

# I migranti ‘golondrinas’ italiani tra fine Ottocento e inizio Novecento: una verifica empirica\*

FRANCESCO DI FILIPPO, STEFANO GALLO

ISMED-CNR Roma

**1. Una categoria.** Nella storiografia delle migrazioni italiane viene spesso ricordata una forma di mobilità dai caratteri assolutamente originali, tanto da affascinare gli studiosi per la sua peculiarità: un migrante transoceanico che partiva e tornava nell’arco di pochi mesi per svolgere lavori agricoli di fatica. Si tratta della cosiddetta emigrazione *golondrina*, diretta dalle campagne italiane verso le estensioni coltivate della pampa argentina: lavoratori che oscillavano con movimenti stagionali tra un continente e un altro, come le rondini da cui prendono il nome, conservando sempre lo stesso mestiere di mietitori. Era anzi proprio il calendario agricolo sfalsato dei tempi di maturazione del grano a spingere questi migranti a seguire le occasioni di lavoro sui due emisferi.

L’aggettivo *golondrina* viene utilizzato in spagnolo per indicare una migrazione ciclica, per lo più stagionale e legata ai lavori della terra. Il termine è utilizzato sia per le mobilità attuali che per quella transoceanica storica tra Ottocento e Novecento. In quest’ultimo caso si fa spesso riferimento ai canali di circolazione tra la penisola iberica – in particolare la Galizia – e l’America Latina (Hoerder 2002, 342). Per quel che riguarda l’Italia invece le migrazioni cosiddette *golondrinas* vennero a colorarsi di caratteristiche peculiari, dai tratti molto decisi, edificando da subito una vera e propria mitologia su questa specie di braccianti giramondo.

Una vera e propria storia dell’adozione e della circolazione di questo termine nel contesto italiano va ancora fatta. Uno dei primi e più autorevoli a farne uso fu sicuramente l’onorevole Luigi Luzzatti, in conclusione di un importante discorso parlamentare del novembre del 1900 (Luzzatti 1900). Pochi anni dopo, una descrizione più ricca e particolareggiata venne data da Pompeo Ghinassi sulle pagine del «Giornale degli economisti» (Ghinassi 1902, 225-226).

Il termine *golondrina*, con la sua valenza positiva, poteva contrapporsi ad altri epiteti meno edificanti. Nei primi del Novecento infatti intorno a quelle stesse tipologie di migranti si potevano trovare descrizioni di tutt’altro tono. Si legga questa:

Con un sacco in ispalla ripieno de’ suoi pochi indumenti egli cammina di colonia in colonia, il deriso *lingera*. [...] Nelle colonie agricole si applica il qualificativo di *lingera* ai

\* Nonostante il saggio sia frutto di un lavoro comune, Francesco Di Filippo ha curato la stesura del paragrafo 2, mentre Stefano Gallo quella dei paragrafi 1 e 3, riprendendo un lavoro precedente (Gallo 2021).

lavoratori che giungono dall'Italia per attendere alla trebbiatura del frumento e lino nei mesi da dicembre a marzo e che, generalmente, terminato questo periodo se ne ritornano in patria. [...] La trebbiatura nei mesi d'estate è, dunque, l'episodio gagliardo; egli, il deriso pellegrino, l'eroe (Scardin 1903, 108).

Il termine *lingera* richiama l'analogo italiano, anzi padano-cremonese, utilizzato in seguito da Danilo Montaldi nelle sue autobiografie della *leggera*, dove indica un mondo di piccola criminalità e marginalità che condivide valori, regole e spazi sociali (Montaldi 1961).

L'immagine ufficiale dell'emigrante italiano non poteva essere quello infamante della *lingera*, mentre riuscì a essere quello edificante della *golondrina*. Proprio alla descrizione di Ghinassi faceva riferimento Ercole Sori nel 1979 nel definire in sede storiografica i lavoratori *golondrinas* come «un esercito di nomadi [...] che inseguiva le estati e i tempi di maturazione delle messi su due emisferi» (Sori 1979, 149). L'altra fonte di Sori era un saggio di Angelo Trento che senza riportare ulteriori fonti collocava la fase più significativa dell'emigrazione *golondrina* tra il 1891 e il 1904, definendola come originaria in particolare dalla pianura padana (Trento 1974).

Nella letteratura internazionale sulle migrazioni quest'ultima raffigurazione è passata senza alcuna modifica, compresa l'esaltazione delle virtù di capacità lavorativa e intraprendenza. L'immagine di un «esercito di nomadi», dei «gagliardi irrequieti indomabili figli del lavoro» che ogni anno vanno e vengono da una parte all'altra dell'oceano per ripetere le stesse operazioni agricole, di cui abbiamo rintracciato le origini italiane di inizio Novecento (ma andrebbero rintracciate anche quelle argentine) e la sua ripresa nella storiografia successiva, è entrata a pieno titolo nel repertorio delle figure migranti più famose e ricordate al mondo. La quantificazione di questo «esercito» rimane estremamente incerta, con stime oscillanti a seconda degli studiosi tra le 7.500 unità e le 250.000 annue (Rey Balmaceda 1985). I *migration studies* danno in buona sostanza per scontata l'esistenza delle *golondrinas*, con le caratteristiche raccontate da Pompeo Ghinassi, assumendo come realistica la cifra di 20.000 migranti l'anno (Baily 1999, 60; Sánchez-Alonso 2006, 386).

