

STEFANIA MARZO

Varietà coerenti e comunità linguistiche: l'italiano ereditario in Belgio

Questo studio affronta la questione dell'unità interna delle varietà di lingua, analizzando l'uniformità dell'italiano parlato all'estero tramite due strumenti quantitativi. Lo studio si basa su un corpus di italiano parlato da italiani di seconda e terza generazione nelle Fiandre (Belgio), stratificato per genere, età e generazione. L'analisi di 35 tratti linguistici, annotati per diversi parametri socio-demografici, esplora la co-occorrenza dei tratti in una prospettiva aggregativa, considerando anche il loro significato sociale (indessicale). I dati evidenziano una coerenza linguistica concepita come *community coherence*, un'eterogeneità ordinata basata su norme condivise da comunità linguistiche definite dalla loro rete sociale. Questa coerenza si manifesta anche a livello indessicale, con l'uniformità dei tratti linguistici all'interno della comunità determinata da dinamiche di prestigio.

Parole chiave: italiano ereditario, coerenza, comunità linguistiche, prestigio.

1. Introduzione

Una questione teorica molto dibattuta nella sociolinguistica contemporanea riguarda la nozione di coerenza (Ghyselen & De Vogelaer 2014; Guy & Hinskens 2016; Cerruti & Vietti 2022). Linguisti e variazionisti in particolare si interrogano da tempo sulla coerenza e l'unità interna di lingue, dialetti e varietà di lingua. Il cosiddetto “dilemma dell'unità” ha rappresentato il punto di partenza per il volume di Beaman & Guy (2022) ed è stato oggetto di indagine attraverso un ampio ventaglio di tratti linguistici, varietà linguistiche, stili di parlato e comunità.

L'interpretazione più comune di “coerenza” nella letteratura attuale è il grado di co-occorrenza tra tratti linguistici in un insieme di dati, caratterizzato da fattori sociali come il grado di formalità (si veda Cerruti & Vietti 2022 per una panoramica degli studi recenti). Negli ultimi anni, la nozione di coerenza è stata estesa anche alle comunità di parlanti (Beaman & Guy 2022), in cui si riferisce alla natura siste-

matica dell'uso della lingua all'interno di una comunità di parlanti. Il concetto di *community coherence* implica quindi che le caratteristiche e le variabili linguistiche non sono distribuite in modo casuale tra i parlanti, ma seguono schemi riconoscibili che si allineano con le norme sociali e linguistiche (Becker 2016).

Nelle pagine seguenti esploreremo il potenziale offerto da due metodi quantitativi per misurare la coerenza interna intesa come le (dis)similarità aggregate tra variabili linguistiche definite da categorie sociali (Wieling *et al.* 2014) all'interno della comunità linguistica italiana all'estero.

Operando in questo paradigma, ci proponiamo di apportare due contributi, uno di ordine più teorico e un altro più linguistico-descrittivo. In primo luogo, analizzeremo la coerenza a livello di comunità, esaminando come i tratti linguistici co-occorrano in un gruppo sociale definito. Questo approccio ci permetterà di valutare la coerenza non solo all'interno di una singola varietà (l'italiano ereditario), ma anche tra le varietà diverse che coesistono all'interno di una comunità di parlanti. In altre parole, guarderemo oltre la variazione intra-varietale, esplorando la distanza tra le varietà che coesistono all'interno di una comunità di parlanti all'estero. Verrà posta altresì attenzione al significato sociale dei tratti linguistici, ovvero la coerenza dei tratti linguistici verrà correlata anche con l'indessicalità che essi comportano.

In secondo luogo, si desidera contribuire specificamente alla descrizione dell'italiano ereditario, ossia l'italiano parlato dai figli delle prima generazioni di migranti. Il termine *heritage speakers* si riferisce alla "second generation immigrants, the children of the original immigrants, who live in a bilingual/multilingual environment from an early age" (Benmamoun *et al.* 2013: 132). L'italiano parlato come lingua ereditaria è da sempre considerato fondamentalmente eterogeneo (Bettoni 1993: 441) perché coinvolge almeno tre sistemi linguistici compresenti nel contesto di emigrazione: il dialetto delle prime generazioni, l'italiano e la lingua locale. Poiché l'emigrazione ha prevalentemente interessato regioni e ceti sociali in cui il dialetto era diffuso, la prima generazione di italiani partita intorno al secondo dopoguerra era principalmente dialettologa (Haller 2002). Una volta insediati nei paesi di accoglienza, i dialettologi si sono spesso trovati costretti a utilizzare più frequentemente l'italiano, attenuando così i tratti dialettali più marcati e adottando una varietà di italiano sovrallocale per comunicare con connazionali di diverse regioni. Questo processo ha

stimolato e rafforzato la diffusione di una varietà di italiano con forte coloritura regionale. Inoltre, il contatto linguistico con la lingua locale ha portato a una graduale sostituzione della varietà d'origine, anche in ambito familiare, esercitando un'influenza crescente sull'italiano.

Questo complesso intreccio di fattori linguistici e parametri sociali e socio-demografici rende comprensibile perché i linguisti abbiano definito l'italiano ereditario come un insieme di tendenze eterogenee, a volte "incontrollabili" (Rovere 1990: 169). Sull'italiano in Australia Clyne (1991: 162) ha suggerito persino che "there are almost as many varieties of migrant Italian as there are Italian migrant speakers". Inoltre, dagli anni Ottanta in poi, l'italiano parlato all'estero è stato esaminato spesso attraverso la lente dell'italiano popolare (Santilli in stampa), che include tratti molto simili a quelli trovati nelle varietà all'estero, cioè forme di interlingua, con apporto residuale dei dialetti di sostrato e processi di semplificazione linguistica (Berruto 1990; Felici 2000). Tuttavia, la categoria dell'italiano popolare, sebbene utile, non coglie completamente le dinamiche linguistiche degli oriundi italiani, non permettendo di vedere la distribuzione dei fenomeni specifici né il loro rapporto con i fattori sociali (Marzo 2015). Studi recenti hanno dimostrato che l'italiano popolare e l'italiano ereditario condividono certe caratteristiche, ma seguono pattern e quindi forme di coerenza molto diverse al livello della comunità (Santilli 2024, in stampa).

