

Questo volume offre una panoramica aggiornata della situazione sociolinguistica di una comunità di sinti lombardi residente a Pavia, i cui membri sono bilingui in sinto lombardo (romaní) e italiano. Sulla base di questionari e interviste sociolinguistiche, il volume esamina il repertorio linguistico della comunità, con particolare attenzione alle pratiche di parlato bilingue, utilizzando metodi propri degli studi sul contatto linguistico e della sociolinguistica variazionista. Viene analizzata una varietà di schemi di parlato bilingue, con particolare attenzione all'inserzione di elementi italiani in discorsi altrimenti in romaní, esaminati secondo diversi criteri. La variazione individuale è inoltre esplorata in dettaglio. Il volume fa dunque luce su una situazione di contatto poco studiata, caratterizzata da bilinguismo asimmetrico e non istituzionalizzato, e colloca il sinto lombardo in una più ampia prospettiva comparativa.

MARCO FORLANO ha conseguito il dottorato di ricerca in Scienze Linguistiche presso l'Università di Bergamo nel 2025 (XXXVII ciclo). Attualmente è assegnista di ricerca in Linguistica presso l'Università di Pavia. I suoi principali interessi di ricerca includono il contatto e la variazione linguistica, la linguistica romaní, il multilinguismo urbano e la linguistica educativa.



Marco Forlano

CONTATTO E VARIAZIONE IN ROMANÍ

88

Marco Forlano

CONTATTO E VARIAZIONE IN ROMANÍ Il caso dei sinti lombardi



UNIVERSITÀ
DEGLI STUDI
DI BERGAMO

2026

Collana della Scuola di Alta Formazione Dottorale

- 88 -

Collana della Scuola di Alta Formazione Dottorale

Diretta da Paolo Cesaretti

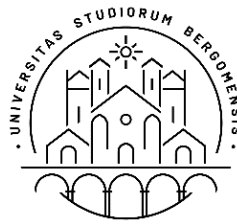
Ogni volume è sottoposto a *blind peer review*.

ISSN: 2611-9927

Sito web: <https://aisberg.unibg.it/handle/10446/130100>

Marco Forlano

CONTATTO E VARIAZIONE IN ROMANÍ
Il caso dei sinti lombardi



Università degli Studi di Bergamo

2026

Contatto e variazione in romaní. Il caso dei sinti lombardi.
/ Marco Forlano. – Bergamo :
Università degli Studi di Bergamo, 2026.
(Collana della Scuola di Alta Formazione Dottorale; 88)

ISBN: 978-88-97253-43-3

DOI: [10.13122/978-88-97253-43-3](https://doi.org/10.13122/978-88-97253-43-3)

Questo volume è rilasciato sotto licenza Creative Commons
Attribuzione - Non commerciale - Non opere derivate 4.0



© 2026 Marco Forlano

Progetto grafico: Servizi Editoriali – Università degli Studi di Bergamo
© 2018 Università degli Studi di Bergamo
via Salvecchio, 19
24129 Bergamo
Cod. Fiscale 80004350163
P. IVA 01612800167

<https://aisberg.unibg.it/handle/10446/330645>

Ringraziamenti

Questo lavoro rappresenta una versione rielaborata della mia tesi di dottorato, che non avrebbe visto la luce senza il supporto di numerose persone.

Desidero ringraziare Ilaria Fiorentini per la sua rigorosa supervisione scientifica, che prosegue da ormai sei anni, per la fiducia riposta in me sin dal principio del mio percorso e per avermi garantito il supporto necessario per portare avanti questo progetto con serenità.

Vorrei esprimere la mia gratitudine a Evangelia Adamou, che mi ha ospitato presso il centro Lacito di Parigi durante il mio terzo anno di Dottorato, spingendomi a superare i miei limiti grazie alla sua guida attenta e stimolante. Il suo libro *A Corpus-Driven Approach to Language Contact. Endangered Languages in a Comparative Perspective* (Routledge, 2016) ha costituito una grande fonte di ispirazione per le mie ricerche e ha molto influenzato la struttura di questo volume.

Un sentito ringraziamento va a Eugenio Gorla per l'interesse dimostrato verso il mio lavoro, i nostri scambi accademici, e le numerose opportunità che mi ha offerto. Ringrazio inoltre Silvia Dal Negro e Simone Ciccolone per aver discusso con me parti del mio lavoro, revisionato la tesi e fornito commenti estremamente utili.

La maggior parte delle mie conoscenze di linguistica romanica sono ancora saldamente radicate là dove tutto è partito, la *Plurilingualism Research Unit* dell'Università di Graz: in questo senso, la mia riconoscenza verso Zuzana Bordnarova, Dieter Halwachs e Chiara Tribulato è immensa. Durante questi anni, ho avuto modo di confrontarmi con altri studiosi nel campo della linguistica romanica, in particolare Andrea Scala e Nicola Duberti, che ringrazio sentitamente.

Sono grato a Elisabetta Della Vigna, Sofia Laganà e don Massimo Mistioli per avermi introdotto alla comunità sinta di Pavia, per la loro fiducia e per il supporto pratico fornito nelle fasi iniziali del lavoro. Ringrazio le lettrici della mia tesi durante il percorso dottorale, dapprima Federica Venier e poi Chiara Zanchi. Desidero inoltre ricordare Gabriele Iannaccaro, che avrebbe dovuto seguire questo lavoro come lettore. Ringrazio infine Martina Colleoni per il suo aiuto nelle analisi statistiche presentate in questo volume.

Ai sinti che mi hanno accolto, che hanno condiviso con me la loro quotidianità e che mi hanno permesso di apprendere dal loro sapere linguistico e culturale, va il mio ringraziamento più sincero. Nel corso degli anni, ho riflettuto continuamente sulla mia presenza e sul mio ruolo nella comunità, come individuo, come ricercatore e, più di recente, come educatore. Sebbene questa tensione non sia ancora risolta, non posso non riconoscere quanto profondamente questa esperienza abbia inciso sulla mia visione del mondo, rafforzando la mia coscienza civile.

Sommario

Introduzione	1
Capitolo 1. Elementi teorici.....	5
1.1 Approcci teorici al contatto linguistico.....	5
1.2 La nozione di repertorio linguistico.....	7
1.3 Temi e approcci allo studio del parlato bilingue	10
1.3.1 Approcci allo studio del <i>code-switching</i> intrasentenziale	11
1.3.1.1 L'approccio sintattico di Poplack (1980).....	11
1.3.1.2 L'approccio psicolinguistico di Myers-Scotton (1993).....	11
1.3.1.3 L'approccio tipologico di Muysken (2000).....	12
1.3.2 Il rapporto tra <i>code-switching</i> intrasentenziale e prestito nel parlato bilingue	14
1.3.3 Il rapporto tra il <i>code-switching</i> intrasentenziale e l'emergere di varietà di contatto	17
1.4 Temi e approcci allo studio del prestito lessicale.....	18
1.4.1 Integrazione.....	19
1.4.2 Gerarchie di prestabilità	21
1.4.3 Motivazioni per il prestito.....	23
1.5 Lingue minacciate, documentazione linguistica, linguistica del contatto.....	26
2.1 I rom: cenni storici e antropologici.....	29
2.2 La lingua romaní	31
2.2.1 Cenni storici e tipologici e il ruolo del contatto linguistico	31
2.2.2 Cenni dialettologici	33
2.2.3 Aspetti sociolinguistici generali	36
2.3 La lingua romaní in Italia.....	38
2.3.1 Rom e sinti in Italia	38
2.3.2 La romaní in Italia: fonti e studi	40
2.3.3 Aspetti sociolinguistici	42
Capitolo 3. Presentazione dell'area di ricerca, metodologia generale d'indagine e dati.....	45

3.1 La comunità sinta di Pavia	45
3.2 Il lavoro sul campo e la raccolta dati.....	50
3.2.1 L'osservazione partecipante	50
3.2.2 Il questionario	51
3.2.3 Le interviste.....	53
3.2.4 Aspetti etici.....	54
3.3 Il campione dei partecipanti.....	55
3.4 Trattamento dei dati di parlato.....	56
3.4.1 L'annotazione per lingua	57
3.4.2 L'annotazione per parte del discorso.....	59
3.4.3 La lemmatizzazione	59
Capitolo 4. L'analisi macrosociolinguistica	61
4.1 Le lingue del repertorio	62
4.1.1 Lingue conosciute	62
4.1.2 Rappresentazioni emiche delle lingue.....	63
4.1.3 Il sinto e l'italiano dei sinti lombardi: alcune note	66
4.2 Periodi di acquisizione di sinto e italiano	67
4.3 Autovalutazione della competenza in sinto e in italiano	68
4.4 Scelte linguistiche e variazione.....	72
4.4.1 Usi endocomunitari ed esocomunitari	72
4.4.2 Usi linguistici nei domini endocomunitari.....	73
4.4.3 Usi linguistici con parenti sinti non lombardi	76
4.4.4 Usi linguistici in altri domini: la tecnologia e i contesti religiosi	79
4.5 Atteggiamenti verso il sinto lombardo e verso il contatto linguistico.....	82
4.5.1 Atteggiamenti verso il sinto lombardo.....	82
4.5.2 Atteggiamenti verso il contatto linguistico e la nozione di 'sinto stretto'	84
4.6 Discussione	89
Capitolo 5. Schemi di parlato bilingue e convenzionalizzazione delle inserzioni italo-romanze.....	93

5.1	Composizione del corpus per lingua	93
5.2	Schemi combinatori.....	96
5.3	Distribuzione delle inserzioni per parte del discorso	99
5.4	Distribuzione dei nomi e dei verbi italo-romanzi per campi semantici	102
5.5	Regolarità.....	106
5.5.1	Parole funzione.....	107
5.5.2	Nomi e verbi	112
5.5.2.1	Nomi.....	114
5.5.2.2	Verbi.....	116
5.6	Calchi lessicali: una panoramica qualitativa	118
5.6.1	Calchi di singole parole	119
5.6.2	Calchi di parole composte o sintagmatiche	120
5.6.3	Calchi di parole funzione multimorfemiche e locuzioni avverbiali	120
5.7	Discussione	121
Capitolo 6.	Strategie di integrazione delle inserzioni italo-romanze	125
6.1	Integrazione a livello fonetico/fonologico	125
6.2	Integrazione dei nomi.....	127
6.2.1	Integrazione morfologica	127
6.2.2	La lingua dei modificatori.....	132
6.3	Integrazione dei verbi	139
6.3.1	Verbi integrati in sinto lombardo.....	140
6.3.2	Verbi non integrati in sinto lombardo	145
6.4	Discussione	149
Capitolo 7.	Variazione tra parlanti	151
7.1	Variazione nella produzione di parole italo-romanze e romaní.....	152
7.2	Variazione nella produzione di inserzioni e alternanze	155
7.3	Variazione nella produzione di nomi e verbi italo-romanzi	158
7.4	Discussione	162

8. Riflessioni conclusive	165
8.1 Integrare la prospettiva macro- e microsociolinguistica	165
8.2 Il parlato bilingue dei sinti lombardi in prospettiva tipologica	170
8.3 Sviluppi di ricerca futuri	174
8.4 Dalla documentazione linguistica, alla scuola, alla società.....	175
Indice delle figure.....	179
Indice delle tabelle.....	181
Bibliografia	183
Appendice 1: Il modulo di consenso informato	197
Appendice 2: Il questionario sociolinguistico	199

Introduzione

Obiettivi e motivazioni del presente volume

Questo volume presenta i risultati di un'indagine sul campo condotta tra i mesi di ottobre 2021 e settembre 2023 presso una comunità di sinti lombardi residente nella città italiana di Pavia. L'attenzione è principalmente rivolta verso aspetti di contatto linguistico, considerati sia a livello 'macrosociolinguistico', sia a livello 'microsociolinguistico'.

Spesso oggetto di rappresentazioni negative nel discorso pubblico e privi di un riconoscimento ufficiale della propria specificità linguistico-culturale, i sinti italiani rappresentano una comunità di antico insediamento nella Penisola. Documentati almeno a partire dal XVII-XVIII secolo (Piasere, 1992), essi risiedono prevalentemente nell'Italia Settentrionale, dove si articolano ulteriormente in sottogruppi su base regionale. In questo quadro, i sinti lombardi rappresentano il sottogruppo stanziato in Lombardia, con propaggini anche nel Piemonte orientale, Emilia e Veneto occidentale (Soravia, 1977; Scala, 2012).

Dal punto di vista sociolinguistico, i sinti lombardi sono oggi bilingui in sinto lombardo (romaní) e italiano, ma tale configurazione rappresenta un'innovazione piuttosto recente nel repertorio linguistico della comunità, storicamente bilingue in sinto lombardo e nei dialetti italo-romanzi parlati nella zona (in particolare il lombardo). La competenza attiva dei sinti lombardi nei dialetti italo-romanzi co-territoriali è riportata oggi come molto ridotta se non assente (Scala, 2012); tuttavia, ne rimangono ampie tracce nei prestiti lessicali integrati storicamente nel sinto lombardo.

Più nello specifico, il presente volume propone un'analisi delle modalità con cui l'italoromanzo (in particolar modo l'italiano) e il sinto lombardo:

- (i) si distribuiscono funzionalmente nel repertorio linguistico dei sinti lombardi;
- (ii) interagiscono nel parlato (semi)spontaneo della comunità;
- (iii) interagiscono nel parlato di singoli individui e gruppi sociali all'interno della comunità.

Muovendo gradualmente dal livello 'macro' al livello 'micro', l'analisi avrà un orientamento empirico e si servirà principalmente di questionari sociolinguistici e di un corpus di parlato, entrambi raccolti parallelamente a una lunga osservazione partecipante. Benché l'approccio adottato per lo studio dei fenomeni di contatto sarà soprattutto sincronico, in quanto basato su dati interazionali, la profondità cronologica e la natura stratificata del contatto tra romaní e italo-romanzo renderà anche necessario distinguere tra fenomeni riconducibili al bilinguismo corrente dei parlanti e fenomeni che

recano le tracce di un contatto ormai cristallizzato. L'analisi si proporrà inoltre di cogliere le relazioni tra queste due dimensioni, concepite non già come categorie discrete, ma come poli di un continuum. La presente ricerca nasce da due motivazioni principali. In primo luogo, nell'ambito della sociolinguistica delle minoranze in Italia (Fiorentini, 2022), le comunità romaní (sia di antico insediamento sia di insediamento più recente) costituiscono tuttora un ambito scarsamente indagato, nonostante proprio negli ultimi anni diverse Università italiane abbiano avviato specifiche attività di documentazione¹.

In secondo luogo, gli studi dedicati alle comunità alloglotte di antico insediamento in Italia si concentrano prevalentemente sulla descrizione delle strutture linguistiche tradizionali, nonostante i parlanti che mantengono una competenza attiva nella lingua d'origine siano oggi pressoché tutti bilingui con l'italiano. Il rischio che ne consegue è trattare queste lingue come oggetti isolati, piuttosto che come parte integrante di repertori plurilingui dinamici (Dal Negro, 2021). In tali lavori, come osservato da Dal Negro (2009), l'attenzione al contatto linguistico è spesso limitata a inventari di prestiti lessicali, con assenza o scarsissima presenza di analisi quantitative basate su dati interazionali (ma cfr. Dal Negro, 2005a, 2005b, 2006, 2009, 2011a, 2024; Perta, 2020; Perta & Tramutoli, 2020 per casi di studio su singole varietà). D'altra parte, nella linguistica romaní, il contatto linguistico è un tema centralissimo (Matras & Adamou, 2020), ma l'attenzione è spesso rivolta agli esiti diacronici del contatto, in una prospettiva tipologica e con metodologie qualitative. Anche in questo contesto, restano dunque ancora pochi i lavori quantitativi basati su corpora di parlato, al punto che, in un recente contributo incentrato su romaní e linguistica del contatto, Matras & Adamou (2020: 348) individuano nell'assenza di "sociolinguistic e discourse-based studies" un'urgente lacuna da colmare (ma cfr. Adamou & Granqvist, 2015; Adamou, 2015, 2016; Mirić & Ćirković, 2022 per contributi recenti in questo senso).

Approcci al contatto linguistico dal taglio sincronico e quantitativo, come quello adottato in questo volume, sono in realtà ben radicati nella linguistica del contatto almeno dagli anni Novanta del secolo scorso. Questi, tuttavia, tendono a riguardare comunità bilingui dal profilo sociolinguistico molto diverso rispetto a quello dei sinti lombardi, come le comunità di recente immigrazione (per esempio la comunità turcofona nei Paesi Bassi; Backus, 1996) o comunità autoctone di ampia diffusione regionale (per esempio la comunità francofona in Québec; Poplack et al., 1988). Tali comunità, pur nella diversità di ciascun contesto, sono relativamente ampie e stabili e presentano spesso reti sociali eterogenee, all'interno delle quali i singoli individui possono esibire profili bilingui anche molto

¹ Si considerino, in particolare, i lavori condotti dall'Unità di Ricerca dell'Università degli Studi di Milano, coordinata da Andrea Scala, all'interno del PRIN 2022 "Contact-induced language change: perspectives from the minority languages in the Italian linguistic space" (PRIN 2022, prot. 20224RFY93, finanziato dal Ministero dell'Università e della Ricerca).

differenziati. D'altra parte, i sinti lombardi di Pavia costituiscono una comunità molto piccola e omogenea, in cui i singoli individui sono fortemente interconnessi e mostrano profili linguistici relativamente simili. Inoltre, i contesti bilingui analizzati secondo prospettive sincroniche e quantitative tendono spesso a riguardare due lingue di ampia diffusione e standardizzate (come il francese e l'inglese, il turco e l'olandese o il russo e il tedesco), mentre il caso dei sinti lombardi fornisce un esempio di situazione di contatto fortemente asimmetrica, in cui una lingua a tradizione orale e priva di una varietà standard di riferimento è in contatto da secoli con una lingua dominante. Alla luce di questo scenario, il presente volume si propone di contribuire alla letteratura esistente in due direzioni principali: da un lato, offrendo una documentazione sociolinguistica dettagliata di una comunità minoritaria finora scarsamente indagata; dall'altro, contribuendo al dibattito accademico sul parlato bilingue in contesti di piccole lingue minoritarie, ponendo anche il sinto lombardo in prospettiva comparativa (cfr. Adamou, 2016).

Struttura del volume

Il presente volume è strutturato come segue. Il Capitolo 1 presenta alcuni elementi teorici centrali per l'analisi. Si introducono dunque la linguistica del contatto e la nozione di repertorio linguistico, per poi passare in rassegna i principali approcci allo studio del parlato bilingue e del prestito diacronico. Il capitolo si conclude con una discussione della nozione di "lingua minacciata", cercando di stabilire intersezioni e punti di divergenza tra la documentazione linguistica (il campo di studi privilegiato delle lingue minoritarie e minacciate) e la linguistica del contatto.

Il Capitolo 2 propone una panoramica relativa alla lingua romaní e ai suoi parlanti, toccando vari aspetti come le origini storiche delle comunità rom e sinte, nonché le caratteristiche storico-tipologiche e sociolinguistiche della romaní. Nella seconda parte di questo Capitolo viene inoltre proposto un approfondimento relativo alla situazione della lingua romaní in Italia.

Il Capitolo 3 presenta l'area di indagine, proponendo una panoramica etnografica della comunità sinta di Pavia. Successivamente, si focalizza sulla metodologia di indagine, presentando gli strumenti utilizzati (in particolare osservazione partecipante, questionari sociolinguistici e interviste semi-strutturate), il campione dei partecipanti e le modalità di annotazione del corpus di parlato oggetto dell'analisi linguistica.

Il Capitolo 4 propone una descrizione della situazione macrosociolinguistica della comunità dei sinti lombardi di Pavia, prevalentemente sulla base dell'analisi dei questionari sociolinguistici. Particolare attenzione viene rivolta alle lingue che i parlanti dichiarano di usare nella vita quotidiana e ai nomi che attribuiscono loro, alle modalità di acquisizione di queste lingue, alla competenza autodichiarata in sinto lombardo e italiano, nonché agli usi linguistici al variare dei domini comunicativi.

Successivamente, si propone un breve approfondimento relativo agli atteggiamenti dei sinti lombardi verso il sinto lombardo e verso il ricorso a pratiche discorsive mistilingui.

Il Capitolo 5 entra nel vivo dell'analisi linguistica, occupandosi di quantificare l'incidenza del materiale italoromanzo nel parlato in sinto lombardo e di distinguere diversi schemi combinatori, dall'inserzione di singoli elementi italoromanzi in discorsi in sinto lombardo ad alternanze più estese verso l'italiano (cfr. Muysken, 2000). Ci si concentra poi sulle inserzioni italoromanze, lo schema quantitativamente dominante, che vengono analizzate dal punto di vista della loro distribuzione per parti del discorso e per campi semantici, nonché per quanto riguarda la loro regolarità nel parlato (in termini di frequenza nel corpus e diffusione tra parlanti). Il Capitolo si conclude con un'analisi preliminare dei casi di calco lessicale totale o parziale (Backus & Dorleijn, 2009), in cui, cioè, un modello lessicale italoromanzo non viene importato come tale (ossia nella sua sostanza fonomorfologica), ma tradotto (del tutto o in parte) con materiale nativo.

Il Capitolo 6 è specificamente dedicato alle strategie di integrazione linguistica delle inserzioni italoromanze in sinto lombardo, un criterio considerato centrale nella letteratura sul contatto linguistico per distinguere tra prestito e casi di *code-switching* nel caso di singole inserzioni alloglotte (Poplack, 2018). Partendo da studi già esistenti (Meli, 2023), ci si concentra prevalentemente sul livello morfologico e sulle categorie lessicali dei nomi e dei verbi. Ne deriva l'idea che l'integrazione non sia sempre un criterio utile a distinguere tra prestito e *code-switching* e che i due fenomeni vadano meglio visti come posti lungo un continuum piuttosto che come categorie discrete.

Infine, il Capitolo 7 propone un'analisi dettagliata della variazione degli schemi di parlato bilingue, presentati come tendenze generali nei capitoli precedenti, attraverso singoli individui e gruppi sociali nella comunità, sulla base di criteri quali genere, età e insediamento di residenza. Viene analizzata la variazione individuale nel ricorso a parole italoromanze, a inserzioni e alternanze, nonché a singoli nomi e verbi italoromanzi (Adamou, 2016). Si cerca di cogliere anche la relazione tra i diversi *pattern* (per esempio, tasso di inserzioni e alternanze) nel parlato dei singoli individui. L'analisi è di taglio quantitativo e si serve dell'ausilio di metodi statistici.

Il volume si conclude, al Capitolo 8, con alcune riflessioni conclusive più generali, delineando i possibili sviluppi della ricerca.

Capitolo 1. Elementi teorici

L'obiettivo di questo capitolo è fornire alcuni strumenti teorici funzionali all'analisi presentata in questo volume. Nello specifico, il § 1.1 introduce la linguistica del contatto e i principali approcci teorici allo studio del contatto linguistico, mentre il § 1.2 discute la nozione di repertorio linguistico. I §§ 1.3 e 1.4 si focalizzano, rispettivamente, sui principali approcci allo studio del parlato bilingue in prospettiva sincronica e sui processi di prestito in diacronia. Infine, il § 1.5 si focalizza sulle lingue minacciate, presentando i principali approcci al loro studio e il rapporto con la linguistica del contatto.

1.1 Approcci teorici al contatto linguistico

Sebbene il contatto tra lingue sia stato oggetto di attenzione nell'ambito della linguistica storica sin dal XIX secolo, la linguistica del contatto come campo di ricerca autonomo si è affermata soltanto relativamente di recente, nella seconda metà del Novecento. Un ruolo fondativo in questo processo è tipicamente attribuito al lavoro di Uriel Weinreich (1953), che ha individuato nel parlante multilingue il *locus* del contatto linguistico. A partire da questo contributo, e attraverso decenni di ricerca successivi, la linguistica del contatto è oggi considerata “an established field of investigation in its own right, rather than just the application of general linguistic methods to data sets that represent contact” (Adamou & Matras, 2021: 1).

Nonostante l'enfasi posta da Weinreich (1953) sul singolo individuo multilingue, la successiva letteratura scientifica nella linguistica del contatto ha prevalentemente concentrato l'attenzione sulle attività e gli scenari sociali più ampi nei quali tali individui sono inseriti. In questa prospettiva, numerosi studi hanno cercato di individuare i fattori “esterni” in grado di influenzare la quantità e la qualità dei fenomeni di contatto linguistico prodotti dai singoli individui all'interno di comunità bi- o multilingui (cfr. p.es. Thomason & Kaufman, 1988; Thomason, 2001; Muysken, 2010; Winford, 2013; Adamou, 2016; Dal Negro, 2024).

Un lavoro di riferimento in questo senso è Thomason & Kaufman (1988), le quali sottolineano come la storia sociolinguistica dei parlanti costituisca il predittore più importante degli esiti del contatto linguistico. In questo quadro, le autrici suggeriscono l'esistenza di un continuum di situazioni di contatto che oscilla da “casual contact” a “very strong cultural pressure”, passando per una serie di stadi intermedi. In una situazione di “casual contact”, si osserveranno al più fenomeni di prestito lessicale (*lexical borrowing*), mentre in situazioni di “very strong cultural pressure”, si potranno avere fenomeni di “heavy structural borrowing” in grado di modificare in maniera sostanziale le strutture

grammaticali della lingua socialmente subordinata. Analogamente, Winford (2013) propone che nelle prime fasi di contatto tra due lingue si osserveranno fenomeni di prestito lessicale, mentre una maggiore intensità del contatto promuoverebbe “increasing structural influence from the dominant language on the subordinate ancestral language” (*ibid.*: 382).

Parallelamente, diversi autori hanno ricentrato l’attenzione sui processi cognitivi a livello del singolo individuo bilingue nel determinare gli esiti del contatto linguistico. In questa prospettiva, i fattori esterni non vengono esclusi, ma assumono un ruolo secondario: la comunità bilingue, con le sue specifiche abitudini interazionali e i suoi atteggiamenti linguistici, consente che le innovazioni linguistiche generate dai processi cognitivi individuali vengano progressivamente accolte e stabilizzate nel lungo termine.

In questo quadro, sulla base di modelli di produzione psicolinguistica, Myers-Scotton (1993, 2002) enfatizza il ruolo di una “lingua matrice” (*matrix language*), cioè la lingua più attiva a livello della produzione linguistica, nel determinare la struttura grammaticale degli enunciati bilingui. Da questa prospettiva, gli esiti del mutamento linguistico indotto dal contatto (quali ad esempio i processi di convergenza sintattica, l’emergere di lingue miste, i processi di *language attrition* e *loss*) deriverebbero da un progressivo cambio diacronico (*turnover*) della lingua matrice.

D’altra parte, secondo Matras (2020), la pressione cognitiva sperimentata dai parlanti durante la produzione discorsiva in modalità bilingue favorirebbe una tendenza sistematica alla semplificazione del repertorio individuale. Essa si manifesterebbe, da un lato, nella selezione stabile di elementi lessicali dalla lingua “pragmaticamente dominante” (come nel caso degli operatori discorsivi; cfr. anche Matras, 1998a) e, dall’altro, in una convergenza delle strutture morfosintattiche. Infine, Zenner et al. (2019) argomentano in favore di una “linguistica del contatto cognitiva” attenta ai processi di categorizzazione mentale dei parlanti bi- e multilingui quali fattori propulsivi della variazione e del mutamento linguistico indotti dal contatto. Gli autori suggeriscono, in particolare, l’importanza di fattori legati all’uso (quali la salienza o la frequenza) nel determinare diversi gradi di *entrenchment* delle unità linguistiche nel lessico mentale dei parlanti bilingui, e dunque una loro maggiore o minore propensione al mutamento linguistico indotto dal contatto (cfr. anche Backus, 2015).

Accanto a questi filoni di ricerca, non mancano approcci che enfatizzano il ruolo dei fattori tipologico-strutturali nel determinare gli esiti del contatto linguistico. Per esempio, numerosi studi relativi alle cosiddette “gerarchie del prestito” (cfr. § 1.4.2) hanno tentato di motivare la maggiore suscettibilità al contatto di alcune unità linguistiche in virtù della loro autonomia morfosintattica, della loro trasparenza semantico-strutturale o sulla base dell’equivalenza strutturale tra le lingue in contatto (cfr. p.es. Moravcsik, 1978; Treffers-Daller, 1994, 1999; Muysken, 2000; Field, 2002; Matras, 2007). Approcci di tipo formale, ispirati al programma generativo, costituiscono inoltre il fondamento

teorico di numerosi contributi recenti allo studio delle cosiddette *heritage languages* (cfr. p.es. Polinsky, 2018).

Infine, nel contesto della sociolinguistica critica, sono emersi negli ultimi decenni approcci che mettono in discussione la concezione tradizionale delle lingue come sistemi discreti. In questo quadro, si fa strada l'idea che i parlanti bi- e plurilingui dispongano di un repertorio integrato costituito da morfemi e costruzioni che giungono a essere associate a specifiche sotto-unità del repertorio (cosiddette “lingua A”, “lingua B”, ecc.) sulla base delle esperienze individuali di socializzazione (Blommaert, 2010, 2013; Blommaert & Baclus, 2013; Matras, 2020). In questo quadro, nozioni tradizionali quali prestito o *code-switching* vengono reinterpretate come manifestazioni dell'impiego flessibile di risorse appartenenti a un repertorio integrato (cfr. anche § 1.2). Su queste basi si sono sviluppate proposte teoriche quali ad esempio il *translanguaging* (Wei, 2018), secondo cui i parlanti plurilingui mobilitano in maniera dinamica e strategica l'insieme delle proprie risorse linguistiche per perseguire specifici obiettivi comunicativi.

Al termine di questa panoramica, è opportuno osservare come gli aspetti cognitivi, sociali e strutturali del contatto linguistico non siano sempre stati concepiti come compartimenti stagni, ma spesso affrontati in modo integrato nella letteratura specialistica. Tale impostazione unitaria appare particolarmente evidente negli approcci più recenti alla linguistica del contatto, che sottolineano la necessità di un approccio integrato che renda conto tanto dei processi sincronici, legati al comportamento linguistico del singolo individuo, quanto degli esiti diacronici, osservabili a livello comunitario o persino su scala areale (cfr., tra gli altri, Adamou & Matras, 2021; Adamou et al., 2023; van Gijn et al., 2023). Questo approccio integrato sarà seguito anche nel presente volume.

1.2 La nozione di repertorio linguistico

La nozione di repertorio linguistico occupa una posizione centrale nella ricerca sociolinguistica. Introdotta negli anni Settanta dal sociolinguista John Gumperz, essa designa l'insieme delle risorse linguistiche (siano esse lingue, dialetti, stili discorsivi) disponibili ai membri di una data comunità linguistica (cfr. Gumperz, 1972, 1982). Sin dalla sua concezione iniziale, tale nozione appare dunque strettamente legata a quella di comunità linguistica, intesa come un insieme di individui che, condividendo le medesime risorse linguistiche e le norme sociali sul loro utilizzo, è in grado di comunicare in modo efficace.

Sulla scia di Gumperz, nella letteratura successiva la nozione di repertorio linguistico è stata soprattutto utilizzata secondo una prospettiva macrosociolinguistica, con l'obiettivo di modellizzare i *pattern* interazionali di comunità bi- e plurilingui sulla base di parametri quali i domini e le situazioni comunicative. In tal senso, una distinzione comune è tra acroletti (le lingue utilizzate nei domini ‘alti’,

ovvero ufficiali, formali e scritti) e basiletti (le lingue utilizzate nei domini ‘bassi’, ovvero non ufficiali e informali, come la comunicazione orale nell’ambito familiare o nella cerchia ristretta di amici). Nel contesto della sociolinguistica italiana, questo modello ha contribuito allo sviluppo di una tipologia di repertori linguistici plurilingui articolata in quattro principali configurazioni (cfr. p.es. Dell’Aquila & Iannàccaro, 2004):

- a) diglossia: tipologia di repertorio linguistico comunitario nel quale coesistono una lingua A, impiegata esclusivamente in domini ufficiali e formali, generalmente scritti, e una lingua B, riservata ai domini informali, generalmente orali. Le due lingue risultano dunque nettamente separate negli usi linguistici dei parlanti e presentano poche occasioni di sovrapposizioni funzionali. Questa tipologia di repertorio linguistico prosegue direttamente la nozione di “diglossia” già introdotta da Ferguson (1959) in rapporto ai modelli di utilizzo dell’arabo standard e dei dialetti locali nel mondo arabofono; un altro esempio frequentemente citato a questo proposito è la relazione tra *dimotikí* e *katharévousa* nella Grecia del XIX e XX secolo;
- b) dilalia: tipologia di repertorio linguistico comunitario nel quale le lingue A (alta, di prestigio) e B (più bassa) non sono mantenute completamente separate negli usi dei parlanti, ma tendono a sovrapporsi soprattutto nei domini informali, dando luogo a una distribuzione più flessibile rispetto alla diglossia. Questa situazione si individua tipicamente per alcuni contesti di area italiana in cui ancora convivono i dialetti italo-romanzi e l’italiano standard, oppure una lingua di minoranza non italo-romanza e l’italiano standard;
- c) bilinguismo sociale: tipologia di repertorio linguistico comunitario opposta alla diglossia, nel quale le lingue A e B possono essere utilizzate interscambiabilmente in ogni contesto comunicativo, senza alcun tipo di gerarchia funzionale. Questo tipo di repertorio linguistico è in realtà considerato molto minoritario nel mondo, in quanto negli usi di una comunità le lingue conoscono di norma un certo grado di differenziazione funzionale;
- d) diacroletta: tipologia di repertorio linguistico comunitario in cui, dati due codici A e B, un solo dei due è utilizzato nei domini informali, mentre entrambi possono essere utilizzati negli usi alti e ufficiali. Questa è una situazione che si può riscontrare a seguito di politiche linguistiche che promuovono l’uso di un codice nei domini alti, come l’amministrazione o la scuola, senza che esso sia necessariamente usato anche nei contesti più informali; un esempio classico in questo senso è il rapporto tra inglese e gaelico in Irlanda.

All’interno di questo quadro, con particolare riferimento al contesto italiano, Dal Negro e Iannàccaro (2003) propongono una tipologia di repertori linguistici plurilingui a grana più fine, ricavabile dall’incrocio di svariati parametri macrosociolinguistici. Tra questi figurano, ad esempio: il numero

di varietà utilizzate; le varietà impiegate nei domini alti e bassi; le varietà percepite dai parlanti come “lingue” o come “dialetti”; le varietà in ascesa o in declino in termini di numero di parlanti e/o di status; le varietà utilizzate come codici endocomunitari ed esocomunitari; le varietà che costituiscono “codici non autonomi”, ossia varietà impiegate in un regime di costante parlato bilingue con lingue dominanti.

In anni recenti, la nozione di repertorio linguistico è tornata alla ribalta nel contesto della sociolinguistica critica, dove è stata oggetto di una profonda rivalutazione anche alla luce del paradigma della *superdiversity* (Vertovec, 2007), che enfatizza la crescente complessità e mobilità delle traiettorie linguistiche individuali. In particolare, Blommaert e Backus (2013) mettono in discussione due nuclei centrali della definizione classica di repertorio linguistico: da un lato, l’idea di comunità linguistica come entità omogenea e stabile, alla quale corrisponderebbe un insieme fisso e invariabile di risorse linguistiche; dall’altro, l’assunzione secondo cui disporre di un repertorio linguistico implicherebbe una competenza piena e uniforme delle risorse che lo compongono.

In alternativa, Blommaert & Backus (2013) propongono di spostare l’attenzione dal livello della comunità a quello dell’individuo, considerato come soggetto attivo che partecipa a una varietà di ambienti sociali, ciascuno caratterizzato dai propri *loci* normativi (cfr. anche Jørgensen, 2008). Inoltre, gli autori intendono la competenza linguistica come un processo dinamico, che non implica di padroneggiare pienamente tutti i registri e le strutture grammaticali di una lingua, ma include anche conoscenze parziali, non focalizzate e temporanee. Ne deriva una riconcettualizzazione del repertorio linguistico come un fatto individuale, biograficamente strutturato e soggetto a variazione diacronica, coincidente, di fatto, con la nozione di “autobiografia linguistica”: “repertoires are individual, biographically organized complexes of resources, and they follow the rhythms of actual human lives” (Blommaert & Backus, 2013: 15).

In sintesi, l’evoluzione diacronica del concetto di repertorio linguistico mostra un passaggio da approcci centrati sulla comunità e sulla distribuzione funzionale delle lingue intese come entità relativamente separate e date *a priori*, a prospettive che privilegiano la dimensione individuale, la non uniformità della competenza e il carattere dinamico delle risorse linguistiche. Nel presente volume, l’analisi si collocherà prevalentemente a livello della comunità dei sinti lombardi, intesa dunque in un’accezione piuttosto tradizionale, anche in virtù della sua natura relativamente omogenea (cfr. Capitolo 3). Al contempo, nel corso dell’analisi, verrà prestata particolare attenzione ad aspetti quali la variazione individuale, il rapporto tra singolo individuo e comunità e le competenze parziali.

1.3 Temi e approcci allo studio del parlato bilingue

Uno degli aspetti più interessanti del comportamento linguistico degli individui bi- e multilingui è la possibilità di comunicare in una “modalità bilingue” (Grosjean, 2001), ovvero attingendo liberamente a tutte le risorse del proprio repertorio linguistico in una stessa situazione comunicativa. I riflessi linguistici di questa modalità comunicativa sono generalmente noti come *parlato bilingue* o, al di fuori del contesto italiano, come *language mixing* (cfr. p.es. Backus, 1996; Poplack, 2018). Con questo termine, si intende tipicamente una produzione linguistica (sia essa un testo, un discorso, un enunciato o una frase) in cui coesistono elementi lessicali e grammaticali di lingue diverse. In quanto termine ombrello, *parlato bilingue* può comprendere un’ampia gamma di manifestazioni.

In particolare, la combinazione di lingue diverse all’interno di un’unità come una frase o un enunciato è spesso definita *code-switching intrasentenziale* (o *code-mixing*), mentre la combinazione di lingue diverse a cavallo di due frasi o enunciati è spesso definita *code-switching intersentenziale* (o semplicemente *code-switching*). Mentre lo studio del *code-switching* intrasentenziale è sovente stato affrontato secondo una prospettiva sintattico-grammaticale, lo studio del *code-switching* intersentenziale è stato più spesso oggetto di approcci pragmatico-discorsivi, volti a motivare il cambio di lingua sulla base di fattori quali la costellazione degli interlocutori o l’organizzazione del discorso. Una nozione simile a *code-switching intrasentenziale* è quella di prestito, sebbene i suoi criteri definitori siano spesso dibattuti in letteratura: per alcuni, la cifra distintiva del prestito rispetto al *code-switching* risiederebbe nell’integrazione linguistica, mentre per altri nel livello di convenzionalizzazione. Un ulteriore tema di rilievo nello studio del parlato bilingue riguarda la distinzione tra processi sincronici di *code-switching* e l’emergere di varietà di contatto, come i *fused lect* e le lingue miste, risultanti dalla convenzionalizzazione degli schemi di *code-switching* nel parlato di una data comunità.

Poiché, come vedremo, gli schemi combinatori nel parlato bilingue dei sinti lombardi riguardano prevalentemente il livello intrasentenziale, in questa sezione ci concentriamo più in dettaglio sugli approcci rilevanti per questa dimensione. Il § 1.3.1 illustra dunque gli approcci più noti allo studio delle caratteristiche grammaticali del *code-switching* intrasentenziale, in particolare quelli di Poplack (1980), Myers-Scotton (1993, 2002) e Muysken (2000). Successivamente, il § 1.3.2 illustra il dibattito relativo alla distinzione tra prestito e *code-switching* nel parlato bilingue. Infine, il § 1.3.3 è dedicato al rapporto tra *code-switching* intrasentenziale e l’emergere di varietà di contatto, come i *fused lect* e le lingue miste.

1.3.1 Approcci allo studio del *code-switching* intrasentenziale

1.3.1.1 L'approccio sintattico di Poplack (1980)

I primi contributi della sociolinguista canadese Shana Poplack alla teoria del *code-switching* intrasentenziale sono volti a formulare vincoli sintattici (o *constraints*) entro i quali l'interazione tra due lingue diverse risulterebbe fortemente sfavorita. In particolare, a partire dai suoi primi lavori sul parlato bilingue inglese-spagnolo di una comunità portoricana a New York, Poplack sviluppa una visione del *code-switching* in termini di alternanza tra costituenti e frasi ben formate delle lingue a contatto. In linea con questa visione, l'autrice teorizza due *constraints* divenuti in seguito molto celebri, il *Free Morpheme Constraint* e l'*Equivalence Constraint*.

In sintesi, il *Free Morpheme Constraint* predice che il *code-switching* non può occorrere tra due morfemi legati, come nel caso della forma fittizia **eatiendo*, contenente una radice lessicale inglese e morfologia flessiva spagnola. L'*Equivalence Constraint* predice invece che il *code-switching* tende a non verificarsi in punti in cui una giustapposizione tra le lingue A e B violerebbe le strutture sintattiche di una delle due o entrambe. Per esempio, nel parlato bilingue inglese-spagnolo, un cambio di lingua tra nome e aggettivo descrittivo risulterebbe fortemente improbabile, in quanto in inglese gli aggettivi sono preposti al nome, mentre in spagnolo sono prevalentemente posposti; la presenza di un aggettivo preposto sarebbe dunque agrammaticale in spagnolo, mentre la presenza di un aggettivo posposto sarebbe agrammaticale in inglese (Poplack, 1980). In sintesi, “simultaneous operation of the free morpheme and the equivalence constraints permits only code-switched utterances which, when translated into either language, are grammatical by both L1 and L2 standards” (Poplack, 1980: 588).

L'approccio di Poplack (1980) è stato ampiamente messo in discussione a seguito dell'individuazione di numerosi controesempi ai suoi *constraints* (cfr. p.es. Gardner-Chloros, 2009: 96-97 per una panoramica). Va tuttavia osservato che, nella successiva produzione scientifica, l'attenzione di Poplack e dei suoi collaboratori si sposta gradualmente dalla formulazione di *constraints* sintattici al *code-switching* intrasentenziale verso la distinzione empirica tra diversi *pattern* di parlato bilingue, cioè il *code-switching* e il prestito, un aspetto che sarà discusso nel dettaglio al § 1.3.2.

1.3.1.2 L'approccio psicolinguistico di Myers-Scotton (1993)

Nel contesto del suo *Matrix Language Frame Model* (Myers-Scotton, 1993), poi divenuto *4-M Model* in successive elaborazioni (a partire da Myers-Scotton & Jake, 2000), Myers-Scotton teorizza un approccio al *code-switching* intrasentenziale basato sulla possibilità di distinguere tra una lingua “matrice” (*Matrix Language*) e una “lingua incassata” (*Embedded Language*). Sebbene quella di lingua matrice sia una nozione primariamente psicolinguistica, indicando la lingua maggiormente

attiva a livello della produzione discorsiva, essa può essere definita empiricamente come la lingua che fornisce la struttura grammaticale complessiva di una frase. Come tale, la lingua matrice determina l'ordine superficiale dei costituenti e la realizzazione di alcuni tipi di morfemi grammaticali, in particolar modo gli *outsider late system morphemes*, responsabili delle relazioni grammaticali extrasintagmatiche, come i morfemi di caso o di tempo-aspetto (Myers-Scotton & Jake, 2016).

In questo quadro, la “lingua incassata” può emergere attraverso due modalità principali: (1) a livello di singoli morfemi contenuto integrati nella struttura grammaticale della lingua matrice (cosiddetti *Embedded Language + Matrix Language Constituents*, caso prototipico); (2) a livello di *Embedded Language Islands*, ovvero sequenze di più morfemi ben formate nella lingua incassata. Secondo Myers-Scotton (1993), queste ultime, più rare, emergerebbero o in presenza di incongruenze strutturali tra le due lingue, per aggirare le quali l'attivazione della lingua matrice viene momentaneamente inibita a vantaggio della lingua incassata, o in presenza di sequenze formulaiche della lingua incassata, che possono essere trasferite in blocco, alla stregua di singole unità lessicali (cfr. anche Backus, 2003).

Nel quadro teorico del *Matrix Language Frame Model*, in ogni frase bilingue sarebbe dunque possibile individuare una lingua matrice e una lingua incassata. Tuttavia, in letteratura questa idea è controversa. Per esempio, sulla base di esempi da diverse coppie di lingue, Auer & Muhamedowa (2005) mostrano come, in una frase bilingue, la lingua matrice possa in parte influenzare la realizzazione grammaticale dei costituenti nella lingua incassata; di conseguenza, i confini tra le due andrebbero intesi come più continui che discreti. Inoltre, in alcuni esiti particolarmente rari del contatto linguistico, noti come *noun-verb mixed languages* (cfr. § 1.3.3), è stato osservato che i nomi e i verbi provengono da lingue diverse mantenendo ciascuno la morfologia originaria (rispettivamente di caso e di tempo-aspetto), dando dunque luogo a uno *split* grammaticale che suggerirebbe la presenza simultanea di due lingue matrici.

In successivi lavori, Myers-Scotton (2002) interpreta questi casi come esempi di “*code-switching* composito”, riconducendoli a situazioni di *language shift* in cui si assiste a un progressivo capovolgimento (*turnover*) della lingua matrice: la precedente lingua matrice si avvia a diventare la nuova lingua incassata, mentre quella precedentemente incassata viene progressivamente promossa a lingua matrice, passando per una fase intermedia di grammatica ‘mista’.

1.3.1.3 L'approccio tipologico di Muysken (2000)

Verso il finire degli anni Novanta, Muysken (2000) perviene a una sintesi dei principali approcci strutturali al *code-switching* intrasentenziale (da lui definito *code-mixing*) proposti fino ad allora. Nel

suo modello, noto anche come “modello triangolare” (Muysken, 2000), l'autore individua tre principali tipi di *code-mixing*: *insertion* (inserzione), *alternation* (alternanza) e *congruent lexicalization* (lessicalizzazione congruente). Questi tre tipi di parlato bilingue possono coesistere all'interno di una stessa comunità linguistica, sebbene una strategia tenda generalmente a prevalere sulle altre in termini quantitativi. Il contesto sociolinguistico e la tipologia delle lingue in contatto possono, in una certa misura, predire quale strategia risulterà dominante in una data comunità in un determinato momento storico. Per distinguere empiricamente i vari *pattern* di *code-mixing*, Muysken (2000) propone una serie di criteri diagnostici, che riepiloghiamo brevemente di seguito.

L'*insertion* consiste nell'incorporazione di singoli elementi di una lingua B all'interno di una frase altrimenti in una lingua A. Questo tipo è compatibile con modelli che postulano un'asimmetria strutturale nel *code-switching*, con una lingua dominante a fornire le strutture grammaticali e l'altra a fornire al più elementi lessicali, come il *Matrix Language Frame Model* di Myers-Scotton (1993). Affinché si possa parlare di *insertion*, l'elemento inserito da una lingua B deve avere un qualche rapporto di dipendenza sintattica con il resto della frase nella lingua A. Si tratterà dunque spesso di singole parole contenuto (come nomi e verbi) integrate morfosintatticamente nelle strutture grammaticali della lingua ricevente, sebbene rientri in questa tipologia anche l'inserzione di interi costituenti alloglotti (cosiddette *Embedded Language Islands* nella terminologia di Myers-Scotton, 1993, o *maximal insertions*, nella successiva proposta di Auer, 2014). Da un punto di vista sociopolitico, Muysken (2000) predice che l'*insertion* si incontri soprattutto in contesti in cui le lingue in contatto sono poste in un rapporto di forte asimmetria, ad esempio tra prime generazioni di immigrati o in società postcoloniali. Tuttavia, la ricerca empirica su svariate coppie di lingue ha mostrato che l'*insertion* rappresenta il *pattern* di parlato bilingue statisticamente più frequente, indipendentemente dal contesto sociolinguistico o dalla tipologia strutturale delle lingue in contatto (cfr. Adamou, 2016; Poplack, 2018).

L'*alternation* comporta un vero e proprio passaggio da una lingua all'altra, che coinvolge sia la grammatica sia il lessico (Muysken, 2000: 5). L'*alternation* appare dunque compatibile con modelli del *code-switching*, come quello di Poplack (1980), che enfatizzano il mantenimento dell'integrità strutturale delle lingue interagenti in una frase bilingue. Nella visione di Muysken, l'*alternation* è tipicamente caratterizzata da una maggiore “lunghezza e complessità” rispetto all'*insertion* (*ibid.*: 97) e dall'autonomia sintattica del materiale alloglotto. Ne deriva l'idea che l'*alternation* coinvolga sequenze di intere frasi oppure singoli costituenti sintatticamente autonomi, quali avverbi, congiunzioni e segnali discorsivi. Tuttavia, alcuni studiosi hanno sostenuto che le inserzioni di singoli elementi sintatticamente autonomi come avverbi e segnali discorsivi andrebbero meglio interpretate come casi di *insertion* poiché, per quanto periferici a livello della frase minima, essi sono integrati a

livello dell'enunciato o del discorso (cfr. p.es. Auer, 1999; Adamou, 2016; Matras, 2020; Ciccolone & Dal Negro, 2021). Dal punto di vista sociolinguistico, secondo Muysken (2000), l'*alternation* si incontrerebbe più spesso in contesti in cui le lingue sono poste in un rapporto di tensione sociopolitica: in questi scenari, i parlanti eviterebbero di 'mescolare' le lingue a un livello così 'intimo' come nel caso dell'*insertion*, mantenendole relativamente separate nella frase.

La terza strategia, *congruent lexicalization*, combina tratti sia dell'*insertion* sia dell'*alternation*. Essa si verifica quando le due lingue presentano una struttura grammaticale sostanzialmente condivisa, che può dunque essere riempita liberamente con elementi lessicali di entrambe le lingue (Muysken, 2000: 6). Tra i tratti diagnostici di *congruent lexicalization* figurano la varietà dei morfemi commutati all'interno di un corpus bilingue e la presenza di *non-constituent switches*, ossia inserzioni che interrompono confini sintattici ben definiti, producendo, ad esempio, *switch* interni alla parola e collocazioni miste. Questo tipo di *code-mixing* sembra ricordare la nozione di "code-switching composito" proposta da Myers-Scotton (2002), in riferimento a situazioni in cui non è possibile individuare una lingua matrice univoca in una frase bilingue. Secondo Muysken, la *congruent lexicalization* si verificherebbe soprattutto tra lingue strutturalmente simili (per prossimità genetica o per convergenza) e in comunità caratterizzate da norme prescrittive poco rigide.

In un successivo lavoro, Muysken (2013) introduce alla sua tipologia un quarto stile di *code-mixing* definito *backflagging*, che si riscontrerebbe in comunità in cui è avvenuto uno *shift* dalla lingua d'origine alla lingua del nuovo paese ospitante. In questo quadro, il *backflagging* consiste nell'utilizzo di elementi della propria lingua d'origine in frasi condotte nell'altra lingua a fini identitari; gli elementi commutati sono spesso segnali discorsivi caratterizzati da una chiara "connotazione etnica" (Muysken, 2013: 713).

1.3.2 Il rapporto tra *code-switching* intrasentenziale e prestito nel parlato bilingue

Uno dei principali problemi empirici affrontati dagli studiosi del parlato bilingue riguarda la distinzione tra *code-switching* intrasentenziale e prestito, soprattutto in riferimento all'inserzione di singoli elementi in una lingua A in una frase altrimenti in una lingua B.

L'individuazione dei prestiti introdotti in una lingua in fasi precedenti di contatto risulta relativamente agevole attraverso l'applicazione del metodo storico-comparativo. Tuttavia, quando le forme alloglotte sono condivise con una lingua di contatto corrente, la distinzione tra prestito e *code-switching* diventa considerevolmente più problematica. Questa difficoltà emerge in modo particolare nei contesti di lingue minoritarie, in cui l'osservazione che la maggior parte dei parlanti è bilingue fa venire meno la soglia di riferimento più solida per distinguere in modo netto tra i due fenomeni, cioè

il parlato di individui monolingui, che produrranno unicamente prestiti entrati nella lingua in fasi precedenti di contatto.

In letteratura, si possono distinguere essenzialmente due approcci alla questione: da un lato, quelli che vedono la distinzione tra prestito e *code-switching* come netta; dall'altro, quelli che interpretano queste due manifestazioni come poste lungo un continuum legato al diverso grado di stabilizzazione nella comunità di singoli lemmi alloglotti.

Tra i sostenitori di una distinzione netta tra prestito e *code-switching* si annoverano Poplack e i suoi collaboratori, che si muovono nel quadro della sociolinguistica variazionista (cfr. Poplack & Meechan, 1998; Poplack, 2018). Per loro, la distinzione principale tra prestito e *code-switching* risiede nell'integrazione dell'elemento di una lingua B in una lingua A a livello morfosintattico (si noti che l'integrazione fonetica non viene considerata rilevante dagli autori; cfr. Poplack et al., 2020). L'idea di fondo è che, se l'elemento alloglotta appare integrato a livello morfosintattico, allora potrà essere considerato come un prestito, in quanto esibisce il medesimo comportamento linguistico del materiale nativo. Sulla base di questo criterio, l'interpretazione come prestiti si applica anche al caso dei cosiddetti *nonce borrowings*, ossia forme idiosincratiche che, pur non soddisfacendo i requisiti sociali degli *established borrowings* (cioè la frequenza e la diffusione in una comunità di parlanti), risultano indistinguibili da questi ultimi sul piano linguistico. Viceversa, un elemento morfosintatticamente non integrato rappresenterà un caso di *code-switching*, fino al caso limite di *single-word switches*, ovvero commutazioni coincidenti con una singola parola.

Confrontando corpora di parlato bilingue inglese-francese raccolti a Ottawa in diversi periodi storici, Poplack & Dion (2012) presentano evidenza diacronica del fatto che le inserzioni di singole parole tendono a essere integrate nella morfosintassi di una lingua ricevente sin dalla loro prima apparizione. Da questa osservazione deriva l'ipotesi secondo cui le inserzioni di singole parole siano generalmente trattate dai parlanti come prestiti, in opposizione alle inserzioni di più parole, le quali tendono a mantenere sempre la grammatica della lingua donatrice, venendo perciò trattate come *code-switching*. Data la sostanziale differenza strutturale tra i due meccanismi, Poplack e colleghi rifiutano l'esistenza di una connessione tra i fenomeni di prestito e i fenomeni di *code-switching*, e dunque l'idea diffusa per cui i primi originino dai secondi (Poplack & Dion, 2012; Poplack, 2018).

Un limite legato a questo approccio, ammesso da Poplack stessa (Poplack, 2018), risiede nel fatto che la possibilità di individuare chiaramente l'integrazione di un elemento alloglotta sia strettamente legata alla presenza di "siti di conflitto" tra due lingue in contatto, ovvero punti in cui le rispettive grammatiche divergono in modo evidente, ad esempio nell'ordine dei costituenti o nella realizzazione superficiale della morfologia flessiva. Ne consegue che il potere diagnostico del criterio dell'integrazione tende a essere maggiore quando le lingue coinvolte sono tipologicamente molto

distanti, come ad esempio nel caso dell'inglese e del tamil (Sankoff et al., 1990), mentre risulta assai più limitato tra lingue caratterizzate da una forte affinità strutturale (o per vicinanza genetica o per convergenza dovuta al contatto linguistico; Dal Negro, 2005) oppure da una ridotta complessità morfologica, come nel caso di lingue isolanti.

Altri approcci, al contrario, sostengono che *code-switching* e prestito si collochino lungo un continuum (cfr., tra gli altri, Myers-Scotton, 1993; Muysken, 2000; Haspelmath, 2009; Backus, 2014; Adamou, 2016; Matras, 2020). Per esempio, Myers-Scotton (1993) propone che prestito e *code-switching* siano soggetti alle medesime procedure di produzione morfosintattica, subordinate a una lingua matrice, e che la differenza tra i due fenomeni vada ricercata a livello psicolinguistico. In tal senso, mentre forme classificabili come *code-switching* appartengono al lessico mentale della sola lingua B, forme classificabili come prestiti sono condivise nel lessico mentale della lingua B (dalla quale provengono originariamente) e della lingua A. Similmente, per distinguere tra prestito e *code-switching*, Muysken (2000) propone il criterio della “listedness”, con la quale intende la memorizzazione di un elemento originario di una lingua B nel lessico mentale della lingua A. Questi secondi approcci si riconducono, più o meno esplicitamente, a una prospettiva *usage-based*, per la quale il continuum prestito-*code-switching* può essere definito sulla base del grado di *entrenchment* dell'unità linguistica alloglotta nel lessico mentale dei parlanti, misurabile empiricamente attraverso criteri sociali come la frequenza e la diffusione nel parlato della comunità (cfr. anche Capitolo 5).

In una prospettiva più ampia, Matras (2020) elabora un modello di continuum multidimensionale per distinguere tra prestito e *code-switching* basato su diversi parametri quali il grado di bilinguismo dell'individuo, la composizione interna dell'unità inserita, la funzione, l'univocità del referente, l'operazionalità grammaticale, la regolarità di occorrenza e il grado di integrazione strutturale. Attraverso l'integrazione di questi criteri, si può ricavare che:

The prototypical, least controversial kind of borrowing thus involves the regular occurrence of a structurally integrated, single lexical item that is as a default expression, often a designation for a unique referent or a grammatical marker, in a used monolingual context. The least controversial codeswitch is an alternational switch at the utterance level, produced by a bilingual consciously. (Matras, 2020: 119)

Accettando l'idea che i fenomeni di prestito e *code-switching* siano posti lungo un continuum, è necessario ammettere che “a considerable degree of ambiguity will (...) always remain in respect of the language mixing patterns of bilinguals and of any single bilingual corpus” (Matras, 2020: 119).

Infine, altri studiosi, come Backus (2014), hanno suggerito che il tentativo di distinguere tra prestito e *code-switching* nel parlato di bilingui abbia poco senso in principio. Backus (2014) osserva infatti

che i parlanti bilingui possono trattare un elemento di una lingua B in una lingua A simultaneamente come prestito e come *code-switching*, scegliendo ad esempio di attribuirgli funzioni pragmatiche in forza della sua associazione a una lingua B (una prerogativa del *code-switching*), anche qualora esso mostri un'elevata stabilità nel parlato della lingua A (una prerogativa del prestito). Inoltre, l'autore nota come ciò che può essere un prestito per un parlante possa costituire un caso di *code-switching* per un altro parlante nella stessa comunità, sulla base di fattori individuali come le attività lavorative svolte o il gruppo di pari, che determinano diversi livelli di *entrenchment* delle forme alloglotte nel lessico mentale degli individui.

1.3.3 Il rapporto tra il *code-switching* intrasentenziale e l'emergere di varietà di contatto

Negli ultimi decenni, diversi studiosi hanno portato all'attenzione il fatto che, date determinate condizioni sociolinguistiche, gli schemi di *code-switching* intrasentenziale possono convenzionalizzarsi presso una comunità bilingue (cfr., tra gli altri, Auer, 1999, 2014; McConvell & Meakins, 2005; McConvell, 2007; Gorla, 2018; Auer & Hakimov, 2021; Adamou, 2021a). In alcuni di questi casi, il risultato può essere l'emergere di una vera e propria varietà mista trasmessa come tale alle nuove generazioni, che potranno utilizzarla senza necessariamente conoscere attivamente le lingue che l'hanno originata.

Un modello volto a cogliere il rapporto dinamico tra il *code-switching* sincronico e l'emergere di varietà di contatto è quello proposto da Auer (1999, 2014), in cui l'autore distingue tra tre stadi di parlato bilingue, posti in un rapporto di implicazione diacronica: *code-switching*, *language mixing* e *fused lect*. Il *code-switching* rappresenta una giustapposizione tra lingue percepita dai parlanti come significativa a livello "locale", ossia a livello del singolo atto linguistico. In tal senso, il *code-switching* è funzionale a ottenere specifici effetti pragmatici, come la selezione dell'interlocutore o la distinzione tra registri narrativi nel discorso. Tuttavia, quando il *code-switching* è molto frequente in una comunità, può accadere che tale significato locale riduca la propria salienza. In questi casi, a risultare pragmaticamente rilevante non sarà tanto il singolo episodio di *code-switching*, quanto il ricorso a entrambe le lingue quale modalità di comunicazione a livello globale della conversazione: si parla a tal proposito di *language mixing*. Se, in principio, il *language mixing* ammette al parlante la possibilità di selezionare liberamente varianti della lingua A e della lingua B in contatto, può accadere che questa possibilità si annulli e che l'uso di una sola delle due lingue diventi obbligatorio per certi costituenti. Il risultato è un *fused lect*, una varietà mista in cui l'utilizzo di specifici costituenti in una sola delle lingue è diventato stabile, e i parlanti non hanno più scelta. A differenza del *code-switching* e del *language mixing*, che presuppongono il bilinguismo dei parlanti, un *fused lect* non presuppone necessariamente il bilinguismo dei parlanti nelle lingue che l'hanno originato. In tal

senso, in un lavoro successivo, Auer (2014) associa lo stadio del *fused lect* a un “ritorno” da parlato bilingue a parlato monolingue.

Secondo Auer (2014), i casi prototipici di *fused lect* sarebbero rappresentati dalle cosiddette lingue miste, in cui la convenzionalizzazione degli schemi di parlato bilingue è ampiamente stabilizzata, anche se, come sostiene Meakins (2013), non vi è ragione di credere che le lingue miste siano esenti dalla normale variazione che caratterizza tutte le lingue storico-naturali. Altri autori, tuttavia, suggeriscono la necessità di distinguere tra *fused lect*, inteso come lingua mista incipiente e non ancora pienamente convenzionalizzata, e lingua mista in senso stretto (Adamou, 2021a). Nella letteratura sulle lingue miste, vengono tipicamente distinti due prototipi (cfr. Mazzoli & Sippola, 2021 per una panoramica recente):

- a) lingue miste “lessico-grammatica” (anche note come lingue miste *intertwined*), nelle quali il lessico proviene in gran parte da una lingua e la grammatica dall'altra; un celebre esempio è la media llengua parlata in Ecuador, in cui il lessico proviene in gran parte dallo spagnolo e la grammatica dal quechua (Muysken, 1981);
- b) le lingue miste “nome-verbo” (*noun-verb*), nelle quali gli elementi relativi al dominio nominale e gli elementi relativi al dominio verbale sono riconducibili a due lingue distinte; un noto esempio è il michif parlato in Canada, in cui gli elementi del sistema nominale provengono in gran parte dal francese, mentre gli elementi del sistema verbale dal cree, una lingua algonchina (Bakker, 1997).

La nozione di *fused lect* è rimasta a lungo controversa in letteratura (Matras & Adamou, 2020: 333-334), ma di recente è stata oggetto di una più attenta sistematizzazione (Auer & Hakimov, 2021). Approcci contemporanei alla nozione di *fused lect* ne enfatizzano la natura graduale, operationalizzabile considerando una varietà di criteri quali ad esempio la quantità di elementi coinvolti nella fusione e il loro grado di “sedimentazione”, cioè regolarizzazione e stabilizzazione in luogo di equivalenti nativi (cfr. anche Capitolo 8). L'idea risultante è di un “continuum di fusione” che oscilla “from complete fusion, which presupposes a complete regularization and conventionalization of mixing patterns and a huge number of grammatical categories involved, to incipient fusion found when conventionalization of mixing formats occurs only to some degree and only in a few elements of the language” (Auer & Hakimov, 2021: 363).

1.4 Temi e approcci allo studio del prestito lessicale

In questa sezione, passiamo in rassegna i principali temi e approcci legati allo studio del prestito lessicale, inteso come processo diacronico e definibile nei seguenti termini:

Borrowing involves a long-term or permanent licence to lift selection constraints on the use of a word-form or structure. Rather than just serve in a limited set of contexts, the word-form or structure now becomes available in a wider set of interactional contexts, perhaps with no limitations at all. (Matras, 2020: 159)

In particolare, il § 1.4.1 discute questioni legate all'integrazione linguistica dei prestiti, già introdotte nella sezione precedente; il § 1.4.2 affronta il tema delle gerarchie del prestito, mentre il § 1.4.3 discute le motivazioni per il prestito. Dal momento che l'approccio seguito in questo volume è molto vicino a modelli di stampo emergentista e *usage-based* (Zenner et al., 2019), anche nel trattare il prestito come processo diacronico presteremo particolare attenzione ai vincoli pragmatici operanti a livello discorsivo quali fattori esplicativi, come ad esempio nel modello di Matras (2020).

1.4.1 Integrazione

Come anticipato, numerosi autori considerano l'integrazione a livello fonetico, morfologico, sintattico e semantico come un criterio definitorio del prestito: l'idea alla base è che, se un elemento alloglotto risulta integrato su tutti questi livelli, allora può essere facilmente considerato come un prestito. Heath (1984) è tra i primi a osservare che tali strategie di integrazione possono costituire meccanismi produttivi a disposizione degli individui bilingui, definiti *borrowing routines*.

Tuttavia, il criterio dell'integrazione per definire un prestito presenta alcuni problemi. Innanzitutto, è stato notato che l'integrazione spesso non avviene simultaneamente su tutti i livelli linguistici. Per esempio, Poplack et al. (2020) hanno osservato che le strategie di integrazione a livello morfosintattico seguono strategie molto stabili e condivise a livello della comunità, mentre l'integrazione a livello fonetico/fonologico mostra una maggiore variabilità individuale.

In secondo luogo, anche all'interno di uno stesso livello di analisi, si possono riscontrare esiti di integrazione variabili. Per esempio, nel dominio della morfologia grammaticale, diversi autori hanno notato che i morfemi di plurale vengono presi a prestito relativamente di frequente insieme alle loro basi nominali, differentemente dai morfemi di caso o di tempo-aspetto, che appaiono particolarmente resistenti al prestito (cfr. p.es. Myers-Scotton, 2002; Backus, 2003; Gardani, 2008; Myers-Scotton & Jake, 2016). Ciò è stato motivato in virtù della maggiore vicinanza dei morfemi di plurale alla derivazione che alla flessione e alla loro possibilità di essere memorizzati nel lessico mentale insieme alla loro testa nominale.

Un'ulteriore difficoltà riguarda il fatto che i prestiti non sempre seguono le stesse strategie di integrazione previste per il materiale nativo, ricevendo strategie di integrazione *ad hoc* o rimanendo non integrati. Un caso emblematico di integrazione *ad hoc* è quello della romaní, in cui i nomi presi a prestito dal greco o da lingue europee successive non vengono integrati nella morfologia indoaria

ereditata (cosiddetta morfologia “oikoclitica”), ma attraverso morfemi indotti dal contatto con il greco (cosiddetta morfologia “xenoclitica”); cfr. Matras, 2002; Elšik , 2000; Elšik & Matras, 2006. In origine i morfemi xenoclitici erano stati probabilmente presi a prestito insieme ai nomi di origine greca, ma successivamente sono stati astratti dalle loro radici lessicali ed estesi a qualsiasi prestito, anche da lingue di contatto successive. Il risultato è dunque una compartimentazione a livello morfologico nel sistema nominale in romaní sulla base dell’etimologia del nome. Myers-Scotton (1993) porta invece l’attenzione su alcuni prestiti chiaramente non integrati morfologicamente in una lingua ricevente, come il caso delle parole di origine latina in inglese *syllabus/syllabi*.

Infine, diverse classi di parole possono mostrare tendenze diverse all’integrazione. Tra queste, i verbi rappresentano probabilmente la categoria che tende ad andare incontro alle procedure di integrazione più vistose. Lo studio più esteso in tal senso è quello di Wohlgemuth (2009), che individua quattro principali strategie di integrazione di prestiti verbali sulla base di un campione tipologico di oltre 500 lingue, riassumibili come seguono:

- 1) *light verb strategy*: una forma nominale del verbo alloglotto è combinata con un verbo supporto della lingua ricevente, come ‘fare’, sintatticamente attivo; questo è il caso di molte varietà immigrate di turco nei paesi germanofoni, dove una forma nominale del verbo alloglotto è accomodata in discorsi in turco attraverso il verbo supporto *yap-* ‘fare’ (cfr. p.es. Backus, 1996);
- 2) *indirect insertion*: una radice verbale alloglotta funziona come verbo effettivo nella lingua ricevente attraverso l’intermediazione di un morfema derivazionale, spesso specializzato nell’integrazione dei prestiti verbali (nel qual caso si parla di *loan verb marker*). Questo è il caso del prestito latino *studieren* in tedesco, in cui la radice latina *stud-* è integrata nella flessione tedesca mediante l’intermediazione del suffisso derivazionale *-ier*, non usato per i verbi nativi.
- 3) *direct insertion*: una radice verbale alloglotta è direttamente inserita nella morfologia flessiva di una lingua ricevente senza alcuna forma di accomodamento esplicita. Questo è il caso del prestito inglese *chattare* in italiano, in cui la radice *chat-* è direttamente inserita nel paradigma flessivo in *-are*.
- 4) *paradigm transfer*: un verbo alloglotto viene trasferito nella lingua ricevente insieme alla sua morfologia flessiva, mantenendo funzioni sintattiche e significati originari; si tratta di una strategia tipologicamente rara, attestata per esempio in alcune varietà di romaní in contatto intenso con il turco (Adamou, 2010).

Per spiegare la preferenza per una strategia di integrazione rispetto a un'altra, Wohlgemuth (2009) evoca fattori di natura macrosociolinguistica, in particolare l'intensità del contatto linguistico all'interno di una comunità. In tal senso, quanto più il contatto all'interno di una comunità è intenso, tanto minore sarebbe il grado di adattamento formale richiesto al prestito verbale, secondo una gerarchia che va dal *paradigm transfer* (dove nessun adattamento è richiesto) alla *light verb strategy* (dove si riscontra il maggior grado di adattamento formale). In alternativa, Matras (2020: 198-200) interpreta tali differenze in termini del modo con cui i parlanti bilingui concettualizzano il verbo alloglotto nell'interazione: come un predicato già disponibile a tutti gli effetti (caso del *paradigm transfer*), oppure come una semplice etichetta lessicale che necessita di adattamenti grammaticali per funzionare come predicato effettivo nella lingua ricevente. Il fatto che la maggior parte dei verbi richieda comunque una qualche forma di adattamento grammaticale quando viene usata in un discorso in un'altra lingua è ricondotto da Matras (2020) alla necessità dei parlanti di ancorare la predicazione, ancorché bilingue, in una lingua matrice univoca (nel senso di Myers-Scotton, 2002). In questo senso, Matras (2020) concorda con Wohlgemuth nel ritenere che l'accettazione di un verbo con morfologia alloglotta come “ancora” della predicazione nel contesto discorsivo di un'altra lingua presupponga una forte accettazione dell'identità bilingue da parte dei parlanti. L'approccio di Matras contribuisce dunque a sottolineare l'importanza di fattori cognitivi e micro-sociolinguistici nel determinare gli esiti e i processi di integrazione dei prestiti a livello interlinguistico.

