeal - appeal to empathy - cancellation - explanation - offer of repair*" come quello prodotto in 22IO.

Un intervento didattico che si fondasse su queste considerazioni sarebbe basato su un *corpus* di produzioni autentiche e in quanto tale orientato ai bisogni degli apprendenti. Tale intervento dovrebbe però far percepire la non normatività delle scelte sociopragmatiche e dovrebbe avere come obiettivo il sostegno della consapevolezza linguistica dell'atto pragmatico e la riflessione sul rapporto interpersonale che si instaura in dipendenza delle diverse scelte.

5. ARCHIVIAZIONE E SVILUPPI DEL CORPUS

Il *corpus* LADDER è un progetto in corso. A progetto terminato (fine 2021) il *corpus* sarà formato da ca. 10.000 *token* per e-mail e ca. 30.000 *token* per l'IM perché tale è la quantità di dati minima per rendere possibile un confronto quantitativamente rilevante anche in *subset*. Le dimensioni del *corpus* sono necessariamente limitate in quantità visto che la raccolta dei dati deve procedere in modo manuale passando per la somministrazione individuale del DCT e del controllo dei metadati. Il collo di bottiglia più stretto risulta al momento l'annotazione di aspetti socio pragmatici che risulta un processo poco automatizzabile e che necessita di un controllo attraverso l'annotazione incrociata di più annotatori.

L'archiviazione del progetto avviene in un duplice formato: la documentazione *in itinere* e l'archiviazione nel lungo periodo. La documentazione *in itinere* ha luogo attraverso il blog scientifico *ladder.hypotheses.org* (<https://ladder.hypotheses.org/>) (Brocca, 2020) dove sono presentati alcuni lavori degli studenti e commenti di esperti. Le ricerche condotte dagli studenti presentano la documentazione della loro analisi in base al paradigma FAIR rendendo tutti i dati rintracciabili (*findable*), accessibili (*accessible*), integrabili con altri dati (*interoperable*) e riutilizzabili (*reusable*) (Wilkinson *et al.*, 2016). Tale documentazione ha scopi di disseminazione dei risultati e si rivolge in particolare a studenti e insegnanti. Inoltre la pubblicazione rientra tra gli obiettivi didattici scelti dal corso in linea con l'approccio del *forschendes Lernen*. L'archiviazione di lungo periodo ha luogo in *zenodo.org*. Al *dataset* è stato attribuito un codice doi (10.5281/zenodo.4895219) (Brocca, 2021). Il materiale è stato convertito dal file Excel (XLSX) ai formati XML/TEI e CSV. La conversione in XML/TEI è avvenuta attraverso un codice scritto in XSLT⁸. Il formato TEI (Burnard, 2014) è un formato *open source* per l'annotazione di testi con una solida tradizione (fondato nel 1987) organizzato da un consorzio che stabilisce linee guida per la codificazione di testi. Un testo in formato TEI è strutturato in parti di testo (*body*) separate dalle metainformazioni sul testo (*header*). In questo modo è possibile separare contenuto dal layout, ma anche introdurre etichette che descrivono il tipo di contenuto secondo i criteri stabiliti dagli annotatori. Il formato TEI è leggibile sia dalla macchina sia dal ricercatore senza particolare conoscenze di programmazione. La conversione da XLSX a CSV è avvenuta con una *library Apache*. Nella conversione sono stati mantenuti le *emojicons* e i *capoversi* come nei messaggi originali.

6. CONCLUSIONI

In questo contributo si è voluto presentare il progetto LADDER che prevede la realizzazione di un *corpus* di scritture digitali private per analisi pragmatiche in italiano di nativi e di apprendenti germanofoni. Il contributo descrive sia il progetto del *corpus* "a regime" sia le tappe della sua costruzione. Quest'ultime sono basate su un innovativo progetto didattico condotto dall'autore presso l'Istituto di didattica dell'Università di Innsbruck che mette gli studenti nella veste di ricercatori e permette loro di raccogliere dati e analizzarli attraverso repliche di studi scientifici. Il progetto ha quindi *in primis* uno scopo didattico per lo sviluppo della competenza pragmatica per insegnanti in formazione. Fenomeni sociopragmatici vengono presentati in modo induttivo e discussi in modo esplicito sulla base del materiale prodotto sia da madrelingua che da apprendenti. In questo modo gli studenti hanno la possibilità di scoprire da materiali estratti personalmente l'efficacia e l'appropriatezza di determinati atti linguistici e di analizzare i bisogni degli apprendenti. Oltre che lo sviluppo di competenze pragmatiche gli studenti esperiscono un modo innovativo per fare didattica (non solo della pragmatica) basandosi sulla co-costruzione e sull'analisi di *corpora*.