Nelle pagine seguenti, esploreremo l'unità presente nell'italiano parlato all'estero, tenendo conto dell'influenza che su di esso esercitano i tre principali sistemi linguistici sopra menzionati. Nel prossimo paragrafo presenteremo le riflessioni sulla coerenza rilevanti per questo studio; successivamente, esamineremo i dati e i metodi applicati; passeremo quindi all'analisi delle componenti principali e della regressione; infine, discuteremo le loro implicazioni per la teoria della coerenza e la descrizione dell'italiano all'estero.

2. Coerenza e comunità linguistica

La coerenza linguistica è un concetto che riguarda sia l'uso della lingua che le dinamiche delle comunità linguistiche: essa si manifesta nel contrasto tra le caratteristiche linguistiche condivise all'interno di una comunità e la diversità tra i singoli parlanti. Le comunità linguistiche possono essere appunto caratterizzate da norme linguistiche salienti,

indicative di una maggiore coerenza, o da norme più diffuse, che suggeriscono una minore coerenza. Ad esempio, gli studi di Labov sull'inglese di New York hanno dimostrato che l'uso sistematico delle variabili linguistiche definisce l'appartenenza alla comunità, suggerendo una certa coerenza sociolinguistica (Labov 1972; v. anche Hinskens & Guy 2016). Questa coerenza linguistica può essere motivata da fattori strutturali o da convenzioni sociali. Allo stesso tempo, le comunità linguistiche sono caratterizzate da una notevole diversità interindividuale in termini di genere, età, istruzione, status e geografia.

Lo studio della coerenza nei sistemi linguistici ha implicazioni significative per la teoria sociolinguistica (Hinskens & Guy 2016; Beaman & Guy 2022). Comprendere la coerenza linguistica serve a chiarire come le comunità linguistiche presentino comportamenti condivisi nonostante la diversità interna. Le domande di ricerca chiave in questo ambito includono l'esame di come diverse caratteristiche linguistiche correlino all'interno di un sistema e quali caratteristiche tendano a co-variare, facendo luce sulla struttura e sulla funzione dei sistemi linguistici (Van Meel *et al.* 2016). Inoltre, è fondamentale esplorare la coerenza con cui i parlanti selezionano e combinano le varianti linguistiche nel loro discorso, nonché il grado di libertà che hanno nella scelta delle risorse linguistiche disponibili (Becker 2016). Questo ci porta alla distinzione tra sistemi linguistici coerenti e i cosiddetti processi di *bricolage* (Eckert 2018: 112), distinzione che riflette una divisione teorico-metodologica più ampia nella sociolinguistica. I sistemi coerenti implicano una covariazione sistematica delle variabili linguistiche, mentre il *bricolage* coinvolge i singoli parlanti che operano una selezione creativa all'interno delle risorse linguistiche disponibili per costruire la loro identità ed esprimere i loro atteggiamenti. In un modello di coerenza variazionista, gli individui all'interno di una comunità linguistica utilizzano tratti linguistici che riflettono i principi di stratificazione sociolinguistica a livello comunitario (Labov 1972). In un modello di *bricolage*, invece, i parlanti combinano varie risorse linguistiche in maniera idiosincratica per generare nuovi significati sociali. Come hanno osservato Guy e Hinskens (2016: 3-4), il concetto di *bricolage* sembra opporsi ad una descrizione unificata delle varietà linguistiche, poiché enfatizza innanzitutto l'adattamento e la variazione idiosincratiche, sfidando la priorità della coerenza comunitaria. Di conseguenza, i sociolinguisti hanno raramente esplorato

possibili tracce di coerenza all'interno di processi di *bricolage*. In altre parole, l'interazione tra le scelte linguistiche individuali e le norme comunitarie rimane una questione complicata; di forte interesse sono quindi gli studi che evidenziano l'importanza dei fattori sia strutturali che sociali nel comprendere come le comunità linguistiche presentano la coerenza nonostante la diversità interna.

Se si parla di *bricolage*, occorre discutere anche l'indessicalità, ovvero il significato sociale delle forme linguistiche (Silverstein 2003; Eckert 2008). Questo implica comprendere come determinati tratti riflettano significati sociali in maniera coerente. Capire come i parlanti esprimano le loro identità sociolinguistiche attraverso il linguaggio e come queste espressioni varino a seconda di dimensioni sociali come genere o età è fondamentale per comprendere la possibile coerenza di un tratto o gruppo di tratti all'interno di una comunità.

Queste riflessioni saranno applicate allo studio dell'italiano all'estero, che di principio è stato percepito come una varietà poco coerente e fortemente eterogenea, proprio in virtù delle diverse condizioni individuali dei parlanti: generazione, input linguistico diverso, ecc. Attraverso quest'analisi, potremo ottenere una comprensione più profonda di come l'italiano venga utilizzato e adattato dalle comunità di parlanti all'estero, e come i fattori sociali e strutturali contribuiscano alla formazione di varietà linguistiche potenzialmente coerenti.

3. *L'italiano parlato nelle Fiandre: dati e metodi*

I dati presentati provengono da un corpus di italiano ereditario parlato nelle Fiandre, Belgio (Marzo 2019). Raccolto tra il 2004 e il 2006, il corpus, composto da quasi 150.000 parole, include interviste sociolinguistiche e narrazioni. Il campione è composto da 48 rispondenti stratificato lungo quattro dimensioni: genere, generazione (seconda e terza), esposizione alla lingua e cultura italiana, non dialettale (contatto frequente e occasionale), e quartiere di residenza (che corrisponde ai due parametri "comunità" e "rete sociale"). Tutti i rispondenti sono nati in Belgio da genitori italiani (nati in Belgio o in Italia), hanno un'età compresa tra i 16 e i 50 anni e risiedono nel Limburgo, una provincia situata nella parte orientale della regione fiamminga. In particolare, i partecipanti vivono in due quartieri specifici, Lindeman e Zwartberg, entrambi situati nelle vecchie aree minerarie, note per

l'alta concentrazione di minoranze etniche, tra le quali le comunità italiane sono state a lungo le più numerose.