**2. Un database.** Una verifica empirica che ci permetta di quantificare le *golondrinas* italiane può essere tentata attraverso l'analisi delle liste di sbarco degli immigrati italiani in Argentina. Il *Centro de Estudios Migratorios Latinoamericanos* (CEMLA) ha costruito negli anni Novanta sotto la direzione di Luigi Favero un database contenente i nominativi relativi a oltre un milione di arrivi avvenuti nel porto di Buenos Aires tra il 2 gennaio 1882 e il 5 dicembre 1920, per la precisione 1.021.391 *records* (Favero 1992). A partire da questa fonte, individuare i migranti che rientrano nel porto di Buenos Aires a intervalli di tempo più o meno lunghi non rappresenta un problema teorico particolarmente complesso. Se si osservano i registri portuali argentini, infatti, si hanno a disposizione molti dei dati essenziali per rintracciare ed eventualmente validare questo tipo di modello di mobilità transoceanico.

Ogni migrante è registrato per nome e cognome e sono riportati la data di ingresso e l'età, da cui si può ricavare l'anno di nascita. Associare l'anno di nascita

Tab. 1. *I viaggi di Raffaele Esposito*

Età	Professione	Nome della nave	Data di arrivo	Porto di imbarco
23	Agricoltore	Poitou	1882-09-25	Napoli
26	Giornaliero	Bengala	1885-04-02	Genova
<b>27*</b>	<b>Giornaliero</b>	<b>La France</b>	<b>1886-11-18</b>	<b>Marsiglia</b>
<b>27*</b>	<b>Giornaliero</b>	<b>La France</b>	<b>1886-11-18</b>	<b>Marsiglia</b>
31	Marinaio	Duca Di Genova	1890-12-16	Genova
<b>40*</b>	<b>Bracciante</b>	<b>Orione</b>	<b>1899-08-03</b>	<b>Genova</b>
<b>40*</b>	<b>Pescatore</b>	<b>Espagne</b>	<b>1899-12-10</b>	<b>Napoli</b>
48	Pescatore	Sin Informacion	1907-09-08	Genova
50	Pescatore	Re Vittorio	1909-10-17	Genova
53	Pescatore	König Friedrich	1912-05-26	Boulogne
56	Pescatore	Cordova	1915-01-29	Genova
61	Pescatore	Re Vittorio	1920-09-04	Genova

Fonte: elaborazione degli autori a partire dal database del CEMLA.

a un nome e cognome permette di ottenere degli indicatori di dettaglio per i singoli individui e di gestire con maggiore accortezza il problema delle omonimie, che in una banca dati di più di un milione di record sono molto consistenti: se uno stesso nominativo appare più volte in anni diversi, una verifica sull'anno di nascita può contribuire a depennare la maggior parte dei casi in cui si tratti di persone diverse. A questo set di informazioni essenziali (nome, cognome e anno di nascita) sono associati ulteriori attributi, come il genere, la professione, le eventuali relazioni parentali, lo stato civile, l'istruzione, la religione, etc. che permettono di esaminare le diverse identità presenti nel dataset, sebbene con minor precisione. Infine, informazioni come il porto di imbarco e il nome della nave con cui hanno viaggiato aiutano a completare la storia personale dei diversi soggetti.

Da un punto di vista teorico, quindi, per validare la teoria della migrazione circolare *golondrina* è sufficiente raggruppare la lista dei passeggeri per i valori principali (l'insieme costituito da nome, cognome e anno di nascita) e osservare quante volte, con quale frequenza e con quali caratteristiche si ripetono questi raggruppamenti. Una volta stabilita la presenza ripetuta di un individuo nei registri portuali, dunque, abbiamo arricchito il modello attraverso ulteriori fattori capaci di gettare luce sulla storia personale del soggetto, come l'elenco delle età che il nostro individuo doveva avere ad ogni rientro al porto di Buenos Aires e l'elenco delle professioni con cui viene registrato. Poiché la banca dati registra le entrate, traducendo gli insiemi ottenuti in termini storici, questi valori di raggruppamento non costituiscono altro se non il numero di arrivi e nuovi arrivi compiuti da ogni individuo.

Al fine di comprendere meglio la complessità del sistema di validazione del nostro modello si veda il seguente esempio riguardante la persona con più viaggi fino ad ora attestata in base all'analisi dei dati, Raffaele Esposito nato nel 1859 (tab. 1).

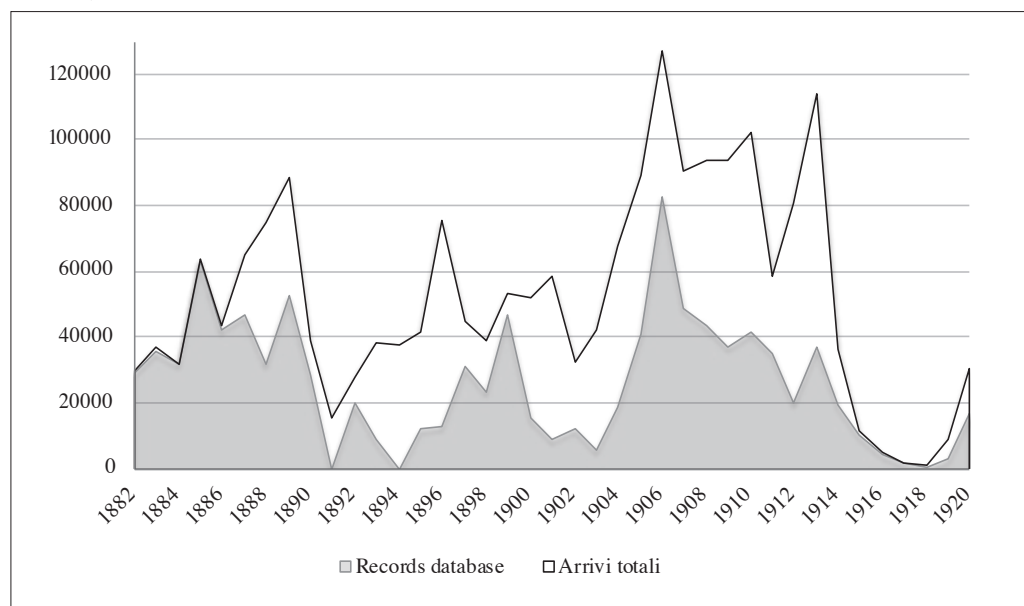
Lo schema sinottico permette di osservare due anomalie significative che possono avere importanti ripercussioni sull'individuazione dei migranti di ritorno. Le occorrenze identificate circa la presunta vicenda migratoria di Raffaele Esposito nascondono in realtà una questione molto più complessa, sia per quanto riguarda il problema delle *golondrinas* sia, più in generale, per quanto riguarda il trattamento semi-automatico di dati storici formalizzati in grandi *repositories* digitali.

