

L'uso dei cognomi nella ricerca biodemografica

PAOLA PIZZETTI, ENZO LUCCHETTI, LAMBERTO SOLIANI

1. Premessa. In quasi tutti i paesi sviluppati, da alcuni decenni il tasso di natalità è sceso ampiamente sotto il livello di sostituzione, mentre i flussi migratori interni e quelli provenienti dall'estero sono fortemente aumentati. La scomparsa di famiglie locali e l'inserimento di nuovi individui determinano un cambiamento rapido ed importante nella composizione etnica, culturale e sociale di quasi tutte le popolazioni. È una trasformazione storica, che ha pochi riscontri nel passato, e in genere su aree geografiche più limitate.

In questa fase d'evoluzione delle comunità, la ricerca demografica tenta di rispondere allargando i propri orizzonti, aprendosi al confronto con conoscenze e strumenti sviluppati da tutte le discipline che analizzano le conseguenze dell'agire umano ed i comportamenti collettivi. È diffusa la convinzione che i nuovi problemi richiedano approcci culturali e metodologici innovativi, strumenti d'indagine che utilizzino anche fonti fino ad ora trascurate.

La biologia e la genetica di popolazioni hanno proposto modelli che partendo dalla scelta nuziale, dalla suddivisione della popolazione in isolati, dall'intensità e direzione dei flussi migratori, ne stimano gli effetti sulla distribuzione delle frequenze dei caratteri ereditari; viceversa, con gli stessi modelli e sulla base delle medesime teorie, è possibile percorrere il cammino inverso, risalire cioè dalle variazioni delle frequenze geniche a indici e misure dei comportamenti sociali che le hanno causate.

Negli ultimi decenni, pure con tutte le limitazioni e le cautele del caso, al posto delle frequenze geniche sono state utilizzate quelle dei cognomi. Essi infatti, almeno nelle società a trasmissione patrilineare, sono assimilabili ad un carattere genetico trasmesso lungo le linee maschili. Varie ricerche del settore antropologico hanno evidenziato l'utilità dell'uso dei cognomi in analisi sui comportamenti sociali e sui fattori che li modificano.

Sono valutazioni e stime che, rispetto alle classiche misure genetiche, hanno un'estensione temporale limitata agli ultimi secoli; se confrontate ai parametri demografici sono meno precise. Tuttavia, per l'analisi sociale, sia attuale sia dei secoli scorsi, presentano alcuni vantaggi: forniscono gli effetti cumulati dei fattori che trasformano la composizione etnica di una popolazione; possono essere utilizzati in assenza di dati specifici sui flussi migratori; permettono analisi retrospettive, estese fino al momento di stabilizzazione dei cognomi.

Analisi basate sui cognomi, come la tipizzazione di una comunità e lo studio delle sue trasformazioni, la similarità tra due o più popolazioni, le stime dei flussi migratori da dati campionari, sono ormai parte fondamentale della ricerca biologica

sulle popolazioni. Dalla prima serie organizzata di articoli comparsa su «Human Biology» nel 1983 (55, 2) al Convegno Internazionale del 1998 (Lione, 6-12 dicembre), i metodi presentati e le loro finalità sono divenuti sempre più complessi e sofisticati. Appare utile una breve sintesi dei metodi fondamentali, per favorire una proficua collaborazione e la convergenza dei temi di ricerca delle discipline bio-antropologiche con quelli delle discipline demografiche, storiche e sociali. Dalla convergenza di interessi e dall'utilizzazione estesa e proficua nelle varie aree di ricerca che studiano le comunità umane, possono derivare validazioni ulteriori del cognome come informazione, oltre che ulteriori sviluppi dei metodi di analisi.

2. Finalità e presupposti culturali. Biologia e Genetica delle popolazioni umane sono discipline tese allo studio della struttura delle popolazioni umane e delle variazioni – nel tempo e nello spazio – che in esse compaiono. Partendo dalla distribuzione di caratteristiche biologiche – in particolare dalle frequenze alleliche, dei diversi modi cioè di essere delle informazioni ereditarie (*geni*), variabili fondamentali nella descrizione della struttura del *pool genico* di una popolazione – queste discipline hanno sviluppato tecniche e metodi idonei all'individuazione e allo studio delle forze e dei meccanismi, deterministici o stocastici, dai quali dipendono le differenziazione nello spazio e le trasformazioni nel tempo (evoluzione) dei gruppi umani. Il grado di affinità esistente tra i *pool genici* di più popolazioni riconduce alla stima del grado di condivisione di uno stesso percorso evolutivo. La proiezione a ritroso nel tempo delle relazioni di similarità/distanza riporta alla ricostruzione del cammino evolutivo conosciuto dalle popolazioni considerate, nell'assunzione che l'affinità sia inversamente correlata al tempo di diversificazione tra le popolazioni¹. La conoscenza della struttura in cognomi di una popolazione permette di affrontare questi stessi problemi? Sulla base di quali presupposti e con quali strumenti? Quanto attendibili sono le risposte alle quali perviene?

3. I cognomi come alleli. I cognomi hanno una derivazione culturale: la loro formazione e le modalità con le quali vengono trasmessi dipendono dal contesto e dalle caratteristiche culturali del gruppo.

Dal momento della loro fissazione nelle popolazioni e con la comparsa di norme precise che ne regolano la trasmissione, essi diventano assimilabili a caratteri ereditari e possono essere indagati come tali. In funzione delle modalità con le quali i cognomi sono conservati e trasmessi al passare delle generazioni, i metodi di analisi propri della Genetica e Biologia delle popolazioni umane possono essere applicati al sistema cognomi. In buona parte delle società occidentali, i cognomi sono trasmessi per via patrilineare: essi possono essere assimilati a un marcatore genetico localizzato sul cromosoma Y; quando i cognomi sono trasmessi per via matrilineare, è possibile rifarsi ai modelli interpretativi utilizzati negli studi sul DNA mitocondriale. Il modello di trasmissione che conserva anche il cognome materno accanto a quello paterno può offrire un'informazione superiore in quantità e qualità.

Il sistema cognomi risulta altamente 'poliformico' – in termini genetici corrisponde ad un carattere dotato di un grande numero di alleli, che si manifesta con un elevato numero di fenotipi (cognomi diversi) – che possiede notevole capacità di

tipizzare e differenziare le popolazioni e i gruppi; la capacità di discriminazione del carattere cognome è superiore a quella dei caratteri genetici, efficaci su grandi aree, ma spesso non in grado di evidenziare differenze entro ambiti geografici limitati o tra popolazioni limitrofe (Lucchetti *et al.* 1989).

L'elevata capacità di tipizzazione trova applicazioni anche nella ricerca sociale, permettendo differenziazioni tra gruppi entro la popolazione. I confronti tra gruppi, normalmente definiti dalla loro collocazione territoriale, possono essere estesi allo studio delle relazioni tra sottogruppi di una stessa popolazione (gruppi o classi sociali, etnici, religiosi, professionali...). L'applicazione appare ricca di prospettive, data la difficoltà di reperire dati idonei a discriminare tra i sottogruppi di una stessa popolazione.

La corrispondenza tra cognomi e geni non è certamente completa. Problemi teorici derivano dall'origine polifiletica dei cognomi: portare lo stesso cognome non significa necessariamente avere un'origine comune. Problemi pratici derivano da illegittimità, adozioni, errori di trascrizione e trasformazioni legate allo sviluppo della lingua: essi introducono variazioni a livello individuale, ma non alterano, in genere, la sostanziale capacità del sistema cognomi di tipizzare una popolazione. Diversa poi è la profondità del tempo indagabile: i cognomi hanno una storia recente, assai breve rispetto ai tempi dell'evoluzione biologica; di contro essi, potendo essere rilevati anche presso popolazioni del passato, offrono la possibilità di seguire direttamente e sul concreto le modificazioni che intervengono al passare delle generazioni.

I cognomi, rispetto ai marcatori biologici, presentano soprattutto il concreto vantaggio di poter essere rilevati con facilità, in tempi brevi, con costi ridotti. La loro rilevazione può essere estesa alla globalità della popolazione; spesso può essere derivata direttamente dagli ormai numerosi elenchi, anche informatizzati, raccolti in varia forma da istituzioni pubbliche e private.

Le fonti dalle quali attingere l'informazione 'cognome' possono essere sia di stato (censimenti, anagrafi, elenchi di presenti o residenti, elenchi dei capifamiglia...), sia di flusso (nascite, matrimoni, decessi). Le diverse fonti si rivelano descrittive di insiemi diversi: in alcuni casi (ad esempio nei cognomi dei nati) sono considerate le persone native nella zona o l'insieme di individui (maschi) che concretamente contribuisce alla continuità in loco della popolazione; in altre (ad esempio nei matrimoni o nei decessi) le persone che in quella zona hanno vissuto momenti particolari, non implicando necessariamente l'effettiva origine o la residenza in loco (ad esempio i cognomi degli sposi nelle società patrilocali non sono indicativi degli individui effettivamente residenti nella popolazione). La durata del periodo d'osservazione introduce inoltre differenze strutturali nelle distribuzioni dei cognomi raccolti da fonti di flusso.

Le fonti di stato offrono descrizioni 'istantanee' della struttura in cognomi: possono considerare tutta la popolazione, implicando nel caso la ripetitività dovuta ai componenti una stessa famiglia, oppure i soli capifamiglia, nel caso vengono trascurati i cognomi delle donne coniugate o di membri aggiunti alla famiglia, utili in particolari tipi di analisi.

Altro vantaggio offerto dai cognomi è la possibilità, come detto, di estenderne la rilevazione alle generazioni passate. Il sistema cognomi offre l'opportunità di se-

guire concretamente – almeno negli ultimi secoli – l’evoluzione di una popolazione, di superare cioè quell’ambiguità di fondo, che sempre persiste quando il passato è ricostruito e interpretato a posteriori mediante proiezione a ritroso della situazione attuale, senza però che vi possa essere alcuna verifica sperimentale.

Dalla ricca bibliografia ormai disponibile, emerge che cognomi e geni seguono dinamiche simili: analisi condotte a partire dai due tipi di dati portano a conclusioni generalmente concordanti (Lucchetti *et al.* 1989; Zei *et al.* 1993).

Applicazioni analoghe possono essere effettuate anche a partire dalla distribuzione dei nomi delle persone, presi separatamente per i due sessi. In tal caso però non è presente una relazione vincolante nelle modalità di trasmissione, cadono quindi gli aspetti correlabili con l’ereditarietà biologica. I nomi forniscono utili indicazioni nel definire l’identità culturale di un gruppo e permettono di inferire sull’intensità e velocità che, nei vari periodi storici o nelle diverse realtà territoriali, caratterizzano l’evoluzione culturale dei gruppi umani.

4. Uso dei cognomi in analisi sulle popolazioni. A livello di intere popolazioni la struttura per cognomi è utile alla definizione dell’identità al tempo stesso culturale e biologica dei gruppi, alla ricostruzione della loro storia, all’individuazione delle relazioni intercorse tra esse.

I cognomi permettono di sviluppare analisi sulla ‘struttura interna’ di una singola popolazione (variabilità in essa presente, suo isolamento o valutazione della mobilità, modalità della formazione delle coppie e comportamento matrimoniale) e della sua evoluzione nel tempo, se la rilevazione dei cognomi è estesa a più generazioni.

Il confronto tra più popolazioni e la stima del grado di similarità o di distanza che le lega permette di inferire sulle interazioni tra esse e sulla loro storia: se derivanti da una comune origine o da una stessa matrice culturale, se tra esse nel tempo vi è stato, in che misura e direzione, interscambio di individui.

4.1. Cognomi e struttura interna di una popolazione. La numerosità relativa delle forme cognominali e la loro distribuzione per ripetitività forniscono informazioni rilevanti sulla struttura e sulle caratteristiche di una popolazione. Sono intuitive le differenze tra le due condizioni estreme: da un lato, una popolazione con pochi cognomi, tutti o quasi con elevata ripetitività; dall’altro, una popolazione con un elevato numero di forme cognominali e, in media, un ridotto numero di individui che portano uno stesso cognome. La prima presuppone un forte isolamento, senza immissione di individui dall’esterno: la continuità è affidata alle sole famiglie locali, il numero di cognomi tende a ridursi, con conseguente aumento della frequenza relativa dei cognomi che permangono. Nella seconda popolazione la maggiore varietà in cognomi lascia presumere l’immissione di persone dall’esterno, che importano nuovi cognomi, la ripetitività media dei quali rimane ridotta.

4.1.1 Numerosità delle forme cognominali: il rapporto S/N. Il numero di cognomi differenti (S) in rapporto al numero di individui (N) può essere usato come misura intuitiva della variabilità interna di una popolazione. A parità di dimensioni

(N), il numero S di forme cognominali dipende dalle caratteristiche della comunità, in particolare dal grado di mobilità, o viceversa di isolamento, che la contraddistingue.