L'*output* del progetto non si limita alla formazione degli insegnanti durante il corso ma ha anche una ricaduta sulla ricerca. Infatti il materiale raccolto viene preparato per la consultazione in formato *open access* ed archiviato diventando un prodotto consultabile dalla comunità scientifica. LADDER fornisce il materiale necessario alla ricerca nell'intersezione tra didattica, pragmatica e linguistica dei *corpora* (Santoro, Vedder,

⁸ Si ringrazia Joseph Wang-Kathrein (Centro di ricerca Archivio del Brennero e Digital Science Center dell'Università di Innsbruck) per la collaborazione. Il codice sorgente usato per la conversione è scaricabile al seguente indirizzo: <https://git.uibk.ac.at/c62456/ladder-archivierung>.

2016). Tale *corpus* trova la sua utilità nello studio di fenomeni pragmatici come atti linguistici, segnali discorsivi, routine conversazionali, uso di *emoticons* etc. LADDER colma un vuoto nelle risorse in mano ai ricercatori perché sazia il bisogno di chi vuole indagare le scritture digitali private e prende in esame fenomeni interlinguistici in una combinazione ancora non molto studiata (italiano e italiano di tedescofoni) ma molto promettente per le differenze cross culturali.

Il contributo ha mostrato un esempio di consultazione del *corpus* per una ricerca durante uno studio di replicazione e ha reso evidente come il *corpus*, seppur nei limiti della sua forma ad oggi non ancora completa, fornisca una prospettiva inedita nella ricerca sulla pragmatica interculturale e crossculturale (Cohen 2018: 11) e possa contribuire in modo innovativo soprattutto alla didattica dell'italiano L2 di tedescofoni.

Un intervento didattico che si fondasse su l'interpretazione dei dati (§ 4.5) sarebbe basato su un *corpus* di produzioni autentiche e in quanto tale orientato ai bisogni degli apprendenti. Tale intervento dovrebbe però far percepire la non normatività delle scelte sociopragmatiche e dovrebbe avere come obiettivo il sostegno della consapevolezza linguistica dell'atto pragmatico e la riflessione sul rapporto interpersonale che si instaura in dipendenza delle diverse scelte. I metodi didattici che possono condurre a questi obiettivi sono eterogenei (Trubnikova, Garofolin 2020: 103) e dovrebbero essere adattati alle esigenze del *setting* di insegnamento. Nonostante la ricerca empirica non sia giunta a una proposta univoca gli studi a disposizione propendono per un intervento didattico che incoraggi un approccio induttivo (Takimoto 2012) e favorisca l'osservazione esplicita delle diversità tra l'uso dei madrelingua e quello degli apprendenti (Taguchi, Roever, 2017: 220). Inoltre una didattica basata sul corpus LADDER potrebbe basarsi sull'aspetto acquisizione dei dati (più evidente una volta che il *corpus* sarà più esteso e bilanciato). In questo modo l'insegnante può sapere quale aspetto sociopragmatico è indicato introdurre in un determinato livello linguistico. I dati di un'analisi acquisizionale potrebbero essere trasferiti nel *curriculum* e obiettivi sociopragmatici e le loro tappe potrebbero essere inseriti nella pianificazione didattica con una base empirica al momento mancante.