In due momenti particolari, nel 1886 e 1899, il *database* registra due volte l'ingresso di un Raffaele Esposito. In un caso, le due occorrenze del Raffaele Esposito quarantenne che arriva in Argentina nel 1899 documentano quasi certamente un esempio di omonimia. Uno è un bracciante, l'altro è un pescatore, ma soprattutto i due approdano a Buenos Aires a quattro mesi di distanza l'uno dall'altro provenendo da due porti di imbarco distinti. Se da un punto di vista temporale questa finestra permetterebbe di tornare in Italia per poi ripartire per il Sud America, tale operazione sarebbe economicamente ingiustificabile vista anche la qualifica lavorativa di basso livello. Ne consegue che l'insieme di attributi costituiti dall'insieme essenziale nome/cognome/anno di nascita e dagli attributi accessori non è sempre sufficiente a definire le diverse identità in modo univoco e, di conseguenza, che molte *golondrinas* identificate attraverso il nostro modello potrebbero rappresentare falsi positivi, mascherando un più banale fenomeno di omonimia, soprattutto nel caso di nomi con frequenze statistiche molto elevate come Raffaele Esposito. Non è un caso se nel 1973 in Italia è stato introdotto il codice fiscale per identificare in modo univoco i contribuenti. Mentre il codice fiscale si basa su otto diversi parametri (nome, cognome, anno, mese e giorno di nascita, sesso, comune di nascita e un carattere di controllo), il nostro lavoro si può avvalere solo di quattro parametri (nome, cognome, anno di nascita, sesso).

Il secondo caso di interesse riguarda un individuo di 27 anni che arriva due volte, esattamente lo stesso giorno, con la medesima nave, dichiarando in entrambi i casi la professione di giornaliero. A differenza dell'altra, il ricorrere di questa occorrenza indica un problema di diversa natura, ovvero la presenza di record duplicati, e permette di svelare un'incoerenza di fondo nel modo in cui è stato assemblato il database dal CEMLA: sembra evidente che per la fase di input non sono stati impiegati adeguati strumenti di controllo per l'inserimento dei *records*.

Quest'anomalia ci ha indotto quindi a esaminare l'intera banca dati attraverso una serie di *window functions* SQL che hanno evidenziato l'esistenza di numerose ripetizioni e la duplicazione di intere sezioni dei registri. Un certo Pasquale Ottaviano, un giornaliero di 33 anni sbarcato il 24 gennaio 1885 dalla nave San Gottardo, compare nel dataset per ben sette volte. Anche in questo caso la presenza di ripetizioni ha un profondo impatto sul modello computazionale concepito per l'individuazione di migranti circolari, inducendo a sovrastimare il fenomeno. Se dunque il problema dell'omonimia può essere semplicemente gestito attraverso un uso combinato dei dati disponibili per ogni migrante, le inconsistenze ereditate dal processo di digitalizzazione, che non riguardano esclusivamente saltuarie ripetizioni, costituiscono lo scoglio maggiore nell'identificazione dei migranti di rientro. L'assenza di studi empirici sul problema storico dei migranti stagionali conosciuti come *golondrinas* può risiedere a nostro avviso proprio nella mancanza di *dataset*

Fig. 1. Arrivi registrati nel database e arrivi totali a Buenos Aires nel periodo 1882-1920 (valori assoluti)



Fonte: elaborazione degli autori a partire dal database del CEMLA; Dirección 1925.

affidabili e ben strutturati. Se, come detto, i dati a nostra disposizione riguardano le registrazioni dei registri di bordo consegnati alle autorità portuali di Buenos Aires, la possibilità di intercettare migranti che periodicamente rientrano in Argentina risiede sostanzialmente nella capacità di identificare in modo univoco i singoli individui. Purtroppo però questo processo si scontra con una serie di problemi non trascurabili e di complessa risoluzione.

In primo luogo il dataset a nostra disposizione è incompleto. I registri portuali sono stati digitalizzati in modo parziale con annate complete o quasi (dal 1882 al 1886), intervalli completamente trascurati (come il 1891 e il 1894), o serie per le quali i record registrati sono meno della metà di quelli documentati nelle statistiche ufficiali argentine (fig. 1). Appare chiaro come questo sistema di acquisizione, per quanto rilevante da un punto di vista statistico, non costituisca una buona base di partenza per uno studio il cui valore si basa sulla possibilità di riconoscere ondate di migranti che ritornano nel paese a distanza di anni: un individuo potrebbe essere a tutti gli effetti un migrante stagionale, ma noi potremmo non intercettare questa caratteristica nel nostro campione.