1.4.2 Gerarchie di prestabilità

Passiamo ora a trattare un ordine di problemi in parte diverso, e cioè la *borrowability*, con la quale si intende la propensione al prestito di diverse classi di parole, morfemi o strutture sintattiche, in diacronia (cfr. p.es. Moravcsik, 1978; Thomason & Kaufman, 1988; Elšik & Matras, 2006; Matras, 2007; Haspelmath & Tadmor, 2009; cfr. anche Whitney, 1881 per un primo approccio).

In letteratura, la *borrowability* è tipicamente rappresentata attraverso “scale” o “gerarchie”, per cui gli elementi posti a sinistra sono maggiormente soggetti al prestito rispetto a quelli a destra. Come sottolineato da Haspelmath (2008), queste scale possono essere interpretate in quattro sensi principali:

- 1) temporale (una lingua prende a prestito elementi a sinistra prima degli elementi posti più a destra);
- 2) implicazionale (una lingua che presenta prestiti di elementi a destra contiene anche prestiti di elementi più a sinistra);
- 3) quantitativo (una lingua prende a prestito più elementi appartenenti ai tipi a sinistra rispetto ai tipi più a destra);

- 4) probabilistico (è più probabile che siano presi a prestito elementi appartenenti ai tipi a sinistra rispetto a quelli appartenenti ai tipi a destra).

Benché ci sia ancora un certo scetticismo tra gli studiosi nell'accettare tendenze assolute relative alla *borrowability*, sono ad oggi piuttosto accettate alcune generalizzazioni (cfr. Matras, 2020: 165-178 per una panoramica complessiva). Per esempio, gli studiosi sono concordi nell'osservare che gli elementi lessicali sono presi a prestito più frequentemente degli elementi grammaticali e che gli elementi morfologici liberi sono più frequentemente presi a prestito degli elementi morfologici legati. Alcuni studiosi hanno inoltre proposto distinzioni a grana più fine all'interno di specifiche categorie lessicali o funzionali. Per esempio, è stato notato che, tra le parole contenuto, i nomi sono maggiormente propensi al prestito rispetto ai verbi e agli aggettivi; che tra i morfemi legati, quelli derivazionali sono maggiormente soggetti al prestito rispetto a quelli flessivi; che tra le parole funzione, operatori connessi all'organizzazione del discorso (come i connettivi e i segnali discorsivi) sono maggiormente soggetti al prestito rispetto a elementi deittici come determinanti o pronomi; e che, nel novero dei connettivi, *ma* è più propenso al prestito rispetto *o* o *ed e* (Matras, 1998a).

Un "classico" sulle gerarchie del prestito è il già citato studio di Thomason & Kaufman (1988), in cui è sostenuto che i processi di prestito sono gradualisti e connessi all'intensità del contatto. In una prima fase di contatto "casuale", soltanto le parole contenuto appartenenti al lessico culturale sarebbero soggette al prestito, mentre, all'aumentare dell'intensità del contatto, "structural features may be borrowed as well – phonological, phonetic and syntactic elements, and even (though more rarely) features of the inflectional morphology" (Thomason & Kaufman, 1988: 37). Queste generalizzazioni sono state tuttavia parzialmente smentite da studi empirici. Per esempio, Adamou (2016) ha mostrato che il prestito di *pattern* grammaticali può essere una strategia molto frequente in una comunità bilingue anche in assenza di un numero significativo di prestiti lessicali. A questo proposito, l'autrice osserva che una preferenza per il prestito di *pattern* grammaticali piuttosto che di materiale lessicale può correlarsi con atteggiamenti prescrittivi dei parlanti e la pressione esercitata da una varietà standard.

Una delle più recenti gerarchie del prestito è stata elaborata in seno a un progetto internazionale guidato dall'Università di Manchester, sulla base di un campione tipologico di 27 lingue (Matras & Sakel, 2007; Matras, 2007). Essa è strutturata come segue e vede come categorie più propense al prestito i nomi e le congiunzioni:

nouns, conjunctions > verbs > discourse markers > adjectives > interjections > adverbs > other particles, adpositions > numerals > pronouns > derivational affixes > inflectional affixes. (Matras, 2007).

Simili domande di ricerca sono anche state perseguite da studiosi interessati all'analisi del parlato bilingue. Questi approcci offrono una prospettiva complementare al tema delle gerarchie del prestito in diacronia, fornendo un'interessante panoramica “in tempo reale” di tali processi (Matras & Adamou, 2023). Come prevedibile, gli studi basati su corpora di parlato tendono a mostrare analogie con gli studi diacronici relativi alle gerarchie di prestabilità, sottolineando spesso la predominanza dei nomi nel novero del materiale alloglotto (cfr. p.es. Myers-Scotton, 1993; Treffers-Daller, 1994; Poplack et al., 1988; Adamou, 2016).

Più nello specifico, analizzando in prospettiva comparativa diversi corpora multilingui da lingue minoritarie e spesso minacciate, Adamou (2016) osserva una correlazione tra la percentuale di parole inserite da una dominante in contatto in discorsi nella lingua minoritaria e la varietà di categorie grammaticali coinvolte nel contatto. In particolare, nei corpora con meno del 5% di parole da una lingua di contatto, soltanto i nomi emergono come la categoria quantitativamente rilevante nel novero delle inserzioni alloglotte; viceversa, i corpora con più del 20% di parole di contatto esibiscono un numero maggiore di classi inserite, come verbi, aggettivi, avverbi, congiunzioni e preposizioni.

Infine, gli studiosi si sono anche occupati di individuare gerarchie del prestito a livello semantico. Il più importante progetto collaborativo in questo senso è il WOLD (*The World Loanword Database*; cfr. Haspelmath & Tadmor, 2009)¹, basato su un campione tipologico di 41 lingue. I risultati di questo progetto hanno mostrato che i domini semantici più propensi al prestito sono quelli connessi alla religione, all'abbigliamento, alla casa e alla legge, mentre i domini semantici meno propensi al prestito sono quelli connessi alla percezione, alle relazioni spaziali, al corpo e alla parentela. L'idea risultante è che concetti connessi all'organizzazione sociale sono più propensi al prestito rispetto a concetti connessi al mondo fisico-naturale e alla sfera personale dei parlanti (cfr. anche Matras, 2020: 188).

1.4.3 Motivazioni per il prestito

Oltre a dar conto delle diverse distribuzioni delle varie classi lessicali e funzionali tra i prestiti delle lingue del mondo, gli studiosi hanno anche cercato di stabilire i fattori generali in grado di spiegare tali fenomeni. In particolare, Matras (2020) sottolinea la necessità di distinguere tra fattori che *motivano* il prestito e fattori che *facilitano* il prestito. Di seguito, seguendo questa distinzione, discutiamo brevemente entrambi gli aspetti.

Le motivazioni per il prestito possono essere definite, seguendo Matras & Adamou (2021: 242), “the factors that drive the use of linguistic matters in interaction settings that are generally associated with

¹ <https://wold.clld.org/>; ultimo accesso: 12/02/2026.

another language”. Un primo fattore spesso evocato a questo proposito è la necessità di colmare lacune lessicali in una lingua ricevente (noto nella letteratura in lingua inglese come *gap filling*). In effetti, in qualsiasi lingua sono presenti termini che designano referenti culturali assenti prima di una qualche forma di contatto con specifiche culture: si pensi ad esempio a prestiti inglesi entrati nell’italiano come *baseball* o *basket*. Queste forme sono anche note come “cultural borrowings” (Myers-Scotton, 1993). Tuttavia, analisi empiriche della distribuzione dei prestiti per campi semantici in specifiche lingue tendono a ridimensionare l’ipotesi del *gap filling*, in quanto i prestiti spesso non si limitano a sfere di attività chiaramente associate a una cultura in contatto (cfr. Poplack et al., 1988; Poplack, 2018). In questo senso, Poplack (2018) sottolinea la preminenza, rispetto al *gap filling*, di fattori di tipo pragmatico, come l’intento di ottenere una maggiore efficacia comunicativa sfruttando tutte le risorse del proprio repertorio.

Attualmente, diversi studiosi tendono a fare convergere queste due motivazioni. Ad esempio, Backus (1996) osserva che i parlanti bilingui olandese-turco nei Paesi Bassi tendono a riferirsi al concetto di “birra” tipicamente in olandese, non perché il turco non disponga di un equivalente adeguato, ma in quanto la birra è associata a un’attività svolta in contesti interazionali tipicamente associati alla lingua olandese: riferirsi alla birra in olandese permette dunque di stabilire un’associazione più precisa con il referente e l’universo di valori da esso evocato. L’idea risultante è che:

Translation equivalents are not always exact equivalents; that is, the encyclopedic knowledge that a speaker has about a certain word will seldom exactly match the connotations she has for the translation equivalent. (Backus, 1996: 115)

Similmente, Matras osserva che:

Gaps are not to be interpreted as deficiencies in the recipient system, but rather as speakers’ attempt to avail themselves of their full inventory of linguistic resources, at all times and in all contexts of interaction. (Matras, 2020: 150)

L’ipotesi del *gap filling* in senso tradizionale è inoltre messa in discussione dalla forte propensione al prestito di alcuni tipi di elementi funzionali, in particolare quelli responsabili dell’organizzazione del discorso, come i connettivi e i segnali discorsivi, definiti *utterance modifiers* da Matras (1998a). Matras (1998a), in particolare, connette la particolare propensione al prestito di questi elementi funzionali non già alla necessità dei parlanti di colmare supposte lacune nella lingua ricevente, né al maggiore prestigio della lingua donatrice, ma alla pressione cognitiva che questi elementi esercitano sui parlanti durante le interazioni in modalità bilingue. Tale pressione originerebbe dal ruolo cruciale degli *utterance modifiers* nel direzionare l’interazione con l’interlocutore e nel gestire le

presupposizioni condivise. Sperimentando tale pressione cognitiva, i parlanti bilingui possono incontrare difficoltà nel mantenere separate le lingue del proprio repertorio in prossimità di questi elementi, tendendo ad attingerli stabilmente da una delle lingue in contatto, definita lingua “pragmaticamente dominante” (Matras, 1998a). Inoltre, il fatto che gli *utterance modifiers* tendano a essere percepiti dai parlanti quasi come parte di un apparato para-linguistico e gestuale, più che strettamente linguistico (come è invece il caso delle parole contenute), può far sì che la loro inserzione in discorsi in un’altra lingua tenda a passare più “inosservata”, con la conseguente possibilità di generalizzarsi più facilmente nel lungo termine anche in comunità caratterizzate da atteggiamenti linguistici più prescrittivi.

In sintesi, secondo Matras (2020), due sarebbero i principali fattori a *motivare* i processi di prestito: la necessità di colmare lacune *lato sensu* sfruttando tutte le risorse del proprio repertorio e la pressione cognitiva che alcuni operatori discorsivi impongono sui parlanti durante l’interazione in modalità bilingue.

In aggiunta ai fattori che motivano il prestito, Matras (2020), come detto, cita anche fattori che *facilitano* il prestito. In particolare, l’autore individua una “utilitarian motivation” (Matras, 2020: 173) per la quale gli elementi più frequentemente presi a prestito sono quelli più frequentemente associati a contesti interazionali in cui viene utilizzata la lingua donatrice. Tale motivazione spiegherebbe, per esempio, la maggiore propensione al prestito dei nomi a referente unico, i quali includono termini istituzionali che il parlante desidera menzionare nella forma originale per evocare una connessione fedele con la realtà da essi evocata (cfr. già Backus, 1996, 2001), o la maggiore propensione al prestito dei numerali cardinali più alti rispetto ai numerali cardinali più bassi, in quanto i primi vengono quasi esclusivamente utilizzati in contesti interazionali associati alla lingua dominante (per esempio nel contesto delle transazioni economiche). Van Hout & Muysken (1994), attraverso un caso di studio sul parlato bilingue spagnolo-quechua, confermano empiricamente che i lessemi più frequenti nella lingua donatrice (lo spagnolo) sono più sensibili al prestito nella lingua ricevente (il quechua). Allo stesso tempo, Matras (2020) osserva come siano più immuni dal prestito elementi che nella lingua ricevente si caratterizzano per elevata frequenza d’uso e “prossimità” al parlante. In tal senso, per esempio, termini di parentela più vicini alla sfera del soggetto (come *madre*, *padre*, *fratello*) sarebbero meno propensi al prestito rispetto ai termini più lontani dalla sfera del soggetto (come *nipote* o *cugino*). In sintesi, la frequenza d’uso nella lingua donatrice e la frequenza/prossimità nella lingua ricevente possono agire con effetti in parte opposti e in interazione dinamica nel determinare la probabilità che una forma alloglotta si stabilizzi diacronicamente come prestito.

Infine, diversi autori concordano nell'osservare che, mentre i fattori linguistici facilitano o motivano il prestito, sono i fattori sociolinguistici quelli decisivi nel consentire che una forma prodotta idiosincraticamente da individui multilingui possa stabilizzarsi in una comunità linguistica, anche al di fuori delle interazioni in modalità bilingue (cfr. p.es. Matras, 2020; Backus, 2014). Pertanto, giocano qui un ruolo primario fattori extralinguistici quali, ad esempio, gli atteggiamenti da parte dei singoli individui verso l'incorporazione di materiale alloglotto, nonché il grado di standardizzazione della lingua e la pressione percepita da parte di eventuali varietà standard, che tendono a promuovere la formazione di neologismi attraverso materiale autoctono piuttosto che il prestito diretto da lingue di contatto (cfr. Adamou, 2016 per uno studio empirico in questa direzione).

1.5 Lingue minacciate, documentazione linguistica, linguistica del contatto

La possibilità delle lingue di “estinguersi” rappresenta oggi una delle principali preoccupazioni dei linguisti. Si stima infatti che, in assenza di interventi *ad hoc*, circa la metà delle lingue parlate nel mondo potrebbe scomparire entro i prossimi cento anni, il che comporterebbe una riduzione della diversità linguistica globale dalle circa 6.000–8.000 lingue documentate oggi a circa 3.000–4.000 entro il 2100 (Adamou, 2024). L'interesse per le lingue minacciate ha conosciuto una crescita particolarmente significativa a partire dalla fine degli anni Novanta del secolo scorso e diversi enti internazionali hanno finanziato progetti volti alla documentazione di lingue in pericolo, dando vita a una vera e propria branca della linguistica denominata documentazione linguistica (per una panoramica sullo stato dell'arte del settore, cfr. Seifart et al., 2018).

Una lingua a rischio di estinzione può essere definita come minacciata o in pericolo (*endangered*). La nozione di lingua minacciata non è perfettamente sovrapponibile a quella di lingua di minoranza: quest'ultimo termine indica più in generale una lingua i cui locutori siano in minoranza demografica all'interno di uno Stato (cfr. Toso, 2008; Fiorentini, 2022). In tal senso, una lingua minoritaria non è necessariamente minacciata; viceversa, una lingua minacciata è quasi certamente minoritaria. Ovviamente, il grado di minaccia varia considerevolmente da lingua a lingua e, per tenere conto di tali differenze, gli studiosi hanno proposto diverse scale di “vitalità” o “endangerment” delle lingue. Una delle più celebri è la *Graded Intergenerational Disruption Scale* (GIDS) proposta da Fishman (1991), che pone l'accento sul grado di trasmissione intergenerazionale della lingua. Nell'ottica dell'autore, la trasmissione intergenerazionale non deriva soltanto dalla scelta dei genitori di non trasmettere la lingua ai figli, ma rappresenta anche il risultato di scelte sociali e istituzionali più generali che regolano il comportamento linguistico dei genitori con i propri figli. Più nello specifico, la scala proposta da Fishman (1991) consta di otto livelli, da “safe” a “endangered”: il primo livello

indica una lingua utilizzata in contesti educativi, lavorativi e nei media, mentre l'ottavo livello indica una lingua i cui ultimi parlanti si rintracciano nella generazione dei nonni.

Più di recente, Lewis & Simons (2010) hanno proposto una versione estesa della GIDS, nota anche EGIDS (*Expanded Graded Intergenerational Disruption Scale*), caratterizzata da dieci diversi livelli. Il primo livello qualifica una lingua come “international”, mentre il decimo livello qualifica una lingua come “extinct”. Secondo Lewis & Simons (2010), una lingua inizia a essere “vulnerable” a partire dal livello 6b, ovvero quando la lingua è utilizzata oralmente da tutte le generazioni, ma soltanto alcune delle generazioni con figli la utilizzano e, dunque, soltanto alcuni bambini la imparano.

Oggi, quasi tutte le lingue minacciate sono anche lingue di contatto, poiché i loro parlanti sono tipicamente bilingui con una o più lingua maggioritarie, generalmente apprese insieme alla lingua locale già durante la socializzazione primaria. Nei progetti di documentazione linguistica, tuttavia, l'attenzione è spesso rivolta alle lingue considerate in isolamento piuttosto che alla loro collocazione nei repertori plurilingui dei parlanti e alla loro interazione con le altre lingue. Di conseguenza, le tangenze tra la linguistica del contatto e la documentazione linguistica rimangono, paradossalmente, ancora poche (Adamou, 2016; Dal Negro, 2021).

In effetti, una delle priorità dei progetti di documentazione linguistica è spesso quella di realizzare dizionari e grammatiche, anche in un'ottica di rivitalizzazione delle lingue per le future generazioni, non di rado all'insegna di un approccio puristico che tende a escludere i numerosi prestiti lessicali (soprattutto quelli più recenti e riconoscibili) dal materiale da documentare (Dal Negro, 2021). Tuttavia, alcuni studiosi attivi nell'ambito della sociolinguistica critica hanno mostrato riserve verso tali “discorsi di preservazione”, i quali possono correre il rischio di fissare “exoticized people in time and place while overlooking their local language practices and ideologies” (Pennycook, 2016: 3). In sintesi, passare da una visione per certi versi reificata delle lingue minacciate a una visione più ampia e dinamica costituisce oggi una delle sfide della documentazione linguistica (Dal Negro, 2021).

L'approccio bilingue adottato in questo volume per lo studio del sinto lombardo intende intraprendere questa direzione, focalizzandosi su come il parlato dei sinti lombardi si adatti dinamicamente alla tensione tra mantenimento della lingua tradizionale e l'influenza crescente di una lingua socialmente dominante come l'italiano.

Capitolo 2. La lingua romaní e i suoi parlanti

Il sinto lombardo, su cui è incentrato il presente lavoro, rappresenta una delle numerose varietà di romaní, una lingua non territorializzata di origine indoaria parlata da milioni di persone rom e sinte in Europa e nel mondo. Questo capitolo è dunque incentrato sulla lingua romaní. Dopo un breve inquadramento storico-antropologico dei suoi parlanti, i rom¹ (§ 2.1), la romaní viene discussa secondo una prospettiva storico-tipologica, dialettologica e sociolinguistica (§ 2.2). Successivamente, viene proposta una disamina approfondita sulla situazione linguistica e sociolinguistica della romaní in Italia (§ 2.3).

2.1 I rom: cenni storici e antropologici

In ogni Paese europeo, così come in America e Australia, sono presenti comunità che parlano (o hanno parlato in passato) varietà indoarie strettamente imparentate e che condividono alcuni tratti culturali salienti. Tali gruppi designano sé stessi con una vasta gamma di etnonimi che può variare a seconda dell'area geografica di insediamento. I più diffusi sono probabilmente *roma* (masc. sg. *rom*, femm. sg. *romni*, femm. pl. *romnja*) e *sinti* (masc. sg. *sinto*, femm. sg. *sinta*, femm. pl. *sinte*), ma sono anche attestati etnonimi quali *calós* / *kalé* / *kaale* (diffusi nella Penisola Iberica e in Finlandia), *manuš* (diffusi tra Francia e Germania) e *romanichals* (nel Regno Unito); cfr. Beníšek (2020) per una panoramica. Per quanto riguarda la distribuzione geografica di questi etnonimi, Piasere (2009) propone di tracciare una linea immaginaria che collega Roma e Helsinki, passando per Vienna e Praga: a est di questa linea si individuano gruppi che si identificano principalmente come *roma*, mentre a ovest si individuano gruppi che si identificano principalmente come *sinti*. Il termine *rom*, di origine indoaria, significa 'uomo' ed è ancora utilizzato in questo senso in tutte le varietà di romaní. L'origine del termine *sinto* è invece controversa: ricerche recenti ne evidenziano una possibile origine europea, riconducendolo alla forma del medio alto tedesco *sint* 'viaggio, strada' (Reitinger, 2023). La comune origine indiana dei rom è oggi riconosciuta universalmente. L'indicatore più evidente di ciò è la lingua da essi storicamente utilizzata, la romaní, la cui origine indoaria e la conseguente vicinanza a lingue indoarie antiche e moderne come il sanscrito e la hindi è stata osservata a partire dallo studio pionieristico di Rüdiger (1782). Tuttavia, le prove linguistiche non hanno permesso di

¹ Poiché *rom* è l'etnonimo più diffuso tra questi gruppi, esso verrà adottato per brevità al fine di descriverne gli aspetti generali e le tendenze comuni (cfr. p.es. Piasere, 2009 per un approccio simile). Nel fare riferimento a gruppi specifici, si utilizzeranno invece i relativi etnonimi.

delineare con chiarezza lo specifico profilo etnico e sociale dei rom all'interno del Subcontinente indiano, né le ragioni storiche della migrazione verso occidente. Va del resto ricordato, con Matras (2002: 18), che “the issue touches on images and self-images of the populations concerned, and so it is likely to remain a point of controversy”.

Per una discussione sull'originario profilo socio-etnico dei rom, è di primaria importanza considerare la presenza di diverse popolazioni di origine indiana al di fuori dell'India, insediate prevalentemente in Medio Oriente e in Asia Centrale, le quali, oltre a parlare varietà indoarie, condividono con i rom d'Europa altri tratti socioeconomici molto salienti. Tra questi gruppi è possibile individuare: i dom, i karači e i kurbati dell'area siro-palestinese-giordana, parlanti domari; gli indu e gli jat dell'Afghanistan, parlanti una varietà indoaria centrale; i parya del Tagikistan, parlanti una varietà di rajasthani; e i doma del Pakistan settentrionale (valle dell'Hunza), parlanti dumaki (Matras, 1998b). I membri di questi gruppi svolgono generalmente attività semi-itineranti volte a fornire servizi alla popolazione maggioritaria, come la lavorazione dei metalli o l'intrattenimento musicale, e vivono spesso in una condizione di relativa marginalità socioeconomica. Inoltre, condividono termini affini per designare quanti non fanno parte della propria comunità: per le popolazioni romaní, questo termine è *gadže* (masc. sg. *gadžo*, femm. sg. *gadži*)² (Benišek, 2020). L'ipotesi che inquadra i rom nel più ampio scenario delle caste dell'India insieme ad altre popolazioni indoarie dell'Asia centrale e occidentale tende a essere oggi ampiamente accettata. Come spiega Matras (2002), essa ha il pregio di attribuire i tratti condivisi dai vari gruppi a una tradizione affine piuttosto che a caratteristiche acquisite separatamente, senza al contempo imporre una comune origine genetica.

Le ragioni e la datazione dell'inizio delle migrazioni dei rom verso ovest rimangono oggetto di dibattito accademico, ma le prove linguistiche sembrano convergere con le fonti storiche nel datare la partenza dei rom dall'India attorno al 500 d.C. (Benišek, 2020). Le tappe intermedie del percorso dei rom verso ovest sono in parte ricostruibili attraverso l'analisi dei prestiti lessicali condivisi nelle varietà romaní parlate oggi, che testimoniarebbero di passate fasi di bilinguismo con le diverse lingue dei diversi Paesi attraversati. Questo tipo di analisi suggerisce che tale percorso ha sicuramente incluso l'attraversamento della Persia e dell'Armenia, fino al raggiungimento dell'Impero bizantino in età medievale, attorno al X secolo d.C. L'impero bizantino ha verosimilmente costituito la principale area di insediamento dei rom per alcuni secoli, come testimoniano numerose fonti storiche (cfr. de Vaux de Foletier, 2010: 46-50) e il forte influsso del greco bizantino sul lessico e sulla sintassi romaní. Tuttavia, probabilmente in risposta all'instabilità portata dall'invasione ottomana dei

² Nel prosieguo di questo volume, questo termine verrà trattato come un prestito adattato e riceverà la grafia italianizzata *gagi/gagio*. La grafia romaní <dž> verrà invece mantenuta negli esempi di parlato in sinto lombardo riportati di volta in volta.

Balcani, attorno al XIV secolo i rom ripresero a spostarsi, questa volta in gruppi individuali, dando vita a una vera e propria “diaspora” nei vari paesi dell’Europa centrale, orientale e occidentale (Piasere, 2009). Questo fatto segna anche l’inizio della differenziazione linguistica delle varietà romaní, che, fino ad allora, si ritiene costituissero un gruppo relativamente omogeneo (Benišek, 2020).

2.2 La lingua romaní

La romaní è una lingua indoeuropea del gruppo indoario, parlata esclusivamente al di fuori del Subcontinente indiano. Il termine romaní è un glottonimo utilizzato in ambito accademico e nei documenti legislativi dell’Unione Europea; si tratta di un aggettivo femminile singolare, presente all’interno del sintagma *romani čhib*, che significa letteralmente ‘lingua dei rom’. Tuttavia, come osserva Scala (2020a: 85), tale denominazione appare “poco o per nulla in uso presso i parlanti di tale lingua e pertanto è sentita spesso come estranea, esogena, imposta da chi non appartiene alla comunità”. I membri delle singole comunità sembrano preferire altri glottonimi quali, ad esempio, *amari čhib* ‘la nostra lingua’ o *romanes*, forma avverbale che può essere grossomodo tradotta ‘alla maniera dei rom’ (Matras, 2002: 5). Nei paragrafi a seguire, come anticipato, discutiamo alcuni aspetti storici-tipologici, dialettologici e sociolinguistici della romaní.

2.2.1 Cenni storici e tipologici e il ruolo del contatto linguistico

Utilizzando il metodo storico-comparativo, i linguisti hanno tentato di ricostruire un antenato comune per le diverse varietà di romaní parlate oggi in Europa. Ciò ha portato alla distinzione tra una fase “proto-romaní” e una fase “early-romaní” (cfr. Matras, 2002: 18-19; Benišek, 2020).

La proto-romaní è un’entità linguistica astratta, definibile come la somma dei mutamenti occorsi durante la fase pre-europea della romaní. Il suo inizio coinciderebbe con il momento in cui tale lingua è divenuta sufficientemente distinta dalle altre lingue indoarie del Subcontinente indiano, mentre la sua fine coinciderebbe con l’arrivo dei rom nell’Impero bizantino in periodo medievale. Qui inizia la early-romaní, che rappresenta l’ultima fase unitaria della romaní prima della diaspora dei rom, avvenuta a seguito dell’invasione ottomana dei Balcani. Mentre la proto-romaní è una nozione inerentemente diacronica, la early-romaní può essere vista come uno stadio sincronico, che costituisce la base per la variazione delle diverse varietà attestate oggi (Benišek, 2020). Un tratto saliente di questa fase è l’acquisizione massiccia di elementi lessicali e morfosintattici dal greco, che sono infatti presenti in tutte le varietà romaní moderne.

In quanto lingua diasporica, la romaní ha conosciuto sostanziali innovazioni su tutti i livelli linguistici a causa del contatto con lingue asiatiche parlate al di fuori del Subcontinente indiano e, soprattutto,

con lingue europee. Di conseguenza, una trattazione, ancorché sommaria, della tipologia strutturale della romaní non può prescindere da riferimenti a fatti di contatto linguistico (Adamou & Matras, 2020).

Il lessico è probabilmente il settore in cui tali contatti sono più evidenti. In generale, gli studiosi individuano almeno tre diversi strati lessicali presenti in tutte le varietà di romaní: (i) lo strato lessicale nativo, conservato dalla fase indoaria; (ii) gli strati iranico e armeno; (iii) lo strato greco (Matras, 2002; Scala, 2020b). In molte varietà si può poi constatare la presenza di ulteriori strati, costituiti dai prestiti dalle varie lingue europee incontrate dopo la diaspora. Il lessico nativo si conserva a vari livelli nelle diverse varietà di romaní, ma, in generale, sembra mantenersi soprattutto a livello di lessemi appartenenti al lessico di base ad alta frequenza (Elšik, 2009; Scala, 2020b). Nell’ambito del già citato progetto WOLD (*World Loanword Database*; Haspelmath & Tadmor, 2009), Elšik (2009) analizza una varietà di romaní parlata in Slovacchia e riporta che i lemmi acquisiti tramite contatti al di fuori del Subcontinente indiano ammontano al 64% del lessico totale. Si tratta del tasso più elevato all’interno di un ampio campione tipologico, il che ha portato a individuare nella romaní un caso prototipico di “high borrower language” (Tadmor, 2009).

Tuttavia, gli effetti del contatto linguistico non si manifestano nella romaní soltanto a livello lessicale, ma anche a livello morfologico e sintattico. Per esempio, come già accennato al Capitolo 1, il contatto prolungato con il greco ha portato al prestito di interi paradigmi nominali, tale per cui morfemi indotti dal greco sono oggi utilizzati per integrare i nuovi prestiti europei, mentre i morfemi nativi rimangono legati al lessico pre-europeo (cfr. Elšik, 2000). Nel dominio della morfologia, sono inoltre ben documentati casi di prestito di morfemi tempo-aspettuali, soprattutto in varietà a contatto con il turco (Friedman, 2020) o con le lingue slave (Rusakov, 2001; Sonnemann, 2022).

Infine, dal punto di vista sintattico, la romaní è stata definita una lingua indoaria “balcanizzata” (Matras, 1994). Tra le innovazioni sintattiche che sembrano essere state motivate dal contatto con le lingue balcaniche, e in particolare il greco, possiamo citare, seguendo Adamou & Matras (2020):

- l’emergere di un allineamento nominativo-accusativo, in opposizione agli allineamenti ergativi o scissi che si individuano nella maggior parte delle lingue indoarie moderne e che alcuni studiosi hanno anche ricostruito per la proto-romaní (cfr. anche Beníšek, 2020);
- l’emergere di un ordine VO non marcato, in opposizione a un ordine OV dominante nelle lingue indoarie moderne, nonché la flessibilità nel posizionamento del soggetto sulla base di fattori pragmatici o connessi alla struttura informativa;
- l’emergere di preposizioni in luogo di posposizioni, che invece sono dominanti nelle lingue indoarie moderne;

- lo sviluppo di un articolo definito, probabilmente a partire dalla grammaticalizzazione del dimostrativo (Matras, 2002);
- la scomparsa dell'infinito verbale, un processo probabilmente iniziato attraverso il contatto con lingue iraniche in Persia (Matras, 2002) e portato a compimento durante il contatto con le lingue balcaniche;
- l'uso di complementatori distinti a seconda del grado di fattualità della subordinata completiva. In particolare, il complementatore *kaj* (ereditato dall'antico indoario) introduce subordinate completive che descrivono eventi fattuali, mentre il complementatore *te*, probabilmente anch'esso di origine indoaria, introduce completive non fattuali o modali (Matras, 2002: 180). Questa distinzione nei complementatori sembra ricalcare quella che si individua nelle lingue balcaniche: per questa regione, *te* è stato spesso discusso in letteratura nel contesto del cosiddetto “dental modal subordinator” balcanico (cfr. p.es. Friedman & Joseph, 2025: 215-216).

Complessivamente, la lingua romaní rappresenta un caso di estremo interesse per gli studiosi in quanto, grazie alla sua ricca storia di contatti, permette di illustrare “the significance of the geographical distribution of typological features over genetic affiliation” (Adamou & Matras, 2020: 223).

2.2.2 Cenni dialettologici

Sebbene le diverse varietà di romaní condividano un nucleo strutturale e lessicale ereditato dalla fase indoaria o indotto da fasi di contatto comuni, il grado di variazione diatopica è oggi notevole, anche a causa della dispersione territoriale dei suoi parlanti e della mancanza di standardizzazione. Di conseguenza, l'intercomprensione tra locutori di varietà diverse è spesso molto ardua. Ciò è stato connesso soprattutto alla presenza di prestiti lessicali accolti individualmente nelle diverse varietà e da variazioni a livello fonetico/fonologico (Elšik & Benišek, 2020). A dimostrazione della difficoltà di intercomprensione tra parlanti di diverse varietà di romaní, Elšik & Benišek (2020) propongono una frase dal significato basilare: “tutti i miei figli sono ancora piccoli”. Mentre un parlante della Bulgaria direbbe *epci me xurde thaa čikone* (*epci* ‘tutto’ e *thaa* ‘ancora’ sono prestiti dal turco), un parlante finlandese direbbe *sāre mo kenti āxena panna peska* (*kenti* ‘bambini’ e *peska* ‘piccolo’ sono prestiti dal basso tedesco): notiamo che quasi nessun lessema è condiviso nelle due varietà. Similmente, Sonnemann (2022) confronta la distanza strutturale tra due varietà di romaní parlate in

Bulgaria e Finlandia sulla base di campioni estratti dal *Database of Romani Dialects*³, già *Romani Morpho-Syntax Database*, e rileva che il lessico condiviso tra le due varietà ammonta a meno del 50%.

L'approccio prevalente nella linguistica romaní per la classificazione dei dialetti romaní è di tipo "genetico" e affonda le radici nel metodo storico-comparativo (cfr. Matras, 2005a; Elšik & Beníšek, 2020). In quest'ottica, i singoli dialetti sono visti come appartenenti a uno dei diversi rami che costituiscono le propaggini di un'originaria proto-varietà, identificata nella early-romaní, mentre la suddivisione tra i vari rami è vista come il risultato di migrazioni di gruppi individuali. Tale modello ha il suo riferimento principale nel lavoro di Miklosich (1872-1880), il quale ha distinto tredici gruppi di dialetti romaní sulla base dei prestiti lessicali accolti dalle varietà di contatto.

Nel 1915, Gilliat-Smith introduce la celebre distinzione tra dialetti "vlax" e "non-vlax". Questa distinzione era stata inizialmente concepita in relazione alle varietà parlate in Bulgaria, ma è stata successivamente estesa dagli studiosi alla romaní più in generale. In particolare, la distinzione tra dialetti vlax e non vlax origina dall'individuazione di correlazioni sistematiche tra variazioni strutturali nella lingua e caratteristiche extralinguistiche di gruppi rom nel nord-est della Bulgaria, in cui i dialetti vlax sono parlati da popolazioni cristiane e nomadi che si autodesignano, appunto, come *vlax* (in quanto originarie della regione storica della Valacchia), mentre i dialetti non vlax sono parlati da gruppi musulmani e sedentari.

Matras (2005a) introduce un approccio areale alla classificazione delle varietà romaní, basato sull'identificazione di innovazioni condivise da varietà parlate in territori geograficamente contigui. In questo modello, l'analisi di tali innovazioni può portare a individuare un continuum di dialetti che riflette la diffusione nel tempo e nello spazio di specifici tratti, rendendo possibile stabilire una correlazione tra vicinanza geografica e variazione strutturale. Sotteso a ciò è il presupposto che le comunità rom non costituiscono "isole", idea prevalente negli approcci di stampo genetico, ma che, almeno in alcune regioni, esse possono formare reti di insediamento stabili attraverso le quali le innovazioni linguistiche hanno la possibilità di diffondersi in maniera graduale.

Come sottolineano Elšik & Beníšek (2020), i raggruppamenti dialettali proposti dagli approcci genetici e areali finiscono di fatto per coincidere in larga parte. Recentemente si è anche cercato di applicare altri metodi, come ad esempio quelli legati alla dialettologia percettiva, basata sul riconoscimento dei confini dialettali da parte dei parlanti (cfr. Friedman, 2017).

³ <https://romani.dch.phil-fak.uni-koeln.de/>; ultimo accesso: 15/03/2026.

Di seguito, presentiamo brevemente la classificazione dialettale proposta da Matras (2002), che distingue quattro gruppi principali: il gruppo balcanico; il gruppo vlax; il gruppo centrale; il gruppo settentrionale.

Il gruppo balcanico ha il suo centro nei Balcani meridionali (Turchia, Grecia, Bulgaria, Macedonia del Nord, Albania, Kosovo). In linea generale, questi dialetti sono molto influenzati dal greco e dal turco; i locutori sono tipicamente musulmani e possono perciò mantenere una conoscenza attiva del turco. Il gruppo balcanico può essere diviso in due sottogruppi: il gruppo meridionale e il gruppo “drindari/kalajdži/burgudži”.

Il gruppo vlax è probabilmente il più prominente per quanto riguarda il numero di parlanti e la documentazione linguistica. La sua zona di irradiazione è identificata nell’attuale Romania, come testimonia la forte influenza del rumeno in queste varietà su tutti i livelli linguistici. Differentemente dai gruppi parlanti varietà balcaniche, i parlanti varietà vlax sono tipicamente cristiani, ma sono documentati anche alcuni gruppi musulmani. Anche questo gruppo può ulteriormente essere distinto in due sottogruppi, comunemente definiti vlax meridionale e vlax settentrionale. I dialetti vlax meridionali più vicini alla loro zona originaria di irradiazione sono probabilmente quelli parlati in Bulgaria, continuatori delle varietà già documentate da Gilliat-Smith (1915); più a sud, distinguiamo ulteriormente un gruppo di varietà parlate in area greca e un gruppo di varietà parlate tra Albania e Bosnia (queste ultime note come “gurbet”). I dialetti vlax settentrionali includono invece varietà parlate nell’attuale Romania; tra queste, si possono segnalare la varietà kalderaš e lovari, quest’ultima parlata nell’attuale Transilvania dove si trova in contatto con l’ungherese.

Il gruppo centrale ha il suo baricentro nelle attuali Repubblica Ceca, Slovacchia e Ungheria (con propaggini in Slovenia e in Austria, Polonia meridionale e Ucraina occidentale). Anche questo gruppo è distinto in un ramo settentrionale e meridionale. Le varietà che ne fanno parte sono tipicamente caratterizzate dall’influenza dall’ungherese e dalle lingue slave (sia settentrionali come il ceco e lo slovacco, sia meridionali come lo sloveno); molte di queste sono (state) soggette a forti processi di *shift* verso le lingue maggioritarie di contatto.

Il gruppo settentrionale è il più eterogeneo e controverso. Tipicamente, viene suddiviso in un gruppo nordoccidentale e nordorientale. Nel primo gruppo rientrano varietà fortemente influenzate dal tedesco, il cui nucleo è probabilmente rappresentato dalle varietà sinti-manuš parlate in Germania, Austria, Olanda, Francia e Italia settentrionale, che risultano simili anche alla romaní finlandese, oggi fortemente minacciata. A questo gruppo appartiene il sinto lombardo, oggetto di questo volume. Nel gruppo nordorientale rientrano invece una serie di dialetti parlati nella Russia occidentale, paesi baltici e Polonia. Vengono talvolta connesse al gruppo occidentale anche la romaní britannica e la

romaní iberica, che oggi hanno lasciato spazio a due varietà para-romaní definite rispettivamente angloromaní e calò (cfr. § 1.3.3).

Infine, Matras (2002) menziona due gruppi isolati. Il primo è la romaní croata, parlata nei Balcani occidentali tra Croazia e Italia orientale da gruppi che si definiscono come *Hrvati* (‘croati’, appunto), mentre il secondo è la romaní parlata nell’Italia meridionale, in Abruzzo, Calabria e Molise, anche definito da Elšik & Benišek (2020: 401) “Appennine Romani”. La Figura 1 di seguito, tratta da Matras (2002: 11), propone una visualizzazione grafica dei principali raggruppamenti dialettali della romaní d’Europa.



Figura 1: Mappa dei principali dialetti romaní (da Matras, 2002: 11).

2.2.3 Aspetti sociolinguistici generali

È difficile ottenere dati certi sul numero di parlanti di romaní, sia in Europa sia in altri continenti, come del resto è difficile stimare il numero complessivo di rom in Europa e nel mondo. Bakker (2001) stima il numero di parlanti romaní a circa 4,6 milioni, con numeri variabili da Stato a Stato. Secondo Matras (2005b: 2), tra le lingue minoritarie rappresentate nel Consiglio d'Europa la romaní è quella con il maggior numero di parlanti dopo il catalano. Con l'ingresso nell'Unione Europea di paesi come la Bulgaria e la Romania nel 2004, si può ragionevolmente considerare che la romaní sia oggi la lingua minoritaria più rappresentata.

Per quanto riguarda la posizione della romaní nel repertorio linguistico dei suoi parlanti, va notato che essa è in prima istanza una lingua orale, generalmente acquisita come prima lingua. L'uso della

romaní tende a non essere regolato e implementato da alcuna agenzia come la scuola o i media, anche a causa dell'assenza di riconoscimenti ufficiali in molti paesi. In secondo luogo, va osservato che tutti i parlanti di romaní sono almeno bilingui con una o più lingue maggioritarie (Matras & Adamou, 2020), la cui acquisizione è spesso simultanea o quasi-simultanea alla romaní (Halwachs, 2020). A livello societario, si tratta di un bilinguismo “unidirezionale” e “asimmetrico” (Matras & Adamou, 2020: 329), nel senso che tutti i parlanti rom sono bilingui con le lingue maggioritarie co-territoriali, mentre i parlanti delle varie lingue maggioritarie co-territoriali non apprendono la romaní.

In questo quadro, gli schemi di utilizzo della romaní sono visti come posti lungo un continuum tra la sfera privata e la sfera pubblica (cfr. Halwachs, 2003; Matras, 2002). In generale, l'uso della romaní è ridotto alla sfera privata (o basilettica), come lingua della comunicazione orale all'interno della famiglia. Sono tuttavia poche (o forse nessuna) le comunità nelle quali gli usi basilettici vengono coperti esclusivamente dalla romaní; in molte è comune una situazione di “multilinguismo basilettico” (Halwachs, 2003) in cui accanto alla romaní è anche presente, a diversi livelli, la/e lingua/e di maggioranza. D'altra parte, il dominio pubblico (o acrolettico) è quasi esclusivamente riservato alla lingua di maggioranza. Gli usi intermedi o mesolettici, legati a contesti semi-informali o ad attività semi-pubbliche quali il lavoro o l'interazione con conoscenti al di fuori della cerchia familiare, possono essere coperti sia dalla lingua di maggioranza sia dalla romaní, in maniera variabile a seconda delle aree. Nei Balcani, ad esempio, non è infrequente che la romaní assuma estese funzioni mesolettiche, mentre nell'Europa occidentale tali funzioni sono tipicamente assegnate alla/e lingua/e di maggioranza (cfr. Matras, 2002). Nuovi domini per l'uso della romaní che hanno attirato l'attenzione degli studiosi sono il web (Halwachs, 2011; Leggio, 2015, 2020) e le cerimonie religiose evangeliche, officiate da pastori rom (van den Heuvel, 2020). Volendo generalizzare molto, nella Tabella 1 viene presentato un modello astratto dei repertori comunitari dei rom europei, sulla base di Halwachs (2003).

ACROLETTO	Lingua/e maggioritaria/e	Comunicazioni con le istituzioni e le autorità, a scuola, nei media, ecc.
MESOLETTO	Lingua/e maggioritaria/e (romaní)	Comunicazioni con la propria cerchia di amici e conoscenti, a lavoro, ecc.
BASILETTO	Romaní (lingua/e maggioritaria/e)	Domini privati informali, in famiglia o con amici.

Tabella 1: Domini d'impiego della romaní (da Halwachs, 2003).

Infine, è importante notare che, presso alcune comunità, la romaní “flessa”⁴ non è più utilizzata, ma è stata sostituita da varietà della/e lingua/e maggioritaria/e co-territoriale/i nelle quali viene integrato un numero variabile di lessemi romaní. Queste varietà, note come para-romaní (cfr. Bakker, 2020 per una panoramica recente), sono state spesso discusse in letteratura come “lingue miste” (cfr. Capitolo 1). Le varietà para-romaní sembrano essere localizzate soprattutto nell’Europa settentrionale e occidentale (Spagna, Francia, Gran Bretagna, ma anche Grecia e Turchia). Tra le meglio documentate, possiamo citare l’angloromaní, parlata nel Regno Unito (Matras, 2010), il calò, parlato nella Penisola Iberica (Gamella et al., 2016), e la scandoromaní, parlata tra Svezia e Norvegia (Carling et al., 2013). Parte del dibattito sulle para-romaní si è concentrato sui loro scenari sociali di sviluppo (cfr. Bakker, 2020 per un’ampia panoramica recente). Un’idea oggi piuttosto diffusa è che esse emergano come risultato di processi relativamente consci, soprattutto tra gruppi inseriti in specifiche nicchie economiche (ad esempio lo spettacolo viaggiante), nel contesto di complessi processi di negoziazione identitaria con altri gruppi professionali (cfr. Matras et al., 2024). Normalmente, nonostante la loro connessione con le lingue miste, si ritiene che le para-romaní non rappresentino *lingue* di per sé, ma piuttosto stili o registri delle lingue co-territoriali, di cui “sfruttano” la morfologia sintassi. Esse non sembrano infatti essere trasmesse come lingue materne e non sono in grado di adempiere a tutti i bisogni comunicativi dei parlanti, ma piuttosto vengono attivate durante lo svolgimento di particolari attività professionali o per veicolare un senso di solidarietà con gli altri membri della comunità (Matras, 1998b, 2010; Matras et al., 2024).

2.3 La lingua romaní in Italia

Dopo aver discusso aspetti generali della lingua romaní, in questa sezione ci addentriamo più nello specifico della situazione italiana. In particolare, il § 2.3.1 offre una panoramica sulle comunità rom e sinte in Italia, mentre il § 2.3.2 si concentra sulle principali fonti per lo studio delle varietà di romaní in Italia. Infine, il § 2.3.3 propone una panoramica aggiornata della situazione sociolinguistica dei principali gruppi rom e sinti presenti in Italia oggi.

2.3.1 Rom e sinti in Italia

In un rapporto del Consiglio d’Europa del 2012⁵, il numero di persone rom e sinte presenti in Italia veniva stimato tra le 120.000 e le 180.000 unità. La presenza di questi gruppi nel territorio italiano è secolare, se si considera che la loro prima attestazione nella penisola è datata al 1422 (Piasere, 2009).

⁴ Con questo termine, in inglese *inflected Romani*, gli studiosi identificano le varietà di romaní che hanno mantenuto la grammatica e in particolare la morfologia flessiva ereditata.

⁵ <https://www.coe.int/en/web/roma-and-travellers/publications>; ultimo accesso: 09/03/2026.

Come nota Piasere (1992), sembra esistere una distinzione piuttosto netta tra i vari gruppi a seconda della loro zona di insediamento: mentre quanti risiedono nelle aree settentrionali si autodenominano generalmente come sinti, quanti risiedono nelle aree centro-meridionali si autodenominano come rom. Il confine tra gli etnonimi è individuato da Piasere (1992), sulla base di fonti storiche, lungo una linea immaginaria che congiunge Cannes e Trieste passando attraverso le Marche. A questi gruppi di storico insediamento si è aggiunta negli ultimi decenni una serie di nuovi gruppi, giunti in Italia in diverse ondate, ma soprattutto a seguito delle guerre balcaniche degli anni Novanta e dopo l'ingresso nell'Unione Europea di Romania e Bulgaria nel 2007 (Bontempelli, 2022).

Concentrandoci ora sulle popolazioni di antico insediamento, i rom dell'Italia meridionale sono probabilmente giunti dai Balcani via mare, già attorno al 1400-1500 d.C. Tale presenza è da intendere come parte di un più ampio flusso migratorio risultante dall'invasione ottomana dell'area balcanica, al quale hanno anche partecipato gruppi greci, albanesi e dalmati (Novi Chavarria, 2007; Pontrandolfo, 2013). Il possibile arrivo via mare dai Balcani è anche sostenuto da prove linguistiche, in quanto le varietà parlate dai rom dell'Italia meridionale sono quasi del tutto prive di prestiti di origine tedesca. Pontrandolfo (2013) distingue alcuni gruppi principali di rom italiani: rom abruzzesi, rom molisani, rom campobassani, rom melfesi, rom calabresi e *romi* di Laterza.

Quanto ai sinti dell'Italia settentrionale, sulla base degli etnonimi essi possono essere distinti in due ulteriori macro-gruppi: i sinti "italiani", residenti nell'Italia nordoccidentale, e i sinti "tedeschi", residenti prevalentemente nel nord-est, nelle aree pedemontane del Veneto e del Trentino Alto-Adige. Questi ultimi si definiscono e sono definiti dai sinti "italiani" come *teič* 'tedeschi' o *gadžkane*, a partire dalla parola *Gadžkano* che nella romaní dei sinti significa 'Germania' (Piasere, 1992: 250). In generale, gli studiosi sono concordi nel ritenere che entrambi i gruppi sinti siano giunti in territorio italiano da aree germanofone, come testimonia l'abbondante presenza di lessico tedesco nelle loro varietà (Piasere, 1992). La differenza tra i due gruppi risiederebbe nel periodo di arrivo in Italia. Come nota Piasere (1992), i sinti tedeschi sarebbero giunti nei territori dell'attuale Italia nordorientale tra il 1800 e il 1900, nel contesto di un più ampio movimento che nel medesimo periodo ha portato gruppi di sinti aventi come baricentro la Germania a sconfinare in territori limitrofi. D'altra parte, l'insediamento dei sinti "italiani" è probabilmente avvenuto prima, tra il 1600 e il 1700.

I sinti "italiani" possono essere distinti in ulteriori gruppi in base alla loro regione di insediamento; in alcuni casi, è anche possibile distinguere gruppi ancora più circoscritti sulla base della città di residenza o che costituisce il nucleo delle attività economiche. In tal senso, è possibile per esempio distinguere tra sinti piemontesi, sinti lombardi, sinti emiliani, sinti marchigiani, sinti reggiani, ecc. A questi è anche da aggiungere il gruppo degli *shinte rosengre*, la cui storia è oltremodo interessante: l'esistenza degli *shinte rosengre* è infatti documentata unicamente nei manoscritti di un tale

Sigismondo Caccini, rinvenuti negli anni Novanta dall'antropologo Leonardo Piasere (cfr. Piasere, 1996; Meli, 2022). Dall'analisi di queste carte emerge che Sigismondo Caccini era un uomo non sinto che si è unito agli *shinte rosengre* tra gli anni Ottanta e Novanta dell'Ottocento, forse a seguito di un legame matrimoniale. Vivendo con questo gruppo, si era anche occupato di scrivere una grammatica della loro lingua. Dagli scritti di Caccini, si deduce come gli *shinte rosengre* praticassero un nomadismo a corto raggio nell'Italia centrale, con una *core-area* nella val di Chiana e valle del Tevere, costituendo pertanto il nucleo più meridionale dei sinti italiani (Meli, 2022). Oggi non sembrano più darsi locutori che proseguano questa varietà.

I sinti "tedeschi" sono invece solitamente distinti in vari gruppi i cui etnonimi rimandano all'area storica di provenienza: abbiamo per esempio i sinti *krassarja* 'del Carso', i sinti *estrexarja* 'austriaci' e i sinti *kranarja* 'della Carnia'; a questi si aggiungono i *sinti eftavagarja*, letteralmente 'delle sette carovane'.

2.3.2 La romaní in Italia: fonti e studi

Il primo tentativo di documentazione di una varietà romaní parlata in Italia si deve a Graziadio Isaia Ascoli: in una sua opera del 1865, intitolata *Zigeunerisches*, l'autore inserisce una sezione denominata *Zigeunerisches aus Süd Italien*, realizzata sulla base di interviste con rom di area molisana (cfr. Pontrandolfo & Piasere, 2002 per la traduzione in italiano). Poco dopo, nel 1889, Adriano Colocci pubblica un dizionario della romaní dei sinti marchigiani (cfr. Colocci, 1889), tuttavia di difficile collocazione (Soravia, 1977). Risale invece al 1936 il noto "rilievo zingaresco" di Ugo Pellis (cfr. Pellis, 1936), un'inchiesta dialettale per il punto 603 dell'Atlante Linguistico Italiano risultante dall'intervista con una donna rom presso l'Annunziata di Giulianova, in provincia di Teramo. In appendice al suo "rilievo zingaresco", Pellis inserisce inoltre una quarantina di lemmi romaní presenti nel gergo parlato dai "cavallari" di Guardiagrele (comune in provincia di Chieti), noto come *baccàì* (cfr. anche Scala, 2014). Ciò si iscrive in un più ampio interesse per lo studio delle voci romaní all'interno dei gerghi di mestiere e della malavita italiani, testimoniato da altri lavori del periodo come Pasquali (1935) e Tagliavini & Menarini (1938) e, successivamente, Cortelazzo (1975) e Soravia (1977, in appendice alla sua grammatica descrittiva).

Un ruolo chiave nella documentazione delle varietà romaní in Italia è poi stato assunto dalla rivista *Lacio Drom* (traducibile come 'Buon Cammino') che per oltre 30 anni, dal 1965 al 1999, ha ospitato contributi relativi a svariati aspetti della lingua e cultura rom e sinta in Italia e altrove. Tra i contributi linguistici, possiamo citare glossari relativi a varietà individuali, poi confluiti nel *Dizionario Sinottico delle Parlate Zingare in Italia* curato da Giulio Soravia e Camillo Fochi nel 1995 (cfr. Soravia & Fochi, 1995).

Il primo studio comprensivo sulla romaní parlata in Italia è fornito da Soravia (1977), il quale si occupa di distinguere diverse varietà di romaní collocate in varie regioni e di descriverne le principali caratteristiche grammaticali. Le varietà romaní individuate da Soravia (1977) sono le seguenti:

- Sinto piemontese: la sua *core-area* è in Piemonte, ma i parlanti di questa varietà traboccano i confini regionali, estendendosi fino in Toscana, Lombardia, Emilia e Liguria; inoltre, Soravia nota che il sinto piemontese è strettamente collegato a varietà parlate nella Francia Meridionale, di cui rappresenta un ramo “al di qua” delle Alpi;
- Sinto lombardo: parlato in Lombardia, con sconfinamenti in Emilia;
- Sinto delle Tre Venezie, ovvero la varietà parlata dai sinti “tedeschi” precedentemente menzionati tra Trentino Alto-Adige e Veneto, alle pendici delle Alpi;
- Sinto istriano: si tratta di una varietà parlata da gruppi originari dell’Istria e residenti in maniera capillare in Italia, particolarmente nel nord-est;
- Rom abruzzese: parlato da gruppi residenti nei principali centri dell’Abruzzo e delle Marche meridionali, con sconfinamenti nel Lazio;
- Rom calabrese, parlato in Calabria;
- Dialetti vlax parlati da gruppi rom di recente insediamento provenienti dalla Romania.

Per quanto di notevole importanza, questa classificazione necessita di alcune precisazioni. Innanzitutto, la designazione “sinti delle Tre Venezie” può risultare ambigua, in quanto rende opaca la distinzione tra i sinti “tedeschi” (*estrexarja*, *eftavagarja* ecc.), a cui Soravia (1977) fa effettivamente riferimento, e i sinti “italiani” residenti in Veneto, non citati da Soravia e della cui lingua si sa tutt’oggi molto poco (Scala, 2012). I sinti veneti italiani costituiscono probabilmente un gruppo strettamente legato al mondo dello spettacolo viaggiante, che ha subito un processo di perdita della romaní e ha adottato come codice endocomunitario varietà italo-romanze locali (cfr. § 2.3.3). In secondo luogo, la denominazione sinto istriano può essere fuorviante, perché, come riconosciuto dallo stesso Soravia (1977), si tratta di una varietà in realtà molto affine ai dialetti arli parlati nei Balcani meridionali e appartenenti al ramo balcanico. Per raggiungere una maggiore completezza, a questa classificazione andrebbe aggiunto il già citato shinto rosengro (per quanto oggi estinto) che, linguisticamente, sembrerebbe rispecchiare la sua posizione geografica intermedia tra dialetti sinti e dialetti rom mostrando tratti condivisi con l’uno e con l’altro gruppo (cfr. Meli, 2022 per una recente descrizione dettagliata della lingua).

Dopo il contributo di Soravia (1977), le varietà romaní italiane sono state oggetto di alcuni studi individuali. Per esempio, si possono menzionare i lavori di Franzese (2023, [2001, 1985]) volti alla descrizione del sinto piemontese, mentre per la romaní abruzzese si trovano diversi lavori di Scala (cfr. p.es. Scala, 2023c). Scala (2023a, 2023b) propone alcune riflessioni sul rapporto tra sinto

piemontese parlato in Piemonte e sinto piemontese parlato in Francia, anche in prospettiva comparativa con il sinto lombardo. Su quest'ultimo, si possono citare i lavori di Scala (2017), focalizzato sul sistema dei numerali in questa varietà, Scala (2021), che presenta un etnotesto in sinto lombardo con alcuni appunti grammaticali, e Sorrenti (2014), che offre una grammatica descrittiva di questa varietà focalizzandosi sugli elementi di maggiore distanza dall'italiano. Meli (2023) analizza in chiave comparativa le strategie di integrazione linguistica dei prestiti italo-romanzi in sinto lombardo e piemontese, uno dei pochi studi esplicitamente incentrati sul contatto linguistico in varietà sinte. Infine, Leggio (2011) offre una descrizione grammaticale della varietà di romaní gurbet parlata da rom kosovari residenti a Palermo, uno dei pochi lavori focalizzati sulle varietà romaní parlate da gruppi rom di immigrazione recente.

Oltre ai materiali raccolti dai linguisti, se ne possono citare altri, talvolta prodotti da rom e sinti. Sul sinto piemontese, Franzese (1985) menziona la raccolta di brevi racconti autobiografici del sinto piemontese Annibale Niemen intitolata *O ker kun le penija* ('La casa con le ruote'), che fu attivo come partigiano durante la Seconda Guerra Mondiale, nonché gli album discografici del cantante Lick Dubois⁶. Di notevole interesse è la registrazione di una messa celebrata in sinto piemontese, disponibile su YouTube⁷. Quanto al sinto lombardo, è importante menzionare la traduzione in questa varietà del Vangelo di Marco ad opera di don Mario Riboldi. Per la romaní abruzzese, si dispone delle produzioni di intellettuali appartenenti alla comunità come Bruno Morelli e Santino Spinelli; a titolo di esempio, si considerino le poesie raccolte in Spinelli (1993). Inoltre, può essere apprezzata una discreta produzione musicale in lingua romaní. Per il sinto lombardo, è possibile menzionare il gruppo *Pisa da Lacio* ('Poco di Buono'), formato da sinti di Pavia, mentre per il sinto o *gadžkano* possiamo menzionare il gruppo *U Romane Ciave* ('I ragazzi romaní'), di stanza a Tortona⁸. Ricchissima è poi la produzione musicale in romaní abruzzese grazie agli album dell'*Alexian Group*, guidato da Santino Spinelli⁹. Infine, i *social media* costituiscono oggi un'interessante fonte di osservazione dell'uso spontaneo e semi-spontaneo della romaní italiana (cfr. a questo proposito gli esempi riportati in Scala, 2015).

2.3.3 Aspetti sociolinguistici

Dopo aver discusso le principali fonti e studi che documentano le varietà di romaní parlate in Italia, possiamo infine a discutere l'attuale situazione sociolinguistica dei gruppi di rom e sinti residenti

⁶ Le cui tracce musicali possono essere ascoltate al seguente indirizzo: https://www.progettoniglo.org/Sinti_Song.htm; ultimo accesso: 12/02/2026.

⁷ <https://www.youtube.com/watch?v=9JELYlqsb8c&t=550s>. Ultimo accesso: 12/02/2026.

⁸ Cfr. ad esempio <https://www.youtube.com/watch?v=nCL7VyLhzbs>; ultimo accesso: 12/02/2026.

⁹ Cfr. ad esempio https://www.youtube.com/watch?app=desktop&v=7R0fAB7G_RQ; ultimo accesso: 12/02/2026.

nella Penisola. A tal fine, ci concentriamo in particolar modo sulla conformazione del loro repertorio linguistico (cfr. Capitolo 1 per una definizione della nozione).

Partendo dai sinti piemontesi, oggi il loro repertorio linguistico è caratterizzato dall'uso dell'italiano come codice esocomunitario e dall'uso del dialetto italoromanzo co-territoriale, il piemontese, come codice endocomunitario (Scala, 2012; Duberti, 2010). In questo scenario, la romaní (nella varietà del sinto piemontese) risulta fortemente minacciata. Si considerino, per esempio, le osservazioni di Sergio Franzese nell'introduzione alla terza edizione della sua grammatica descrittiva di sinto piemontese, pubblicata nel 2021:

Eccoci infine giunti, a ben 35 anni dalla prima pubblicazione, alla terza edizione di questa grammatica di sinto piemontese. Molte cose sono cambiate da allora, il processo di perdita della lingua materna tra i Sinti Piemontesi in Piemonte è proseguito inesorabilmente fino a costituire purtroppo una triste e irreversibile realtà; anche tra le comunità attestate in Francia le cose paiono non andare meglio, nonostante alcune “sacche di resistenza”. La revisione e la ristampa di questa grammatica potrebbero pertanto non avere senso se il suo scopo fosse unicamente quello di recuperare l'uso della lingua sinta tra le giovani generazioni. (Franzese, 2021: 1)

Viceversa, tra i sinti lombardi, la romaní (nella varietà del sinto lombardo) sembra essere oggi ancora ben conservata. Per questa comunità l'italiano è oggi naturalmente il codice esocomunitario, mentre la trasmissione del dialetto lombardo risulta poco o per nulla incentivata (Scala, 2012).

Scala (2012) connette la divergente situazione repertoriale dei sinti piemontesi e dei sinti lombardi ai mutamenti sociolinguistici più ampi avvenuti in Italia nel secondo dopoguerra, quando ai dialetti italoromanzi locali si è affiancato e gradualmente sostituito l'italiano anche negli usi orali quotidiani. In questo scenario, i parlanti di sinto lombardo e piemontese si sarebbero trovati a gestire tre lingue: un dialetto italoromanzo, l'italiano e la romaní. A fronte di questa situazione, i sinti dei due gruppi hanno adottato strategie diverse: i sinti piemontesi hanno mantenuto il dialetto piemontese, rifunzionalizzandolo però come codice endocomunitario a scapito della romaní, mentre i sinti lombardi hanno mantenuto la romaní a scapito del dialetto lombardo. Tuttavia, entrambi i gruppi sono rimasti significativamente bilingui, utilizzando come codice endocomunitario una lingua non utilizzata (nel caso dei lombardi) o sempre meno utilizzata (nel caso dei piemontesi) dalla popolazione maggioritaria co-territoriale.

Duberti (2010) riporta che i sinti piemontesi possono attivare un particolare registro del piemontese caratterizzato da inserzioni lessicali romaní in una cornice grammaticale romanza. Si tratta di un fenomeno osservato in altre comunità di romaní europee e noto con il nome di para-romaní (cfr. § 2.2.3 in questo capitolo). A causa dell'esiguità degli studi in materia, tuttavia, si sa molto poco a proposito di questo particolare registro misto, né dal punto di vista strutturale (per esempio l'ampiezza

del corpus di parole romaní ancora disponibile ai parlanti), né dal punto di vista sociolinguistico e pragmatico (ossia la tipologia di utenti che ne fa uso e le sue funzioni comunicative).

Per quanto riguarda i sinti emiliani e veneti, Scala (2012) osserva per entrambi i gruppi una situazione di perdita della romaní a favore di una varietà italo-romanza, similmente a quanto osservato nei sinti piemontesi. In un recente intervento presso *l'International Conference on Romani Linguistics* di Vienna, Ziano & Trevisan (2025) hanno osservato che la varietà italo-romanza parlata dai sinti emiliani rappresenta non già un dialetto emiliano, bensì un dialetto veneto, grossomodo riconducibile all'area del Polesine ma caratterizzato da alcuni tratti irriducibili a una specifica varietà, indicando un processo di *levelling*.

Purtroppo, non conosciamo quasi nulla delle comunità sinte del Trentino Alto-Adige. Scala (2012: 144) traccia una situazione di trilinguismo, osservando che queste verosimilmente “parlano l'italiano di Bolzano, il tedesco del Südtirol e con ogni probabilità conservano ottimamente la romaní”. Tuttavia, Tauber (2008: 157-158) riporta un atteggiamento negativo da parte dei sinti dell'Alto Adige verso la comunità tedescofona della zona e verso il dialetto tedesco, il che porta a una conoscenza e a un uso attivo dell'italiano e della romaní, ma in misura assai minore del dialetto tedesco, nonostante la popolazione maggioritaria co-territoriale sia dominante in questa lingua.

Quanto ai rom dell'Italia meridionale, presso i rom abruzzesi la romaní appare ben conservata (Scala, 2023c; Agresti, 2023), accanto all'italiano regionale della zona e, verosimilmente, anche al dialetto abruzzese. Presso i rom calabresi, invece, un processo di perdita della romaní a favore del dialetto calabrese è già riportato in Soravia (1977). In un recente intervento, Cosentino (2023) riporta che il codice endocomunitario utilizzato dai rom calabresi residenti a Cosenza è una varietà italo-romanza (definita *zingarišku* dai suoi parlanti) che mostra alcune peculiarità rispetto agli altri dialetti della Calabria settentrionale. Infine, a mia conoscenza, sono del tutto assenti studi relativi alla composizione dei repertori dei gruppi rom di arrivo più o meno recente dall'Europa orientale.

In sintesi, una conoscenza dettagliata dei repertori linguistici dei sinti e rom di antico e nuovo insediamento in Italia rimane un obiettivo non ancora pienamente raggiunto. Tuttavia, dalla breve rassegna proposta in questa sezione, sembrano emergere alcune tendenze comuni di notevole interesse. Da una parte, si assiste alla perdita della romaní, spesso sostituita da dialetti italo-romanzi, che garantiscono così il mantenimento di repertori plurilingui e una distinzione ben chiara tra codici endocomunitari ed esocomunitari (cfr. Scala, 2012). Dall'altra, i dialetti adottati come codici endocomunitari spesso non coincidono esattamente con quelli parlati dalla popolazione maggioritaria co-territoriale e possono recare tratti di livellamento. Complessivamente, questi aspetti confermano le complesse dinamiche identitarie che modellano la strutturazione del repertorio linguistico in un gruppo come quello dei rom e dei sinti in Italia

Capitolo 3. Presentazione dell'area di ricerca, metodologia generale d'indagine e dati

Poiché il principale obiettivo di questo volume è indagare la composizione del repertorio linguistico dei sinti lombardi e analizzarne gli usi linguistici quotidiani, all'avvio della ricerca si è optato per una metodologia che permettesse di osservare e raccogliere i fenomeni di interesse nel loro contesto naturale, avviando quindi una collaborazione diretta con la comunità sinta di Pavia. In questo capitolo, viene introdotta la comunità dei sinti lombardi di Pavia da un punto di vista etnografico (§ 3.1) e viene descritta la metodologia che ha guidato la ricerca sul campo e la raccolta dati (§ 3.2). Il § 3.3 offre poi una panoramica quantitativa dei dati raccolti e del campione dei partecipanti, mentre il § 3.4 illustra più nello specifico l'annotazione dei dati di parlato.

3.1 La comunità sinta di Pavia

La comunità sinta della città di Pavia è distribuita in due insediamenti (cfr. Figura 2). Il più grande si trova nelle immediate vicinanze del centro cittadino, in piazzale Europa, una zona di proprietà pubblica abitata da famiglie sinte a partire dal 1985. Il secondo insediamento, di dimensioni più modeste, sorge invece nella prima periferia sud della città, in via Bramante.

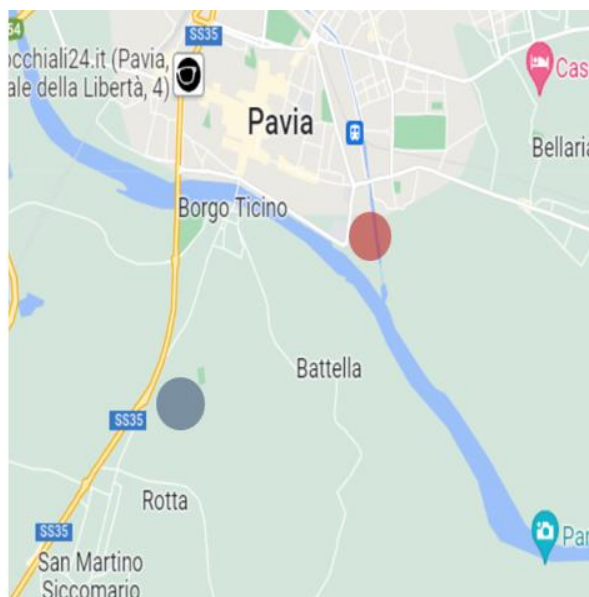


Figura 2: Collocazione degli insediamenti sinti nel comune di Pavia: il cerchio rosso identifica l'insediamento di piazzale Europa, mentre il cerchio grigio l'insediamento di via Bramante.

Circa quindici anni fa, Membretti (2009) stimava la presenza di circa 400 persone sinte a Pavia, segnalando difficoltà nell'individuare un numero più preciso anche a causa della presenza di famiglie residenti negli insediamenti in maniera informale. Un censimento operato dal comune di Pavia nel 2018¹ ha confermato le stime di Membretti, rivelando la presenza di 398 persone sinte distribuite in 130 nuclei familiari tra piazzale Europa e via Bramante. Questo censimento fornisce anche informazioni più dettagliate sulla composizione socio-anagrafica dei due insediamenti, riportate nelle Tabelle 2 e 3.

	Donne	Uomini	Totale
Europa	126	139	265
Bramante	68	65	133
Totale	194	204	398

Tabella 2: Composizione della comunità di sinti pavesi per genere.

	<18	18-65	65+	Totale
Europa	105	152	8	265
Bramante	44	81	8	133
Totale	149	233	16	398

Tabella 3: Composizione della comunità di sinti pavesi per fascia d'età.

Pur trattandosi di dati da considerare con cautela, essi mostrano alcune peculiarità demografiche della comunità sinta di Pavia, ad esempio l'incidenza molto elevata dei minori di diciotto anni sul totale della popolazione (quasi il 40%) e, di converso, la scarsa presenza di persone di più di 65 anni (circa il 4%).

I sinti di piazzale Europa e di via Bramante definiscono sé stessi come *lombardi*, rivendicando un rapporto storico con il territorio della Lombardia occidentale. La loro storia meno recente, come accade per i sinti e rom più in generale, è tuttavia difficile da ricostruire con precisione. Come emerge dalle testimonianze dei parlanti stessi e da alcune fonti (cfr. p.es. Membretti, 2009; Tribulato & Pastori, 2022), fino agli anni Sessanta le forme di sostentamento dei sinti lombardi erano prevalentemente inserite in contesti rurali e caratterizzate da economie informali e stagionali, in parte anche basate sull'attività dell'elemosina. A partire dagli anni Sessanta, un numero crescente di sinti ha poi trovato nello spettacolo viaggiante un ambito lavorativo relativamente stabile. Questa opportunità è stata favorita dalla Legge Corona del 1968, che riconosce la funzione sociale dei circhi

¹ <https://laprovinciapavese.gelocal.it/pavia/cronaca/2019/07/28/news/rom-e-sinti-in-provincia-i-comuni-hanno-iniziato-il-censimento-salvini-1.37234307>; ultimo accesso 10/06/2024.

equestri e dello spettacolo viaggiante e ne demanda la regolamentazione direttamente alle amministrazioni comunali (cfr. anche Tribulato, 2020). I sinti lombardi hanno così mantenuto il loro stile di vita itinerante, ma la natura di tali spostamenti è mutata notevolmente, passando dalle aree rurali ai centri urbani di grande e media estensione, determinando una ridefinizione del rapporto con il territorio circostante e un contatto più intensivo con la popolazione maggioritaria co-territoriale (si vedano, in questo senso, le interessantissime pagine del diario del sinto Jucki Herzerembergher, riportato integralmente in Tribulato & Pastori, 2022).