Il rapporto S/N è un parametro elementare, seppure grezzo, che stima la 'ricchezza' in cognomi di una popolazione. Tanto più il valore del rapporto si approssima all'unità, tanto maggiore è la ricchezza in forme cognominali e la variabilità interna della comunità; al limite, il valore unitario del rapporto indicherebbe una popolazione, i componenti della quale sono tutti portatori di cognomi differenti. Viceversa, bassi valori di S/N indicano la presenza, nella comunità, di poche forme cognominali, ognuna delle quali portata da un numero relativamente elevato di individui. Nel primo caso, il cognome possiede un elevato potere di identificazione del soggetto all'interno del gruppo, mentre scarso è il suo potere di identificazione della popolazione cui l'individuo appartiene; nel secondo, il cognome non permette l'identificazione dell'individuo all'interno del gruppo, ma può assumere elevata capacità di identificazione della comunità.

Il rapporto S/N risente anche del tipo di fonti utilizzate. Per le fonti di flusso è rilevante la durata del periodo di osservazione: il numero di cognomi diversi (S) non gode della proprietà additiva, come il numero di eventi o di persone (N) considerati. Come già ricordato le diverse possibili fonti (nascite, matrimoni, decessi) prendono in esame insieme con differente significato. Negli atti di matrimonio – i più usati per queste ricerche – i valori del rapporto S/N risultano più elevati, anche in relazione alla loro minore numerosità. Si notano differenze in funzione dell'isolamento della popolazione e delle dimensioni del mercato matrimoniale, ma anche in funzione del sesso: S/N presenta valori più elevati tra gli sposi, più spesso di provenienza da altri paesi, che tra le spose, generalmente del posto. Ricordiamo che nelle società patrilocali solitamente il matrimonio viene celebrato nella parrocchia della sposa, indipendentemente dalla località ove la coppia prende poi residenza: i cognomi degli sposi desunti dagli atti di matrimonio risultano quindi poco rappresentativi della popolazione residente.

Nella tabella 1 vengono riportati i valori ottenuti dalle serie storiche dei dati di flusso in due popolazioni prese come esempio: Alia in provincia di Palermo e Civitella del Tronto in provincia di Teramo (De Iasio 2000). Analoghe per dimensioni, esse sono caratterizzate da un diverso grado di isolamento: più chiusa la popolazione di Alia, più aperta la popolazione di Civitella del Tronto. Il periodo considerato e la suddivisione in sottoperiodi sono identici.

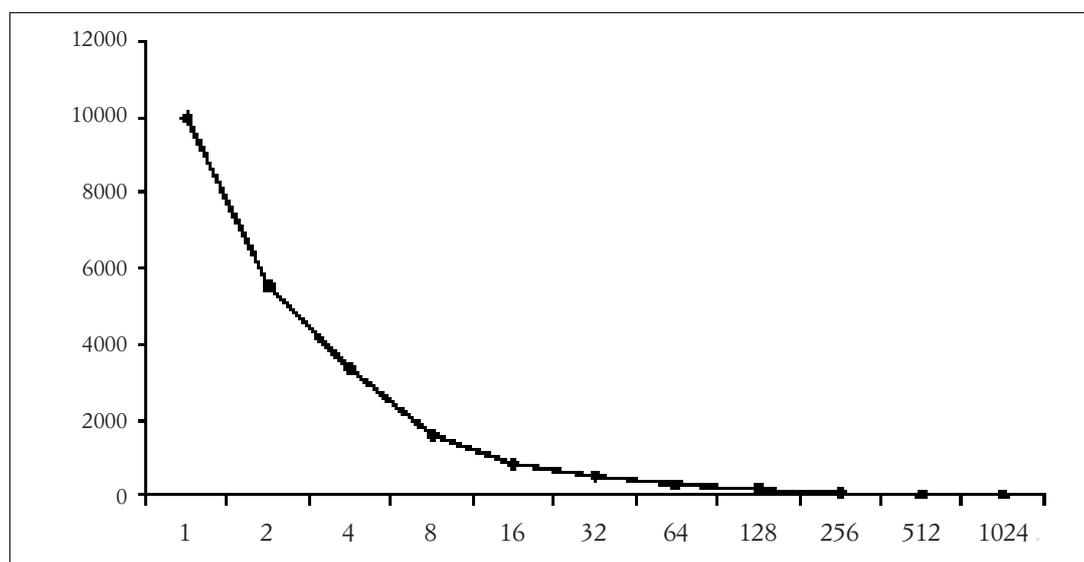
I valori ottenuti considerando nella sua globalità tutto il periodo, a causa della non additività di S, risultano ovviamente inferiori rispetto a quelli ricavati per ogni singolo sottoperiodo.

Le fonti di stato (censimenti, *Status Animarum*, elenchi vari di residenti o di famiglie) possono fare riferimento a tutti gli individui della popolazione oppure possono considerare i soli capifamiglia. Un esempio specifico di utilizzazione delle fonti di stato verrà proposto di seguito, con riferimento alla popolazione attuale residente nel comune di Parma.

Tab. 1. Valori di S/N ricavati da fonti di flusso presso le popolazioni di Alia (PA) e Civitella del Tronto (TE)

Periodo	NASCITE			SPOSI			SPOSE			MORTI		
	S	N	S/N	S	N	S/N	S	N	S/N	S	N	S/N
ALIA												
1815-29	315	2513	0,125	185	501	0,369	171	501	0,341	310	1746	0,178
1830-44	265	2396	0,111	219	607	0,361	197	607	0,325	383	2004	0,191
1845-59	309	2984	0,104	228	760	0,300	212	760	0,279	351	2083	0,169
1860-74	453	3901	0,116	259	824	0,314	244	824	0,296	290	1386	0,209
Totale	674	11794	0,057	429	2692	0,159	377	2692	0,140	644	7219	0,089
CIVITELLA												
1815-29	528	2624	0,201	294	547	0,537	236	547	0,431	639	2300	0,278
1830-44	608	2792	0,218	266	426	0,624	228	426	0,535	633	1832	0,346
1845-59	543	2814	0,193	350	643	0,544	299	644	0,464	713	2159	0,330
1860-74	693	3321	0,209	339	597	0,568	316	597	0,529	748	2498	0,299
Totale	1163	11551	0,101	697	2213	0,315	577	2214	0,261	1407	8789	0,160

Fig. 1 - Classi di frequenza dei cognomi presenti al 01-01-2000 a Parma



4.1.2 *Distribuzione dei cognomi per classi di frequenza.* Il rapporto S/N, basandosi sul solo numero di cognomi di una popolazione, non sempre risulta capace di discriminare tra situazioni diverse; esso assume lo stesso valore sia quando gli S cognomi sono presenti tutti con uguale frequenza, sia quando tra gli S cognomi una parte ha ricorrenza elevata e un'altra parte ha frequenza ridotta o solo unitaria. È necessario valutare la distribuzione delle forme cognominali in funzione della loro ricorrenza.

za o ripetitività (distribuzione per frequenza o per classi di frequenza); essa mostra in genere un andamento esponenziale negativo, come evidenziato nella figura 1.

Rilevante è la incidenza dei cognomi che compaiono una sola volta (cognomi a frequenza unitaria); essa si diversifica in funzione delle caratteristiche della popolazione e anche della tipologia della fonte utilizzata. Nelle fonti di flusso la loro incidenza è solitamente più elevata, dipendendo dalla numerosità degli eventi e dalla durata del periodo di osservazione. La incidenza dei cognomi a frequenza unitaria si differenzia in funzione del sesso. In popolazioni storiche, specie se isolate, a causa delle modalità di trasmissione dei cognomi, ma soprattutto delle tradizioni patri-locali delle popolazioni, i cognomi a frequenza unitaria risultano in buona misura portati da donne coniugate, inserite nella popolazione col matrimonio. In popolazioni attuali la incidenza dei cognomi a frequenza unitaria, o comunque a bassa ripetitività, è maggiormente correlata alla mobilità di individui singoli, più frequentemente maschi, che migrano per motivi di lavoro.

Zei e altri (1983) hanno mostrato che alla distribuzione per frequenza dei cognomi può essere applicata la funzione di Karlin McGregor (1967), proposta per studiare la distribuzione di alleli neutrali in una popolazione finita, sottoposta a mutazioni e all'azione della deriva. Il modello proposto da Karlin McGregor

$$E(N_k) = \frac{1}{k} \frac{Nv}{1-v} \binom{M-k}{N-k} \binom{M}{N}$$

permette di valutare il numero atteso di alleli (cognomi, nella applicazione) portati da k individui in una popolazione di dimensione N . Il parametro v , ricavabile dalla relazione

$$\frac{S}{N} = \frac{v \ln v}{(v-1)}$$

stima il tasso di immissione di nuovi alleli per mutazione.

Nell'applicazione ai cognomi, v è interpretato come misura del tasso di immissione di cognomi nuovi. La comparsa di nuovi cognomi resta imputabile, fondamentalmente, all'arrivo di nuovi individui: è quindi ragionevole assumere che v rappresenti una valutazione del tasso di immigrazione.

Alla distribuzione per frequenza dei cognomi può essere applicata anche la funzione proposta da Fisher (1943) per stimare il numero di specie e di individui per specie in campioni di insetti

$$E(N_k) = \frac{\alpha x^k}{k}$$

ove $E(N_k)$ è il numero atteso di specie rappresentate da k individui (il numero di cognomi presenti k volte nella popolazione).

Il parametro x viene ricavato dalla relazione (Krebs 1989)

$$\frac{S}{N} = \frac{1-x}{x(-\ln(1-x))}$$

e quindi α , parametro che fornisce una stima della ricchezza in specie del campione (di cognomi nella popolazione), è dato da

$$\alpha = \frac{N(1-x)}{x}$$

I parametri v e α , tra loro correlati, descrittivi il primo della probabilità di immissione di nuovi cognomi e il secondo della ricchezza in cognomi, risultano dipendere da S/N e risentono quindi dei limiti interpretativi, sopra considerati, di S e N nelle singole popolazioni.

La struttura in cognomi di una popolazione può essere analizzata mediante indici biologici messi a punto per stimare la variabilità biologica (biodiversità) in popolazioni naturali o manipolate dall'uomo. Si fa riferimento alla numerosità delle specie (dei cognomi nell'applicazione) e al numero di appartenenti a ciascuna di esse (numero di individui che portano ciascun cognome). La 'biodiversità' è tanto maggiore quanto più numerose sono le forme cognominali e quanto più omogeneamente sono tra esse ripartiti gli individui, raggiungendo il massimo teorico quando tutti gli individui portano cognomi diversi.

L'indice di *species richness* stima il numero atteso \hat{S}_n di specie (cognomi) in un campione di n individui tratto da una popolazione di dimensioni N , ove N_i è il numero di rappresentanti della i -ma specie (portatori dell' i -mo cognome):

$$E(S_n) = \sum_{i=1}^S \left[1 - \frac{\binom{N-N_i}{n}}{\binom{N}{n}} \right]$$

L'indice di Shannon-Wiener stima l'entropia della distribuzione

$$H = - \sum_{i=1}^S p_i \log p_i$$

dove S è il numero di cognomi e p_i la frequenza relativa di ciascuno di essi.

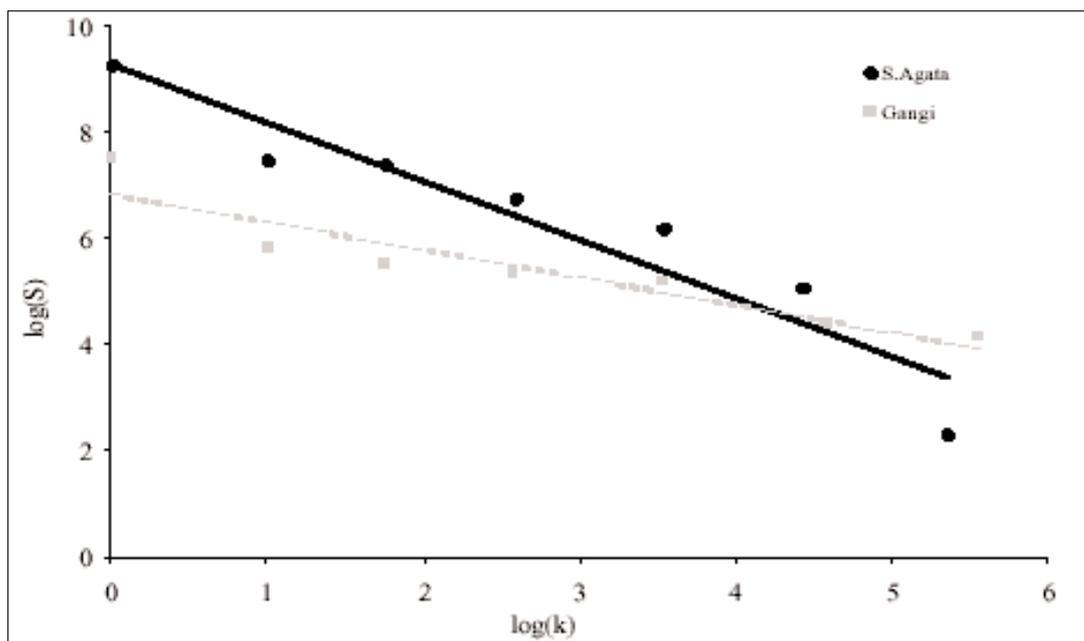
L'indice di Brillouin (preferibile per campioni piccoli con pochi cognomi) stima la eterogeneità della distribuzione,

$$H = \frac{1}{N} \log \left(\frac{N!}{n_1! n_2! n_3! \dots n_S!} \right)$$

ove N è il numero totale di individui, S è il numero di cognomi, n_i il numero di individui che portano l' i -mo cognome.