Ancor più importante, due aspetti fondamentali caratterizzano in negativo il dataset. In primo luogo alcuni campi del registro non sono stati presi in considerazione nel processo di acquisizione digitale degli originali cartacei, in particolare quelli che permetterebbero di definire meglio tipologia e natura delle imbarcazioni. I registri cartacei originali del porto di Buenos Aires riportano il nome dell'imbarcazione, la sua classe, la matricola, la provenienza (ma non le eventuali tappe), i nomi del capitano e del medico di bordo, il tonnellaggio, il numero dei membri dell'equi-

paggio e dei passeggeri, il giorno di arrivo. Solo quest'ultimo e la provenienza sono stati registrati nel processo di digitalizzazione: gli altri, tuttavia, sarebbero stati di estrema importanza per la definizione dei caratteri del viaggio e dunque per meglio distinguere la storia personale dei migranti.

In secondo luogo le informazioni a nostra disposizione sono caratterizzate da un eccesso di 'rumore di fondo'. Molti attributi, specialmente quelli che presentano un più sostanziale livello di variabilità come nomi e cognomi, sono caratterizzati da un elevatissimo livello di errore, siano essi generati originariamente dagli ufficiali censuari e dalle compagnie di navigazione o introdotti a distanza di un secolo durante il processo di digitalizzazione. Se osserviamo ad esempio l'incidenza dei cognomi unici all'interno del *dataset* questi costituiscono il 21,87% di quelli registrati, una misura decisamente troppo elevata se comparata con altri *corpora* onomastici. I nomi di persona presentano un tasso di unicità decisamente meno marcato: i nomi che compaiono una sola volta incidono per circa il 5% sul totale di quelli attestati. Gli altri attributi, come il sesso o la qualifica lavorativa, riportano anch'essi numerose incongruenze e, a volte, inspiegabili caratterizzazioni: ricorrono ad esempio 25 categorie diverse per la definizione del genere. A questo si aggiunga un importante *bias* di natura linguistica: molti, troppi nomi di italiani sono registrati con l'alternativa in spagnolo (anche in questo caso non è facile stabilire quando l'errore sia stato introdotto) o comunque per mezzo di una fonetica non italiana; ad esempio le consonanti doppie che in italiano hanno valore fonemico, e quindi distintivo, spesso cadono.

Se dunque il modello computazionale per identificare modi e qualità della circolarità dei rientri da un punto di vista concettuale è di facile implementazione, sono i dati stessi a costituire l'ostacolo maggiore: da un lato abbiamo troppi pochi attributi per definire univocamente i migranti, dall'altra notiamo troppi errori di compilazione nella banca dati. Quest'ultimo aspetto non deve stupire. Se prendiamo in considerazione altri approcci empirici alle migrazioni transoceaniche, la povera consistenza dei dati rappresenta quasi una costante, sia essa dipendente da una confusa registrazione originaria sia da un modesto processo di digitalizzazione a un secolo di distanza. Nello studio di Brum sulla migrazione degli italiani negli Stati Uniti, ad esempio, l'autore osserva diverse inconsistenze (Brum 2019; Abramitzky *et al.* 2012; Feigenbaum 2016; Spitzer, Zimran 2018). Nel *dataset* *Italians to America Passenger Data File* il 60% dei dati relativi alla residenza sono mancanti o invalidi (Brum 2019, 11), mentre dal confronto tra quest'ultimo *dataset* e il *Full Count 1900 United States Census* emerge che circa il 60% dei cognomi sono registrati con errori di varia natura che impediscono il confronto tra le due banche dati (*ivi*, p. 60).

Per risolvere un problema costante nella formalizzazione di grandi *repositories* testuali, nelle ultime due decadi è però stata introdotta un'ampia varietà di approcci per il *fuzzy names matching*, ovvero un ventaglio di tecniche computazionali per determinare il grado di somiglianza tra stringhe e, dunque, nel nostro caso tra nomi, cognomi, professioni, etc. In realtà alcuni strumenti sono disponibili da più di un secolo: l'algoritmo fonetico *Soundex*, ad esempio, venne sviluppato da Robert C. Russell e Margaret King Odell tra il 1918 e il 1922, anche se la sua fortuna venne definitivamente sancita negli anni Settanta (Knuth 1973). Per una trattazione gene-

rale dei vari approcci si rimanda in generale alla letteratura dedicata al tema delle *similarity measures* (Christen 2006; Bandyopadhyay, Saha 2013; Wang, Dong 2020). Sebbene gli esempi siano numerosissimi, gli approcci più diffusi di *fuzzy names matching* si fondano sul computo di misure di similarità sulla base principi fonetici (*phonetic*), delle distanze tra i termini in esame (*edited-based*) o dei loro costituenti (*token-based*), sulla base di particolari domini di conoscenza (*domain-dependent*), o infine su approcci ibridi. Contestualmente va sottolineato come non esista un metodo più valido di altri: la loro applicazione dipende di volta in volta dalla particolare esigenza del tema di ricerca. Così, ad esempio, in numerosi studi si è fatto ricorso ad algoritmi fonetici (*phonetic similarity*) per normalizzare le incoerenze delle trascrizioni di nomi e attributi (Parmar, Kumbharana 2014; Coluccia, Spadavecchia 2021). Altri, come Feigenbaum (2016), in uno studio comparativo sui censimenti, applicano l'algoritmo di similarità bi-grammi (*token-based measure*) in combinazione con Jaro-Winkler (*edited-based measure*) per istruire un processo di *machine-learning* (Dong, Wang 2019; Embley *et al.* 2021; Hadwan *et al.* 2021).