Il progressivo declino delle attività legate allo spettacolo viaggiante tra gli anni Ottanta e Novanta ha poi portato a una progressiva sedentarizzazione dei sinti lombardi di Pavia nei campi di sosta urbani menzionati precedentemente. In questo contesto, i sinti lombardi di Pavia si dedicano prevalentemente alla raccolta del ferrovecchio, alla vendita ambulante, al bracciantato agricolo su base stagionale o a servizi di pulizia per conto di enti pubblici e privati. Alcuni membri della comunità hanno inoltre contratti come operai in fabbriche locali. Tuttavia, il tasso di disoccupazione rimane elevato. Tra i giovani in età scolare, anche il tasso di abbandono scolastico è molto alto: sebbene molti completino la scuola secondaria di primo grado, pochissimi proseguono fino a conseguire un diploma di scuola superiore (cfr. anche Rossi, 2020; Laganà, 2024).

La condizione di marginalità socioeconomica di molti sinti lombardi limita la loro mobilità sociale. A ciò si aggiunge una mobilità geografica relativamente ridotta, anche a causa dell'importanza attribuita ai legami familiari e del desiderio di non separarsi dal proprio nucleo. La mobilità geografica, quando si verifica, può essere connessa a legami matrimoniali: è infatti relativamente comune che i sinti lombardi si sposino con altri sinti di città vicine e che, dopo il matrimonio (sancito attraverso una breve fuga, definita *fuitina*, dalla famiglia), la moglie si trasferisca nell'insediamento in cui risiede la famiglia del marito, continuando a visitare frequentemente la famiglia d'origine.

Oltre a condividere gli spazi abitativi, i sinti lombardi trascorrono molto tempo insieme anche fuori dall'insediamento, sia durante il tempo libero che nello svolgimento delle attività lavorative: ad esempio, la raccolta del ferrovecchio viene tipicamente svolta in piccoli gruppi dagli uomini della comunità. Anche i culti religiosi si svolgono nelle chiese situate all'interno degli insediamenti e non prevedono interazioni con la popolazione maggioritaria, nonostante i giovani sinti di piazzale Europa frequentino da generazioni (soprattutto in passato) l'oratorio del quartiere in cui il loro insediamento è ubicato. D'altra parte, i contatti con la popolazione maggioritaria fuori dal campo sono frequentissimi e avvengono su base quotidiana, in contesti come il lavoro, la scuola, le attività sportive (che riguardano soprattutto gli uomini), oltre che nelle interazioni con le amministrazioni e i commercianti della città. Tuttavia, questi contatti sono effimeri, salvo nei casi (non frequentissimi, ma documentati) di matrimoni con non sinti.

In sintesi, la comunità sinta di Pavia costituisce un ottimo esempio di comunità caratterizzata da reti sociali “dense” e “multiple” (Milroy, 2002). Per reti sociali dense, intendiamo il fatto che i rapporti tra i membri della comunità sono caratterizzati da un alto livello di “transitività”: dato un soggetto al centro della rete, le persone che questo/a conosce tendono a conoscersi a loro volta. Per reti sociali multiple, intendiamo il fatto che i membri della rete interagiscono tra loro in più ruoli: per esempio, il vicino di casa di un soggetto può anche essere un suo parente o un collega al lavoro.

Fornito questo quadro generale, procediamo ora ad analizzare in maggiore dettaglio le caratteristiche di ciascuna comunità. La comunità di piazzale Europa è costituita grossomodo da un unico nucleo familiare allargato: i suoi membri sono i figli, i nipoti e i pronipoti di Mafalda Navone (sinta di origine piemontese) e Guarino Casagrande (sinto di origine marchigiana), considerati i capostipiti della comunità, i cui primi insediamenti stabili nella città di Pavia si datano, come anticipato, al 1985. In particolare, Mafalda Navone, deceduta nel 1998, è ricordata come la *leader* indiscussa della comunità dei sinti pavesi. Ciò le è valso l'appellativo di “regina”, con il quale è anche divenuta nota alla comunità pavese maggioritaria, a testimonianza di un lungo rapporto di coesistenza e di (lenta) conoscenza reciproca tra la comunità sinta e la comunità cittadina circostante². Le storie e gli aneddoti sulla regina Mafalda sono numerosi a piazzale Europa, dove è ricordata come figura chiave per l'integrazione della comunità nel tessuto sociale della città. I sinti di piazzale Europa sono prevalentemente cattolici e partecipano alla messa celebrata da un sacerdote non sinto presso una chiesetta collocata nell'insediamento.

Più nello specifico, l'insediamento di Piazzale Europa è ulteriormente suddiviso in due insediamenti, separati da una piazza in cui sorge il Palazzo delle Esposizioni, luogo in cui vengono organizzati abitualmente eventi cui partecipa la cittadinanza pavese. Durante la mia ricerca sul campo ho avuto modo di collaborare quasi esclusivamente con gli individui residenti nell'insediamento posto più vicino al fiume Ticino, alla sinistra del Palazzo delle Esposizioni.

E' importante notare che alcuni dei sinti di piazzale Europa (in particolare quelli dell'insediamento più vicino al fiume Ticino, con cui ho avuto modo di collaborare) si autodefiniscono, in aggiunta a *lombardi*, come *mucini* (cfr. Partisani, 1984, che all'interno della rivista *Lacio Drom* propone una breve descrizione di una varietà di romaní da lui definita ‘sinto mucino’, lombarda nella grammatica e piemontese nel lessico). Come osservato anche da Scala (2012), nel microcosmo sinto questo termine assume talvolta una connotazione dispregiativa (almeno per chi non si identifica con esso). Alcuni parlanti intervistati (che non si identificano come *mucini* e utilizzano questo appellativo in

² Si notino a questo proposito i titoli di alcuni articoli del quotidiano locale *La Provincia Pavese*: <https://laprovinciapavese.gelocal.it/pavia/cronaca/2017/05/19/news/oliviero-toscani-dai-sinti-per-il-reportage-su-mafalda-1.15359972>; <https://laprovinciapavese.gelocal.it/pavia/cronaca/2012/04/10/news/piazza-europa-ecco-i-figli-di-mafalda-1.3795674>; ultimo accesso: 17/09/2024.

sensu deteriore) collegano l'etimologia del termine alla radice lessicale *mucin*, indicante il “fondo della sigaretta” che i sinti *mucini* raccoglierebbero da terra per fumare, non potendo permettersi di acquistare normali pacchetti di sigarette.

La comunità di via Bramante, meno numerosa di quella di piazzale Europa, è invece costituita da gruppi familiari più autonomi, che non si identificano in un unico capostipite. Generalmente, le famiglie con cui è stato possibile collaborare riportano legami storici, spesso connessi alle attività economiche dello spettacolo viaggiante, con le province di Pavia, Milano, Varese (area di Busto Arsizio) e Bergamo. A differenza dei sinti di piazzale Europa, molti membri della comunità di via Bramante sono evangelici, avendo aderito alla MEZ (“Missione Evangelica Zigana”), un movimento religioso nato dal basso tra i rom e i sinti stessi, oggi ben organizzato in Italia e in Europa³. La piccola chiesa evangelica posta al centro dell'insediamento, dove la messa viene celebrata settimanalmente, rappresenta un importante punto di ritrovo per i sinti di via Bramante.

Per i sinti di via Bramante, l'appartenenza alla confessione evangelica costituisce un tratto molto significativo per la percezione di sé stessi, anche in opposizione agli altri sinti. La *conversione* all'evangelismo, sancita attraverso un battesimo, impone rigide norme di condotta, come la rinuncia alle “cose del mondo” (*u koa du veltu*) quali ad esempio l'alcol, il tabacco e i piccoli crimini, spesso associate dai sinti evangelici ai sinti “non convertiti”, cioè cattolici. Dal canto loro, i sinti cattolici possono percepire la conversione all'evangelismo come una scelta eccessivamente radicale e settaria. Infine, sulla base dell'osservazione partecipante condotta, mi sembra possibile sostenere che presso i sinti di via Bramante viga un senso di maggiore identificazione con il mondo dello spettacolo viaggiante rispetto ai sinti di piazzale Europa. In tal senso, i sinti di via Bramante sembrano riconoscersi come più prossimi ai sinti “dei mestieri” (il termine *mestiere* viene usato dai sinti nel significato di *giostre*) che ai sinti “da campo” – un'etichetta, quest'ultima, che tende a coincidere con quella di *mucini*, assegnata ai sinti di piazzale Europa (sulla differenza tra “sinti dei mestieri” e “sinti da campo”, cfr. anche Tribulato, 2019). L'auto-associazione dei sinti di via Bramante con i *mestieri* sembra anche coincidere con l'idea di una loro maggiore predisposizione all'imprenditorialità e alla gestione delle risorse economiche rispetto ai sinti *mucini*.

Nel complesso, la comunità sinta di Pavia si configura come un contesto socialmente coeso, ma al tempo stesso segnato da profonde differenze tra i due insediamenti, tali da giustificare un'analisi sia unitaria sia distinta per sottogruppi, come verrà fatto in questo volume.

³ Il movimento pentecostale tra i rom e sinti è sorto agli inizi degli anni Cinquanta in Francia, con le predicazioni del pastore evangelico Clement Le Cossec (cfr. van den Heuvel, 2020: 464 per una panoramica generale), per poi diffondersi rapidamente altrove in Europa. In Italia, le prime riunioni ufficiali della Missione Evangelica Zigana si datano ai primi anni Ottanta. Per una panoramica sui sinti italiani e il pentecostalismo, cfr. p.es. Simonelli (1996), nonché il sito web <https://www.sastajnas.it/Mezadi/index.html> (ultimo accesso: 08/02/2026).

Infine, è importante ricordare che nella città di Pavia risiedono anche gruppi rom di migrazione più recente, prevalentemente originari della Romania (zona di Craiova). Sebbene alcuni gruppi rom siano attestati nella città già dagli anni Sessanta del Novecento, la loro presenza si è decisamente infittita a partire dai primi anni Duemila, con l'ingresso della Romania nell'Unione Europea (Membretti, 2009). Attualmente, molti rom a Pavia risiedono in appartamenti comunali sparsi nella città. Le frequentazioni tra la comunità sinta e la comunità rom sono pressoché inesistenti e non si osserva, da parte dei sinti, la valorizzazione di una comune origine (comunque nota) con i rom. Questi ultimi, infatti, sono spesso definiti dai sinti come “slavi” o “rumeni” (in opposizione alla propria “italianità” storica) e considerati portatori di un'identità culturale piuttosto diversa.

3.2 Il lavoro sul campo e la raccolta dati

L'approccio alla ricerca sul campo e alla raccolta dei dati si è basato, sin dalla sua concezione iniziale, su una combinazione di metodi propri di diverse discipline delle scienze linguistiche, in particolare la documentazione linguistica, l'etnografia e la sociolinguistica variazionista (cfr. Gorla & Gasparini, 2025: 40–41). In linea con l'approccio della documentazione linguistica, la priorità è stata la realizzazione di un corpus di parlato che cercasse di cogliere, per quanto possibile, il parlato spontaneo e irriflesso dei locutori. Tale obiettivo è stato perseguito in accordo con i principi della sociolinguistica variazionista, che sottolinea la necessità di lavorare con un campione di parlanti il più possibile bilanciato e stratificato sociolinguisticamente (ma sulle difficoltà connesse all'applicazione di questo approccio in contesti di lingue minoritarie, cfr. Adamou, 2016). Parallelamente, l'approccio etnografico ha guidato l'intensa partecipazione di chi scrive a momenti comunitari e l'attenzione alle riflessioni metalinguistiche dei parlanti. Come vedremo, queste ultime consentono di risalire alle prospettive ‘emiche’ circa i rapporti percepiti tra le lingue del repertorio e sui fenomeni di contatto linguistico oggetto di questo volume.

Nei paragrafi seguenti, ci si focalizza nel dettaglio sui principali strumenti di raccolta dati utilizzati, ossia l'osservazione partecipante (§3.2.1), il questionario sociolinguistico (§3.2.2) e l'intervista sociolinguistica (§3.2.3). Si conclude poi con una discussione degli aspetti etici legati alla ricerca (§3.2.4).

3.2.1 L'osservazione partecipante

L'osservazione partecipante ha rappresentato sia uno strumento di accesso alla comunità, sia uno strumento conoscitivo per l'intera durata della ricerca. Per avviare la collaborazione con entrambe le comunità sinte di Pavia, si è scelto di avvalersi dell'intermediazione di enti terzi da tempo attivi nel territorio pavese. In particolare, la collaborazione con la comunità sinta di via Bramante è iniziata nel

mele di ottobre 2021 tramite la collaborazione con un'associazione di volontariato locale, Babele ODV⁴, che da diversi anni offre un servizio di doposcuola che ha luogo nella chiesa evangelica dell'insediamento. La collaborazione come volontario con Babele ODV mi ha permesso di effettuare visite settimanali alla comunità di via Bramante per circa due anni, fino al mese di settembre 2023. La collaborazione con la comunità di piazzale Europa è invece iniziata nel mese di aprile 2022, dopo aver preso contatti con un sacerdote cattolico che vive nell'insediamento insieme alla comunità sinta, nella quale svolge attività missionaria. Se il primo ingresso nella comunità di piazzale Europa è dunque avvenuto nell'ambito di attività di catechismo con ragazzi delle scuole elementari e medie, la collaborazione si è successivamente estesa ad attività di doposcuola con un gruppo più ampio di bambini e adolescenti. Queste attività sono state portate avanti in maniera continuativa fino al mese di settembre 2023.

Sin dalle prime visite a entrambe le comunità, è stata esplicitamente comunicata l'intenzione di condurre una ricerca per conto dell'Università di Pavia sulla lingua e sulla cultura dei sinti lombardi. Nelle prime fasi della ricerca, è stata fondamentale l'intermediazione di E. A., membro della comunità di via Bramante e parlante nativa di sinto lombardo, nonché membro dell'associazione di volontariato attiva nel servizio compiti. Grazie alla sua assistenza durante il servizio compiti in via Bramante, nell'arco di poche settimane sono state effettuate alcune registrazioni pilota con alcuni giovani parlanti impegnati nei compiti. L'ascolto e la trascrizione di queste registrazioni sono stati cruciali per il miglioramento della competenza di chi scrive nel sinto lombardo.

Nel procedere della ricerca, sono stati gradualmente sviluppati rapporti di fiducia con specifiche famiglie di entrambe le comunità, il che ha anche permesso la partecipazione a svariati momenti di vita comunitaria, in aggiunta ai compiti e alle attività di catechismo, favorendo la continua esposizione a pratiche linguistiche ecologicamente valide. Per esempio, è stato possibile prendere parte alla festa di Natale del 2022 presso l'insediamento di piazzale Europa; inoltre, grazie alla partecipazione alle attività di catechismo, si è potuto partecipare a svariate attività connesse alla celebrazione delle cresime e delle comunioni.

3.2.2 Il questionario

Per ottenere una panoramica dettagliata della situazione macrosociolinguistica della comunità sinta di Pavia, si è scelto di avvalersi di un questionario (cfr. Appendice 2). Il questionario era composto da 35 domande a risposta aperta o multipla, distribuite su due fogli (fronte e retro) in formato A4. Le

⁴ <https://www.babeleonlus.it/>; ultimo accesso 03/02/2026.

domande erano precedute da una sezione anagrafica in cui si chiedeva il nome del parlante, l'età, il luogo di nascita, e informazioni sul periodo di residenza a Pavia.

Le 35 domande del questionario erano volte a elicitarle, grossomodo, i seguenti aspetti:

- (1) le lingue che sono percepite dai parlanti come parte del loro repertorio;
- (2) l'autovalutazione dei parlanti in produzione orale, lettura e scrittura in sinto lombardo e in italiano;
- (3) le scelte linguistiche dei parlanti al variare di interlocutori e domini comunicativi;
- (4) gli orientamenti verso il sinto lombardo, verso il bilinguismo con l'italiano e verso l'uso simultaneo delle lingue nella conversazione, nonché verso possibili politiche linguistiche volte a rafforzare il sinto lombardo, come lo sviluppo di uno standard scritto o l'introduzione del sinto lombardo a scuola.

La maggior parte delle domande si rivolgeva all'esperienza linguistica diretta del parlante (p.es. "Quale lingua parli con i tuoi figli?") o sollecitava valutazioni o giudizi personali (p.es. "Mischiare il sinto e l'italiano quando si parla è normale?"). Di norma, mentre le domande relative all'esperienza linguistica prevedevano risposte aperte (non si voleva infatti correre il rischio di limitare le possibili risposte), le domande volte a sollecitare un giudizio proponevano risposte chiuse (del tipo "Sì" / "No" oppure "Molto", "Poco", "Abbastanza", ecc.). Soltanto in tre casi è stato richiesto ai rispondenti di riferirsi all'esperienza linguistica non propria ma altrui, in particolare con le domande "Quali lingue parlano o parlavano i tuoi genitori tra di loro?", "La lingua parlata dai tuoi genitori era o è diversa rispetto a quella che si parla oggi?" e "La lingua parlata dai tuoi nonni era o è diversa rispetto a quella che si parla oggi?". Una prima versione del questionario è stata testata con E. A. e successivamente sono state modificate alcune domande sulla base dei suoi commenti.

Il questionario è stato sottoposto ai parlanti tra i mesi di luglio e settembre 2023. Ai parlanti era data la possibilità di scegliere se compilare il questionario in autonomia o se farlo leggere e poi compilare da chi scrive sulla base delle risposte fornite. Questa seconda opzione è stata preferita dalla maggior parte dei partecipanti. Quando somministrate, talvolta le domande del questionario offrivano lo spunto per riflessioni più ampie da parte dei soggetti intervistati. Sebbene il registratore non fosse stato attivato durante lo svolgimento dei questionari per non scoraggiare la partecipazione, alcuni elementi chiave di tali riflessioni venivano appuntati a mano da chi scrive. Generalmente, la compilazione del questionario secondo le modalità descritte durava circa dieci minuti. Una volta raccolte, le risposte alle domande dei questionari sono state inserite in un foglio Excel e abbinate ai singoli parlanti e alle loro caratteristiche socio-anagrafiche.

3.2.3 Le interviste

Le interviste sociolinguistiche hanno rappresentato lo strumento per raccogliere i dati di parlato spontaneo in sinto lombardo che costituiscono l'oggetto principale di questo volume. Come anticipato, con la progressiva conoscenza della comunità e della lingua sinta, dopo pochi mesi dall'inizio della ricerca sul campo mi è stato possibile iniziare a condurre interviste sociolinguistiche in maniera autonoma. Ciò ha permesso una notevole flessibilità nel lavoro, sfruttando le disponibilità dei partecipanti indipendentemente dalla presenza della collaboratrice E. A., l'aiuto della quale era comunque subordinato alla sua personale disponibilità e non vincolato formalmente.

La lingua di base delle interviste era stabilita nel sinto lombardo, senza che fossero 'censurati' passaggi all'italiano, anche al fine di permettere il pieno fluire del repertorio linguistico dei parlanti. Il fatto che io non fossi un membro della comunità non sembra aver generalmente creato imbarazzo o difficoltà nei partecipanti, una volta stabilito un rapporto di fiducia con loro. Le interviste erano di natura semi-strutturata, basate cioè su un canovaccio di argomenti prestabiliti ma modificabili nel corso dell'interazione anche secondo le preferenze dei parlanti. In alcuni casi, nel procedere dell'intervista, erano gli stessi parlanti a proporre nuovi argomenti di conversazione. In generale, gli argomenti di conversazione hanno incluso le tradizioni dei sinti, la vita quotidiana, riflessioni sul rapporto con la società maggioritaria, gli atteggiamenti nei confronti della propria lingua, aneddoti personali, ecc. Le interviste si sono svolte nelle chiese di entrambi gli insediamenti o nelle abitazioni dei singoli parlanti. Durante le interviste, è capitato sovente che altri individui della comunità passassero e si fermassero ad avviare una breve interazione con i soggetti intervistati, come accade di frequente nelle comunità romaní.

In aggiunta alle interviste, sono state registrate una messa evangelica celebrata nella chiesa di via Bramante e una sessione integrale di compiti che ha visto la partecipazione di alcuni ragazzi sinti dello stesso insediamento. Tuttavia, poiché la lingua utilizzata nella messa era pressoché esclusivamente l'italiano, mentre la registrazione della sessione compiti è risultata particolarmente caotica, questi dati non sono stati in ultima analisi considerati ai fini della presente analisi.

Prima di concludere, va notato che le modalità descritte di raccolta dei dati di parlato presentano dei potenziali rischi metodologici, principalmente connessi alla presenza di un intervistatore esterno alla comunità che si pone nel ruolo di interlocutore in sinto, creando una situazione comunicativa fortemente marcata. Da un punto di vista del dato linguistico, una conseguenza problematica di ciò potrebbe risiedere nell'eventualità che i parlanti abbiano limitato consapevolmente il ricorso a fenomeni di *language mixing* al fine di restituire una versione il più possibile "pura" della propria lingua. Tuttavia, Adamou (2016) riporta che, anche in contesti di documentazione linguistica, i parlanti tendono a riprodurre i pattern di *language mixing* dominanti nella loro comunità, e che

dunque “the avoidance of contact words or codeswitching is not necessarily achieved or maybe even aimed at” (Adamou, 2016: 208). Questa idea emerge anche in alcune interviste. Si notino di seguito le parole di un parlante (MOR), riportate nell’esempio (1) che riflette sull’inopportunità, rispetto all’obiettivo percepito della sessione di registrazione, di cercare di riprodurre un “sinto stretto” ormai obsoleto (sulla nozione di “sinto stretto”, cfr. Capitolo 4). In questo esempio e nei seguenti, le parole preromanze (romaní) sono in corsivo, quelle italoromanze in stampato minuscolo, quelle codificate come “Altro” in stampato minuscolo sottolineato (cfr. § 3.4.1 per una descrizione dettagliata della metodologia di annotazione per lingua).

- (1) MON *i ja sintu, leskur barba (...)* / *jo džinel u sintu* proprio *u sintu* stretu da *ja* volta, stretu stretu
 ‘c’è un sinto, suo zio (...). Lui sa il sinto proprio, il sinto stretto di una volta, stretto stretto’
- MOR *i anca mar bibi kaj che rakarela pisa pape* / però *na servol=tu* miga perché ormai nianca *men na džinas=li kala kua kaj, tu kamea rakares ar u sinti d'akana*
 ‘c’è anche mia zia qui che parla un po’ ancora. però non ti serve perché ormai neanche noi sappiamo queste cose qui, tu vuoi parlare come i sinti di oggi’

Inoltre, riteniamo che tali limiti intrinseci siano stati almeno parzialmente compensati dalla lunga osservazione partecipante condotta, che ha permesso di sviluppare una conoscenza qualitativa delle pratiche linguistiche della comunità e di validare la qualità dei dati che venivano via via raccolti nei contesti ‘formali’ di intervista. Appare inoltre opportuno sottolineare come, in contesti sensibili come quello oggetto di analisi, in assenza di altri possibili strumenti di raccolta, sia prioritario privilegiare la trasparenza nella raccolta dei dati e il benessere dei partecipanti, anche a scapito dell’idea, spesso illusoria, della totale spontaneità del dato.

3.2.4 Aspetti etici

Prima di compilare il questionario e/o di prendere parte all’intervista sociolinguistica, i parlanti hanno letto e firmato un modulo di consenso informato. Tale modulo era scritto in stampato maiuscolo, con un linguaggio molto semplice per garantire la piena comprensione anche ai partecipanti meno scolarizzati (cfr. Appendice 1). Esso includeva la descrizione per sommi capi degli scopi della ricerca, i contatti del ricercatore e alcuni chiarimenti relativi alla *privacy*. Per esempio, era specificato che i dati sarebbero stati mantenuti anonimi e che le tracce audio non sarebbero state condivise con terzi. Inoltre, era garantita ai partecipanti la possibilità di ritirare in ogni momento la propria adesione alla

ricerca e di ottenere l'eliminazione dei propri dati, circostanza che non si è mai verificata. I partecipanti non hanno ricevuto un compenso finanziario per il loro coinvolgimento nella ricerca.

3.3 Il campione dei partecipanti

Questa sezione offre una panoramica quantitativa dei dati che costituiscono l'oggetto dell'analisi presentata nei successivi capitoli di questo volume.

Hanno risposto al questionario sociolinguistico 47 partecipanti, un numero piuttosto elevato, pari circa al 12% della comunità sinta di Pavia. La Tabella 4 mostra la loro ripartizione per genere, età e insediamento di residenza.

Fascia d'età	Europa		Bramante		Totale
	Uomo	Donna	Uomo	Donna	
<18	6	5	4	1	16
18-35	0	6	1	8	15
35-50	0	5	4	4	13
50+	0	1	1	1	3
Totale	6	17	10	14	47

Tabella 4: Caratteristiche socio-anagrafiche dei rispondenti al questionario.

Come si può notare, il campione risulta ben bilanciato per quanto riguarda l'insediamento di residenza, con circa la metà dei rispondenti residente in via Bramante (N=24) e l'altra metà (N=23) in piazzale Europa. Per quanto riguarda il genere, notiamo che le donne (N=31) sono più rappresentate rispetto agli uomini (N=16), una circostanza legata alla maggiore possibilità di interagire con loro durante la ricerca sul campo (in quanto molte sono le madri dei ragazzi partecipanti alle attività di doposcuola). Infine, quanto alla fascia d'età, i partecipanti sono piuttosto ben distribuiti, salvo per la fascia d'età 50+, rappresentata da soli tre parlanti. Se questo deriva in parte da una contingenza connessa al contesto di ricerca, ossia una minore possibilità di interazione attraverso le attività di catechismo e di doposcuola con gli appartenenti alle generazioni più anziane, tale distribuzione riflette anche la composizione demografica della comunità, nella quale, come visto, i minori di 18 anni costituiscono il 40% del totale mentre i parlanti più anziani sono scarsamente rappresentati. Quanto alle interviste sociolinguistiche, ai fini dell'analisi è stato selezionato un campione di testi prodotti da 26 parlanti, di cui la Tabella 5 indica la ripartizione per genere, età e insediamento di residenza.

Fascia d'età	Europa		Bramante		Totale
	Uomo	Donna	Uomo	Donna	
<18	2	5	1	1	9
18-35	0	4	2	2	8
35-50	0	2	1	1	4
50+	0	2	2	1	5
Totale	2	13	6	5	26

Tabella 5: Caratteristiche socio-anagrafiche dei partecipanti alle interviste.

Come si osserva, anche in questo caso i parlanti sono piuttosto equamente distribuiti tra i due insediamenti, con 15 parlanti residenti in piazzale Europa e i 11 in via Bramante, un dato che riflette le effettive differenze demografiche tra i due insediamenti. Inoltre, analogamente a quanto riscontrato per i questionari, le donne (N=18) sono più numerose degli uomini (N=8). Quanto alla fascia d'età, i gruppi di parlanti di meno di 18 anni e di età compresa tra i 18 e i 35 anni sono composti da un numero simile di parlanti, rispettivamente 9 e 8, mentre, per le ultime due fasce d'età, diminuisce il numero di parlanti intervistati.

3.4 Trattamento dei dati di parlato

Il corpus di parlato utilizzato per la presente analisi consta di 38.671 parole, un numero da considerare necessariamente come approssimativo data l'assenza di norme ben stabilite per la resa grafica del sinto lombardo. I dati di parlato sono stati trascritti nel *software* ELAN e successivamente annotati in maniera funzionale alle domande di ricerca. Per quanto riguarda la trascrizione, le parole di origine italo-romanza sono state trascritte seguendo le norme della grafia italiana, mentre le parole romaní sono state trascritte attendendosi alle norme utilizzate nel *Database of Romani Dialects* (cfr. Capitolo 2).

Quanto all'annotazione, i dati sono stati in primo luogo tokenizzati (ovvero divisi in *token* di parole) attraverso gli strumenti semi-automatici di ELAN. Successivamente, ogni parola è stata annotata su tre livelli gerarchici:

- 1) lingua;
- 2) parte del discorso;
- 3) lemma.

Poiché non è stato possibile utilizzare metodi automatici o semi-automatici, tali annotazioni sono state effettuate in maniera interamente manuale. Malgrado l'estrema dispendiosità in termini di

risorse temporali, questo triplice livello di annotazione ha costituito una solida base a partire dalla quale poter individuare gli schemi di combinazione linguistica maggiormente ricorrenti nel parlato dei sinti lombardi di Pavia.

La scelta di annotare l'intero corpus, anziché i soli fenomeni oggetto di analisi (nel nostro caso, ad esempio, le inserzioni e i casi di *code-switching* verso l'italiano), deriva dall'intento di applicare all'analisi il cosiddetto *principle of accountability* di matrice laboviana (cfr. p.es. Poplack, 2018). Tale principio metodologico impone di considerare non soltanto i fenomeni oggetto di interesse quando effettivamente si realizzano, ma anche i contesti in cui essi avrebbero potuto realizzarsi senza tuttavia verificarsi (nel nostro caso, ad esempio, i passaggi monolingui in sinto lombardo). Ciò consente di contestualizzare i fenomeni analizzati rispetto all'insieme delle occorrenze, ragionando cioè in termini proporzionali e dunque relativi, piuttosto che assoluti.

Va notato che, mentre l'annotazione per lingua ha riguardato tutto il corpus (in modo da determinare l'incidenza generale di parole dalle lingue a contatto nel parlato della comunità), l'annotazione per parte del discorso e per lemma non è stata effettuata per passaggi monolingui estesi in italiano, in quanto essa non sarebbe stata funzionale alle domande di ricerca. Gli ulteriori livelli di annotazione, relativi a domande di ricerca e campioni del corpus più specifici (p.es. il campo semantico dei nomi e dei verbi italo-romanzi), sono poi stati effettuati separatamente in fogli Excel.

Per quanto apparentemente chiari, i tre livelli di annotazione descritti (lingua, parte del discorso, lemma) hanno posto una serie di questioni metodologiche che discutiamo qui brevemente, in quanto rilevanti anche da un punto di vista teorico.

3.4.1 L'annotazione per lingua

In merito all'annotazione per lingua, un primo problema che si è posto ha riguardato la tensione tra un'analisi etimologica delle singole forme e il loro statuto psicolinguistico presso i parlanti. Un secondo problema è poi stato posto dalla notevole stratificazione linguistica, anche recente, caratterizzante il contesto di contatto in analisi.

In questo quadro, per garantire massima oggettività, è sembrato opportuno ricorrere a un'annotazione di tipo etimologico, distinguendo tra parole annotate come “ROM” e parole annotate come “ITA”, a cui è stata aggiunta un'etichetta residuale, “ALTRO”.

L'etichetta “ROM” è stata assegnata a parole romaní, appartenenti, cioè, allo strato indoario ereditato o ai successivi strati storici di contatto precedenti la fase italo-romanza, come l'armeno, il greco, lo slavo e il tedesco (cfr. Adamou & Granqvist, 2015; Adamou, 2016; Mirić & Ćirković, 2022 per un simile approccio in altri contesti romaní). Dal momento che non hanno più alcuna competenza, né

attiva né passiva, nelle lingue di contatto precedenti, i parlanti considerano le parole da esse prese a prestito come generalmente “sinte”.

L’etichetta “ITA” è stata invece utilizzata per parole di origine italo-romanza, sia italiane sia di origine dialettale (prevalentemente lombarda). Nel contesto di contatto in analisi, la distinzione tra forme italiane e dialettali sarebbe piuttosto rilevante su un piano teorico a causa del diverso statuto delle due lingue nel repertorio dei parlanti: come vedremo anche al Capitolo 4, a differenza dell’italiano il dialetto lombardo non è più una lingua attiva nella comunità, e dunque il ricorso a parole di origine lombarda rappresenta l’esito di un contatto ormai cristallizzato, non (più) frutto del bilinguismo dei parlanti. In pratica, tuttavia, effettuare una chiara distinzione tra parole italiane e lombarde si è rivelato in molti casi difficile a causa della vicinanza genetica delle due lingue e dei processi di integrazione a livello fonetico che rendono ulteriormente opaca questa distinzione (cfr. Capitolo 6). In questo quadro, laddove possibile, le forme inequivocabilmente dialettali sono state etichettate più specificamente come “I-Dial” (come l’avverbio negativo *miga* ‘mica’ o i sostantivi *funaral* ‘funerale’ e *barba* ‘zio’), mentre in presenza di forme ambigue non si è potuto che limitarsi all’etichetta più generale “ITA”. Per questa ragione, pur senza escludere occasionali distinzioni a grana più fine, durante l’analisi si farà soprattutto riferimento a “italoromanzo” come categoria generale. Occorre anche notare che, laddove specifiche parole contenute in italo-romanze risultassero integrate nella morfologia del sinto lombardo (come nel caso dei verbi), nell’annotazione per lingua si è considerata soltanto la lingua d’origine della radice lessicale; di conseguenza, forme di questo tipo sono state annotate come “ITA”, mentre l’analisi delle strategie di integrazione a livello morfologico è stata oggetto di una successiva analisi dettagliata (cfr. Capitolo 6).

Se la distinzione tra forme italiane e dialettali ha dunque posto particolari problemi, non si sono invece riscontrate difficoltà nella distinzione tra forme etimologicamente italo-romanze e romaní. Tuttavia, l’etichetta “ALTRO” è stata utilizzata per alcune parole non immediatamente attribuibili all’italoromanzo o al sinto. In particolare, sotto questa etichetta sono stati inclusi: vocaboli palesemente stranieri (come *rock’n’roll*); nomi propri (principalmente antroponomi e marchionimi), che per definizione sono dati e non specifici per lingua; riempitivi come *eh* o *mhmh*; parole funzionalmente miste, composte da segmenti riconducibili a entrambe le lingue, come la congiunzione *eta* ‘e’, frutto dell’unione della forma italiana *e* e della forma di origine indo-aria *ta*.

In conclusione, è opportuno ribadire come il criterio scelto per l’annotazione linguistica sia stato di natura diacronico-etimologica e *non* sincronico-psicolinguistica. In questo senso, il fatto che una forma sia annotata come “italoromanza” non equivale a dire che essa sia necessariamente considerata come tale dai parlanti. L’allocazione emica delle unità linguistiche ai diversi segmenti del repertorio bilingue risulterà piuttosto un’approssimazione dell’indagine empirica sulla base di fattori come la

frequenza, la diffusione e l'integrazione linguistica, in linea con lo spirito *corpus-driven* di questo lavoro (cfr. Adamou, 2016). Inoltre, si riconosce che i parlanti bilingui non attribuiscono necessariamente etichette linguistiche univoche a singole forme che, pur etimologicamente legate a una sola lingua, giungono a essere condivise da due o più, come peraltro evidenziato da diversi approcci che enfatizzano la natura integrata del repertorio bilingue (cfr. Capitolo 1).

3.4.2 L'annotazione per parte del discorso

Anche l'annotazione per parti del discorso ha comportato alcune difficoltà. Esse sono state motivate da vari fattori, tra i quali si può menzionare l'assenza di una tradizione scritta per il sinto, circostanza che rende l'individuazione stessa dei confini di parola non sempre ovvia, e la non sempre perfetta corrispondenza delle categorie grammaticali del sinto lombardo e dell'italiano. In questo quadro, è sembrato un buon compromesso utilizzare etichette piuttosto ampie, definite su base sintattico-distribuzionale. In particolare, abbiamo distinto le seguenti classi di parole, seguendo Dal Negro (2009): Nomi; Nomi Propri; Verbi; Aggettivi; Avverbi; Pronomi; Determinanti; Connettivi e Segnali Discorsivi; Preposizioni. L'etichetta "Connettivi e Segnali Discorsivi" necessita di alcuni chiarimenti: si tratta di una categoria, volutamente molto ampia ed eterogenea, che include congiunzioni vere e proprie (subordinanti e coordinanti, del tipo *ma* o *perché*), avverbi connettivi (come *allora* e *però*), segnali discorsivi come *vabbè* e *dai* e particelle affermative e negative come *ua* 'sì' e *na* 'no'. Si tratta di elementi funzionali all'organizzazione del discorso e sensibili al contesto dell'interazione, che tendono grossomodo a coincidere con la categoria degli "utterance modifiers" individuata da Matras (1998a).

3.4.3 La lemmatizzazione

L'ultimo livello di annotazione, la lemmatizzazione, permette di dare conto della frequenza di specifici tipi lessicali nel parlato della comunità, anche in opposizione alle forme 'native' (cfr. Capitolo 5). A tal fine, le forme italiane sono state ricondotte alla forma di citazione normalmente indicata nei dizionari della lingua italiana, mentre le forme romaní sono state lemmatizzate seguendo le norme diffuse nella linguistica romaní (Matras, 2002). Anche per questo livello si è posto il problema della distinzione tra forme dialettali e forme italiane. Come nell'annotazione per lingua, così anche per questo livello le forme chiaramente dialettali sono state ricondotte a un lemma indipendente rispetto alla forma italiana (p.es. *nivudo* > *nivudo* 'nipote' vs *nipote* > *nipote*), mentre le forme ambigue sono state ricondotte a un unico lemma.

Capitolo 4. L'analisi macrosociolinguistica

Questo capitolo si propone di analizzare la comunità dei sinti lombardi di Pavia da un punto di vista macrosociolinguistico, principalmente sulla base delle risposte dei partecipanti ai questionari (cfr. Capitolo 3). A queste si affiancano, quale completamento qualitativo, i passaggi delle interviste in sinto lombardo in cui emergono le riflessioni metalinguistiche dei parlanti, nonché le osservazioni etnografiche effettuate nel corso della ricerca sul campo. Piuttosto che offrire una panoramica completa di tutte le domande del questionario, l'analisi si concentra su quelle ritenute maggiormente informative per comprendere le dinamiche sociolinguistiche più rilevanti in atto nella comunità, anche attraverso una triangolazione con i metodi qualitativi citati.

A nostra conoscenza, l'unico studio di taglio macrosociolinguistico a oggi disponibile sulla comunità dei sinti lombardi è quello di Scala (2012), in cui la conformazione del repertorio linguistico viene confrontata con quella di altri gruppi sinti dell'Italia settentrionale, anche in chiave diacronica. Pur trattandosi di un contributo di notevole qualità, tale analisi si basa prevalentemente sulle osservazioni dirette dell'autore e non su dati raccolti in modo sistematico tramite strumenti di indagine empirica all'interno di una comunità specifica. Il presente capitolo intende dunque approfondire l'analisi proposta da Scala (2012), offrendo uno spaccato degli usi linguistici in una comunità specifica in un momento storico definito.

Il capitolo è strutturato come segue. Il § 4.1 analizza le lingue che i parlanti dichiarano come parte del proprio repertorio linguistico, soffermandosi anche sulle rappresentazioni emiche relative alle nozioni di "lingua madre" e alle denominazioni attribuite alle diverse lingue. Il § 4.2 illustra brevemente i periodi di acquisizione del sinto lombardo e dell'italiano, con l'obiettivo di ricostruire le modalità di socializzazione linguistica nella prima infanzia. Il § 4.3 prende in esame le autovalutazioni fornite dai parlanti in merito alla loro competenza in sinto e in italiano su diversi livelli (parlato, scrittura, lettura). Il § 4.4, che costituisce uno dei nuclei di questo capitolo, si concentra sulle scelte linguistiche al variare dei domini comunicativi e degli interlocutori. Il § 4.5 analizza poi gli atteggiamenti dei sinti lombardi di Pavia verso il sinto lombardo e verso il contatto linguistico con l'italiano, prendendo in considerazione sia le valutazioni esplicite emerse dal questionario sia le riflessioni metalinguistiche ricorrenti nelle interviste. Infine, il § 4.6 propone una discussione dei principali risultati emersi.

4.1 Le lingue del repertorio

Questa sezione è dedicata all'analisi delle risposte alle prime due domande, entrambe aperte, con cui si apriva il questionario: “Qual è la tua lingua madre?” e “Quali altre lingue e/o dialetti parli?”. Nel § 4.1.1, le risposte alle due domande vengono considerate congiuntamente, in modo da ottenere una panoramica quantitativa delle lingue che sono percepite dai sinti lombardi come parte del proprio repertorio linguistico. Il paragrafo 4.1.2 si concentra invece esplicitamente sulla prospettiva dei parlanti, analizzando le rappresentazioni emiche relative alla nozione di “lingua madre” (§ 4.1.2.1) e alle denominazioni attribuite alle diverse lingue del repertorio (§ 4.1.2.2). Infine, il § 4.1.3 propone una breve descrizione delle caratteristiche strutturali del sinto e dell'italiano parlato dai sinti lombardi, combinando una prospettiva etica con le osservazioni e le categorizzazioni emerse dal discorso dei parlanti stessi.

4.1.1 Lingue conosciute

Le risposte combinate alle domande “Qual è la tua lingua madre?” e “Quali altre lingue e/o dialetti parli?” sono rappresentate nella Figura 3.

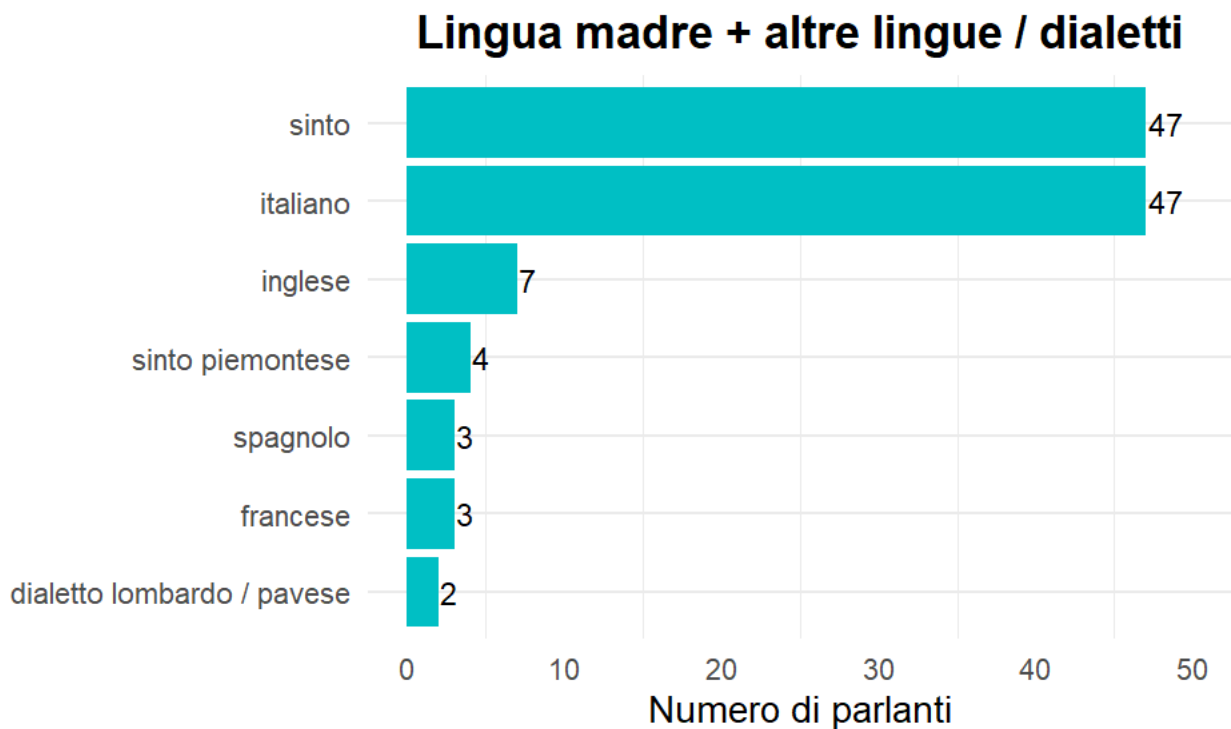


Figura 3: Risultati accorpati delle domande “Qual è la tua lingua madre?” e “Quali altre lingue o dialetti parli?”.

Come si può osservare, tutti i 47 rispondenti dichiarano di parlare sia il sinto lombardo sia l'italiano.

Al di fuori di queste due lingue, le altre lingue conosciute risultano tutto sommato marginali. Quattro rispondenti dichiarano di conoscere anche il sinto piemontese, un dato che solleva interessanti interrogativi su quale varietà essi intendano effettivamente, alla luce della complessa situazione sociolinguistica dei sinti piemontesi delineata al Capitolo 2. In particolare, alla richiesta di fornire un esempio di sinto piemontese, tre parlanti hanno proposto esempi che fanno chiaramente intendere di riferirsi non già alla *romaní*, bensì alla varietà italoromanza adottata dai sinti piemontesi negli ultimi decenni. Una parlante, inoltre, osserva esplicitamente che il sinto piemontese a cui si riferisce è “il dialetto parlato dai *gagi* piemontesi”. L’indicazione di questa varietà è stata motivata dai rispondenti, tutti adulti, sulla base del fatto di avere parenti (in alcuni casi i genitori) parlanti questa varietà.

È inoltre significativo notare che soltanto due parlanti dichiarano di avere una competenza attiva nei dialetti italoromanzi della zona, definiti in un caso dialetto pavese e nell’altro caso dialetto lombardo. Questo dato conferma ampiamente le osservazioni di Scala (2012) relative all’interruzione della trasmissione del dialetto presso i sinti lombardi. Va tuttavia precisato che l’assenza, nel campione, di parlanti appartenenti alla fascia d’età più avanzata (oltre i 60 anni) non consente di verificare la tenuta del dialetto lombardo presso tale gruppo.

Infine, alcuni parlanti dichiarano competenze attive in lingue della comunicazione internazionale: prima su tutte l’inglese (7 risposte), seguito da spagnolo e francese. In tutti i casi si tratta di parlanti giovani e frequentanti la scuola, per i quali queste lingue rappresentano materie di studio.

In sintesi, questi dati confermano l’italiano e il sinto lombardo come le lingue principali del repertorio linguistico dei sinti lombardi.

4.1.2 Rappresentazioni emiche delle lingue

4.1.2.1 La ‘lingua madre’

La prima domanda del questionario recitava: “Qual è la tua lingua madre?”. Come è noto, quella di “lingua madre” non è una categoria scientifica in senso stretto, ma una nozione *folk* che può tuttavia essere utilmente impiegata come strumento per accedere alle rappresentazioni linguistiche dei parlanti. A questo proposito, sulla base di questionari sociolinguistici somministrati a individui di diverse comunità minoritarie in Serbia, Mirić et al. (2024) mostrano che, al netto di una marcata variazione individuale, i parlanti tendono ad associare la lingua madre alla prima lingua acquisita temporalmente e/o trasmessa in ambito familiare. Iannàccaro & Dell’Aquila (2000) osservano inoltre che la lingua madre dichiarata dai parlanti tende a indicare la comunità linguistica nella quale essi sentono di o desiderano appartenere.

L'obiettivo di questa analisi è comprendere quale/i lingua/e i sinti lombardi considerano come propria 'lingua madre' e su quali basi motivino tale scelta. I risultati delle risposte a questa domanda sono riportati nella Figura 4.

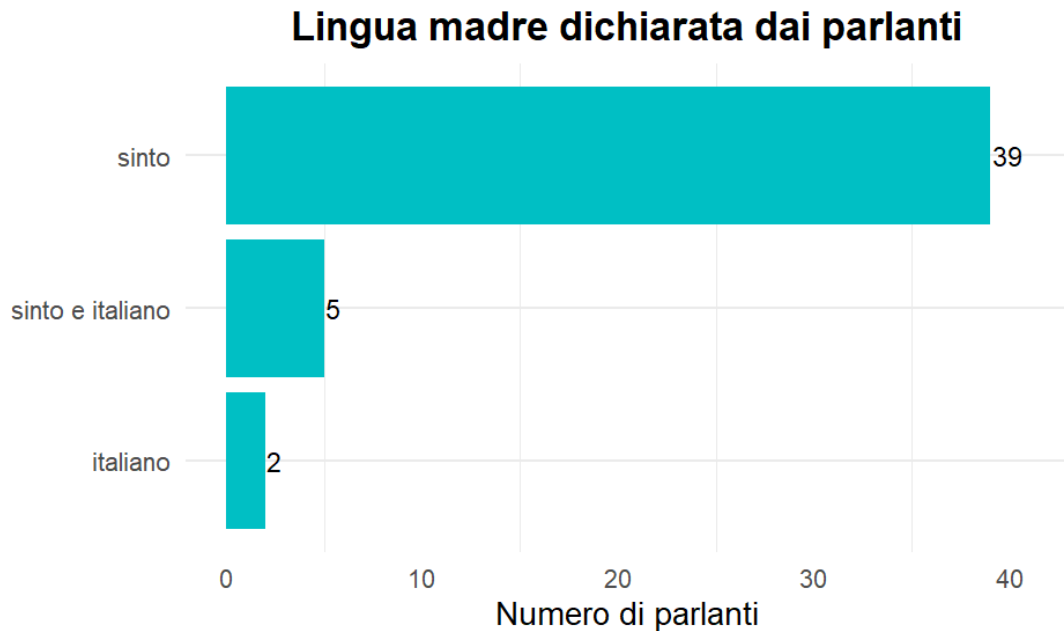


Figura 4: Risposte alla domanda: “Qual è la tua lingua madre?”.

Come si osserva, la netta maggioranza dei rispondenti (N=39) identifica la propria lingua madre nel sinto, più raramente definito sinto lombardo; 5 parlanti indicano sia il sinto lombardo sia l'italiano, mentre soltanto due dichiarano l'italiano come unica lingua madre. Si segnala inoltre che una partecipante non ha fornito risposta a questa domanda, dichiarando di non comprendere chiaramente il significato della nozione di lingua madre.

Occasionalmente, i parlanti hanno accompagnato la propria risposta con motivazioni esplicite. Coloro che indicano il sinto lombardo come lingua madre tendono a enfatizzare il fatto che si tratti della *propria* lingua (sull'uso di questa etichetta, cfr. § 4.1.2.2), vale a dire la lingua trasmessa e utilizzata nella propria famiglia. I due partecipanti che hanno risposto italiano tendono invece a enfatizzare la sua ufficialità, quale lingua della nazione in cui vivono e insegnata a scuola. Quanti, infine, dichiarano come lingue madre sia il sinto lombardo sia l'italiano evocano la complementarità funzionale di queste due lingue nella loro esperienza quotidiana, nonché il loro essere al contempo sinti e italiani. Nel complesso, i risultati confermano le osservazioni di Mirić et al. (2024) circa la pluralità di rappresentazioni associate alla nozione di lingua madre anche all'interno di una stessa comunità. Diversamente da quanto osservato per altre comunità minoritarie (cfr. Dell'Aquila & Iannàccaro, 2006 sulla comunità ladina in Alto Adige), il fatto che il sinto lombardo non sia una lingua ufficiale

e standardizzata non sembra ostacolare la sua categorizzazione come lingua madre da parte dei parlanti. In questo contesto, altri fattori, quali la forte distanziamento strutturale del sinto lombardo dall'italiano, il marcato senso di specificità etnico-culturale rispetto alla società maggioritaria (cfr. Capitolo 3), e il fatto che il sinto lombardo costituisca la lingua della comunicazione familiare sembrano costituire fattori più rilevanti nelle rappresentazioni della nozione di lingua madre presso i sinti lombardi. Infine, le cinque risposte che indicano sia il sinto lombardo sia l'italiano come lingua madre risultano piuttosto significative, anche considerando che la domanda faceva riferimento a una sola lingua madre al singolare, e che l'idea dell'esistenza di una sola lingua madre rappresenta un costrutto ideologico particolarmente radicato negli stati-nazione occidentali (Dal Negro, 2011b). Tali risposte potrebbero dunque già suggerire una valutazione positiva del bilinguismo da parte dei sinti lombardi, un aspetto che verrà ripreso più in dettaglio al § 4.5.2.

4.1.2.2 I nomi delle lingue

Un ulteriore aspetto di interesse relativo alle rappresentazioni linguistiche dei parlanti riguarda le denominazioni che essi attribuiscono alle loro lingue. L'analisi dei nomi delle lingue consente infatti di “risalire all'allocazione emica dei codici in repertori individuali e comunitari” (Lupica Spagnolo, 2019: 142).

Nelle risposte alle prime due domande del questionario, i parlanti hanno sistematicamente utilizzato il termine “italiano” per riferirsi all'italiano e “sinto”, o “sinto lombardo”, per riferirsi alla lingua di minoranza. Tali denominazioni rientrano nella categoria dei “glottonomi”, ovvero nomi propri utilizzati per classificare le lingue (Lupica Spagnolo, 2019). È interessante notare come il sinto non venga mai classificato attraverso un'etichetta più ampia e “istituzionale”, quale ad esempio *romaní*, che, come osserva Scala (2020b), rappresenta una denominazione prevalentemente accademica, poco o per nulla utilizzata dai parlanti stessi, in gran parte estranei anche ai discorsi ufficiali sulla lingua portati avanti da intellettuali appartenenti alle comunità (cfr. p.es. Abercrombie, 2018).

Il confronto dei dati del questionario con i dati delle interviste rivela inoltre differenze rilevanti nelle denominazioni. In queste ultime, l'italiano è infatti frequentemente indicato dai parlanti come *gagio* (si ricordi che il termine *gagio*, in *romaní*, designa primariamente l'uomo non sinto). Il contesto del questionario, probabilmente percepito come più ‘ufficiale’, e la scelta dell'italiano come lingua di compilazione (il termine *gagio* tende infatti a essere concepito come il traduttore sinto di ‘italiano’) potrebbero aver motivato la differenza tra le designazioni utilizzate nei due contesti di elicitazione. Analogamente, nelle interviste emerge una maggiore varietà di denominazioni anche per il sinto. Accanto al termine “sinto”, comunque molto frequente, ricorrono espressioni quali *mengar chib* ‘la nostra lingua’ o, più spesso, *mengur rakarpe*, letteralmente ‘la nostra parlata’ (dal verbo *rakar-*

‘parlare’). Queste espressioni rientrano nella categoria delle “apposizioni”, ovvero nomi comuni utilizzati non tanto per classificare le lingue, quanto per descriverle (Lupica Spagnolo 2019). Non è al momento possibile stabilire se sussista presso i parlanti una qualche gerarchizzazione o specializzazione semantica tra i termini *chib* e *rakarpe*. In ogni caso, il contrasto ricorrente tra *mengur rakarpe* ‘la nostra lingua’ e *gagio* (la lingua dei *gagi*, cioè gli altri per eccellenza) sembra indicare in modo chiaro la funzione identitaria attribuita al sinto.

4.1.3 Il sinto e l’italiano dei sinti lombardi: alcune note

Questa sezione ha l’obiettivo di tracciare brevemente un profilo del sinto e dell’italiano parlati dai sinti lombardi di Pavia, al fine di inquadrare meglio quali varietà siano effettivamente in contatto nel contesto in analisi, lasciando spazio anche alle prospettive emiche dei parlanti.

La varietà di sinto parlato dai sinti lombardi di Pavia è chiaramente riconducibile al sinto lombardo, così come descritto da Soravia (1977). Sebbene molte persone residenti nell’insediamento di piazzale Europa si autodefiniscano come *mucini*, la varietà di sinto che esse utilizzano non appare riconducibile a ciò che Partisani (1984), in un glossario pubblicato dalla rivista *Lacio Drom*, ha descritto come *sinto mucino*, ovvero una sorta di varietà sinta di compromesso, lombarda nella grammatica e piemontese nel lessico (cfr. Scala, 2012 per simili osservazioni).

È interessante notare che i parlanti, pur consci di parlare la medesima varietà di sinto lombardo, riconoscono l’esistenza di differenze tra le varietà utilizzate dalle varie famiglie all’interno della comunità. Tali differenze sembrano correlarsi in particolare all’appartenenza all’insediamento di via Bramante o di piazzale Europa. In particolare, i parlanti molto spesso osservano che la comunità di piazzale Europa parla “con la o”, mentre quella di via Bramante parla “con la u”. Dagli esempi forniti emerge che tale distinzione riguarda la selezione dell’una o dell’altra vocale nelle sillabe aperte o chiuse in finale di parola, tanto nel lessico romaní (p.es. *lačo* vs. *laču* ‘bello’), quanto nel lessico italo-romanzo (p.es. *barbon* vs. *barbun* ‘barbone’). In particolare, si può osservare che la variante -o risulta in alcuni casi stigmatizzata dai parlanti di via Bramante, in quanto associata ai sinti *mucini*. Talvolta, nel corso della mia ricerca sul campo, mi è addirittura capitato di essere ironicamente redarguito da alcuni parlanti di via Bramante per un uso eccessivo della -o durante le interviste con loro, con l’“accusa” di “parlare come un *mucino*”. Nel microcosmo sinto, si potrebbe dunque concludere che la selezione delle vocali la -o o -u rappresenti un vero e proprio indicatore sociale, nel senso di Labov (1972), del gruppo di appartenenza. Inoltre, è molto diffusa tra i parlanti l’idea di servirsi di un sinto fortemente interferito dall’italiano a livello lessicale, un aspetto su cui torneremo più in dettaglio al § 4.5.

Per quanto riguarda l'italiano, la varietà utilizzata dai sinti lombardi non è mai stata oggetto di studi specifici, ma può essere preliminarmente descritta come una varietà di italiano regionale di area settentrionale caratterizzata da alcuni tratti etnolettali a diversi livelli linguistici. A livello fonetico, per esempio, si osserva la tendenza alla realizzazione dei suoni affricati [ts] e [dz], assenti nell'inventario fonologico del sinto lombardo, come [s] e [z]. Come già segnalato da Sorrenti (2014), questo tratto è riscontrabile anche presso parlanti italiani di area settentrionale prevalentemente dialettofoni e può talvolta essere associato, nella percezione della popolazione maggioritaria, a una conoscenza approssimativa dell'italiano standard. A livello lessicale, oltre all'impiego di inserzioni 'culturali' dalla romaní (come, per esempio, *gagio* o *kampina* 'roulotte'), si osserva il ricorso occasionale a calchi di modelli lessicali romaní, come le espressioni ingiuriose *mangiamorti* (cfr. sinto *xala mule*) o *mangiamerda* (cfr. sinto *xala ful*), nonché i verbi *prendere* con significato di 'sposarsi' (cfr. sinto *l- rom/romli* 'prendere marito/moglie') o *mangiare via* (cfr. sinto *x- vek*), nel senso di affrontare aggressivamente qualcuno dopo essersi arrabbiato con lui.

Da un punto di vista emico, l'uso dell'italiano non sembra invece essere oggetto dello stesso grado di riflessione metalinguistica che caratterizza il discorso sul sinto lombardo. Tuttavia, in alcuni casi i parlanti segnalano difficoltà nella realizzazione dei suoni affricati [ts] e [dz] mentre parlano in italiano, associando una maggiore frequenza di questo tratto agli usi linguistici delle persone più anziane e scarsamente scolari.

4.2 Periodi di acquisizione di sinto e italiano

In aggiunta alle lingue conosciute, altre domande, sempre aperte, del questionario erano volte a raccogliere informazioni sui processi di acquisizione del sinto e dell'italiano (cfr. domande 13 e 14). In questa sezione, ci concentriamo in particolare sulle risposte alla domanda "Quale/i lingua/e hai imparato per prima/e da piccolo?". I risultati sono riportati nella Figura 5.

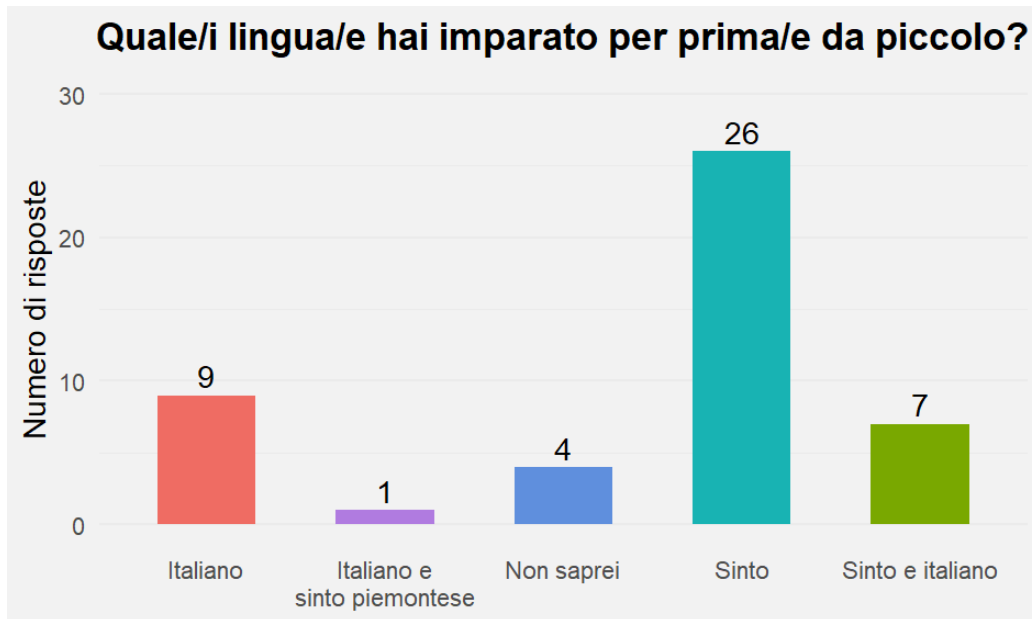


Figura 5: Risposte alla domanda “Quale/i lingua/e hai imparato per prima/e da piccolo?”.

Dallo spoglio delle risposte emerge una situazione piuttosto articolata: 26 parlanti su 47 dichiarano di aver appreso prima il sinto, nove di aver appreso prima l’italiano, sette di avere appreso simultaneamente il sinto e l’italiano. Quattro parlanti dichiarano inoltre di non essere in grado di stabilire quale lingua abbiano appreso per prima. Infine, una parlante di più di cinquant’anni dichiara di avere appreso insieme al sinto lombardo anche il sinto piemontese, in quanto sua madre era una sinta piemontese.

È rilevante notare che, anche nei casi in cui i parlanti dichiarano di aver appreso inizialmente una lingua specifica (ad esempio il sinto lombardo), essi osservano generalmente di aver acquisito l’altra lingua poco dopo, sempre nel corso della prima infanzia. Coerentemente con quanto emerge dai dati del questionario, l’esperienza etnografica sul campo mi ha permesso di osservare che anche bambini molto piccoli, che non hanno ancora fatto il loro ingresso nella scuola primaria (ed è raro che frequentino la scuola dell’infanzia), utilizzano attivamente sia il sinto lombardo sia l’italiano.

In sintesi, questi dati suggeriscono una condizione diffusa di bilinguismo simultaneo, in cui sia il sinto sia l’italiano vengono apprese durante la socializzazione primaria, prima dell’ingresso a scuola, in linea con la situazione riportata per la maggior parte delle comunità rom in Europa (Matras, 2002).

4.3 Autovalutazione della competenza in sinto e in italiano

Il questionario richiedeva inoltre ai parlanti di autovalutare la propria competenza nel sinto lombardo e nell’italiano, assegnando un punteggio da uno a dieci a tre diversi livelli di competenza comunicativa: parlato, scrittura e lettura (cfr. domande 24 e 25). I grafici nella Figura 6 mostrano i valori medi delle autovalutazioni fornite dai parlanti, insieme alla relativa deviazione standard.

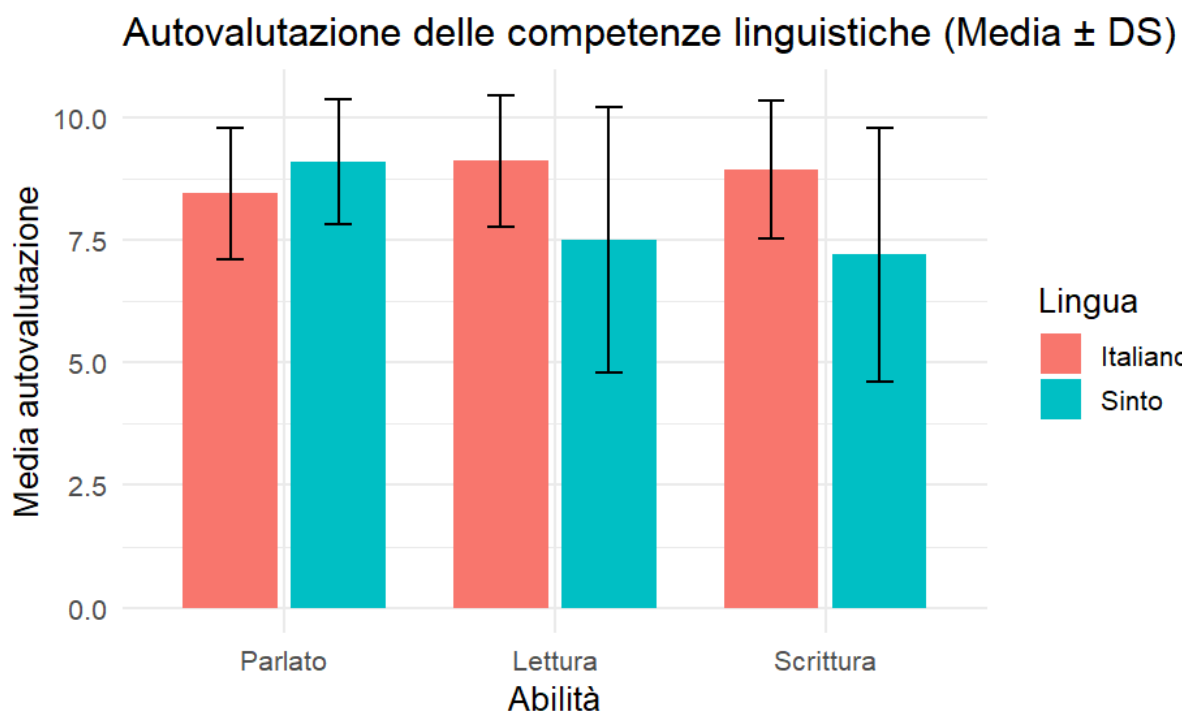


Figura 6: Competenze autodichiarate in sinto lombardo e italiano (parlato, lettura, scrittura).

I risultati che emergono dalla visualizzazione di questa Figura sono piuttosto interessanti. Si può osservare che, per quanto riguarda il sinto, la competenza media risulta più elevata nel parlato (9.1) rispetto alle abilità di lettura (7.84) e scrittura (7.49). Queste ultime presentano inoltre una maggiore deviazione standard (2.6 nel caso della scrittura e 2.7 nel caso della lettura), indicando una forte variabilità nelle valutazioni individuali. Per quanto riguarda l'italiano, i parlanti si auto-assegnano livelli di competenza complessivamente elevati e più omogenei sui tre livelli. Tuttavia, emerge una gerarchia di competenze opposta rispetto a quella osservata per il sinto, in quanto la lettura (9.13) e la scrittura (8.93) ricevono valutazioni medie più alte del parlato (8.49). Inoltre, la deviazione standard risulta più contenuta su tutti e tre i livelli, suggerendo una minore variazione individuale nella competenza percepita dell'italiano. L'autovalutazione più bassa nelle abilità di lettoscrittura nella lingua minoritaria rispetto alla lingua standard dominante è stata riportata in numerose altre situazioni di contatto linguistico asimmetrico: a fini comparativi con il nostro caso di studio, cfr. Mirić (2025) su una comunità rom in Serbia.

Alla luce della variabilità osservata, si è ritenuto opportuno approfondire in che misura le differenze nelle autovalutazioni della competenza in sinto lombardo e in italiano siano correlate a tre fattori socio-anagrafici: genere, fascia d'età e zona di insediamento (cfr. Figure 7–9).

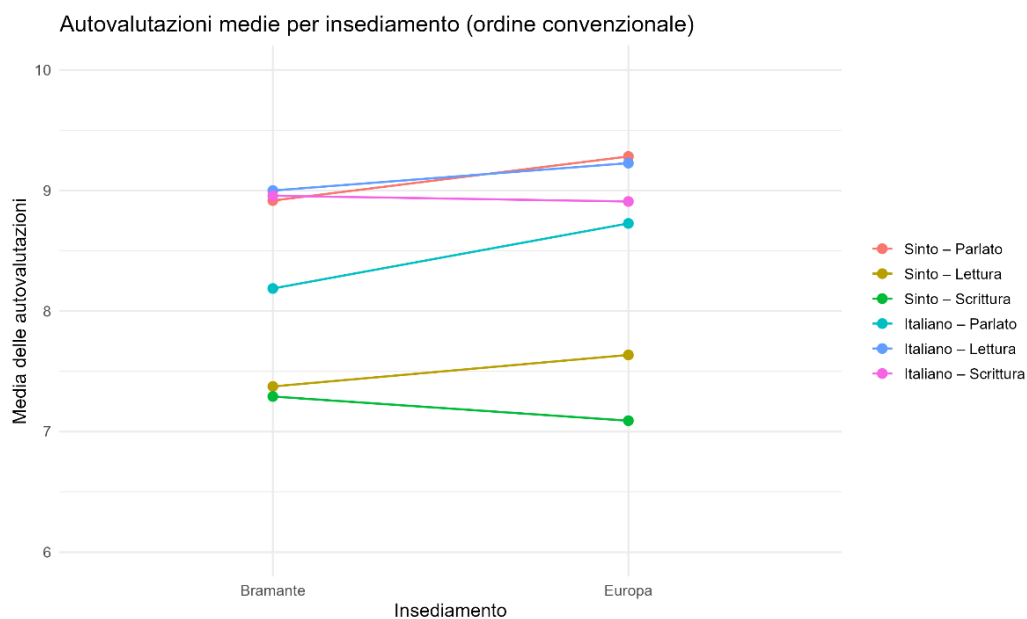


Figura 7: Variazione nelle autovalutazioni per insediamento.