L'indice di Simpson stima anch'esso la eterogeneità, uguale a 1-D. Con la stessa simbologia si ha

$$D = \sum_{i=1}^S p_i^2$$

Fig. 2. Metodo $\log_2\text{-}\log_2$ applicato a due comuni della Sicilia

D è una misura dell'isonimia, dovuta al caso, presente nelle popolazioni (probabilità che due individui presi a caso portino lo stesso cognome); l'indice 1-D di Simpson indica la probabilità che due individui, scelti a caso nella popolazione, siano portatori di cognomi diversi.

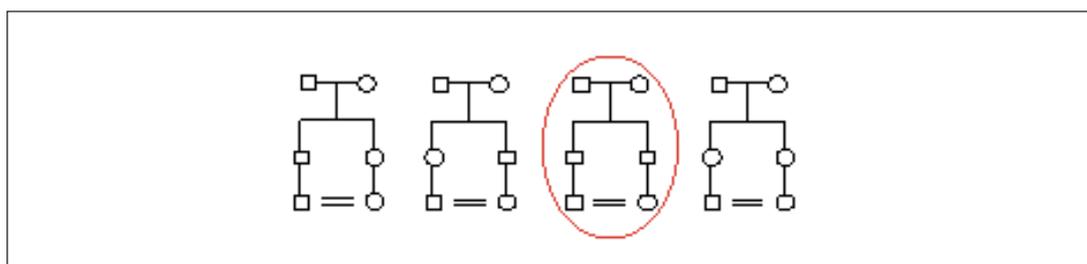
Barrai e altri (1987; 1989), riportando in un diagramma cartesiano sull'asse delle ascisse il \log_2 di k e sull'asse delle ordinate il \log_2 di S_k (S_k rappresenta il numero di cognomi con ripetitività k), hanno mostrato che i punti tendono a disporsi linearmente. In popolazioni diverse le rette differiscono per la loro pendenza, in rapporto alla numerosità dei cognomi rispettivamente a bassa o elevata ripetitività, fornendo un indice di valutazione del grado di apertura e di mobilità della popolazione.

La figura 2 mostra una applicazione del metodo a due comuni siciliani, S. Agata (CT) e Gangi (PA) (Lucchetti *et al.* 1990): la prima popolazione evidenzia una distribuzione cognominale caratterizzata da un più grande numero di cognomi scarsamente ripetuti e pochi cognomi ad elevata ripetitività; la seconda presenta una maggiore omogeneità nella consistenza numerica delle varie classi di frequenza dei cognomi. Si può pertanto dedurre che la popolazione residente a S. Agata è maggiormente sottoposta a flussi d'ingresso, mentre quella di Gangi appare più chiusa. L'applicazione a 386 comuni della Sicilia, con dati ricavati dagli elenchi telefonici, ha mostrato una correlazione significativa tra stima del tasso di migrazione predetta dal modello e valori reali ottenuti da fonti demografiche ufficiali.

4.1.3 Isonimia, consanguineità e modalità di formazione delle coppie. La prima utilizzazione dei cognomi in biologia delle popolazioni umane è attribuita a G. Darwin (1875), che stimò la frequenza dei matrimoni tra cugini a partire dalla incidenza dei

matrimoni isonimici, dei matrimoni cioè tra persone portanti lo stesso cognome. Crow e Mange (1965) individuarono una precisa relazione tra consanguineità ed isonimia. La corretta stima della consanguineità deve considerare non solo il numero dei matrimoni consanguinei, ma anche l'effettivo grado di consanguineità tra i coniugi; la probabilità che due coniugi consanguinei portino lo stesso cognome varia in funzione del livello di consanguineità e della distribuzione per sesso degli antenati intermedi. Ad esempio: due fratelli, legati da consanguineità $F_1=1/4$, portano sempre lo stesso cognome ($I_1=1$); due cugini primi, legati da consanguineità $F_1=1/16$, portano lo stesso cognome solo quando figli di due fratelli, cioè in una sola delle 4 possibili tipologie di alberi genealogici ($I_1=1/4$) (fig. 3); sempre in relazione al sesso degli antenati intermedi esistono 16 tipologie per i matrimoni tra cugini secondi (consanguineità $F_1=1/64$): i due coniugi portano lo stesso cognome solo nel caso in cui tutti gli antenati intermedi siano maschi ($I_1=1/16$). Crow e Mange osservarono che F_1/I_1 , il rapporto tra il grado di consanguineità che lega i due coniugi e la probabilità che essi portino lo stesso cognome, è costante, pari a $1/4$.

Fig. 3. Consanguineità tra cugini primi



La frequenza con la quale coniugi consanguinei portano lo stesso cognome permette dunque di inferire l'ammontare della consanguineità totale presente nella popolazione.

Applicazioni che prendono in considerazione la ripetitività di coppie di cognomi tra coniugi permettono di estendere l'analisi del comportamento matrimoniale (modalità di formazione delle coppie) a tutta la popolazione.

Lasker e Kaplan (1985), allo scopo di individuare e valutare la presenza di sottopopolazioni entro le quali i matrimoni avvengono in modo preferenziale, calcolano il numero di ripetizioni negli abbinamenti dei cognomi dei coniugi (RP: *Repeated Pairs*):

$$RP = \frac{\sum [S_{ij}(S_{ij} - 1)]}{N(N - 1)}$$

dove S_{ij} rappresenta il numero di coppie nelle quali il marito ha il cognome i -mo e la moglie il cognome j -mo e N è il numero totale di coppie.

L'indice RP varia da 0 (tutti gli abbinamenti di cognomi compaiono solo una volta) a 1 (tutte le coppie hanno la stessa combinazione di cognomi). Il valore atteso $E(RP)$ per abbinamenti casuali dei cognomi può essere calcolato secondo il metodo proposto da Chakraborty (1985). Il confronto tra i valori osservati e quelli attesi in base all'ipotesi di unioni casuali permette di ponderare la componente non

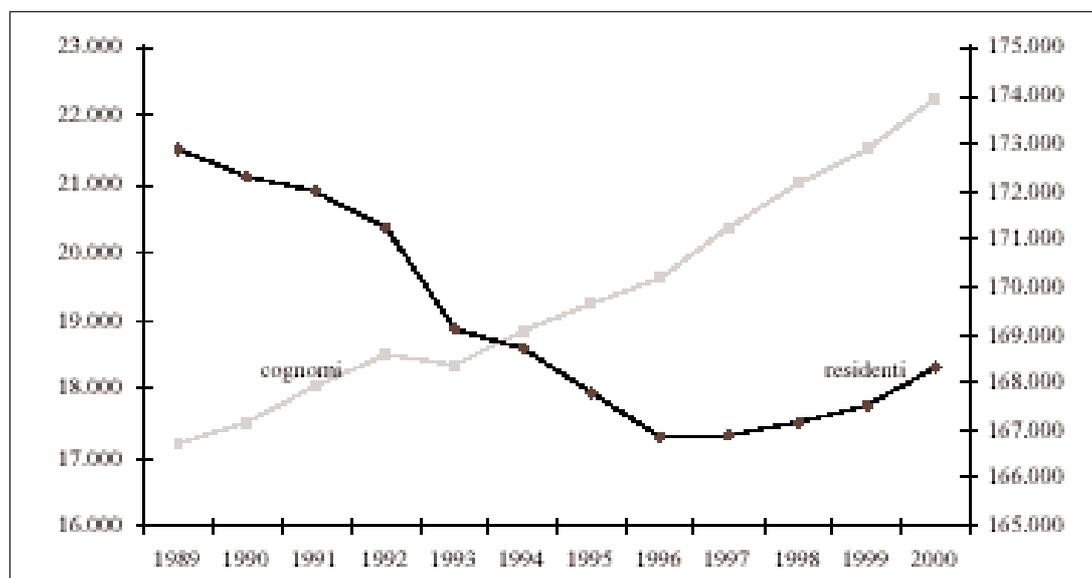
random rispetto a quella random; quindi di valutare l'incidenza di scelte preferenziali nella formazione delle coppie, espressione della presenza di suddivisioni all'interno della popolazione (Mascie-Taylor *et al.* 1987). Applicazioni a popolazioni storiche (Relethford 1992) del Massachusetts mostrano che la componente random di RP è associata negativamente alle dimensioni della popolazione e all'esogamia. L'eccesso relativo di RP – la componente non random – risulta tuttavia correlato positivamente con le stesse variabili, evidenziando come la probabilità di stratificazioni o suddivisione aumenti al crescere delle dimensioni della popolazione.

Applicazioni a popolazioni con diverso grado di isolamento (Gueresi *et al.* 1996) hanno permesso valutazioni del livello di suddivisione riconducibile a barriere geografiche e/o culturali o etniche.

4.2 Uso delle fonti di stato: analisi sulla popolazione di Parma (1989-2000). Dal 1989 al 2000, all'inizio di ogni anno, sono stati rilevati dall'anagrafe i cognomi dei residenti nella città di Parma (Pizzetti 2000). Le analisi sviluppate a partire dalla struttura in cognomi e dalle variazioni da essa subite nel decennio 1989-2000 riguardano sia la descrizione della struttura della popolazione, sia valutazioni sulla mobilità degli individui, sull'entità e direzione dei flussi migratori.

4.2.1 Numerosità dei residenti, numerosità e frequenza dei cognomi. Il numero di residenti nel comune di Parma è diminuito, dal 1989 al 2000, da 172.854 a 168.336 (il calo è continuo e costante dal 1989 al 1996; la rilevazione del 1.1.1993 mostra anomalie a causa delle verifiche effettuate a seguito del censimento 1991; negli ultimi anni si osserva un lieve recupero). Al contrario, il numero di cognomi mostra un continuo e costante aumento, indipendentemente dalla diminuzione del volume della popolazione (fig. 4).

Fig. 4. Distribuzione del numero di individui residenti (scala a destra) e del numero di cognomi (scala a sinistra) a Parma dal 1989 al 2000



Tab. 2. Residenti, cognomi rapporto S/N e altri parametri calcolati sulla popolazione residente a Parma dal 1989 al 2000

anno	S	N	S/N	v	alfa
1989	17.194	172.854	0,0995	0,0267	4,7418
1990	17.528	172.313	0,1017	0,0275	4,8726
1991	18.031	171.989	0,1048	0,0287	5,0819
1992	18.513	171.228	0,1081	0,0230	4,0310
1993	18.340	169.129	0,1084	0,0300	5,2308
1994	18.853	168.700	0,1118	0,0312	5,4329
1995	19.248	167.785	0,1147	0,0324	5,6183
1996	19.609	166.847	0,1175	0,0334	5,7652
1997	20.376	166.916	0,1221	0,0352	6,0898
1998	21.016	167.165	0,1257	0,0366	6,3507
1999	21.511	167.523	0,1284	0,0377	6,5630
2000	22.253	168.336	0,1322	0,0392	6,8680

Il rapporto tra il numero (S) di cognomi ed il numero (N) di residenti (tab. 2) aumenta da 0,0995 (1989) a 0,1322 (2000). Nel 1989 su 100 persone prese a caso si trovano 9,9 forme cognominali diverse (o ciascun cognome risulta in media portato da 10,05 persone); al 2000 ogni cognome è portato, in media, da 7,56 residenti, con la presenza di 13,2 cognomi diversi ogni 100 residenti. Nella tabella vengono anche riportati i valori di v , stima del tasso di immissione di nuovi cognomi, e di α , stima della ricchezza in cognomi, entrambi in evidente crescita lungo il decennio.