RIFERIMENTI BIBLIOGRAFICI

- Abel A., Glaznieks A. (2020), "Kohärenz digital: Zum Konnektoregebrauch in der Online-Kommunikation und dessen Repräsentation in Sprachressourcen bzw. Bildungskontexten", in *Textqualität im digitalen Zeitalter*, 2, pp. 146-173.
- Alcón Soler E. (2013), "Pragmatic variation in British and international English language users' e-mail communication: a focus on requests", in *Revista española de lingüística aplicada*, 26, pp. 25-44.
- Antonelli G. (2018), "L'e-taliano tra storia e leggende", in Lubello S. (a cura di), *L'e-taliano: scriventi e scritture nell'era digitale*, Franco Cesati Editore, Firenze, pp. 11-18.
- Artoni D., Benigni V., Nuzzo E. (2020), "Pragmatic instruction in L2-Russian: a study on requests and advice", in *Instructed Second Language Acquisition*, 4, 1, pp. 62-95.
- Artstein R., Poesio M. (2008), "Inter-Coder Agreement for Computational Linguistics", in *Computational linguistics - Association for Computational Linguistics*, 34, 4, pp. 555-596: <https://direct.mit.edu/coli/article/34/4/555/1999/Inter-Coder-Agreement-for-Computational>.
- Bali M., Rizzo G. (2006), *Espresso: ein Italienischkurs: 2, Lehr- und Arbeitsbuch, Niveaustufe A2* (1. Aufl., 5. [Dr.] ed.), Hueber, Ismaning.

- Bardovi-Harlig K. (1996), "Pragmatics and Language Teaching: Bringing Pragmatics and Pedagogy Together", in Bouton L. F. (ed.), *Pragmatics and Language Learning*. (Vol. 7), University of Illinois, Division of English as an International Language: Monograph, Urbana-Champaign, pp. p. 21-39.
- Barr R. B., Tagg J. (1995), "From Teaching to Learning: A New Paradigm for Undergraduate Education", in *Change*, Nov./Dec, pp. 13-25.
- Block J., Kuckertz A. (2018), "Seven principles of effective replication studies: strengthening the evidence base of management research", in *Management Review Quarterly*, 68, 4, pp. 355-359:
<https://link.springer.com/content/pdf/10.1007/s11301-018-0149-3.pdf>.
- Blum-Kulka S., House J., Kasper G. (2011), "Cross-cultural pragmatics: Requests and apologies", in Archer D., Grundy P. (eds.), *The pragmatics reader*, Routledge, London, pp. 344-353.
- Boulton A. (2017), "Corpora in language teaching and learning", in *Language Teaching*, 50, 4, pp. 483-506:
https://www.researchgate.net/publication/319774344_Corpora_in_language_teaching_and_learning.
- Brocca, N. (2021). LADDER. Learners' digital communication: a database for pragmatic competences in Italian L1/L2 (Version 2021.06.02) [Data set]. Zenodo.
<http://doi.org/10.5281/zenodo.4895219>.
- Brocca N. (2020), "Implizite Bedeutung in sozialen Netzwerken. Nachhaltige Bildung in L2 Französisch", in Kapelari S., Schnabl T., Taglieber J. (eds.), *Interdisziplinäre fachdidaktische Diskurse zur Bildung für nachhaltige Entwicklung*, innsbruck University Press, Innsbruck, pp. 177-205:
https://www.uibk.ac.at/iup/buch_pdfs/fachdidaktik_8/10.15203-99106-019-2-09.pdf.
- Brocca N. (in stampa), "Insegnare italiano secondo l'approccio per Task attraverso media digitali", in Hirzinger-Unterrainer E. M. (ed.), *Aufgabenorientierung im Italienischunterricht. Ein theoretischer Einblick mit praktischen Beispielen*, Narr Francke Attempto Verlag, Tübingen.
- Brocca N. (ed.) (2020), *Ladder. Learners Digital communication: a Database for pragmatic competence in L2*: <https://ladder.hypotheses.org/author/costruttivis>.
- Brocca N., Masia V., Kucelman E. (2020), "Didactics of Pragmatics as a way to improve Social Media Literacy: An experiment proposal to Polish and Italian students in L1", in Brocca N. (Ed.), *HeiEDUCATION Journal, Sonderheft, Social Networks in Education and Linguistics*, 5, pp. 81-107:
[h Heidelberg.de/journals/index.php/heied/issue/view/2400/heied%20Journal%20%7C%20Heft%205](https://www.h Heidelberg.de/journals/index.php/heied/issue/view/2400/heied%20Journal%20%7C%20Heft%205).
- Burnard, L. (2014), *What is the text encoding initiative? How to add intelligent markup to digital resources*, OpenEdition Press, Marseille.
- Castagneto M., Ravetto M. (2015), "The variability of compliment responses: Italian and German data", in Gesuato S., Bianchi F., Cheng W. (eds.), *Teaching, learning and investigating pragmatics: Principles, methods and practices*, Cambridge University Press, Cambridge, pp. 387-413.
- Chambers A. (2019), "Towards the corpus revolution? Bridging the research–practice gap", in *Language Teaching*, 52, 4, pp. 460-475.
- Cohen A. (2018), *Learning pragmatics from native and nonnative language teachers*, Multilingual Matters, Bristol.
- Corino E., Marengo C. (2009), "Didattica con i corpora di italiano per stranieri", in *Italiano LinguaDue*, 1, pp. 279-285:
<https://riviste.unimi.it/index.php/promoitals/article/view/444>.