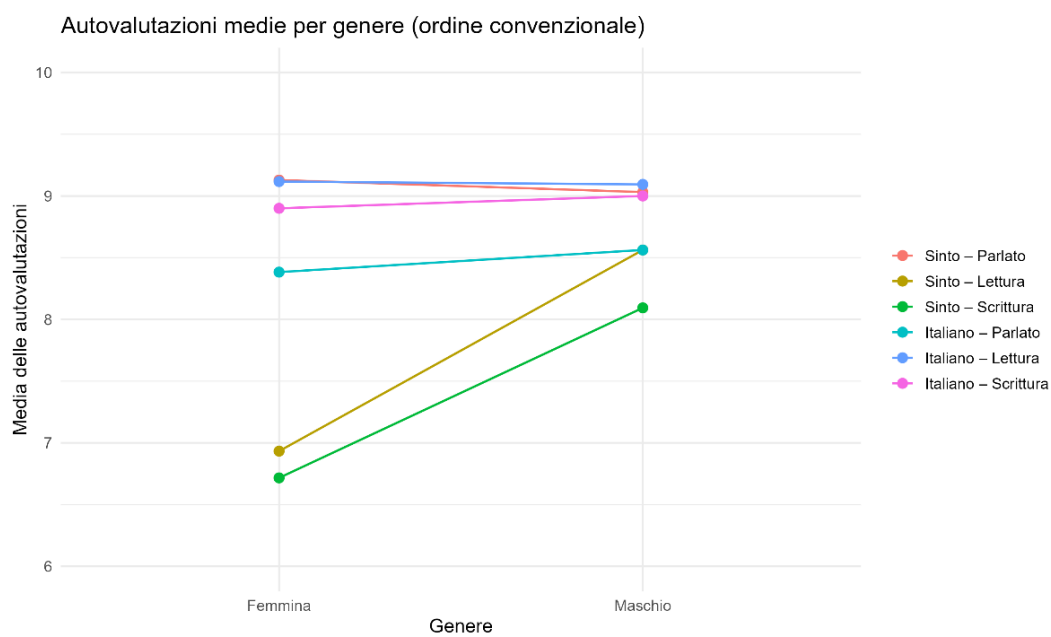


Figura 8: Variazione nelle autovalutazioni per genere.

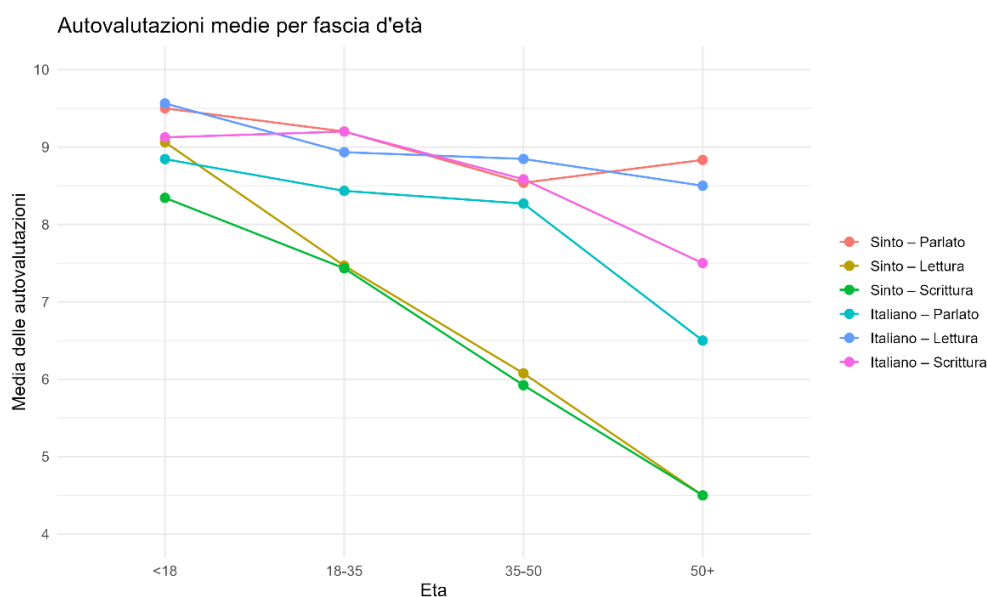


Figura 9: Variazione nelle autovalutazioni per fascia d'età.

I grafici mostrano alcune tendenze degne di nota. Si può notare che le differenze legate all'insediamento di residenza (Figura 7) risultano particolarmente limitate: in entrambi i contesti, le autovalutazioni medie sulle sei competenze considerate si collocano in intervalli molto simili, con scarti generalmente inferiori a mezzo punto. Anche per quanto riguarda il genere (Figura 8), le differenze appaiono contenute: in media, i punteggi attribuiti da uomini e donne differiscono di meno di un punto su tutti i livelli di competenza, sebbene gli uomini tendano a dichiararsi leggermente più competenti nella lettura e nella scrittura in sinto lombardo rispetto alle donne.

La variazione più marcata nelle autovalutazioni emerge invece considerando la fascia d'età (Figura 9). Tra i parlanti con meno di 18 anni e i parlanti con più di 50 anni si osservano infatti differenze che possono superare i due punti nelle competenze di lettura e scrittura in sinto lombardo e, in misura minore, in italiano. Più nello specifico, la scrittura e la lettura in sinto lombardo mostrano una diminuzione progressiva passando dalle fasce più giovani a quelle più anziane, fino a raggiungere valori medi prossimi al limite inferiore della scala per i parlanti oltre i 50 anni.

Tale andamento sembra potersi mettere in relazione con le diverse esperienze di pratica della lettoscrittura nei vari gruppi d'età, un aspetto strettamente connesso alla scolarizzazione e all'accesso alle tecnologie. Mentre i parlanti più giovani frequentano la scuola e praticano la lettoscrittura su base quotidiana, anche attraverso i *social network*, i parlanti di età superiore ai 50 anni non frequentano la scuola, spesso l'hanno frequentata per un periodo limitato in passato e tendono a non esercitare ulteriormente la lettoscrittura in contesti informali come i *social network*. Nella comunità dei sinti lombardi di Pavia, non è così infrequente trovare parlanti di più di cinquant'anni analfabeti: si tratta soprattutto degli appartenenti alla generazione che ha vissuto la propria infanzia e adolescenza prima

degli anni Ottanta, decennio in cui i sinti di Piazzale Europa si sono stabiliti a Pavia rinunciando alle attività itineranti connesse allo spettacolo viaggiante, una circostanza che non consentiva la frequentazione stabile della scuola. Si noti inoltre che in contesti quali la scuola e i *social network* la lettoscrittura tende a essere praticata in italiano (cfr. § 4.4.4.1): sembra dunque che la percezione di un'elevata competenza nella lettoscrittura in italiano da parte dei parlanti più giovani venga trasferita anche al sinto su questi stessi livelli, nonostante lo scarso utilizzo del sinto per tali funzioni. Viceversa, la percezione di una scarsa competenza nella lettoscrittura in italiano da parte dei parlanti più anziani sembra essere trasferita, e anzi amplificata, nel caso del sinto lombardo.

4.4 Scelte linguistiche e variazione

Dopo aver discusso le lingue che i sinti lombardi di Pavia percepiscono come parte del proprio repertorio e la loro competenza in sinto lombardo e in italiano, questa sezione si propone di ricostruire la distribuzione funzionale di tali lingue al variare dei domini e delle situazioni comunicative. L'attenzione è rivolta in particolare agli usi orali all'interno e all'esterno della comunità (§ 4.4.1), alle scelte linguistiche al variare degli interlocutori all'interno della comunità (§ 4.4.2) e agli usi nella comunicazione con parenti sinti non lombardi (§ 4.4.3), per concludere con alcuni domini specifici, quali la tecnologia, il web e le pratiche religiose (§ 4.4.4). Si noti che tutte le domande analizzate in questa sezione prevedevano risposte aperte.

4.4.1 Usi endocomunitari ed esocomunitari

Le domande 3-13 del questionario erano volte a indagare gli usi linguistici dei sinti lombardi nella comunicazione orale all'interno della comunità (domande 3-10) e all'esterno (domande 11-13). Definiremo il primo tipo di situazioni comunicative come “endocomunitarie” e il secondo come “esocomunitarie” (cfr. p.es. Scala, 2012). Più nello specifico, le domande relative ai domini endocomunitari indagavano gli usi con diversi locutori relativi della comunità, tra cui padre, madre, fratelli, nonni, figli, partner, amici sinti, bambini piccoli e parlanti anziani della comunità. Le domande relative ai domini esocomunitari includevano invece tre situazioni caratterizzate da un grado decrescente di formalità: la lingua usata nel dialogo con l'amministrazione (p.es. in comune); la lingua usata con i commercianti della città; la lingua usata con gli amici non sinti.

Una panoramica aggregata delle risposte relative ai due macro-domini è presentata di seguito nella Figura 10: in pratica, per ciascun dominio, tutte le risposte “italiano” sono state incluse in un gruppo, tutte le risposte “sinto” in un altro gruppo, tutte le risposte “sinto e italiano” in un altro ancora, e così via.

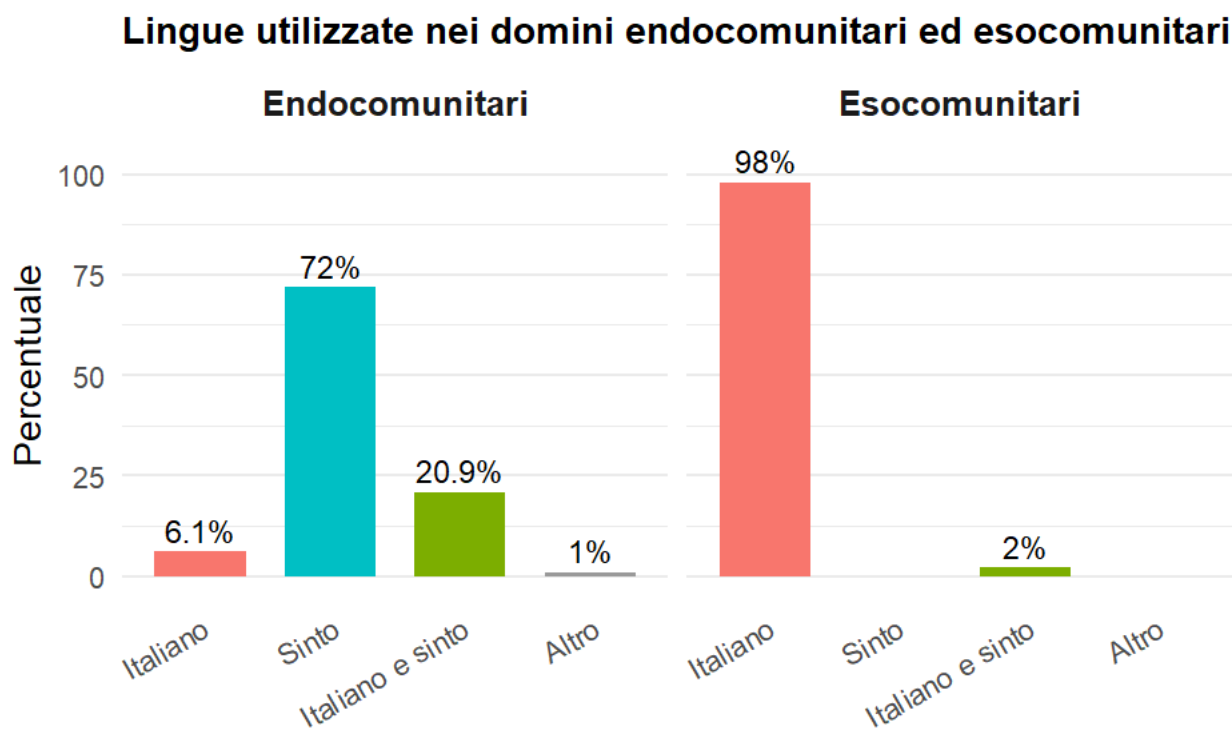


Figura 10: Lingue utilizzate nei domini endocomunitari ed esocomunitari (aggregato).

Una visualizzazione della Figura mostra che, negli usi endocomunitari, il sinto lombardo da solo costituisce la lingua quantitativamente dominante (72%). Tuttavia, una percentuale non trascurabile di risposte (20.9%) indica l'uso sia del sinto lombardo che dell'italiano; in alcuni casi, si segnala anche l'uso esclusivo dell'italiano (6.1%). La presenza relativamente elevata di risposte che includono sia il sinto lombardo che l'italiano sembra suggerire che i parlanti percepiscano come frequente il ricorso a pratiche bilingui nella comunicazione endocomunitaria. Per quanto riguarda invece gli usi esocomunitari si osserva, come del tutto prevedibile, un impiego pressoché esclusivo dell'italiano (98%), sebbene il 2% delle risposte indichi anche l'utilizzo simultaneo del sinto lombardo e dell'italiano. Uno sguardo più attento a queste risposte mostra che esse si riferiscono sempre alla comunicazione con amici non sinti, in grado di comprendere alcune parole o espressioni di base in sinto lombardo. Data la maggiore variazione osservata nei contesti endocomunitari, nella prossima sezione viene proposta un'analisi più dettagliata degli usi linguistici in questo dominio.

4.4.2 Usi linguistici nei domini endocomunitari

In questo paragrafo vengono approfondite le scelte linguistiche dei sinti lombardi nei domini endocomunitari, attraverso un'analisi dettagliata delle risposte alle domande relative alla lingua utilizzata con madre, padre, fratelli/sorelle, partner/coniuge, figli, parenti anziani, bambini e amici (cfr. domande 3-10). Le risposte sono indicate nella Figura 11. Si noti che le domande relative alla

comunicazione con partner e figli hanno raccolto un numero inferiore di risposte, poiché non tutti i partecipanti al questionario hanno partner o figli.

Variazione degli usi linguistici nei contesti endocomunitari

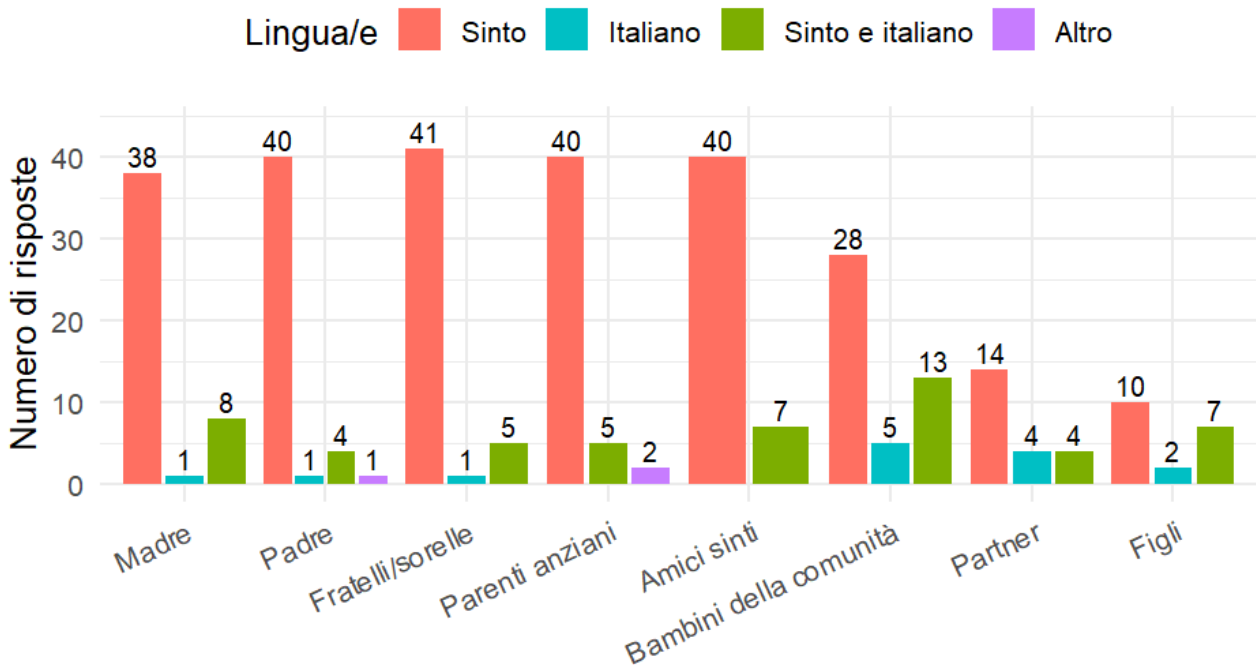


Figura 11: Lingue utilizzate nei domini endocomunitari al variare degli interlocutori.

A livello generale, i dati mostrano che il sinto lombardo si conferma come lingua dominante della comunicazione endocomunitaria, risultando la lingua più frequentemente selezionata in relazione a tutte le categorie di interlocutori considerate. Tuttavia, mentre il numero di quanti dichiarano di utilizzare soltanto il sinto lombardo nella comunicazione con madre, padre, fratelli e sorelle, parenti anziani e amici si mantiene molto stabile (circa 40 osservazioni per categoria), il numero di quanti dichiarano di utilizzare soltanto il sinto con i bambini si abbassa piuttosto notevolmente. In questo caso, soltanto 28 parlanti dichiarano di utilizzare esclusivamente il sinto lombardo, mentre aumenta il numero di coloro che dichiarano di utilizzare sia il sinto lombardo che l'italiano (N=13) o soltanto l'italiano (N=5). Un andamento analogo sembra emergere dalle risposte relative alla comunicazione con i figli: dieci parlanti dichiarano di utilizzare esclusivamente il sinto lombardo, sette il sinto lombardo e l'italiano e due l'italiano. Tali dati devono tuttavia essere interpretati con cautela, poiché l'età dei "figli" è inevitabilmente correlata all'età dei genitori che hanno risposto al questionario, rendendo le due domande solo parzialmente confrontabili.

Per verificare se le differenze osservate nelle scelte linguistiche endocomunitarie siano statisticamente significative, è stato condotto un test del chi-quadro di indipendenza. A tal fine, sono

state escluse dall'analisi le risposte relative alla comunicazione con figli e partner a causa del minore numero di risposte. Il test risulta significativo ($\chi^2 = 23.91$, $p \approx 0.007$), indicando una dipendenza tra interlocutore e uso linguistico. L'analisi dei residui standardizzati conferma che la significatività è principalmente legata alla categoria “bambini della comunità”, nella comunicazione con i quali l'uso esclusivo del sinto risulta fortemente sottorappresentato rispetto ai valori attesi, mentre l'italiano e il bilinguismo sinto-italiano risultano sovrarappresentati; cfr. Tabella 6.

Interlocutore	Solo sinto	Sinto + italiano	Solo italiano
Madre	-0.2148647	-0.3416183	0.38992747
Padre	1.3225369	-0.2914383	-1.28214697
Fratelli/sorelle	1.0337548	-0.3416183	-0.94903834
Parenti anziani	1.3225369	-1.2640696	-0.82798494
Amici sinti	0.6175483	-1.2974593	-0.05639447
Bambini della comunità	-4.0706434	3.5396050	2.71239378

Tabella 6: Residui standardizzati del test χ^2 per interlocutore e scelta linguistica nei domini endocomunitari.

L'idea di un uso relativamente ridotto del sinto lombardo nella comunicazione con i bambini della comunità è peraltro ampiamente condivisa dai parlanti stessi, come emerge anche dalle interviste. Si considerino, a titolo esemplificativo, gli estratti riportati negli esempi (2–4).

- (2) *me quando jom cun u tine čave na vel=ma miga spontaneo ta rakara lensal in sintu, rakarava pena in gadžu*

‘io quando sono con i bambini, non mi viene spontaneo parlare con loro in sinto, parlo solo in gagio’

- (3) *mengar čave d'akana, mur nivudu i=lu sei anni sette / na rakarela miga in sintu rakarela pena in gadžu (.) me pinava che i i televisione il cellulare*

‘i nostri bambini di adesso, mio nipote ha sette anni, non parla in sinto, parla solo il gagio. Io dico che è la televisione, il cellulare’

- (4) *mur čau rakarela pena, rakarela pena gadžu / u sintu xoel=lu krol ma na rakarel=lu miga / infatti akana che vel=lo pisa più baru i=ta inisiara ta sikara=lu*
 ‘mio figlio parla soltanto, parla soltanto in gagio. Il sinto lo capisce tutto, ma non lo parla. Infatti adesso che diventa un po’ più grande, devo iniziare a insegnarglielo’

Alla luce di quanto osservato, sembra possibile ipotizzare un potenziale indebolimento della trasmissione intergenerazionale del sinto lombardo, qualificando la varietà come un possibile “codice in declino di parlanti”, secondo la tipologia proposta da Dal Negro & Iannàccaro (2003); cfr. anche Dell’Aquila & Iannàccaro (2006: 65). Va tuttavia sottolineato che pressoché tutti i giovani e giovanissimi con cui ho avuto modo di collaborare più da vicino (in particolare bambini delle scuole elementari e ragazzi delle scuole medie) mantengono un uso attivo della lingua sinto, mentre una vera e propria interruzione della trasmissione intergenerazionale del sinto appare limitata ad alcune famiglie isolate.

4.4.3 Usi linguistici con parenti sinti non lombardi

Dopo aver analizzato gli usi linguistici all’interno e all’esterno della comunità, questa sezione è dedicata agli usi linguistici dei sinti lombardi nella comunicazione con parenti sinti non lombardi, sulla base delle risposte alla domanda “Hai parenti o amici sinti che parlano lingue diverse dal sinto lombardo? Se sì, che lingua/e usi con loro?” (cfr. domanda 23). L’ipotesi di partenza era che l’esistenza di reti parentali relativamente dense con sinti appartenenti ad altri gruppi, che non parlano sinto lombardo, possa favorire presso i sinti lombardi l’acquisizione di competenze parziali in altre varietà di romaní. Considerare questo aspetto è rilevante soprattutto alla luce degli sviluppi più recenti della nozione di repertorio linguistico, che sottolineano l’importanza delle competenze linguistiche situate e parziali (Blommaert & Backus, 2013; cfr. anche Capitolo 1).

Le risposte alla domanda 23 sono presentate in due grafici distinti. Il primo (Figura 12) riporta le risposte, positive o negative, relative alla presenza di parenti sinti di altri gruppi. Il secondo (Figura 13) mostra le lingue che i parlanti che hanno risposto affermativamente dichiarano di utilizzare nella comunicazione con tali parenti. Data la forte variabilità delle risposte alla seconda domanda, i dati sono presentati in forma aggregata e percentuale, al fine di fornire una prima visualizzazione d’insieme, senza distinguere se i parlanti abbiano indicato una sola lingua o più lingue; tale distinzione verrà discussa nel commento ai risultati.

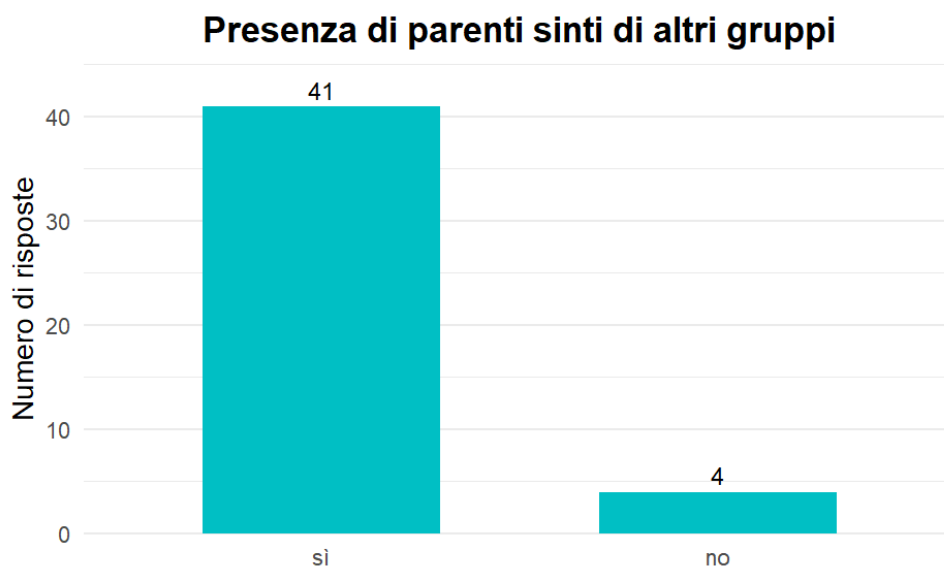


Figura 12: Risposte alla domanda “Hai parenti o amici sinti che parlano lingue diverse dal sinto lombardo?”.

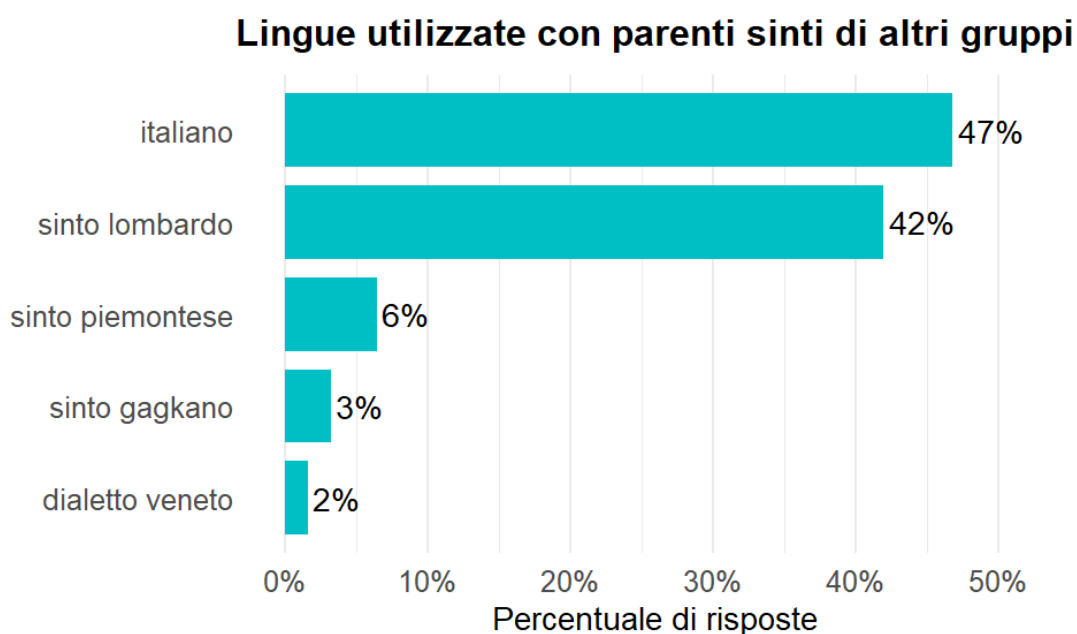


Figura 13: Lingue dichiarate con parenti sinti non lombardi (aggregato).

Dal grafico in Figura 12 emerge innanzitutto che 41 rispondenti su 45 (in due casi non è stato possibile registrare una risposta) dichiarano di avere parenti che appartengono ad altri gruppi sinti che usano varietà differenti dal sinto lombardo. Tale dato sembra dunque confermare l'importanza di considerare questi contesti comunicativi in una panoramica delle scelte linguistiche quotidiane dei sinti lombardi. Dai commenti dei rispondenti emerge che tali parenti sinti appartengono al gruppo

*gadžkano*¹, piemontese o emiliano. In generale, i parlanti riferiscono di incontrare i propri parenti non lombardi soprattutto in occasione di feste; tuttavia, laddove sussistano legami parentali più stretti, le visite possono anche avvenire con cadenza settimanale.

Per quanto riguarda le lingue usate nella comunicazione con parlanti sinti di varietà diverse, si osserva, come anticipato, una notevole variazione nelle risposte (cfr. Figura 13). Inoltre, 21 delle 41 risposte considerate riportano l'utilizzo di più di una lingua con i parenti sinti non lombardi: in alcuni casi, i parlanti distinguono la lingua in base all'interlocutore (p.es. *con i sinti piemontesi parlo piemontese, con i sinti gadžkani parlo il sinto lombardo*); in altri, dichiarano l'utilizzo di più lingue con la stessa tipologia di interlocutori, suggerendo il ricorso a pratiche comunicative multilingui.

In particolare, le lingue menzionate con maggiore frequenza dai rispondenti nella comunicazione con i parenti non lombardi sono l'italiano e il sinto lombardo. Se l'uso dell'italiano non stupisce, in quanto unica risorsa certamente condivisa nei repertori dei vari gruppi sinti, la frequenza con cui viene menzionato il sinto lombardo sembra indicare una conoscenza, almeno passiva, di questa varietà da parte di sinti appartenenti ad altri gruppi. A sostegno di questa interpretazione, è forse interessante menzionare che, durante la mia ricerca sul campo, ho potuto assistere a una conversazione tra un giovane sinto di Reggio Emilia, di circa di 18 anni, in visita a Pavia, e i suoi cugini, un gruppo di adolescenti residenti nell'insediamento di piazzale Europa. Durante questa interazione, la lingua utilizzata nelle comunicazioni dirette tra il giovane sinto reggiano e i cugini pavesi era quasi esclusivamente l'italiano; tra loro, tuttavia, i sinti pavesi alternavano liberamente l'italiano e il sinto lombardo, senza mostrare particolare preoccupazione per la comprensione da parte del cugino, probabilmente confidando in una qualche competenza passiva della varietà da parte di quest'ultimo. Tornando ai dati dal questionario, cinque rispondenti dichiarano inoltre di utilizzare il sinto piemontese (verosimilmente inteso come varietà italoromanza, come già visto precedentemente) nella comunicazione con i sinti piemontesi; tuttavia, numerose risposte indicano anche l'utilizzo dell'italiano con questi interlocutori, e più raramente del sinto lombardo. A questo proposito, una parlante segnala come per i sinti piemontesi sia particolarmente complesso esprimersi in sinto lombardo e comprenderlo, in quanto essi hanno abbandonato la *romaní*, mentre per i sinti lombardi utilizzare e comprendere il dialetto piemontese può risultare relativamente agevole a causa della sua vicinanza strutturale con l'italiano.

Nella comunicazione con i sinti *gağkani*, sembra invece prevalere l'uso dell'italiano e del sinto lombardo, probabilmente facendo affidamento su una competenza passiva di quest'ultimo da parte

¹ Benché l'area di residenza dei sinti *gadžkani* sia tipicamente il nord-est italiano, va notato che nella città di Tortona, situata a meno di 60 km di Pavia, risiede una comunità di sinti *gadžkani* che intrattiene rapporti di parentela con i sinti di Pavia di entrambi gli insediamenti.

degli interlocutori *gaḡkani*. Soltanto due parlanti dichiarano di utilizzare il sinto *gaḡkano* nella comunicazione con i parenti *gaḡkani*. Se, da un lato, non è infrequente l'osservazione dei sinti lombardi che il sinto *gaḡkano* costituisce una varietà di romaní particolarmente complessa da comprendere e produrre, dall'altro un'esposizione a questa varietà è garantita da una buona produzione musicale di alcuni artisti *gaḡkani*, in primo luogo il già citato gruppo dei *Romane Chave* (cfr. Capitolo 2).

Infine, una parlante segnala l'uso del dialetto veneto nella comunicazione con parenti giostrai, appartenenti al mondo dello spettacolo viaggiante. Si tratta verosimilmente della varietà italomanzesca a base veneta diffusa tra i giostrai sinti dell'Italia settentrionale (cfr. Capitolo 2).

In sintesi, le risposte al questionario delineano una situazione fortemente eterogenea, che sembra indicare l'attivazione di strategie plurilingui complesse e situate, fondate su un impiego flessibile delle risorse disponibili nei repertori linguistici dei parlanti. In linea con l'ipotesi di partenza, l'esistenza di reti parentali tra diversi gruppi sinti sembra dunque contribuire ad ampliare tali repertori, favorendo l'acquisizione di competenze, spesso di tipo passivo, in altre varietà di romaní o in varietà italomanzesche.

4.4.4 Usi linguistici in altri domini: la tecnologia e i contesti religiosi

A conclusione della sezione dedicata alle scelte linguistiche e alla variazione nei domini d'uso, accenniamo di seguito agli usi linguistici in due domini specifici, ovvero la tecnologia e il web (§ 4.4.4.1) e le cerimonie religiose (§ 4.4.4.2).

4.4.4.1 La tecnologia e il web

Come mostrato da numerosi studi negli ultimi decenni, lo sviluppo della tecnologia e del Web 2.0 ha aperto nuovi spazi di utilizzo per le lingue minoritarie, favorendone impieghi creativi e funzionalmente differenziati (per la romaní, cfr. almeno Halwachs, 2012; Leggio, 2020). Per indagare le scelte linguistiche dei sinti lombardi in questo particolare dominio, il questionario includeva due domande: “In quale lingua parli al telefono con altre persone sinte?” e “In quale lingua mandi messaggi ad altre persone sinte?” (cfr. domande 20-21). Le risposte a queste domande sono riportate nelle Figure 14 e 15.

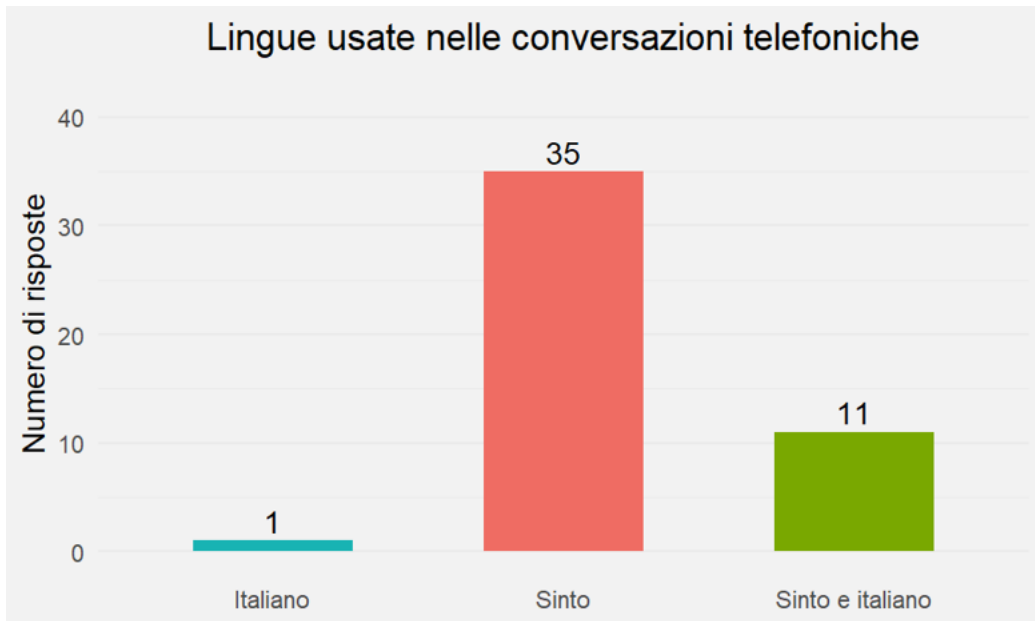


Figura 14: Risposte alla domanda “In quale lingua parli al telefono con altre persone sinte?”.

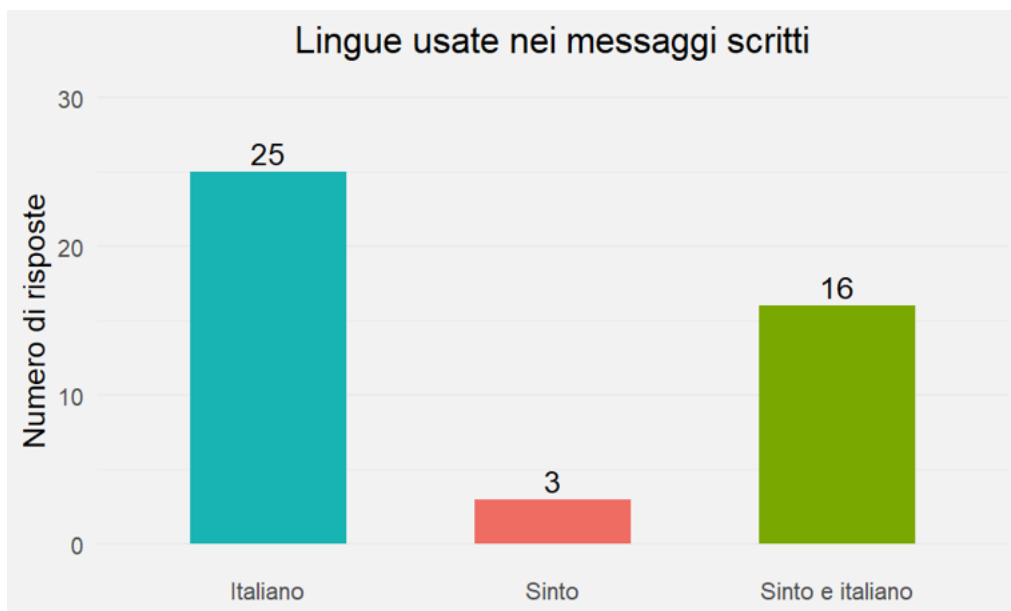


Figura 15: Risposte alla domanda “In quale lingua scrivi messaggi ad altre persone sinte?”.

Come si può osservare, nonostante si trattasse di domande aperte, i parlanti hanno indicato esclusivamente tre possibili configurazioni linguistiche: l’uso del solo sinto lombardo, del solo italiano oppure l’uso simultaneo di entrambe le lingue.

Più nello specifico, la maggior parte dei rispondenti dichiara di utilizzare prevalentemente il sinto (N=35), più raramente il sinto insieme all’italiano (N=11), durante le conversazioni telefoniche con altri sinti (cfr. Figura 14). Viceversa, la maggior parte dei rispondenti dichiara di utilizzare soltanto l’italiano (N=25), più raramente il sinto insieme all’italiano (N=16), nei messaggi scritti (cfr. Figura 15).

Questa distribuzione conferma la persistenza, anche all'interno del dominio tecnologico, di una distinzione funzionale tra una dimensione orale, principalmente associata al sinto lombardo, e una dimensione scritta, prevalentemente affidata all'italiano (cfr. Scala, 2015). A conferma di ciò, diversi parlanti osservano che su piattaforme come WhatsApp, che consentono comunicazioni sia scritte sia vocali, essi tendono a utilizzare l'italiano nei messaggi testuali e il sinto lombardo nei messaggi vocali.

I sinti lombardi di Pavia utilizzano inoltre abitualmente *social network* quali Facebook, Instagram o TikTok. Sulla base del materiale osservato nel corso della ricerca sul campo nelle pagine degli utenti (non condiviso qui per ovvie ragioni di *privacy*), si nota che sui *social* l'italiano costituisce la lingua largamente dominante, sebbene non manchino inserzioni in sinto. Tali inserzioni assumono spesso la forma di formule fisse con funzioni identitarie o espressive. È ad esempio frequente l'uso, all'interno di post interamente in italiano, di sintagmi in romaní del tipo [bare X anni] (p.es. *bare 24 anni*), impiegati come augurio di compleanno. Come verrà discusso nel Capitolo 5, l'inserzione di sintagmi interamente italiani indicanti l'età può essere considerata una modalità “non marcata” di parlare in sinto lombardo.

4.4.4.2 I contesti religiosi

Come già accennato in precedenza, i sinti di piazzale Europa e quelli di via Bramante partecipano alle cerimonie religiose (rispettivamente cattoliche ed evangeliche) nelle chiese collocate all'interno dei loro rispettivi insediamenti.

Le cerimonie cattoliche sono celebrate da un sacerdote non sinto che vive da tempo all'interno della comunità. Esse avvengono in italiano, sebbene alcuni canti siano in sinto lombardo. Tali testi sono stati composti da Don Mario Riboldi, sacerdote che ha operato a lungo presso diverse comunità rom e sinte con finalità evangelizzatrici, producendo canti, preghiere e traduzioni di testi religiosi in diverse varietà di romaní, tra cui anche il vangelo di Marco in sinto lombardo, menzionato nel Capitolo 2. È interessante notare che questi testi circolano anche nell'ambito delle attività di catechismo di piazzale Europa, comunque svolte in italiano da catechisti *gagi*, e rappresentano, a mia conoscenza, le uniche testimonianze scritte (su supporto cartaceo) del sinto lombardo.

Le cerimonie evangeliche sono invece officiate da un pastore evangelico che vive all'interno della comunità, il quale ha partecipato direttamente alla ricerca prendendo parte al questionario e alle interviste. Anche queste cerimonie si svolgono in italiano, compresi i canti e le preghiere. Secondo le testimonianze del pastore e di altri parlanti, la scelta dell'italiano risponde alla volontà di non escludere dai culti i sinti che non conoscono la romaní e, più in generale, eventuali *gagi* interessati a partecipare. In occasioni particolari, come ad esempio durante i battesimi, può tuttavia verificarsi un

ricorso al sinto lombardo. Diversamente da quanto osservato per il contesto cattolico di piazzale Europa, non mi risultano esistere testi scritti in sinto lombardo destinati al culto evangelico.

Nel complesso, l'uso della lingua maggioritaria co-territoriale durante le cerimonie religiose è in linea con quanto riscontrato nella maggior parte delle comunità romaní cattoliche ed evangeliche in Europa (cfr. van den Heuvel, 2020: 465-467 per una panoramica).

È infine interessante notare che le pratiche religiose costituiscono per i sinti lombardi anche un'importante occasione di contatto con sinti di altri gruppi o rom. I sinti evangelici partecipano, ad esempio, a grandi raduni organizzati dalla Missione Evangelica Zigana, talvolta della durata di più settimane, che coinvolgono rom e sinti "convertiti" provenienti da diverse aree d'Italia e dall'estero. Analogamente, i sinti cattolici di piazzale Europa possono prendere parte a incontri con altri sinti cattolici, spesso *gadžkani* o piemontesi, organizzati dalla rete di sacerdoti che operano a contatto con queste comunità. Al momento, tuttavia, non si dispone di dati sugli usi linguistici prevalenti in tali contesti intercomunitari.

4.5 Atteggiamenti verso il sinto lombardo e verso il contatto linguistico

Quest'ultima sezione si propone di analizzare gli atteggiamenti dei sinti lombardi nei confronti del sinto lombardo e del contatto linguistico tra sinto lombardo e italiano. "Atteggiamento" è qui impiegato come termine ombrello per indicare sia le opinioni esplicite su lingue o comportamenti linguistici, così come elicitate tramite le domande del questionario, sia le riflessioni metalinguistiche che emergono nel corso delle interviste. Si è tuttavia consapevoli che, nell'accezione utilizzata in questo lavoro, la nozione di atteggiamento non coincide con quella utilizzata soprattutto nell'ambito della psicologia sociale (cfr. p.es. Baker, 1992), dove il termine rimanda a disposizioni mentalmente stabili e latenti, difficilmente indagabili mediante gli strumenti qui adottati. Nello specifico, il § 4.5.1 prende in esame gli atteggiamenti dei parlanti nei confronti del sinto lombardo, mentre il § 4.5.2 è dedicato agli atteggiamenti verso il contatto linguistico con l'italiano e il parlato bilingue.

4.5.1 Atteggiamenti verso il sinto lombardo

Le analisi presentate nelle precedenti sezioni hanno mostrato che il sinto lombardo occupa una posizione centrale nel repertorio linguistico dei sinti lombardi, configurandosi non soltanto come la lingua primaria della comunicazione endocomunitaria, ma anche come la "lingua madre" dichiarata dalla maggior parte dei parlanti. Seguendo questa linea, altre domande del questionario erano focalizzate più direttamente sul ruolo del sinto lombardo nella definizione dell'identità linguistica dei parlanti. In particolare, analizziamo qui le risposte alle domande: "Sei fiero di parlare sinto?" (cfr.

domanda 34) e “Conoscere il sinto per essere sinti è?...” (cfr. domanda 29). I grafici nelle Figure 16 e 17 mostrano le risposte alle rispettive domande.

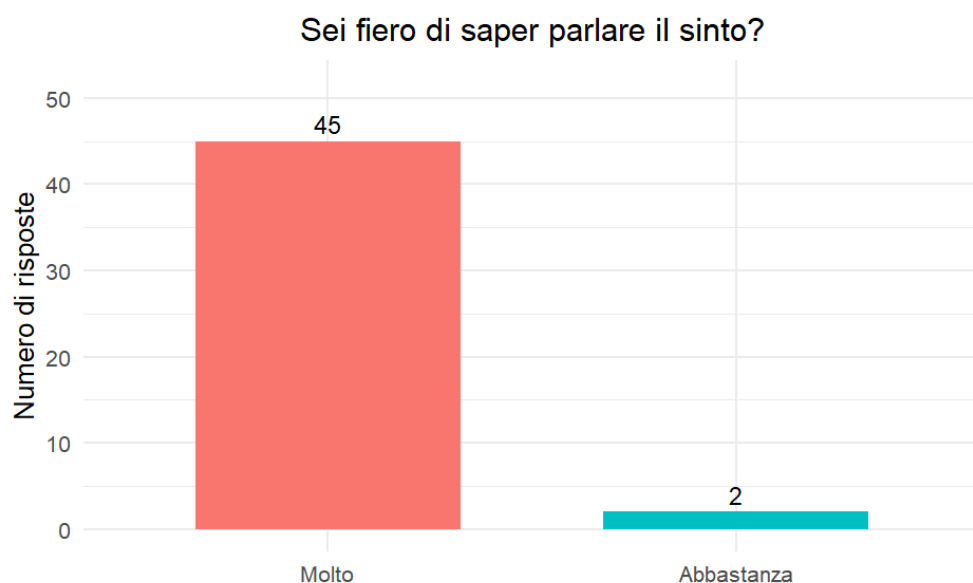


Figura 16: Risposte alla domanda “Sei fiero di parlare il sinto?”.

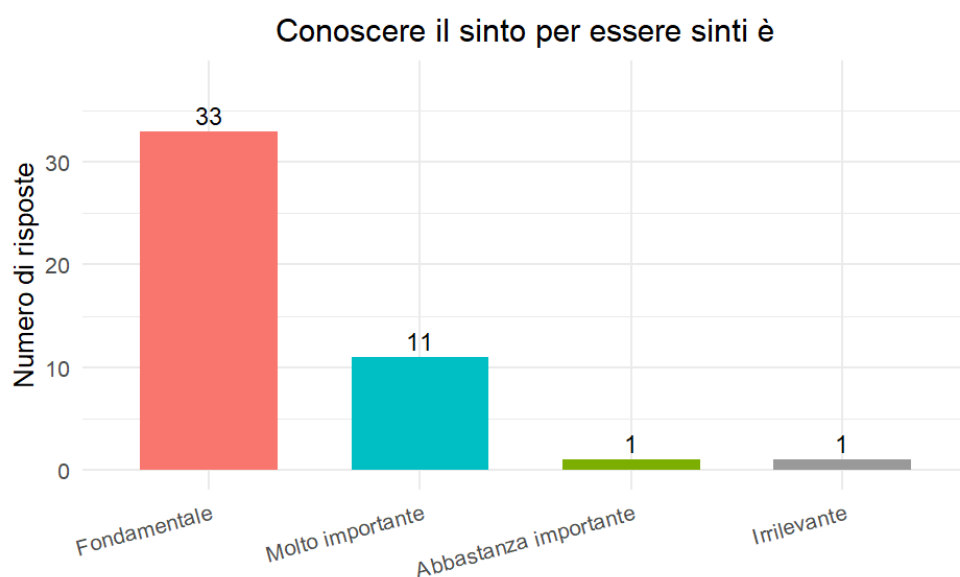


Figura 17: Risposte alla domanda “Conoscere il sinto per essere sinti è...?”.

Quanto alla prima domanda (cfr. Figura 16), 45 parlanti su 47 dichiarano di essere “molto fieri” di parlare il sinto. Analogamente, alla seconda domanda, 44 parlanti su 47 dichiarano che conoscere il sinto per essere sinti è “fondamentale” o “molto importante” (cfr. Figura 17). Può essere rilevante notare che i due parlanti che si discostano dalle tendenze generali in entrambe le domande, mostrando atteggiamenti apparentemente più negativi nei confronti del sinto, sono i medesimi e appartengono all'insediamento di via Bramante, dove sono molto attivi nella missione evangelica. Un'ipotesi di

ricerca, che potrebbe essere esplorata in studi futuri, è che l'adesione alla religione evangelica possa comportare un cambiamento negli atteggiamenti verso il sinto lombardo. Tuttavia, allo stato attuale, la dimensione del campione non consente di trarre conclusioni affidabili in merito a tale relazione. Nel complesso, questi dati sembrano confermare ulteriormente il ruolo centrale del sinto lombardo nel repertorio linguistico dei sinti lombardi e la presenza di atteggiamenti positivi nei suoi confronti. È tuttavia necessaria una particolare cautela metodologica nell'interpretazione di tali risultati. Le due domande considerate erano infatti particolarmente sensibili, in quanto miravano a elicitare direttamente le opinioni su una lingua, il sinto lombardo, che i parlanti sapevano essere al centro degli obiettivi della ricerca. Non si può pertanto escludere che, in alcuni casi, i rispondenti siano stati indotti a fornire valutazioni più positive rispetto a quelle che emergerebbero in contesti meno monitorati. Ciò non invalida i risultati emersi, ma suggerisce di interpretarli come indicatori di rappresentazioni linguistiche in qualche modo co-costruite con il ricercatore, più che come riflessi non mediati di atteggiamenti individuali stabili.

4.5.2 Atteggiamenti verso il contatto linguistico e la nozione di 'sinto stretto'

Questo paragrafo è dedicato agli atteggiamenti dei sinti lombardi nei confronti di un altro elemento centrale della loro identità linguistica, ovvero il contatto linguistico tra sinto lombardo e italiano e il ricorso a pratiche discorsive bilingui. Anche in questo caso, l'analisi prende avvio dai dati del questionario per poi lasciare spazio alla voce dei parlanti attraverso estratti di interviste.

La prima domanda rilevante in questo senso (cfr. domanda 31) recitava quanto segue: “Prima di andare a scuola, i bambini dovrebbero crescere parlando...?”. Le risposte sono riportate nella Figura 18.

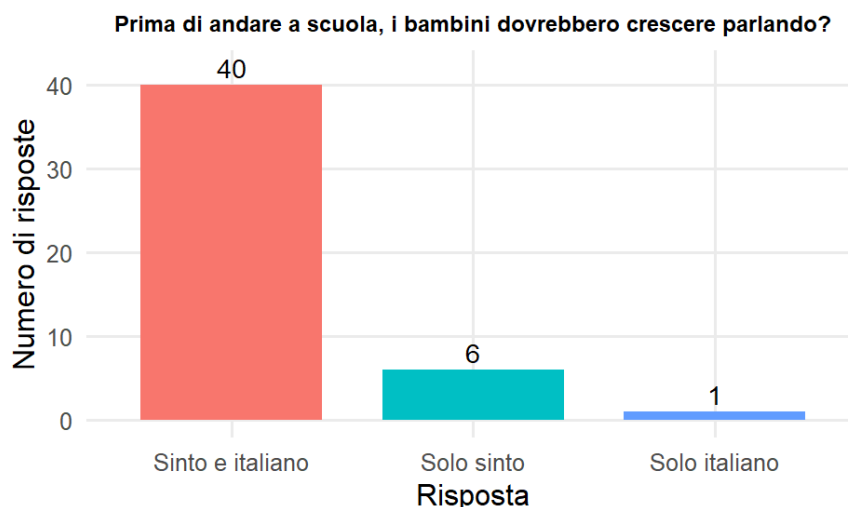


Figura 18: Risposte alla domanda “Prima di andare a scuola, i bambini dovrebbero crescere parlando...?”.

Come si osserva, 40 parlanti su 47 dichiarano che sarebbe preferibile che i bambini crescessero parlando sia il sinto lombardo che l'italiano prima dell'ingresso a scuola; sei rispondenti indicano invece una preferenza per l'acquisizione esclusiva del sinto, mentre in un solo caso una parlante ritiene preferibile l'apprendimento del solo italiano. Nel complesso, questi dati indicano un atteggiamento favorevole verso il ricorso a pratiche multilingui durante la socializzazione primaria. Questo risultato trova riscontro in un recente sondaggio condotto tra parlanti rom rumeni da Padure & Adamou (2024): in questo studio, gli autori riportano che il 96% dei rispondenti rumeni ritiene che i bambini dovrebbero crescere imparando romaní, rumeno e almeno un'altra lingua in ambito domestico.

Le due domande successive (cfr. domande 32-33) erano volte a elicitar più nello specifico gli atteggiamenti dei parlanti nei confronti del parlato bilingue. La prima, di carattere più valutativo, mirava a raccogliere un'opinione sulla pratica di “mescolare” il sinto lombardo e l'italiano nella conversazione spontanea; la seconda, ancorata all'esperienza linguistica diretta dei rispondenti, intendeva invece rilevare la frequenza con cui tale mescolamento avviene nei loro usi quotidiani. Le Figure 19 e 20 consentono di confrontare le risposte alle due domande.

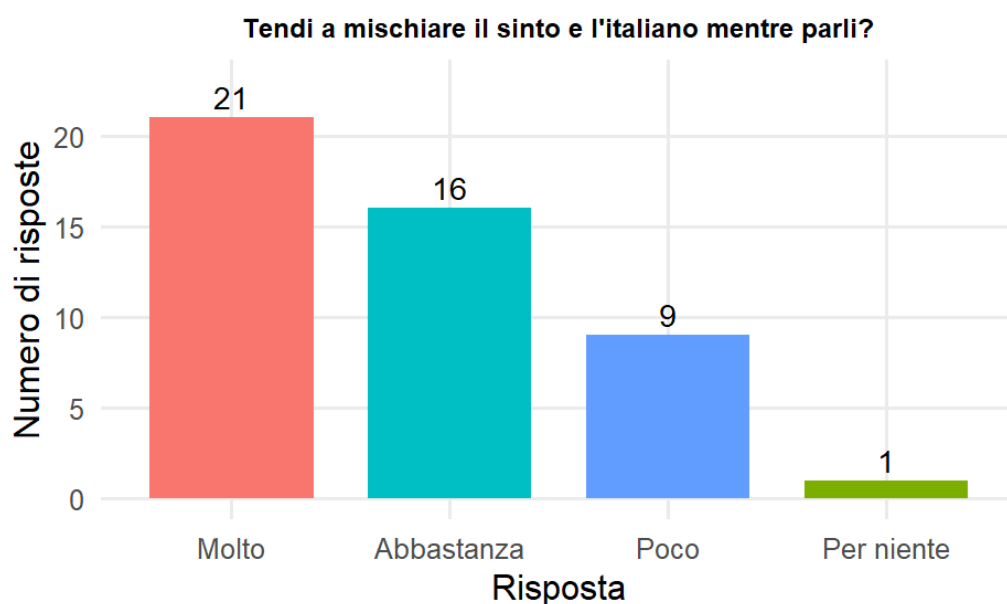


Figura 19: Risposte alla domanda “Tendi a mischiare il sinto e l'italiano mentre parli?”.

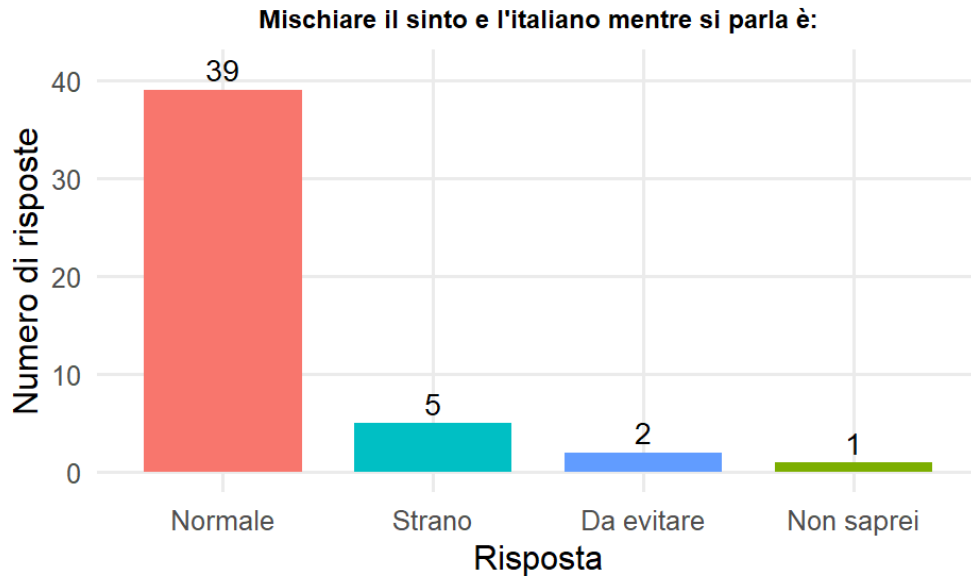


Figura 20: Risposte alla domanda “Mischiare il sinto e l’italiano mentre si parla è...”.

Per quanto riguarda la prima domanda (cfr. Figura 19), la maggioranza dei parlanti (N=37) dichiara di mescolare “molto” o “abbastanza” le lingue del proprio repertorio nelle conversazioni quotidiane. Significativamente, una rispondente osserva che mescolare le lingue è “strano” per le generazioni più anziane, ma “normale” per i bambini. Analogamente, alla seconda domanda (cfr. Figura 20), 40 rispondenti dichiarano di considerare “normale” l’utilizzo simultaneo del sinto lombardo e dell’italiano nella conversazione.

Nel corso delle interviste, le riflessioni metalinguistiche dei parlanti sul “mescolamento” del sinto lombardo e dell’italiano nella conversazione risultano particolarmente frequenti. In questo quadro, ricorre spesso la distinzione tra il sinto parlato oggi e un *sinto stretto* generalmente associato alle generazioni più anziane. Si considerino i tre esempi di seguito (5-7):

- (5) se dopu *sunea i* registrasiun, *sunea* che *n’i miga ja čaču sinto*
 ‘se dopo ascolti la registrazione, senti che non è un vero sinto’
 i but i but pure che *rakarena u sintu* stretu
 ‘ci sono molti, ci sono molti anziani che parlano il sinto stretto’
 ma u sintu- stretu i pisa pure pape che *rakarena=lo*
 ‘ma il sinto stretto ci sono ancora pochi anziani che lo parlano’
 cioè *u pure ua, u terne na*
 ‘cioè, gli anziani sì, i giovani no’

- (6) MF ma *u terni rakarena* diverso confronto *u pure*?
'ma i giovani parlano diversamente in confronto agli anziani?'
- B ehm *na gina* / certi paroli *ua*, certi parola *ua*
'ehm, non so, certe parole sì, certe parole sì'
tipo *men pinaja* magari *pape* / *rakaraja* più moderno, *gagizzaraja* u *sinto*
'tipo noi diciamo magari ancora... parliamo più moderno, gagizziamo il *sinto*'
ehm magari *me pinava i luce* rivola magari *jek* da *mar zii* più *puri pinena i lehta*
'ehm, magari io dico *la luce*, arriva magari uno dei miei zii più anziani e dicono *la lehta*'
me pinava: (.) magari *me pinava* macchina *vela jek* più *puro* e *pinela massina*
'io dico... magari io dico *macchina*, arriva uno più vecchio e dice *massina*'
- (7) K ma anca perché *na rakaraja* mai *u sinto* stretto
J 'ma anche perché non parliamo mai il *sinto* stretto'
tipo *men ginaja* pena fino a *so* / sette otto nove dieci fino al cento *ginaj=li* normale
'tipo noi contiamo solo fino a sei... sette, otto, nove, dieci, fino al cento, li contiamo normale'
- MF *ar u gadže*
'come i gagi'
- K *ar u gadže*
'come i gagi'
- J perché se no *i=li* u numari / *jek duj trin star panč sov exta oxta enja diese*
'perché se no ci sono i numeri: *jek duj trin star panč sov exta oxta enja diese*'
- MF e perché *na pinena=li*?
- J eh perché: / boh *i ja* generasion *nevi*, *jam* moderni
'eh perché, boh, è una generazione nuova, siamo moderni'
- K eh *jam* moderni *na džinas=lu* miga
'eh siamo moderni, non lo sappiamo'

Come questi esempi mostrano, la nozione di *sinto stretto* sembra dunque fungere da categoria emica per concettualizzare la variazione e il cambiamento linguistico indotti dal contatto. A questo proposito, è interessante notare che, nell'estratto in (6), una parlante definisce il *sinto* parlato oggi

come un sinto “gagizzato”, ricorrendo alla neoformazione verbale su base italiana *gagizzaraja* ‘gagizziamo’².

Più in generale, anche sulla base di estratti non riportati qui, è possibile osservare che gli esempi lessicali forniti dai parlanti per confrontare il sinto parlato oggi con il *sinto stretto* sono piuttosto ricorrenti (p.es. *caffè* vs. *xantigo*, *luce* vs. *lexta*, *macchina* vs. *massina*, *farina* vs. *jarxu*). Inoltre, come si evince dall’esempio in (7), la distinzione tra il sinto parlato oggi e il *sinto stretto* si estende anche al dominio dei numerali: i parlanti sono consapevoli di utilizzare prestiti di origine italiana per i numerali superiori al sei nella conversazione spontanea (cfr. anche Capitoli 5 e 6), nonostante siano consci dell’esistenza di equivalenti preromanzi che coprono alcune cifre successive (in particolare *eftà* ‘sette’, *otxò* ‘otto’ e *injà* ‘nove’, di origine greco bizantina). Complessivamente, questi esempi ricorrenti sembrano suggerire la possibile circolazione comunitaria di un discorso metalinguistico sul cambiamento linguistico e sul contatto con l’italiano.

La nozione di *sinto stretto* è attestata anche in altre comunità sinte. In particolare, Trevisan (2005) osserva che, presso i sinti emiliani, essa si riferisce a una pratica discorsiva in cui elementi lessicali romaní vengono inseriti all’interno di una varietà italo-romanza a base veneta che funge da codice endocomunitario (cfr. Capitolo 2). In quel contesto, il *sinto stretto* si oppone dunque alla stessa varietà italo-romanza senza inserzioni romaní. Nel caso dei sinti lombardi, invece, il contrasto tra sinto e *sinto stretto* rimane interno al dominio della romaní, opponendo una varietà caratterizzata da un ricorso limitato a inserzioni italo-romanze a una varietà che ne fa un uso esteso.

Sebbene il *sinto stretto* sia comunemente associato dai parlanti alle generazioni anziane o al parlato di specifici individui, in alcuni casi questa nozione sembra caricarsi di una connotazione quasi “mitica”, evocando un passato percepito come più tradizionale, caratterizzato da una maggiore distanza dalla società maggioritaria e da un rapporto più stretto con la natura. Si considerino, per esempio, le osservazioni di un parlante emerse poco dopo che egli ha introdotto la nozione di *sinto stretto* (8).

- (8) *mengar sinti pure cun pengar garavani pengar graj dinglan*
 ‘i nostri sinti vecchi con le loro carovane, il loro cavallo davanti’
fermonas maskar u feldi e čenas
 ‘si fermavano in mezzo ai campi e stavano’
karenas penge u jak in maskar / čuenas=pe dendur tine panja (...)
 ‘si facevano il fuoco in mezzo, si lavavano dentro l’acqua’

² Sulle strategie di integrazione dei prestiti verbali italo-romanzi in sinto lombardo, cfr. Capitolo 6.

kaja is i vita / i sukar vita

‘quella era la vita, la bella vita’

Nonostante i valori di autenticità assegnati al *sinto stretto* in opposizione al *sinto parlato* oggi, un’analisi più attenta degli esempi mostra che le valutazioni associate dai sinti lombardi al *sinto parlato* oggi sono comunque tendenzialmente neutre o positive, come emerge, per esempio, dall’associazione del *sinto* odierno a nozioni di “modernità” (cfr. 6-7). Alcuni parlanti sottolineano inoltre l’utilità pratica dell’inserzione di elementi italo-romanzi nel parlato in *sinto*, contrapponendo questa pratica allo sforzo cognitivo che sarebbe invece richiesto dalla produzione di un discorso monolingue in *sinto*; cfr. (9).

- (9) *na na i fedar kjake* perchè in italiano trovi l’aggettivo per ogni cosa
‘no no è meglio così³ perché in italiano trovi l’aggettivo per ogni cosa’
indur sintu n’i=lu miga / vela trop confuso kjake
‘in *sinto* non c’è... viene troppo confuso così’

Non si registrano invece, nei dati raccolti, atteggiamenti esplicitamente negativi nei confronti del mescolamento del *sinto lombardo* e dell’italiano.

In sintesi, questa sezione ha mostrato come i sinti lombardi di Pavia esprimano atteggiamenti tendenzialmente molto positivi sia nei confronti del bilinguismo con l’italiano, sia verso il ricorso al parlato bilingue. Tale risultato, sebbene in contrasto con quanto osservato in molte altre comunità multilingui, dove il parlato bilingue è frequentemente stigmatizzato (cfr. Gardner-Chloros, 2009: 14–16 per una panoramica), è coerente con la letteratura sulle comunità romaní, che documenta in questi contesti una diffusa accettazione dell’identità bilingue e una marcata tolleranza nei confronti della variazione e del mutamento linguistico indotti dal contatto (cfr., tra gli altri, Matras, 1998a; Rusakov, 2001; Elšík, 2007; Adamou, 2010, 2016; Padure & Adamou, 2024).

4.6 Discussione

In questo capitolo abbiamo proposto un’analisi dettagliata delle risposte al questionario sociolinguistico somministrato a un ampio campione di sinti lombardi della comunità di Pavia, con l’obiettivo di individuare alcune delle principali tendenze sociolinguistiche in atto nella comunità.

³ Mescolare il *sinto lombardo* e l’italiano piuttosto che parlare soltanto in *sinto* evitando forme italo-romanze.

L'analisi quantitativa dei dati è stata integrata, dove possibile, da informazioni di natura qualitativa, tratte dalle interviste e dalle osservazioni etnografiche svolte nel corso della ricerca sul campo.

Nel loro insieme, i risultati confermano che il repertorio linguistico dei sinti lombardi di Pavia si fonda principalmente sulla compresenza del sinto lombardo e dell'italiano. In questo quadro, in linea con quanto già osservato da Scala (2012), i dialetti italo-romanzi co-territoriali risultano marginali e non trasmessi alle nuove generazioni. Inoltre, l'esistenza di reti parentali molto dense con sinti appartenenti ad altri gruppi contribuisce ad ampliare tali repertori, favorendo la conoscenza di altre varietà di romaní o di varietà italo-romanze parlate da altri gruppi sinti.

Dal punto di vista funzionale, il sinto lombardo e l'italiano appaiono relativamente ben distinti: l'italiano costituisce il codice esclusivo degli usi esocomunitari, mentre il sinto lombardo rappresenta la lingua principale della comunicazione endocomunitaria. Al contempo, i dati mostrano come entrambe le lingue possano essere utilizzate simultaneamente in numerose situazioni endocomunitarie, suggerendo il ricorso frequente a pratiche discorsive bilingui. Questo quadro è coerente con precedenti caratterizzazioni del repertorio dei sinti lombardi come dilalico (Scala, 2012), nonché con l'osservazione che gli schemi di interazione dei sinti lombardi prevedono il ricorso a forme di multilinguismo basilettico (Halwachs, 1993).

Un ulteriore elemento di rilievo emerso dall'analisi riguarda la variazione generazionale negli usi linguistici. I dati hanno mostrato che il sinto lombardo viene utilizzato con minore frequenza nella comunicazione con i bambini piccoli, a vantaggio dell'italiano, rispetto che nella comunicazione con tutti gli altri gruppi di interlocutori nella comunità. In questo senso, il sinto lombardo potrebbe essere descritto, seguendo la tipologia proposta da Dal Negro & Iannàccaro (2003), come un possibile “codice in declino di parlanti”, mentre l'italiano si configurerebbe come possibile “codice in ascesa” (Dell'Aquila & Iannàccaro, 2006) nel repertorio della comunità. Tale quadro non implica tuttavia una perdita immediata della competenza in sinto da parte delle generazioni più giovani, che di norma continuano a utilizzare attivamente la lingua. In questo senso, sulla base della scala di *endangerment* proposta da Lewis & Simons (2010) e descritta al Capitolo 1, il sinto lombardo parlato a Pavia potrebbe essere identificato come “vigorous” (livello 6a), in quanto “the language is used orally by all generations and is being learned by children as their first language” (*ibid.*: 28); tuttavia, non è escluso che nell'arco di qualche generazione essa possa diventare “threatened” (livello 6b), etichetta utilizzata in riferimento a situazioni in cui “the language is used orally by all generations but only some of the child-bearing generation are transmitting it to their children” (*ibid.*). Si noti come, nella classificazione di Lewis & Simons (2010), il passaggio dal livello 6a al livello 6b marchi il passaggio cruciale da lingua “safe” a lingua “vulnerable”.

Infine, è emersa una narrazione metalinguistica condivisa che oppone il sinto parlato oggi a un sinto “puro”, associato al parlato di alcuni locutori più anziani e talvolta collocato in un passato percepito come più autentico. Questa rappresentazione appare in linea con il “discourse of loss” che Matras et al. (2024) hanno riportato in merito ad alcune varietà para-romaní in Europa; la presente ricerca mostra dunque che tali discorsi possono essere ricorrenti anche tra parlanti di varietà “flesse” di romaní. È tuttavia importante sottolineare che a tali rappresentazioni non corrisponde in alcun modo, presso i parlanti, un atteggiamento di allarme o di rigetto nei confronti del contatto linguistico: al contrario, il bilinguismo e il ricorso a pratiche discorsive miste sono generalmente valutati in modo positivo e considerati funzionali alla vita quotidiana, in linea con la letteratura sociolinguistica su altre comunità romaní.

A conclusione di questo capitolo, sulla base di quanto discusso, la Tabella 7 propone una schematizzazione del repertorio linguistico dei sinti lombardi di Pavia, ispirata al modello di Dal Negro & Iannàccaro (2003), introdotto al Capitolo 1.

Parametro	Lingua
Acroletto	Italiano
Basileto	Italiano, sinto lombardo
Codice esocomunitario	Italiano
Codice endocomunitario	Italiano, sinto lombardo
Codice in ascesa	Italiano
Codice in declino di parlanti	Sinto lombardo
Codice non autonomo	Sinto lombardo?

Tabella 7: Schematizzazione del repertorio linguistico comunitario dei sinti lombardi sulla base di alcuni parametri proposti in Dal Negro & Iannàccaro (2003).

In sintesi, nel repertorio linguistico dei sinti lombardi, il sinto si configura come basileto, codice endocomunitario e codice in declino di parlanti, mentre l'italiano si configura come acroletto, codice esocomunitario e codice in ascesa. Le funzioni di basileto e di codice endocomunitario appaiono pertanto spartite con l'italiano. Come si può osservare, è stato inserito un punto interrogativo accanto alla categoria di “codice non autonomo”. Con questa etichetta, Dal Negro & Iannàccaro (2003) indicano una lingua minoritaria che non è (più) utilizzata senza il costante apporto di elementi provenienti dalla lingua o dalle lingue maggioritarie in contatto. Nel caso del sinto lombardo, i dati macro-sociolinguistici qui analizzati non consentono di fornire una risposta definitiva a tale interrogativo. Come sottolineano gli stessi autori, la valutazione dell’“autonomia” di una lingua

minoritaria richiede un'analisi approfondita delle pratiche linguistiche effettive a livello microsociolinguistico. Questo sarà l'obiettivo del capitolo successivo, dedicato all'analisi del contatto tra italoromanzo e sinto lombardo nella conversazione spontanea

Capitolo 5. Schemi di parlato bilingue e convenzionalizzazione delle inserzioni italoromanze

Questo capitolo è dedicato all'analisi degli schemi di parlato bilingue che emergono nel discorso spontaneo dei sinti lombardi e del grado di convenzionalizzazione delle inserzioni di origine italoromanza. In particolare, il § 5.1 descrive la composizione linguistica del corpus a livello dei singoli *token* di parole, secondo il criterio di annotazione etimologica descritto nel dettaglio al Capitolo 3, mentre il § 5.2 propone una distinzione del materiale italoromanzo secondo un'interpretazione larga delle nozioni di “inserzione” e “alternanza” proposte da Muysken (2000). Le sezioni successive sono dedicate all'analisi delle inserzioni, che rappresentano la manifestazione più prominente del contatto linguistico nel corpus in analisi: in particolare, il § 5.3 ne illustra la distribuzione per parte del discorso, il § 5.4 presenta la distribuzione dei nomi e dei verbi italoromanzi per campo semantico, mentre il § 5.5 discute in dettaglio la regolarità delle inserzioni italoromanze, sulla base di criteri come la frequenza assoluta nel corpus, la diffusione tra i parlanti e la competizione con equivalenti nativi. Infine, il § 5.6 propone quindi una panoramica qualitativa dei fenomeni di calco, ossia casi in cui un modello lessicale italoromanzo non è riprodotto come tale, ma replicato (del tutto o in parte) attraverso materiale nativo. Infine, il § 5.7 propone una discussione dei principali risultati. La descrizione del parlato bilingue dei sinti lombardi proseguirà poi nel Capitolo 6, incentrato sulle strategie di integrazione a livello fonomorfológico dei nomi e dei verbi italoromanzi.