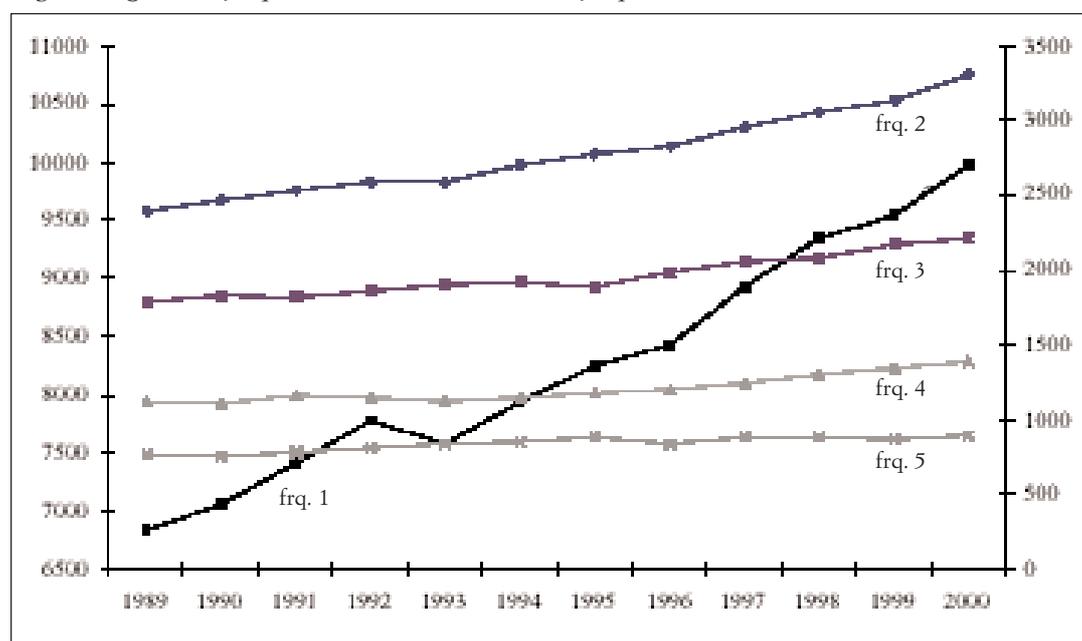
La figura 5 mostra l'evoluzione del numero assoluto dei cognomi con ricorrenza da 1 a 5: la frequenza dei cognomi presenti con una sola ricorrenza aumenta da meno di 7.000 a circa 10.000. Dal 1989 al 2000 la incidenza dei cognomi a frequenza unitaria, sempre superiore al 40% dei cognomi presenti, aumenta di quattro punti percentuali. Il numero di cognomi appartenenti alle classi con ripetitività superiore mostra un incremento sempre più ridotto al crescere della ripetitività stessa; già per i cognomi a ripetitività 5 l'incremento lungo il decennio è praticamente trascurabile. La distribuzione per classi di frequenza al 2000 (fig. 1) evidenzia l'andamento esponenziale negativo della numerosità dei cognomi al crescere della loro ripetitività.

I sette cognomi a più elevata ripetitività sono sempre gli stessi in ognuna delle rilevazioni annuali, ma il numero dei loro portatori va continuamente diminuendo (dai 6.265 del 1989 ai 5.525 del 2000).

4.2.2 Mobilità in cognomi e mobilità degli individui. La tabella 3 mostra, per ogni anno, il numero di cognomi che risultano nuovi rispetto alla prima rilevazione (del 1989) e di quelli che risultano nuovi rispetto alla rilevazione dell'anno precedente. In modo analogo, viene ricostruita la dinamica della scomparsa dei cognomi, per ciascun anno, rispetto sia alla rilevazione iniziale sia a quella appena antecedente.

Ad ogni rilevazione una quota di cognomi compresa tra il 4 e il 6% risulta nuova rispetto all'anno precedente; in numero assoluto tra i residenti nella città com-

Fig. 5. Cognomi a frequenza 1 (scala a sinistra) e frequenza da 2-5 (scala a destra)



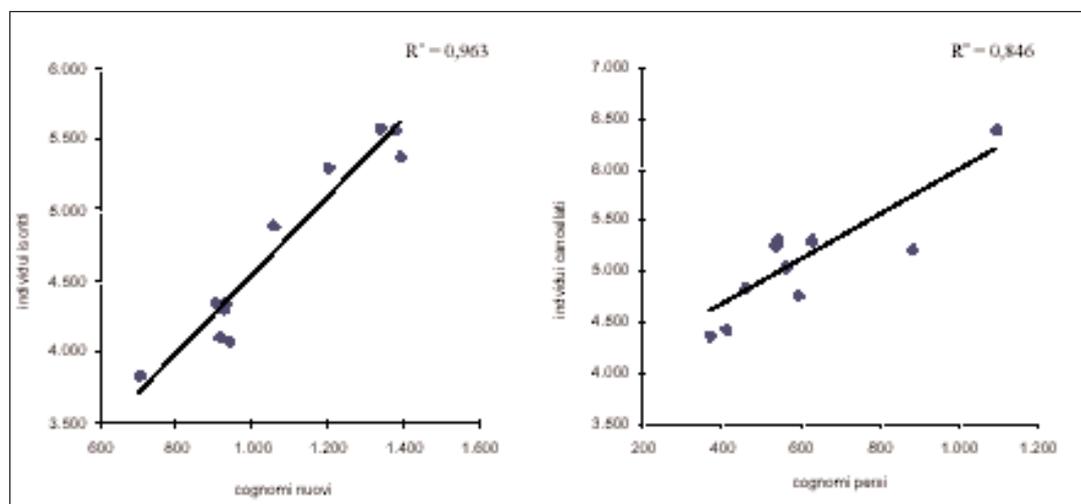
Tab. 3. Cognomi nuovi e cognomi scomparsi rispetto alla prima rilevazione del 1989 e rispetto all'anno precedente

	residenti		cognomi nuovi rispetto				cognomi scomparsi rispetto			
			cognomi	1989	anno precedente	1989	anno precedente			
			cognomi	%	cognomi	%	cognomi	%	cognomi	%
1990	172.313	17.528	699	3,99%	699	3,99%	365	2,08%	365	2,08%
1991	171.989	18.031	1.493	8,28%	910	5,05%	656	3,74%	407	2,26%
1992	171.228	18.513	2.248	12,14%	936	5,06%	929	5,30%	454	2,45%
1993	169.129	18.340	2.735	14,91%	921	5,02%	1.589	9,07%	1.094	5,97%
1994	168.700	18.853	3.465	18,38%	1.050	5,57%	1.806	10,30%	537	2,85%
1995	167.785	19.248	4.006	20,81%	924	4,80%	1.952	11,14%	529	2,75%
1996	166.847	19.609	4.528	23,09%	900	4,59%	2.113	12,05%	539	2,75%
1997	166.916	20.376	5.448	26,74%	1.389	6,82%	2.266	12,93%	622	3,05%
1998	167.165	21.016	6.180	29,41%	1.196	5,69%	2.358	13,45%	556	2,65%
1999	167.523	21.511	6.892	32,04%	1.372	6,38%	2.575	14,69%	877	4,08%
2000	168.336	22.253	7.709	34,64%	1.331	5,98%	2.650	15,12%	589	2,65%

paiono ogni anno 1.000 e più cognomi nuovi; il trend è in aumento. Dei 22.253 cognomi presenti alla rilevazione del 1.1.2000, 7.709 (34,6%) risultano nuovi rispetto alla rilevazione iniziale: più di un terzo di quelli attualmente presenti sono stati introdotti nel corso dell'ultimo decennio. Di essi il 72% è comparso con frequenza unitaria; i cognomi portati da un solo individuo comprendono, ogni anno, l'86-87% dei cognomi nuovi.

Ogni anno scompaiono dalla popolazione cittadina una quota di cognomi pari

Fig. 6. Correlazione tra numero di cognomi nuovi e numero di individui nuovi iscritti (a sinistra) e tra numero di cognomi scomparsi e numero di individui cancellati



a circa al 2-4%; il confronto con la rilevazione del 1989 evidenzia che il 15% (2.650 in numero assoluto) di quelli allora presenti è scomparso nel corso degli ultimi dieci anni. Circa l'80% dei cognomi presenti al 1989 e non più presenti al 2000 aveva frequenza unitaria; tra i cognomi che scompaiono anno dopo anno, quelli a frequenza unitaria incidono per l'88-90%.

Le distribuzioni per frequenza dei cognomi rispettivamente scomparsi e scomparsi mette in risalto la grande incidenza dei cognomi a frequenza unitaria e la loro importanza nella valutazione dei flussi: sono essi fondamentalmente che determinano le variazioni del numero dei cognomi nella popolazione.

L'analisi può condurre a inferenze sulla entità delle migrazioni: esiste una relazione tra la dinamica di comparsa/scomparsa dei cognomi e la mobilità in entrata e in uscita degli individui? La mobilità osservata per i cognomi può fornire indicazioni sulla mobilità degli individui?

La conoscenza del numero dei nuovi iscritti e dei cancellati, ricavata dai dati anagrafici, permette di verificare la presenza di una regressione lineare tra la mobilità in cognomi e mobilità in individui. La regressione lineare ($b = 2,74$) tra il numero di nuovi cognomi e il numero di individui iscritti mostra una elevata capacità predittiva ($R^2 = 0,963$); la regressione lineare tra il numero di cognomi scomparsi e la cancellazione di individui ($b = 2,18$) ha una predittività ($R^2 = 0,846$) un poco inferiore (fig. 6).

Un nuovo immigrato non introduce necessariamente un nuovo cognome, così come una cancellazione non implica necessariamente la scomparsa di un cognome; certo che la comparsa di un nuovo cognome implica l'iscrizione di almeno un nuovo individuo, così come la scomparsa di un cognome implica l'uscita di almeno una persona. La regressione tra mobilità in cognomi e mobilità in individui mostra una diversa capacità predittiva in funzione della direzione della mobilità stessa. Sulle nuove iscrizioni le nascite, che in genere non comportano l'immissione di nuovi cognomi, hanno incidenza minore dell'immigrazione. Durante il 1999 l'82% delle

nuove iscrizioni è dovuta a immigrazione; i 4.593 immigrati introducono 1.279 nuovi cognomi, dei quali 1.097 sono a frequenza unitaria. Le cancellazioni sono dovute a morte o ad emigrazione. Le prime, che con minore probabilità portano alla scomparsa di un cognome, hanno sulle cancellazioni una rilevanza maggiore di quella delle nascite sulle nuove iscrizioni. Inoltre, l'uscita di autoctoni, che ha una qualche rilevanza tra le cancellazioni per emigrazione, meno probabilmente implica la scomparsa di un cognome.

L'inserimento di un cognome non implica che lo stesso permanga poi con continuità nella popolazione. Cognomi che compaiono per la prima volta possono scomparire negli anni successivi e quindi ricomparire di nuovo dopo qualche anno di assenza. La differenza esistente tra il valore cumulato di cognomi nuovi comparso ciascun anno e quello di cognomi nuovi rispetto al 1989 fornisce una valutazione dell'andirivieni dei cognomi. Anche la regressione tra l'andirivieni dei cognomi e quello degli individui (usciti che dopo qualche anno rientrano nella popolazione) conferma l'elevata capacità predittiva dei cognomi, almeno nel contesto delle popolazioni moderne.

Date le caratteristiche dei flussi migratori attuali – relativi più a singoli che a gruppi familiari, con più elevata mobilità di stranieri, più facilmente caratterizzati da cognomi nuovi per la popolazione d'arrivo – i cognomi si rivelano uno strumento adeguato per inferenze sulla mobilità individuale. Nel caso, è conveniente che i cognomi siano rilevati su tutta la popolazione, essendo importante l'incidenza dei cognomi a frequenza unitaria. Per popolazioni nelle quali la mobilità avviene in prevalenza per nuclei familiari – non a livello individuale – l'analisi potrebbe essere più coerentemente sviluppata utilizzando i cognomi dei soli capifamiglia. Per le popolazioni storiche, la regressione esistente potrebbe rappresentarne una modalità di stima, una volta che ne siano stati definiti i parametri in gruppi campione.

4.2.3 Natalità presso i portatori di nuovi cognomi. Le nascite, come già osservato, in genere non introducono nuovi cognomi; per ogni anno comunque vi sono 20-40 nascite di individui che risultano portatori di un cognome non presente nella rilevazione dell'anno precedente; lungo il decennio il loro numero tende ad aumentare.

I portatori di cognomi nuovi sono immigrati; anche se certo non ne rappresentano la totalità, essi possono essere assunti come un insieme – facilmente identificabile – dal quale trarre inferenze sui comportamenti degli immigrati. Dal rapporto tra il numero di nati portatori di cognomi nuovi e il totale degli individui portatori di cognomi nuovi è possibile stimare un tasso generico di natalità che, seppure certamente improprio, dà indicazione sulla natalità presso gli immigrati.

La tabella 4 riporta, a sinistra, il tasso generico di natalità nella popolazione totale, mantenutosi lungo il decennio intorno al 6‰; a destra vengono riportati il numero dei nati e degli individui che risultano portatori di cognomi nuovi, rispetto sia alla rilevazione iniziale del 1989, sia alla rilevazione dell'anno precedente ed i rispettivi rapporti per mille.