Appartengono alla seconda generazione i soggetti i cui genitori sono emigrati in età adulta o in un'età in cui la competenza linguistica era già acquisita (a partire dagli otto anni). La terza generazione è rappresentata dagli informatori che hanno almeno uno dei genitori nato in Belgio o emigrato prima di aver acquisito pienamente la competenza linguistica (prima degli otto anni).

La dimensione "esposizione alla lingua e cultura italiana" è definita secondo quattro parametri: (1) viaggi e soggiorni in Italia (una volta all'anno/più di una volta all'anno/meno di una volta all'anno); (2) presenza in casa di una persona italiana di prima generazione, come partner, genitori o nonni (sì/no); (3) origine italiana del partner (sì/no); (4) situazione professionale o impegno personale che implica un contatto regolare con la lingua o la cultura italiana (sì/no). Si considera un contatto "frequente" con la lingua e la cultura italiana quando in almeno tre dei quattro parametri viene indicato il valore che implica un maggiore contatto, ossia "più di una volta all'anno" per il primo parametro e "sì" per i restanti tre parametri.

3.1. *Tratti linguistici e parametri sociali*

Riprendendo gli schemi degli studi già compiuti sull'italiano e sulle altre lingue in situazione di contatto, sono state individuate tre dimensioni di variabilità linguistica: tratti regionali, tratti di semplificazione e livellamento e tratti di contatto. I fenomeni esaminati riguardano aspetti fonologici, morfosintattici e lessicali.

I tratti regionali raggruppano i fenomeni provenienti dai dialetti della regione di origine. Nel corpus occorrono sia tratti centromeridionali (ad esempio, la caduta della finale *-re* nelle forme dell'infinito, l'uso del possessivo posposto) sia, in minor parte, tratti più tipicamente settentrionali (come l'uso del *che* dopo le congiunzioni *quando* e *mentre*).

I tratti di semplificazione e di livellamento comprendono i tratti relativi ai fenomeni di semplificazione, spesso associati all'italiano popolare (Berruto 1983, 1990). Questa categoria include fenomeni di livellamento pronominale (uso di *ci* al posto di *gli*, *le* e *loro*), la semplificazione dell'articolo maschile (uso di *il* in ogni contesto) e l'uso di forme analitiche al posto di quelle sintetiche per i pronomi possessivi, relativi e personali (es. *i bambini che gli ho dato il premio*).

I tratti di contatto derivano dal contatto con la varietà locale del neerlandese (il fiammingo), lingua ormai dominante presso le nuove generazioni degli italiani del Limburgo. Esempi includono la caduta dell'articolo davanti ai pronomi possessivi (es. *mia casa*) e l'uso della preposizione dopo costrutti impersonali seguiti dall'infinito (es. *è bello di andare*).

La lista in Tabella 1 presenta tutti i fenomeni analizzati, con le abbreviazioni utilizzate nelle pagine seguenti. Mediante l'analisi delle componenti principali e l'analisi della regressione si cercherà di comprendere la variazione nell'uso di questi tratti linguistici, identificando chi utilizza quali varianti e con quale frequenza. In questo modo sarà possibile controllare la formazione di determinati gruppi di persone con un uso linguistico simile.

Tabella 1 - *Variabili linguistiche*

Tratti regionali		
	Complementatore doppiamente riempito (<i>quando che – quando</i>)	N.cong.che
<i>Centromeridionali</i>		
	Conservazione della -e latina (<i>de – di</i>)	R.de
	Apocope dell'infinito (<i>parla' – parlare</i>)	R.inf.ap
	Apocope di <i>sono</i> (<i>so' – sono</i>)	R.so
	Apocope di <i>due</i> (<i>du' – due</i>)	R. du
	Apocope della preposizione (<i>pe' – per, co' – con</i>)	R.pe R.co
	<i>Stare + a + infinito</i> (<i>sto a guardare – sto guardando</i>)	R.stare.inf
	<i>a + oggetto diretto</i> (<i>aspettare a Paolo – aspettare Paolo</i>)	R.acc.prep
	Posposizione dei possessivi (<i>fratello mio – mio fratello</i>)	R.pos.post
	Ausiliare <i>avere</i> con i verbi riflessivi (<i>mi ho lavato – mi sono lavato</i>)	R.av.ess

Tratti regionali	Variabili lessicali (<i>mo' – adesso – ora / stare – essere / pure – anche</i>)	R.mo' R.stare.ess R.pure
Livellamento sintagmatico e paradigmatico	Articolo definito singolare + cons. (<i>il studente – lo studente</i>)	L.il.s
	Articolo indefinito singolare + cons. (<i>un studente – uno studente</i>)	L.un.s
	Definiti e indefiniti plurali + vocale (<i>i/dei studenti – gli/degli studenti</i>)	L.i.voc.
	Introduttori di relative (<i>che – cui</i>)	L.che.pol
	Pronome dativo <i>ci</i> (<i>ci – gli/le/loro</i>)	L.ci.pol
	Analogia nelle desinenze verbali (<i>ono – ano</i>) <i>parlono – parlano</i>	L.ono
	Accordo fra soggetto e verbo (<i>assenza/presenza di accordo</i>)	L.accordo
	Uso avverbiale degli aggettivi (<i>parlare diverso – parlare diversamente</i>)	L.agg.avv
Tratti di interferenza	Verbi e costrutti impersonali + Prep + Infinito (<i>è bello di andare – è bello andare</i>)	C.imp.prep
	Omissione della preposizione dopo <i>andare, venire e continuare</i> (<i>continuare fare – continuare a fare</i>)	C.verbo.inf
	Omissione dell'articolo definito con i possessivi (<i>mia casa – la mia casa</i>)	C.art.pos
	Omissione dell'articolo definito con i nomi di paesi (<i>Belgio – il Belgio</i>)	C.art.paes
	Ausiliare <i>essere</i> invece di <i>avere</i> (<i>siamo smessi – abbiamo smesso</i>)	C.es.av
	Posizione post-focale del focalizzatore <i>anche</i> (<i>io anche vengo – vengo anch'io</i>)	C.anche

Per ogni variabile è stato calcolato un valore relativo basato sulla somma delle frequenze relative delle varianti standard e non standard (regionali, semplificate o di contatto), secondo il principio variazionista della *envelope of variation*¹ (Speelman *et al.* 2003; Tagliamonte 2006).