Nel nostro studio abbiamo deciso di affrontare la risoluzione delle grafie anomale attraverso un approccio più tradizionale, ovvero tramite l'impiego di algoritmi fonetici. La ragione di questa scelta risiede principalmente nel tipo di architettura informatica scelta per l'immagazzinamento e la gestione dei dati dei registri portuali, un RDBMS (PostgreSQL nella fattispecie). Come evidenziato in precedenza, il nostro modello per il tracciamento dei migranti *golondrinas* si basa sulla possibilità di raggruppare le occorrenze dello stesso individuo. Questo processo, in un'architettura data-centrica, viene realizzato per mezzo della clausola SQL 'group by' che prende come argomento alcune relazioni di equivalenza (in senso matematico) e le utilizza per suddividere una relazione SQL in classi di equivalenza: questo tipo di relazione si definisce come simmetrica, riflessiva e transitiva. Impiegando misure di similarità diverse da quelle fonetiche, ad esempio quelle *edited-based* o *token-based* (Feigenbaum 2016; Brum 2019), la nostra relazione avrebbe disatteso uno dei suoi caratteri essenziali, ovvero non potrebbe più considerarsi transitiva. In altri termini, l'impiego di un parametro di raggruppamento basato sul predicato di raggruppamento «se due stringhe si trovano entro una certa distanza di modifica l'una dall'altra», per quanto intellegibile per un essere umano, non può essere gestita dalla clausola SQL 'group by' (almeno non in modo efficiente). Per definire la distanza di resa tra due individui, ad esempio Agostino e Agustin Rossi, per poi raggruppare quelli considerati entro una determinata soglia di differenza, si dovrebbe confrontare ogni individuo con tutti gli altri presenti nella banca dati, generando una matrice di tipo esponenziale e sottoponendo il processore a uno sforzo eccessivo.

Un algoritmo fonetico, al contrario, si basa su un principio profondamente diverso all'interno di un costrutto SQL. Prima di raggruppare, infatti, un modello basato su un algoritmo fonetico di riduzione restituisce una stringa comune per i nomi considerati simili (ad esempio sia la parola 'Agostino' che 'Agustin' vengono convertiti nella stringa convenzionale 'AKST') e solo a quel punto il processo prosegue con il vero e proprio raggruppamento. In questo caso la relazione mantiene tutti e tre i caratteri essenziali per un efficiente e corretto raggruppamento: simmetria, riflessività e transitività.

Rimane solo da definire quale sia il miglior algoritmo fonetico da utilizzare (Christen 2012; Parmar, Kumbharana 2014; Smiley, Kübler 2017). Nel nostro caso abbiamo optato per due diversi approcci. Se è vero che la maggior parte degli algoritmi fonetici oggi in uso sono stati pensati per la resa di contesti anglofoni, il *soundex* e il *double-metaphone* si sono rivelati di particolare utilità nel nostro caso.

Il *soundex* permette la transcodifica in una medesima stringa di nomi con pronuncia uguale, ma diversa rappresentazione grafica. La sua scarsa sofisticazione lo rende forse uno dei più adattabili a contesti non anglofoni. Abbiamo dunque proceduto alla modifica delle regole elementari dell'algoritmo per adattarlo alla fonologia italiana sulla base delle osservazioni di P. Borsa (2006). L'unica sostanziale differenza rispetto al modello per un *soundex* italiano di Borsa riguarda l'uso delle doppie: visto che queste spesso cadono nelle trascrizioni a nostra disposizione abbiamo volutamente ridotto l'importanza di questa caratteristica nel sistema di transcodifica (ad esempio il cognome Piazzolla sarà uguale a Piazzola).

Il secondo algoritmo che si è dimostrato particolarmente utile è il *double-metaphone*. Sebbene anche questo sia stato pensato originariamente per una fonetica anglosassone, prende però in considerazione anche le particolarità di pronuncia di altre lingue. A differenza di altri algoritmi fonetici, il *double-metaphone* può restituire sia una transcodifica primaria sia una secondaria per gestire con più accuratezza i casi di ambiguità (Christen 2012, 78).

Il risultato finale del nostro modello può essere riassunto dalla seguente *query* SQL, qui presentata nella sua forma semplificata:

```
WITH dataset AS (
  SELECT *, EXTRACT(YEAR FROM data_arrivo) - eta AS anno_nascita FROM
  dati
  WHERE eta >= 10 AND eta <=90
  AND sex IN ('M')
)
SELECT
  dmetaphone(nome) AS nome,
  _cognomex(cognome) AS cognome
  anno_nascita,
  count(*) AS num_viaggi,
  array_agg( (data_arrivo, porto_imbarco, nave) ORDER BY data_arrivo),
  array_agg( professione ORDER BY data_arrivo)
  array_iter(array_agg(eta ORDER BY data_arrivo)) AS incremento
FROM dataset
GROUP BY dmetaphone(nome_norm), _cognomex(cognome), anno_nascita
HAVING count(*) > 2
ORDER BY num_viaggi DESC, anno_nascita ASC
```

La *query* SQL si basa su una tabella temporanea introdotta da una *Common Table Expressions* (CTE *clause*) che ha lo scopo di pre-filtrare i dati su cui condurremo l'analisi e mira a: 1) definire l'anno di nascita dei migranti sottraendo l'età dichiarata all'anno di arrivo; 2) filtrare gli individui in età lavorativa<sup>1</sup>; 3) limitare il

campione ai soli uomini. Quest'ultimo limite, l'esclusione delle donne, ci ha portato a eliminare circa un quarto del campione disponibile, ma è stato altresì osservato che il costume di cambiare cognome dopo il matrimonio ha importanti ripercussioni sulla riconoscibilità delle donne all'interno di grandi banche dati censuarie (Brum 2019, 16).