5.1 Composizione del corpus per lingua

Come anticipato, al fine di inquadrare l'incidenza delle lingue in contatto nel parlato spontaneo dei sinti lombardi e di consentire confronti interlinguistici, un primo approccio consiste nell'analisi della composizione del corpus per lingua a livello delle singole parole. Come indicato da Adamou (2016: 21), “in order to determine the rates of words depending on their origin, one has to count all the words which have previously been tagged as ‘words of the current-contact language(s)’ and calculate what percentage of the total number of words in the corpus they represent”.

Applicando questo metodo al corpus di sinto lombardo, notiamo che il 54.2% delle parole è romaní, mentre il 42.4% è di origine italoromanza; il restante 3.4% è costituito da forme annotate come “Altro” in quanto straniere o di ambigua classificazione; cfr. Tabella 8

Lingua	Token (N.)	Token (%)
Romaní	20.960	54.2%
Italoromanzo	16.396	42.4%
Altro	1.315	3.4%
Totale	38.671	100%

Tabella 8: Composizione del corpus per lingua.

Questi dati mostrano che, nonostante il sinto lombardo fosse stato stabilito come lingua di base dell'interazione nelle interviste, l'incidenza di parole italoromanze nel discorso risulta estremamente elevata, rappresentando quasi la metà dei *token* complessivi. In linea generale, questo dato appare coerente con la percezione dei parlanti, discussa nel capitolo precedente, di utilizzare un sinto fortemente “misto” e “gagizzato”. L'estratto in (10) offre un esempio piuttosto prototipico di parlato in sinto lombardo, caratterizzato dalla frequente combinazione di elementi romaní e italoromanzi.

- (10) *me labatardom anca but dar fabbrica dua karavas ciuculata*
 ‘io ho lavorato anche tanto in fabbrica, dove facevo cioccolata’
čomas sempre me čom sempre inketane cun u gadže perché me jomas ja giustraia
 ‘stavo sempre, io sono sempre stata insieme ai *gagi* perché io ero una giostraia’
quindi i fin da quando jomas ja tini čai che jomas maskar u gage xajudal?
 ‘quindi è fin da quando ero una bambina che ero in mezzo ai *gagi*, capito?’
is=ma mar giostri džajas in giro sempre cun mar- krol mar famiglia
 ‘avevo le mie giostre, andavamo in giro sempre con la mia, tutta la mia famiglia’
krol u nial / u vent vajas kaj ki nona mafalda / e čajas kaj/
 ‘tutta l'estate... l'inverno, venivamo qui dalla nonna Mafalda, e stavamo qui’
e u nial partajas, via cun i kampina
 ‘e l'estate partivamo, via con la roulotte’
girarajas krol u gava quand'is u festi piantarajas u giostri
 ‘giravamo tutte le città quando c'erano le feste, piantavamo le giostre’

Con le dovute cautele metodologiche, la composizione linguistica del corpus di sinto lombardo può essere anche contestualizzata in ottica comparativa, facendo riferimento ad altri studi relativi a lingue minoritarie in contatto con lingue dominanti. In questo senso, un termine di confronto molto utile è rappresentato dal lavoro di Adamou (2016), che confronta la composizione linguistica di corpora relativi a sette lingue minoritarie: due varietà di romaní (una parlata nella Tracia greca, in contatto

con il greco e il turco, e l'altra in Finlandia, in contatto con il finlandese); quattro varietà slave (la *nashta* parlata nel comune di Liti presso Salonicco, in contatto con il greco; il croato del Burgenland parlato nell'Austria orientale, in contatto con il tedesco; il sorabo superiore parlato nell'Alta Lusazia, in contatto con il tedesco; lo slavo molisano parlato nei comuni di Montemitro, Acquaviva Collecroce e San Felice del Molise, in provincia di Campobasso, in contatto con l'italiano); e l'iscateco parlato a Santa María Ixcatlán, nello stato messicano di Oaxaca, in contatto con lo spagnolo. A tali dati si aggiungono a fini comparativi i risultati riportati da Dal Negro (2009) sul walser parlato a Issime (in contatto piemontese, italiano, francese e francoprovenzale) e quelli relativi a un'altra varietà di *romani* parlata nella Serbia Orientale in contatto con il serbo discussi da Mirić & Ćirković (2022). In quest'ultimo caso, è tuttavia necessaria una precisazione: mentre gli studi precedentemente menzionati analizzano i corpora nella loro interezza, Mirić & Ćirković (2022) riportano soltanto i dati relativi alle inserzioni dal serbo, escludendo i casi di *code-switching* alternante; ciò comporta con ogni probabilità una sottostima delle parole serbe complessive nel corpus. I risultati del confronto con i corpora menzionati sono riportati nella Figura 21.

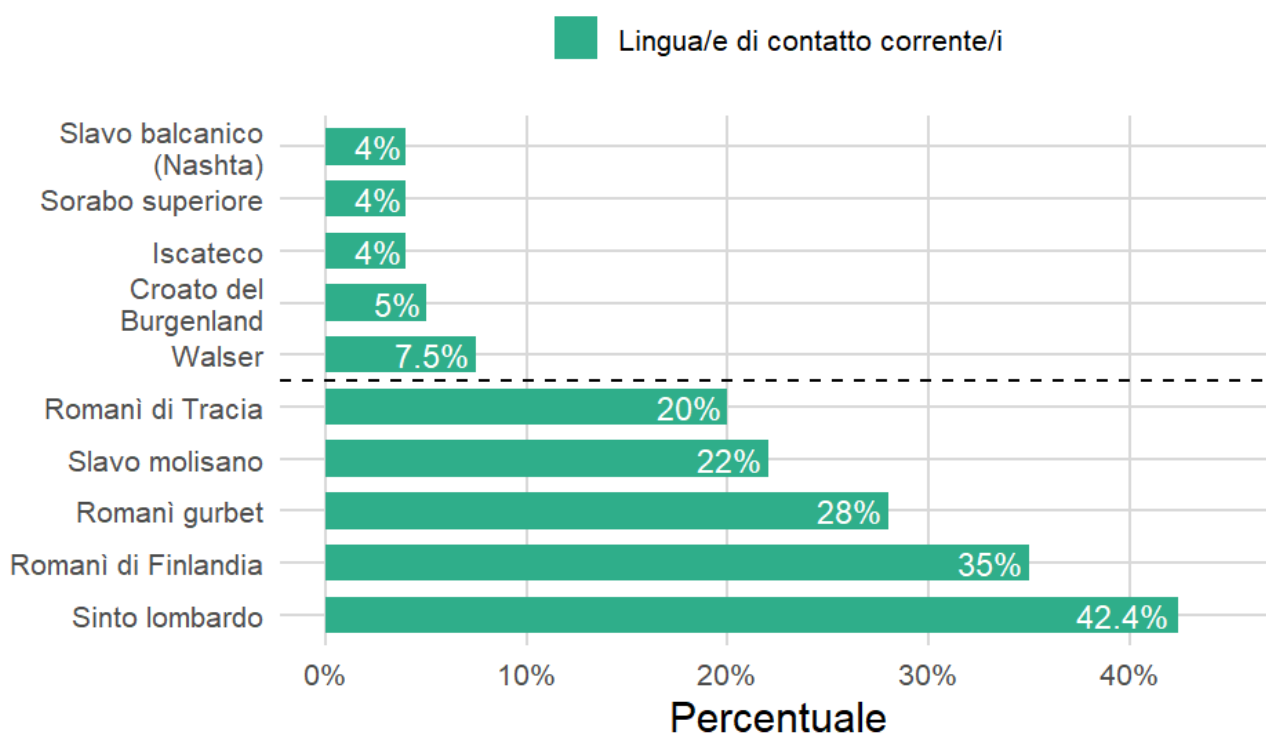


Figura 21: Confronto della composizione linguistica del corpus di sinto lombardo con corpora di altre lingue minoritarie (Dal Negro, 2009; Adamou, 2016; Mirić & Ćirković, 2022).

Come si può osservare, il sinto lombardo presenta l'incidenza più elevata di parole di contatto tra i corpora considerati in questa rassegna. Si potrebbe obiettare che ciò sia dovuto al fatto che, in sede di annotazione, sono state incluse tra le "parole di contatto" sia parole italiane sia parole provenienti

dal dialetto lombardo, la cui trasmissione si è interrotta nella comunità dei sinti lombardi. Tuttavia, anche gli altri corpora con cui è stato stabilito tale confronto riguardano lingue minoritarie in contatto plurisecolare con una o più lingue dominanti. Più in generale, la Figura 21 mostra che la distinzione più evidente è quella tra corpora con meno del 10% di parole provenienti dalla/e lingua/e di contatto e corpora con una percentuale superiore al 20%. Significativamente, tutti i corpora di romaní finora analizzati in letteratura rientrano nel secondo gruppo, sebbene, come detto, il sinto lombardo si distingue anche all'interno di questo gruppo per l'elevata incidenza delle parole di contatto.

In sintesi, questi risultati sembrano indicare per il sinto lombardo una situazione di contatto particolarmente intensa, anche in ottica comparativa. Le sezioni successive presentano un'analisi a grana più fine del parlato bilingue documentato dal corpus in analisi.

5.2 Schemi combinatori

I dati relativi alla composizione di un corpus per lingua consentono di ottenere una prima idea del parlato bi- o plurilingue di una data comunità, ma necessitano di essere integrati con ulteriori livelli di analisi. La presenza di una proporzione più o meno simile di parole da una lingua di contatto in diversi corpora può infatti nascondere preferenze per stili di parlato bilingue molto diversi sul piano formale e funzionale (Muysken, 2000; Adamou, 2016; Bullock et al., 2021).

Anche nel caso del corpus di sinto lombardo, l'etichetta “parole italoromanze” comprende una gamma molto ampia di fenomeni, evidenziati in grassetto negli esempi di seguito: casi di *code-switching* evidente verso l'italiano, che presuppongono un cambio (ancorché spesso momentaneo) nella lingua di base dell'interazione (cfr. 11); inserzioni di costituenti italiani ben formati in un contesto frasale prevalentemente romaní (cfr. 12); inserzioni di singole forme italiane apparentemente non integrate sul piano morfologico (cfr. 13); inserzioni di forme italoromanze pienamente integrate a livello fonomorfologico (cfr. 14); forme di chiara origine dialettale (cfr. 15). A un primo ascolto, le inserzioni in (14) e (15) possono non essere distinguibili dal materiale nativo.

- (11) *jam kola che pinena **deve passare il fiore dell'età***
 ‘siamo quelli che dicono ‘deve passare il fiore dell'età’

- (12) **il giorno di pasqua** *džas=lu peske /*
 ‘il giorno di Pasqua se n'è andato’

- (13) *jon nadal na karena kek **albero***
'loro a Natale non fanno nessun albero'
- (14) *u nial **partajas***
'l'estate partivamo'
- (15) *karaja **pena** u regali e bom*
'facciamo solo (appena) i regali e basta'

Per tracciare una prima distinzione quantitativa tra questi diversi tipi combinatori, si è scelto di seguire l'approccio proposto in Dal Negro (2006, 2009), che distingue tra:

- (1) casi di *code-switching* evidente;
- (2) inserzioni di singoli sintagmi o singole parole alloglotte all'interno di contesti discorsivi nella lingua minoritaria.

Seguendo questa metodologia, l'esempio in (11) rientra dunque nella prima categoria, mentre gli esempi (12)-(15) rientra nella seconda categoria. Più nello specifico, nella presente analisi, i casi di *code-switching* evidente verso l'italiano sono stati operazionalizzati in base alla presenza di almeno una frase minima contenente un verbo finito in italiano, come appunto nella frase in (11). I contesti discorsivi dominanti in sinto lombardo sono invece stati individuati sulla base di alcuni parametri strutturali, come il numero di morfemi romaní ma, soprattutto, la presenza di morfemi che compaiono tipicamente in romaní nel parlato della comunità, come la morfologia flessiva del verbo finito, gli articoli o i pronomi personali (cfr. § 5.3 per ulteriori dettagli). Tale scelta metodologica coincide con un'interpretazione molto larga delle nozioni di "inserzione" e "alternanza" proposte da Muysken (2000) e discusse nel dettaglio al Capitolo 1.

In alcuni casi la distinzione tra queste due categorie, ancorché definite in maniera larga, non risulta netta. Si consideri per esempio l'estratto in (16), nel quale l'inserzione del sintagma locativo italiano *a Roma* ha innescato un passaggio più esteso, di tipo alternante, verso l'italiano (*vedi che non è così*).

- (16) *se tu džae a Roma vedi che non è così*
'se tu vai a Roma, vedi che non è così'

In un caso come questo, il sintagma *a Roma* è stato considerato come un esempio di inserzione, mentre la frase a seguire come un caso di alternanza. Questa decisione è stata informata dalla conoscenza degli schemi di parlato bilingui tipici della comunità, in cui il ricorso a sintagmi

interamente italiani in corrispondenza di toponimi costituisce un *pattern* frequente, osservabile anche in contesti discorsivi altrimenti romaní (cfr. Capitolo 6).

Tenendo ben presenti queste scelte metodologiche, l'esclusione dei casi di alternanza evidente dal novero del materiale italoromanzo riduce la sua incidenza sui *token* complessivi del corpus dal 42.4% al 33.8%. Ciò indica dunque che le inserzioni italoromanze costituiscono il meccanismo di combinazione linguistica di gran lunga più ricorrente nel parlato bilingue dei sinti lombardi.

Uno sguardo più attento alle alternanze mostra come molto spesso queste siano associate a specifiche funzioni pragmatiche. Per esempio, in (17), il passaggio dal sinto all'italiano, peraltro in corrispondenza di una cesura prosodica e di un cambio di intonazione, è motivato dall'intento della parlante di chiedere conferma all'intervistatore (non parlante nativo di sinto) circa la corretta gestione dell'intervista.

- (17) *anca mur rom xal=la cun u panini u maro (.) vado bene avanti così?*
 ‘anche mio marito la mangia con i panini, il pane... Vado bene avanti così?’

In (18), invece, una parlante passa all'italiano per rispondere alla richiesta dell'intervistatore di descrivere il ‘capannone’ nel quale i sinti svolgono le proprie feste.

- (18) *u capanon sarebbe (.) ehm te lo dico in italiano*
 ‘il capannone sarebbe...’
 una tenda gigante bianca chiusa / dove ci siamo dentro tutti noi a bere ballare sentire
 la nostra musica / e andiamo avanti per giorni...

Tale scelta potrebbe essere interpretata sia come strategia per agevolare la comprensione dell'intervistatore, sia come conseguenza dell'indisponibilità di lessico romaní per effettuare la descrizione richiesta, circostanza che potrebbe rendere preferibile un passaggio completo all'italiano. Un aspetto particolarmente interessante in questo esempio è che il passaggio all'italiano è segnalato dalla parlante attraverso un commento metalinguistico (*te lo dico in italiano*), un meccanismo noto come *flagging* (Poplack, 1988; Rosignoli, 2011).

Infine, un'ultima funzione ricorrente delle alternanze è quella di riportare una citazione prodotta in un contesto di interazione italiano, sia dal parlante stesso sia da altri possibili locutori, come nell'esempio in (19).

- (19) *tipo doj fermardas=ma ja gagi e bučas proprio mandar, pindas ma chi è che è morto?*
 tipo lì mi ha fermato *una* gagi e ha chiesto proprio a me, ha detto: ma chi è che è morto

In sintesi, il *code-switching* alternante nel parlato bilingue dei sinti lombardi sembra rappresentare un fenomeno piuttosto saliente, in virtù non soltanto della sua minore frequenza rispetto alle inserzioni, ma anche in forza delle evidenti funzioni pragmatiche, connesse all'organizzazione dell'interazione o del discorso, a cui è associato. Nelle sezioni successive, l'analisi si concentra più in dettaglio sulle inserzioni, data la loro particolare prominenza nel parlato bilingue dei sinti lombardi.

5.3 Distribuzione delle inserzioni per parte del discorso

In questa sezione, analizziamo la distribuzione delle inserzioni italoromanze presenti all'interno di contesti frasali dominanti romaní per parte del discorso, sulla base delle categorie di annotazione descritte al Capitolo 3. L'obiettivo è collocare i dati del corpus all'interno del più ampio dibattito sulla maggiore o minore suscettibilità al contatto linguistico di specifiche classi di parole, un tema ampiamente discusso sia in studi sincronici *corpus-based* basati su corpora di parlato, sia in indagini tipologiche (cfr. Capitolo 1).

Come sottolineato, tra gli altri, da Muysken & van Hout (1994) e Poplack (2018), un'analisi limitata alla distribuzione interna del solo materiale alloglotto nel corpus di una lingua rischia di produrre risultati fuorvianti, in quanto la predominanza di una classe di parole sulle altre potrebbe semplicemente rifletterne la maggiore frequenza nel parlato in quella lingua. Per poter sostenere che una determinata classe di parole abbia effettivamente una particolare propensione al prestito, occorrerebbe adottare una soglia di riferimento esterna (Muysken & van Hout, 1994): questa può consistere nel confronto tra distribuzione delle classi di parole nel parlato monolingue e nel materiale alloglotto (cfr. Poplack et al., 1988), oppure nel calcolo della proporzione di parole di contatto rispetto alle parole native all'interno di ciascuna classe di parole (cfr. Adamou & Granqvist, 2015; Adamou, 2016; Mirić & Ćirković, 2022). Nel presente studio, si è scelto di seguire il secondo approccio, anche considerando che l'idea di un presunto “parlato monolingue” che funga da soglia di riferimento esterno non risulta applicabile nel caso del sinto lombardo.

La Figura 22 mostra la percentuale delle parole italoromanze rispetto alle parole romaní in ciascuna classe di parole all'interno di contesti discorsivi dominanti in sinto lombardo.

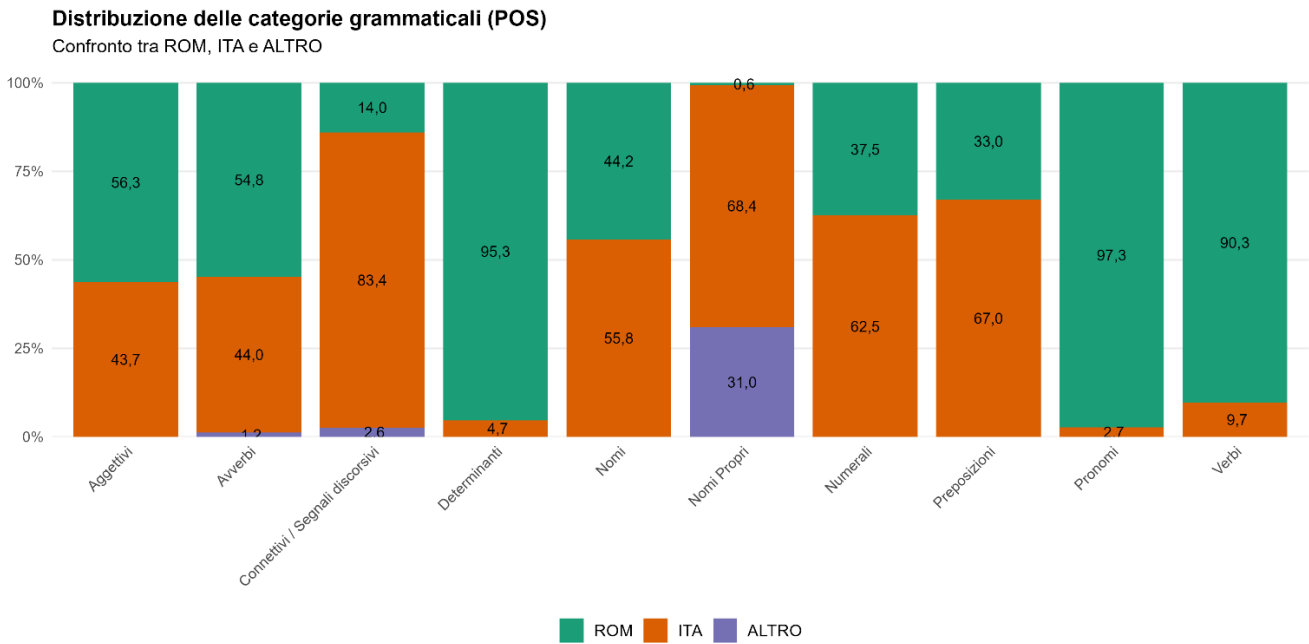


Figura 22: Distribuzione delle parole per lingua e parte del discorso nel corpus, in contesti dominanti romaní (percentuale).

Prima di commentare i dati, va fatta una precisazione: la Figura 22 riporta i valori relativi a *tutte* le parole italo-romanze presenti nel corpus, inclusi i sintagmi (nominali e preposizionali) interamente italo-romanzi presenti in contesti frasali romaní, mentre sono stati esclusi segmenti italiani coincidenti almeno con una frase minima. Ciò potrebbe influire, almeno in parte, su una maggiore incidenza di determinanti e preposizioni nei dati.

Come si può osservare dalla Figura 22, gli elementi italo-romanzi nel parlato in sinto lombardo possono appartenere a tutte le classi di parole; tuttavia, la loro distribuzione risulta fortemente sbilanciata. Alcune classi di parole presentano infatti una netta predominanza di materiale italo-romanzo: è il caso dei connettivi e segnali discorsivi (83.4%), delle preposizioni (67%) e dei numerali (62.5%). Altre classi mostrano una distribuzione più equilibrata tra materiale italo-romanzo e romaní: si tratta dei nomi (55.8% di parole italo-romanze), degli aggettivi (44% di parole italo-romanze) e degli avverbi (44% di parole italo-romanze). Infine, alcune classi di parole risultano costituite in larga maggioranza da materiale romaní: si tratta dei verbi (solo 9.1% di parole italo-romanze), dei determinanti (solo 4.7% di parole italo-romanze) e dei pronomi (solo 2.7% di parole italo-romanze). In altre parole, un discorso tipico in sinto lombardo prevede l'utilizzo maggioritario di connettivi/segnali discorsivi, preposizioni e numerali italo-romanzi; l'utilizzo in misura più o meno eguale di nomi, aggettivi e avverbi italo-romanzi e romaní; e l'utilizzo maggioritario di verbi, pronomi e determinanti romaní.

Alcuni di questi risultati appaiono in linea con le tendenze interlinguistiche più generali riportate in situazioni di contatto. Per esempio, la forte presenza di connettivi e di segnali discorsivi italoromanzi è coerente con la gerarchia di prestabilità individuata da Matras (2007), che colloca questi elementi al primo posto. Tale tendenza, connessa alla salienza di questi elementi nella strutturazione del discorso (Matras, 1998a), appare particolarmente marcata in contesti di contatto asimmetrico come quello in analisi, in cui sono coinvolte una lingua maggioritaria e una lingua minoritaria a tradizione orale, quest'ultima limitata ai domini di interazione più informali nei quali l'attivazione di una modalità bilingue è generalmente accettata (Matras, 2020: 209). Non a caso, tendenze analoghe sono documentate in altre varietà di romaní: Ćirković & Mirić (2022a) riportano infatti che nella romaní di Knjaževac le congiunzioni e le particelle pragmatiche di origine serba costituiscono oltre il 78% dei *token* della categoria, mentre Adamou & Granqvist (2014) mostrano che le congiunzioni finlandesi nel parlato in romaní finlandese ammontano al 72% dei *token*. I connettivi costituiscono, in entrambi i casi, la categoria lessicale maggiormente rappresentata da materiale non romaní.

Anche la significativa presenza di nomi italoromanzi nel corpus di sinto lombardo è in linea con le tendenze interlinguistiche più generali, spesso motivate dalla funzione referenziale dei nomi e dal loro impiego per denominare concetti nuovi o culturalmente salienti (cfr. p.es. Myers-Scotton, 1993; Backus, 1996). Nella sezione successiva, un'analisi più dettagliata dei nomi per campi semantici permetterà di verificare più nel dettaglio questa ipotesi. Anche la scarsa propensione al prestito dei verbi, determinanti e soprattutto pronomi è in linea con le principali gerarchie del prestito e gli studi sul parlato bilingue (cfr. p.es. Adamou & Granqvist, 2014; Matras, 2020; Dal Negro, 2009).

Un dato interessante nel corpus in analisi è la particolare propensione al prestito dei numerali e delle preposizioni. In contesti di lingue minoritarie, il prestito dei numerali non appare infrequente, soprattutto per quanto riguarda i numerali alti, che tendono a essere utilizzati unicamente in contesti esterni alla comunità, come le transazioni economiche (Matras, 2020: 173). Come già accennato al Capitolo 4 e come vedremo meglio successivamente, questo è anche il caso del sinto lombardo, in cui i numerali nativi si mantengono fino al sei, mentre dal sette in poi la numerazione prosegue attraverso prestiti italoromanzi. Comparativamente più rara appare forse l'inserzione massiccia di preposizioni, le quali, come vedremo meglio in seguito, non si limitano a comparire all'interno di sintagmi interamente italoromanzi, ma spesso anche all'interno di sintagmi misti dove selezionano nomi romaní. Un ricorso così elevato a preposizioni dalle lingue di contatto non è riportato negli altri studi relativi alla romaní e, in ottica tipologica, è stato associato a fasi di contatto particolarmente intense (Thomason & Kaufman, 1988). Si può ritenere che la situazione del sinto lombardo sia da inquadrare alla luce di un cambiamento sintattico più ampio, ovvero la progressiva erosione del

sistema dei casi per la flessione di nomi e la riassegnazione delle relative funzioni sintattiche a preposizioni di provenienza italo-romanza (cfr. già Soravia, 1977); cfr. anche § 5.5.

In ultimo, è opportuno ribadire che l'analisi presentata in questa sezione si basa esclusivamente su singoli *token* di parole e non è dunque sensibile alla diversa frequenza di specifici tipi lessicali. Un'analisi in termini di lemmi restituirebbe un risultato piuttosto diverso: a questo proposito, al § 5.5 verrà proposta un'analisi in termini di regolarità di specifici lemmi per parte del discorso.

5.4 Distribuzione dei nomi e dei verbi italo-romanzi per campi semantici

La letteratura sul contatto linguistico concorda nel riconoscere che le parole contenute, e in particolare i nomi, presentano una maggiore suscettibilità al prestito rispetto alle altre classi di parole. Questa tendenza è stata confermata dai dati di sinto lombardo, in cui le forme di origine italo-romanza costituiscono più del 50% dei nomi, sebbene nel contesto in analisi altre classi di parole mostrino tendenze al prestito ancora maggiori. Come anticipato più volte, la particolare propensione al prestito dei nomi è stata spesso ricondotta al fatto che i nomi consentono di riferirsi a concetti culturalmente specifici, talvolta assenti nella lingua ricevente o comunque più efficacemente espressi attraverso il ricorso alla lingua originaria (cfr. Matras, 2020; Backus, 1996, 2001; Myers-Scotton, 1993).

In questa sezione, ci occupiamo di analizzare la distribuzione per campi semantici dei nomi e dei verbi di origine italo-romanza nel parlato bilingue dei sinti lombardi, in contesti dominanti romani. Per questo studio, a tutti i nomi e verbi italo-romanzi sono state assegnate le 24 categorie semantiche utilizzate nel già citato WOLD (Haspelmath & Tadmor, 2009; cfr. Capitolo 1), pensato per stabilire gerarchie di prestito a livello interlinguistico in ottica diacronica. Nel nostro caso, l'approccio è invece sincronico e si limita alle forme di origine italo-romanza, senza estendersi a fasi di contatto precedenti (cfr. Adamou, 2016 per un approccio analogo su dati di parlato bilingue).

Va sottolineato che la classificazione semantica adoperata comporta un certo grado di soggettività. Operativamente, laddove lemmi equivalenti a quelli incontrati nel corpus di sinto lombardo fossero inclusi nel WOLD, è stata riutilizzata la stessa etichetta impiegata nel *database*. Infine, nomi e verbi di ambigua classificazione sono stati inclusi nella categoria “Miscellaneous”.

Fatte queste premesse, la Tabella 9 mostra la ripartizione dei nomi e dei verbi italo-romanzi per campo semantico (considerati sia a livello di singoli *token*, sia a livello di lemmi), insieme ad alcuni esempi illustrativi. La lista dei campi semantici segue un ordine di frequenza in termini di *token*. Accanto alle forme dialettali o foneticamente integrate in sinto lombardo è riportata la traduzione in italiano.

Campo semantico	N. token	N. lemmi	Esempi
Social and political relations	427	107	<i>Aiuto, amicizia, festa, capo, comune</i>
Time	407	49	<i>Anno, volta, dicembre, domenica, pomeriggio</i>
Cognition	331	96	<i>Assenza, bisogno, comodità, consiglio, decisione</i>
Kinship	317	36	<i>Antenato, nono ‘nonno’, bisnono ‘bisnonno’, genitore, nivudo ‘nipote’</i>
The modern world	307	78	<i>Albergo, fabbrica, cellulare, televisium ‘televisione’, discoteca</i>
Food and drink	244	88	<i>Aglio, alcol, bira ‘birra’, cacao, carota</i>
Emotions and values	178	74	<i>Agonia, allegria, amore, divertimento, libertà</i>
Speech and language	161	41	<i>Accento, canto, colloquio, discorso, parola</i>
Motion	151	39	<i>Azione, ballo, ginnastica, missione, viaggio</i>
Religion and belief	123	38	<i>Battesimo, bibbia, cimitero, conversione, culto</i>
Basic actions and technology	103	25	<i>Alluminio, baraca ‘baracca’, cesto, corda, oro</i>
Spatial relations	102	31	<i>Campo, piasa ‘piazza’, teren ‘terreno’, posto, via</i>
The body	80	34	<i>Cervello, coma, malattia, sesso, nausea</i>
Quantity	67	13	<i>Conto, numero, parte, quintal ‘quintale’, pesare</i>
The house	49	30	<i>Bagno, bicier(o) ‘bicchiere’, cana ‘rubinetto’, doccia, frigo</i>
Agriculture and vegetation	24	11	<i>Albero, cereale, fiore, pianta, ramo</i>
The physical world	23	13	<i>Ambiente, deserto, goccia, luce, luna</i>
Possession	19	13	<i>Lusso, pension ‘pensione’, prodotto, richesa ‘ricchezza’, guadagnare</i>

Clothing and grooming	15	8	<i>Borsa, maia ‘maglia’, mascherina, perla, stivale</i>
Law	13	9	<i>Atto, avvocato, crimine, formulario, galera</i>
Sense perception	12	9	<i>Rumur ‘rumore’, savur ‘sapore’, suono, vista, assaggiare</i>
Animals	8	3	<i>Gallina, lupo, zanzara</i>
Warfare and hunting	4	4	<i>Battaglia, bomba, guerra, difendere</i>
Miscellaneous	133	55	<i>Gagizzare, facsimile, prassi</i>
Totale	3298	910	

Tabella 9: Distribuzione dei nomi e dei verbi italo-romanzi sulla base delle categorie del WOLD (Haspelmath & Tadmor, 2009).

Contrariamente a quanto ci si potrebbe aspettare, la Tabella 9 mostra che i campi semantici più popolati da materiale italo-romanzo non sono necessariamente “culturalmente specifici”. Per esempio, come si può anche evincere dagli esempi forniti nella Tabella 9, i due campi semantici più frequenti, “Social and political relations” e “Time”, non evocano concetti strettamente associati alla cultura maggioritaria. Va però notato che la particolare rappresentazione del campo semantico “Time” è anche legata all’elevata frequenza dei lemmi ‘anno’ e ‘volta’ (cfr. anche § 5.5), che insieme costituiscono una porzione molto ampia dei *token* riconducibili a questo dominio. Un’analisi in termini di lemmi restituisce infatti un risultato in parte diverso, riducendo il numero di forme uniche appartenenti a questo campo semantico a 49. Una situazione simile si ravvisa per il campo semantico “Kinship”, che consta di un numero limitato di forme piuttosto regolari. In questo quadro, il campo semantico “The modern world”, che tende a includere molti referenti associati al dominio della tecnologia moderna e del mondo istituzionale o scolastico, connessi dunque alla cultura maggioritaria, si colloca soltanto al quinto posto nei ranghi di frequenza per *token* (e al quarto per lemmi), preceduto da “Cognition” e “Kinship”, altri due campi semantici che vengono tipicamente considerati resistenti al contatto (Tadmor, 2009).

È anche interessante analizzare la distribuzione per campi semantici dell’italiano e del dialetto lombardo. A questo proposito, possiamo per esempio notare che il campo semantico “Kinship” è dominato da elementi di chiara origine dialettale (p.es. *mama* ‘mamma’, *nono* ‘nonno’, *cugnado* ‘cognato’, *nevudo* ‘nipote’), testimoniando che non si tratta di acquisizioni recenti nel lessico. Anche i campi semantici “Sense perception”, “Spatial relations” e “The house” sembrano ospitare molte parole dialettali. Viceversa, nel campo semantico “The modern world” domina come prevedibile

l'italiano. Nondimeno, un'analisi più attenta mostra che gli elementi dialettali hanno una distribuzione piuttosto ampia attraverso i diversi campi semantici, non limitandosi ai concetti più tradizionali. Per esempio, anche nel campo semantico "The modern world" sono attestate forme dall'aspetto dialettale come *televisiun* 'televisione' o *registrasiun* 'registrazione'. In questi casi, tuttavia, sembra quasi di avere a che fare con forme italiane che subiscono una sorta di trafila di "dialettizzazione" per risultare "più sinte" quando inserite in discorsi in questa lingua, che prevede la soppressione della vocale finale e talvolta il passaggio da *o* a *u*. Si noti che un simile meccanismo è stato discusso da Dal Negro (2009, 2011) in relazione alle inserzioni italiane nel parlato in walser.

Un'analisi dettagliata della composizione di ogni campo semantico per lingua, contrapponendo italoromanzo e romaní, risulta difficile da effettuare in quanto i dati di parlato che costituiscono l'oggetto di questo volume sono relativi ad argomenti abbastanza ricorrenti, che favoriscono il ricorso a particolari campi semantici e la scarsa attestazione di altri. A titolo di esempio, possiamo però discutere il caso del campo semantico "Kinship", particolarmente frequente nel discorso, dal momento che una domanda delle interviste includeva la descrizione della famiglia. In questo caso, la distribuzione dei lemmi tra elementi italoromanzi e romaní risulta emblematica. I lemmi che indicano relazioni familiari più strette tendono a rimanere nativi: *daj* 'madre', *dat* 'padre', *pral* 'fratello', *pen* 'sorella', *čao/i* 'figlio/a'. Viceversa, lemmi di origine italoromanza indicano gradi di parentela meno prossimi al soggetto (si vedano i già discussi *zia*, *zia*, *nono*, *nonu*, *cugnado*, *nivudu*, *suocera*, ecc.). La forma *zia* si trova in variazione con l'equivalente nativo *bibi*, sul quale prevale in termini di frequenza e diffusione, mentre al maschile le forme *zio* o *barba* (quest'ultima, più rara, originaria del dialetto piemontese) non risultano in variazione con lemmi più antichi. Anche *nono* 'nonno' e *nona* 'nonna' non risultano in variazione con altri lemmi nel corpus, sebbene la parola romaní *papu* 'nonno', prestito di origine greco-bizantina, sia nota ai parlanti (cfr. anche § 5.5). Complessivamente, questa distribuzione è coerente con le tendenze generali rilevate in altri contesti di contatto linguistico (Matras, 2020), in cui i termini che indicano le relazioni familiari più strette tendono a resistere maggiormente al prestito.

Infine, l'analisi per campi semantici consente di osservare che alcuni lemmi vengono adottati nella lingua minoritaria soltanto in accezioni specifiche (sul ruolo della specificità semantica nel facilitare i processi di prestito, cfr. in particolare Backus, 1996, 2001). Per esempio, nel parlato in sinto lombardo il nome italiano *fuoco* compare esclusivamente per indicare il 'fuoco d'artificio', mentre la parola romaní *jak* è stabilmente utilizzata per indicare il 'fuoco' inteso come entità fisica. Similmente, *ora* ricorre soltanto nell'accezione di 'ora scolastica' (quasi sempre all'interno di sequenze italiane più ampie come *quinta ora* o *sesta ora*), mentre, per indicare l'ora della giornata, il termine *stunda* (un prestito dal tedesco, della cui origine i parlanti non hanno consapevolezza) prevale. Infine, la

parola *campo* appare nella sua accezione di ‘campo nomadi’, mentre la forma di origine tedesca *felda* è utilizzata per riferirsi al campo agricolo.

5.5 Regolarità

Un’analisi delle “gerarchie del prestito” consente di individuare tendenze generali e facilmente confrontabili a livello interlinguistico, ma non fornisce indicazioni sufficienti sulla stabilità di specifiche forme a livello della comunità bilingue e dei singoli individui. La presente sezione intende dunque approfondire la collocazione delle inserzioni italoromanze lungo il continuum prestito-*code-switching* nel parlato bilingue dei sinti lombardi, un tema discusso da un punto di vista teorico al Capitolo 1.

La regolarità è stata considerata uno dei criteri centrali per distinguere tra prestito e *code-switching* nel parlato bilingue, soprattutto negli approcci di stampo *usage-based*, che considerano la regolarità di una forma come un indicatore del suo grado di *entrenchment*. In un corpus bilingue, la nozione di regolarità può essere operazionalizzata sulla base di due criteri principali: la frequenza e la diffusione (Poplack et al., 1988). Per frequenza intendiamo il numero assoluto di occorrenze di un lemma in un corpus; per diffusione intendiamo il numero di parlanti che lo impiegano. A questi parametri si aggiunge la nozione di “frequenza relativa” (Myers-Scotton, 1993), ossia la frequenza di una forma alloglotta rispetto al suo eventuale equivalente nativo. In sintesi, il criterio della regolarità per distinguere tra prestito e *code-switching* può essere visto come risultante dall’interazione di tre criteri, schematizzati nella Tabella 10.

Prestito prototipico	Code-switching prototipico
-Equivalente nativo	+Equivalente nativo
+Frequenza	-Frequenza
+Diffusione	-Diffusione

Tabella 10: Rappresentazione del continuum prestito-*code-switching* sulla base del criterio della regolarità.

Questa schematizzazione, tuttavia, non è esente da limiti. Come osserva Adamou (2016), il criterio della regolarità può risultare particolarmente informativo per le parole funzione, la cui probabilità di occorrenza nel discorso non dipende dall’argomento di conversazione. Viceversa, per le parole contenuto, la frequenza è inevitabilmente influenzata dalla ricorrenza tematica: un lemma può risultare altamente frequente in un corpus semplicemente perché l’argomento che lo richiede è stato discusso ripetutamente, ma può risultare infrequente (o persino mai attestato) perché l’argomento che

lo attiva è stato poco o per nulla discusso. D'altra parte, anche nel parlato monolingue diversi elementi lessicali possono mostrare una frequenza di occorrenza molto ridotta.

Un criterio alternativo per misurare la regolarità di una forma di una lingua B in un discorso in una lingua A è rappresentato dalla nozione *listedness* (Deuchar et al., 2007, sulla base di Muysken, 2000), con cui si intende l'inclusione di una forma alloglotta nei dizionari di una lingua ricevente. Tale criterio, tuttavia, presenta diversi limiti. Da un lato, diversi autori hanno sottolineato come non vi sia necessariamente una corrispondenza tra la presenza di una forma in un dizionario e il suo effettivo *entrenchment* nel lessico mentale dei parlanti (Poplack, 2018; Treffers-Daller, 2023); dall'altro, in molti contesti di lingue minoritarie la disponibilità di dizionari aggiornati è limitata o assente. Per quanto riguarda le varietà di romaní in Italia, è disponibile il dizionario di Soravia & Fochi (1995), descritto al Capitolo 1. Tuttavia, oltre alla sua datazione ormai non recente, una sua consultazione mostra come le voci di origine italoromanza tendano a essere sistematicamente escluse a favore del lessico nativo, rendendo questo strumento poco adatto. Nel corso della presente analisi, il dizionario di Soravia & Fochi (1995) sarà utilizzato come base di confronto.

Per chiarezza di trattazione, nell'analisi a seguire distinguiamo tra parole funzione e parole contenuto. Nel primo gruppo rientrano le categorie dei connettivi/segnali discorsivi, preposizioni, pronomi, determinanti e numerali, mentre nel secondo gruppo rientrano le categorie dei nomi e dei verbi. L'esame della regolarità costituisce il primo passo nell'analisi del continuum prestito-*code-switching*; esso sarà integrato nel Capitolo 6 dall'analisi delle strategie di integrazione morfologica, criterio considerato centrale in alcuni approcci (Poplack, 2018) per la distinzione tra i due fenomeni.

5.5.1 Parole funzione

In linea con quanto osservato in letteratura (cfr. Adamou, 2016), il criterio della regolarità si rivela piuttosto efficace se applicato all'analisi delle parole funzione di origine italoromanza nel corpus, in quanto consente di individuare alcune forme altamente regolari non solo a livello dell'individuo, ma anche della comunità, e dunque interpretabili come prestiti acclimatati nel sinto lombardo.

La Tabella 11 di seguito offre una panoramica delle parole funzione italoromanze più regolari nel corpus di sinto lombardo. Si riporta inoltre l'eventuale presenza di un equivalente 'nativo', cioè preromano, sia nel corpus sia nel dizionario di Soravia & Fochi (1995)¹. Ovviamente, l'assenza di forme native indica soltanto l'assenza di queste forme nel dizionario, non necessariamente in stadi precedenti della lingua.

¹ Si riportano soltanto le forme che, nel dizionario, sono esplicitamente indicate come appartenenti al sinto lombardo.

Categoria	Forme italoromanze nel corpus	Equivalente ‘nativo’ nel corpus	Equivalente ‘nativo’ in Soravia & Fochi (1995)
Connettivi	<i>E</i> <i>O</i> <i>Ma</i> <i>Invece</i> <i>Perché / parché</i> <i>Però / parò</i> <i>Allora</i> <i>Insomma</i> <i>Che</i> <i>Poi</i> <i>Quindi</i> <i>Se</i> <i>Quando</i> <i>Cioè</i>	<i>/</i> <i>/</i> <i>/</i> <i>/</i> <i>Oske</i> <i>/</i> <i>/</i> <i>/</i> <i>/</i> <i>/</i> <i>/</i> <i>/</i> <i>/</i> <i>/</i>	<i>Ta</i> <i>/</i> <i>/</i> <i>/</i> <i>Oske</i> <i>/</i> <i>/</i> <i>/</i> <i>Os (pron. rel.), ta (cong.)</i> <i>/</i> <i>/</i> <i>Ta</i> <i>Kana</i> <i>/</i>
Preposizioni	<i>Da ‘di, da’</i> <i>Cun</i> <i>Par</i> <i>In</i> <i>A</i> <i>Fra</i>	<i>/</i> <i>/</i> <i>/</i> <i>Dar, indar, dendar, andar</i> <i>Ke</i> <i>Maskar (loc.), katar (temp.)</i>	<i>/</i> <i>/</i> <i>/</i> <i>/</i> <i>/</i> <i>Maskar</i>
Focalizzatori	<i>Anca ‘anche’</i> <i>Nianca ‘neanche’</i> <i>Pena ‘solo’</i>	<i>/</i> <i>/</i> <i>/</i>	<i>Nina, pandar</i> <i>/</i> <i>/</i>
Avverbi fasali	<i>Bela ‘già’</i>	<i>/</i>	<i>/</i>
Intensificatore negativo	<i>Miga</i>	<i>/</i>	<i>/</i>
Modalizzatori epistemici	<i>Forse</i> <i>Magari</i>	<i>/</i> <i>/</i>	<i>/</i> <i>/</i>
Avverbi temporali	<i>Prima</i> <i>Dopo</i> <i>Mai</i> <i>Sempre</i>	<i>/</i> <i>/</i> <i>/</i> <i>Xa</i>	<i>/</i> <i>Pale</i> <i>/</i> <i>/</i>
Indefiniti	<i>Qualche</i>	<i>/</i>	<i>/</i>

	<i>Stesso</i>	/	/
	<i>Ogni</i>	/	/
	<i>Ognidun</i> ‘ognuno’	/	/
Particella comparativa	<i>Più</i>	/	/

Tabella 11: Parole funzione italoromanze nel corpus (lemmi).

In generale, la maggior parte delle parole funzione italoromanze elencate nella Tabella 11 non appare in variazione con equivalenti nativi nel corpus, né si registrano frequentemente equivalenti nativi nel dizionario di Soravia & Fochi (1995).

Più nello specifico, è possibile osservare che il sinto lombardo ha sostanzialmente preso a prestito dall’italoromanzo l’intero sistema di connettivi, la maggior parte dei quali è utilizzata stabilmente senza essere posta in variazione con equivalenti nativi nella conversazione spontanea (questo è il caso di *e*, *o*, *ma*, *però*, *allora*, *insomma*, *invece*, *che*, *poi*, *quindi*, *quando*, *se*). Di questi, alcuni equivalenti nativi sono riportati nel dizionario di Soravia & Fochi (1995): *os* (nel significato di ‘che’ come pronomi relativo), *ta* (nel significato di ‘se’) e *kana* (nel significato di ‘quando’).

Come si osserva nella Tabella 11, solo *perché* mostra nel corpus spontaneo tracce di competizione con una forma nativa, *oske*. Quest’ultima, significativamente, è prodotta una sola volta da un solo parlante, mentre la forma romanza *perché/parché* occorre 382 volte per 24 parlanti, un dato che indica chiaramente la sua stabilizzazione completa. Si segnalano inoltre alcune forme miste, quali *eta*, esito della fusione tra la congiunzione italoromanza *e* e la congiunzione indoaria *ta* (mai attestata come tale nel corpus, ma indicata nel dizionario di Soravia & Fochi, 1995) e *invece ta* ‘invece di’, in cui la forma romanza *invece* si combina al complementatore modale di origine indoaria *ta*. Di fatto, l’unico connettivo romaní attestato nel corpus è *ta*, di origine indoaria, utilizzato in combinazione con verbi finiti (coniugati al modo congiuntivo) per introdurre complementive non fattuali o modali: tale funzione è assente in italiano, per cui sarebbe utilizzato un infinito verbale.

Diversi studi relativi a lingue minoritarie, molti dei quali condotti in contesto italiano (cfr. Dal Negro, 2005b; Stolz, 2007; Fiorentini, 2017), hanno osservato che i connettivi adottati da una lingua dominante a una lingua minoritaria tendono (almeno inizialmente) a non coprire l’intero spettro funzionale attestato per la lingua dominante. In particolare, è stato rilevato che tali elementi tendono a essere mutuati in primo luogo nella loro funzione di segnali discorsivi, connessi all’organizzazione dell’interazione, e solo in un secondo tempo (eventualmente) nella funzione di connettivi semantici, connessi all’organizzazione del testo. I dati del corpus mostrano che questo schema non si applica al sinto lombardo, in cui i connettivi di origine italoromanza appaiono impiegati lungo l’intero spettro delle funzioni associate alla lingua di partenza. Ciò suggerisce che il processo di integrazione di questi elementi in sinto lombardo è oggi molto avanzato e non limitato alla dimensione interazionale. Questa

polifunzionalità è innanzitutto osservabile nel caso di *che*, il connettivo più frequente in assoluto nel corpus, con 771 occorrenze da parte di tutti i locutori, attestato nel parlato in sinto lombardo in tutte le principali funzioni documentate per l'italiano: come complementatore (20), come subordinatore generico con sfumatura causale (21), come pronome relativo (22), come pronome esclamativo (23) e come pronome interrogativo (24).

(20) *jon pinden mange **che** džēn ku funeral*
 'loro mi hanno detto che sono andati al funerale'

(21) *rakar in sinto **che** jo kamela impararel=lo*
 'parla in sinto che lui vuole impararlo'

(22) *i jek **che** seguarela u stildi*
 'è uno che segue i carerati'

(23) *dik **che** sukar*
 'guarda che bello'

(24) ***che** stunda i?*
 'che ore sono?'

Una situazione analoga si osserva per *ma* e *allora*. Negli esempi (25) e (26), *allora* è rispettivamente utilizzato come segnale discorsivo di presa di turno e come connettivo con funzione consecutiva. Analogamente, negli esempi (27) e (28), *ma* è rispettivamente utilizzato come connettivo con funzione avversativa e come segnale discorsivo di presa di turno.

(25) ***allora** / prima is più / pinas che lenas rom prima u čave*
 'allora, prima era più... Diciamo che si sposavano prima i ragazzi'

(26) *is kedžano che sardelas mēge pren i bari giostra **allora** mur dat pindas u karaja?*
 non c'era nessuno che ci portava avanti la giostra grande, allora mio padre ha detto: che facciamo?

- (27) *u sintu xoel=lu krol* **ma** *na rakarel=lu miga*
 ‘il sinto lo capisce tutto ma non lo parla’
- (28) **ma** *dik / men* essendo cristiani evangelici *na festegiaraj=lu miga*
 ‘ma guarda, noi essendo cristiani evangelici non lo festeggiamo’

Quanto alle preposizioni, si osserva che *par* ‘per’, *cun* ‘con’ e *da* ‘di, da’, altamente regolari nel corpus, non risultano mai in variazione con equivalenti nativi. Queste forme hanno una chiara origine dialettale, il che testimonia che non si tratta di acquisizioni recenti nel sinto lombardo; a riprova di ciò, notiamo che *cun* e *par* sono anche registrate come voci sinte lombarde nel dizionario di Soravia & Fochi (1995).

Come anticipato al § 5.3, una presenza così massiccia di preposizioni può essere messa in relazione con l’erosione del sistema dei casi nella flessione nominale del sinto lombardo. Queste preposizioni possono anche selezionare pronomi personali che, a differenza dei nomi, hanno mantenuto il sistema casuale ereditato, producendo una marcatura, per così dire, “ridondante”. Si consideri l’esempio di seguito (29), in cui il ruolo semantico del pronome personale di prima persona *me* (*man* alla forma obliqua) è marcato sia dalla preposizione italoromanza *par*, sia dal caso locativo *-de*.

- (29) *no is=li kamli venas narvali* **par mande**
 ‘no, erano bravi, diventavano matti per me’

Le preposizioni italoromanze *a* e *in* mostrano un’elevata frequenza e presentano complesse dinamiche di variazione con equivalenti nativi, che sembrano dipendere sia dalla semantica specifica con cui vengono impiegate (temporale vs. spaziale), sia dalla natura del nome che selezionano (italoromanzo vs. romaní). Questo aspetto sarà ulteriormente ripreso al Capitolo 6.

Passando alle altre categorie, si può notare che il sinto lombardo ha mutuato dall’italoromanzo i focalizzatori *anca* ‘anche’, *nianca* ‘neanche’ e *proprio* ‘proprio’, nonché l’avverbio fasale *bela* ‘già’. Quest’ultimo deriva verosimilmente dalla costruzione lombarda *bèl e* ‘bello e’ all’interno di espressioni del tipo *l’è bèl e andré a mangià* ‘sta già mangiando’, lett. ‘è bello e andato a mangiare’ (Galli, 1996: 37). Anche in questi casi, l’origine dialettale di molte forme indica una loro penetrazione non recente nel lessico del sinto lombardo. Si nota inoltre che nel dizionario di Soravia & Fochi (1995) sono indicate le forme native *nina* e *pandar* per ‘anche’ per il sinto lombardo, ma queste non occorrono mai nel corpus in analisi.

Forme regolari e generalmente prive di equivalenti nativi sono inoltre gli avverbi temporali *prima*, *dopo* (usato anche come preposizione), *sempre* e *mai*, i modalizzatori epistemici *magari* e *forse*, la particella comparativa *più* e l'intensificatore negativo *miga*. Nel corpus, *sempre* risulta in variazione con la forma nativa *xa*, riportata anche in Soravia & Fochi (1995) come caratteristica del sinto piemontese. Tuttavia, mentre *sempre* mostra 111 occorrenze nel corpus, *xa* registra soltanto due occorrenze, confermando la stabilità della forma *sempre* in sinto lombardo e la sua interpretazione come prestito. Per le altre forme non si registrano equivalenti nativi nel dizionario di Soravia & Fochi (1995), tranne per l'avverbio *dopo*, in cui è indicato il lemma *pale*, non attestato nel corpus spontaneo di sinto lombardo.

Infine, nella Tabella 11 non sono stati riportati i numerali, in quanto costituiscono una classe potenzialmente aperta difficilmente rappresentabile attraverso un elenco di lemmi. Tuttavia, come già anticipato, un'analisi del corpus mostra che non sono mai utilizzati numerali cardinali superiori al sei di origine romaní: tali forme appaiono esclusivamente in italiano, salvo in momenti di riflessione metalinguistica in cui i parlanti associano al "sinto stretto" le forme di origine greca *eftà* 'sette', *oxtò* 'otto' e *injà* 'nove', salvo dichiarare di non ricorrervi mai nel parlato quotidiano (cfr. anche Capitolo 4). Infine, tutti i numerali ordinali (p.es. *primo*, *secondo*, *terzo*) appaiono sistematicamente selezionati dall'italoromanzo.

5.5.2 Nomi e verbi

Passiamo ora all'analisi della regolarità delle parole contenuto, concentrandoci su nomi e verbi. Ai fini dell'analisi, tutti i lemmi nominali e verbali nel corpus sono stati distinti in classi di regolarità e frequenza; cfr. Figura 23.

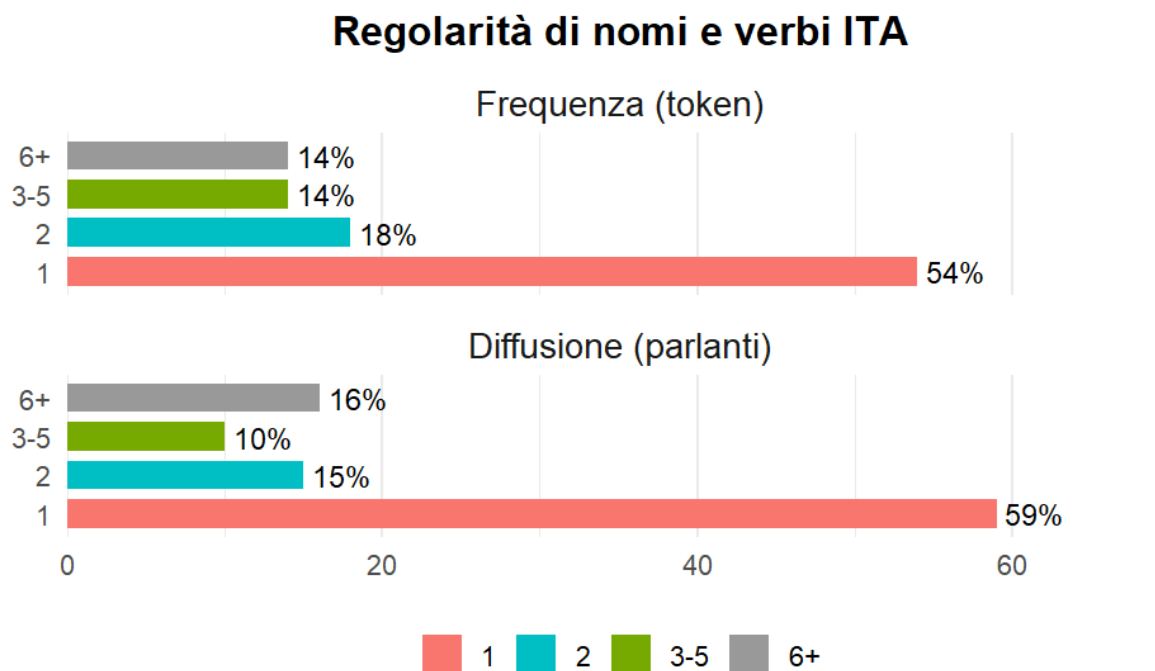


Figura 23: Regolarità dei nomi e dei verbi italoromanzi.

La Figura 23 mostra che la maggior parte dei nomi e dei verbi italoromanzi (il 54%) occorre una sola volta nel corpus, e una percentuale ancora più elevata (il 59%) è prodotta da un solo parlante. Al contrario, soltanto una quota ridotta di lemmi (circa il 15%) è attestata almeno sei volte o nel parlato di almeno sei locutori². Complessivamente, questi dati suggeriscono che le parole contenute sono meno regolari nel parlato in sinto lombardo rispetto alle parole funzione. Di conseguenza, la loro inserzione in un discorso in sinto lombardo appare costituire un meccanismo più dinamico, subordinato alle esigenze referenziali del discorso.

Nondimeno, anche nel caso dei nomi e dei verbi è possibile individuare alcuni lemmi altamente frequenti, diffusi e privi di equivalenti nativi nel corpus, che possono essere interpretati con buona sicurezza come prestiti acclimatati in sinto lombardo. Allo stesso tempo, si riscontra in altri casi variazione tra un nome o un verbo italoromanzo e i suoi equivalenti romaní; l'analisi di questi casi può dunque essere interessante rispetto al continuum prestito-*code-switching*. Le sezioni successive discutono più in dettaglio la regolarità dei nomi (§ 5.5.2.1) e dei verbi (§ 5.5.2.2).

² La soglia di sei occorrenze/sei parlanti non ha valore assoluto e viene adottata come criterio operativo per distinguere forme relativamente diffuse da forme sporadiche nel corpus.

5.5.2.1 Nomi

Dopo aver analizzato nel dettaglio le parole funzione, la Tabella 12 riporta i quindici nomi di origine italo-romanza più regolari nel corpus, indicandone frequenza (in termini di *token*) e diffusione (in termini di parlanti), nonché l'eventuale presenza di un equivalente nativo nel corpus o nel dizionario di Soravia & Fochi (1995).

Forma	Freq.	Diff.	Equivalente 'nativo' nel corpus	Equivalente 'nativo' in Soravia & Fochi (1995)
Anno	103	24	<i>Bers</i>	<i>Bers</i>
Volta	82	21	<i>Molo</i>	<i>Molo</i>
Nonno 'nonno'	53	14	/	<i>Papo</i>
Persona	50	18	/	/
Parte	46	16	/	/
Giro	39	18	/	/
Scuola	37	15	/	/
Festa	36	13	/	/
Vita	26	9	/	<i>Ĝipe</i>
Giostra	25	11	/	<i>Mambruka, milo</i>
Cultura	25	2	/	/
Cinema 'film, cinema'	24	11	/	/
Cusin 'cugino'	24	15	/	/
Piasa 'piazza'	23	10	/	/
Tradizione	23	12	/	/

Tabella 12: Nomi italo-romanzi più regolari in sinto lombardo.

Come si osserva nella Tabella 12, la maggior parte dei nomi più regolari nel parlato in sinto lombardo appartiene a campi semantici relativamente di base (come, ad esempio, la parentela o il tempo) e non mostra variazione con equivalenti nativi nel corpus: questi fattori permettono di qualificare tali forme come prestiti stabilizzati nel sinto lombardo. Un'eccezione è rappresentata dal lemma *anno*, posto in variazione con l'equivalente nativo *bers*, più frequente: come vedremo nel prossimo capitolo, tale variazione è condizionata dalla lingua del numerale precedente. Similmente, *volta* appare in variazione con l'equivalente nativo *molo*: tuttavia, il fatto che *volta* occorra 82 volte, mentre *molo* soltanto 4, mostra chiaramente come la forma italo-romanza rappresenti di fatto l'opzione più disponibile ai parlanti nel sinto contemporaneo, qualificandosi più vicina al polo del prestito. Si

consideri l'estratto di seguito in (30), in cui un parlante fornisce una riflessione metalinguistica esplicita sulla competizione tra la forma italoromanza *volta* e l'equivalente *molo*.

- (30) allora, *men ja volta* / tipo ecco già volta non dicevamo volta *na pinajas=lu miga / pinajas molo*
 'allora, noi una volta, tipo ecco già volta non dicevamo *volta*, non lo dicevamo; dicevamo *molo*'

Similmente a quanto già visto per le parole funzione, si può inoltre notare che il dizionario di Soravia & Fochi (1995) talvolta registra la presenza di forme romaní che di fatto non compaiono mai nelle conversazioni dei parlanti, sebbene essi, in alcuni casi, possano essere consci della loro esistenza. Questo è per esempio il caso della forma *papo/u* per 'nonno', un prestito di origine greco-bizantina, nota ad alcuni parlanti, che tuttavia utilizzano l'equivalente dialettale *nono* come unica opzione.

Un'analisi estesa oltre ai lemmi più frequenti mostra che in diversi casi un nome italoromanzo, nel corpus, alterna con un equivalente romaní caratterizzato da un diverso grado di regolarità; a tal proposito, è possibile distinguere almeno tre configurazioni ricorrenti.

Un primo gruppo è costituito da nomi italoromanzi ad alta frequenza e diffusione usati in variazione con equivalenti romaní meno frequenti e diffusi. In questi casi, il criterio della regolarità orienta l'interpretazione delle forme italoromanze verso il polo del prestito. Oltre al già citato esempio di *volta* vs. *molo*, altri esempi in questo senso sono:

- *tempo* (12 occorrenze per 7 parlanti) vs *siro* (4 occorrenze per 3 parlanti);
- *lavoro* (14 occorrenze per 8 parlanti) vs *labatarpe* (6 occorrenze per 2 parlanti);
- *zia* (12 occorrenze per 6 parlanti) vs *bibi* (4 occorrenze per 2 parlanti).

L'esempio in (31) mostra un esempio di eterocorrezione dell'uso della parola *lavoro* con l'equivalente romaní *labatarpe*. Come si nota, una parlante, MON, inserisce l'inserzione *lavoro* in un discorso in sinto lombardo, e l'altro parlante intervistato, MOR, interviene suggerendo la forma nativa *labatarpe*, che viene ripetuta da MON nel turno successivo.

- (31) MON *is sukar* parchè prima da *krol is=tu tur lavoro*
 'era bello perché prima di tutto avevi il tuo lavoro'
 MOR *tur labatarpe*
 'il tuo lavoro'
 MON *tur labatarpe* / e poi *prindžareas but* personi

‘il tuo lavoro... E poi conoscevi molte persone’

Un secondo gruppo è poi costituito da nomi italoromanzi a bassa frequenza e diffusione (molto spesso *hapax*) che alternano con equivalenti romaní più regolari. In questi casi, il criterio della regolarità sembra orientare l’interpretazione di tali inserzioni verso il polo del *code-switching*. Rientrano in questo gruppo, ad esempio:

- *ragazzo* (1 occorrenza) vs. *raklo* (10 occorrenze per 5 parlanti);
- *città* (1 occorrenza) vs. *gau* (16 occorrenze per 11 parlanti);
- *oro* (1 occorrenza) vs. *sunakaj* (5 occorrenze per 3 parlanti);
- *dio* (1 occorrenza) vs. *deval* (44 occorrenze per 8 parlanti).

Un terzo gruppo comprende infine casi in cui nome italoromanzo ed equivalente romaní mostrano livelli di frequenza e diffusione simili. Qui il criterio della regolarità non consente una chiara collocazione lungo il continuum. Esempi includono:

- *carro* (2 occorrenze per 1 parlante) vs. *vago* (2 occorrenze per 2 parlanti);
- *ballo* (1 occorrenza) vs. *kilape* (1 occorrenza);
- *bar* (1 occorrenza) vs. *vjerta* (3 occorrenze per 2 parlanti)
- *colpa* (2 occorrenze per 2 parlanti) vs. *dos* (3 occorrenze per 2 parlanti)

5.5.2.2 Verbi

Passando ai verbi, la Tabella 13 fornisce una lista dei quindici verbi italoromanzi più frequentemente utilizzati nel parlato in sinto lombardo, insieme alle stime relative alla loro diffusione tra parlanti e la presenza o assenza di un equivalente nativo nel corpus e nel dizionario di Soravia & Fochi (1995). I verbi italoromanzi sono riportati loro forma dell’infinito, mentre le loro controparti romaní nella forma radicale.

Forma	Freq.	Diff.	Equivalente ‘nativo’ nel corpus	Equivalente ‘nativo’ in Soravia & Fochi (1995)
Dipendere	53	18	/	/
Dire	24	7	<i>Pin-</i>	<i>Pin-</i>
Arrivare	22	17	/	/
Girare	22	17	/	/
Pensare	18	11	<i>Mand-</i>	<i>Tinkar-</i>
Finire	18	10	/	/
Capitare	14	7	/	/
Aiutare	14	4	/	/
Perdonare	12	9	/	/
Frequentare	11	7	/	/
Odiare	11	6	/	/
Fermare	10	6	/	<i>Vatar- pe</i>
Scrivere	10	6	/	<i>Seivar-</i>
Servire	10	6	/	/
Cominciare	9	8	/	/

Tabella 13: Verbi italoromanzi più regolari in sinto lombardo.

Come si osserva, i verbi più frequenti sono *dipendere* e *dire*. Tuttavia, l’analisi dell’integrazione morfologica e della struttura argomentale proposta al Capitolo 6 mostrerà che questi verbi tendono a non essere mai integrati nella morfosintassi del sinto lombardo, indicando un possibile utilizzo come segnali discorsivi piuttosto che come verbi prototipici.

Tra i verbi più prototipici, che, come vedremo, appaiono morfologicamente integrati nel sinto lombardo, il più frequente è ‘arrivare’, foneticamente articolato come *(a)rivar-* ‘arrivare’, verosimilmente di origine dialettale. Esso non presenta variazione con equivalenti nativi nel corpus e nel dizionario di Soravia & Fochi (1995) non è indicato alcun suo possibile equivalente preromanzo; inoltre, si nota che la radice romanza *riv-* è indicata con significato di ‘arrivare’ per il vicino sinto piemontese. È interessante notare che lo stesso verbo è indicato come il più frequente nel walser issimese da Dal Negro (2006), la quale osserva che *arrivare* è uno dei verbi più frequentemente presi a prestito dall’italoromanzo in diverse varietà alloglotte in Italia. I dati del sinto lombardo confermano dunque questa tendenza, le cui motivazioni restano tuttavia da chiarire. Successivamente nei ranghi di frequenza compaiono i verbi *girare* e *pensare*. Mentre il primo non appare in variazione con equivalenti nativi nel corpus, né si registra nel dizionario di Soravia & Fochi (1995) un suo

equivalente preromanzo, il secondo appare nel corpus in variazione con il verbo nativo *mand-*. Tuttavia, mentre *pensare* mostra 18 occorrenze per 11 parlanti, l'equivalente nativo *mand-* mostra solo 4 occorrenze per 2 parlanti. Ciò sembra indicare che il verbo romanzo è maggiormente regolare rispetto alla sua controparte nativa, qualificandosi dunque molto vicino al polo del prestito. Il dizionario di Soravia & Fochi (1995) indica come forma lombarda con significato di 'pensare' il verbo *tinkar*, che tuttavia i parlanti nel corpus utilizzano nel senso di 'ricordare'. Infine, segue in ordine di frequenza il verbo *finire*, per cui non si riscontra alcun equivalente nativo né nel corpus né in Soravia & Fochi (1995), menzionato da Dal Negro (2006) come il secondo lemma più regolare dopo 'arrivare' nel walser issimese. In generale, analogamente a quanto osservato per i nomi, la maggior parte dei verbi più regolari non mostra dunque alternanza con equivalenti nativi, suggerendo un avanzato grado di stabilizzazione.

Accanto a questi casi, si osservano tuttavia configurazioni di variazione tra verbi italoromanzi ed equivalenti romaní con diversi livelli di regolarità. Anche per i verbi è possibile distinguere tre configurazioni principali, similmente a quanto visto per i nomi. In particolare, si individua un gruppo di verbi italoromanzi a bassa frequenza (1 o 2 occorrenze) che si trovano in variazione con equivalenti romaní più frequenti e più diffusi. In tali casi, l'inserzione italoromanza appare più orientata verso il polo del *code-switching* che del prestito. Esempi includono:

- *insegnare* (2 occorrenze per 2 parlanti) vs. *sikar-* (14 occorrenze per 6 parlanti);
- *mantenere* (1 occorrenza) vs. *rikar-* (23 occorrenze 7 parlanti);
- *cercare* (1 occorrenza) vs. *rud-* (3 occorrenze per 3 parlanti);
- *fumare* (2 occorrenze per 1 parlante) vs. *faifar-* (4 occorrenze per 4 parlanti)³.

Infine, si danno anche casi di verbi italoromanzi a bassa frequenza che alternano con equivalenti romaní con livelli di regolarità simili. Esempi sono:

- *contare* (1 occorrenza) vs. *gin-* (2 occorrenze per 2 parlanti)
- *crescere* (3 occorrenze per 1 parlante) vs. *v- baro* (4 occorrenze per 2 parlanti)

5.6 Calchi lessicali: una panoramica qualitativa

Nelle sezioni precedenti di questo capitolo, ci siamo occupati di analizzare l'interazione tra elementi italoromanzi ed elementi romaní nel parlato bilingue dei sinti lombardi. Tuttavia, il trasferimento di forme nella loro sostanza fonomorfologica da una lingua B a una lingua A non rappresenta l'unica

³ A proposito di questa alternanza, i consulenti riportano che i parlanti di piazzale Europa utilizzano la forma romanza *fumare* mentre i parlanti di via Bramante utilizzano la forma preromanza *faifar-*: i dati del corpus sembrano confermare parzialmente questa osservazione, ma sarebbero necessari ulteriori studi per validarla.

strategia a disposizione dei parlanti bilingui, che possono riprodurre modelli lessicali allogeni traducendoli, del tutto o in parte, attraverso materiale nativo. È a questo insieme di fenomeni che ci riferiamo qui con il termine di calco lessicale.