- Cortés Velásquez D., Nuzzo E. (2017), “Disdire un appuntamento: spunti per la didattica dell’italiano L2 a partire da un *corpus* di parlanti nativi”, in *Italiano Lingua Due*, 9, 1, pp. 17-36: <https://riviste.unimi.it/index.php/promoitals/article/view/8761>.
- Council of Europe, (2020), *Common European Framework of Reference for Languages: learning, teaching, Assessment. Companion Volume*: <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4>.
Trad. it. a cura di Barsi M., Lugarini E., Cardinaletti A., *Quadro Comune Europeo di Riferimento per le Lingue: apprendimento, insegnamento, valutazione. Volume complementare*, in *Italiano LinguaDue*, 2. 2020:
<https://riviste.unimi.it/index.php/promoitals/article/view/15120>.
- de la Cruz B. L. (2019), “Making a little go a long way: A *corpus*-based analysis of a high-frequency word and some pedagogical implications for young Spanish learners”, in *International journal of English studies*, 19, pp. 1-20:
<https://revistas.um.es/ijes/article/view/349311/266631>.
- Di Eugenio B., Jordan P. W., Thomason R. H., Moore J. D. (2000), “The agreement process: an empirical investigation of human–human computer-mediated collaborative dialogs”, in *International journal of human-computer studies*, 53, 6, pp. 1017-1076.
- Economidou-Kogetsidis M. (2011), “«Please answer me as soon as possible»: Pragmatic failure in non-native speakers’ e-mail requests to faculty”, in *Journal of Pragmatics*, 43, 13, pp. 3193-3215.
- Farr F. (2008), “Evaluating the use of *corpus*-based instruction in a language teacher education context: Perspectives from the users”, in *Language Awareness*, 17, 1, 2pp. 5-43.
- Flores-Salgado E., Castineira-Benitez T. A. (2018), “The use of politeness in WhatsApp discourse and move ‘requests’”, in *Journal of Pragmatics*, 133, pp. 79-92.
- Flowerdew J. (2017), “*Corpus*-based approaches to language description for specialized academic writing”, in *Language Teaching*, 50, 1, pp. 90-106.
- Frankenberg-Garcia, A. (2012), “Raising teachers’ awareness of *corpora*”, in *Language Teaching*, 45, 4, pp. 475-489.
- Garassino D., Masia V., Brocca N. (in stampa), “Is implicit communication quantifiable? A *corpus*-based analysis of English and Italian political tweets”, in *Journal of Pragmatics*.
- Ghezzi C., Piantoni M., Bozzone Costa R. (2015), *Nuovo contatto: corso di lingua e civiltà italiana per stranieri: B1*, Loescher, Torino.
- Glaser K. (2020), “Assessing the L2 pragmatic awareness of non-native EFL teacher candidates: Is spotting a problem enough?”, in *Lodz Papers in Pragmatics*, 16, 1, pp. 33-65.
- Goffmann I. (1956), *The Presentation of Self in Everyday Life*, Anchor Book, New York.
- Granger S. (2009), “The contribution of learner *corpora* to second language acquisition and foreign language teaching: A critical evaluation”, in Aijmer K. (ed.), *Corpora and Language Teaching*, Benjamins, Amsterdam-Philadelphia, pp. 13-32.
- Hawkins J. A., Buttery P. (2009), “Using learner language from *corpora* to profile levels of proficiency: Insights from the English Profile Programme”, in Taylor L., Weir C. J. (eds.), *Language Testing Matters: Investigating the Wider Social and Educational Impact of Assessment*, Cambridge University Press, Cambridge, pp. 158-175.
- Hoek J., Scholman M. C. J. (2017), “Evaluating discourse annotation: Some recent insights and new approaches.”, in *Proceedings of the 13th Joint ISO-ACL Workshop on Interoperable Semantic Annotation*, isa-13, pp. 1-13.
- König K. (2019), “Stance taking with ‘laugh’ particles and emojis – Sequential and functional patterns of ‘laughter’ in a *corpus* of German WhatsApp chats”, in *Journal of Pragmatics*, 142, 156-170.