I fattori sociali che influiscono sulla scelta di varianti standard o non standard in contesto migratorio sono molteplici. Generalmente si considerano il periodo di emigrazione, la generazione e la lingua di contatto. Altri elementi rilevanti includono genere, età, livello di istruzione e atteggiamento del parlante verso le proprie origini. Oltre ai parametri socio-demografici, questo articolo tiene conto anche della rete sociale dei rispondenti, identificata con il quartiere di residenza. La definizione di rete sociale si ispira alla teoria di Milroy (1987, 2002) e agli studi di Gumperz (1964) e Labov (1972). I concetti chiave sono la molteplicità e la densità: la densità di una rete sociale dipende dal grado di conoscenza reciproca tra i membri della rete di un individuo, mentre la molteplicità riguarda il numero di relazioni diverse (parentela, lavoro, amicizia) tra i membri della rete. Reti ad alta densità e molteplicità (*close-knit*) rafforzano le norme linguistiche interne e quindi anche la coerenza, mentre reti a bassa densità (*loose-knit*) facilitano la diffusione di innovazioni linguistiche.

Sulla base della *network strength scale* di Milroy (2002), si è calcolata la forza della rete sociale nei due quartieri italiani studiati, con un punteggio numerico basato su indicatori di densità e molteplicità (si veda Marzo & Vanvolsem 2009 per ulteriori dettagli). Il quartiere Lindeman (LM) emerge come la comunità più densa e molteplice (*close-knit*), mentre il quartiere Zwartberg (ZW) appare meno coeso (*loose-knit*).

¹ In ambito sociolinguistico, il termine *envelope of variation* indica l'insieme dei contesti in cui si manifesta la variazione di una variabile linguistica. Questo concetto risulta centrale nella sociolinguistica variazionista, poiché evidenzia la necessità di definire con precisione i contesti entro i quali possono essere osservate diverse realizzazioni della variabile, al fine di garantire un'analisi accurata e approfondita dell'uso linguistico (Tagliamonte 2006: 86-87).

Tabella 2 - *Parametri sociali e sociolinguistici*

Caratteristiche socio-demografiche	Contatto linguistico	Rete sociale
Genere uomo-donna	Origine regionale (Nord – Centro – Sud)	Close-knit vs. Loose-knit
Generazione 2a – 3a	Esposizione all’italiano parlato in Italia (Occasionale – Frequente)	
Età 15-30 31-50	Contatto con varietà regionale e/o dialetto (Occasionale – Frequente)	
Livello di studio	Corsi di lingua italiana	

3.2 *Analisi delle componenti principali*

L’analisi delle componenti principali (*Principle Component Analysis* o PCA) è una tecnica statistica descrittiva ed esplorativa, spesso utilizzata nelle prime fasi di elaborazione dei dati (si veda Marzo & Vanvolsem 2009; Cerruti & Vietti 2022). Il metodo serve a fornire una visione generale del problema, a comprendere le relazioni tra gli oggetti o le variabili e a dare un’indicazione preliminare sul loro ruolo. La PCA viene impiegata per la riduzione dimensionale del campione: in termini tecnici, permette di individuare in un gruppo di variabili correlate (x_1, x_2, \dots, x_q) poche nuove variabili sintetiche, dette componenti, non correlate tra loro (y_1, y_2, \dots, y_q), ciascuna delle quali è una combinazione lineare delle variabili originali, mantenendo tuttavia le informazioni essenziali sul fenomeno studiato. Il metodo consente quindi di identificare e di estrarre una struttura che raggruppa le variabili più importanti, talvolta latenti.

Un esempio potrebbe riguardare lo studio delle abitudini di consumo di un gruppo di persone. Ogni individuo potrebbe essere valutato in base a diverse categorie di consumo (alimentazione, intrattenimento, istruzione, trasporti, ecc.) i cui valori costituiscono le variabili osservate. Dopo aver raccolto i dati, si dispone di un gran numero di variabili; sarà quindi utile ridurle a poche variabili sintetiche. Si potrebbe costruire una nuova variabile, come “spesa complessiva”, che

sia una combinazione lineare delle variabili di spesa originarie, riassumendo le abitudini di consumo degli individui.

La PCA si basa quindi sulla riduzione della complessità delle variabili a una serie di componenti principali facilmente interpretabili, capaci di evidenziare e sintetizzare l'informazione contenuta nella matrice originale, minimizzando la perdita di informazione. La prima componente principale è in grado di spiegare la maggior parte della varianza (intesa come la variazione spiegata), la seconda ne spiega un po' meno, la terza ancora meno, e così via. Le ultime componenti contribuiscono a spiegare poco o nulla della variabilità presente nel set di variabili originali. L'obiettivo è ottenere una quantità accettabile di varianza spiegata. Considerando solo un numero limitato di componenti, solitamente due o tre, ci si può concentrare sull'informazione rilevante, eliminando il cosiddetto "rumore", ovvero la variabilità residua che accompagna la struttura di base senza aggiungere informazioni essenziali.

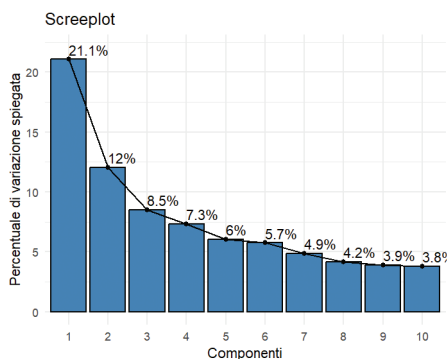
Un vantaggio importante di questa riduzione è che le componenti principali, in quanto nuove variabili sintetiche, possono essere utilizzate in analisi della regressione per misurare la correlazione tra variabili linguistiche e fattori sociali. L'uso delle componenti principali, anziché delle numerose variabili originali, elimina il rischio di multicollinearità. Per tornare all'esempio sulle abitudini di consumo, se si vuole analizzare l'effetto della spesa per alimentazione e intrattenimento sul risparmio, si potrebbero utilizzare le componenti principali derivate dalle varie categorie di spesa per semplificare il modello di regressione. Questo garantisce che le variabili non siano correlate tra loro e fornisce una stima più precisa delle correlazioni, permettendo di comprendere meglio come le diverse abitudini di consumo influenzino la capacità di risparmio degli individui.