Seguendo Backus & Dorleijn (2009), limitiamo l'etichetta di calco ai casi in cui la fonte della replicazione è identificabile in una struttura lessicale, distinguendoli dalla replicazione di strutture sintattiche astratte e schematiche (come l'ordine dei costituenti nella frase). In una prospettiva *usage-based*, i modelli lessicali replicati non coincidono necessariamente con singoli morfemi, ma possono anche comprendere sequenze multimorfemiche relativamente fisse, potenzialmente memorizzate come unità (cfr. Backus, 2014). Sul piano descrittivo, in letteratura si distinguono comunemente calchi totali e parziali. Nel primo caso, il modello originario viene interamente tradotto attraverso materiale della lingua ricevente (nella letteratura anglofona si parla di *loan translations*); nel secondo caso, solo una parte del modello viene tradotta, mentre altri elementi rimangono inalterati (*loan blends*). Tuttavia, in entrambi i casi, il meccanismo sottostante rimane il medesimo: “the speaker wants to say something in a base language in the way that it is said in the other language” (Backus & Dorleijn, 2009: 90).

Un problema metodologico non trascurabile nell'analisi dei calchi riguarda la necessità di escludere che strutture apparentemente replicate non siano in realtà preesistenti nella lingua ricevente, indipendentemente dal contatto. Nel caso del sinto lombardo, l'assenza di documentazione diacronica rende questo controllo particolarmente complesso. Di conseguenza, l'analisi che segue ha un taglio volutamente qualitativo: essa non mira a quantificare il fenomeno, ma a discutere i casi più trasparenti.

Per chiarezza di trattazione, distinguiamo i calchi in base al tipo di struttura che funge da modello: singole parole (§ 5.6.1), parole composte o sintagmatiche (§ 5.6.2), parole funzione (§ 5.6.3).

5.6.1 Calchi di singole parole

I calchi di singole parole sono spesso noti in letteratura come “calchi semantici” (Cerruti & Regis, 2015), in quanto ciò che viene trasferito da una lingua B a una lingua A è il significato e non la struttura interna del lessema. Sebbene casi di questo tipo appaiano piuttosto rari nel corpus di sinto lombardo, è possibile citare almeno due esempi: *bisarel* ‘buttare’, usato nel senso specifico di ‘buttare la pasta’ (32), e *karel* ‘fare’, usato nel senso italiano di ‘dire’ (33).

- (32) poi *bisarea* i pasta e *čivea drent u lon*
 ‘poi butti la pasta e ci metti dentro il sale’

- (33) *u gadžo **karela** na den miga xole par tumengar daj*
 ‘il gagio fa: non arrabbiatevi per vostra madre’

5.6.2 Calchi di parole composte o sintagmatiche

Nel corpus sono anche attestati calchi, totali o più spesso parziali, di parole contenute composte o sintagmatiche. Come esempio di calco totale si può menzionare la forma *pirani indur ker*, chiaramente modellata sull’italiano ‘fidanzata in casa’ (34), mentre un esempio di calco parziale è *terza stunda* ‘terza ora’ (35), dove il nome *stunda* è romaní (si tratta in particolare di un prestito dal tedesco) e il numerale italiano.

- (34) *me jomas **pirani indur ker***
 ‘io ero fidanzata in casa’

- (35) ***terza stunda** kardam matematica*
 ‘terza ora abbiamo fatto matematica’

Infine, un esempio di calco lessicale a livello di verbo complesso può essere la forma *stela vrin* ‘salta fuori’ (nel senso di ‘emerge’, ‘si scopre’, con una sfumatura mirativa), risultante dall’unione del verbo *st-* ‘saltare’ e *vrin* ‘fuori’; l’estratto in (36) ne mostra un uso in contesto.

- (36) *poi **stela vrin** che bistardas=pe*
 ‘poi salta fuori che si è dimenticato’

In generale, le forme qui discusse risultano piuttosto infrequenti (si tratta nella maggior parte di *hapax*) e sembrano dunque emergere spontaneamente nel discorso, come dimostra anche la loro libera variazione con equivalenti direttamente copiati dall’italiano e non tradotti (come la forma *fidanzato in casa*, attestata due volte nel corpus). Un’analisi più dettagliata dei calchi parziali (come nel caso di *terza stunda*) sembra mostrare che gli elementi tradotti sono tipicamente quelli caratterizzati da un equivalente disponibile in sinto lombardo, mentre quelli non tradotti ne sono privi.

5.6.3 Calchi di parole funzione multimorfemiche e locuzioni avverbiali

Un gruppo particolarmente interessante di calchi è costituito da parole funzioni multimorfemiche che presentano una struttura internamente analizzabile e dunque suscettibile di replicazione parziale. Si può citare a questo proposito la forma *daparkrol* ‘dappertutto’, realizzata sul modello dialettale

dapartüt, in cui l'ultimo membro ('tutto') è stato tradotto con il materiale romaní disponibile (ossia l'indefinito *krol*), mentre i due segmenti precedenti (*da* e *par*) rimangono invariati; cfr. (37) per un esempio in contesto.

- (37) *girarena pisal daparkrol*
 'girano un po' dappertutto'

Un esempio simile è *qualche jek* 'qualcuno', che sembra originare dallo stesso meccanismo: la forma è stata analizzata dai parlanti nei suoi elementi costitutivi e l'ultimo elemento, *uno*, è stato tradotto attraverso l'equivalente romaní *jek* (cfr. 38).

- (38) magari **qualche jek** *ljas rom* a quindici anni
 'magari qualcuno si è sposato a quindici anni'

Diversamente dai casi visti al § 5.6.2, che appaiono tendenzialmente idiosincratici e spesso riconducibili a occorrenze isolate nel corpus, forme come *dapartut* e *qualchejek* mostrano una maggiore stabilità nel parlato dei sinti lombardi, pur essendo utilizzati in variazione con le forme interamente italoromanze *dapartuto* e *qualchedun*.

Infine, si rintracciano alcuni casi di calchi di locuzioni avverbiali italiane: un esempio è la forma *akana come akana* 'ora come ora'. Si noti che la parola italiana *come* possiede un equivalente ben attestato in sinto, *ar*. Differentemente da quanto visto in precedenza, in questo caso l'assenza di una forma equivalente nella lingua ricevente non sembra motivare il perché tale forma non sia stata interamente tradotta dal locutore.

5.7 Discussione

In questo capitolo abbiamo proposto una panoramica dell'incidenza quantitativa e delle caratteristiche qualitative del parlato bilingue dei sinti lombardi, a partire dall'analisi del corpus spontaneo descritto al Capitolo 3.

Un'analisi della composizione linguistica del corpus a livello di singoli *token* di parole ha mostrato che le forme di origine italoromanza costituiscono circa il 42% del totale complessivo, un dato che risulta tanto più rilevante se si considera che il sinto lombardo era stato stabilito come lingua di base dell'interazione nelle interviste (cfr. Capitolo 3). Un confronto con altri corpora multilingui in situazioni sociolinguisticamente simili ha mostrato che il corpus di sinto lombardo presenta, tra quelli considerati, la più alta incidenza di parole provenienti dalla/e lingua/e di contatto, in questo caso

l'italoromanzo, pur collocandosi vicino ad altri corpora di romaní (Adamou, 2016; Mirić & Ćirković, 2022). Nel Capitolo precedente, il sinto lombardo era stato caratterizzato come una lingua “vital” nella terminologia di Lewis & Simons (2010), in quanto la sua trasmissione alle generazioni più giovani è ancora attiva, nonostante alcuni segnali di indebolimento. Il fatto che, nonostante questa relativa vitalità, il parlato in sinto lombardo mostri maggiore incidenza di parole di contatto rispetto a molte lingue la cui trasmissione si è già interrotta o è fortemente in regressione (si considerino, tra gli altri, i casi di iscateco, walser o nashta) conferma l'osservazione di Adamou (2016) secondo cui il grado di *endangerment* di una lingua non correla necessariamente con la quantità di discorso bilingue prodotto dai suoi parlanti.

In una seconda fase, si è potuto osservare che il parlato bilingue dei sinti lombardi si caratterizza, nell'ampia maggioranza dei casi, per l'inserzione di singole parole o costituenti in un discorso prevalentemente in romaní. Questo meccanismo combinatorio è compatibile sia con il *code-mixing* di tipo insertivo (Muysken, 2000), sia con il prestito lessicale. Più in generale, questo risultato è in linea con la letteratura sul contatto linguistico, che osserva come brevi inserzioni di singole parole o costituenti rappresentino il *pattern* dominante di parlato bilingue indipendentemente dalla situazione sociolinguistica di una comunità o dalla tipologia delle lingue coinvolte (Poplack, 2018; Adamou, 2016). Viceversa, i casi di *code-switching* alternante verso l'italiano appaiono molto più rari e spesso pragmaticamente marcati, venendo utilizzati dai parlanti per selezionare un interlocutore specifico o per ottenere particolari effetti discorsivi.

L'analisi della distribuzione delle inserzioni per parte del discorso e per campo semantico ha poi mostrato che la probabilità di un'inserzione italoromanza di appartenere a una determinata classe di parole o campo semantico è piuttosto variabile. Per esempio, per quanto riguarda le classi di parole, i connettivi e i segnali discorsivi hanno molta più probabilità di provenire dall'italoromanzo in un discorso in sinto lombardo rispetto ai verbi, ai pronomi e ai determinanti. Le differenze osservate rispecchiano in parte le gerarchie di prestabilità individuate in letteratura e discusse al Capitolo 1, che evidenziano la particolare propensione al prestito dei connettivi e segnali discorsivi e dei nomi sulle altre classi di parole. Allo stesso tempo, nel corpus di sinto lombardo è stata anche notata la presenza massiccia di elementi italoromanzi, come numerali e preposizioni, il cui prestito è considerato come relativamente raro nella letteratura sul contatto linguistico. Per esempio, Thomason & Kaufman (1988) collocano il prestito delle preposizioni al terzo e ultimo stadio del prestito lessicale nel loro modello relativo alla *borrowability*, discusso al Capitolo 1. Sebbene il modello di Thomason & Kaufman (1988) sia concepito soprattutto in una prospettiva diacronica, la sua applicazione ai dati di sinto lombardo mostra come, in una situazione di contatto sufficientemente prolungato e intensivo,

non emergano limiti assoluti alla possibilità di adottare elementi alloglotti (cfr. anche Dal Negro, 2009: 92).

Un'analisi della regolarità delle inserzioni italoromanze, operazionalizzata in termini di frequenza, diffusione e variazione con equivalenti nativi, ha poi permesso di collocare più precisamente queste forme lungo il continuum prestito-*code-switching*. Uno dei risultati più significativi emersi riguarda la fondamentale differenza, in termini distributivi, tra parole funzione e parole contenuto.

Le parole funzione italoromanze appaiono generalmente molto stabili nel parlato in sinto lombardo e spesso non alternano con equivalenti nativi preromanzi, rappresentando pertanto l'unica risorsa a disposizione dei sinti lombardi. Un ulteriore indice di integrazione delle parole funzione italoromanze nel sinto lombardo è dato dal fatto che i connettivi sono di norma utilizzati sia nella loro funzione interazionale di segnali discorsivi, sia nella loro funzione di congiunzioni. Poiché le congiunzioni sono maggiormente incassate nella sintassi della frase, il loro prestito viene generalmente considerato successivo a quello dei segnali discorsivi in situazioni di contatto linguistico (cfr. Dal Negro, 2005b; Stolz, 2007). Il caso del sinto lombardo conferma, dunque, una situazione di contatto avanzata. Viceversa, le parole contenuto mostrano una regolarità molto minore, il che è strettamente legato alla natura stessa delle parole contenuto, una classe aperta la cui possibilità di occorrenza è limitata alla ricorrenza tematica. In diversi casi, si osserva inoltre la variazione di nomi e verbi italoromanzi con equivalenti nativi. Nel complesso, questi risultati suggeriscono che, in sinto lombardo, le parole funzione di origine italoromanza tendono ad avvicinarsi al polo del prestito stabilizzato, mentre le parole contenuto (pur includendo alcuni lemmi chiaramente acclimatati) realizzano più spesso processi inserzionali dinamici, collocabili più vicino al polo del *code-switching*.

Un ultimo risultato, pur prodotto in maniera esclusivamente qualitativa, riguarda l'abbondante presenza di calchi lessicali nel parlato in sinto lombardo, che affiancano il trasferimento materiale di lessemi e costituenti. Particolarmente significativo è il fatto che alcuni calchi sono parziali e coinvolgono sequenze multimorfemiche idiomatiche, "spezzando" dunque costituenti relativamente fissi. Questo tipo di comportamento suggerisce che tra le due lingue sia percepito un elevato grado di isomorfismo strutturale, che consente ai parlanti di stabilire corrispondenze punto per punto tra segmenti delle due strutture e di operare traduzioni interne. Tali riflessioni verranno riprese al Capitolo 8, relativo alla classificazione più generale del sinto lombardo nella tipologia del discorso bilingue. Il prossimo capitolo si focalizza invece su un altro aspetto centrale del parlato bilingue dei sinti lombardi, e cioè le strategie di integrazione linguistica delle inserzioni italoromanze.

Capitolo 6. Strategie di integrazione delle inserzioni italoromanze

Dopo aver analizzato gli schemi ricorrenti del parlato bilingue dei sinti lombardi, questo capitolo si concentra sulle strategie di integrazione, prevalentemente morfologica e morfosintattica, delle inserzioni italoromanze nel parlato in sinto lombardo. Nella linguistica del contatto, il criterio dell'integrazione a livello linguistico è stato considerato particolarmente rilevante per distinguere tra casi di *code-switching*, che mantengono le strutture grammaticali della lingua donatrice, e casi di prestito, che implicano invece un'integrazione nelle strutture linguistiche della lingua ricevente (Poplack, 2018). Tuttavia, come discusso nel dettaglio al Capitolo 1, i meccanismi di integrazione non sempre agiscono uniformemente su tutti i livelli linguistici (fonetico, morfologico, sintattico), classi di parole o diversi strati etimologici di contatto.

In questo capitolo, vengono analizzate le strategie che emergono dal corpus di sinto lombardo per integrare linguisticamente le inserzioni italoromanze. In particolare, nel § 6.1 viene proposta un'analisi preliminare dell'integrazione a livello fonetico delle parole italoromanze, mentre i §§ 6.2 e 6.3 si focalizzano sul livello morfosintattico e sono dedicati, rispettivamente, ai nomi e ai verbi. Infine, il § 6.4 conclude con una discussione dei principali risultati.

6.1 Integrazione a livello fonetico/fonologico

Pur non avendo condotto un'analisi sistematica dell'integrazione delle parole di origine italoromanza a livello fonetico/fonologico, alcuni esempi selezionati dal corpus confermano la variabilità rilevata dagli studiosi in altri contesti bilingui (Poplack et al., 2020).

A questo riguardo, può essere utile approfondire brevemente che cosa capita con parole italiane contenenti i suoni affricati [ts] e [dz], assenti nell'inventario fonologico del sinto lombardo (Sorrenti, 2014). L'analisi del corpus mostra che, in alcuni casi, [ts] e [dz] vengono mantenuti, mentre in altri casi vengono resi attraverso i fonemi più vicini nell'inventario fonologico del sinto lombardo, ovvero [s] e [z]. Significativamente, può esistere variazione nelle produzioni di uno stesso parlante a livello di uno stesso lemma. Si considerino a questo proposito gli esempi (40) e (41), che catturano momenti diversi nel parlato degli stessi individui.

- (40) perché i tradizione [tradi'tsjone] mukjam=lu ja pisal (...)
'perché la tradizione l'abbiamo lasciata un po'

mengar tradizione [tradi'sjone] *paradas=pe ja pisal*

‘la nostra tradizione è cambiata un po’

- (41) *u zingari* [‘zingari] *i ja vavar* tipo /
 ‘gli zingari sono un altro tipo’
men jam miga zingari [‘dzingari] no? /
 ‘noi non siamo zingari, no?’

In (40), notiamo che il suono affricato [ts] presente nel nome italiano *tradizione* è mantenuto come tale nella prima occorrenza (dunque la forma non apparirebbe come integrata), mentre nella seconda occorrenza è reso attraverso la sibilante sorda [s] (dunque la forma apparirebbe come integrata). Similmente, in (41), una parlante, a brevissima distanza, pronuncia la parola *zingari* prima con la sibilante sonora [z] e successivamente con l’affricata [dz].

Tuttavia, va anche precisato che, in molti casi, risulta arduo stabilire se le parole italoromanze contenenti i suoni [s] e [z] nel parlato in sinto lombardo siano effettivamente parole italiane adattate alla fonetica del sinto lombardo, o se si tratti piuttosto di forme di origine dialettale, dal momento che anche nel dialetto lombardo sono assenti i fonemi [ts] e [dz] e spesso la differenza tra una forma italiana e dialettale risiede proprio nella loro realizzazione. Questa somiglianza nei sistemi fonologici del sinto lombardo e del dialetto lombardo spiega anche la difficoltà nell’operare, in sede di annotazione dei dati, distinzioni chiare tra forme di origine dialettale e forme italiane potenzialmente integrate nel sistema fonologico del sinto lombardo.

Più in generale, è importante notare che i sinti lombardi possono mostrare variazione nella resa di questi suoni affricati anche nel loro parlato monolingue in italiano, realizzando cioè il suono [ts] come [s] e il suono [dz] come [z], un tratto che diventa talvolta oggetto di esplicite riflessioni metalinguistiche da parte dei parlanti (cfr. Capitolo 4). Ciò è anche documentato nell’esempio (42) di seguito, che riporta un caso di *code-switching* alternante verso l’italiano.

- (42) *ringrasiamo* [ringra'sjamo] l’evangelo

Poiché una certa variabilità nella resa dei suoni [ts] e [dz] emerge anche nel parlato monolingue italiano, la loro resa come [s] e [z] nel parlato in sinto lombardo potrebbe dunque indicare ciò che Poplack e colleghi definiscono un’“integrazione involontaria”: non già l’intenzione di integrare la parola italoromanza nel sistema fonologico della lingua ricevente, bensì la riproduzione di una variabilità già esistente nella lingua donatrice. In definitiva, dati i numerosi problemi discussi in

questa sezione, il criterio dell'integrazione fonetica appare poco rilevante per distinguere tra prestito e *code-switching*. Nelle sezioni a seguire, affrontiamo in maniera più sistematica l'integrazione a livello morfologico dei nomi e dei verbi.

6.2 Integrazione dei nomi

Dopo aver discusso brevemente le problematiche connesse all'applicazione del criterio dell'integrazione a livello fonetico/fonologico per distinguere tra prestito e *code-switching*, passiamo ora ad analizzare più in dettaglio l'integrazione dei nomi a livello morfosintattico. Al § 6.2.1, discutiamo la loro integrazione morfologica, mentre al § 6.2.2 analizziamo la lingua con cui vengono espressi i loro modificatori all'interno del sintagma nominale.

6.2.1 Integrazione morfologica

Per affrontare la questione dell'integrazione morfologica dei nomi italoromanzi nel sinto lombardo, è utile introdurre brevemente il funzionamento della morfologia nominale in romaní.

In romaní, i nomi hanno un genere inerente e si flettono per numero e per caso (Elšik, 2020). Il sistema dei casi non è tuttavia mantenuto in sinto lombardo per la flessione dei nomi, in cui le uniche forme del paradigma procedono dal caso nominativo (Meli, 2023). Dal punto di vista della loro flessione, nella maggior parte delle varietà di romaní i nomi si dividono in due gruppi principali: i nomi "oikoclitici" e i nomi "xenoclitici" (cfr. Elšik & Matras, 2006). I nomi oikoclitici sono di origine indoaria o comunque pre-europea e vengono flessi attraverso morfemi indoari ereditati, che danno origine a svariate classi flessive. I nomi xenoclitici, di origine greca e post-greca, si flettono attraverso morfemi indotti dal contatto con il greco. Di fatto, dopo il contatto con il greco, i morfemi xenoclitici rimangono dunque gli unici produttivi in romaní.

In sinto lombardo, come in molte varietà di romaní, i nomi xenoclitici hanno generalmente un'uscita in *-o* al maschile singolare, in *-a* al femminile singolare e in *-i* al plurale ambigenere, ovvero sia maschile che femminile (Meli, 2023). A causa di una tendenza più generale alla neutralizzazione dei fonemi */-i/* ed */-e/* in finale di parola in sinto lombardo, Meli (2023) osserva che il plurale ambigenere in *-i* può essere anche realizzato come *-e*. Come riporta Meli (2023), sono flessi secondo questo schema prestiti greci come *foro* 'città' (pl. *fori*) o tedeschi come *vago* 'carro' (pl. *vagi/e*); un esempio al femminile è *berga* 'montagna' (cfr. ted. *Berg*). In questo scenario, è interessante osservare che cosa accade con i nomi di origine italoromanza sulla base dei dati presenti nel corpus. Un'analisi in questa direzione è già stata intrapresa da Meli (2023); l'obiettivo di questa sezione è ampliarla sulla base dei dati presenti nel corpus di parlato spontaneo.

L'analisi dei dati mostra che i nomi di origine chiaramente dialettale usati in discorsi in sinto lombardo tendono generalmente a essere integrati nelle classi xenoclitiche. In particolare, i nomi femminili di origine dialettale in *-a* sono generalmente assimilati alla classe dei femminili xenoclitici con singolare in *-a* e plurale in *-e/-i* (p.es. *spal-a*, *spal-e/i* 'spalla/e'; *pias-a*, *pias-e/i* 'piazze'). Mentre la forma del singolare in *-a* coincide tra la lingua d'arrivo e di partenza e non permette di stabilire l'integrazione, la forma del plurale in *-i* indica integrazione, dal momento che nel dialetto lombardo esso sarebbe realizzato attraverso la soppressione della vocale finale. A questa classe in *-a* / *-i* vengono anche ricondotti alcuni maschili dialettali in *-a*, come *barba* 'zio' (pl. *barbi*); il genere maschile viene mantenuto e disambiguato dall'accordo con i determinanti o gli aggettivi (*u barba* 'lo zio').

I maschili dialettali in consonante, invece, possono pienamente assimilarsi ai maschili xenoclitici con singolare in *-o* e plurale in *-i* (p.es. *nivud-o*, *nivud-i* 'nipote/i') oppure mantenere l'uscita in consonante al singolare (p.es. *cusin* 'cugino', *ospidal* 'ospedale') ma con plurale romaní in *-i* (p.es. *cusin-i* 'cugini'). Tendono a confluire in questa classe con singolare in consonante e plurale in *-i* anche alcune forme dialettali femminili in consonante come *televisiun* 'televisione' o *generasiun* 'generazione'; anche in questo caso, il genere femminile è mantenuto in sinto lombardo, come si evince dall'accordo con determinanti o aggettivi (*i televisiun* 'la televisione', *ja generasiun nevi* 'una generazione nuova').

Non è ben chiaro che cosa motivi la variazione che si riscontra al maschile singolare delle forme dialettali in consonante, dal momento che non sembrano valide ragioni di ordine fonetico. Per esempio, malgrado sia *nivud* che *cusin* siano nomi bisillabici terminanti in consonante e accentati sulla sillaba finale, sulla base delle occorrenze nel corpus il primo risulta sempre integrato morfologicamente con l'aggiunta di *-o*, mentre il secondo si mantiene come tale. Notiamo inoltre che può esistere variazione nell'integrazione di uno stesso lemma, e talvolta anche nel parlato di uno stesso locutore. Si considerino i casi di *bicier* / *biciero* 'bicchiere' e *macel* / *macelo* 'macello', negli esempi (43-46) di seguito. È importante notare che gli esempi (43-44) sono stati prodotti da uno stesso parlante, così come gli esempi (45-46).

- (43) *men ja **bicier** da mol pal u xabe piaj=lo* li stes
'noi un bicchiere di vino con il cibo lo beviamo lo stesso'

- (44) *finchè pjea ja tinu **biciero** e bon džala misto*
'finché bevi un bicchierino e bon, va bene'

- (45) prima volta *pinden ch'is=li lace* ma i secunda e i tersa *kardom vrin ja macel*
 ‘la prima volta hanno detto che erano buoni, ma la seconda e la terza ho fatto fuori un macello’
- (46) eh prima *is pisa ja macelo* cun u festi u koa du sinti
 ‘eh prima era un macello con le feste, le cose dei sinti’

Passiamo ora a discutere più in dettaglio le strategie di integrazione dei nomi italiani, particolarmente importanti in quanto rappresentano l’unico serbatoio lessicale da cui i sinti lombardi oggi continuano ad attingere produttivamente. A ben vedere, tale integrazione è spesso impossibile da stabilire, in quanto le classi flessive più produttive in italiano (con uscita in *-o/-i* al maschile singolare e plurale e in *-a/-e* al femminile singolare e plurale) di fatto coincidono con le uscite xenoclitiche del sinto lombardo. Dunque, a fronte di esempi come quelli riportati in (47-48) di seguito, risulta impossibile stabilire se le forme singolari *quaderno* e *messa* vengano mantenute morfologicamente intatte dall’italiano o se acquisiscano la morfologia romaní.

- (47) *joi disegnarel=li sempre pur diario*
 ‘lei disegna sempre sul diario’
- (48) *džajas ke messa*
 ‘andavamo a messa’

Anche al plurale, stabilire l’integrazione morfologica dei nomi italiani è difficile in quanto, come anticipato, in sinto lombardo la distinzione tra *-i* ed *-e* è neutralizzata in finale di parola ed entrambe le realizzazioni sono possibili sia al maschile sia al femminile (Meli 2023). Un’analisi più dettagliata del corpus permette di rilevare che, nonostante la neutralizzazione tra *-e* e *-i* sia possibile, nel sinto lombardo di Pavia l’uscita in *-i* è più frequente sia al maschile sia al femminile (alcuni esempi di femminili plurali includono *u pianti*, *u caroti*, *u patati*). Se ciò al maschile è prevedibile, dato l’influsso del modello del maschile plurale italiano in *-i*, nel caso del femminile sembra indicare una certa aderenza al modello tradizionale romaní (originariamente in *-i*) nonostante il possibile influsso del modello del femminile plurale italiano in *-e*.

Si individuano tuttavia alcuni casi specifici in cui è possibile valutare l’integrazione o la mancata integrazione di nomi italiani. Un esempio è costituito dai nomi italiani maschili e femminili con

singolare in *-e*, come *dottore, preside, parte, luce* o *occasione*, che mantengono questa uscita anche nel parlato in sinto lombardo; cfr. esempio (49).

- (49) magari *sena* ja **occasione** / *dikjal ar i* no?
'magari trovano un'occasione... hai visto com'è, no?'

In questo quadro, può essere interessante analizzare la realizzazione dei nomi in *-e* al plurale, che in italiano è realizzato in *-i*. Come osserva infatti Meli (2023):

“se la neutralizzazione [tra *-i* ed *-e*] fosse attiva, queste parole potrebbero fare capo a una classe a genere nascosto, invariabile per numero; se invece ulteriori dati evidenziassero che le due uscite non sono sovrapponibili, potremmo trovarci di fronte ad almeno una nuova classe flessiva con singolare *-e* e plurale *-i*, in cui la neutralizzazione sarebbe dunque bloccata. Considerato anche, come già detto, che la lingua tetto dei sinti lombardi è l'italiano, tale blocco potrebbe doversi all'attivazione delle regole della varietà sovraordinata, come in un costante *code-mixing*”. (Meli, 2023: 416)

Attraverso uno spoglio dei dati presenti nel corpus, si nota che, per questa classe di nomi, nella maggior parte dei casi i parlanti mantengono la distinzione tra singolare in *-e* e plurale in *-i*. Tuttavia, in due casi (peraltro con la stessa parola, *comune*), l'uscita al singolare è chiaramente realizzata come *-i*; cfr. esempio (50). Il fatto che, in (50), *comuni* si trovi al singolare è disambiguato dal verbo *činel* 'tagliare', coniugato alla terza singolare.

- (50) *i u comuni* ch'*i da činel u* pianti
'è il comune che deve tagliare le piante'

L'analisi del corpus permette dunque di osservare che, per i nomi italiani in *-e/-i*, nonostante la distinzione tra *-e* e *-i* al singolare e plurale sia generalmente mantenuta, in alcuni casi è anche possibile una neutralizzazione secondo uno schema tipico del sinto lombardo.

Al termine di questa panoramica, la Tabella 14 schematizza i processi di adattamento dei nomi di origine italo-romanza nel parlato in sinto lombardo per classe flessiva.

Italoromanzo (desinenza singolare)	Esito sinto lombardo (singolare)	Esito sinto lombardo (plurale)
Consonante <i>Nivud</i> (M.) <i>Cusin</i> (M.) <i>Televisiun</i> (F.)	Consonante / -o <i>Nivud-o</i> (M.) <i>Cusin</i> (M.) <i>Televisiun</i> (F.)	-i <i>Nivud-i</i> <i>Cusin-i</i> <i>Televisiun-i</i>
-o <i>Quadern-o</i> (M.)	-o <i>Quadern-o</i>	-i <i>Quadern-i</i>
-a <i>Fest-a</i> (F.) <i>Barb-a</i> (M.)	-a <i>Festa</i> (F.) <i>Barb-a</i> (M.)	-e/-i <i>Fest-e/-i</i> <i>Barb-i</i>
-e <i>Luc-e</i> (F.) <i>Comun-e</i> (F.)	-e/(-i) <i>Luce-e</i> (F.) <i>Comun-e/-i</i> (M.)	-i <i>Luc-i</i> <i>Comun-i</i>

Tabella 14: Principali strategie di integrazione dei nomi italoromanzi.

Dalla panoramica presentata, emerge dunque come i nomi italoromanzi (e in particolare italiani) necessitino di pochi e talvolta nessun adattamento formale per essere impiegati nel parlato in sinto lombardo. Ciò, come abbiamo osservato, è anche dovuto a mutamenti del sinto lombardo, come la perdita del sistema dei casi o l’annullamento del valore distintivo dei fonemi *i* ed *e*. Questi mutamenti hanno portato a un crescente isomorfismo tra i paradigmi nominali del sinto lombardo e dell’italiano, al punto che, in numerosi casi, risulta impossibile stabilire se una forma sia morfologicamente integrata o mantenuta non integrata dai parlanti. Riprendendo la terminologia di Matras (2020), si potrebbe parlare di una vera e propria “fusione” a livello morfologico, in cui cioè la distinzione strutturale tra le due lingue in contatto, e dunque l’assegnazione dei singoli morfemi all’una o all’altra, tende ad annullarsi. In tale scenario, appare evidente come il criterio dell’integrazione a livello morfologico, che fa leva sull’esistenza di “siti di conflitto” ben identificabili tra le due lingue in contatto (Poplack, 2018), non costituisca sempre un buon criterio per distinguere tra casi di prestito e casi di *code-switching* nel parlato bilingue in sinto lombardo.

6.2.2 La lingua dei modificatori

Data la scarsa utilità del criterio dell'integrazione morfologica per i nomi italoromanzi, l'analisi del continuum tra prestito e *code-switching* nel parlato bilingue in sinto lombardo può anche essere affrontata analizzando la lingua in cui sono realizzati i modificatori che li accompagnano. Come osserva Dal Negro (2005a: 170-171) in relazione alle alloglossie storiche in contatto con l'italoromanzo in Italia, "si potrebbe sostenere che la presenza di articolo romanzo rappresenti un passaggio, seppur breve, all'altro codice, mentre la presenza di articolo nella lingua di minoranza dovrebbe indicare una maggiore integrazione morfosintattica del nome, trattato dunque come prestito". In un lavoro successivo, Dal Negro (2009) mostra che l'integrazione variabile dei nomi romanzi nel parlato in walser issimese correla con la loro frequenza d'uso nella lingua minoritaria: i nomi romanzi tendono a ricevere determinanti walser, mentre i nomi romanzi meno frequenti tendono ad attivare attorno a sé un contesto linguistico romanzo, selezionando articoli romanzi. L'autrice conclude che i casi del primo tipo sono assimilabili a prestiti, mentre i casi del secondo tipo sono assimilabili a episodi di *code-switching*. Tuttavia, all'interno del medesimo studio, l'autrice non riscontra le stesse tendenze per i nomi italoromanzi in un corpus di arbëresh.

In questa sezione, analizziamo come le dinamiche delineate si manifestino nel contesto di contatto in analisi, analizzando la lingua in cui sono realizzati i modificatori dei nomi italoromanzi nel parlato in sinto lombardo. Ci concentriamo in particolare sui modificatori del sintagma nominale, distinti operativamente in determinanti (articoli e dimostrativi), indefiniti e numerali. I risultati relativi a queste tre tipologie di modificatori sono presentati nelle sezioni a seguire (§ 6.2.2.1 – 6.2.2.3), mentre il § 6.2.2.4 discute brevemente il caso dei sintagmi preposizionali.

6.2.2.1 Sequenze Determinante-Nome

La Figura 24 indica la lingua in cui, dato un nome italoromanzo o romaní nel corpus, vengono realizzati i determinanti che lo accompagnano.

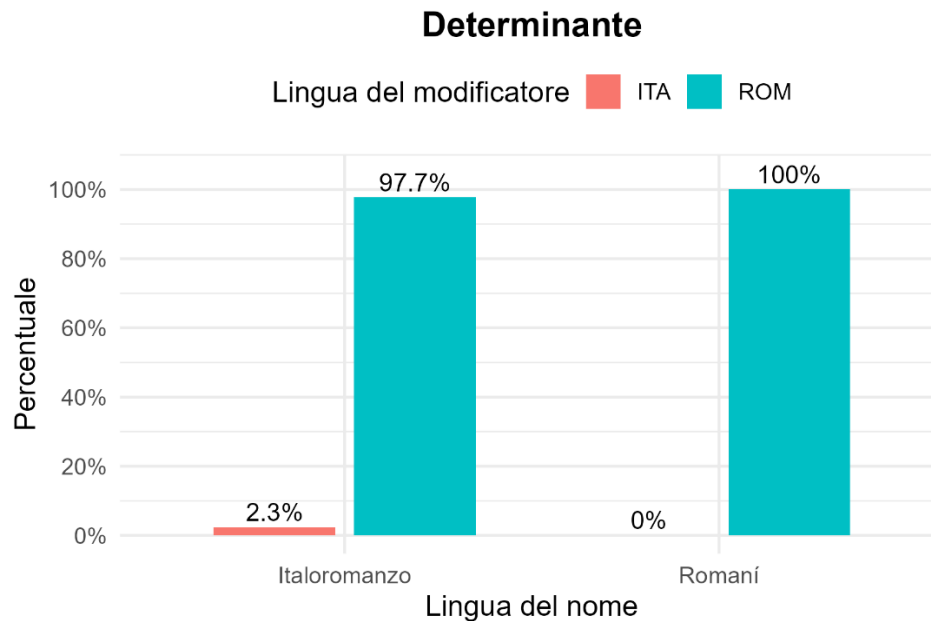


Figura 24: Distribuzione dei sintagmi nominali del corpus per lingua del nome e del determinante.

Come si può osservare dalla Figura 24, nel 97.7% dei casi i nomi italoromanzi sono preceduti da determinanti romaní, rispecchiando il comportamento dei nomi romaní, che sono preceduti da determinanti romaní nel 100% dei casi. La natura sostanzialmente categorica di questo *pattern* permette quindi di concludere che, in sinto lombardo, i nomi italoromanzi tendono a essere trattati come prestiti, ricevendo cioè determinanti romaní alla stregua delle loro controparti romaní, indipendentemente dalla loro frequenza nel corpus.

Tra le eccezioni che vedono l’inserzione di un nome italoromanzo con articolo italiano in contesti dominanti sinto lombardo (creando vere e proprie “isole” italoromanze, seguendo la terminologia di Myers-Scotton, 1993) possiamo innanzitutto individuare espressioni indicanti date o ricorrenze, come per esempio *il ventitré dicembre, il giorno di Pasqua, il tredici*; cfr. esempio (51).

- (51) **il tredici** dena u mandarini ku tine čave
 ‘il tredici danno i mandarini ai bambini’

La relativa regolarità con cui questo *pattern* si verifica sembrerebbe suggerire che, in espressioni di questo tipo, l’articolo tende a essere memorizzato nel lessico mentale dei parlanti insieme al nome ed essere dunque attivato insieme a esso come parte dell’unità più ampia. Questa tendenza è del resto stata ampiamente discussa nella letteratura sul contatto linguistico, sia su un piano teorico (Matras, 2020), sia in specifici studi empirici (cfr. p.es. Dal Negro, 2006: 95; Padure & Adamou, 2019), ed è stata ricondotta all’associazione di date e ricorrenze a contesti interazionali associati alla lingua maggioritaria.

Un secondo gruppo, numericamente più ridotto (circa una ventina di casi), comprende sintagmi nominali semplici (p.es. *la polizza, l'evangelo, la possibilità, il cenone, il risveglio*) che non sembrano essere ripresi dai parlanti come espressioni fisse. Inoltre, differentemente dai casi visti prima, questi sintagmi tendono a occorrere in posizione argomentale, subito prima o subito dopo il verbo in romaní; cfr. esempio (52).

- (52) in francia vjas pru velto **il risveglio**
 ‘in Francia è nato il risveglio’

È interessante notare che circa la metà di questi casi si concentra nel parlato di un singolo locutore (N), che, come vedremo al capitolo 7, è colui che produce il maggior numero di parole italoromanze nel campione e anche il maggior numero di alternanze.

Ciò sembrerebbe suggerire che la produzione di sintagmi nominali interamente italoromanzi (all’infuori di espressioni fisse) potrebbe correlare con una maggiore attivazione del bilinguismo a livello della conversazione (e nella mente del parlante bilingue). Alternativamente, si potrebbe ipotizzare che questo parlante abbia riprodotto più liberamente, nel contesto dell’interazione con il ricercatore, gli schemi bilingui prevalenti nella conversazione *in-group*, che prevedono una maggiore libertà combinatoria e dunque la possibilità di ricorrere più frequentemente a sequenze estese di articolo e nome in italiano¹.

6.2.2.2 Sequenze Indefinito-Nome

La Figura 25 mostra la lingua in cui, dato un nome italoromanzo o romaní nel corpus, vengono realizzati gli indefiniti che li accompagnano.

¹ Ringrazio Simone Ciccolone per questo suggerimento.

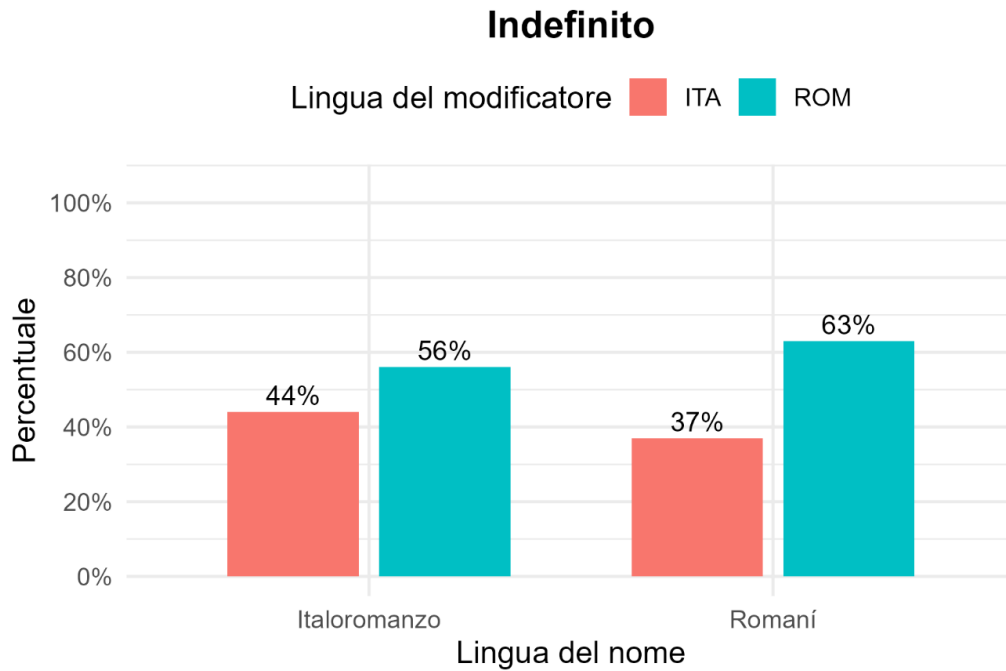


Figura 25: Distribuzione dei sintagmi nominali del corpus per lingua del nome e dell'indefinito.

Come si osserva, la percentuale di nomi italoromanzi preceduti da indefiniti romaní si riduce al 56%, ma questo tasso ricalca grossomodo quanto si osserva anche per i nomi romaní, preceduti da indefiniti romaní nel 63% dei casi. Questa distribuzione si può spiegare facilmente considerando che alcuni indefiniti (come *qualche*, *alcuno* e *ogni*) sono sprovvisti di equivalenti nativi, qualificandosi come prestiti nel sinto lombardo (cfr. Capitolo 5): come tali, essi possono combinarsi tanto con elementi romaní quanto con altri elementi italoromanzi. Si veda l'esempio in (53), in cui *qualche* si combina con il nome romaní *bers* 'anni', e (54), in cui *qualche* si combina invece con il nome italoromanzo *parola*.

- (53) **qualche bers** fa kardom=lo
 'qualche anno fa l'ho fatto'

- (54) **qualche parola** vela mange in gadžo
 'qualche parola mi viene in italiano'

Un esempio come quello in (54) mostra come il semplice apparire in sequenza di due elementi italoromanzi non implic necessariamente un passaggio all'italiano, quindi un caso di *code-switching*. Questa osservazione sembra dunque rendere necessaria una rilettura dell'opposizione tra prestito e *code-switching* intesa in termini strettamente lineari, cioè in termini di *token* consecutivi da una lingua

B in una frase altrimenti in una lingua A, una concezione tipica degli approcci variazionisti allo studio del parlato bilingue (cfr. Poplack, 2018).

6.2.2.3 Sequenze Numerale-Nome

Passiamo, infine, a considerare le combinazioni dei nomi con i numerali; cfr. Figura 26.

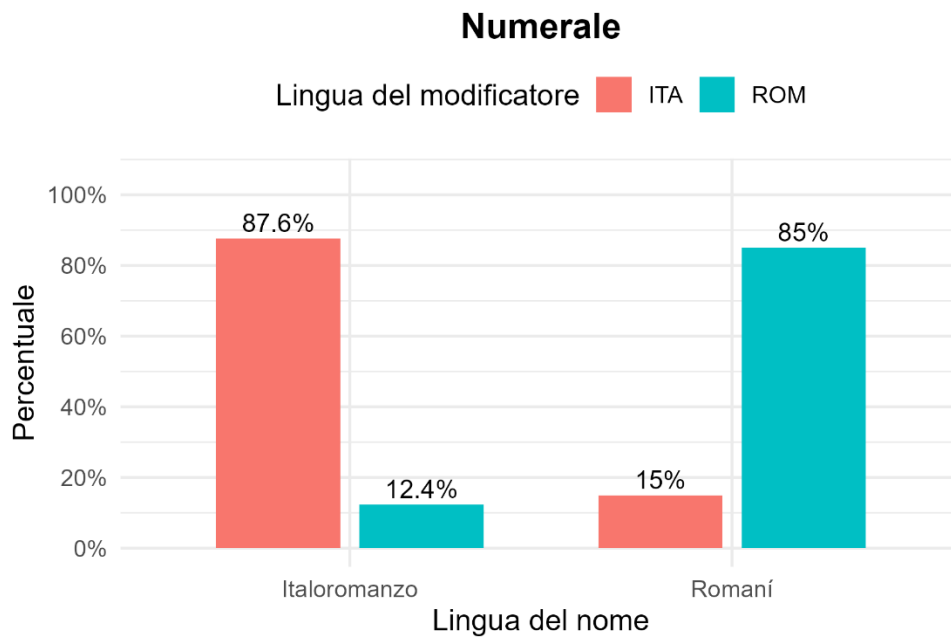


Figura 26: Distribuzione dei sintagmi nominali del corpus per lingua del nome e del numerale.

La Figura 26 mostra che i nomi italo-romanzi sono nella grande maggioranza dei casi preceduti da numerali italiani (87.6% dei casi), mentre i nomi romaní sono prevalentemente preceduti da numerali romaní (85% dei casi). Ciò sembra dunque indicare, nella terminologia di Matras (2020), uno scenario di “compartimentazione” linguistica, ossia un meccanismo per cui le unità linguistiche in una lingua mostrano distribuzioni diverse a seconda della loro origine etimologica.

Un’analisi più dettagliata dei dati rivela che i sintagmi interamente italiani includono sempre numerali superiori al sei, mentre i sintagmi interamente romaní includono sempre numerali inferiori al sei. Come già osservato in precedenza, in sinto lombardo i numerali fino al sei sono di origine romaní, mentre quelli superiori al sei sono prestati dall’italiano, nonostante i parlanti possano mantenere una conoscenza metalinguistica di alcune forme preromanze. Sulla base di questa osservazione, sembra dunque possibile ipotizzare che lo strato etimologico di appartenenza del numerale in sinto lombardo influenzi il contesto discorsivo, determinando la selezione della lingua del nome successivo.

Ciò è molto evidente nel caso delle costruzioni che combinano un numerale e le parole per ‘anno’, ‘mese’ e ‘giorno’, in quanto queste ultime sono disponibili sia in italiano sia in romaní

(rispettivamente *bers*, *monato*, *divas*): tale condizione permetterebbe in teoria ai parlanti di selezionare liberamente tra forme dell'una o dell'altra lingua. In questi contesti, i nomi 'anno', 'mese' e 'giorno' appaiono sempre in romaní quando il numerale è disponibile in romaní (dunque inferiore al 6), mentre appaiono sempre in italiano quando il numerale è disponibile soltanto in italiano (dunque superiore a 6). Combinazioni miste (ossia con numerale italiano superiore al sei e nome romaní, o numerale romaní inferiore al sei e nome italiano) non sono mai attestate. Questo tipo di compartimentazione è chiaramente illustrato nell'esempio in (55), in cui una parlante usa la forma italiana *mesi* (in luogo della forma romaní *monato*) con il numerale italiano *undici*, ma la forma romaní *bers* 'anni' con i numerali romaní *jek* 'uno', *duj* 'due' e *trin* 'tre'.

- (55) *kamaj=mi ja bers undici mesi duj bers trin pers / e dopu nasea*
 'ci amiamo un anno, undici mesi, due anni, tre anni... E dopo scappi'

Risulta interessante notare che tendenze molto simili sono state osservate in altre tre lingue di contatto geograficamente distanti: la domari in contatto con l'arabo parlata nei pressi di Gerusalemme (Matras, 2012); una varietà di berbero settentrionale in contatto con l'arabo parlato al confine tra Algeria e Marocco (Souag & Kherbach, 2016); e il tetun dili, una lingua veicolare austronesiana parlata a Timor Est in contatto con indonesiano e spagnolo (Williams-Van Klinken & Hajek, 2018). Nel caso della domari, ad esempio, i numerali superiori al cinque sono prestatati dall'arabo e selezionano stabilmente nomi arabi, laddove i numerali inferiori al cinque, ereditati, selezionano nomi domari. Mentre Souag & Kherbach (2016) riconducono questo schema a meccanismi di produzione *online*, Matras (2012) delinea per la domari uno scenario di "suppletivismo bilingue" a livello di sistema linguistico. In precedenti lavori (cfr. Fiorentini & Forlano, 2025), similmente a quanto sostenuto da Souag & Kherbach (2016), il *pattern* riscontrato nel sinto lombardo era stato motivato sulla base di processi di co-attivazione (o *triggering*; Clyne, 2003) a livello della produzione sincronica, tali per cui l'accesso a un numerale disponibile soltanto in italiano tende a innescare il nome successivo nella stessa lingua. Tuttavia, in linea con le osservazioni di Matras (2012), il fatto che questo schema sia estremamente stabile in sinto lombardo non appare in effetti completamente compatibile con un'interpretazione esclusivamente discorsiva e idiosincratICA del fenomeno.

Di fatto, il caso delle sequenze Numerale-Nome in sinto lombardo mostra perfettamente come i confini tra fenomeni legati al discorso (*code-switching*) e legati al sistema (prestito) nel parlato bilingue siano intrinsecamente sfumati: in sinto lombardo, la produzione di sintagmi interamente italoromanzi con numerali superiori al sei è infatti altamente stabile e prevedibile, una prerogativa del prestito; tuttavia, tale regolarità è probabilmente ottenuta grazie alla conoscenza attiva dei parlanti

dell'italiano, attraverso la quale essi possono selezionare svariate forme nominali italiane in luogo delle equivalenti romaní in funzione del numerale utilizzato.

6.2.2.4 Il caso dei sintagmi preposizionali: alcune osservazioni qualitative

Concludiamo questa sezione considerando brevemente le strategie di integrazione dei nomi itoloromanzi all'interno dei sintagmi preposizionali. In linea generale, si riscontrano in questo ambito tendenze piuttosto simili a quelle osservate per i sintagmi nominali. Tendono infatti a comparire stabilmente insieme alla preposizione itoloromanza sequenze relativamente fisse, ad esempio indicanti date ed età; cfr. esempi (56) e (57) di seguito.

- (56) tipo i *Ellison ljas rom a diciannove anni*
 'tipo la Ellison si è sposata a diciannove anni'

- (57) *vela vrin alle quattro e trenta quattro e quarantacinque kjake*
 'esce alle quattro e mezza, quattro e quarantacinque, così'

Va però notato che alcune delle preposizioni itoloromanze presenti in questi sintagmi, come *a* o *dopo*, tendono anche ad apparire in combinazione con nomi romaní (p.es. *a star bers* 'a quattro anni' o *dopu duj divas* 'dopo due giorni'), qualificandosi come prestiti stabiliti nel sinto lombardo (cfr. anche Capitolo 5). Pertanto, similmente a quanto visto con le sequenze Indefinito-Nome, nel parlato bilingue dei sinti lombardi la mera successione lineare di una preposizione e di un nome itoloromanzo non implica necessariamente un passaggio all'italiano.

Infine, è interessante notare che anche i toponimi, espressi quasi esclusivamente tramite nomi italiani, tendono sistematicamente a preferire la preposizione in italiano, come negli esempi (58) e (59).

- (58) però *akana i=la doj ki peskur čau a ivrea*
 'però adesso è lì da suo figlio a Ivrea'

- (59) *jon džan'in serbia džan'in albania*
 'loro vanno in Serbia, vanno in Albania'

Differentemente dal caso delle preposizioni indicanti date ed età, in sinto lombardo tendono a essere mantenute le preposizioni native indicanti stato in luogo (p.es. *indar, dendar*) e moto a luogo (*ke*), che vengono utilizzate stabilmente con nomi romaní. Il fatto che i toponimi richiedano invece

stabilmente preposizioni italoromanze sembra dunque indicare un processo di co-attivazione simile a quello che abbiamo riscontrato per i numerali. Al contrario dei numerali, la co-attivazione sembrerebbe in questo caso agire dal segmento che segue (il nome) al segmento che precede (la preposizione). Appare interessante notare che *pattern* molto simili in presenza di toponimi sono stati riportati in letteratura da Clyne (2003), sul parlato bilingue inglese-tedesco in Australia, e da Dal Negro (2024), in relazione al cimbro e al walser parlato in Italia. Clyne (2003) motiva questa tendenza osservando che i nomi di luogo sono di norma condivisi tra le due lingue e hanno per questo una maggiore probabilità di attivare ulteriore materiale di una lingua B in una frase condotta in una lingua A, ancorché spesso momentaneamente e senza attivare un effettivo *switch* di tipo alternante verso la lingua B. Alternativamente, Backus (1996) nota che le sequenze formate da nome di luogo e preposizione sono di norma altamente specifiche e per questo sono passibili di essere memorizzate come unità olistica nel lessico mentale dei parlanti.

6.3 Integrazione dei verbi

Nelle sezioni precedenti abbiamo analizzato nel dettaglio le strategie di integrazione dei nomi di origine italoromanza nel parlato in sinto lombardo. In questa sezione ci concentriamo invece sulle strategie di integrazione dei verbi.

In un recente studio, Meli (2023) ha osservato che i parlanti di sinto lombardo adottano strategie di “indirect insertion” (nella tipologia di Wohlgemuth, 2009; cfr. anche Capitolo 1) per l’integrazione dei verbi italoromanzi, una tendenza molto diffusa nelle varietà di romaní (cfr. Elšik, 2020; Ćirković & Mirić, 2022). L’autrice, in particolare, riporta la presenza di due strategie: la prima prevede l’integrazione della radice lessicale italoromanza nei paradigmi flessivi romaní attraverso il verbalizzatore di origine indoaria *-ar-* (p.es. *perdun-ar-ava* ‘perdono’); la seconda prevede l’impiego del verbalizzatore di origine indoaria *-o-*, derivato da un originario **-ov-* (p.es. *cuminci-o-la* ‘comincia’ < **cuminci-ov-ela*). I due morfemi erano originariamente derivazionali e sensibili alla valenza verbale, con *-ar-* usato per derivare verbi transitivi (causativi e fattitivi) e *-ov-* usato per derivare verbi intransitivi a partire da altre basi verbali o nominali (Meli, 2023). Tuttavia, tale funzione di derivazione da materiale nativo non appare oggi più produttiva, e l’unica funzione sincronica di questi morfemi sembra essere quella di adattare prestiti di origine italoromanza; in questo senso, *-ar-* e *-ov-* si confermerebbero a pieno titolo come “loan verb markers” (nella terminologia Wohlgemuth, 2009). Nonostante la tradizionale associazione di *-ar-* e *-o-* alla valenza del verbo, Meli (2023) osserva che questa distinzione non sembra (più) determinare pienamente la scelta del suffisso *-ar-* o *-o-* in sinto lombardo, in quanto *-ar-* si sta espandendo anche a verbi intransitivi a scapito di *-o-* (p.es. *tus-ar-ava* ‘tossisco’; *ibid.*: 429).

Nei paragrafi successivi analizziamo, a partire dal lavoro di Meli (2023), le strategie di integrazione dei verbi italoromanzi nel parlato in sinto lombardo. Come primo passo dell'analisi, la Figura 27 presenta una distinzione binaria tra forme verbali italoromanze integrate nella morfologia flessiva romaní e forme che non risultano integrate.

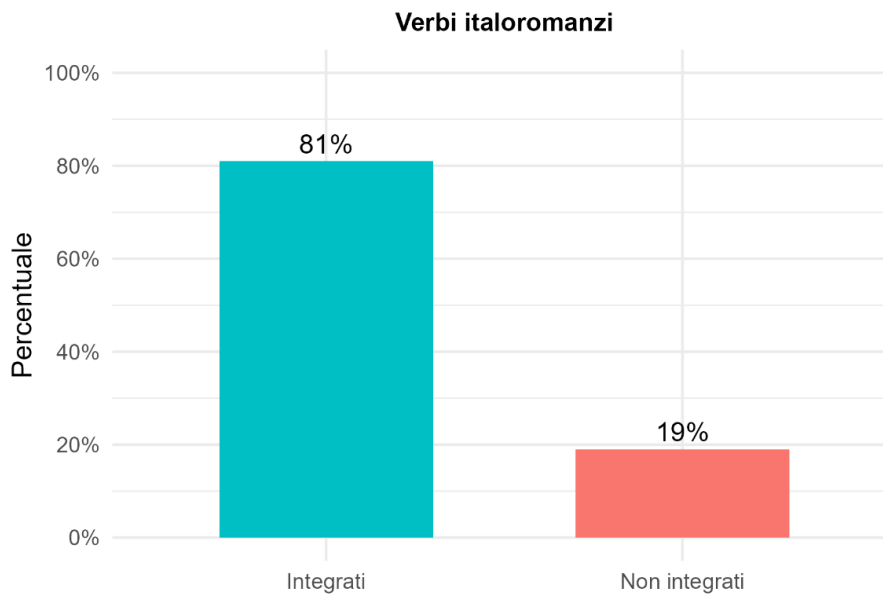


Figura 27: Distribuzione dei verbi italoromanzi morfologicamente integrati e non integrati nel corpus.

Come si nota, l'81% (N=526) dei verbi italoromanzi utilizzati all'interno di contesti discorsivi in sinto lombardo è integrato nei paradigmi flessivi di questa lingua, mentre il restante 19% (N=117) non risulta integrato. La presenza di verbi non integrati indica già una maggiore varietà di possibilità rispetto a quelle finora riportate in letteratura; l'analisi che segue considera pertanto queste due categorie separatamente.

6.3.1 Verbi integrati in sinto lombardo

Le forme verbali italoromanze morfologicamente integrate nel sinto lombardo ammontano a 526 *token*, riconducibili a 181 lemmi, la maggior parte dei quali (il 51%; N=92) si configura come *hapax*. Da un punto di vista strutturale, le singole forme flesse sono attestate nei modi finiti dell'indicativo e del congiuntivo, nei tempi del presente e del perfetto, e in tutte le persone.

Un'analisi più dettagliata delle singole occorrenze conferma in generale l'impiego dei due verbalizzatori *-ar-* e *-o-* per accomodare le radici verbali italoromanze nei paradigmi romaní (Meli, 2023). Tuttavia, sono necessarie alcune precisazioni, anche legate alla specifica forma del paradigma in cui questi verbi compaiono.

L'esempio in (60) mostra la radice del verbo italoromanzo *tentare* regolarmente integrata, attraverso il verbalizzatore *-ar-*, nella morfologia flessiva del sinto lombardo, rappresentata in questo caso dal morfema *-aja*, che segnala la prima persona plurale del presente indicativo.

- (60) *but volti men **tentaraja** ta rakaras u sinto*
 ‘molte volte noi cerchiamo di parlare il sinto’

L'esempio in (61) mostra che il verbo *convertirsi* è integrato a partire dalla forma radicale mediante lo stesso verbalizzatore nella morfologia flessiva romaní, segnalata dall'uscita in *-dam*, che indica la prima persona plurale del perfetto. Come si osserva in questo esempio, i verbi italoromanzi usati in discorsi in sinto lombardo possono anche selezionare clitici romaní (in questo caso il clitico riflessivo di prima persona plurale *mì*), ulteriore indizio della loro piena integrazione morfosintattica.

- (61) *perché men **cunvertardam**=(m)i, jam evangelici*
 ‘perché noi ci siamo convertiti, siamo evangelici’

L'esempio (62) mostra invece la radice del verbo italoromanzo *cambiare* integrata mediante *-o-* alla terza persona plurale del presente indicativo.

- (62) ***cambiona** pisal paroli*
 ‘cambiano poche parole’

Come nota Meli (2023), la forma originaria del verbo sarebbe **cambiovena*, che ha successivamente subito la caduta della consonante *[v]* e della vocale *[e]*. Alla luce di questi processi di semplificazione, attestati anche in altre varietà di romaní, lo statuto del morfema *-o* è oggetto di discussione nella letteratura specialistica, con alcuni studiosi che lo considerano ancora come morfema derivazionale, mentre altri lo considerano come distintivo di una nuova classe flessiva intransitiva in *-o* (cfr. Matras, 2002: 126). In questo quadro, Matras (2002) conclude che lo statuto del morfema *-o* come derivazionale o flessivo deve essere stabilito caso per caso nelle singole varietà.

Nel corpus in analisi emerge un ulteriore problema per la classificazione dello statuto di *-o-*. Nella quasi totalità dei casi, i verbi che seguono questa strategia di integrazione occorrono alla terza persona singolare o plurale del presente, in cui la vocale *-o* risulta sempre generalizzata a scapito della vocale *-e*, come nel caso di *cambiona* nell'esempio (62). Ciò non permette di verificare la stabilità di questo *pattern* nelle altre forme del paradigma al presente. L'unica occorrenza attestata in una persona

diversa dalla terza singolare o plurale è *rivoja* (**riv-ov-aja*), coniugata alla prima persona plurale del presente indicativo. Questo dato sembrerebbe suggerire che anche nelle altre forme del paradigma l'originaria vocale tematica venga eliminata a favore della generalizzazione di *-o-*. Tuttavia, sarebbero necessari più dati per confermare questa ipotesi.

Se ulteriori dati confermassero che *-o-* sia stato rianalizzato come morfema flessivo in sinto lombardo, risulterebbe problematico considerare, in sincronia, forme quali *cambiona* o *rivoja* come casi di *indirect insertion*, che presuppone per definizione il ricorso a morfemi derivazionali di accomodamento. Piuttosto, una forma come *cambiona* sembrerebbe istanziare un particolare sottotipo di *direct insertion*, che Wohlgemuth (2009: 91-92) definisce “direct insertion via verbal inflection class markers”. Con questa espressione l'autore si riferisce a casi in cui un verbo prestatato viene assegnato direttamente a una classe flessiva segnalata attraverso uno specifico marcatore, come ad esempio la vocale *-a-* in italiano che segnala la prima coniugazione (*-are*).

Un'interpretazione come *direct insertion* è invece l'unica possibile nelle forme del perfetto di questi verbi intransitivi. In tali forme scompare infatti qualsiasi traccia di un verbalizzatore e la radice italo-romanza si combina direttamente con la morfologia flessiva romaní, analogamente a quanto avviene per i verbi nativi. Si consideri, a tal proposito, l'esempio (63), in cui la radice italo-romanza *camb-* è immediatamente seguita dalla flessione *-jas*, indicante la terza persona singolare del perfetto indicativo.

- (63) *kana* ormai ***cambjas*** *u koa*
 ‘ora ormai la cosa è cambiata’

Indipendentemente dalla classificazione delle diverse strategie in ottica tipologica, i dati del corpus confermano le osservazioni di Meli (2023) secondo cui *-ar-* può essere impiegato sia con verbi transitivi sia con verbi intransitivi, mentre *-o-* rimane confinato ad alcuni verbi intransitivi. Nel corpus si osserva tuttavia un notevole grado di variazione, anche a livello di uno stesso verbo. Si considerino, a questo proposito, le coppie di esempi in (64) e (65), in cui il verbo *girare* è integrato dapprima mediante *-ar-* e successivamente mediante *-o-*, a parità di uso intransitivo.

- (64) *karena* *u tornei a tortona* poi magari a busto / ***girarena*** *pisal* *daparkrol*
 ‘fanno i tornei a Tortona, poi magari a Busto... Girano un po’ dappertutto’

- (65) *is jek* *du famigli più bari* che ***gironas*** *maskar u sinti*
 ‘era una delle famiglie più grandi che giravano tra i sinti’

Un'analisi più approfondita suggerisce che tale variazione non è casuale, ma sensibile a fattori semantico-sintattici che prescindono da una distinzione dicotomica tra verbi transitivi e intransitivi (Forlano, 2025). In particolare, sembra giocare un ruolo rilevante nella variazione osservata il grado di agentività del soggetto (cfr. Ćirković & Mirić, 2022 per osservazioni molto simili su una varietà gurbet di romani parlata nella Serbia orientale). Quando il soggetto è animato ed esercita un chiaro controllo sull'azione espressa dal verbo, il verbo tende a selezionare *-ar-*; viceversa, quando il soggetto è inanimato e ricopre il ruolo di tema o paziente, il verbo tende a selezionare *-o-*. Di conseguenza, la selezione del verbalizzatore *-ar-* è più frequente alla prima e seconda persona, intrinsecamente agentive, mentre la selezione di *-o-* appare più frequente alla terza persona, in cui è più frequente incontrare soggetti animati (Forlano, 2025). Ciò risulta particolarmente evidente negli esempi in (66)-(67) e (68)-(69), che coinvolgono, rispettivamente, i verbi italoromanzi *mancare* e *cambiare*, usati intransitivamente.

(66) *na **mancardam** mai da rispetto ke mengar genitori*
 ‘non abbiamo mai mancato di rispetto ai nostri genitori’

(67) *u tut na **mancolas** miga parché is u guramja*
 ‘il latte non mancava perché c'erano le mucche’

(68) *forse **cambiarava** da scuola*
 ‘forse cambio di scuola’

(69) ***cambiona** dola pisa paroli*
 ‘cambiano quelle poche parole’

In (66), *mancare* ha un soggetto chiaramente agentivo e animato (sebbene non espresso sintatticamente) e seleziona il verbalizzatore *-ar-*; viceversa, in (67), lo stesso verbo ha come soggetto un tema inanimato e seleziona *-o-*. Analogamente, in (68), il verbo *cambiare* seleziona *-ar-* in presenza di un soggetto agentivo e animato (sintatticamente non espresso), mentre in (69) seleziona *-o-* in presenza di un soggetto inanimato e paziente.

Tuttavia, il fatto che *-ar-* compaia nel corpus anche in presenza di soggetti non agentivi e inanimati, come nell'esempio in (70), sembrerebbe suggerire una sua progressiva affermazione come verbalizzatore *di default* in sinto lombardo (cfr. Meli, 2023).

- (70) dua **finarelas** u drom is ja rettangolo
 ‘dove finiva la strada c’era un rettangolo’

Indipendentemente dalle ragioni precise alla base della variazione osservata con verbi intransitivi (che necessiteranno di essere approfondite in studi futuri), è importante ribadire che le strategie di integrazione descritte rappresentano degli schemi particolarmente produttivi, come dimostra anche la loro applicazione a un numero molto elevato di *hapax*, come ad esempio la forma *bigiarava* ‘bigio’, costruita a partire dal lemma dell’italiano regionale lombardo *bigio* ‘marino la scuola’. Se sottoposte ai parlanti, queste forme idiosincratiche tendono a essere segnalate come forme “miste”, che emergono occasionalmente nel discorso soprattutto dei parlanti più giovani, sbilanciati verso l’italiano. Si tratta di esempi prototipici di “nonce borrowings” nella terminologia di Poplack (2018), ovvero forme che, pur occasionali, si presentano perfettamente integrate sul piano linguistico, al pari del materiale nativo

Da un punto di vista psicolinguistico, questi meccanismi di integrazione costituiscono delle strategie che permettono ai parlanti di mantenere una sorta di “ancoraggio” (Matras, 2020) alla romaní come lingua di base dell’interazione, attingendo al contempo a tutto il repertorio lessicale. Il ruolo strategico assegnato alla morfologia flessiva per ancorare la predicazione in romaní si può osservare chiaramente nell’esempio di seguito; cfr. (71).

- (71) urliamo buon anno (.) ehm, **urlaraja** buon anno
 ‘urliamo buon anno... ehm, urliamo buon anno’

Nell’estratto proposto, il parlante sta parlando in italiano, come si evince anche dall’uso del verbo italiano con morfologia flessiva italiana *urliamo*. Successivamente, rendendosi conto di dover parlare in sinto lombardo ai fini dell’intervista, il parlante ricorre a un rapido processo di trasformazione del verbo italiano *urliamo* nella forma *urlaraja* in romaní, lasciando invariato il resto della frase (si noti che una forma nativa del verbo *urlare*, *d- golde/i*, è attestata nel corpus). La lingua di base delle due frasi, altrimenti identiche, può dunque essere stabilita soltanto sulla base della morfologia flessiva del verbo finito. In altre parole, l’integrazione dei verbi italoromanzi appare un buon criterio per distinguere un discorso in sinto lombardo da un *code-switching* di tipo alternante verso l’italiano. Questa osservazione contribuisce anche a spiegare, nel corpus, la sostanziale assenza di verbi finiti che, occorrendo in contesti frasali sinto lombardo, non appaiono integrati nella sua morfologia. Questi casi verranno discussi più in dettaglio nella sezione successiva.

6.3.2 Verbi non integrati in sinto lombardo

Oltre ai verbi integrati nella morfologia flessiva del sinto lombardo, nel corpus si riscontrano anche alcuni casi di verbi che, pur inseriti in un discorso dominante in sinto lombardo, mantengono la morfologia flessiva italoromanza. Un'analisi più approfondita mostra che tali occorrenze possono essere ricondotte a due gruppi principali:

- (1) verbi non finiti, in particolare participi passati e gerundi;
- (2) verbi finiti che, come si vedrà, presentano alcune caratteristiche ricorrenti, che li collocano vicino al polo dei segnali discorsivi.

Le sezioni seguenti discutono le due categorie in maggiore dettaglio.

6.3.2.1 Verbi non finiti

L'inserzione di forme verbali non finite morfologicamente non adattate da una lingua B in un discorso in lingua A è un fenomeno ampiamente documentato in corpora bilingui (cfr. Muysken, 2000: 174–175; Myers-Scotton & Jake, 2014). Nel caso dei participi, tale comportamento è stato ricondotto alla natura delle marche di participio, considerate più prossime alla derivazione che alla flessione e dunque più facilmente attivabili insieme alla loro base lessicale, in modo analogo ad altri morfemi derivazionali come le marche di plurale (cfr. Backus, 2003).

In sinto lombardo, i participi passati italoromanzi tendono spesso a comparire all'interno di costruzioni perifrastiche miste, in combinazione con ausiliari romaní (cfr. esempi 72-75).

- (72) *men vaja **rigenerati***
'noi veniamo rigenerati'

- (73) *čas=lu **abandunado** di romli*
'è stato abbandonato dalla moglie'

- (74) *me jom **nata** kaj e čava kaj* tutta la vita
'io sono nata qui e sto qui tutta la vita'

- (75) *me n'i=ma mai **scritto** čĩ*
'io non ho mai scritto niente'

Come si può osservare, tali costruzioni perifrastiche svolgono funzione passiva (72-73) oppure di passato analitico (74-75). In alcuni casi, sembra plausibile ipotizzare che l'intera costruzione perifrastica in cui il participio italomanzano si inserisce costituisca un calco strutturale di modelli italomanzani. Nell'esempio (73), ad esempio, il verbo romaní č- 'stare' compare al perfetto con funzione di ausiliare in una costruzione passiva, seguito dal participio italomanzano (*čas=lu abandonado*). In questo contesto, č- 'stare' sembra replicare funzionalmente l'italiano *è stato*, forma che ricorrerebbe nella medesima costruzione in italiano, benché in italiano *è stato* costituisca il passato prossimo di *essere* e non di *stare*. A mia conoscenza, costruzioni passive che al passato impiegano il verbo č- 'stare' in funzione di ausiliare non sono attestate in altre varietà di romaní, un elemento che rafforza l'ipotesi di uno sviluppo dovuto al contatto con l'italiano. Analogamente, nell'esempio (75), il participio passato *scritto* compare in combinazione con la forma *i=ma* 'ho', dando luogo a una costruzione di passato analitico (*i=ma scritto* 'ho scritto'). Poiché la romaní non dispone di un passato analitico, anche questa configurazione può dunque essere interpretata come un possibile calco dall'italiano.

Infine, l'esempio (76) mostra l'inserzione della forma del gerundio *essendo* in un contesto dominante romaní.

- (76) *ma dik men essendo cristiani evangelici na festegiaraj=lu miga u nadal*
 'ma guarda, noi, essendo cristiani evangelici, non lo festeggiamo il Natale'

Come osservato da Matras (2002: 160), le varietà sinte di romaní sono sprovviste di un modo equivalente al gerundio, per la realizzazione delle cui funzioni è necessario ricorrere a strutture perifrastiche con verbo finito. L'assenza di una categoria grammaticale corrispondente può dunque costituire una condizione favorevole per l'adozione delle forme italiane di gerundio non adattate.