- Krulatz A. M. (2014), "Electronic Requests in Native and Non-Native Russian: Insights into Foreign Language Learners' Sociolinguistic Competence", in *Journal of Linguistics and Language Teaching*, 5, 1, pp. 87-97.
- Krummes C., Ensslin A. (2015), "Formulaic language and collocations in German essays: from *corpus*-driven data to *corpus*-based materials", in *Language learning journal*, 43, 1, pp. 110-127.
- Leech G. (1998), "Preface: Learner *corpora*: what they are and what can be done with them", in Granger S. (ed.), *Learner English on Computer*, Addison Wesley Longman, London-New York, pp. xiv-xx.
- Lewis M. (1997), *Implementing the lexical approach: putting theory into practice*, Language Teaching Publ., Hove (UK).
- Matsumoto N. (2020), "Developing one *corpus*-based grammar textbook about irregular multi-verb sequences in English", in *International Journal of Applied Linguistics*, s.p.: [doi:10.1111/ijal.12322](https://doi.org/10.1111/ijal.12322).
- McCarthy M. (2016), "Putting the CEFR to good use: Designing grammars based on learner-*corpus* evidence", in *Language Teaching*, 49, 1, pp. 99-115.
- Natale S., Kunz A., Krakenberger E. (2020), "I seminari di ricerca: approccio ad una linguistica empirica e partecipativa", in Sansò A. (a cura di), *Insegnare Linguistica: basi epistemologiche, metodi, applicazioni*. Atti del LIII Congresso della Società di Linguistica Italiana (Università dell'Insubria, 19-21 settembre 2019), Officinaventuno, Milano, pp. 335-348:
https://www.societadilinguisticaitaliana.net/wp-content/uploads/2021/01/021_Natale_Kunz_Krakenberger_Atti_SLI_LIII_Insubria.pdf.
- Nuzzo E. (2013), "La pragmatica nei manuali d'italiano L2: una prima indagine sull'atto linguistico del ringraziare", in *Revista De Italianistica*, 2, 26, pp. 5-29:
<https://www.revistas.usp.br/italianistica/article/view/116077/113731>.
- Nuzzo E., Cortés Velásquez D. (2020), "Canceling Last Minute in Italian and Colombian Spanish: A Cross-Cultural Account of Pragmalinguistic Strategies", in *Corpus Pragmatics*, 4, pp. 333-358:
https://www.researchgate.net/publication/341455808_Canceling_Last_Minute_in_Italian_and_Colombian_Spanish_A_Cross-Cultural_Account_of_Pragmalinguistic_Strategies.
- Nuzzo E., Santoro E. (2017), "Apprendimento, insegnamento e uso di competenze pragmatiche in italiano L2/LS: la ricerca a partire dagli anni Duemila", in *E-journal ALL*, 4, 1-27:
http://www.e-journal.org/vol4_issue2_2017-2376905x-7-116/.
- Orlandino E., Bali M., Ziglio L. (2018), *Espresso ragazzi 3 : 3 B1 : Ein Italienischkurs/Lehr- und Arbeitsbuch mit DVD und Audio-CD* (1. Auflage. ed.), Hueber Verlag, München.
- Pagliara F. (2020), "L'atto della richiesta nelle e-mail inviate dagli studenti universitari ai docenti: un confronto tra parlanti nativi e non nativi", in Grassi R. (a cura di), *La scrittura per l'apprendimento dell'Italiano L2*, Franco Cesati Editore, Firenze, pp. 225-254.
- Pallotti G. (2019), "Assessing Tasks: The Case of Interactional Difficulty", in *Applied Linguistics*, 40, 1, pp. 176-197.
- Payne S. (2013), "Compliment responses of female German and Italian university students: A contrastive study", in *Language Studies Working Papers*, 5, pp. 22-31.
- Petitjean C., Morel E. (2017), "«Hahaha»: Laughter as a resource to manage WhatsApp conversations", in *Journal of Pragmatics*, 110, pp. 1-19.
- Pistolessi E. (2004), *Il parlar spedito : l'italiano di chat, e-mail e SMS*, Esedra Editrice, Padova.