Il numero delle componenti da selezionare dipende da vari criteri e può essere stabilito con diversi metodi. Occorre identificare il numero di componenti che spiegano la maggior parte della varianza comune a tutte le variabili originali. I due criteri principali da considerare sono: la proporzione di varianza, ovvero la variazione totale spiegata dalle componenti selezionate, e la proporzione cumulativa, ovvero la quantità di variazione estratta dalle componenti ricavate in successione. Secondo quest'ultimo principio, le componenti che non aggiungono

una quantità considerevole di variazione spiegata non risultano particolarmente utili per l'interpretazione globale della variazione.

Questi criteri possono essere integrati con il cosiddetto *scree-test*, una rappresentazione grafica che visualizza quante componenti occorre individuare: in un istogramma (detto *screeplot*, Figura 1) si riportano sull'ordinata le diverse proporzioni di varianza e sull'ascissa il numero delle componenti. Nell'istogramma è possibile individuare il punto in cui le barre tendono ad appiattirsi, formando quasi una linea orizzontale. A partire da questo punto, le componenti non spiegano più una quantità significativa di variazione e possono essere tralasciate. La prima componente principale spiega il 21,1% della variazione nel campione di dati, alla quale la seconda aggiunge un ulteriore 12%, la terza componente vi aggiunge l'8,5% di variazione spiegata. Nella Figura 1, le barre si appiattiscono all'altezza delle componenti 3 e 4, indicando che è opportuno estrarre le prime tre componenti. In tutto le prime tre componenti spiegano il 41,6% della variazione.

Figura 1 - *Screeplot*

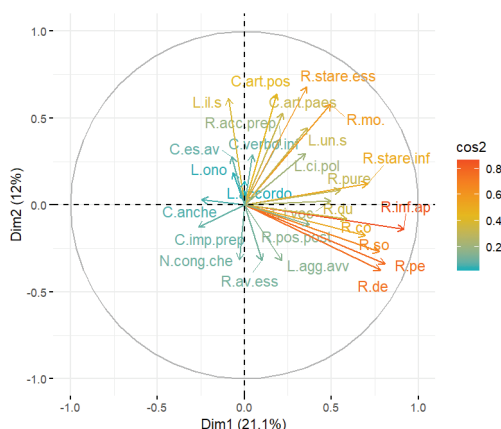


4. Risultati

4.1. Eterogeneità strutturata

Oltre ai vari criteri e metodi di rappresentazione, l'individuazione e la definizione delle componenti richiede anche una visualizzazione concreta delle variabili linguistiche e sociali attraverso una serie di grafici o *biplo*t, che verranno discussi nella sezione dei risultati.

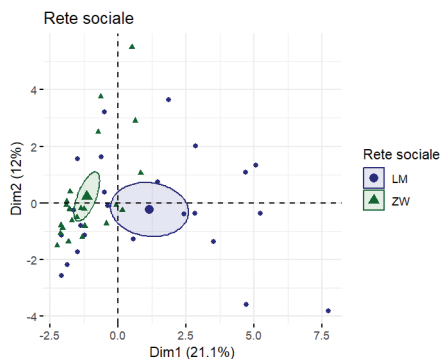
Figura 2 - Biplot con le due prime componenti principali e le variabili linguistiche



Il *biplot* in Figura 2 visualizza le prime due componenti principali (PC1 e PC2) del *dataset*, con le variabili linguistiche rappresentate da frecce colorate in base ai loro punteggi *cos2*, che indicano la qualità della rappresentazione delle variabili sulle componenti principali. Le frecce più lunghe e colorate indicano variabili che contribuiscono maggiormente alle componenti principali. La direzione delle frecce mostra come ciascuna variabile influisce su PC1 e PC2. Ad esempio, variabili come R.inf.ap (apocope dell'infinito) e R.de (conservazione della *-e* latina) rivelano un alto contributo, come indicato dalle frecce più lunghe e intensamente colorate. Questa visualizzazione è utile per identificare le variabili più influenti e capire come sono correlate alle prime due componenti principali.

Dal *biplot* possiamo immediatamente dedurre che si forma nei due quadranti a destra un *cluster* di tratti prevalentemente di area centromeridionale (come l'infinito apocopato e altre variabili di tipo R.x) (CP 1), mentre nei quadranti a sinistra si trovano più tratti di livellamento e di interferenza (L.il.s e L.ono e altre variabili di tipo L.x e C.x) (CP 2).

Figura 3 - *Distribuzione dei rispondenti a seconda del quartiere o rete sociale.*
La numerazione corrisponde a ciascun partecipante



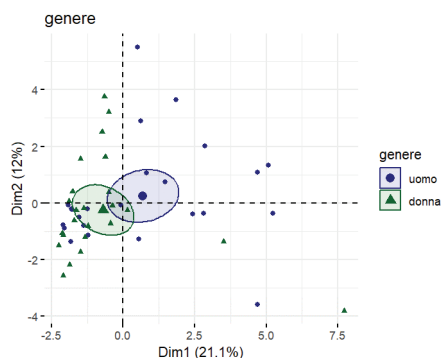
Il grafico delle variabili linguistiche può essere messo in relazione con i soggetti del campione, creando così un altro grafico che mette in luce le analogie e le correlazioni tra i rispondenti stessi, in funzione dell'occorrenza dei tratti linguistici nella loro parlata. Il grafico nella Figura 3 mostra che il *cluster* delle caratteristiche regionali coincide principalmente con i rispondenti provenienti dalla rete chiusa (LM), mentre i tratti di livellamento corrispondono maggiormente ai rispondenti della rete aperta (ZW). Occorre sottolineare che i rispondenti dalla rete chiusa non hanno tutti origini centromeridionali. Al contrario, alcuni di loro provengono da regioni settentrionali e hanno semplicemente adottato caratteristiche del Centro-Sud (Marzo 2008). Nel quartiere sociale denso, le caratteristiche centromeridionali quindi formavano una sorta di nuova *koinè* regionale (Trudgill 1986; Kerswill 2013; Cerruti & Tsipakou 2020), utilizzata principalmente dagli uomini, come evidenziato anche dal grafico della Figura 4, che illustra come siano esclusivamente questi a utilizzare tali varianti.