A questo punto segue il processo di aggregazione vero e proprio e si basa su quanto definito fino ad ora. I nomi sono gestiti dall'algoritmo fonetico *double-metaphone*, i cognomi dalla versione adattata alla fonologia italiana del *soundex*<sup>2</sup>. Inoltre sono stati aggiunti alcuni campi aggregati di controllo come il set composto da data di arrivo, porto di imbarco e nome della nave, nonché una lista di tutte le professioni con cui un individuo viene registrato. Entrambi i set, ordinati per la data di arrivo a Buenos Aires, possono essere messi in relazione ai diversi sbarchi di ogni presunto individuo. Infine, una funzione *custom* realizzata attraverso il linguaggio procedurale PL/Python ('array\_iter') permette di ottenere una lista con la differenza numerica espressa in anni di ogni nuovo sbarco. Qualora tale incremento sia minore di uno, sono molto elevate le possibilità che la presunta *golondrina* individuata tramite il raggruppamento sia in realtà un caso di omonimia o, come evidenziato in precedenza, una duplicazione del record.

L'*output* preliminare si presenta come segue:

nome*	cognome*	anno_nascita	viaggi	età	incremento
RAFAELE, RAFFAELE	ESPOSITO	1859	12	23, 26, 27, 27, 31, 40, 40, 48, 50, 53, 56, 61	3, 1, 0, 4, 9, 0, 8, 2, 3, 3, 5
GIUSEPPE	ROSSI	1860	9	25, 25, 26, 26, 27, 28, 29, 29, 30	0, 1, 0, 1, 1, 1, 0, 1
AGOSTIN, AGOSTINO, AGUSTIN, AGUSTINE, AUGUSTIN, AUGUSTINO	ROSSI, ROSI, ROSSI	1883	9	31, 32, 32, 32, 33, 33, 34, 34, 34	1, 0, 0, 1, 0, 1, 0, 0
GIUSEPPE	GRECCO, GRECO	1857	8	25, 27, 29, 32, 36, 42, 48, 52	2, 2, 3, 4, 6, 6, 4
ANTONIO	ROSSI	1858	8	25, 27, 28, 28, 29, 30, 31, 39	2, 1, 0, 1, 1, 1, 8
GIUSEPPE	ROSI, ROSSI	1873	8	12, 14, 16, 26, 27, 28, 33, 40	2, 2, 10, 1, 1, 5, 7
GIUSEPPE, GUISEPE	ROSSI	1875	8	14, 21, 22, 23, 24, 27, 30, 37	7, 1, 1, 1, 3, 3, 7
PASQUALE	OTTAVIANO	1852	7	33, 33, 33, 33, 33, 33, 33	0, 0, 0, 0, 0, 0

Sebbene il metodo di raggruppamento sia ancora da affinare, come è stato possibile osservare durante tutta la nostra discussione il problema maggiore che

Tab. 2. *Una stima degli arrivi multipli*

Numero di arrivi	Casi riscontrati	Falsi positivi	Casi plausibili
2	23.722	7.150	16.572
3	2.373	35	2.338
4	468	4	464
5	131	0	131
6	58	1	57
7	19	0	19
8	2	0	2
9	2	0	2
12	1	0	1
Totale dei "ritornati"	26.776		19.586
TOT subset	667.491		667.491
%	4,01		2,93
TOT db	1.021.391		1.021.391

Fonte: elaborazione degli autori a partire dal database del CEMLA.

abbiamo incontrato riguarda prevalentemente l'individuazione di falsi positivi. In altri termini, il nostro modello tende a sovrastimare la presenza di migranti che a distanza di tempo rientrano nel porto di Buenos Aires.

Ciononostante i risultati mostrano anche che la presenza di ritorni in Argentina ha un'incidenza statistica estremamente bassa sul totale dei dati a disposizione (tab. 2).

**3. Un'ipotesi.** I risultati emersi tendono a ridimensionare in maniera sostanziale la consistenza effettiva del fenomeno delle *golondrinas*. La gran parte dei rientri non supera i due sbarchi a Buenos Aires su una mobilità circolare totale poi non così folta. Pur tenendo conto degli inevitabili buchi causati dall'incompletezza della fonte, che potrebbero essere in parte colmati con delle proiezioni matematiche, non sembra di essere di fronte a un fenomeno di massa, tanto da poter ipotizzare una sua scarsa rilevanza statistica.

Fin dai primi del Novecento, in effetti, l'esistenza reale del concetto di *golondrina* era stata oggetto di forti critiche. Luis Moltedo, ex agente marittimo per la Navigazione Generale Italiana e la Lloyd, a partire dall'estate del 1911 pubblicò sulla stampa argentina alcuni interventi in cui contestava la forma con cui era comunemente presentata la migrazione transoceanica *golondrina*. La tesi di Moltedo era che le operazioni di raccolta agricola venivano sostenute in larga parte da lavoratori che già risiedevano in Argentina: il numero di immigrati che arrivavano in ottobre, all'inizio della stagione, attirati da questa possibilità di lavoro, doveva considerarsi molto minore di quello comunemente inteso. I guadagni che potevano essere tratti

dal lavoro bracciantile infatti non erano tali da sostenere – pur tenendo in considerazione eventuali sussidi governativi – due viaggi transoceanici l'anno: chi arrivava per la mietitura rimaneva nel paese oppure se ne andava per non tornare l'anno successivo.