- Pistolesi E. (2016), “Aspetti diamesici”, in Lubello S. (a cura di), *Manuale di linguistica italiana*, De Gruyter, Berlin-Boston, pp. 442-458.
- Rapp A. (2017), “Manuelle und automatische Annotation”, in Jannidis F., Kohle H., Rehbein M. (eds.), *Digital Humanities: Eine Einführung*, J. B. Metzler, Stuttgart, pp. 253-267.
- Ravetto M. (2012), “Le risposte al complimento in italiano e in tedesco. Analisi di corpora di parlato”, in *Studi Italiani di Linguistica Teorica e Applicata*, 1, pp. 85-122.
- Reinmann G., Lubcke E., Heudorfer A. (2019), *Forschendes Lernen in der Studieneingangsphase: Empirische Befunde, Fallbeispiele und individuelle Perspektiven*, Springer, Wiesbaden.
- Sandel T. L., Ou C., Wangchuk D., Ju B., Duque M. (2019), “Unpacking and describing interaction on Chinese WeChat: A methodological approach”, in *Journal of Pragmatics*, 143, pp. 228-241.
- Santoro E., Vedder I. (2016), *Pragmatica e interculturalità in italiano lingua seconda*, Franco Cesati Editore, Firenze.
- Savić M. (2018), “Lecturer perceptions of im/politeness and in/appropriateness in student e-mail requests: A Norwegian perspective”, in *Journal of Pragmatics*, 124, pp. 52-72.
- Schmiderer K., Zanasi L., Konecny C., Autelli E. (2020), “Sviluppare la competenza lessicale e fraseologica tramite i task. Un contributo allo sviluppo di materiale didattico per l'italiano L2/LS”, in Serena E., Bernhard G., Gallerani I., Visser J. (a cura di), *L'italiano in contesti plurilingui: contatto, acquisizione, insegnamento*, Atti del III Convegno Internazionale di Linguistica e Glottodidattica Italiana Ruhr-Universität Bochum, Romanisches Seminar, 11-13 ottobre 2018, *Quaderni di Italiano linguaDue*, 4, in *ILD*, 12, 2, pp. 238-256:
<https://riviste.unimi.it/index.php/promoitals/article/view/15069>.
- Serra Borneto C., Bonvino E. (2003⁵), *C'era una volta il metodo : tendenze attuali nella didattica delle lingue straniere*, Carocci, Roma.
- Spooren W. P. M. S., Degand L. (2010), “Coding coherence relations: Reliability and validity”, in *Corpus Linguistics and Linguistic Theory*, 6, pp. 241-266:
https://dial.uclouvain.be/pr/boreal/object/boreal%3A72718/datastream/PDF_01/view.
- Stadler W., Will L., Eloff I. (eds.), (in stampa), *Authenticity across Languages and Cultures – Themes of Identity in Foreign Language Teaching & Learning*, Narr Verlag, Tübingen.
- Taguchi N., Roever C. (2017), *Second language pragmatics*, Oxford University Press, Oxford.
- Takimoto M. (2012), “Metapragmatic discussion in interlanguage pragmatics”, in *Journal of pragmatics*, 44, 10, pp. 1240-1253.
- Thomas J. (1983), “Cross-Cultural Pragmatic Failure”, in *Applied Linguistics*, 4, 2, pp. 91-112.
- Trubnikova V., Garofolin B. (2020), *Lingua e interazione. Insegnare la pragmatica a scuola*, ETS, Pisa.
- Ueberwasser S., Stark E. (2017), “What’s up, Switzerland? A corpus-based research project in a multilingual country”, in *Linguistik online*, 84, pp. 105-126.
- Venuti I. (2013), *A comparison of the realization of requestive speech acts in Italian and German*. Tesi di dottorato non pubblicata, Università Ca’Foscari, Venezia. Retrieved from <http://dspace.unive.it/handle/10579/3501>.
- Wilkinson M. D., Dumontier M., Aalbersberg I. J., Appleton G., Axton M., Baak A., Mons B. et al. (2016), “The FAIR Guiding Principles for scientific data management and stewardship”, in *Scientific Data*, 3, 1:
https://www.researchgate.net/publication/298345883_The_FAIR_Guiding_Principles_for_scientific_data_management_and_stewardship.