Nella Figura 3 i rispondenti indicati in celeste corrispondono agli abitanti di LM (quartiere a forte coesione interna), quelli in verde agli abitanti di ZW (quartiere meno coeso). Il grafico evidenzia tre gruppi relativamente congruenti, uno collocato a destra lungo l'asse della prima componente principale e due altri gruppi omogenei situati invece a sinistra.

La Figura 4 ripropone la distribuzione dei rispondenti, ma suddivisi per genere; è interessante notare che nei quadranti a destra

prevalgono i parlanti di genere maschile. Ciò offre una prospettiva interessante sul significato sociale dei fenomeni linguistici che si manifestano prevalentemente in quell'area, argomento che verrà approfondito nel §4.2.

Figura 4 - *Distribuzione dei rispondenti a seconda del genere.*
La numerazione corrisponde a ciascun partecipante



Questi due grafici permettono già di trarre una serie di conclusioni provvisorie. Proiettando la posizione degli informanti su quella delle variabili linguistiche delle Figure 3 e 4 risulta chiaro che nel corpus si forma un *cluster* (sull'asse della PC1) di tratti prevalentemente centromeridionali, usati da un gruppo di informanti, soprattutto provenienti dal quartiere LM. La parlata del gruppo sull'asse della PC2, un gruppo di informanti la cui maggioranza appartiene al quartiere ZW, sembra invece caratterizzarsi soprattutto da fenomeni di livellamento e di interferenza. Per contro, il gruppo di parlanti in basso sullo stesso asse, è difficilmente associabile a determinati fenomeni linguistici, poiché in quella posizione il grafico non presenta nessun *cluster* particolare di variabili linguistiche.

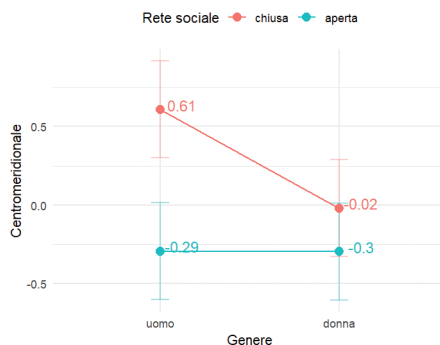
4.2. Coerenza e indessicalità

L'analisi delle componenti principali ha identificato i principali tratti linguistici distintivi all'interno del *dataset*. In particolare, l'analisi ha rivelato l'esistenza di due *cluster* principali di tratti linguistici: uno composto da tratti prevalentemente centromeridionali e l'altro da tratti di livellamento.

Il primo *cluster* include caratteristiche linguistiche comuni nelle regioni centrali e meridionali. La coerenza interna di questo *cluster* suggerisce una forte omogeneità nei tratti linguistici che lo compongono. Analogamente, il secondo *cluster*, con i tratti di livellamento, raggruppa tratti che indicano un processo di uniformità linguistica, riducendo la variazione regionale.

Nella fase successiva di questo studio, si procede con la creazione di variabili linguistiche basate sui due *cluster* individuati. In particolare, si è calcolata la media dei tratti linguistici più importanti di ciascun *cluster*, ottenendo così due nuove variabili: la media dei tratti centromeridionali e la media dei tratti di livellamento e di interferenza. Queste variabili medie saranno successivamente utilizzate come variabili dipendenti in un modello di regressione lineare, che ci permetterà di esplorare le relazioni tra i tratti linguistici e altre variabili indipendenti presentate nel §3.1. L'analisi della regressione mira a comprendere meglio i meccanismi che influiscono sulla distribuzione dei tratti linguistici all'interno delle due comunità studiate e, di conseguenza, a capire meglio eventuali livelli di indesicalità che ne determinano la coerenza.

La regressione lineare per i tratti centromeridionali dimostra un'importante interazione tra genere e rete sociale, come illustrato dalla Figura 5. Dal grafico si evince che, nelle reti sociali dense, vi è una differenza altamente significativa nell'uso delle forme regionali tra uomini e donne ($F(3,44)=7,655, p=0,0003$); in effetti, sono gli uomini a diffondere maggiormente le forme regionali. Questo fenomeno può essere spiegato attraverso il concetto di *covert prestige* per gli uomini, mentre le donne, che ricorrono meno alle forme centromeridionali, manifestano un *overt prestige*. Questa indesicalità appare rilevante solo nei quartieri ad alta densità sociale; vi è maggiore coerenza nelle comunità con forti interazioni e dinamiche di italianità, dove nascono le cosiddette norme condivise, anche in termini di prestigio. In questo caso particolare, gli uomini tendono a imitare e utilizzare tratti centromeridionali.

Figura 5 - *Grafico della regressione con i tratti centromeridionali*

Per quanto riguarda il secondo *cluster*, si osservano dinamiche più tipicamente associate al contatto linguistico. La regressione presenta una correlazione significativa con il contatto linguistico con l'italiano: coloro che hanno un contatto più occasionale con l'italiano (ad esempio, leggono poco in italiano, visitano raramente l'Italia, ecc.) tendono a usare più frequentemente tratti di livellamento e di contatto ($F(1,46)=2,841, p=0,09864$). In questo caso, è quindi l'input della lingua a determinarne l'occorrenza. Anche le generazioni sembrano avere un ruolo importante, sebbene meno significativo, così come il genere: le donne sembrano meno propense ad usare tratti di livellamento, ma con solo lieve significatività rispetto ai tratti regionali. La coerenza per questi tratti sembra, di fatto, essere meno fortemente determinata da norme condivise all'interno di un quartiere. I tratti di livellamento e di contatto sono più frequenti nelle reti sociali meno dense, probabilmente a causa del minor contatto con l'Italia.

5. Interpretazione e conclusione

A partire dagli anni Ottanta del secolo scorso, si è diffusa la convinzione che l'italiano parlato all'estero sia estremamente eterogeneo a causa dei profili altamente diversi dei parlanti. L'obiettivo principale di questo studio è stato di cercare una forma di coerenza linguistica a livello della comunità dei cosiddetti "parlanti ereditari". Inizialmente, è stata condotta un'analisi delle componenti principali per ridurre la variabilità del *dataset* raccolto e per identificare nuove componenti

omogenee (come già presentato in Marzo & Vanvolsem 2009). Sulla base delle componenti individuate, sono state eseguite delle regressioni lineari per comprendere la variazione interna a questi *cluster* e rivelare il significato sociale.