Gli interventi di Moltedo vanno contestualizzati in una fase di difficili relazioni diplomatiche tra i due paesi proprio intorno alla questione migratoria, il cui culmine venne raggiunto con l'epidemia di colera del 1911 e lo scontro sui controlli da effettuarsi sui vettori provenienti dalla Penisola. Poteva quindi risultare politicamente conveniente per un opinionista minimizzare l'apporto degli italiani all'economia argentina (Choate 2008). Meno di un anno prima, il 5 ottobre 1910, Giuseppe Bevione era intervenuto dalle colonne de «La Stampa» per sostenere al contrario che «se il Governo italiano proibisse per una sola annata l'emigrazione 'golondrina' i raccolti marcirebbero per tre quarti nei campi, e l'Argentina soffrirebbe più che se una immensa invasione di cavallette avesse straziato le sue culture da Tucuman allo stretto di Magellano, senza risparmiare un pollice di terra» (Corradini 1911, 23).

Tuttavia le argomentazioni dell'ex agente marittimo argentino hanno trovato alcune significative conferme in sede storiografica. Già nel 1981 Raúl C. Rey Balmaceda, in un lavoro presentato nel corso delle «primeras jornadas nacionales de estudios sobre inmigración en Argentina» (Rey Balmaceda 1985), riportava le tesi di Moltedo e provava a farne una sommaria verifica. Qualche anno più tardi un lavoro più compiuto e approfondito veniva svolto dallo storico del lavoro Jeremy Adelman, il quale confermava l'inconsistenza del mito delle *golondrinas*. Negli archivi statali argentini riuscì a reperire un'indagine condotta dal dipartimento dell'immigrazione alla fine dell'Ottocento, da cui risultava come solo in circostanze particolarmente fortunate un bracciante italiano poteva guadagnare abbastanza dalla mietitura da poter tornare l'anno successivo. Gli ulteriori calcoli formulati da Adelman supportavano anche matematicamente questa tesi. In Argentina la mietitura iniziava a ottobre nel nord e finiva a febbraio nel sud, in ogni distretto durava al massimo un mese. I lavoratori si spostavano da nord a sud seguendo più campagne di mietitura, dunque potevano trovare lavoro nel settore per cinque mesi consecutivi, tuttavia – considerando le possibilità reali di spostamento – una stima di due mesi di lavoro appare più ragionevole (Adelman 1992, 99). Difficilmente i proventi sarebbero bastati da soli per tornare in patria portando con sé anche qualche risparmio. La maggior parte dei salariati agricoli occupati proveniva dall'interno della Repubblica, in particolare da Buenos Aires e dintorni.

Più che la *golondrina* europea dei viaggi transoceanici, dunque, a dominare la scena sarebbe stata la *golondrina criolla* degli spostamenti interni, figura – *mutatis mutandis* – estremamente diffusa anche all'interno dei confini italiani dell'epoca. Gli immigrati italiani dunque si sarebbero mischiati a quelli interni. Dopo gli studi di Adelman, altri studiosi si sono rivelati in effetti più prudenti nell'utilizzo della categoria delle *golondrinas* per i braccianti immigrati in Argentina, pur senza aggiungere ulteriori materiali per chiarire la questione (Bernasconi 2000, 88; Devoto 2006, 237-239; Belich 2009, 531-532). Per poter sciogliere in maniera incontrovertibile la questione, non resta che continuare a lavorare sui dati per pulirli e possibilmente integrarli con nuove campagne di raccolta.

<sup>1</sup> Alcune età molto elevate potrebbero non essere consistenti dato il tema della ricerca, ma servono ad escludere molti individui per i quali il database riporta l'inspiegabile età di novecento anni.

<sup>2</sup> A questo sono state apportate ulteriori modifiche per la pulizia dei dati. Un esempio su tutti: molto di frequente al posto della O viene usato 0, uno zero (ROSSI ~ ROSSI).