Si ha indicazione di una più elevata fecondità tra gli immigrati, specie nell'anno stesso del loro arrivo. La considerazione dei cognomi nuovi rispetto al 1989 indica

Tab. 4. *Tasso di natalità generico e tassi di natalità desunti dai cognomi*

	individui che portano cognomi nuovi rispetto al						
	popolazione totale	1989			anno precedente		
		nati	residenti	tasso	nati	residenti	tasso
1990	5,86	23	823	27,95	23	823	27,95
1991	6,22	48	1.859	25,82	30	1.062	28,25
1992	5,79	64	2.916	21,95	26	1.130	23,01
1993	5,91	44	3.639	12,09	17	1.111	15,30
1994	5,47	74	4.701	15,74	29	1.253	23,14
1995	5,81	69	5.484	12,58	18	1.083	16,62
1996	5,88	107	6.321	16,93	24	1.096	21,90
1997	6,23	139	7.881	17,64	45	1.680	26,79
1998	6,23	157	9.172	17,12	34	1.429	23,79
1999	6,08	174	10.514	16,55	34	1.626	20,91
2000	5,94	183	12.036	15,20	36	1.331	27,05

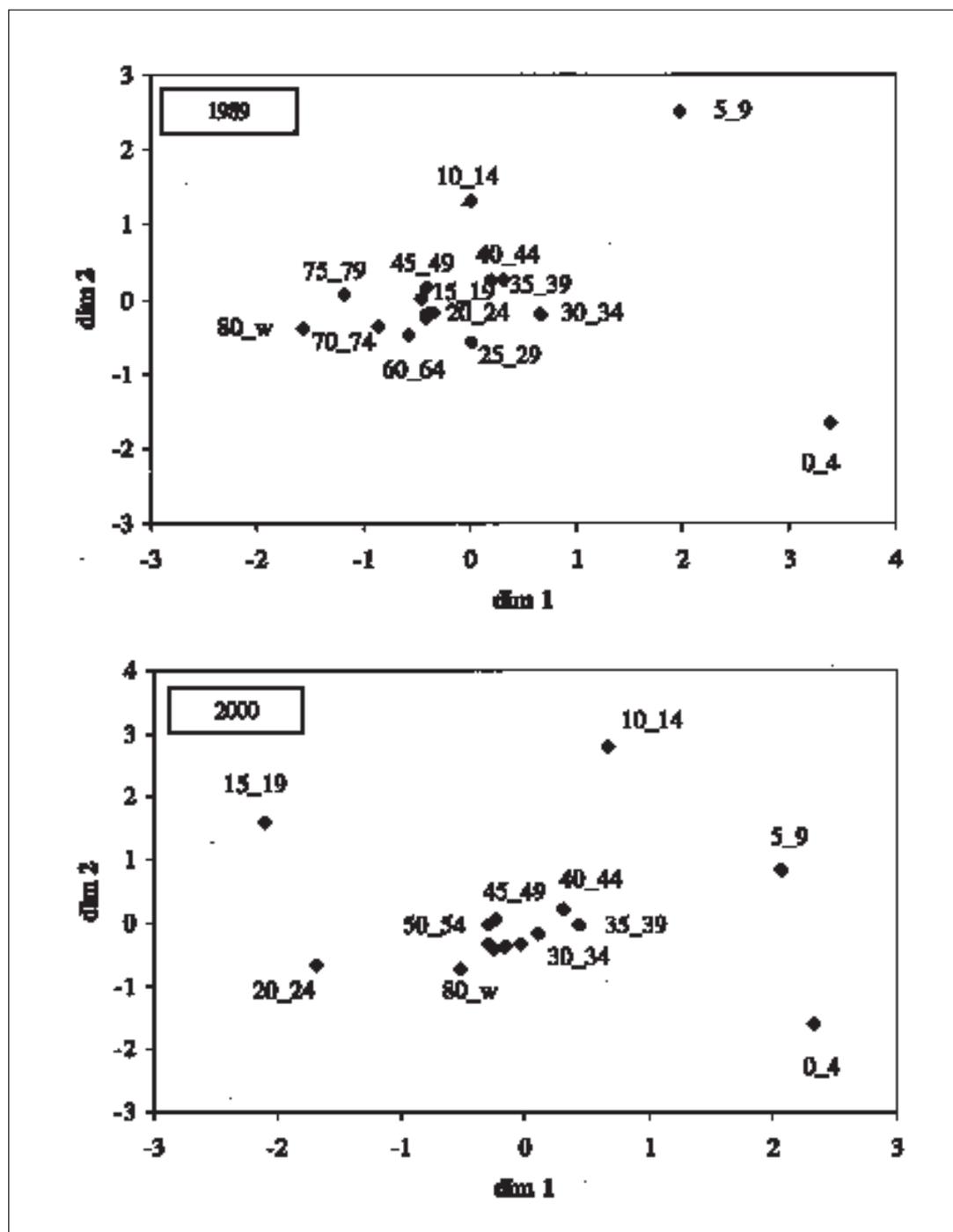
che la fecondità dei portatori di cognomi nuovi, pur mantenendosi sempre su livelli circa tre volte maggiori di quelli della popolazione totale, tende a diminuire al perdurare della presenza nella popolazione d'arrivo. La comparsa di cognomi nuovi è in gran parte dovuta a immigrazione dall'estero, per lo più da paesi nei quali la fecondità è decisamente più elevata. Come già ricordato, nell'analisi vengono considerati solo gli immigrati regolarmente iscritti all'anagrafe; spesso la nascita di un figlio è anche occasione per la regolarizzazione della presenza degli immigrati.

4.2.4 Struttura per cognomi e classi d'età. È possibile ricavare la struttura per cognomi di ciascuna classe di età e indagare quindi quali siano le età nelle quali maggiore è la differenziazione in cognomi indotta dalle variazioni comportamentali e dai flussi migratori. Il confronto può esser fatto calcolando una matrice di similarità tra classi di età e, da essa, una rappresentazione topologica (fig. 7) che sintetizzi le relazioni tra le classi (maggiori dettagli sul metodo vengono dati al paragrafo 4.3).

La rappresentazione bidimensionale della matrice di similarità tra classi di età quinquennali mostra che, a Parma, alla rilevazione del 1989 le classi di età infantili e giovanili si separano dal cluster che comprende la maggior parte delle classi, evidenziando per esse una più elevata eterogeneità in cognomi. La disposizione delle classi lungo la prima dimensione segue un ordinamento funzione dell'età, sfilando da destra verso sinistra all'aumentare dell'età. Le classi 0-4 e 5-9 si staccano nettamente a mostrare la più elevata variabilità in cognomi in esse presente. Seguono, da destra verso sinistra sulla prima dimensione, le classi dai 30 ai 40 anni, quelle cioè dei genitori degli individui 0-4 e 5-9; la classe 10-14 occupa una posizione più centrale sulla prima dimensione (più simile a quella delle classi di età dei loro probabili genitori), distinguendosi però sulla seconda. Le classi anziane si distanziano lievemente dal cluster centrale, posizionandosi dal lato opposto rispetto alle classi infantili.

La maggior differenziazione tra le classi di età estreme, entrambe caratterizzate

Fig. 7. *Similarità tra i cognomi dei residenti al 1989 ed al 2000 suddivisi per classi d'età quinquennali*



da una minore numerosità rispetto alle classi centrali, testimonia la loro diversa composizione in cognomi: tra gli anziani sono presenti in modo pressoché esclusivo cognomi locali, presso le classi infantili, come prima visto, compaiono con incidenza crescente cognomi di immigrati.

Nella rappresentazione ottenuta considerando i cognomi al 2000 si nota l'usc-

ta dal gruppo centrale delle classi di età 15-19 e 20-24: sono le classi 5-9 e 10-14 del 1989, già allora differenziate. Le classi 0-4 e 5-9, composte da individui nati durante il decennio di osservazione, risultano le più differenziate, a testimoniare la loro eterogeneità (la bassa fecondità autoctona, che implica una bassa ripetitività dei cognomi tra i nati, e la crescente incidenza di nati da immigrati). La differenziazione lungo la prima componente, complessivamente sempre connessa all'età, viene integrata dalle differenziazioni sulla seconda componente, probabilmente influenzata dal combinarsi dell'immigrazione dalle regioni dell'Italia meridionale, in corso ormai da più decenni, e l'immigrazione dall'estero, iniziata con gli anni '90.

4.3 Similarità/distanza tra strutture per cognomi di più popolazioni. Le forme cognominali derivano dalla matrice culturale delle popolazioni nelle quali si sono formati. Spesso richiamano patronimici, mestieri, toponimi, caratteristiche somatiche...; anche quando derivano dalla stessa caratteristica, sovente i cognomi si presentano con forme diversificate nelle varie regioni. Esempio classico in Italia è la distribuzione dei cognomi correlati alla attività di lavorazione del ferro: *Ferrari* nelle regioni centro settentrionali, *Fabbri* e *Magnani* nelle regioni centro orientali, varianti tipo *Ferraris* o *Ferrero* in alcune regioni nord occidentali, *Forgione* nelle regioni meridionali.

Zone geografiche diverse si distinguono per i propri cognomi peculiari, oltre che per le differenti incidenze dei cognomi che condividono. Una elevata similarità tra le strutture per cognomi di due popolazioni lascia supporre una loro derivazione da una comune matrice culturale. Gli scambi migratori aumentano la similarità delle strutture in cognomi tra popolazioni. Migrazioni in entrambe le direzioni portano ad una più rapida omogeneizzazione dei cognomi, mentre migrazioni unidirezionali comportano introduzione di nuovi cognomi nella popolazione di arrivo o variazioni nelle loro frequenze se già presenti. Popolazioni sottoposte a sola emigrazione possono essere soggette, in funzione delle loro dimensioni, a fenomeni di deriva, con estinzione casuale di alcuni cognomi e aumento dell'incidenza relativa di altri, incrementando così la propria differenziazione rispetto alle popolazioni limitrofe.

4.3.1 Matrici di similarità / distanza e rappresentazioni topologiche. In genetica l'affinità tra coppie di popolazioni viene valutata in base alla probabilità che alleli di un dato locus, presi dall'uno e dall'altro dei pool genici messi a confronto, siano tra loro uguali. Nell'applicazione ai cognomi l'indice di similarità è dato dalla probabilità che due individui, presi a caso dall'una e dall'altra popolazione, portino lo stesso cognome.

L'indice di Lasker (1977; 1985) calcola la similarità tra ciascuna coppia di popolazioni

$$R_{ij} = \frac{\sum (S_{ik} S_{jk})}{2N_i N_j}$$

dove S_{ik} e S_{jk} sono rispettivamente il numero di individui appartenenti alle popolazioni i e j che portano il k -mo cognome, e N_i e N_j sono il numero dei componenti le due popolazioni. R_{ij} , indice di similarità, diventa nullo quando le due popolazioni non possiedono forme cognominali in comune. Il valore dell'indice risente del

numero S di forme cognominali presenti nelle due popolazioni: già la presenza di due cognomi, gli stessi in entrambe le popolazioni in ognuna delle quali siano portati ciascuno da una metà dei componenti, riconduce il valore dell'indice a 0,5. Assunto che due popolazioni siano tra loro identiche per struttura dei cognomi, presentino cioè gli stessi cognomi con le stesse frequenze, il valore dell'indice diminuisce al crescere del numero S di cognomi, assumendo al più il valore $1/S$.

Chen e Cavalli-Sforza (1983) hanno introdotto un indice standardizzato

$$R_{ij} = \frac{\sum p_{ik} p_{jk}}{\sqrt{\sum p_{ik}^2 \sum p_{jk}^2}}$$

ove p_{ik} e p_{jk} sono le frequenze del k -mo cognome rispettivamente nelle popolazioni i e j messe a confronto. R_{ij} è zero quando la due popolazioni non hanno alcun cognome comune, diventa uno quando le due popolazioni condividono gli stessi cognomi con frequenze uguali, indipendentemente dal numero di cognomi.