L'analisi delle componenti principali è stata fondamentale per isolare i tratti linguistici più significativi e comprenderne la distribuzione all'interno della comunità. Successivamente, le regressioni lineari sono state utilizzate per analizzare la variazione interna ai *cluster* identificati. Questo approccio ha permesso di esplorare il significato sociale dei tratti linguistici e di capire come questi siano distribuiti tra i parlanti. In particolare, si è cercato di spiegare perché alcuni parlanti utilizzano più varianti regionali rispetto ad altri, indagando i possibili significati sociali associati.

Nella comunità densa, caratterizzata dai tratti linguistici regionali centromeridionali, gli uomini sono i principali diffusori di norme linguistiche *covert*, ossia forme che, pur non standard, godono di prestigio all'interno del gruppo sociale. Le donne, invece, sono le principali promotrici di elementi linguistici standard. Questo fenomeno riflette delle dinamiche di prestigio già osservate negli studi di Milroy sulla comunità operaia di Belfast (Milroy 2002), in cui gli uomini con reti sociali dense tendono a preservare e diffondere forme linguistiche regionali, percepite come segni di appartenenza e autenticità nella comunità.

La coerenza riscontrata nei tratti di livellamento e di interferenza è di un ordine diverso rispetto a quella dei tratti regionali. In questo caso, i meccanismi di contatto linguistico giocano un ruolo preponderante. Ad esempio, i parlanti con meno contatto diretto con l'Italia tendono a utilizzare più frequentemente questo genere di tratti. Questo indica che il grado di esposizione e interazione con la lingua italiana influisce significativamente sulla variazione linguistica, portando a una maggiore uniformità tra i parlanti che hanno meno opportunità di mantenere forme linguistiche regionali specifiche.

In entrambi i casi si può parlare di una coerenza al livello della comunità linguistica determinata dal tipo di quartiere e dalla rete sociale, nonostante la variazione interna. Questo implica che, nonostante l'apparente eterogeneità, esistono norme condivise all'interno delle comunità linguistiche che guidano l'uso delle varianti linguistiche. Comprendere questi meccanismi offre una prospettiva più approfondita

dita sulle varianti linguistiche e sulle norme che le regolano in una comunità linguistica (si veda anche Tamminga & Wade 2022).

L'integrazione dell'analisi delle componenti principali con la regressione lineare ci ha permesso, da un lato, di esplorare la struttura della variazione linguistica e, dall'altro, di comprendere i meccanismi sottostanti che ne spiegano l'origine. L'uso delle varianti regionali sembra essere influenzato principalmente da significati sociali o indessicali associati a forme di prestigio sia tradizionali (*overt*) che non tradizionali (*covert*), presenti e rilevati in numerosi altri studi socio-linguistici. I tratti di livellamento e di interferenza sembrano meno legati a meccanismi di prestigio. Qui emergono fattori tipici delle lingue ereditarie, poiché l'input linguistico e il contatto con l'italiano svolgono un ruolo significativo.

Riccollegandoci alla parte teorica di questo contributo, in cui abbiamo descritto il concetto di *bricolage* in relazione alla teoria della coerenza, possiamo affermare, basandoci su un campione di dati limitato e un'interpretazione cauta, che la varietà linguistica ereditaria nelle Fiandre mostra forme di coerenza a seconda della comunità. Questa coerenza si manifesta all'interno delle singole comunità, ognuna delle quali presenta forme specifiche di variazione interna: nelle reti sociali dense valgono più chiaramente dinamiche indessicali di prestigio, con la creazione di *koinè* regionali; nelle reti sociali più aperte si osservano tendenze più legate al livellamento sintagmatico e paradigmatico e al ridotto contatto con la lingua ereditaria.

Nonostante i risultati emersi dalle analisi offrano nuovi spunti per la riflessione teorica sulle varietà italiane usate all'estero, essi sono tutt'altro che definitivi. Ulteriori ricerche sono indispensabili per comprendere appieno le dinamiche interne dell'italiano parlato all'estero. Sono necessarie ricerche sull'italiano ereditario, includendo analisi di dati di conversazione tra parlanti, simili a quelli dei corpora recenti creati in Italia, come il KIParla (Mauri *et al.* 2019; Ballarè *et al.* 2022), e questo per individuare forme di coerenza e indessicalità che emergono dalle interazioni spontanee. Dati interazionali sono particolarmente rilevanti per le nuove (quarte e quinte) generazioni discendenti dai migranti del dopoguerra, poiché esse tendono a parlare meno frequentemente l'italiano.

Per concludere, ci sarebbe ancora molto da dire sulla relazione e il confronto tra i tratti di livellamento dell'italiano popolare in Italia

e quelli osservati all'estero, attualmente oggetto di studio (Santilli 2024). Questo confronto sarà fondamentale per delineare le caratteristiche distintive dell'italiano popolare e i fenomeni di livellamento linguistico in generale. Inoltre, questo paragone fornirà risultati più solidi sulle possibili differenze tra le varietà di italiano popolare parlate all'interno e al di fuori dei confini italiani, offrendo così una panoramica più completa dei processi linguistici in atto.