6.3.2.2 Verbi finiti

Il secondo gruppo di verbi italomanzani inseriti morfologicamente intatti in un discorso in sinto lombardo comprende una serie di forme che, in italiano, hanno sviluppato funzioni in parte o del tutto indipendenti dai loro paradigmi primari, avvicinandosi alla categoria eterogenea dei segnali discorsivi (cfr. Ghezzi & Molinelli, 2014; Lo Baido, 2024). La Tabella 15 riporta le forme più frequenti a partire dai lemmi di riferimento.

Lemma	Type
Dare	Può darsi (che) (x3)
Dipendere	Dipende (x53)
Dire	Diciamo (che) (x8), per dire (x15)
Guardare	Guarda che (x2)
Pensare	Pensa (x1), penso (x3)
Sapere	Mi sa (che) (x2), si sa (x2), sai (che) (x1), so (x2)

Tabella 15: Distribuzione per lemma e type dei principali verbi finiti non integrati in sinto lombardo.

Come si osserva, si tratta prevalentemente di verbi di cognizione, percezione o legati agli atti del dire e del parlare. Sebbene il numero di lemmi e di *type* coinvolti sia relativamente limitato, il numero di *token* risulta invece piuttosto elevato, come nel caso del verbo *dipendere*, che occorre per 53 volte nella forma di terza singolare *dipende*. Ciò differenzia queste forme, su un piano distribuzionale, dai verbi finiti morfologicamente integrati, che si riconducono a un numero elevato di lemmi a bassa frequenza.

Si riportano di seguito alcuni esempi in contesto con le forme verbali *dipende* (77), *sai* (78), *pensa* (79), *diciamo (che)* (80) e *mi sa (che)* (81).

(77) *činej=la ja toco o mukej=la anca intrega (.) dipende ar kamea kares=la*
 ‘la tagli un pezzo o la mangi anche intera, dipende come la vuoi fare’

(78) *sai krol u predicatori d’italia sena=pe una volta al mese*
 ‘sai, tutti i predicatori d’Italia si incontrano una volta al mese’

(79) *n’is pura ta suen inketane pensa*
 ‘non potevano dormire insieme, pensa’

(80) *paple akana però diciamo che akana vjen anca pisa più gosare*
 ‘ancora adesso, però diciamo che adesso sono anche diventati più furbi’

(81) *mi sa che i dapal mende u sinti*
 ‘mi sa che dietro di noi ci sono i sinti’

Le diverse forme esemplificate sembrano mostrare diversi livelli di autonomia semantica rispetto al loro significato primario e di autonomia sintattica rispetto al resto dell'enunciato, segnalando diversi livelli di grammaticalizzazione. Il grado di grammaticalizzazione appare molto avanzato nel caso di *sai* e *pensa* (esempi 78-79), che, nelle specifiche forme in cui compaiono, hanno sviluppato un significato altamente autonomo rispetto al resto del paradigma ed esercitano uno *scope* sull'intero enunciato che precede o segue, senza attivare una reale struttura argomentale. Tale processo appare invece meno avanzato negli esempi in (77), (80) e (81), in cui i verbi *dipende*, *diciamo* e *mi sa* attivano una struttura argomentale. Significativamente, i verbi *dipende* e *diciamo* sono seguiti dal complementatore di origine italo-romanza *che*, con cui potrebbero costituire un'unità fissa.

Risulta interessante notare che la forma *dipende*, a differenza delle altre, che compaiono sempre nella loro forma integrata, può trovarsi in variazione con la forma integrata *dipendola*. Tuttavia, a fronte delle 53 occorrenze della forma *dipende*, la variante morfologicamente integrata è attestata soltanto due volte, peraltro nel parlato di un medesimo parlante, che in un caso vi ricorre come autocorrezione dopo aver prodotto la forma non integrata; cfr. (82).

- (82) *na dipende du dat / dipendola du dat*
 ‘no, dipende dal padre, dipende dal padre’

In sintesi, per i casi discussi sembra plausibile ipotizzare che queste forme siano memorizzate dai parlanti come unità non analizzate, attivabili come sequenze “pronte all'uso” indipendentemente dal contesto linguistico dell'interazione (italiano o sinto lombardo). In questo senso, tali forme si differenziano dai verbi finiti prototipici, che mostrano regolarmente integrazione nella lingua ricevente al fine di ancorare nel sinto lombardo la lingua di base della predicazione.

In questo quadro, i verbi prototipicamente referenziali usati in un contesto dominante sinto lombardo, ma non integrati nella sua morfologia flessiva, risultano estremamente rari. Quando compaiono, sono spesso separati dal resto dell'enunciato attraverso cesure prosodiche marcate, oppure presentano un rapporto di dipendenza sintattica ambiguo, come nell'esempio (84).

- (84) *tumen / entrate*, comportatevi da italiani
 ‘voi... entrate, comportatevi da italiani’

In questo caso, il pronome romaní al caso nominativo *tumen* è seguito dal verbo italiano con morfologia italiana *entrate*, una configurazione che rappresenta un'apparente eccezione alle previsioni del *Matrix Language Frame Model* (Myers-Scotton, 1993), secondo cui gli *outsider late*

system morphemes (cioè le marche di caso e di tempo-aspetto) dovrebbero provenire da un'unica lingua. Tuttavia, a un'analisi più attenta, notiamo che *tumen* ed *entrate* sono separati da una pausa prosodica nettamente percepibile; inoltre, si può osservare che la selezione del verbo italiano morfologicamente intatto innesca un passaggio più esteso all'italiano (*comportatevi da italiani*).

6.4 Discussione

In questo capitolo abbiamo analizzato le strategie di integrazione delle inserzioni italoromanze nel parlato in sinto lombardo, considerando dapprima il livello fonetico/fonologico e successivamente quello morfologico e morfosintattico, con particolare attenzione a nomi e verbi.

In linea con la letteratura sull'argomento (Poplack, 2018; Poplack et al., 2020; Adamou, 2016; Matras, 2020), l'analisi esplorativa delle strategie di integrazione a livello fonetico/fonologico ha mostrato che queste tendono a essere variabili a livello di uno stesso lemma e persino a livello del singolo individuo. Quanto invece all'integrazione morfologica, essa si mantiene più stabile a livello della comunità. Tuttavia, è stato notato come la possibilità stessa di osservare l'integrazione dipenda dal grado di somiglianza strutturale tra le due lingue a livello di specifiche classi di parole.

In particolare, nel caso dei nomi l'analisi ha mostrato che l'integrazione morfologica è spesso impossibile da stabilire a causa di un forte isomorfismo tra i paradigmi nominali del sinto lombardo e dell'italiano, come già osservato da Meli (2023). In tal senso, soprattutto per le forme più recenti di origine italiana, prestito e *code-switching* tendono a coincidere largamente per i nomi nell'uso. L'analisi della lingua con cui sono realizzati i modificatori (Dal Negro, 2005) ha invece mostrato che i nomi italoromanzi, quando utilizzati in un discorso in sinto lombardo, ricevono quasi categoricamente i determinanti del sinto lombardo, venendo dunque trattati come prestiti dai parlanti (sulla base del criterio dell'integrazione). Questo risultato è in linea con le tendenze interlinguistiche più generali, che osservano come i nomi alloglotti tendano a ricevere i determinanti della lingua dominante della frase nel parlato bilingue (Myers-Scotton, 2002; Torres Cacoullos et al., 2022). Infine, si è potuto osservare un interessante fenomeno di compartimentazione per quanto riguarda i numerali, tale per cui i numerali superiori al sei, disponibili soltanto in italiano, tendono a co-attivare sistematicamente la parola successiva in italiano, anche laddove questa sarebbe disponibile in sinto lombardo.

Quanto ai verbi, l'analisi ha mostrato come questi, a differenza dei nomi, tendano a essere sottoposti a strategie di integrazione particolarmente sistematiche e vistose, tramite schemi di adattamento molto produttivi (Heath, 1984). I casi di verbi finiti non integrati risultano invece limitati e appaiono spesso associati a usi pragmatico-discorsivi (forme come *dipende*, *guarda*, *mi sa*), interpretabili come sequenze ad alta disponibilità e potenzialmente memorizzate come unità. In un certo senso, si

potrebbe sostenere che sia proprio il parlato bilingue a fungere da “reagente” (Dal Negro, 2015: 96) per distinguere tra diversi meccanismi di processazione e produzione delle unità linguistiche, rendendo queste differenze molto più visibili rispetto che nel parlato monolingue.

La lingua in cui è realizzata la morfologia flessiva dei verbi è stata considerata centrale in numerosi approcci allo studio del parlato bilingue, che vedono riflessa in essa la scelta della lingua in cui i parlanti desiderano ancorare la loro predicazione (su tutti, cfr. Myers-Scotton, 1993, 2002; Matras, 2020). Anche nel caso del sinto lombardo, la morfologia flessiva del verbo appare come un eccellente indicatore per stabilire la lingua di base della frase; detto altrimenti, si può sostenere che il sinto lombardo rappresenti la “lingua matrice” delle frasi bilingui.

Infine, malgrado la morfologia grammaticale possa offrire interessanti osservazioni sullo studio del parlato bilingue, i risultati del capitolo mostrano anche come questa non sia sempre un criterio affidabile per distinguere tra prestito e *code-switching*. Due fattori appaiono decisivi in questo senso. Da un lato, un crescente isomorfismo strutturale tra i sistemi (favorito sia da fattori legati al contatto, come l’erosione del sistema dei casi, sia da fattori interni, come la perdita del valore distintivo di alcuni fonemi) riduce i “siti di conflitto” in cui l’integrazione sarebbe osservabile in modo netto (cfr. Dal Negro, 2005 per riflessioni molto simili sul parlato bilingue in altre alloglossie di area italo-romanza). Dall’altro, la presenza di fenomeni di co-attivazione e di ripresa di sequenze relativamente fisse o semifisse (numerale+nome, preposizione+toponimo, formule temporali, alcuni verbi usati come segnali discorsivi) mostra che l’emergere di sequenze italo-romanze superiori al singolo morfema non implica necessariamente un cambio discreto di codice. Questa osservazione contrasta con gli approcci di stampo variazionista all’analisi del parlato bilingue (Polack, 2018), che vedono nell’estensione lineare in termini di *token* dell’elemento alloglotto il migliore criterio per distinguere tra prestito e *code-switching* nel parlato bilingue.

Capitolo 7. Variazione tra parlanti

Nei capitoli precedenti, abbiamo analizzato l'incidenza quantitativa e le caratteristiche qualitative di svariati schemi combinatori nel parlato bilingue dei sinti lombardi. I dati presentati, tuttavia, rappresentano delle medie calcolate sull'insieme delle produzioni di tutti i parlanti: ciò significa che il grado di variazione individuale o per gruppi può risultare piuttosto elevato (Adamou, 2016).

Questo capitolo adotta un approccio variazionista all'analisi del parlato bilingue dei sinti lombardi, con l'obiettivo di individuare la variazione degli schemi combinatori sia a livello individuale, sia all'interno di specifici gruppi di parlanti. Questi ultimi vengono definiti sulla base di tre variabili sociolinguistiche: il genere, la fascia d'età e l'area di residenza dei parlanti.

Come osserva Adamou (2016), in piccole comunità bilingui caratterizzate da reti sociali “dense” e “transitive” (Milroy, 2002), la variazione individuale del parlato bilingue tende a essere relativamente contenuta, in quanto i singoli locutori riproducono modelli stabiliti a livello comunitario (cfr. anche Adamou et al., 2016). Similmente, Nagy (2011) e Meyerhoff (2014) non trovano evidenze di un aumento significativo nell'uso di prestiti nelle generazioni più giovani in due diverse comunità bilingui (rispettivamente italiano-faetano in Italia e nkep-bislama nelle isole Vanuatu), nonostante la diffusa percezione tra i parlanti di un sempre maggiore deterioramento strutturale della loro lingua per l'influsso lessicale della lingua dominante. Per testare tali osservazioni sui dati di sinto lombardo, in questo capitolo la variazione negli schemi di parlato bilingue nel corpus è analizzata sia mediante *boxplot*, che consentono di visualizzare la variazione individuale e di identificare possibili *outlier*, sia tramite test ANOVA a una direzione, utili a rilevare differenze significative tra gruppi di parlanti. Poiché il corpus su cui si basa l'analisi è piuttosto limitato, questi risultati dovranno essere intesi come esplorativi.

Più nello specifico, il § 7.1 analizza la variazione rispetto alla produzione di parole italoromanze e romaní; il § 7.2 si concentra sulla variazione nel ricorso a inserzioni e alternanze (cfr. Capitolo 4); il § 7.3 esamina più dettagliatamente la variazione nella produzione di nomi e di verbi italoromanzi; infine, il § 7.4 propone una discussione dei risultati.

7.1 Variazione nella produzione di parole italoromanze e romaní

Nel capitolo precedente, abbiamo notato che il 42.3% delle parole nel corpus è di origine italoromanza, un dato particolarmente elevato anche se considerato in ottica comparativa con altri corpora bi- e plurilingui (Adamou, 2016).

Un'analisi più approfondita di questo dato mostra un certo grado di variazione individuale: nei loro discorsi, i singoli locutori producono un tasso di parole italoromanze che oscilla dal 30.2% al 57.8%, e un tasso di parole romaní compreso tra il 40.8% e il 67%. Tuttavia, una rappresentazione grafica tramite *boxplot*, riportata nella Figura 28, mostra l'assenza di *outlier* per le tre categorie linguistiche “ITA”, “ROM” e “ALTRO”, suggerendo che, pur in presenza di differenze individuali, nessun parlante si discosta in modo estremo dalla distribuzione complessiva.

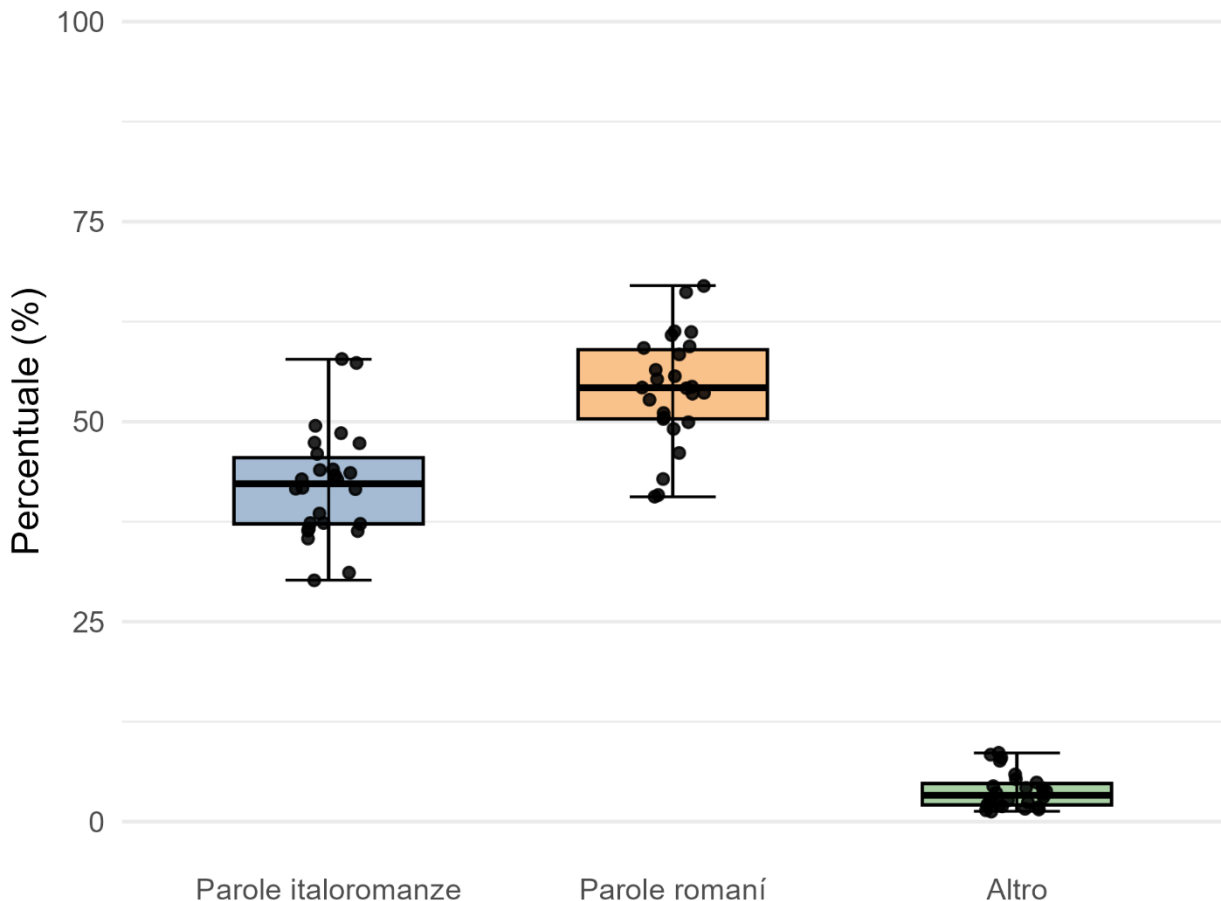


Figura 28: Variazione nella produzione di parole italoromanze, romaní e ‘altro’ per locutore.

I grafici a dispersione nelle Figure 29-31 permettono di visualizzare più in dettaglio le relazioni tra il tasso di parole italoromanze nei discorsi dei singoli parlanti e caratteristiche socioanagrafiche quali genere, età e insediamento di residenza.

Variazione tra parlanti

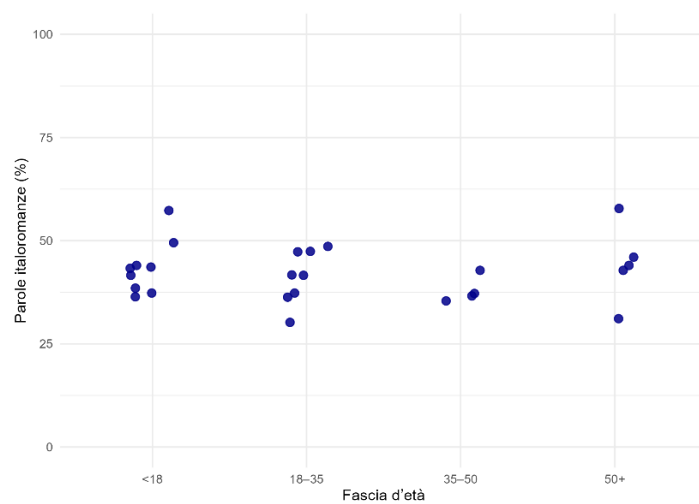


Figura 29: Variazione del tasso di parole italo-romanze per fascia d'età.

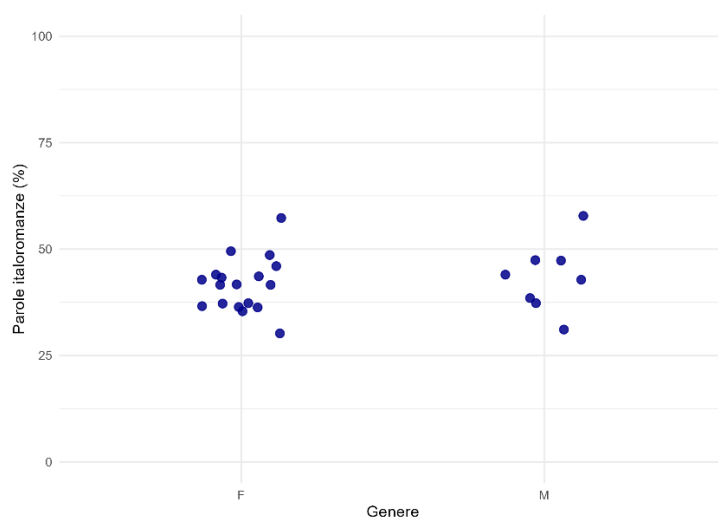


Figura 30: Distribuzione del tasso di parole italo-romanze per genere.

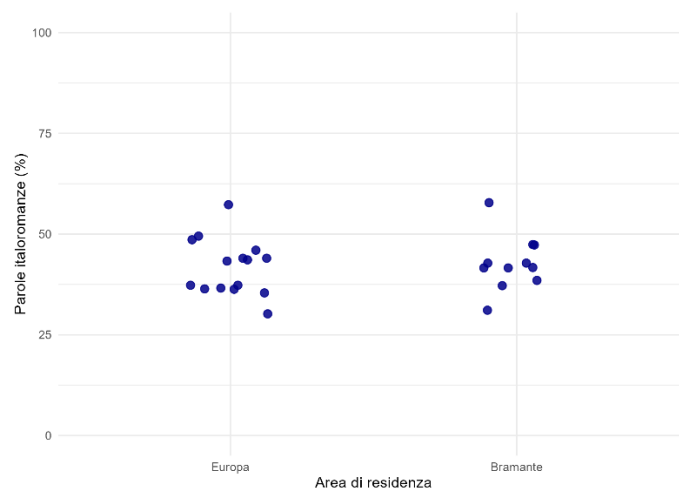


Figura 31: Distribuzione del tasso di parole italo-romanze per area di residenza.

Dalla visualizzazione di questi dati non sembrano emergere differenze chiare tra i diversi sottogruppi. Pertanto, al fine di comprendere più dettagliatamente questa variazione, è stato condotto un test ANOVA. Seguendo Adamou et al. (2016), il rapporto tra le parole italoromanze e le parole romaní nei discorsi dei singoli parlanti è stato considerato la variabile dipendente, mentre sono state considerate come variabili indipendenti il genere, l'età e l'insediamento di residenza. Il test ANOVA non ha rilevato differenze significative tra i gruppi considerati (Età: $p = 0.675$; Genere: $p = 0.545$; Insediamento: $p = 0.801$). Di conseguenza, questi fattori non sembrano governare la variazione individuata.

Infine, ci si potrebbe chiedere in quale misura l'incidenza di parole italoromanze nelle produzioni dei singoli parlanti possa meramente dipendere dalla quantità di parlato prodotto. Come visto al Capitolo 3, infatti, i singoli parlanti mostrano notevoli differenze nella quantità di *token* prodotti, con alcuni parlanti che hanno prodotto meno di 1.000 parole e altri che ne hanno prodotte più di 3.000. A questo proposito, la Figura 32 mostra la variazione del tasso delle parole italoromanze in funzione del numero di *token* complessivi prodotti da ogni parlante.



Figura 32: Relazione tra tasso di parole italoromanze e parole totali per parlante.

La linea di tendenza riportata nella Figura 32 riassume l'andamento complessivo dei dati, mostrando come la percentuale di parole italoromanze vari in media all'aumentare del numero di parole prodotte. Come si osserva, tale linea, pur mostrando un lieve incremento all'aumentare dell'estensione nelle singole produzioni, rimane relativamente stabile, indicando l'assenza di una correlazione forte tra tasso di parole italoromanze e quantità di *token* complessiva prodotta dai singoli individui.

7.2 Variazione nella produzione di inserzioni e alternanze

Nel precedente paragrafo, abbiamo analizzato la variazione tra parlanti relativamente alla produzione di parole italoromanze, mostrando l'assenza di correlazioni significative tra il tasso di parole italoromanze nel parlato dei singoli locutori e i fattori socioanagrafici considerati, al netto di una certa variazione individuale. Nella categoria "parole italoromanze" erano incluse sia le inserzioni di singole parole e sintagmi, sia casi di *code-switching* alternante verso l'italiano. Tuttavia, come abbiamo avuto modo di osservare al Capitolo 4, si tratta di due manifestazioni di contatto piuttosto diverse tra loro, sia in termini qualitativi che in termini quantitativi. In questo paragrafo, analizziamo dunque più in dettaglio la variazione tra parlanti rispetto alla produzione di inserzioni e alternanze.

A questo proposito, il *boxplot* nella Figura 33 mostra la variazione individuale nella produzione di inserzioni e alternanze nel parlato dei singoli locutori.

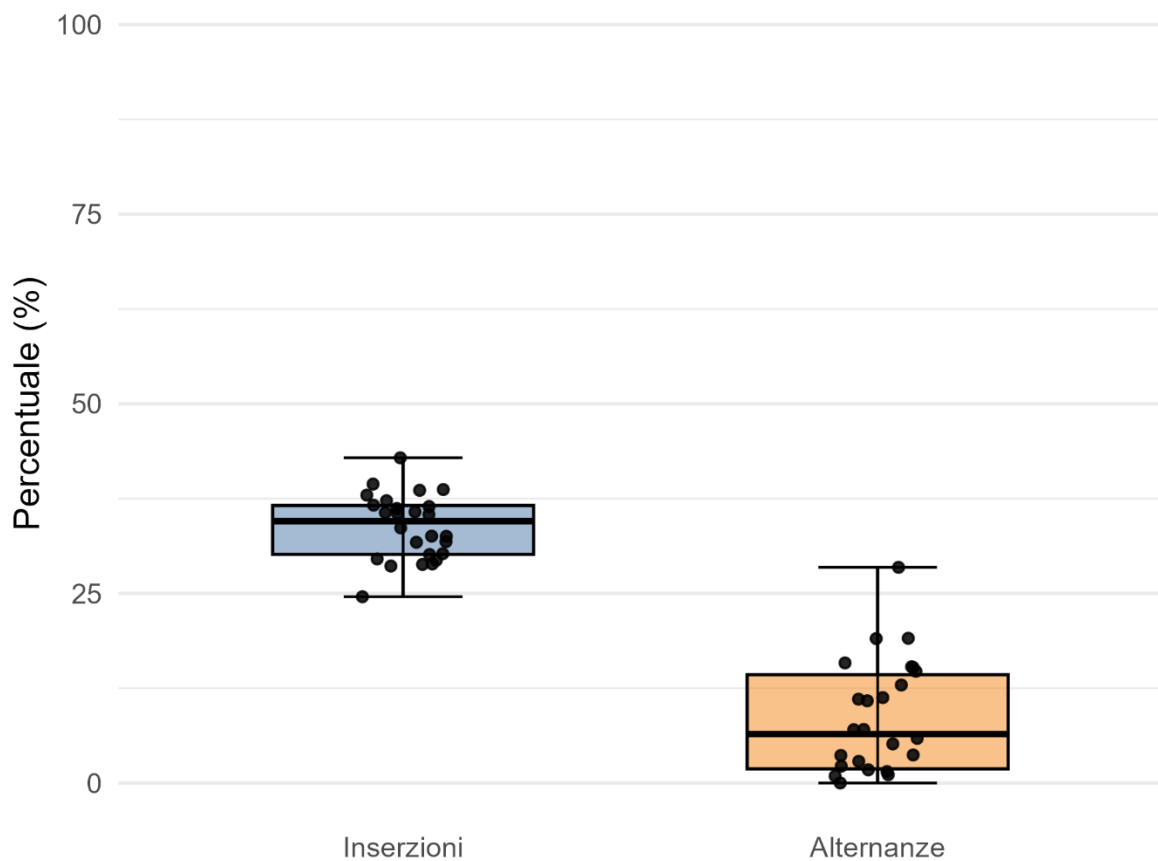


Figura 33: Variazione nella produzione di inserzioni e alternanze per locutore.

Come si osserva, la variazione nella produzione di inserzioni italoromanze risulta relativamente limitata, oscillando tra circa il 25% a circa il 40%. Viceversa, il ricorso ad alternanze mostra maggiore

variabilità: alcuni parlanti non ne fanno mai ricorso, mentre per altri esse giungono a coprire fino al 25% delle loro produzioni. Per entrambe le categorie, si rileva comunque l'assenza di *outlier*.

Questi dati permettono inoltre di connettere la variazione nella produzione di parole italoromanze osservata nella sezione precedente a un maggiore o minore ricorso ad alternanze. Per esempio, C., la parlante che produce il minor numero di parole italoromanze in generale (circa 30%) non ricorre mai ad alternanze nel suo discorso; viceversa, N., il parlante che produce il maggior numero di parole italoromanze in generale (circa 58%), ricorre ampiamente ad alternanze verso l'italiano (che costituiscono circa il 25% del suo *output* verbale).

A questo proposito, può essere interessante valutare se sussista una correlazione tra il tasso di inserzioni e il tasso di alternanze nel parlato dei singoli locutori; cfr. Figura 34.

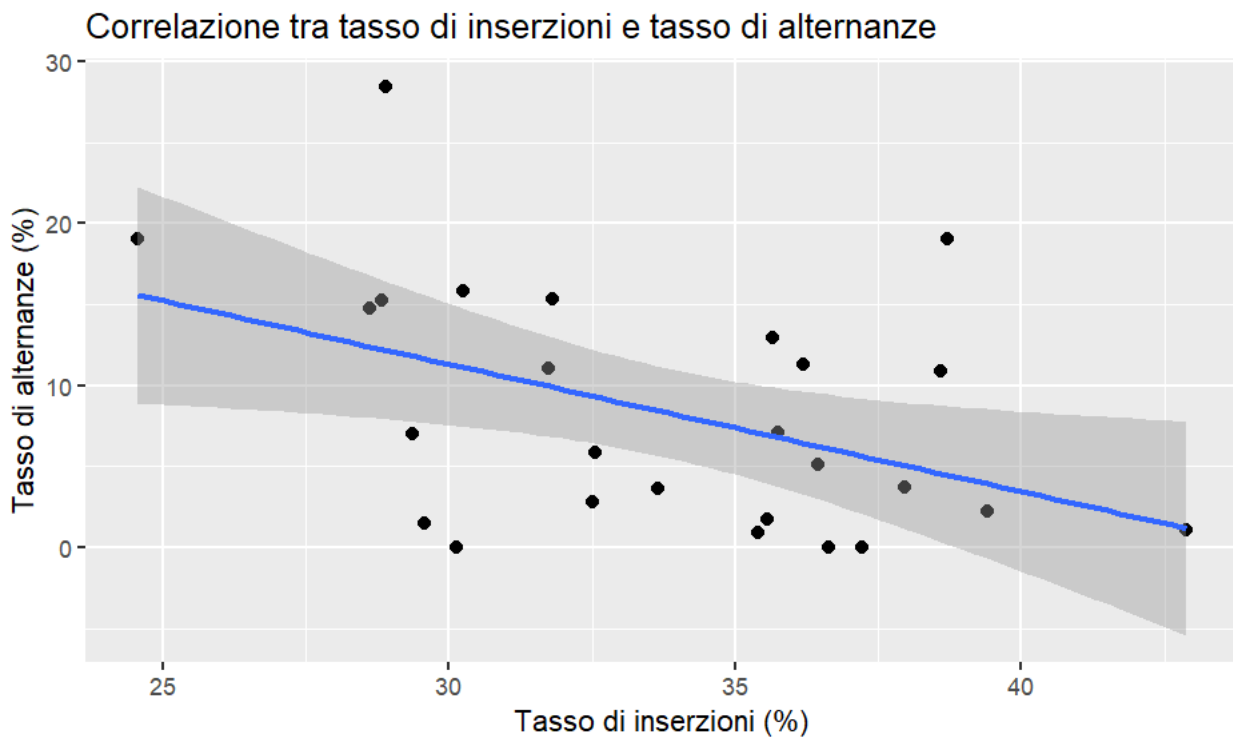


Figura 34: Correlazione tra tasso di inserzioni e tasso di alternanze nel parlato dei singoli locutori.

Come si osserva dalla Figura, al crescere del tasso di inserzioni, il tasso di alternanze appare generalmente più basso, come evidenziato dalla linea di tendenza, tracciata in blu. Un test di correlazione di Spearman conferma una correlazione negativa statisticamente significativa tra tasso di inserzioni e tasso di alternanze nel parlato dei singoli locutori ($\rho = -0.39$, $p = 0.047$). Questo risultato suggerisce che i parlanti che ricorrono maggiormente alle inserzioni tendono a ricorrere meno alle alternanze, indicando una possibile relazione inversa tra le due strategie di parlato bilingue.

Variazione tra parlanti

Dal momento che le inserzioni rappresentano la strategia di parlato bilingue dominante, ci focalizziamo ora più in dettaglio su di esse. Le Figure 35, 36 e 37 mostrano dunque la variazione nella produzione di inserzioni italoromanze per gruppi, sulla base di età, genere e insediamento di residenza dei parlanti.

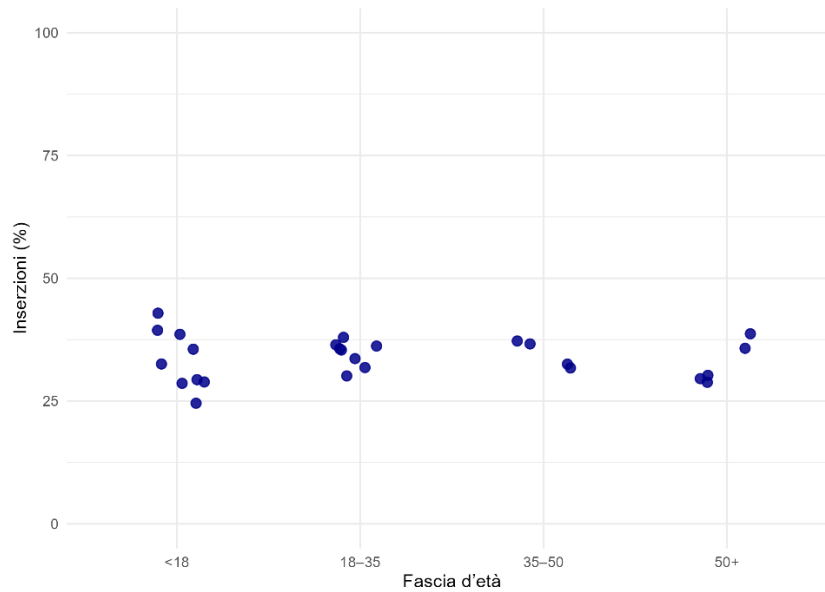


Figura 35: Distribuzione del tasso di inserzioni per fascia d'età.

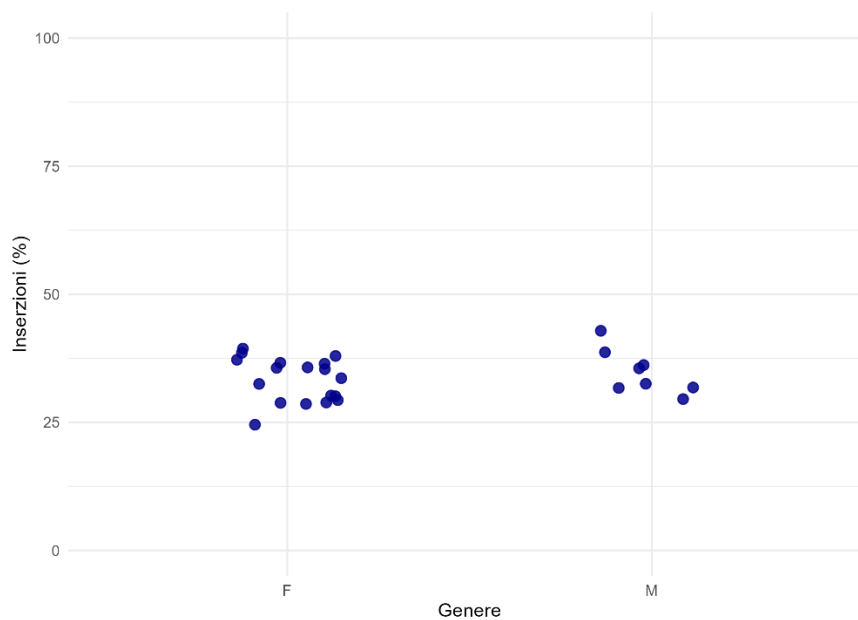


Figura 36: Distribuzione del tasso di inserzioni per genere.

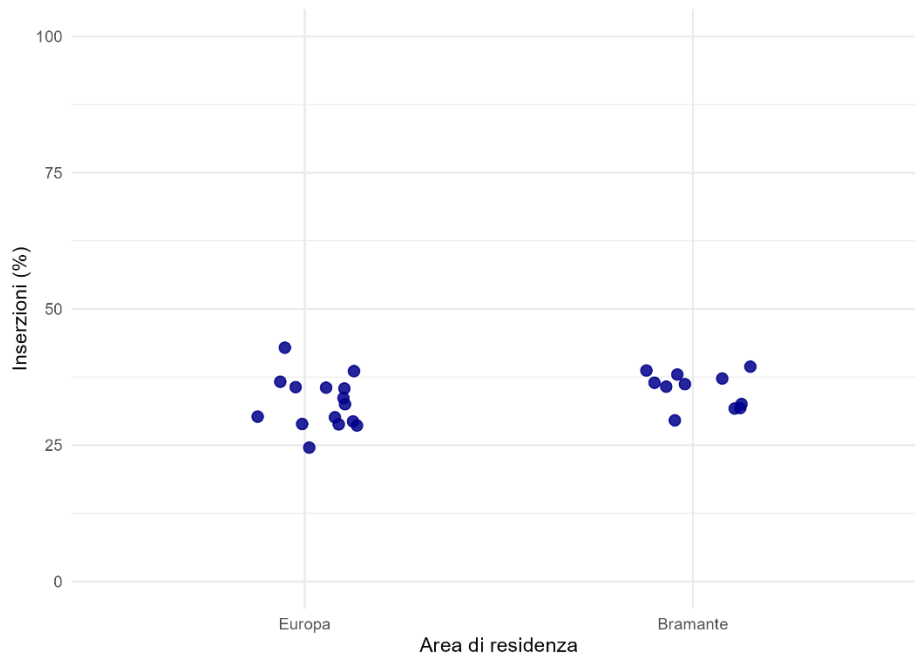


Figura 37: Distribuzione del tasso di inserzioni per insediamento.

Come si può osservare nelle Figure, i tassi di inserzioni appaiono molto stabili attraverso i vari gruppi sociali. Un test ANOVA (considerando il rapporto tra le inserzioni italo-romanze e il totale delle parole per locutore come variabile dipendente, ed età, genere e insediamento di residenza come variabile indipendente) mostra l'assenza di effetti significativi per età ($p = 0.953$), genere ($p = 0.357$) e insediamento ($p = 0.104$).

7.3 Variazione nella produzione di nomi e verbi italo-romanzi

Nel paragrafo precedente, abbiamo analizzato la variazione tra parlanti nella produzione di inserzioni italo-romanze e di alternanze verso l'italiano nel parlato in sinto lombardo. In questa sezione, ci concentriamo più in dettaglio sulle inserzioni, esaminando in particolare la variazione nel ricorso a nomi e verbi italo-romanzi. Come osservato da Adamou et al. (2016), poiché l'occorrenza di un nome e di un verbo di contatto tende a essere legata all'argomento di conversazione, si può prevedere una maggiore variazione per queste categorie lessicali sia rispetto alle "parole di contatto" in generale, sia rispetto alle inserzioni che comprendono tutte le parti del discorso.

A questo proposito, il *boxplot* nella Figura 38 mostra la variazione individuale nella produzione di nomi e di verbi italo-romanzi per locutore.

Variazione tra parlanti

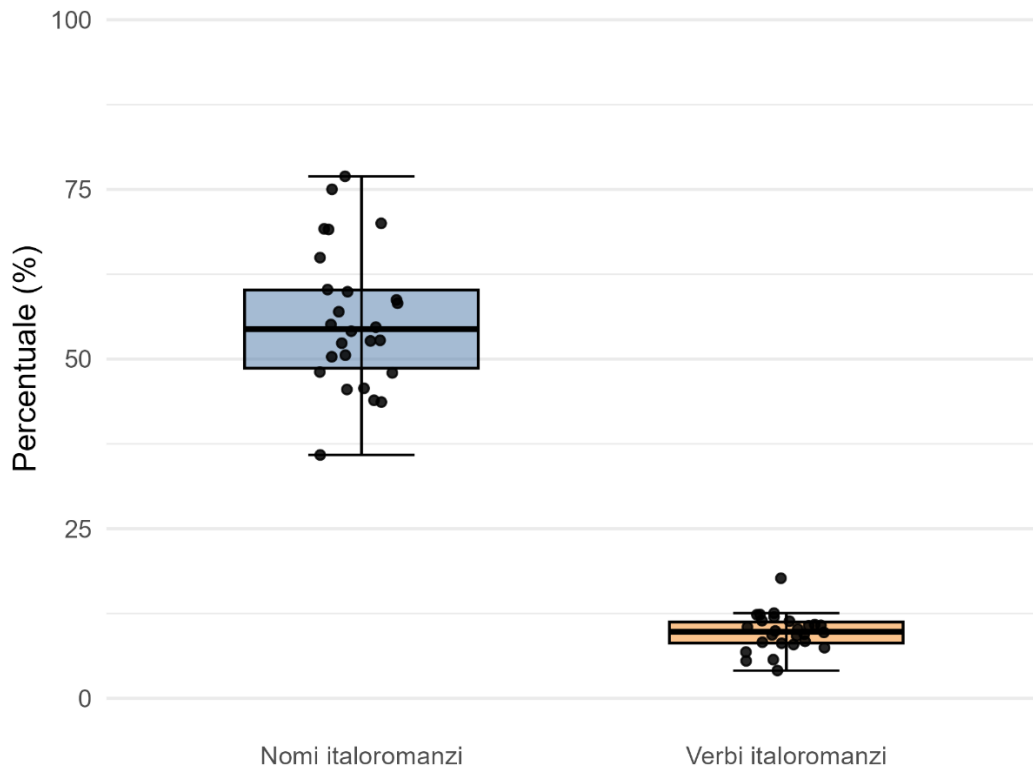


Figura 38: Variazione nella produzione di nomi e verbi italo-romanzi per locutore.

La Figura 38 mostra che tutti i parlanti producono nomi italo-romanzi a tassi molto maggiori rispetto ai verbi italo-romanzi. Inoltre, la produzione di verbi italo-romanzi appare molto stabile attraverso i singoli parlanti, con tassi che oscillano da 4.1% al 17.7% sul totale dei verbi per locutore, nonostante di fatto la maggior parte dei parlanti si concentri nella fascia 5%-10%. Viceversa, il tasso dei nomi italo-romanzi sul totale dei nomi appare molto più variabile, oscillando tra il 36% e il 77%. Nuovamente, notiamo la sostanziale assenza di *outlier* per entrambe le categorie, salvo per il parlante che produce il 17.7% di verbi italo-romanzi.

Dal momento che i nomi mostrano maggiore variabilità, si è ulteriormente esaminata la loro variazione sulla base di genere, età e insediamento dei parlanti; cfr. Figure 39, 40 e 41.

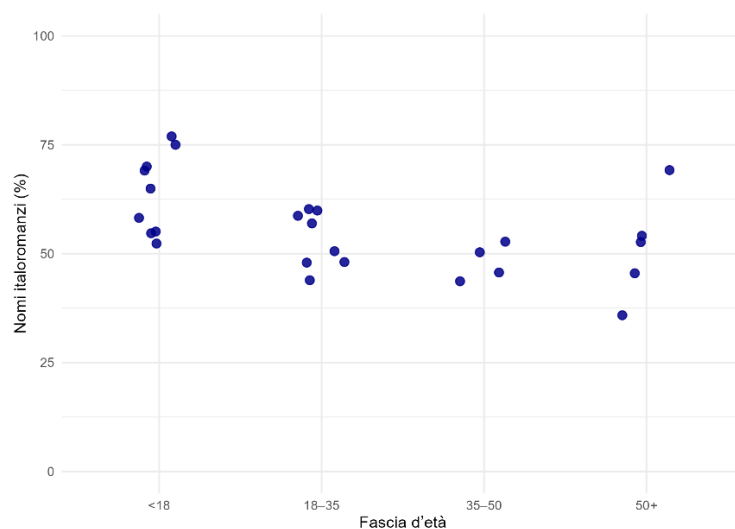


Figura 39: Distribuzione del tasso di nomi italo-romanzi per fascia d'età.

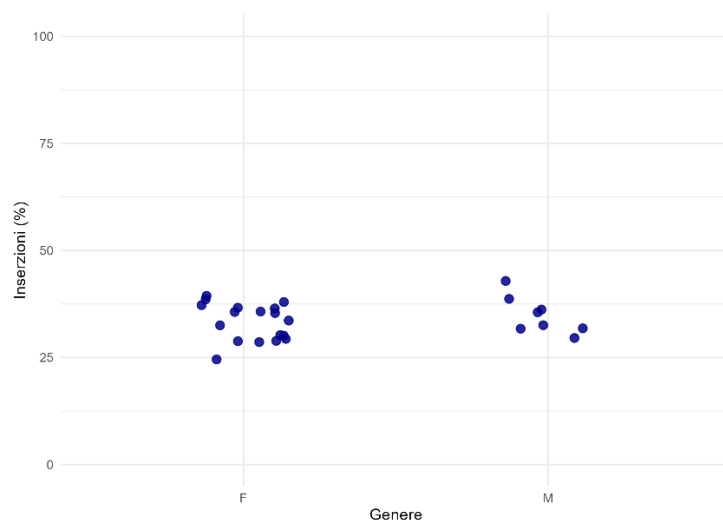


Figura 40: Distribuzione del tasso di nomi italo-romanzi per genere.

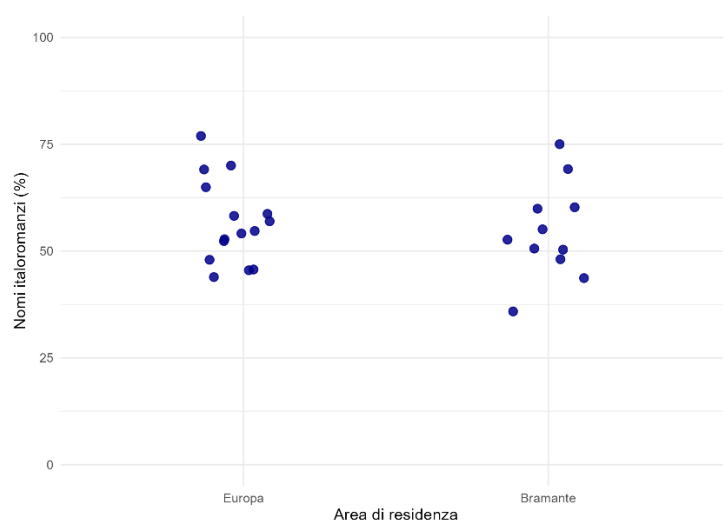


Figura 41: Distribuzione del tasso di nomi italo-romanzi per insediamento.

La Figura 39 sembra indicare alcune tendenze interessanti rispetto all'età: in particolare, i parlanti di meno di 18 anni producono i tassi più alti di inserzioni di nomi, seguiti dalle fasce d'età intermedie (18-35 e 35-50). D'altra parte, i parlanti di più di 50 anni mostrano un comportamento più variabile. Quanto al genere e all'insediamento (Figure 38-39), emergono tendenze meno diversificate, sebbene si possa notare che i parlanti uomini mostrano una maggiore dispersione interna rispetto alle parlanti donne.

Un test ANOVA (utilizzando il rapporto dei nomi italiani sul totale dei nomi come variabile dipendente, ed età, genere e insediamento come variabile indipendente) mostra effetti significativi sia per il genere ($p = 0.0020$) sia per l'età ($p = 0.0016$), mentre la variabile 'insediamento' non appare significativa ($p = 0.474$). Test post-hoc a coppie hanno mostrato che l'unica differenza statisticamente significativa emerge tra i parlanti di meno di 18 anni e i parlanti di età compresa tra i 35-50 anni ($p = 0.00390$). Inoltre, il confronto tra i parlanti di meno di 18 anni e quelli di età compresa tra i 18 e 35 anni restituisce un valore vicino alla soglia di significatività, stabilita a 0.05 ($p = 0.0534$). Complessivamente, i test post-hoc a coppie sembrano confermare il comportamento divergente dei minori di 18 anni rispetto agli altri gruppi di età per quanto riguarda la produzione di nomi italo-romanzi.

Data la significatività dei fattori 'età' e 'genere', si è infine proceduto ad un'analisi della loro interazione, visualizzata nella Figura 42.

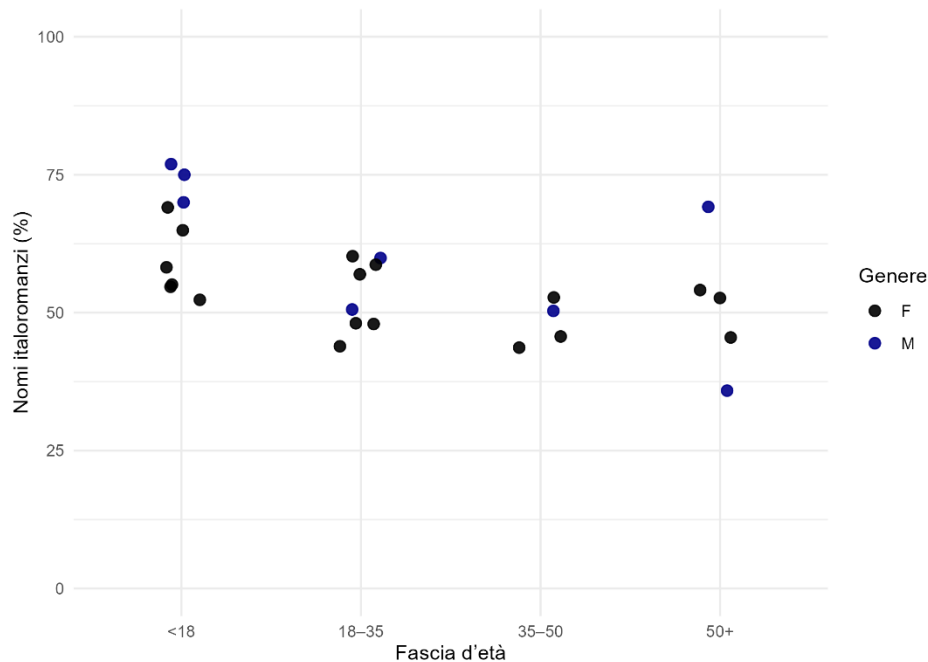


Figura 42: Interazione tra fascia d'età e genere nella distribuzione del tasso di nomi italo-romanzi.

Come si osserva, i parlanti uomini di meno di 18 anni producono un tasso di nomi italo-romanzi molto simile, compreso tra il 70% e il 76.9%, il più alto tra tutti i soggetti del campione. Anche le parlanti donne di meno di 18 anni mostrano tendenze simili, producendo un numero di nomi italo-romanzi minore rispetto ai coetanei uomini, ma compreso tra il 52% al 69%. Ciò sembrerebbe suggerire che il gruppo di pari (inteso come combinazione di genere ed età) possa essere rilevante nella variazione degli schemi di parlato bilingue analizzati. Tuttavia, passando alle fasce d'età superiori, è più difficile individuare tendenze nette. Per esempio, i parlanti uomini di più di 50 anni sono massimamente differenziati, con un parlante che produce meno del 35,9% di nomi italo-romanzi (il dato più basso di tutti i parlanti), mentre l'altro parlante produce il 69% di nomi italo-romanzi (uno dei valori più alti tra tutti i parlanti).

7.4 Discussione

In questo capitolo abbiamo adottato un approccio variazionista all'analisi del parlato bilingue dei sinti lombardi, al fine di verificare in che misura le tendenze quantitative, presentate nei capitoli precedenti come medie complessive, trovino riscontro a livello individuale e attraverso diversi gruppi sociali. Nel complesso, i risultati mostrano una variazione relativamente limitata negli schemi di parlato bilingue, pur con alcune differenze legate ai singoli *pattern* (inserzioni vs. alternanze) e alle categorie lessicali considerate.

Più nello specifico, l'analisi ha mostrato come il ricorso a inserzioni italo-romanze all'interno di discorsi in sinto lombardo rappresenti la strategia dominante di parlato bilingue per tutti i locutori, per i quali le inserzioni costituiscono almeno il 25% dell'*output* verbale. Inoltre, esse appaiono molto stabili tra i parlanti, con una variazione di circa il 15% tra il parlante che produce più inserzioni e il parlante che ne produce di meno. Viceversa, il ricorso al *code-switching* alternante risulta più variabile: alcuni parlanti non vi ricorrono, mentre altri vi ricorrono in maniera relativamente cospicua. Questi risultati supportano l'idea per cui il ricorso a inserzioni italo-romanze rappresenti una modalità “non marcata” di parlare il sinto lombardo, mentre le alternanze costituiscono una strategia opzionale, connessa a specifici fattori contestuali o discorsivi, come osservato al Capitolo 5. Un risultato interessante emerso dall'analisi, sebbene ancora da comprendere del tutto, riguarda la correlazione inversa tra inserzioni e alternanze, tale per cui i parlanti che producono più inserzioni producono meno alternanze (cfr. Adamou, 2016 per risultati simili su un corpus bilingue romaní-finlandese). Ciò sembra indicare che le due strategie rappresentano, almeno in parte, modalità alternative di gestione del repertorio bilingue.

Un'analisi più dettagliata della produzione di nomi e verbi italo-romanzi ha inoltre mostrato come il ricorso a nomi italo-romanzi presenti una forte variabilità tra parlanti, mentre il ricorso a verbi

italoromanzi appare più stabile. Questa differenza può verosimilmente essere ricondotta alla maggiore sensibilità dei nomi agli argomenti di conversazione trattati.

Per quanto riguarda la variabilità individuale e sociale, l'assenza pressoché sistematica di *outlier* nei *boxplot* e la mancata osservazione di associazioni significative tra variazione negli schemi bilingui e le variabili sociali considerate sembrano indicare che il singolo individuo, l'età, il genere e l'insediamento non governano in modo sistematico i *pattern* generali di parlato bilingue nel corpus analizzato. In questo senso, l'analisi conferma le osservazioni di Adamou (2016) per cui piccole comunità minoritarie, caratterizzate da reti sociali “dense” e “transitive”, tendono a produrre un parlato bilingue relativamente stabile, convergendo verso le norme stabilite a livello comunitario.

L'unica parziale eccezione è rappresentata dai nomi, che mostrano maggiore variazione individuale e una differenza significativa legata all'età e, in misura minore, al genere, con i parlanti di meno di diciotto anni a produrre un maggior numero di nomi italo-romanzi rispetto ad alcune fasce adulte. Questo risultato potrebbe essere correlato a un contatto più intensivo dei parlanti giovani con l'italiano (ad esempio attraverso la scuola) e appare in linea con le osservazioni metalinguistiche dei parlanti secondo cui i giovani sinti si servono di un sinto “gagizzato”, ovvero fortemente mescolato all'italiano (cfr. Capitolo 4).

Tuttavia, è necessaria a questo proposito molta cautela interpretativa, per diverse ragioni. In primo luogo, alcuni gruppi di età mostrano una forte variabilità interna nella produzione di nomi italo-romanzi (in particolare i parlanti di più di 50 anni), rendendo difficile individuare una tendenza lineare rispetto all'età. In secondo luogo, le differenze riscontrate potrebbero dipendere almeno in parte dagli argomenti di conversazione trattati: per esempio, non è infrequente che i parlanti più giovani abbiano scelto di parlare, nelle interviste, delle proprie giornate scolastiche o delle proprie attività sportive, ambiti semantici connessi alla realtà sociale e istituzionale italiana e caratterizzati dunque da una forte incidenza di lessico italo-romanzo. Pertanto, per confermare l'ipotesi di un graduale mutamento nei *pattern* di parlato bilingue nella comunità in base alla fascia generazionale, sarebbe necessario lavorare con dati comparabili per dominio tematico o con test controllati in cui i parlanti possano selezionare tra forme italo-romanze e romaní equivalenti, un aspetto che non è stato possibile controllare nella presente analisi.

8. Riflessioni conclusive

Giunti al termine dell'analisi, questo capitolo propone alcune riflessioni conclusive. In particolare, il § 8.1 riprende, ponendoli in relazione, i principali risultati emersi dall'analisi micro e macrosociolinguistica, mentre il § 8.2 colloca il parlato bilingue dei sinti lombardi in una prospettiva tipologica più ampia. Il § 8.3 delinea possibili sviluppi futuri della ricerca, mentre il § 8.4 riflette sulle ricadute educative e sociali maturate nel corso del lavoro sul campo.

8.1 Integrare la prospettiva macro- e macrosociolinguistica

In questo volume, ci eravamo proposti di presentare un'analisi sociolinguistica relativa a una comunità sinta del nord Italia, residente in due diversi insediamenti della città di Pavia. L'obiettivo era offrire una panoramica delle dinamiche di contatto tra il sinto lombardo e l'italiano, sia su un piano macrosociolinguistico, ossia a livello della distribuzione funzionale delle lingue nel repertorio della comunità, sia su un piano macrosociolinguistico, analizzando cioè come esse interagiscono concretamente nel parlato dei singoli individui.

La motivazione di fondo per questo studio nasceva dalla constatazione che, nonostante l'importanza attribuita al tema del contatto linguistico nella linguistica romaní, gli studi in materia adottano prospettive principalmente tipologico-diacroniche, trascurando spesso una prospettiva sincronica basata su dati di parlato (semi)spontaneo all'interno di comunità ben definite (Matras & Adamou, 2020). Abbiamo inoltre notato che questa stessa mancanza si riscontra per le alloglossie di antico insediamento in Italia, dove prevalgono ricerche dal taglio documentaristico e descrittivo, ma raramente studi basati su corpora contenenti produzioni linguistiche spontanee analizzati mediante metodologie quantitative (cfr. Dal Negro, 2009). Il nostro contributo intendeva dunque colmare, almeno parzialmente, tali lacune, collocandosi all'incrocio tra questi due filoni di ricerca.

Per raggiungere questo obiettivo generale è stata adottata una metodologia mista che ha previsto la somministrazione di un questionario sociolinguistico e la realizzazione di interviste semi-strutturate, condotte direttamente da chi scrive, nelle quali il sinto lombardo è stato utilizzato come lingua di base. Le interviste semi-strutturate hanno portato alla realizzazione di un corpus di interviste di circa 40.000 *token*, in cui ogni *token* è stato annotato per lingua e parte del discorso e ricondotto al lemma di riferimento. La trascrizione e l'annotazione interamente manuale del corpus di parlato, che hanno

occupato gran parte del lavoro di ricerca, ne rappresentano anche uno degli esiti più significativi, data la generale mancanza di risorse di questo tipo per la lingua romani.

Per riassumere brevemente i principali risultati emersi dall'analisi, al Capitolo 4 abbiamo osservato che il sinto lombardo e l'italiano costituiscono le lingue dominanti del repertorio della comunità. Le due lingue mantengono una distribuzione funzionale relativamente distinta, ma con possibilità di sovrapposizione nella comunicazione interna, cioè con altri membri della famiglia estesa. Come abbiamo avuto modo di osservare, ciò è coerente con studi precedenti che identificano il repertorio dei sinti lombardi come "dilalico" (Scala, 2012) e, più in generale, con la situazione di "multilinguismo basilettico" (Halwachs, 2003) individuata dagli studiosi per la maggior parte delle comunità rom in Europa. In tale contesto, il sinto lombardo mantiene un livello piuttosto stabile di trasmissione intergenerazionale, anche grazie alle importanti funzioni identitarie che la comunità sembra attribuirgli (spesso in opposizione all'italiano), mentre il dialetto lombardo locale si conferma avere un ruolo estremamente marginale, come già riportato da Scala (2012).

L'approccio etnografico adottato ha inoltre permesso di fare luce sulle rappresentazioni emiche dei parlanti rispetto a specifici fatti di comportamento linguistico, come ad esempio il bilinguismo con l'italiano e il ricorso a pratiche discorsive mistilingui. A questo proposito, abbiamo potuto rilevare che i parlanti mostrano una notevole consapevolezza metalinguistica relativa alla propria condizione di parlanti bilingui e, soprattutto, al ricorso a pratiche discorsive mistilingui, spesso contrapposte a una versione più "stretta" di sinto parlato un tempo e non più in uso. Abbiamo inoltre osservato che gli atteggiamenti dei sinti lombardi verso il discorso bilingue appaiono tutt'altro che negativi, in quanto questa pratica discorsiva viene associata a nozioni di modernità e civilizzazione. Si potrebbe dunque concludere che l'apertura verso pratiche discorsive multilingui rappresenti una strategia adattiva per il mantenimento del sinto lombardo in una situazione di contatto asimmetrico con una lingua maggioritaria dominante quale l'italiano.

Passando al piano microsociolinguistico, l'analisi presentata nei Capitoli 5, 6 e 7 ha esaminato in dettaglio diversi aspetti del contatto linguistico che emergono nella conversazione interna alla comunità.

In primo luogo, sulla base di un conteggio delle parole di origine italoromanza nel corpus, è stato possibile collocare il sinto lombardo in una prospettiva comparativa con altre lingue minoritarie. In questo senso, abbiamo notato che il corpus di sinto lombardo, includendo più del 40% di parole italoromanze, presenta una quantità di elementi provenienti dalle lingue con cui è in contatto molto elevata anche in prospettiva comparativa (Adamou, 2016). Un'analisi più approfondita dei principali schemi combinatori, distinguendo grossomodo tra inserzioni e alternanze (Muysken, 2000), ha mostrato che il materiale italoromanzo è prevalentemente costituito da singole inserzioni in contesti

altrimenti romaní, confermando che l'elevata proporzione di parole italoromanze non deriva semplicemente da meccanismi di alternanza di codice legati, per esempio, alla selezione di un interlocutore monolingue.

Alla luce del contesto macrosociolinguistico precedentemente delineato, per spiegare una così massiccia incorporazione di elementi italoromanzi nel sinto lombardo sembra rilevante menzionare una serie di fattori tra loro strettamente interconnessi:

- (1) il contatto plurisecolare, intensivo ed estensivo, dei sinti lombardi, al netto della loro condizione di relativa marginalità socioeconomica (cfr. anche Adamou, 2010): “intensivo”, in quanto il contatto con l'italiano avviene su base quotidiana (a scuola, sul luogo di lavoro, con i commercianti della città, ecc.); ed “estensivo”, in quanto coinvolge ogni singolo membro della comunità;
- (2) l'assenza di una varietà di standard che possa fungere da riferimento normativo per i parlanti, promuovendo, tra le altre cose, la creazione di neologismi attraverso materiale nativo piuttosto che attraverso il ricorso a inserzioni evidenti dalle lingue di contatto (cfr. Tadmor, 2009; Adamou, 2016; Dal Negro, 2024). Come si è potuto osservare dall'analisi dei calchi lessicali al Capitolo 5, la possibilità di creare neologismi attraverso materiale nativo è in realtà presente in sinto lombardo, ma appare decisamente meno frequente rispetto al ricorso a materiale superficiale italoromanzo;
- (3) l'atteggiamento positivo verso il bilinguismo e verso l'attivazione di una modalità bilingue nelle conversazioni interne al gruppo, nonché una certa apertura verso la variazione e il cambiamento linguistico indotti dal contatto.

Globalmente, questi risultati appaiono in linea con l'abbondante letteratura sul contatto linguistico che sottolinea l'importanza dei fattori ecologici e della storia sociolinguistica dei parlanti nel motivare la quantità e la qualità dei fenomeni di contatto prodotti in una comunità bilingue (cfr., tra gli altri, Thomason & Kaufman, 1988 e Winford, 2013). Come osserva Matras (2020), queste condizioni sociolinguistiche non innescano di per sé tali processi (la cui origine è piuttosto connessa ai meccanismi cognitivi attivi durante la produzione linguistica), ma ne “licenziano” la diffusione nella comunità di parlanti.

Nondimeno, il fatto che anche le altre comunità rom presentino le stesse caratteristiche sociolinguistiche delineate, ma il sinto lombardo mostri comparativamente una maggiore incidenza di “parole di contatto”, sembrerebbe richiedere studi comparativi a grana più fine. In questo senso, risulterebbe rilevante considerare svariati parametri come la durata cronologica del contatto, la dimensione della comunità, la sua natura urbana o rurale, sedentaria o nomade, le attività economiche

storicamente praticate, il contatto con altre comunità rom, ecc. Tuttavia, intraprendere confronti di questo tipo risulta particolarmente complesso, sia per la specificità di ciascuna situazione sociolinguistica, sia per la necessità di analisi storico-etnografiche approfondite, spesso difficili da intraprendere a causa della scarsità di fonti documentarie.

Nei Capitoli 5 e 6 è stata poi dedicata particolare attenzione alle caratteristiche strutturali e sociali delle inserzioni italoromanze nel parlato in sinto lombardo.

In particolare, è stato osservato che le inserzioni italoromanze in discorsi in sinto lombardo provengono da tutte le parti del discorso, sebbene con probabilità diverse, e che le parole contenuto appartengono ai più svariati campi semantici, inclusi ambiti associati al cosiddetto lessico “di base”. Questo profilo ricorda da vicino i *pattern* di parlato bilingue osservati da Dal Negro (2024) per le isole linguistiche di area italoromanza caratterizzate da contatto prolungato e asimmetrico con l’italoromanzo, ad esempio quelle walser, cimbre e mochene. In tali comunità, il parlato bilingue tende infatti a presentare un’elevata presenza di inserzioni romane sia “referenziali” (ossia parole contenuto come nomi e verbi), sia “tetiche” (parole funzione connesse all’organizzazione del discorso, come i segnali discorsivi), sia “funzionali” (cioè elementi legati all’organizzazione sintattico-testuale dell’enunciato, come le congiunzioni e le preposizioni). Come osserva Dal Negro (2024), questi *pattern* contrastano con quelli osservabili in comunità bilingui in cui le lingue a contatto sono poste in un rapporto di contatto meno intenso (come la comunità tedescofona dell’Alto Adige), in cui le inserzioni alloglotte si limitano ad (alcuni tipi di) elementi referenziali e tetici, ma mai funzionali. Il fatto che il sinto lombardo presenti caratteristiche strutturali affini alle isole linguistiche potrebbe costituire un ulteriore argomento per avvicinare questa varietà a tale tipologia di lingue minoritarie (come già sostenuto in Scala, 2020c), nonostante nella letteratura sociolinguistica italiana la romaní sia spesso trattata separatamente, a causa della presunta (ma non effettiva) non territorialità dei suoi parlanti (cfr. Toso, 2008; Vietti, 2014).

Per quanto riguarda la regolarità di occorrenza di specifici elementi italoromanzi, numerosi elementi di alcune categorie funzionali, come connettivi e segnali discorsivi, ma anche preposizioni, focalizzatori e avverbi, risultano altamente regolari nella conversazione, rappresentando spesso l’unica risorsa a disposizione dei parlanti bilingui. In alcuni casi, queste forme sono di origine dialettale, indicando che non si tratta di recenti acquisizioni nel lessico, e appaiono persino registrate come forme sinte lombarde nel dizionario di Soravia & Fochi (1995). Viceversa, le parole contenuto mostrano una regolarità di occorrenza più ridotta, subordinata all’argomento di conversazione; in molti casi si tratta di *hapax* che alternano con equivalenti nativi e assai raramente appaiono registrate nel dizionario di Soravia & Fochi (1995).

Lo studio delle strategie di integrazione del materiale italoromanzo in sinto lombardo ha inoltre mostrato l'esistenza di un'ampia varietà di strategie di (non) integrazione legate allo strato etimologico del contatto, alla categoria grammaticale e alla classe flessiva di appartenenza di una specifica parola. Tra le strategie più stabili a livello individuale e comunitario si è notato che, quando utilizzati in una frase in sinto lombardo, i nomi italoromanzi tendono a ricevere quasi categoricamente articoli romaní, e che i verbi italoromanzi tendono a essere stabilmente integrati all'interno della morfologia flessiva romaní (salvo alcune eccezioni che possono essere spiegate sulla base di altri fattori, come la loro lessicalizzazione come unità non analizzate).

Questi risultati consentono di ottenere importanti generalizzazioni rispetto all'identificazione di una "lingua di base" o "matrice" (Myers-Scotton, 2002; Matras, 2020) nelle frasi bilingui dei sinti lombardi. Sulla base di quanto discusso, essa va a nostro avviso chiaramente identificata nel sinto lombardo, indipendentemente dal numero di morfemi etimologicamente ascrivibili a esso in una determinata frase o enunciato. A questo proposito, elemento di particolare interesse è l'assenza di casi a direzionalità inversa rispetto a quelli discussi, ossia di radici verbali romaní integrate nei paradigmi flessivi romanzi o di nomi romaní che ricevono determinanti romanzi. Questo risultato potrebbe in parte dipendere da scelte di metodo, in quanto ai parlanti era richiesto di parlare in sinto lombardo ai fini delle interviste; tuttavia, anche durante la ricerca sul campo non mi è mai capitato di sentire frasi che presentassero queste strutture a direzionalità inversa (salvo nei rarissimi casi di inserzioni culturali o con connotazione tabuistica di origine romaní in discorsi in italiano, come nella frase *quella lì mangia la ful* 'quella lì mangia la merda'). Strutture miste con lessico romaní e grammatica della lingua maggioritaria si possono piuttosto riscontrare nelle varietà para-romaní o in altri tipi di varietà di contatto che hanno completato uno *shift* verso la lingua maggioritaria, come risultato di un "capovolgimento" (*turnover*) della lingua matrice (cfr. Myers-Scotton, 1998, 2002). In questi casi, il materiale lessicale della lingua minoritaria emerge come in una sorta di "influenza da sostrato" (Dal Negro 2009: 193), incapsulato nelle strutture grammaticali della nuova lingua matrice. Complessivamente, ciò indica che il sinto lombardo si trova in una fase antecedente a tale *turnover*, e che gli schemi di parlato bilingue riscontrati possono ancora essere visti come strategie di innovazione della lingua di minoranza attraverso le potenzialità del proprio sistema.

Infine, l'analisi statistica della variazione individuale presentata al Capitolo 7 ha permesso di verificare, a livello del singolo individuo, la tenuta delle generalizzazioni formulate nei capitoli precedenti sui dati aggregati dell'intera comunità. L'analisi ha confermato che l'inserzione di elementi italoromanzi costituisce lo schema di parlato bilingue dominante per tutti gli individui nella comunità, e che le inserzioni si mantengono relativamente stabili nei singoli locutori se misurate in termini della loro proporzione rispetto al parlato complessivo. Questo risultato permette di osservare

che i sinti lombardi, una comunità caratterizzata da reti sociali altamente dense e transitive, tendono a seguire gli schemi di parlato bilingue stabiliti nella propria comunità, con scarse differenze legate a fattori socio-anagrafici come sesso, età e area di residenza (Adamou, 2016).

8.2 Il parlato bilingue dei sinti lombardi in prospettiva tipologica

Dopo aver riepilogato le principali tendenze emerse dall'analisi, possiamo ora considerare più in dettaglio il parlato bilingue dei sinti lombardi in una prospettiva comparativa.

Pur essendo partiti da un approccio analitico che ha considerato le due lingue in contatto come sistemi relativamente separati e chiaramente distinguibili, l'analisi dei dati ha mostrato come l'italoromanzo e il sinto lombardo costituiscano, almeno in parte, un repertorio fortemente integrato. Questa impressione è anche confermata dalle rappresentazioni emiche dei parlanti che, pur identificando chiaramente la propria lingua come “sinto lombardo”, osservano che parlare in sinto implica un ricorso sistematico a inserzioni italoromanze.