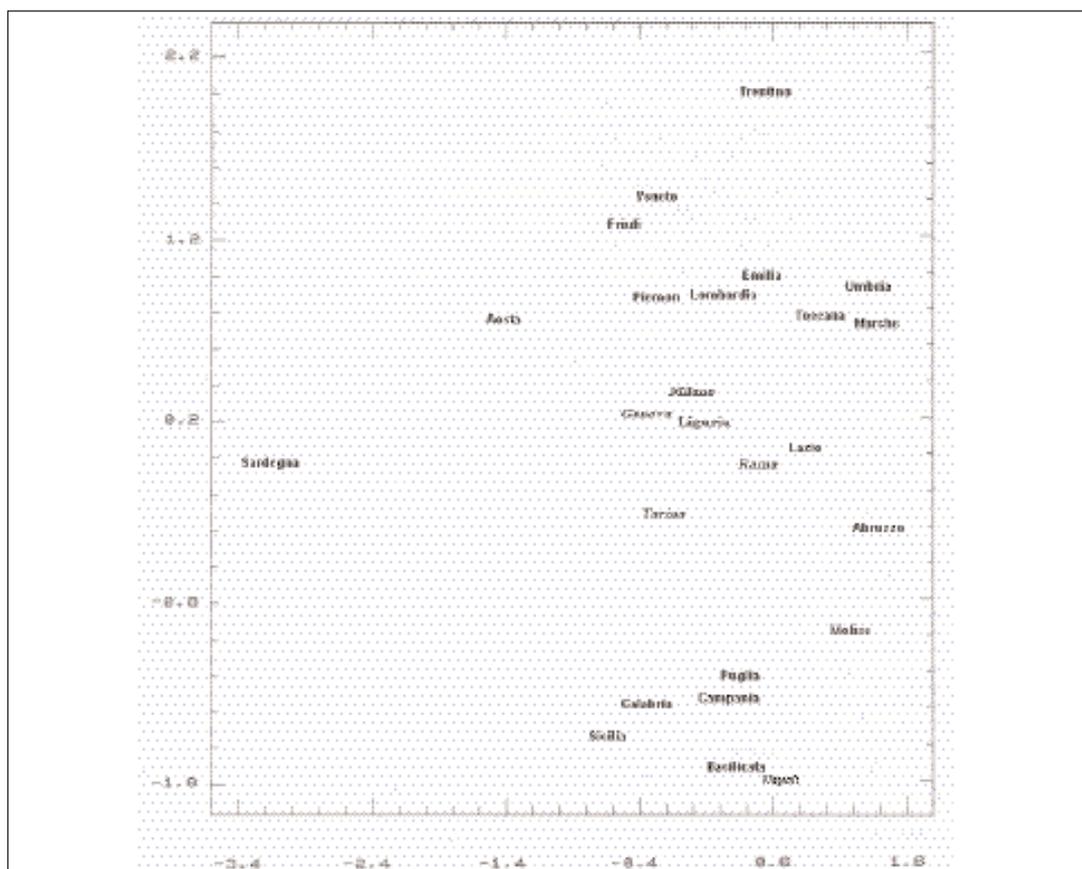
I valori delle relazioni di similarità tra popolazioni possono essere organizzati in una matrice quadrata e simmetrica: i valori della diagonale principale rappresentano l'isonimia interna di ogni popolazione quando si usa l'indice di Lasker, sono unitari quando si usa l'indice s di Chen-Cavalli Sforza.

Le matrici di similarità/distanza per cognomi possono essere sottoposte agli stessi metodi (costruzione di alberi filettici, ottenibili con metodi diversi; rappresentazioni topologiche ottenute mediante l'analisi per componenti principali o il nm-MDS) utilizzati per rappresentare sinteticamente le relazioni tra oggetti, ai quali fa normalmente ricorso la genetica di popolazione per inferire sulle relazioni filettiche tra popolazioni. Applicazioni alla popolazione italiana hanno evidenziato la correlazione con la storia passata delle popolazioni regionali italiane (Zei *et al.* 1993), l'influenza di immigrazioni storiche, gli effetti delle separazioni storico-amministrative (Barbujani *et al.* 1991; 1992). Le analisi condotte a partire dai cognomi trovano anche un buon accordo con le mappe ottenute a partire dalle frequenze alleliche (Piazza *et al.* 1987).

4.3.2 Similarità tra le popolazioni regionali italiane; effetti dei flussi migratori Sud-Nord. Una prima esemplificazione deriva da analisi sulla distribuzione (ottenuta dalla SEAT, la società che gestisce gli elenchi degli abbonati al telefono, dopo cancellazione di tutti i nominativi non corrispondenti a persone fisiche) dei cognomi degli abbonati al telefono nell'anno 1993, classificati per regione (Lucchetti *et al.* 1989). La diffusione (1993) del telefono e la numerosità degli abbonati, in media un telefono per famiglia, è garanzia di una buona rappresentatività dei residenti totali. Dalle matrici di similarità tra regioni ricavata dalle distribuzioni dei cognomi, sono state ottenute le rappresentazioni topologiche sottoillustrate.

La disposizione relativa delle regioni – con l'avvertenza di porre la prima dimensione sull'asse delle ordinate, ad individuare la direzione Sud-Nord, e la seconda dimensione sull'asse delle ascisse ad individuare la direzione Ovest-Est – rispecchia a grandi linee la loro posizione geografica (fig. 8) e evidenzia al contempo le peculiarità di qualche regione. Si nota la separazione della regione Sardegna, collocata in posizione isolata dalle altre popolazioni regionali italiane, come peraltro

Fig. 9. *Mappa delle regioni italiane e delle cinque maggiori città (Milano, Torino, Genova, Roma, Napoli) derivata dalla distribuzione dei cognomi al 1993*

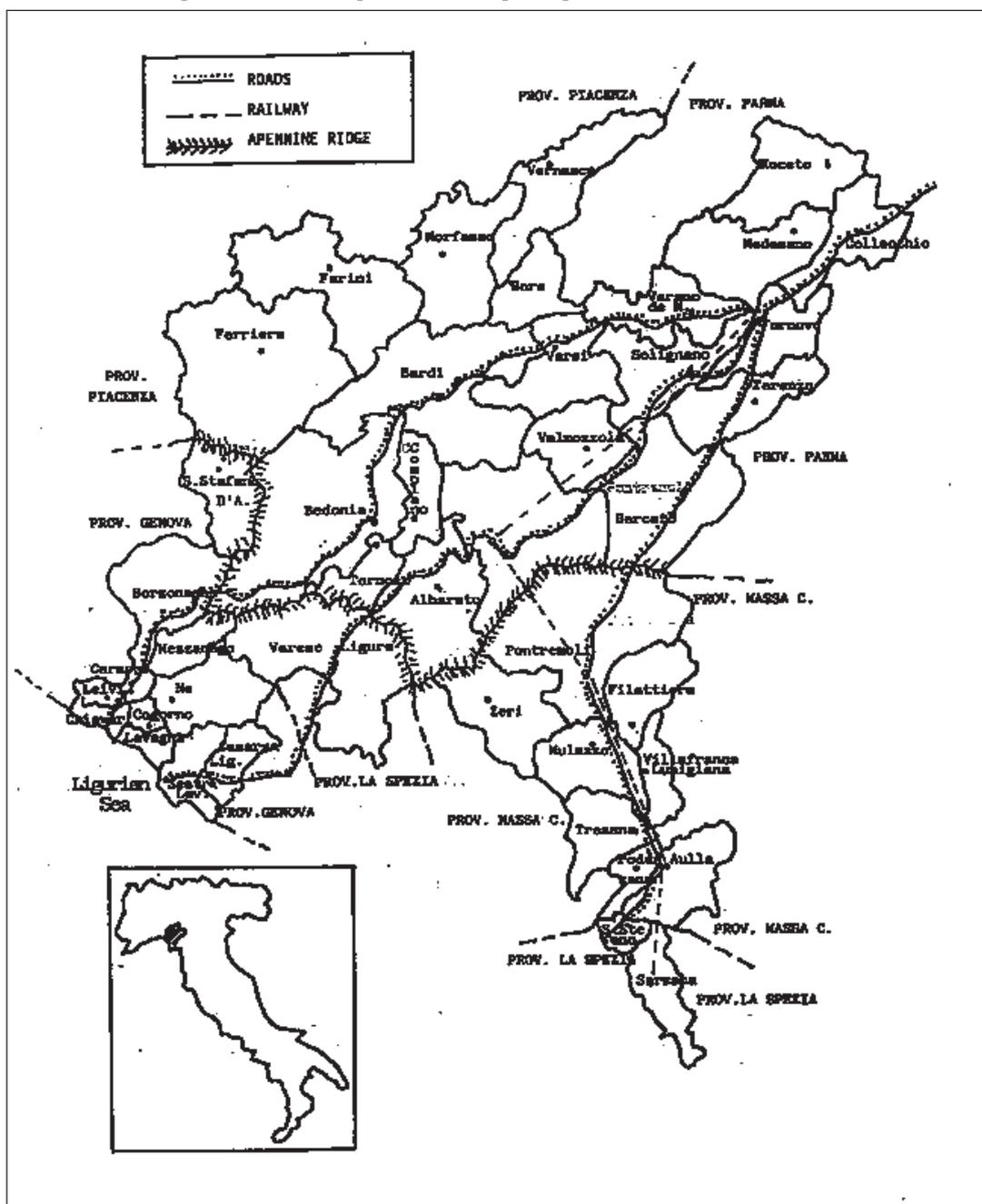


zione delle quali rimane prossima a quella della rimanente parte della Liguria e del Lazio. Per Genova e la Liguria la causa va ricercata in una dispersione maggiormente omogenea degli immigrati su tutto il territorio regionale, mentre per Roma nella presenza di immigrati provenienti da tutte le regioni italiane.

4.3.3 Struttura per cognomi, mobilità e caratteristiche geografiche del territorio. Le caratteristiche geomorfologiche del territorio (esistenza di crinali montuosi, di fiumi, di laghi), la presenza di più facili e accessibili vie di comunicazione e trasporto, di suddivisioni amministrative attuali o passate, in che misura possono aver influenzato la mobilità delle persone tra le zone considerate? La affinità tra strutture cognominali, evidenziando il grado di somiglianza o di mescolamento tra le popolazioni interessate, possono dare indicazioni sulla presenza e sugli effetti di questi fattori.

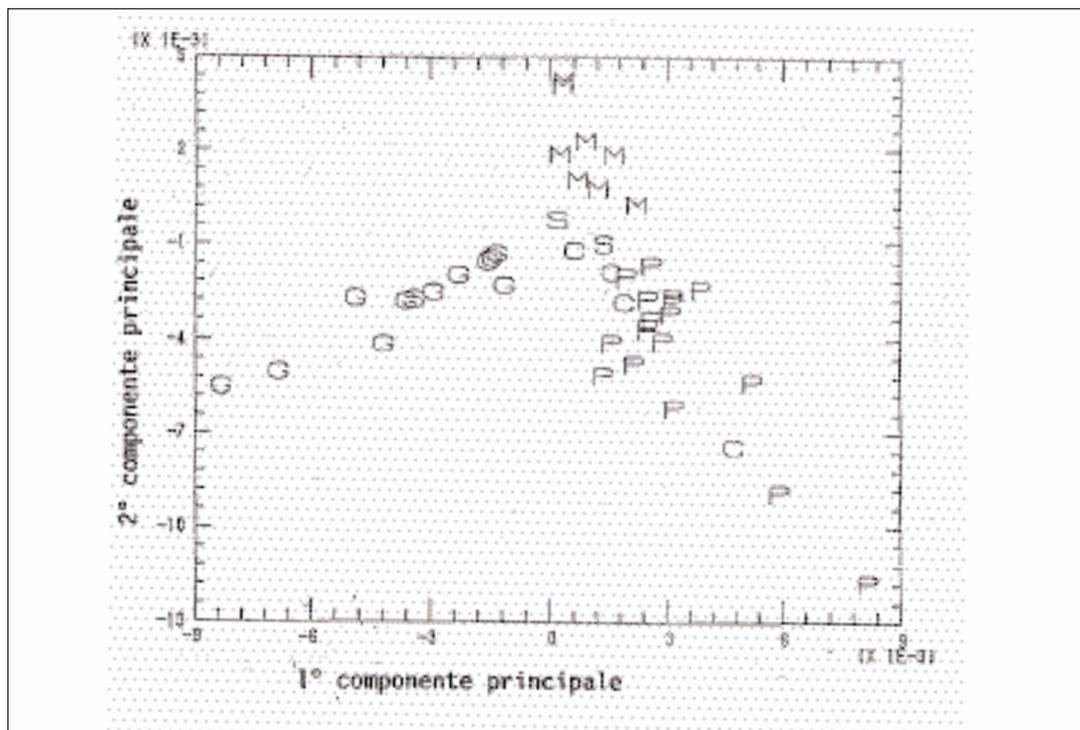
Allo scopo sono state considerate le strutture in cognomi di alcune popolazioni comunali poste a cavallo del crinale appenninico, appartenenti alle province di Parma e Piacenza (a nord del crinale appenninico) e di Massa Carrara, La Spezia e Genova (a sud). Le popolazioni afferiscono ad amministrazioni regionali e provinciali diverse, sono state caratterizzate nel passato dall'appartenenza a stati diversi, sono tra loro separate dalla presenza del crinale; tra alcune la comunicazione è favorita dalla presenza di antiche vie di comunicazione e di una linea ferroviaria (fig. 10).