Riferimenti bibliografici

- Ballarè, Silvia & Goria, Eugenio & Mauri, Caterina. 2022. *Italiano parlato e variazione linguistica. Teoria e prassi nella costruzione del corpus KIParla*. Bologna: Pàtron.
- Beaman, Karen & Guy, Gregory R. (a cura di). 2022. *The coherence of linguistic communities: Orderly heterogeneity and social meaning*. New York: Routledge.
- Becker, Kara. 2016. Linking community coherence, individual coherence, and bricolage: The co-occurrence of (r), raised bought and raised bad in New York City English. *Lingua* 172/173. 87–99.
- Benmamoun, Elabbas & Montrul, Silvina & Polinsky, Maria. 2013. Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics* 39(3-4). 129–181.
- Berruto, Gaetano. 1983. L'italiano popolare e la semplificazione linguistica. *Vox Romanica* 42. 38–79.
- Berruto, Gaetano. 1990. Semplificazione linguistica e varietà sub-standard. In Holtus, Günther & Radtke, Edgar (a cura di), *Sprachlicher Substandard. III. Standard, Substandard und Varietätenlinguistik*, 17–43. Tübingen: Niemeyer.
- Bettoni, Camilla. 1993. L'italiano fuori d'Italia. In Sobrero, Alberto A. (a cura di), *Introduzione all'italiano contemporaneo. II. La variazione e gli usi*, 411–460. Roma-Bari: Laterza.
- Cerruti, Massimo & Tsiplakou, Stavroula. 2020. Koinai and regional standard varieties in Europe: An introduction. In Cerruti, Massimo & Tsiplakou, Stavroula (a cura di), *Intermediate language varieties. Koinai and regional standards in Europe*, 1–29. Amsterdam/Philadelphia: Benjamins.
- Cerruti, Massimo & Vietti, Alessandro. 2022. Identifying language varieties. Coexisting standards in spoken Italian. In Beaman, Karen V. & Guy,

- Gregory R. (a cura di), *The coherence of linguistic communities: Orderly heterogeneity and social meaning*, 261-280. New York: Routledge.
- Clyne, Michael. 1991. *Community languages. The Australian experience*. Cambridge University Press. Cambridge.
- Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12(4). 453-476.
- Eckert, Penelope. 2018. *Meaning and linguistic variation: The third wave in Sociolinguistics*. Cambridge: Cambridge University Press.
- Ghyselen, Anne-Sophie & De Vogelaer, Gunther. 2018. Seeking systematicity in variation: theoretical and methodological considerations on the variety concept. *Frontiers in Psychology* 9. (<https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2018.00385/full>) (Consultato il 21.11.2024.)
- Gumperz, John J. 1964. Linguistic and social interaction in two communities. *American Anthropologist* 66. 137-154.
- Guy, Gregory R. & Hinskens, Frans. 2016. Linguistic coherence: Systems, repertoires and speech communities. *Lingua* 172/173. 1-9.
- Haller, Hermann W. 2002. Italian in New York. In García, Ofelia & Fishman, Joshua A. (a cura di), *The Multilingual Apple: Languages in New York City*, 119-142. Berlin/New York: de Gruyter.
- Kerswill, Paul. 2013. Koineization. In Chambers, J.K. & Schilling, Natalie (a cura di), *The handbook of language variation and change*, 669-702. Oxford: Wiley-Blackwell.
- KIParla = <https://www.kiparla.it/> (Consultato il 21.11.2024.)
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Marzo, Stefania. 2008. Italiano, lingua franca in due comunità di Italiani all'estero. *Italienisch: Zeitschrift für Italienische Sprache und Literatur* 1. 48-65.
- Marzo, Stefania. 2019. From flamano to urban vernacular. Linguistic and meta-linguistic heritage of first generation miners in Flemish Limburg. *International Journal of the Sociology of Language* 258. 99-199.
- Marzo, Stefania & Vanvolsem, Serge. 2009. Tecniche statistiche per lo studio variazionale dell'italiano. In Ferrari, Giacomo & Benatti, Ruben & Mosca, Monica (a cura di), *Linguistica e modelli tecnologici di ricerca. Atti del XL Congresso Internazionale di Studi della Società di Linguistica Italiana (Vercelli 21-23 settembre 2006)*, 149-168. Roma: Bulzoni.

- Mauri, Caterina & Ballarè, Silvia, & Gorla, Eugenio & Cerruti, Massimo & Suriano, Francesco. 2019. KIParla Corpus: A New Resource for Spoken Italian. In Bernardi, Raffaella & Navigli, Roberto & Semeraro, Giovanni (a cura di), *CLiC-it 2019 – Italian conference on Computational Linguistics. Proceedings of the sixth Italian conference on Computational Linguistics*. (<https://ceur-ws.org/Vol-2481/>) (Consultato il 21.11.2024.)
- Milroy, Lesley. 1987. *Language and social networks*. New York: Basil Blackwell.
- Milroy, Lesley. 2002. Social networks. In Chambers, J.K. & Schilling, Natalie (a cura di), *The handbook of language variation and change*, 549–72. Oxford: Wiley-Blackwell.
- Rovere, Giovanni. 1990. Gli studi sull'emigrazione veneta in una prospettiva sociolinguistica. In Padoan, Giorgio (a cura di), *Presenza, cultura, lingua e tradizioni dei veneti nel mondo. II. I paesi di lingua inglese*, 151–213. Venezia: Giunta Regionale del Veneto.
- Santilli, Enzo. 2024. *Contrasting italiano popolare and heritage Italian. The case of definite and indefinite determiners*. (Paper presented at ICLaVe 12 Vienna).
- Santilli, Enzo. In stampa. L'italiano popolare nel parlato di giovani della Marsica orientale: evidenze dal corpus Nec Sine. *Rivista Italiana di Dialettologia* 48 (2025).
- Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language & Communication* 23(3-4). 193–229.
- Speelman, Dirk & Grondelaers, Stefan & Geeraerts, Dirk. 2003. Profile-Based Linguistic Uniformity as a Generic Method for Comparing Language Varieties. *Computers and the Humanities* 37(3). 317–337.
- Tagliamonte, Sali A. 2006. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.
- Tamminga, Meredith & Wade, Lacey. 2022. Coherence across social and temporal scales. In Beaman, Karen V. & Guy, Gregory R. (a cura di), *The coherence of linguistic communities: Orderly heterogeneity and social meaning*, 34–52. New York: Routledge.
- Trudgill, Peter J. 1986. *Dialects in contact*. Oxford: Wiley-Blackwell.
- Van Meel, Linda & Hinsken, Frans & Van Hout, Roeland. 2016. Co-variation and varieties in modern Dutch ethnolects. *Lingua* 172/173. 72–86.
- Wieling, Martijn, Montemagni, Simonetta, Nerbonne, John & Baayen, R. Harald. 2014. Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and socio-demographic variation using generalized additive mixed modeling. *Language* 90(3). 669–692