## Riferimenti bibliografici

- R. Abramitzky, L.P. Boustan, K. Eriksson 2012, *Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration*, «American Economic Review», 102, 1832-1856.
- J. Adelman 1992, *The Harvest Hand: Wage-Labouring on the Pampas, 1880-1914*, in J. Adelman (ed.), *Essays in Argentine Labour History, 1870-1930*, Palgrave Macmillan, Basingstoke-London, 91-111
- S.L. Baily 1999, *Immigrants in the lands of promise. Italians in Buenos Aires and New York City, 1870-1914*, Cornell University Press, Ithaca.
- S. Bandyopadhyay, S. Saha 2013, *Unsupervised Classification: Similarity Measures, Classical and Metaheuristic Approaches, and Applications*, Springer, Berlin.
- J. Belich 2009, *Replenishing the Earth. The Settler Revolution and the Rise of the Anglo-World 1783-1939*, Oxford University Press, Oxford.
- A. Bernasconi 2000, *Imigrantes Italianos na Argentina (1880-1930): Uma Aproximação*, in B. Fausto (org.), *Fazer a América. A Imigração em Massa para a América Latina*, Editora da Universidade de São Paulo, São Paulo, 61-92.
- P. Borsa 2006, *Problemi di metodologia statistica: la funzione Soundex*, working paper Università di Milano.
- M. Brum 2019, *Italian Migration to the United States: The Role of Pioneers' Locations*, RIDGE working paper.
- M.I. Choate 2008, *Emigrant Nation. The Making of Italy Abroad*, Harvard University Press, Cambridge-London.
- P. Christen 2012, *The Data Matching Process. Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection, Data-Centric Systems and Applications*, Springer, Berlin-Heidelberg
- P. Christen 2006, *A Comparison of Personal Name Matching: Techniques and Practical Issues*, in *Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06)*, IEEE, Hong Kong, 290-294.
- D.M. Coluccia, L. Spadavecchia 2021, *The Economic Effects of Immigration Restriction Policies – Evidence from the Italian Mass Migration to the US*, CESifo Working Paper No. 9361.
- E. Corradini 1911, *Classi proletarie: socialismo, nazioni proletarie: nazionalismo*, in G. Castellini (a cura di), *Il Nazionalismo italiano. Atti del congresso di Firenze*, Quattrini, Firenze, 22-35.
- F. Devoto 2006, *Historia de los italianos en la Argentina*, Editorial Biblos, Buenos Aires.
- Dirección General de Inmigración 1925, *Resumen estadístico del movimiento migratorio de la Republica Argentina. Años 1857-1924*, Talleres Gráficos del Ministerio de Agricultura de la Nación, Buenos Aires.
- B. Dong, H.W. Wang 2019, *Efficient Authentication of Approximate Record Matching for Outsourced Databases*, in *Theory and Application of Reuse, Integration, and Data Science. Presented at the International Conference on Information Reuse and Integration*, Springer, Cham, 119-168.
- D.W. Embley, S.W. Liddle, D.W. Lonsdale, S.N. Woodfield 2021, *Inter-Generational Family Reconstitution with Enriched Ontologies*, in *Advances in Conceptual Modeling: ER 2021 Workshops CoMoNoS, EmpER, CMLS St. John's, NL, Canada, October 18-21, 2021, Proceedings*, Springer, Cham, 61-74.
- L. Favero 1992, *Le liste di sbarco degli immigrati in Argentina*, «Altreltalie», 7, 126-138.
- J.J. Feigenbaum 2016, *A Machine Learning Approach to Census Record Linking*, Harvard working paper.
- S. Gallo 2021, *La grande emigrazione dei rurali italiani e il mito delle golondrinas*, «Archivio Storico delle Emigrazioni Italiane», 20-21, 29-38.
- P. Ghinassi 1902, *Gli agricoltori italiani nell'Argentina. (Continuazione; v. il fascicolo di luglio)*, «Giornale degli economisti», luglio, 13, 25, 42-66.
- M. Hadwan, M.A. Al-Hagery, M. Al-Sanabani, S. Al-Hagree 2021, *Soft Bigram distance for names matching*, «PeerJ Computer Science», 7, e465.

- D. Hoerder 2002, *Cultures in contact. World migrations in the second millennium*, Duke University Press, Durham-London.
- D. Knuth 1973, *The Art of Computer Programming*, vol. 3, *Sorting and searching*, Addison-Wesley, Boston.
- L. Luzzatti 1900, *Sull'emigrazione. Discorso del deputato Luigi Luzzatti pronunciato alla Camera dei deputati nella tornata del 29 novembre 1900*, Camera dei Deputati, Roma.
- D. Montaldi 1961, *Autobiografie della leggera*, Einaudi, Torino.
- V.P. Parmar, C. Kumbharana 2014, *Study Existing Various Phonetic Algorithms and Designing and Development of a working model for the New Developed Algorithm and Comparison by implementing it with Existing Algorithm(s)*, «International Journal of Computer Applications», 98, 45-49.
- R.C. Rey Balmaceda 1985, *Acerca de la 'migración golondrina' en la República Argentina*, in Ministerio de Cultura y Educación, *Primeras Jornadas Nacionales de estudios sobre inmigración en Argentina. 5, 6 y 7 de noviembre de 1981*, Eudeba, Buenos Aires, 529-47.
- B. Sánchez-Alonso 2006, *Labor and Immigration*, in Victor Bulmer-Thomas, John H. Coatsworth, Roberto Cortés Conde (eds.), *The Cambridge Economic History of Latin America*, vol. 2, *The Long Twentieth Century*, Cambridge University Press, Cambridge, 377-426.
- F. Scardin 1903, *Vita italiana nell'Argentina. Impressioni e note di viaggio*, vol. 2, Talleres Gráficos de L.J. Rosso, Buenos Aires.
- C. Smiley, S. Kübler 2017, *Native Language Identification using Phonetic Algorithms*, in J. Tetreault, J. Burstein, C. Leacock, H. Yannakoudakis (eds.), *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications. Association for Computational Linguistics*, ACL, Copenhagen, 405-412.
- E. Sori 1979, *L'emigrazione italiana dall'Unità alla seconda guerra mondiale*, Il Mulino, Bologna.
- Y. Spitzer, A. Zimran 2018, *Migrant self-selection: Anthropometric evidence from the mass migration of Italians to the United States, 1907-1925*, «Journal of Development Economics», 134, 226-247.
- A. Trento 1974, *Appunti sull'emigrazione italiana a Buenos Aires agli inizi del secolo e sul suo apporto al movimento operaio argentino*, «Affari sociali internazionali», 1-2, 145-169.
- J. Wang, Y. Dong 2020, *Measurement of Text Similarity: A Survey*, «Information» 11, 421, 1-17.