Fig. 10. *Distribuzione geografica dei comuni considerati, con indicazione del crinale appenninico, dei confini provinciali e dei percorsi delle principali vie di comunicazione*



Il confronto tra le strutture per cognomi porta ad una rappresentazione topologica (fig. 11) nella quale i comuni attraversati dalle antica strada statale e dalla ferrovia, indipendentemente dalla provincia di appartenenza e dalla loro collocazione rispetto al crinale, assumono valori simili sulla prima dimensione collocandosi verso destra; quelli invece esterni a queste vie di comunicazioni si collocano a sinistra. Sulla seconda componente si separano verso il basso i comuni posti a nord del crinale appenninico, verso l'alto quelli posti a sud del crinale, in modo più marcato quelli della provin-

Fig. 11. Rappresentazione bidimensionale delle relazioni di affinità tra i comuni. I comuni sono indicati per provincia di appartenenza (P=Parma, C=Piacenza, S=La Spezia, G=Genova, M=Massa Carrara)



cia di Massa Carrara e i due comuni limitrofi della provincia di La Spezia, in modo meno marcato quelli della provincia di Genova e il terzo comune spezzino distante dalla strada statale e dalla ferrovia e più prossimo al genovese. L'interpretazione congiunta delle due dimensioni evidenzia l'importanza sulla mobilità delle vie di comunicazione, l'effetto contrastante del crinale appenninico, l'effetto dell'appartenenza a amministrazioni diverse o comunque, fino alla metà dell'Ottocento, a stati differenti.

La figura 12 mostra le relazioni tra le popolazioni comunali considerate utilizzando tre dimensioni. La terza sottolinea gli effetti dell'isolamento: vi è complessivamente una buona corrispondenza tra l'altitudine dei comuni e il valore assunto dalla terza dimensione, venendo così a delineare il profilo dei monti. È interessante notare che i comuni posti sulle vie di comunicazione assumono sulla terza dimensione valori inferiori, definendo così un avvallamento in corrispondenza del tracciato delle vie di comunicazione, alla presenza delle quali si determina un evidente calo dell'isolamento.

4.3.4 Modificazioni nel tempo della struttura per cognomi di popolazioni umane. Le serie storiche dei cognomi permettono di estendere le analisi alle generazioni passate e di valutare in modo rapido e sintetico l'evoluzione della struttura della popolazione. Sono state analizzate – ad esempio – le strutture per cognome di 29 parrocchie dell'alta Val Parma dal 1810 al 1959 (Soliani *et al.* 1986). Il periodo complessivo, 150 anni, è stato suddiviso in 15 periodi di 10 anni ciascuno. La matrice di similarità confronta quindi le strutture presentate dalla stessa popolazione in periodi successivi; la

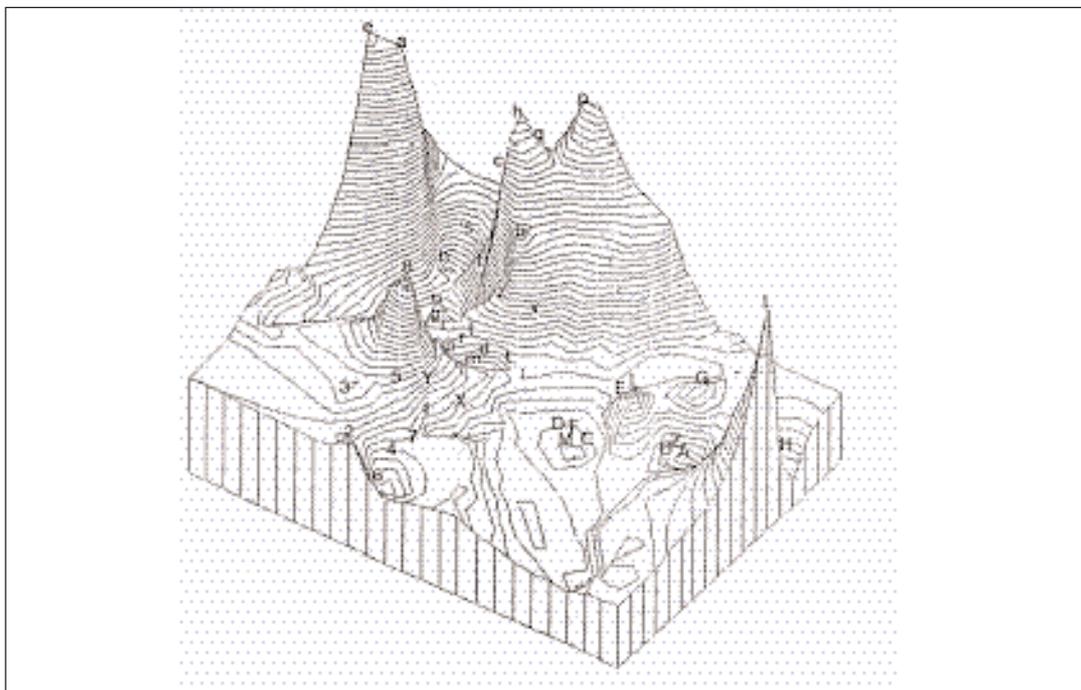
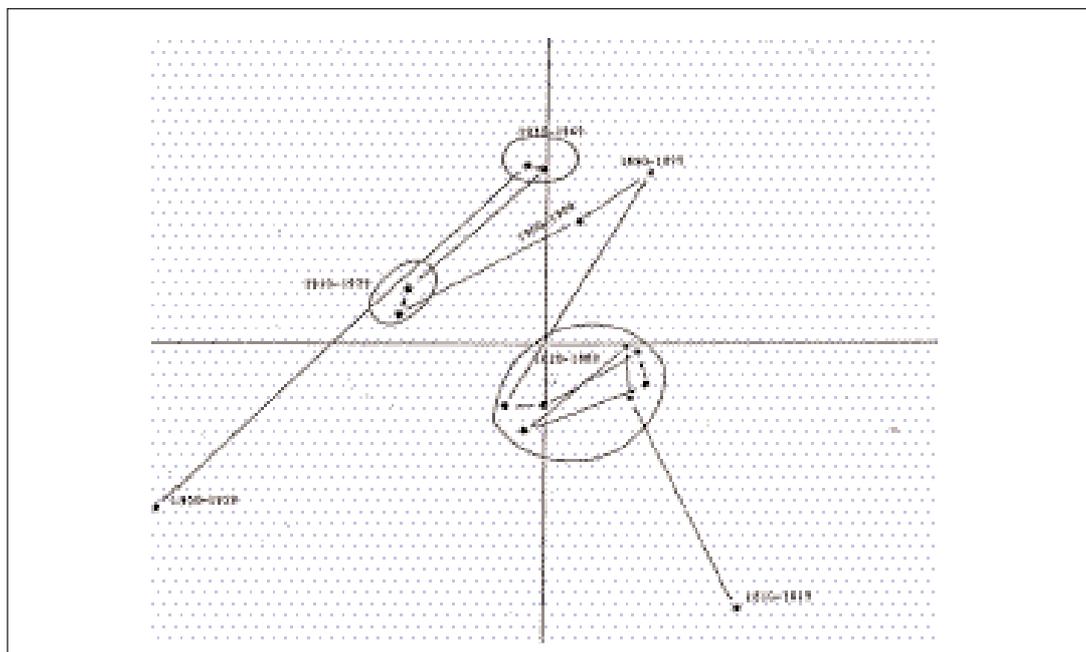
Fig. 12. *Rappresentazione tridimensionale delle relazioni tra le popolazioni comunali*

figura 13 mostra la rappresentazione topologica ottenuta. I punti corrispondenti ai vari periodi evidenziano momenti di relativa stabilità (periodi successivi collocati l'uno in prossimità dell'altro), alternati a momenti durante i quali sono avvenute rapide variazioni (periodi successivi tra loro nettamente distanziati); le variazioni sono in genere correlate con vicende storico-demografiche della popolazione.

La sensibile distanza tra i punti che individuano i periodi 1810-19 e 1820-29 può essere imputata alla crisi di mortalità del 1817, conseguenza di una grave epidemia di tifo petecchiale che, dopo anni di ricorrenti carestie, eliminò circa il 20% della popolazione, provocando negli anni successivi consistenti variazioni nella struttura e nei comportamenti demografici. Dal 1820 al 1889 la struttura in cognomi della popolazione rimane sostanzialmente stabile: le variazioni sono ridotte e casuali, i punti che identificano questi periodi risultano tutti abbastanza prossimi tra loro. Alla fine del XIX secolo e nel primo decennio del XX si manifestano intensi movimenti migratori, interrotti dallo scoppio della prima guerra mondiale; essi riprenderanno dopo il 1920, per ridursi di nuovo a causa delle leggi del 1930 sulle migrazioni e dello scoppio della seconda guerra mondiale. La rilevante distanza del punto che individua il periodo 1950-59 evidenzia le conseguenze dei movimenti migratori più recenti, della cosiddetta «rottura degli isolati» e dello spopolamento delle zone appenniniche. L'esame della distribuzione, alle varie date, della frequenza dei singoli cognomi mostra quali gruppi familiari o discendenze sono scomparsi e quali sono percentualmente aumentati nei 150 anni, determinando la evoluzione descritta.

5. I nuovi sviluppi per l'individuazione dei flussi e le matrici di migrazione. Sviluppi più recenti si propongono scopi ancora più interessanti per la ricerca demografica

Fig. 13. Rappresentazione bidimensionale delle affinità tra periodi successivi, ricavata dal confronto delle strutture in cognomi di ciascun periodo



e sociale: ricostruire i flussi tra popolazioni e seguirne l'evoluzione (Poulain *et al.* 1998; Wijsman 1984), quando per ognuna si dispone delle distribuzioni dei cognomi in almeno due periodi diversi.

Lo scopo viene perseguito con modalità diverse secondo due metodologie: la prima, dopo aver identificato alcuni cognomi tipici nella zona di partenza, individua i flussi che legano le zone d'emigrazione a quelle d'immigrazione e con rilevazioni successive segue la dispersione territoriale degli immigrati; la seconda, considerando congiuntamente la struttura in cognomi di più popolazioni, determina l'intensità e la direzione degli scambi migratori avvenuti tra esse nel periodo considerato. Rispetto all'analisi demografica, che spesso ha limiti evidenti nello studio della mobilità, entrambi i metodi hanno l'ulteriore vantaggio di usare dati raccolti con finalità totalmente differenti.

Le analisi sulla direzione dei flussi migratori attraverso i cognomi sono fondate su un'ipotesi semplice: l'etimologia di un cognome è indicativa dell'area linguistico-culturale dalla quale deriva. Scegliendo i cognomi adeguati, si possono evidenziare i movimenti migratori internazionali e seguire la distribuzione degli immigrati sul territorio. Nelle Fiandre, zona depressa che per decenni ha alimentato flussi migratori consistenti verso la Francia e il Belgio, sono stati identificati 143 cognomi con il prefisso tipicamente fiammingo *Van* oppure *Ver* hanno permesso di identificare 143 cognomi, attualmente portati da 650.000 persone nelle Fiandre (11% della popolazione totale); questi cognomi (Poulain *et al.* 1998) hanno evidenziato la localizzazione dell'immigrazione fiamminga in Belgio e in Francia, e permesso di quantificarne l'incidenza nelle singole zone e infine la loro dispersione successiva.

Una ricerca in collaborazione tra genetisti italiani e francesi (Degioanni *et al.*

1996) ha utilizzato i cognomi dei nati nei 36.500 comuni di Francia raccolti da l'INSEE, successivamente raggruppati per dipartimento, e per l'Italia quelli delle donne che hanno avuto un matrimonio consanguineo, provenienti dalle dispense matrimoniali del Vaticano, suddivisi per diocesi. Studiando queste distribuzioni ed i loro cambiamenti nel periodo 1891-1940, si voleva indagare sull'origine delle presenze italiane in Francia e proporre una stima dei flussi migratori dall'Italia verso la Francia. Nonostante le difficoltà tecniche derivanti dal tipo di fonti disponibili per l'Italia, gli autori affermano che il metodo permette sia di quantificare i flussi migratori, benché in modo approssimato, sia di chiarire aspetti sull'origine geografica dei migranti, che generalmente i dati censuari non forniscono. Piemontesi e Liguri si sono installati per primi nel Sud della Francia; a partire dall'inizio del XX secolo, essi hanno lasciato la supremazia ai Veneti nella vallata della Garonne e lungo l'asse Nantes-Rennes, ai Lombardi e ai Veneti in Lorraine, ai Lombardi in Franche-Comté, mentre Parigi accoglie tutta la diversità italiana (con una componente prevalente dell'Emilia-Romagna); la regione di Marsiglia accoglie soprattutto italiani del Centro-Sud, provenienti da Lazio, Sicilia, Sardegna e Toscana.

Il metodo proposto da Wijsman e altri (1984) si propone, sulla base delle trasformazioni nella distribuzione di frequenza dei cognomi, di ricavare stime sul comportamento migratorio di più popolazioni.

La struttura in cognomi di più popolazioni può essere descritta mediante una matrice \mathbf{N} , di dimensioni $P \times S$, ove P è il numero di popolazioni considerate e S è il numero di cognomi differenti (quindi molte migliaia nel caso di province o regioni). Ogni riga corrisponde a una popolazione e ogni colonna ad un cognome; l'elemento generico n_{pk} indica il numero di individui della p -ma popolazione che portano il k -mo cognome.