Queste osservazioni richiamano categorie analitiche come quella di *fused lect*, proposta inizialmente da Auer (1999), con cui si fa riferimento a una varietà mista stabilizzata nella quale, a differenza di quanto accade nel *language mixing*, la scelta di ricorrere a elementi da una lingua A e da una lingua B per certi costituenti si annulla, in quanto i parlanti non dispongono più di alternative (cfr. anche Matras, 1998a). Approcci più recenti hanno proposto una concezione più dinamica di *fused lect* (Auer & Hakimov, 2021; Adamou, 2021a), proponendo anche criteri che consentano di classificare con maggiore precisione la quantità di fusione individuabile una varietà di contatto. Consideriamo in particolare i tre criteri recentemente proposti da Auer & Hakimov (2021), che verranno di seguito applicati al caso specifico del sinto lombardo:

- (1) *quantità di struttura coinvolta da fusione*, che può oscillare da singoli elementi appartenenti a una specifica classe funzionale fino a interi sottosistemi della grammatica e del lessico;
- (2) *grado di sedimentazione del processo di fusione*, che oscilla da una situazione in cui gli elementi della lingua di contatto continuano ad alternarsi con equivalenti nativi, a una situazione in cui numerosi elementi diventano obbligatori e non presentano più alternative, passando per una fase intermedia che può anche mostrare variabilità individuabile;
- (3) *grado di compartimentazione della fusione*, riferito nello specifico alla morfologia: si passa da casi in cui la morfologia alloglotta si accompagna esclusivamente al lessico prestato (compartimentazione) a casi in cui la morfologia prestata si estende anche al lessico nativo (fusione).

Per quanto riguarda il primo criterio, nel caso del sinto lombardo il livello di fusione sembra variare sensibilmente a seconda della classe di parole considerata. Alcune classi (in particolare i connettivi, le preposizioni e i numerali) mostrano un grado di fusione molto avanzato, essendo costituite in larga misura da materiale italoromanzo; altre classi, come nomi, aggettivi e avverbi, presentano invece un livello di fusione intermedio; infine, classi quali verbi, pronomi e determinanti mostrano un livello di fusione piuttosto limitato, poiché sono composte prevalentemente da elementi romaní.

Il secondo criterio consente di osservare con maggiore precisione il comportamento dei singoli lemmi di origine italoromanza. Nel caso del sinto lombardo, la maggior parte delle parole funzione tende a non competere con equivalenti nativi, suggerendo un grado molto elevato di stabilizzazione. Per quanto riguarda invece le parole contenuto, emerge una situazione più sfumata: accanto ad elementi ormai completamente stabilizzati si trovano infatti forme che continuano a competere con equivalenti romaní. Anche questo dato conferma che il processo di fusione appare più avanzato nel dominio delle parole funzione.

Quanto al terzo criterio, a livello della morfologia verbale non si dà alcuna traccia di compartimentazione o fusione: al contrario, come osservato più volte, i verbi italoromanzi tendono a essere integrati stabilmente nella morfologia flessiva del sinto lombardo, anche se usati occasionalmente. Nel caso dei nomi, l'inserzione di forme italiane appartenenti a specifiche classi flessive insieme alla loro morfologia flessiva (p.es. i nomi della classe in *-e*, come *occasione*) potrebbero costituire un primo passo verso una compartimentazione.

Nel complesso, in linea con approcci contemporanei (Auer & Hakimov, 2021), il sinto lombardo sembra configurarsi come una varietà di contatto dinamica, nella quale le inserzioni italoromanze si collocano variamente lungo il continuum *language mixing* – *fused lect*. Si noti che molte delle inserzioni italoromanze sono di origine lombarda, piuttosto che italiana, una lingua di cui la maggior parte dei parlanti non ha più conoscenza attiva. Questo è un ulteriore elemento a favore dell'interpretazione del sinto lombardo come *fused lect*, in quanto, nel suo modello, Auer (1999) specifica che i parlanti di un *fused lect* AB possono, ma non necessariamente devono, essere parlanti attivi della lingua A e/o B.

Nella sua definizione di *fused lect*, in aggiunta all'obbligatorietà di specifici costituenti, Auer (1999) richiama ulteriori criteri di natura strutturale, che indicherebbero un processo di ristrutturazione del nuovo sistema misto AB in maniera relativamente indipendente rispetto alle lingue che l'hanno originato. Alla luce delle analisi condotte nei capitoli precedenti, il sinto lombardo sembra presentare almeno tre meccanismi di questo tipo, frequentemente citati in letteratura come possibili diagnostiche di *fused lect*:

- 1) marcatura grammaticale ridondante;

- 2) sviluppo di nuove funzioni da parte di specifiche parole funzione, originariamente assenti o con diverso valore nella lingua di partenza;
- 3) presenza di calchi parziali di strutture multimorfemiche.

Un esempio di marcatura grammaticale ridondante (1) è rappresentato dall'utilizzo simultaneo di marche di caso romaní e preposizioni di origine italoromanza per codificare le funzioni sintattico-semantiche dei pronomi, come discusso al Capitolo 5 (p.es. *par mande* 'per me', dove il ruolo semantico del pronome di prima singolare *me* è codificato sia dalla marca *-de* di locativo, applicata alla base obliqua *man*, sia dalla preposizione di origine italoromanza *par*). Fenomeni analoghi sono stati descritti in letteratura per altri *fused lect*, ad esempio in una varietà mista tagico-uzbeca parlata in Afghanistan, in cui le relazioni di caso possono essere marcate contemporaneamente da morfologia tagica e da preposizioni uzbeche (Boeschoten & Backus, 1997: 57, citato in Auer, 1999).

Per quanto riguarda il secondo punto (2), i dati suggeriscono che alcune parole funzione italoromanze hanno sviluppato in sinto lombardo nuovi valori semantico-funzionali non attestati nel dialetto lombardo, anch'essi discussi al Capitolo 5. È il caso, ad esempio, di *bela* 'già', che deriva probabilmente dalla costruzione lombarda *bel e* 'bello e' (all'interno di strutture come *l'è bel e arivà* 'è già arrivato') e che nel sinto lombardo ha acquisito pienamente la funzione di avverbio fasale (come nella frase *jom bela kaj* 'sono già qui'). Si segnala inoltre l'uso del focalizzatore italoromanzo *pena* 'solo' (dal lombardo *apena*) in combinazione con il complementatore italoromanzo *che* a formare la congiunzione etimologicamente romanza *pena che* 'solo che', la quale, a mia conoscenza, non si rintraccia nei dialetti italoromanzi.

Infine, la presenza di calchi parziali su modelli chiaramente italoromanzi (3), discussi al Capitolo 5, è piuttosto prominente nel sinto lombardo. Esempi includono sia espressioni abbastanza stabilizzate, come *qualche jek* 'qualcuno' o *daparkrol* 'dappertutto', sia costruzioni che emergono spontaneamente nel discorso, eventualmente in variazione con le forme italoromanze direttamente copiate come tali. Esempi simili sono stati discussi in relazione ad altri *fused lect*, come il kildin saami parlato in Russia (Hakimov & Reissler, 2021), e sembrerebbero suggerire l'abilità dei parlanti di stabilire corrispondenze formali punto per punto tra le lingue coinvolte (cfr. anche Dal Negro, 2005a: 168).

Il ricorso a calchi parziali è stato inoltre menzionato da Muysken (2000) come diagnostico di un meccanismo di "congruent lexicalization" nel parlato bilingue, sulla base del parametro "mixed idioms and collocations" (*ibid.*:134). Per esemplificare questo parametro, l'autore cita verbi frasali con radice lessicale in una lingua A e particella in una lingua B. Secondo Muysken (2000), l'emergere di queste strutture è favorito da una convergenza delle lingue in contatto. Benché la convergenza

sintattica delle strutture grammaticali in sinto lombardo non sia stata affrontata nel dettaglio in questo volume, si notano numerosi esempi di questo tipo (cfr. § 8.3).

Analogamente, il comparire di combinazioni di morfemi che sarebbero inattese in situazioni di “*code-switching* classico” (come nel caso della marcatura grammaticale ridondante sui pronomi o dei calchi parziali multimorfemici) è stato considerato da Myers-Scotton (2002) come un criterio distintivo di “*composite code-switching*”, nel quale la struttura grammaticale astratta degli enunciati bilingui non si riconduce a una lingua matrice univoca. Tuttavia, come osservato più volte, la produttività della morfologia verbale romaní per integrare inserzioni italoromanze suggerisce che il sinto lombardo continua chiaramente a costituire la lingua matrice degli enunciati bilingui, in linea con l’interpretazione più ristretta di lingua matrice proposta da Matras (2020).

Nel parlato bilingue in sinto lombardo si possono inoltre individuare diversi altri parametri compatibili con il modello di “*congruent lexicalization*” proposto da Muysken (2000). Tra questi si menzionano: la presenza di svariati morfemi o classi di parole dalla lingua B, incluse molte parole funzione (cfr. parametro “*function words*”; *ibid.*: 130); la presenza di fenomeni di co-attivazione, come ipotizzato per le sequenze nome-numerale o preposizione-nome in presenza di toponimi (cfr. parametro “*triggering*”; *ibid.*: 134); e la presenza di *switch* interni alla parola, particolarmente evidenti nel caso dei verbi (cfr. parametro “*morphological integration*”; *ibid.*: 134).

Infine, Muysken (2000) definisce come tratto diagnostico di *congruent lexicalization* l’impossibilità di stabilire con certezza l’attribuzione linguistica di certi segmenti discorsivi a causa della presenza di *cognates* o elementi prestati da una lingua all’altra. Anche nel caso del sinto lombardo emergono tratti compatibili con questo: la varietà è infatti caratterizzata da un’elevata presenza di lessemi condivisi con l’italoromanzo che, pur risultando chiaramente riconducibili a quest’ultimo sul piano etimologico, presentano uno statuto più ambiguo in prospettiva sincronica. In molti casi tali forme appaiono pienamente integrate nel repertorio dei parlanti e non sembrano essere necessariamente percepite come appartenenti a un sistema linguistico distinto. Significativamente, nell’ambito del supporto fornitomi durante la trascrizione e annotazione dei dati (cfr. Capitolo 3), la consulente E.A. ha spesso manifestato difficoltà nell’attribuire determinate parole o sintagmi al ‘sinto’ o all’‘italiano’, proprio perché tali forme risultavano condivise nel suo repertorio bilingue.

L’analisi del sinto lombardo suggerisce pertanto la possibilità di stabilire un collegamento tra le nozioni di *fused lect* e *congruent lexicalization*, che raramente sono state messe in relazione in letteratura (ma cfr. Ciccolone, 2014). Un’esplorazione sistematica delle convergenze e delle differenze tra questi modelli su nuovi dati sembra dunque costituire una promettente direzione di ricerca futura. Inoltre, le osservazioni fornite in questo Capitolo sembrano compatibili con prospettive

più recenti sul plurilinguismo, quali ad esempio il *translanguaging* (descritto al Capitolo 1), che enfatizzano il carattere integrato dei repertori linguistici degli individui plurilingui.

8.3 Sviluppi di ricerca futuri

Questo volume ha offerto un'analisi approfondita delle pratiche linguistiche dei sinti lombardi di Pavia, ma i suoi possibili sviluppi futuri sono numerosi.

Una prima prospettiva riguarda l'ampliamento del corpus di parlato. Ciò potrebbe permettere di osservare con maggiore precisione alcune tendenze emerse dall'analisi, ma sottorappresentate nei dati attualmente disponibili. Nel contesto di questa espansione, sarebbe auspicabile includere una maggiore varietà di situazioni comunicative, caratterizzate da diversi gradi di formalità, nonché da livelli di attivazione lungo il continuum tra modalità monolingue e modalità bilingue (Grosjean, 2001).

Parallelamente all'ampliamento del corpus, un ulteriore sviluppo per integrare l'analisi qui presentata riguarda l'impiego di metodi sperimentali. Recenti studi hanno infatti mostrato l'efficacia di metodi sperimentali applicati a lingue minacciate al fine di confermare o smentire ipotesi formulate a partire dall'osservazione e dall'analisi di dati ecologici (Adamou, 2021b). Ad esempio, l'ipotesi che il sinto lombardo possa essere interpretato come un *fused lect* potrebbe essere verificata attraverso esperimenti di processamento linguistico. Se, come suggerito da Auer (2014), un *fused lect* rappresenta una forma di parlato bilingue stabilizzato potenzialmente assimilabile al parlato monolingue, la processazione degli enunciati mistilingui non dovrebbe comportare per i parlanti costi cognitivi maggiori rispetto alla processazione del parlato monolingue (Adamou & Shen, 2017; Adamou, 2021a). Più nello specifico, alla luce della concezione dinamica di *fused lect* proposta nella letteratura recente ed estesa all'interpretazione del sinto lombardo nel § 8.2, si dovrebbero prevedere differenze molto fini nei costi di processazione di singoli enunciati, modulate dal grado di stabilizzazione di specifici elementi lessicali italo-romanzi nel parlato in sinto lombardo (Adamou, 2021a).

Similmente, nel Capitolo 6 è stato osservato che le strategie di integrazione dei verbi italo-romanzi nel parlato in sinto lombardo seguono meccanismi complessi, con la competizione tra i verbalizzatori *-ar-* e *-o-* nei verbi intransitivi potenzialmente legata a fattori all'interfaccia tra semantica e sintassi come il grado di agentività del soggetto o la persona del verbo. Test sperimentali che controllino attentamente questi parametri (che non sarebbero comunque stati intuibili senza un'analisi dettagliata del corpus raccolto) potrebbero contribuire a chiarire i meccanismi alla base di questa alternanza.

La presente ricerca si è quasi esclusivamente concentrata sull'analisi dei fenomeni di parlato bilingue considerati a livello di replicazione di “matter”, cioè di forme esogene nella loro sostanza

fonomorfologica (Matras & Sakel, 2007). In futuro, sarebbe auspicabile ampliare l'analisi degli effetti del contatto ad altri livelli linguistici, analizzando la convergenza tra strutture sintattiche. A questo proposito, Forlano & Brigada Villa (in stampa) hanno osservato che l'ordine reciproco del nome e dell'aggettivo in sinto lombardo ha sviluppato una chiara convergenza con l'italiano, in quanto, nonostante per la romaní si ricostruisca un ordine nativo Aggettivo-Nome (Adamou & Matras, 2020), in sinto lombardo gli aggettivi appaiono più frequentemente in posizione postnominale, similmente all'italiano. Inoltre, gli autori osservano che la variazione nel posizionamento di uno stesso aggettivo, indipendentemente dalla sua etimologia, può essere motivata sulla base degli stessi fattori semantico-pragmatici attivi per l'italiano, come la natura restrittiva o non restrittiva della modificazione.

Vi sono tuttavia numerosi altri punti in cui la grammatica del sinto lombardo e dell'italiano sembrano aver sviluppato una convergenza. Abbiamo già menzionato al Capitolo 6 il possibile sviluppo di costruzioni verbali analitiche per l'espressione del passato o del passivo, sotto l'influsso dell'italiano, attraverso l'impiego creativo di materiale ereditato. Un altro esempio a livello sintattico è rappresentato dalla costruzione negativa, che in sinto lombardo può essere realizzata mediante diverse configurazioni a seguito del prestito dell'intensificatore negativo italoromanzo *miga* (cfr. Capitolo 5): il costrutto con nativo *na + verbo*, il costrutto *na + verbo + miga*, oppure il costrutto *verbo + miga*, che replica un *pattern* tipico dei dialetti italoromanzi co-territoriali (cfr. anche Scala, 2021). L'esplorazione del corpus raccolto consentirebbe dunque di comprendere con maggiore precisione i fattori alla base della variazione a livello di una grande varietà di costruzioni sintattiche.

Infine, sarebbe auspicabile un numero crescente di studi *corpus-based* sul parlato bilingue delle comunità romaní. Ricerche di questo tipo permetterebbero di tracciare distinzioni più fini tra i diversi tipi di parlato bilingue praticati nelle varie comunità rom e sinte e di mettere tali differenze in relazione con la storia sociolinguistica dei parlanti, mantenendo al contempo relativamente costanti variabili come la struttura della lingua minoritaria e la natura asimmetrica del contatto (cfr. Rusakov, 2001; Adamou, 2010, 2016).

8.4 Dalla documentazione linguistica, alla scuola, alla società

Al termine di questa trattazione, mi sia consentito un breve bilancio relativo all'esperienza di ricerca sul campo con i sinti lombardi e alla sua evoluzione nel tempo.

Quando ho avviato questa ricerca, nel 2021, i miei interessi erano principalmente orientati alla somministrazione di questionari sociolinguistici e alla costruzione del corpus di parlato: pur essendo certamente motivati da un forte interesse per la cultura sinta, essi si collocavano dunque strettamente nell'ambito della documentazione linguistica. Ben presto, tuttavia, il contesto nel quale mi sono trovato a operare ha sollecitato ulteriori priorità e prospettive.

In particolare, la collaborazione nell'ambito delle attività di doposcuola mi ha reso particolarmente sensibile alla questione scolastica. Come già anticipato al Capitolo 3, molti studenti sinti sperimentano percorsi educativi segnati da forme di discriminazione che spesso si traducono in un elevato tasso di dispersione scolastica. In tale contesto, il conseguimento di un diploma di scuola secondaria di secondo grado rimane un traguardo estremamente raro. In questo quadro, provando a effettuare un bilancio del tempo trascorso con la comunità in questi anni, constato che il numero di ore dedicate all'attività educativa è stato esponenzialmente maggiore di quelle strettamente impiegate nella raccolta dei dati linguistici.

Nel 2025, l'associazione presso la quale svolgevo attività di volontariato, Babele ODV, è entrata come partner nel progetto "A scuola con Mafalda - Progetto per l'inclusione ed integrazione di bambine, bambini e adolescenti Rom Sinti Caminanti", sostenuto dai finanziamenti del Programma Nazionale Inclusione e Lotta alla Povertà 2021–2027²³. In tale contesto, la coordinatrice del centro, Elisabetta Della Vigna, mi ha offerto la preziosa opportunità di entrare a far parte del progetto con l'obiettivo di operare formalmente sia in ambito scolastico sia extrascolastico. A partire dall'anno scolastico 2025–2026, parallelamente alle attività di sostegno scolastico negli insediamenti, sono così stati avviati numerosi percorsi di formazione rivolti ai docenti delle scuole della città di Pavia con una significativa presenza di alunni sinti, nonché laboratori destinati alle classi.

Tra le iniziative realizzate, mi piace citare il laboratorio "Le lingue siamo noi", organizzato in due incontri di due ore ciascuno in una classe quarta della scuola primaria Gabelli di Pavia e finalizzato alla promozione del plurilinguismo attraverso attività di riflessione linguistica che includevano anche numerosi riferimenti alla lingua romani. Al termine del percorso, l'insegnante ha osservato come l'alunno sinto presente in classe avesse manifestato un livello di partecipazione e motivazione raramente riscontrato in precedenza durante le attività scolastiche, avendo trovato uno spazio nel quale la propria esperienza linguistica e culturale potesse essere riconosciuta e valorizzata.

Parallelamente, nell'ambito del progetto "A scuola con Mafalda" sono stati organizzati eventi aperti alla cittadinanza dedicati alla memoria del *Samudaripen*, il genocidio dei rom e dei sinti durante la Seconda guerra mondiale, nonché iniziative legate alla Giornata Internazionale dei Rom, dei Sinti e dei Caminanti, celebrata l'8 aprile. In occasione di uno di questi eventi, ho anche avuto l'opportunità di presentare alla comunità una prima restituzione dei risultati emersi dalla ricerca. Analogamente, in occasione della Giornata mondiale della lingua romani, che ha luogo il 5 novembre, è previsto lo svolgimento di un ciclo di eventi legati, appunto, al tema della lingua.

²³ Fondo FSE+, Priorità 2 "Child Guarantee" - CUP: J21H24000060006.

Riflessioni conclusive

Il percorso è certamente ancora lungo e molti obiettivi restano da raggiungere. Con questo volume, spero di aver contribuito a mostrare come la documentazione linguistica, se perseguita attraverso una collaborazione stretta con la comunità, possa rappresentare un punto di partenza per la costruzione di relazioni, percorsi e progetti più ampi, condivisi tra ricerca, scuola e società. Non vedo l'ora di proseguire in questa direzione.

Indice delle figure

Figura 1: Mappa dei principali dialetti romaní (da Matras, 2002: 11).....	36
Figura 2: Collocazione degli insediamenti sinti nel comune di Pavia: il cerchio rosso identifica l'insediamento di piazzale Europa, mentre il cerchio grigio l'insediamento di via Bramante.	45
Figura 3: Risultati accorpati delle domande “Qual è la tua lingua madre?” e “Quali altre lingue o dialetti parli?”.....	62
Figura 4: Risposte alla domanda: “Qual è la tua lingua madre?”.....	64
Figura 5: Risposte alla domanda “Quale/i lingua/e hai imparato per prima/e da piccolo?”.....	68
Figura 6: Competenze autodichiarate in sinto lombardo e italiano (parlato, lettura, scrittura).	69
Figura 7: Variazione nelle autovalutazioni per insediamento.....	70
Figura 8: Variazione nelle autovalutazioni per genere.	70
Figura 9: Variazione nelle autovalutazioni per fascia d'età.....	71
Figura 10: Lingue utilizzate nei domini endocomunitari ed esocomunitari (aggregato).....	73
Figura 11: Lingue utilizzate nei domini endocomunitari al variare degli interlocutori.....	74
Figura 12: Risposte alla domanda “Hai parenti o amici sinti che parlano lingue diverse dal sinto lombardo?”.....	77
Figura 13: Lingue dichiarate con parenti sinti non lombardi (aggregato).	77
Figura 14: Risposte alla domanda “In quale lingua parli al telefono con altre persone sinte?”.....	80
Figura 15: Risposte alla domanda “In quale lingua scrivi messaggi ad altre persone sinte?”.....	80
Figura 16: Risposte alla domanda “Sei fiero di parlare il sinto?”.....	83
Figura 17: Risposte alla domanda “Conoscere il sinto per essere sinti è...?”.....	83
Figura 18: Risposte alla domanda “Prima di andare a scuola, i bambini dovrebbero crescere parlando...?”.....	84
Figura 19: Risposte alla domanda “Tendi a mischiare il sinto e l'italiano mentre parli?”.....	85
Figura 20: Risposte alla domanda “Mischiare il sinto e l'italiano mentre si parla è...”.....	86

Figura 21: Confronto della composizione linguistica del corpus di sinto lombardo con corpora di altre lingue minoritarie (Dal Negro, 2009; Adamou, 2016; Mirić & Ćirković, 2022).	95
Figura 22: Distribuzione delle parole per lingua e parte del discorso nel corpus, in contesti dominanti romaní (percentuale).	100
Figura 23: Regolarità dei nomi e dei verbi italoromanzi.	113
Figura 24: Distribuzione dei sintagmi nominali del corpus per lingua del nome e del determinante.	133
Figura 25: Distribuzione dei sintagmi nominali del corpus per lingua del nome e dell'indefinito.	135
Figura 26: Distribuzione dei sintagmi nominali del corpus per lingua del nome e del numerale. ..	136
Figura 27: Distribuzione dei verbi italoromanzi morfologicamente integrati e non integrati nel corpus.	140
Figura 28: Variazione nella produzione di parole italoromanze, romaní e 'altro' per locutore.	152
Figura 29: Variazione del tasso di parole italoromanze per fascia d'età.	153
Figura 30: Distribuzione del tasso di parole italoromanze per genere.	153
Figura 31: Distribuzione del tasso di parole italoromanze per area di residenza.	153
Figura 32: Relazione tra tasso di parole italoromanze e parole totali per parlante.	154
Figura 33: Variazione nella produzione di inserzioni e alternanze per locutore.	155
Figura 34: Correlazione tra tasso di inserzioni e tasso di alternanze nel parlato dei singoli locutori.	156
Figura 35: Distribuzione del tasso di inserzioni per fascia d'età.	157
Figura 36: Distribuzione del tasso di inserzioni per genere.	157
Figura 37: Distribuzione del tasso di inserzioni per insediamento.	158
Figura 38: Variazione nella produzione di nomi e verbi italoromanzi per locutore.	159
Figura 39: Distribuzione del tasso di nomi italoromanzi per fascia d'età.	160
Figura 40: Distribuzione del tasso di nomi italoromanzi per genere.	160
Figura 41: Distribuzione del tasso di nomi italoromanzi per insediamento.	160
Figura 42: Interazione tra fascia d'età e genere nella distribuzione del tasso di nomi italoromanzi	161

Indice delle tabelle

Tabella 1: Domini d'impiego della romaní (da Halwachs, 2003).....	37
Tabella 2: Composizione della comunità di sinti pavesi per genere.....	46
Tabella 3: Composizione della comunità di sinti pavesi per fascia d'età.	46
Tabella 4: Caratteristiche socio-anagrafiche dei rispondenti al questionario.	55
Tabella 5: Caratteristiche socio-anagrafiche dei partecipanti alle interviste.	56
Tabella 6: Residui standardizzati del test χ^2 per interlocutore e scelta linguistica nei domini endocomunitari.....	75
Tabella 7: Schematizzazione del repertorio linguistico comunitario dei sinti lombardi sulla base di alcuni parametri proposti in Dal Negro & Iannàccaro (2003).	91
Tabella 8: Composizione del corpus per lingua.....	94
Tabella 9: Distribuzione dei nomi e dei verbi italoromanzi sulla base delle categorie del WOLD (Haspelmath & Tadmor, 2009).	104
Tabella 10: Rappresentazione del continuum prestito- <i>code-switching</i> sulla base del criterio della regolarità.	106
Tabella 11: Parole funzione italoromanze nel corpus (lemmi).	109
Tabella 12: Nomi italoromanzi più regolari in sinto lombardo.	114
Tabella 13: Verbi italoromanzi più regolari in sinto lombardo.	117
Tabella 14: Principali strategie di integrazione dei nomi italoromanzi.	131
Tabella 15: Distribuzione per lemma e type dei principali verbi finiti non integrati in sinto lombardo.	147

Bibliografia

- Abercrombie A., 2018. Language purism and social hierarchies: Making a Romani standard in Prizren. *Language in Society*, 47(5): 741-761.
- Adamou E., 2015. A corpus-driven analysis of Romani in contact with Turkish and Greek. In Torgersen E., S. Hårstad, B. Mæhlum, U. Røyneland (a cura di), *Proceedings of the 7th International Conference on Language Variation in Europe (ICLaVE 7)*. Amsterdam & Philadelphia: Benjamins, pp. 1-16.
- Adamou E., 2016. *A Corpus-driven Approach to Language Contact. Endangered Languages in a Comparative Perspective*. Boston & Berlin: de Gruyter Mouton.
- Adamou E., 2021a. How sentence processing sheds light on mixed language creation. In Mazzoli M., E. Sippola (a cura di), *New Perspectives on Mixed Languages. From Core to Fringe*. Boston & Berlin: de Gruyter Mouton, pp. 93-119.
- Adamou E., 2021b. *The Adaptive Bilingual Mind: Insights from Endangered Languages*. Cambridge: Cambridge University Press.
- Adamou E., 2024. *Endangered Languages*. Cambridge, MA: MIT Press.
- Adamou E., K. Granqvist, 2015. Unevenly Mixed Romani Languages. In *The International Journal of Bilingualism*, 19(5): 525-547.
- Adamou E., W. Breu, L. Scholze, R.X. Shen., 2016. Borrowing and contact intensity: A corpus- driven approach from four Slavic minority languages. In *Journal of Language Contact*, 9(3): 515- 544.
- Adamou E., R. Shen, 2017. There are no switching costs when codeswitching is frequent. In *International Journal of Bilingualism*, 23: 53-70.
- Adamou E., Y. Matras, 2020. Romani Syntactic Typology. In Matras Y., A. Tenser (a cura di), *The Palgrave Handbook of Romani Linguistics*. London, Palgrave MacMillan, pp 187–227
- Adamou E., Y. Matras, 2021. *The Routledge Handbook of Language Contact*. London: Routledge.
- Adamou E., Bullock B., Toribio A. J., 2023. *Understanding Language Contact*. London: Routledge.
- Agresti G., 2023. Appunti per una storia sociale della romaní d'Abruzzo. Il problema delle rappresentazioni sociali di lingua e identità. In D. Di Virgilio (a cura di), *Musiche tradizionali in Abruzzo: le registrazioni di Giorgio Nataletti (1948-49)*. Roma: Squilibri, pp. 78-99.
- Auer P., 1999. From code-switching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. In *The International Journal of Bilingualism*, 3: 309-332.
- Auer P., 2014. Language mixing and language fusion: When bilingual talk becomes monolingual. In Besters-Dilger J., C. Dermarkar, S. Pfänder, A. Rabus (a cura di), *Congruence in contact- induced*

- language change, language families, typological resemblance, and perceived similarity* (Linguae & Litterae 27). Berlin: De Gruyter, pp. 294-334.
- Auer P., R. Muhamedowa, 2005. “Embedded language” and “matrix language” in insertional language mixing: Some problematic cases. In *Journal of Italian Linguistics*, 17(1): 35-54.
- Auer P., N. Hakimov, 2021. From language mixing to fused lects: The process and its outcomes. In *International Journal of Bilingualism*, 25(2): 361-368.
- Backus A., 1996. *Two in one: Bilingual speech of Turkish immigrants in the Netherlands*. Tilburg: Tilburg University Press.
- Backus A., 2001. The role of semantic specificity in insertional codeswitching: Evidence from Dutch-Turkish. In R. Jacobson (a cura di), *Codeswitching Worldwide II*, 125-154. Berlin and New York: Mouton de Gruyter.
- Backus A., 2003. Units in code switching: Evidence for multimorphemic elements in the lexicon. In *Linguistics*, 41(1): 83-132.
- Backus A., 2015. A usage-based approach to code-switching: The need for reconciling structure and function. In Stell G., K. Yakpo (a cura di), *Codeswitching between structural and sociolinguistic perspectives* (Linguae & Litterae). Berlin: De Gruyter, 43: 19-38.
- Backus A., M. Dorleijn, 2009. Loan translations versus code-switching. In B. E. Bullock, A. J. Toribio (a cura di), *The Cambridge Handbook of Linguistic Code-Switching*, pp. 54-75.
- Baker C., 1992. *Attitudes and Languages*. Bristol: Multilingual Matters.
- Bakker P., 1997. *A language of our own: The genesis of Michif, the mixed Cree-French language of the Canadian Métis* (Oxford studies in anthropological linguistics, 10). Oxford & New York: Oxford University Press.
- Bakker P., 2001. Romani in Europe. In Extra G., D. Gorder (a cura di), *The other languages of Europe: Demographic, sociolinguistic and educational perspectives*. Bristol: Multilingual Matters, pp. 293-313.
- Benišek M., 2020. The Historical Origins of Romani. In Matras Y., A. Tenser, (a cura di), *The Palgrave Handbook of Romani Language and Linguistics*. London: Palgrave MacMillan, pp. 13-46.
- Blommaert J., 2010. *The Sociolinguistics of Globalization*. Cambridge: Cambridge University Press.
- Blommaert J., 2013. *Ethnography, Superdiversity and Linguistic Landscapes: Chronicles of Complexity*. Bristol: Multilingual Matters.
- Blommaert J., A. Backus, 2013. Superdiverse Repertoires and the Individual. In Saint-Georges, I.d., Weber JJ. (a cura di), *Multilingualism and Multimodality. The Future of Education Research*. SensePublishers, Rotterdam.
- Boeschoten H., A. Backus, 1997. Code-switching and ongoing linguistic change. In *Turkic Languages* 1: 41-62.
- Bontempelli S., 2022. *I rom. Una storia*. Roma: Carocci Editore.

- Bullock B., J. A. Toribio, J. Serigos, G. Guzman, 2021. Processing multilingual data. In Matras Y., A. Tenser (a cura di), *The Routledge Handbook of Language Contact*. London: Routledge, pp. 7-27.
- Carling G., L. Lindell, G. Ambrazaitis, 2013. *Scandoromani – Remnants of a Mixed Language*. Leiden: Brill.
- Cerruti M., R. Regis, 2015. Dal discorso alla norma: prestiti e calchi tra i fenomeni di contatto linguistico. In *Vox Romanica* 74: 20-45.
- Ciccolone S., 2014. Classificare il code mixing: una reinterpretazione dei parametri di constituency del modello di Muysken. In *Linguistica e Filologia*, 34: 95-134.
- Ciccolone S., S. Dal Negro, 2021. *Comunità bilingui e lingue in contatto. Uno studio sul parlato bilingue in Alto Adige*. Cesena/Bologna: Caissa.
- Ćirković S., M. Mirić, 2022. Serbian Loanverb Adaptation in the Gurbet Romani Variety in Eastern Serbia. In *Mediterranean Language Review* 29: 89-127.
- Clyne M., 2003. *Dynamics of Language Contact. English and Immigrant Languages*. Cambridge, Cambridge University Press.
- Colocci A., 1889. *Gli Zingari. Storia di un popolo errante*. Torino: Loescher.
- Cortellazzo M., 1975. Voci zingare nei gerghi padani. In *Linguistica (In Memoriam Stanko Škerli Oblata)*, 15(1): 29-40.
- Cosentino M., 2023. U zingarišku: il dialetto neolatino parlato dalla comunità Rom stanziata di Cosenza. In *L'identità/le identità linguistiche e culturali delle comunità Rom italiane: storia e prospettive - Cosenza, Villa Rendano, APS "Lav Romanò"/Fondazione ETS "Attilio e Elena Giuliani"*.
- Dal Negro S., G. Iannàccaro, 2003. "Qui parliamo tutti uguale, ma diverso". Repertori complessi e interventi sulle lingue. In Valentini A., G. Bernini, P. Molinelli, P. Cuzzolin (a cura di), *Ecologia linguistica*. Roma: Bulzoni, pp. 431-450.
- Dal Negro S., 2005a. Il code-switching in contesti minoritari soggetti a regressione linguistica. In *Italian Journal of Linguistics*, 17(1): 151-178.
- Dal Negro S., 2005b. Lingue in contatto: il caso speciale dei segnali discorsivi. In Banti G., A. Marra, E. Vineis (a cura di), *Atti del 4° congresso di studi dell'Associazione Italiana di Linguistica Applicata*. Perugia: Guerra, pp. 73-88.
- Dal Negro S., 2006. Sistemi linguistici a contatto: il caso di Issime. In *Diglossie et interférences linguistiques: néologismes, emprunts, calques. Actes de la conférence annuelle sur l'activité scientifique du Centre d'études francoprovençales*, Saint Nicolas, 17(1): 157-178.
- Dal Negro S., 2009. Fenomeni di contatto tra discorso e sistema. In *Alloglossie e comunità alloglotte nell'Italia contemporanea: teorie, applicazioni e prospettive. Atti del XLI Congresso internazionale di studi della società di linguistica italiana (SLI)*. Roma: Bulzoni, pp. 81-94.
- Dal Negro S., 2011a. Tedesco di contatto in Italia. In Fazzini E. (a cura di), *Il tedesco superiore: Tradizione scritta e varietà parlate*. Alessandria: dell'Orso, pp. 179-199.

- Dal Negro S., 2011b. *Überdacht o dachlos? Di vitalità e di coperture linguistiche*. In B. Moretti, E. Maria Pandolfi & M. Casoni (a cura di), *Vitalità di una lingua minoritaria. Aspetti e proposte metodologiche*. Bellinzona: Osservatorio Linguistico della Svizzera Italiana, pp. 183-210.
- Dal Negro S., 2021. Documentazione linguistica e lingue in contatto. In Favilla M. A., S. Machetti (a cura di), *Lingue in contatto e linguistica applicata: individui e società*, pp. 9-23.
- Dal Negro S., 2024. Sociolinguistic typology and language contact in Northern Italy. In B. Hans-Bianchi, C. Truppi, B. Vogt (a cura di), *Speakers and Structures in Language Contact: Pluralistic Approaches to Change and Variation* (Language Contact and Bilingualism, 31). Berlin/Boston: Mouton de Gruyter, pp. 199-224.
- De Vaux de Foletier F., 2010. *Mille anni di storia degli Zingari*, Milano: Jaca Book (ed. or. *Mille ans d'histoire des Tsiganes*, Paris 1970).
- Dell'Aquila V., Iannàccaro G. 2004. *La pianificazione linguistica. Lingua, società e istituzioni*. Roma: Carocci.
- Dell'Aquila V., G. Iannàccaro, 2006. *Survey Ladins. Usi linguistici nelle Valli Ladine*. Trento: Regione Autonoma Trentino-Alto Adige.
- Duberti N., 2010. *Trin kamlé tikné: studenti sinti a Rocca de' Baldi (Cuneo)*. In *Bolettino dell'atlante linguistico italiano*, 3(34): 37-78.
- Elšík V., 2000. *Romani nominal paradigms: Their structure, diversity, and development*. In Elšík V., Y. Matras (a cura di), *Grammatical relations in Romani: The noun phrase*. Amsterdam: John Benjamins, pp. 9-30.
- Elšík V., 2007. Grammatical borrowing in Hungarian Rumungro. In Y. Matras & J. Sakel (a cura di), *Grammatical borrowing in cross-linguistic perspective*. Berlin: Mouton de Gruyter, pp. 261–282.
- Elšík V., M. Beníšek, 2020. Romani Dialectology. In Matras Y., A. Tenser (a cura di), *The Palgrave Handbook of Romani Language and Linguistics*. London: Palgrave MacMillan, pp. 389-428.
- Elšík V., Y. Matras, 2006. *Markedness and Language Change: The Romani Sample*. Berlin: de Gruyter.
- Ferguson C.A., 1959. Diglossia, in *WORD*, 15(2): 325-340.
- Field F., 2002. *Linguistic borrowings in bilingual contexts*. Amsterdam & Philadelphia: Benjamins.
- Fiorentini I., 2017. *Segnali di Contatto. Italiano e ladino nelle valli dolomitiche*. Milano: Franco Angeli.
- Fiorentini I., 2022. *Sociolinguistica delle minoranze in Italia. Un'introduzione*. Roma: Carocci Editore.
- Fiorentini I., Forlano M., 2025. “Duj bers et’u vavar tredici anni”. Il continuum tra *insertion* e *alternation* nel parlato bilingue italiano-romaní. In Cerruti M., C. Onesti C., R. Regis, D. Ricca (a cura di), *Atti del LVI Congresso Internazionale di Studi della Società di Linguistica Italiana (Torino, 14-16 settembre 2023)*, pp. 243-264. Milano: Officinaventuno.
- Fishman J.A., 1991, *Reversing language shift: theoretical and empirical foundations of assistance to threatened languages*. Avon, UK: Multilingual Matters.

- Forlano M., 2025. Verbal borrowing in Lombard Sinti. A corpus-driven approach. Paper presentato presso 16th *International Conference on Romani Linguistics*, Austrian Academy of Sciences, Vienna, 23-25 settembre 2025.
- Franzese S., 2023 [2001, 1985]. *Grammatica di sinto piemontese*. Disponibile da: <http://www.progettoniglo.org>.
- Friedman V., 2017. Seven varieties of Arli: Skopje as a center of convergence and divergence of Romani dialects. In *Romani Studies*, 27(1): 29-45.
- Friedman V., 2020. The Impact of Turkish on Romani. In Y. Matras & A. Tenser (a cura di), *The Palgrave Handbook of Romani Linguistics*. London: Palgrave MacMillan, pp. 231-260.
- Friedman V., B. Joseph, 2024. *The Balkan Languages*. Cambridge: Cambridge University Press.
- Galli E., 1996. *Dizionario pavese-italiano*. Pavia: Studia ghisleriana; pubblicazioni dell'Associazione alunni del Collegio Ghislieri in Pavia.
- Gamella J. F., C. Fernandez Ortega, I. X. Adiego, 2016. The long agony of Hispanoromani; The remains of Caló in the speech of Spanish Gitanos. In *Romani Studies*, 25(1): 53-93.
- Gardani F., 2008. *Borrowing of Inflectional Morphemes*. Frankfurt: Peter Lang.
- Gardner-Chloros P., 2009. *Code-switching*. Cambridge: Cambridge University Press.
- Gilliat-Smith B.J., 1915. A report on the Gypsy tribes of North East Bulgaria. In *Journal of the Gypsy Lore Society*, 9: 69-109.
- Ghezzi C., Molinelli P., 2015. Segnali allocutivi di richiamo: percorsi pragmatici e sviluppi diacronici tra latino e italiano. In *Cuadernos de Filología Italiana* 22: 21-47.
- Goria E., 2018. *Inglese e spagnolo a Gibilterra. Le dinamiche del discorso bilingue*. Bologna: Caissa Italia.
- Goria E., Gasparini F., 2025. Towards an Oral Corpus for Heritage Piedmontese. In *Oral Archives Journal*, 1, 35-55.
- Grosjean F., 2001. The bilingual's language modes. In Nicol J. L. (a cura di), *One mind, two languages. Bilingual language processing*. Oxford: Blackwell, pp. 1-22.
- Gumperz J., 1972. Introduction. In Gumperz J., Hymes D. (a cura di), *Directions in Sociolinguistics: The Ethnography of Communication*. London: Blackwell, pp. 1-25.
- Gumperz J., 1982. *Discourse strategies*. Cambridge: Cambridge University Press.
- Hakimov N., M. Rießler, 2021. Partial fusion in long-term bilingualism: The case of vernacular Kildin Saami. In *International Journal of Bilingualism*, 25(2), 401-424.
- Halwachs D. W., 2003. *The Changing Status of Romani in Europe*. In Hogan-Brun G., S. Wolff, (a cura di), *Minority Languages in Europe*. London: Palgrave MacMillan, pp. 192-207.
- Halwachs D. W., 2011. Language planning and media: The case of Romani. In *Current Issues in Language Planning*, 12(3): 381-401.

- Halwachs D. W., 2020. Language Policy and Planning in Romani. In Matras Y., A. Tenser (a cura di), *The Palgrave Handbook of Romani Language and Linguistics*. London: Palgrave MacMillan, pp. 429-457.
- Haspelmath M., 2008. Loanword Typology: Steps towards a systematic cross-linguistic study of lexical borrowability. In Stolz T., D. Bakker, R. Salas Palomo (a cura di), *Aspects of language contact: New theoretical, methodological and empirical findings with special focus on Romanisation processes*. Berlin: Mouton de Gruyter, pp. 43-62.
- Haspelmath M., 2009. Lexical borrowing: Concepts and issues. In Haspelmath M., U. Tadmor (a cura di), *Loanwords in the World's Languages: A Comparative Handbook*. Berlin: Mouton de Gruyter, pp. 35-54.
- Heath J., 1984. Language contact and language change. *Annual Review of Anthropology*, 13: 367–384.
- Iannàccaro G., V. Dell'Aquila, 2000. Alla ricerca della comunità linguistica: spunti dal concetto di lingua madre. In Marcato G. (a cura di), *Isole linguistiche? Per un'analisi dei sistemi in contatto*. Padova: Unipress, pp. 361-371.
- Jorgensen J.N., 2008. Polylingual languaging around and among children and adolescents. *International Journal of Multilingualism* 5(3):161-176.
- Labov W., 1972. Some principles of linguistic methodology. In *Language in Society*, 1:97-120.
- Laganà S., 2024. *I sinti di Pavia: un approccio interculturale in ambito scolastico ed educativo*. Tesi di laurea inedita, Università degli Studi di Milano – Bicocca.
- Leggio D.V., 2011. The dialect of the Mitrovica Roma. In *Romani Studies*, 21(1): 57-114.
- Leggio D.V., 2015. Radio Romani Mahala: Romani identities and languages in a virtual space. In Alietti A., M. Olivera, V. Riniolo (a cura di), *Virtual citizenship? Roma communities, inclusion policies, participation and ICT tools*. Milano: McGraw-Hill Education, pp. 95-114.
- Leggio D. V., 2020. Romani on the Internet. In Matras Y., A. Tenser (a cura di), *The Palgrave Handbook of Romani Language and Linguistics*. London: Palgrave MacMillan, pp. 515-537.
- Lewis P., G. Simons, 2010. Assessing endangerment: Expanding Fishman's GIDS. In *Revue Roumaine de Linguistique*, 55(2):103–120.
- Wei L., 2018. Translanguaging as a practical theory of language. *Applied Linguistics*, 39(1): 9–30.
- Lo Baido M.C., 2024. *Forms and Functions of Meta-Discourse. The Case of Comment Clauses in Present-Day Italian*. Boston/Berlin: Mouton de Gruyter.
- Lupica Spagnolo M., 2019. *Storie di confine. Biografie linguistiche e ristrutturazione dei repertori tra Alto Adige e Balcani*. Stuttgart: Franz Steiner Verlag
- Matras Y., 1994. Structural Balkanisms in Romani. In Reiter N., U. Hinrichs, J. van Leeuwen- Turnocová (a cura di), *Sprachlicher Standard und Substandard in Südosteuropa und Osteuropa*. Wiesbaden: Harrassowitz, pp. 195-210.

- Matras Y., 1998a. Utterance modifiers and universals of grammatical borrowing. In *Linguistics*, 36(2): 281-331.
- Matras Y., 1998b. Para-Romani revisited. In Matras Y. (a cura di), *The Romani element in non- standard speech*. Wiesbaden: Harrassowitz.
- Matras Y., 1999. Writing Romani: The Pragmatics of Codification in a Stateless Language. In *Applied Linguistics*, 20(4): 481-502.
- Matras Y., 2002. *Romani: A Linguistic Introduction*. Cambridg: Cambridge University Press.
- Matras Y., 2005a. The classification of Romani dialects: A geographic-historical perspective. In Schrammel B., D. Halwachs, G. Ambrosch (eds), *General and applied Romani linguistics: Proceedings from the 6th International Conference on Romani Linguistics*. Munich: Lincom, pp. 7- 22.
- Matras Y., 2005b. *The status of Romani in Europe*. Report commissioned by the Council of Europe's Language policy Division.
- Matras Y., 2007. The borrowability of structural categories. In Matras Y., J. Sakel (a cura di), *Grammatical Borrowing in Cross-Linguistic Perspective*. Berlin: Mouton de Gruyter, pp. 31-73.
- Matras Y., 2010. *Romani in Britain: The afterlife of a language*. Edinburgh: Edinburgh University Press.
- Matras Y., 2012. *A grammar of Domari*. Berlin: Mouton de Gruyter.
- Matras Y., 2020 [2009]. *Language contact*. Cambridge: Cambridge University Press.
- Matras Y., J. Sakel, 2007. *Grammatical Borrowing in Cross-Linguistic Perspective*. Berlin, New York: de Gruyter.
- Matras Y., E. Adamou, 2020. Romani and Contact Linguistics. In Matras Y., A. Tenser (a cura di), *The Palgrave Handbook of Romani Language and Linguistics*. London: Palgrave MacMillan, pp. 329-352.
- Matras Y., E. Adamou, 2021. Borrowing. Adamou E., Y. Matras (a cura di), *The Routledge Handbook of Language Contact*. London: Routledge, pp. 237-251.
- Matras Y., E. Adamou, 2023. Word classes in language contact. In van Lier E. (ed), *The Oxford Handbook of Word Classes*. Oxford: Oxford University Press, pp. 887-898.
- Mazzoli M., Sippoli E., 2021. *New Perspectives on Mixed Languages*. Boston/Berlin: Mouton de Gruyter.
- McConvell P., F. Meakins, 2005. Gurindji Kriol: A mixed language emerges from code-switching. *Australian Journal of Linguistics*, 25(1): 9-30.
- McConvell P., 2007. Mixed languages as outcomes of code-switching: recent examples from Australia and their implications. In *Journal of Language Contact*: 187-212.
- Meakins F., 2013. Mixed Languages. In P. Bakker, Y. Matras (a cura di), *Contact languages*. Boston & Berlin: Walter de Gruyter, pp. 159-228.

- Meli G., 2022. *Il dialetto degli shinte rosengre. Esame delle fonti e analisi della morfologia tra sincronia e diacronia*. Alessandria: Edizioni dell'Orso.
- Meli G., 2023. L'integrazione morfologica dei prestiti romanzi in sinto piemontese di Francia e in sinto lombardo. In Faraoni V., L. Filipponio, T. Paciaroni, S. Schmid (a cura di), *Prospettive di ricerca in linguistica italiana e romanza: studi offerti a Michele Loporcaro dagli allievi e dai collaboratori zurighesi*. Pisa: Edizioni ETS, pp. 407-424.
- Membretti A., 2009. Rom e sinti a Pavia: la negazione istituzionale di una questione pubblica. In Vitale T. (a cura di), *Politiche possibili. Abitare le città con i rom e i sinti*. Roma: Carocci Editore, pp. 69-79.
- Meyerhoff M., 2014. Borrowing in Apparent Time: With some comments on attitudes and universals. *University of Pennsylvania Working Papers in Linguistics*, 20(2): 121-128.
- Miklosich F., 1872-1880. *Über die Mundarten und Wanderungen der Zigeuner Europas. I–XII*. Vienna: Karl Gerold's Sohn.
- Milroy L., 2002. Social Networks. In Chambers J. K., P. Trudgill, N. Schilling-Estes (a cura di), *The Handbook of Language Variation and Change*. Malden, Mass: Blackwell, pp. 549-572.
- Mirić M., 2025. Literacy practices in Romani: pilot data from Serbia. In M. Dinić Marinković, N. Panić Cerovski, B. Kovačević (a cura di), *Proceedings of The Eighth International Congress of Applied Linguistics Today – Modern Approaches to Old and New Challenges, 23–25 May 2025, Belgrade*, pp. 337-358.
- Mirić M., S. Ćirković, 2022. *Гурбетски ромски у контакту: анализа балканизама и позајмљеница из српског језика*. Београд: Балканолошки институт САНУ. (*Gurbetski romski u kontaktu analiza balkanizama i pozajmljenica iz srpskog jezika*. Beograd: Balkanoloski Institut.)
- Mirić M., V. Sokolovska, A. Sorescu-Marinković, 2024. Mother tongue in Serbia: A speakers' perspective on the meaning of the concept. In *Open Linguistics* 10/1, pp. 1–24.
- Moravcsik E., 1978. Language Contact. In Greenberg J. H. (a cura di), *Universals of Human Language, I. Method and Theory*. Stanford: Stanford University Press, pp. 93-122.
- Muysken P., 1981. Halfway between Quecha and Spanish: The Case for Relexification. In *Historicity and Variation in Creole Studies*, pp. 52-78.
- Muysken P., 2000. *Bilingual Speech: a typology of code-mixing*. Cambridge: Cambridge University Press.
- Muysken P., 2010. Scenarios for Language Contact. In Hickey R. (a cura di), *The Handbook of Language Contact*. Oxford: Blackwell, pp. 265-281.
- Muysken P., 2013. Language contact outcomes as the result of bilingual optimization strategies. In *Bilingualism: Language and Cognition*, 16(4): 709-730.
- Myers-Scotton C., 1993. *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: Clarendon Press.

- Myers-Scotton C., 1998. A Way to Dusty Death: The Matrix Language Turnover Hypothesis. In Grenoble L., L. L. Whaley (a cura di), *Endangered languages: Language loss and community response*. Cambridge: Cambridge University Press, pp. 289-316.
- Myers-Scotton C., 2002. *Contact linguistics, bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.
- Myers-Scotton C., J. L. Jake, 2000. Four types of morpheme: Evidence from aphasia, code-switching and second-language acquisition. In *Linguistics*, 38(6): 1053-1100.
- Myers-Scotton C., J. L. Jake, 2014. Nonfinite verbs and negotiating codeswitching: Implications for a language production model. In *Bilingualism, Language and Cognition*, 17, 511–525.
- Myers-Scotton C., J. L. Jake., 2016. Revisiting the 4-M model: Codeswitching and morpheme election at the abstract level. In *International Journal of Bilingualism*, 21(3): 340-366.
- Nagy N., 2011. Lexical change and language contact: Faetar in Italy and Canada. In *Journal of Sociolinguistics*, 15: 366-382.
- Novi Chavarria E., 2007. *Sulle tracce degli Zingari*. Napoli: Guida.
- Padure C., E. Adamou, 2019. Matter replication phenomena in Romani spoken in Romania. halshs02420669f
- Padure C., Adamou E, 2024. The activity spheres framework applied to language policy: Dispositions towards Romani revitalization in Romania. *Identities, Representations and Policies: State-Minority Relations in the 20th and 21st Centuries*. fhalshs-03268416f
- Partisani S., 1984. I mucini. In *Lacio Drom* 20(4): 23-24.
- Pasquali P.S., 1935, Romani Words in Italian Slangs. In *Journal of the Gypsy Lore Society*, 3(14): 44-51.
- Pellis U., 1936. Il rilievo zingaresco all'Annunziata di Giulianova (Teramo). In *Bollettino dell'Atlante Linguistico Italiano*, 2.
- Perta C., 2020. Sociolinguistic Aspects and Language Contact. In *Journal of Language Contact*, 13(2): 271–288
- Perta C., L. Tramutoli, 2020. Strategie discorsive dei bilingui e contesto sociolinguistico. Vecchie e nuove minoranze a confronto. In *Lingue e linguaggi*, 39: 311-327.
- Pennycook A., 2016. Language Policy and Local Practices. In O. García, N. Flores, M. Spott (a cura di), *The Oxford Handbook of Language and Society*. Oxford Handbooks Online,
- Piasere L., 1992. Considerazioni sulla presenza zingara nel nord Italia nel XIX secolo sulla base di alcuni documenti linguistici (con appendici di testi di James Pincherle). In *Ce fastu?*, 58(2): 233- 267.
- Piasere L., 1996. Sigismondo Caccini e gli Šinte Rosengre. In *Italia Romani*, vol. 1. Roma: CISU.
- Piasere L., 2009. *I Rom d'Europa. Una storia moderna*. Roma: Laterza.
- Polinsky M., 2018. *Heritage Languages and Their Speakers*. Cambridge: Cambridge University Press.
- Pontrandolfo S., 2013. *Rom dell'Italia meridionale*. Roma: CISU.

- Poplack S., 1980. Sometimes I'll start a sentence in Spanish y termino en espanol. In *Linguistics*, 18: 581-618.
- Poplack S., 1988 Contrasting patterns of code-switching in two communities. In M. Heller (a cura di), *Codeswitching. Anthropological and Sociolinguistic Perspectives*. Berlin: Mouton de Gruyter, pp. 215-224.
- Poplack S., 2018. *Borrowing. Loanwords in the Speech and in the Grammar*. New York: Oxford University Press.
- Poplack S., D. Sankoff, C. Miller, 1988. The social correlates and linguistic processes of lexical borrowing and assimilation. In *Linguistics*, 26: 47-104.
- Poplack S., M. Meechan, 1998. Introduction: How Languages Fit Together in Codemixing. In *International Journal of Bilingualism*, 2(2): 127-138.
- Poplack S., N. Dion, 2012. Myths and Facts about Loanword Development. In *Language Variation and Change*, 24(3): 279-315.
- Poplack S., S. Robillard, N. Dion, J. Paolillo, 2020. Revisiting phonetic integration in bilingual borrowing. In *Language*, 96(1): 126-159.
- Reitinger D., 2023. The origin of the self-appellation Sinti: A historical and linguistic examination. In *Romani Studies* 33 (2): 157-87.
- Rosignoli A., 2011. *Flagging in English-Italian Code-Switching*. Tesi di dottorato inedita, University of Bangor.
- Rossi A., 2020. Metodo militante nel contrasto alla dispersione scolastica dei minori sinti. In S. Boni, A. Koensler, A. Rossi (a cura di), *Etnografie militanti: Prospettive e dilemmi*. Meltemi, pp. 203-216.
- Rusakov A., 2001. The North Russian Romani dialect: interference and code switching. In Dahl Ö., M. Koptjevskaja-Tamm (a cura di), *Circum-Baltic languages*. Amsterdam & Philadelphia: Benjamins, pp. 313-338.
- Rüdiger J. C. C., 1782. Von der Sprache und Herkunft der Zigeuner aus Indien. In Leipzig S. E., P. G. Kummer (a cura di), *Neuester Zuwachs der teutschen, fremden und allgemeinen Sprachkunde in eigenen Aufsätzen*, pp. 37-84.
- Sankoff D, Poplack S., Vanniarajan S., 1990. The case of the nonce loan in Tamil. In *Language Variation and Change*, 2(1): 71-101.
- Scala A., 2012. Purché la lingua non sia una sola... Trasformazione dei repertori e conservazione del plurilinguismo presso i Sinti italiani dall'Unità ad oggi. In Telmon T., G. Raimondi, L. Revelli (a cura di), *Coesistenze linguistiche nell'Italia pre- e postunitaria, Atti del 45° Congresso della società internazionale della Società di linguistica italiana (SLI): Aosta, Bard, Torino, 26-28 settembre 2011*. Roma: Bulzoni.

- Scala A., 2014. La componente romaní nel baccà di Guardiagrele: rileggendo le raccolte di Ugo Pellis ed Ernesto Giammarco. In F. Cugno, L. Mantovani, M. Rivoira, M. S. Specchia (a cura di), *Studi in onore di Lorenzo Massobrio*. Torino: Istituto dell'Atlante Linguistico Italiano, pp. 909-921.
- Scala A., 2015. “Se proprio dobbiamo scrivere, almeno facciamolo come gli altri italiani”. I Sinti dell'Italia settentrionale e la grafizzazione della loro lingua. In Dal Negro S., F. Guerini, G. Iannàccaro (a cura di), *Elaborazione ortografica delle varietà non standard. Esperienze spontanee in Italia e all'estero*. Bergamo: Bergamo University Press.
- Scala A., 2017. I numerali da 1 a 10 in sinto lombardo. In Prada M., G. Sergio (a cura di), *Italiani di Milano: Studi in onore di Silvia Morgana*. Milano: Ledizioni, pp. 789-797.
- Scala A., 2020a. La romaní. In Fiorentini I., C. Gianollo, N. Grandi (a cura di), *La classe plurilingue*. Bologna: Bononia University Press.
- Scala A., 2020b. Romani Lexicon. In Matras Y., A. Tenser (a cura di), *The Palgrave Handbook of Romani Language and Linguistics*. London: Palgrave MacMillan.
- Scala A., 2020c. La romaní in Italia tra rappresentazione e legittimazione, In *Language Problems & Language Planning*, 44(3): 346-370.
- Scala A., 2021. A Lombard Sinti ethno-text on mourning and marriage. In *Ethnorêma*, 16: 45-57.
- Scala A., 2023a. Per una migliore comprensione dei rapporti storici tra sinto piemontese di Piemonte e sinto piemontese di Francia: alcune considerazioni a partire da evidenze morfologiche e lessicali. In *Quaderns d'Italià*, 28: 175-188
- Scala A., 2023b. Di qua e di là dal confine: la romaní dei sinti piemontesi di Piemonte e quella dei sinti piemontesi di Francia a confronto, in S. Dal Negro, D. Mereu (a cura di), *Confini nelle lingue e tra le lingue. Atti del LV Congresso Internazionale di Studi della Società di Linguistica Italiana (Bressanone, 8-10 settembre 2022)*. Milano: Officina Ventuno, pp. 47-58.
- Scala A., 2023c. La romaní d'Abruzzo: una tessera linguistica indoaria nello spazio culturale aprutino (Con qualche nota in margine alla lingua dei canti registrati da Giorgio Nataletti), in D. Di Virgilio (a cura di), *Musiche tradizionali in Abruzzo: le registrazioni di Giorgio Nataletti (1948-49)*. Roma: Squilibri, pp. 78-99.
- Seifart F., N. Evans, H. Hammarström, S.C. Levinson, 2018. Language documentation twenty-five years on. In *Language*, 94, e324–e345.
- Simonelli C., 1996. «Ci sarà annunciato Dio da uno di noi»: lo sviluppo del movimento “vangelista” in una comunità di sinti. In L. Piasere (a cura di), *Italia Romanì*. CISU, Roma, pp. 71-92.
- Sonnemann A., 2022. *Language Families in Contact: The Mutual Impact of Slavic and Romani*. Berlin, Boston: De Gruyter Mouton.
- Soravia G., 1977. *Dialetti degli zingari italiani*. Pisa: Pacini.

- Soravia G., C. Fochi, 1995. *Vocabolario Sinottico delle Parlate Zingare in Italia*. Roma: Centro Studi Zingari.
- Sorrenti E., 2014. Italiano e sinto lombardo a confronto: somiglianze, divergenze e prospettive didattiche. In *Italiano LinguaDue*, 1: 117-147.
- Spinelli S., 1993. *Romanipè. Raccolta di poesia zingara*, Solfanelli Edizioni, Chieti.
- Stolz T., 2007. Allora: On the recurrence of function-word borrowing in contact situations with Italian as donor language. In Rehbein J., C. Hohenstein, L. Pietsch (a cura di), *Connectivity in grammar and discourse*. Amsterdam: Benjamins, pp. 75-99
- Tadmor U., 2009. Loanwords in the World's Languages: Findings and results. In Haspelmath M., U. Tadmor (a cura di), *Loanwords in the World's Languages*. Berlin: Mouton de Gruyter, pp. 35-54.
- Tagliavini C., A. Menarini, 1938. Voci zingare nel gergo bolognese. In *Archivium Romanicum*,
- Tauber E., 2008. "Do you remember the time we went begging and selling" The Ethnography of Transformations in Female Economic Activities and Its Narrative in the Context of Memory and Respect among the Sinti in North Italy. In F. Jacobs, J. Ries (a cura di), *Roma-/Zigeunkulturen in neuen Perspektiven*, Leipzig, Leipziger Universitätsverlag GmbH, pp. 155-176.
- Thomason S. 2001. Linguistic areas and language history. In Haspelmath M., Koenig E., Oesterreicher W., Raible W. (a cura di), *Language typology and language universals, Sprachtypologie und sprachliche Universalien: An international handbook*. Berlin & New York: Walter de Gruyter.
- Thomason S. G., T. Kaufman, 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Torres Cacoullous R., N. Dion, D. LaCasse, S. Poplack, 2022. How to mix. Confronting "mixed" NP models and bilinguals' choices. In *Linguistic Approaches to Bilingualism*, 12(5): 628-656.
- Toso F., 2008, *Le minoranze linguistiche in Italia*. Bologna: Il Mulino
- Treffers-Daller J., 1994. *Mixing two languages: French-Dutch contact in a comparative perspective*. Berlin: Mouton de Gruyter.
- Treffers-Daller J., 1999. Borrowing and shift-induced interference: Contrasting patterns in French-Germanic contact in Brussels and Strasbourg. In *Bilingualism: Language and Cognition*, 2(1): 1-22.
- Treffers-Daller J., 2023. The Simple View of borrowing and code-switching. In *International Journal of Bilingualism* 29(2): 347-370.
- Trevisan P., 2005. *Storie e vite dei sinti dell'Emilia*. Roma, CISU.

- Tribulato, C., 2019. *Qui in mezzo a noi. I sinti nello spettacolo viaggiante*. Tesi di Dottorato inedita, Università di Venezia/Padova/Verona.
- Tribulato C., 2020. Dritto in contatto: elementi romaní nel gergo di una comunità girovaga italiana. *Argotica* 1 (9): 81-102.
- Tribulato C., B. Pastori, 2022. *Il diario di Jucki*. Padova: Edizioni Diodati.
- Van den Heuvel W., 2020. Romani Bible Translation and the Use of Romani in Religious Contexts. In Matras Y., A. Tenser (a cura di), *The Palgrave Handbook of Romani Language and Linguistics*. London: Palgrave MacMillan, pp. 459-486.
- Van Hout R., P. Muysken, 1994. Modelling lexical borrowability. In *Language Variation and Change*, 6: 39-62.
- Van Gijn R., H. Ruch, M. Wahlström, A. Hasse (a cura di), 2023. *Language contact. Bridging the gap between individual interactions and areal patterns*. Berlin: Language Science Press.
- Vertovec S., 2007. Super-diversity and its implications. In *Ethnic and Racial Studies*, 30: 1024-1054.
- Vietti A., 2014. Minoranze non territoriali. In Iannàccaro G. (a cura di), *La linguistica italiana all'alba del terzo millennio*. Roma: Bulzoni, pp. 341-368.
- Weinreich U., 1953. *Languages in contact*. The Hague: Mouton.
- Whitney W. D., 1881. On Mixture in Language. *Transactions of the American Philological Association* (1869-1896) 12: 1-26.
- Winford D., 2013. Social factors in contact languages. In P. Bakker, Y. Matras (a cura di), *Contact languages*, 363-416. Boston & Berlin: Walter de Gruyter.
- Wohlgemuth J., 2009. *A Typology of Verbal Borrowings*. Berlin: Mouton de Gruyter.
- Zenner E., A. Backus, E. Winter-Froemel (a cura di), 2019. *Cognitive Contact Linguistics: Placing Usage, Meaning and Mind at the Core of Contact-Induced Variation and Change*. Berlin: Mouton de Gruyter.
- Ziano C., P. Trevisan, 2025. A Veneto Dialect in Emilia: Linguistic traces of Sinti mobility across Northern Italy. Paper presentato presso *16th International Conference on Romani Linguistics*, Vienna, 23-25 Settembre 2025.

Appendice 1: Il modulo di consenso informato



UNIVERSITÀ
DI PAVIA

GENTILE SIGNORE / SIGNORA,

LEI E' INVITATO/A A PARTECIPARE A UNA RICERCA INTITOLATA "SINTI LOMBARDI: UN'INDAGINE LINGUISTICA", ESEGUITA DA ME (MARCO FORLANO) PER L'UNIVERSITA' DI PAVIA.

QUESTA RICERCA HA COME OBIETTIVO STUDIARE LA LINGUA PARLATA DAI SINTI LOMBARDI. PER RAGGIUNGERE QUESTO OBIETTIVO, VORREI FARE ALCUNE REGISTRAZIONI IN CUI LE PERSONE REGISTRATE PARLERANNO IN SINTO E RISPONDERANNO A DOMANDE LEGATE ALLA LINGUA SINTA.

QUINDI VORREI CHIEDERE IL CONSENSO, PER LEI O PER EVENTUALI MINORI A SUO CARICO, DI PARTECIPARE ALLE REGISTRAZIONI.

LA PARTICIPAZIONE ALLA RICERCA E' VOLONTARIA E NON COMPORTA NESSUN RISCHIO E DISAGIO. LE PERSONE CHE PARTECIPERANNO ALLA RICERCA RIMARRANNO ANONIME. POTRANNO RICHIEDERE IN QUALSIASI MOMENTO DI NON PARTECIPARE PIU' E DI NON UTILIZZARE I PROPRI DATI.

PER QUALSIASI INFORMAZIONE O CHIARIMENTO, MI POTRA' CONTATTARE AL NUMERO: [REDACTED].

(firma)

(firma di chi raccoglie il consenso)

Appendice 2: Il questionario sociolinguistico

1. Quale lingua consideri la tua lingua madre? *(è possibile più di una risposta)*
2. Quali altre lingue e/o dialetti parli? *(è possibile più di una risposta)*
3. In quali lingue e/o dialetti parli (o parlavi) con tua madre? *(è possibile più di una risposta)*
4. In quali lingue e/o dialetti parli (o parlavi) con tuo padre? *(è possibile più di una risposta)*
5. In quali lingue e/o dialetti parli con fratelli e/o sorelle? *(è possibile più di una risposta)*
6. In quali lingue e/o dialetti parli con il tuo partner/coniuge? *(è possibile più di una risposta)*
7. In quali lingue e/o dialetti parli con i tuoi figli? *(è possibile più di una risposta)*
8. In quali lingue e/o dialetti parli con i parenti anziani? *(è possibile più di una risposta)*
9. In quali lingue e/o dialetti parli con i bambini della tua comunità? *(è possibile più di una risposta)*
10. Che lingue e/o dialetti usi con i tuoi amici sinti? *(è possibile più di una risposta)*
11. Che lingue e/o dialetti usi scuola o negli uffici pubblici? *(è possibile più di una risposta)*
12. Che lingue e/o dialetti usi con i commercianti della città? *(è possibile più di una risposta)*
13. Che lingue e/o dialetti usi con gli amici non sinti? *(è possibile più di una risposta)*
14. Quali lingue e/o dialetti hai imparato per primi da piccolo? *(è possibile più di una risposta)*
15. Quali lingue e/o dialetti parlavano tra loro i tuoi genitori? *(è possibile più di una risposta)*
16. Il sinto parlato dai tuoi nonni era diverso rispetto a quello che si parla oggi?
 - a. Molto
 - b. Abbastanza
 - c. Poco
 - d. No
 - e. Non saprei
17. Il sinto parlato dai tuoi genitori era diverso rispetto a quello che si parla oggi?
 - a. Molto
 - b. Abbastanza
 - c. Poco
 - d. No
 - e. Non saprei
18. Quale lingua o dialetto usi di più nell'arco di una giornata? *(è possibile più di una risposta)*
19. In quali lingue pensi tra te e te? *(è possibile più di una risposta)*
20. In quale lingua parli al telefono? *(è possibile dare più di una risposta)*
21. In quale lingua scrivi messaggi sul cellulare? *(è possibile più di una risposta)*
22. Come capisci altri dialetti sinti?
 - a. Bene
 - b. Abbastanza bene
 - c. Poco
 - d. Niente
23. Hai parenti o amici che parlano altri dialetti sinti? Se sì, in che lingua o dialetto parli con loro? *(è possibile più di una risposta)*
- 24.

Competenza	Da 1 a 10
Quanto sai parlare il sinto?	
Quanto sai leggere il sinto?	
Quanto sai scrivere il sinto?	

25.

Competenza	Da 1 a 10
Quanto sai parlare l'italiano?	
Quanto sai leggere l'italiano?	

Quanto sai scrivere l'italiano?	
---------------------------------	--

26. Una lingua scritta per il futuro del sinto sarebbe:
- a. Una risorsa
 - b. Indifferente
 - c. Un pericolo
 - d. Non so
27. L'insegnamento del sinto a scuola sarebbe:
- a. Una risorsa
 - b. Indifferente
 - c. Un pericolo
 - d. Non so
28. Un sito Internet per studiare e fare ricerche sul sinto a disposizione dei gagì sarebbe:
- a. Una risorsa
 - b. Indifferente
 - c. Un pericolo
 - d. Non so
29. Conoscere il sinto per essere sinti è:
- a. Fondamentale
 - b. Importante
 - c. Abbastanza importante
 - d. Poco importante
 - e. Del tutto irrilevante
30. Credi che in sinto sia possibile dire tutto?
- a. Sì
 - b. No
 - c. Non so
31. Prima di andare a scuola, i bambini dovrebbero crescere parlando:
- a. Solo il sinto
 - b. Solo l'italiano
 - c. Il sinto e l'italiano
 - d. Non saprei
32. Tendi a mischiare molto il sinto e l'italiano mentre parli?
- a. Molto
 - b. Abbastanza
 - c. Poco
 - d. Per niente
33. Mischiare il sinto e l'italiano mentre si parla è:
- a. Normale
 - b. Strano
 - c. Da evitare
 - d. Non saprei
34. Sei fiero di sapere parlare il sinto?
- a. Molto
 - b. Abbastanza
 - c. Poco
 - d. Per niente
35. Come vedi il futuro del sinto?
- a. Meglio del presente
 - b. Uguale al presente
 - c. Peggioro del presente
 - d. Non so