La struttura in cognomi delle P popolazioni rilevata in 2 tempi diversi, t_1 e t_2 , può essere descritta con due matrici \mathbf{N}_1 e \mathbf{N}_2 . Se le P popolazioni rappresentano un sistema relativamente isolato, \mathbf{N}_2 può essere espressa in funzione di \mathbf{N}_1 secondo la relazione

$$\mathbf{N}_2 = \mathbf{B} * \mathbf{N}_1 + \mathbf{E}$$

ove la matrice \mathbf{E} raccoglie gli effetti dovuti al caso o comunque a cause diverse dalla migrazione e fecondità.

La matrice \mathbf{B} , quadrata, di dimensioni $P \times P$, raccoglie gli effetti congiunti dei comportamenti migratori e riproduttivi; può essere scomposta nelle due matrici \mathbf{R} e \mathbf{M} , tali che

$$\mathbf{B} = \mathbf{R} * \mathbf{M}.$$

\mathbf{R} è una matrice diagonale: l'elemento generico r_{jj} descrive il comportamento riproduttivo della j -ma popolazione; \mathbf{M} è una matrice quadrata, anch'essa di dimensioni $P \times P$, nelle quali l'elemento generico m_{jk} indica la quota di residenti nella popolazione j provenienti dalla popolazione k ; gli elementi della sua diagonale principale stimano la quota di residenti nativi.

Trascurando \mathbf{E} , considerando cioè solo gli effetti di migrazione e riproduzione, si può scrivere

$$\mathbf{N}_2 = \mathbf{R} * \mathbf{M} * \mathbf{N}_1$$

ove il termine generico n_{pk} di N_2 , numero di individui che al tempo $t=2$ appartengono alla popolazione p e portano il cognome k , è espresso dalla sommatoria degli individui con cognome k -mo al tempo t_1 , migrati da ogni popolazione nella popolazione p e in essa riprodottisi con tasso r_{pp} .

Wijsman (1984) suggerisce il modo di ricavare la matrice \mathbf{B} attraverso le matrici di isonimia tra popolazioni e di ricavare da essa la matrice \mathbf{M} , che descrive le probabilità di migrazioni tra le popolazioni considerate. L'applicazione del metodo a dati storici della Sardegna ha evidenziato un buon accordo con quanto noto sui movimenti migratori della zona.

¹ È possibile trovarne un'esposizione sintetica, finalizzata in modo specifico all'uso delle tecniche e dei concetti della genetica di popolazioni da parte del demografo, nella prima stesura del *Trattato di Demografia*, al paragrafo *Genetica e Demografia* (di Lucchetti e Soliani in *Actes du Séminaire de San Miniato*, Pise 17-

19 décembre 1997. Dipartimento di Scienze Demografiche, Università degli Studi «La Sapienza» di Roma; Institut national d'études démographiques, Paris; Département des sciences de la population et du développement, Université catholique de Louvain, 1, 159-182).

Bibliografia

- G. Barbujani, R.R. Sokal 1991, *Genetic population structure of Italy. II. Physical and cultural barriers to gene flow*, «American Journal of Human Genetics», 48, 398-411.
- G. Barbujani, P. Vian, L. Fabbri 1992, *Cultural barriers associated with large gene frequency differences among Italian populations* «Human Biology», 64, 479-495.
- I. Barraï, G. Barbujani, M. Beretta, I. Maestri, A. Russo, G. Formica, I. Pinto-Cisternas 1987, *Surnames in Ferrara: distribution, isonymy and levels of inbreeding*, «Annals of Human Biology», 14, 415-423.
- I. Barraï, G. Formica, R. Barale, M. Beretta 1989, *Isonymy and migration distance*, «Annals of Human Genetics» 53, 249-262.
- I. Barraï, G. Formica, R. Barale, C. Scapoli, R. Canalla, M. Beretta 1990, *Isonymy in emigrants from Ferrara in 1981-1988*, «Annals of Human Biology», 17, 1, 7-18.
- G. Biondi, G.W. Lasker, P. Raspe, C.G.N. Mascie-Taylor 1993, *Inbreeding coefficients from the surnames of grandparents of the schoolchildren in Albanian-speaking Italian villages*, «Journal Biosocial Science», 25, 63-71.
- G. Biondi, P. Raspe, C.G.N. Mascie-Taylor, G.W. Lasker 1996, *Ripetition of the same pair of surnames in marriages in Albanian Italians, Greek Italians and the Italian population of Campobasso province*, «Human Biology» 68, 4, 573-583.
- R. Chakraborty 1985, *A note on the calculation of random RP and its sampling variance*, «Human Biology», 57, 4, 713-717.
- K. Chen, L.L. Cavalli-Sforza 1983, *Surnames in Taiwan: Interpretations based on geography and history*, «Human Biology» 55, 2, 367-374.
- J.F. Crow 1980, *The estimation of inbreeding from isonymy*, «Human Biology» 52, 1, 1-14.
- J.F. Crow 1989, *Update to "the estimation of inbreeding from isonymy"*, «Human Biology» 61, 5/6, 949-954.
- J.F. Crow, A.P. Mange 1965, *Measurement of inbreeding from the frequency of marriages between persons of the same surname*. «Eugenic Quart» 12, 199-203.
- G. Darwin 1875, *Marriages between first cousins in England and their effects*, «Journal of the Statistical Society», 38, 153-184.
- P. Darlu, J. Ruffie 1992, *Relationships between consanguinity and migration rate from surname distributions and isonymy in France*, «An-

- nals of Human Biology», 19, 2, 133-137.
- A. Degioanni, A. Lisa, G. Zei, P. Darlu 1996, *Patronymes italiens et migration italienne en France entre 1891 et 1940*, «Population», 6, 1153-1180.
- S. De Iasio 2000, *Analisi dei cognomi ricavati da diverse serie storiche*. Atti Convegno Irgoli: Biodemografia e Genetica. Irgoli, 11-12 marzo.
- S. De Iasio, E. Farnetti, M. Manfredini, E. Lucchetti 1993, *Struttura e dinamica di una popolazione analizzata tramite serie storiche di cognomi Madregolo (PR), XVII-XX secolo*, «Antropologia Contemporanea» 16, 1/4, 87-93.
- R.A. Fisher 1943, *The relation between the number of species and the number of individuals in a random sample of an animal population. Part. 3*, «Journal Animal Ecology», 12, 42-58.
- M.G. Franceschi, G. Paoli 1994, *Isolation factors and kinship by isonymy in a group of parishes in Northern Tuscany Italy: influence of within-parish similarity level on between-parish similarity pattern*, «Human Biology» 66, 5, 905-916.
- P. Gueresi, D. Pettener, F. Martuzzi-Veronesi 1996, *Il metodo delle coppie ripetute nell'analisi del grado di suddivisione di popolazioni*, «Bollettino di Demografia Storica», 24/25, 85-96.
- C.R. Guglielmino, A. De Silvestri 1995, *Surname sampling for the study of the genetic structure of an Italian province*, «Human Biology» 67, 4, 613-628.
- S. Karlin, J. McGregor 1967, *The number of mutant forms maintained in a population*. Proc. Fifth Berkeley Symp. Math. Stat. Prob. 4, 415-438.
- C.J. Krebs 1989, *Ecological Methodology*, Harper & Row, New York.
- G.W. Lasker 1977, *A coefficient of relationship by isonymy: a method for estimating the genetic relationship between populations*, «Human Biology» 49, 3, 489-493.
- G.W. Lasker, B.A. Kaplan 1985, *Surnames and genetic structure: repetition of the same pairs of names of married couples, a measure of subdivision of the population*, «Human Biology» 57, 3, 431-440.
- E. Lucchetti, C.M. Battisti, G. Ghisolfi, L. Soliani 1989, *Delimitazione e aggregazione tra popolazioni analizzate mediante la struttura per cognomi*, «Antropologia Contemporanea» 12, 3, 1-16.
- E. Lucchetti, L. Soliani 1989, *Similarità tra popolazioni esaminate mediante i cognomi*, «Rivista di Antropologia», 67, 181-198.
- E. Lucchetti 1990, *Uso dei cognomi in analisi sulla struttura ed evoluzione di popolazioni umane*, «Antropologia Contemporanea» 13, 2/3, 107-119.
- E. Lucchetti, F. Conterio, L. Soliani, I. Barraì 1990, *Assessment of isonymy and migration in Sicily by log-log transformation*. Atti Associazione Genetica Italiana. Perugia, 22-26 ottobre.
- C.G.N. Mascie-Taylor, G.W. Lasker, A.J. Boyce 1987, *Repetition of the same surnames in different marriages as an indication of the structure of the population of Sanday, Orkney Islands*, «Human Biology», 58, 97-102.
- G. Paoli, M.G. Franceschi 1993, *L'uso dei cognomi negli studi sulla struttura genetica delle popolazioni: verifica dell'assunzione di neutralità selettiva*, «Antropologia Contemporanea», 16, 1/4, 247-251.
- D. Pettener 1990, *Temporal Trends in marital structure and isonymy in S.Paolo Albanese, Italy*, «Human Biology», 62, 6, 837-851.
- A. Piazza, S. Rendine, G. Zei, A. Moroni, L.L. Cavalli-Sforza 1987, *Migration rates of human populations from surname distributions*. «Nature», 329, 714-716.
- J. Pinto-Cisternas, H. Arvelo, R. Martinez, D. Castro De Guerra 1997, *Coefficient of relationship RI within and between four black Venezuelan populations*, «International Journal of Anthropology» 12, 2, 55-62.
- P. Pizzetti, S. De Iasio, E. Lucchetti 2000, *Surnames as means for biodemographic analysis*, «Rivista di Antropologia», 78, 243-252.
- M. Poulain, M. Foulon, A. Degioanni, P. Darlu 1998, *L'immigration flamande en Wallonie et en France: l'apport des patronymes*. Contribution aux entretiens Jacques Cartier, Lyon, Dicembre 1998.
- J. Relethford 1988, *Estimation of kinship and genetic distance from surnames*, «Human Biology», 60, 3, 475-492.
- J. Relethford 1992, *Analysis of marital structure in Massachusetts using repeating pairs of surnames*, «Human Biology», 64, 25-33.
- A. Rodriguez-Larralde, A. Pavesi, C. Scapoli, F. Conterio, E. Siri, I. Barraì 1994, *Isonymy and the genetic structure of Sicily*, «Journal of Biosocial Science», 26, 9-24.
- C. Scapoli, A. Rodriguez-Larralde, M. Beretta, C. Nesti, E. Lucchetti, I. Barraì 1997, *Corre-*

- lations between isonymy parameters*, «International Journal of Anthropology», 12, 1, 17-37.
- L. Soliani, A. Anelli, W. Anghinetti, E. Siri 1988, *Analisi con cognomi e nomi delle variazioni demografiche, biologiche, culturali di comunità e popolazioni*, «Genus», 92, 3/4, 55-88.
- A. Tagarelli, C. Brancati 1995, *Analisi della struttura dei cognomi nella minoranza etnica albanese delle provincia di Cosenza*, «Antropologia Contemporanea», 18, 1/2, 35-42.
- J. Wang, A. Caballero 1998, *Developments in predicting the effective size of subdivided populations*, «Hereditas», 82, 212-226.
- M.C. Whitlock, D.E. McCauley 1998, *Indirect measures of gene flow and migration*, «Hereditas», 82, 117-125.
- E. Wijsman, G. Zei, A. Moroni, L.L. Cavalli-Sforza 1984, *Surnames in Sardinia II: Computation of migration matrices from surname distribution in different periods*, «Annals of Human Genetics», 48, 65-78.
- G. Zei, R. Guglielmino Matessi, E. Siri, A. Moroni, L.L. Cavalli-Sforza 1983, *Surnames in Sardinia I*, «Annals of Human Genetics», 47, 329-352.
- G. Zei, G. Barbujani, A. Lisa, O. Fiorani, P. Menozzi, E. Siri, L.L. Cavalli-Sforza 1993, *Barriers to gene flow estimated by surnames distribution in Italy*, «Annals of Human Genetics», 57, 123-140.

Riassunto

L'uso dei cognomi nella ricerca biodemografica

L'origine dei cognomi, le loro tipologie, dipendono dal contesto e dalle caratteristiche culturali della popolazione nella quale si sono formati. Dal momento della loro fissazione e con la comparazione di norme specifiche che ne regolano la trasmissione essi sono divenuti assimilabili a caratteri genetici e possono essere indagati come tali. I cognomi possono così contribuire a definire l'identità di una popolazione che è al tempo stesso culturale e biologica: la loro tipologia, il loro numero, la distribuzione per frequenza si rivelano dati assai utili – facili da rilevare e ricchi di informazione – per lo studio delle popolazioni umane.

La conoscenza infatti della struttura per cognomi permette di sviluppare analisi sulla struttura interna della popolazione, sulla variabilità in essa presente, sul grado di isolamento, sulle modalità di formazione delle coppie, sul comportamento matrimoniale, sui flussi migratori sia in entrata che in uscita. La estendibilità della rilevazione dei cognomi a generazioni del passato consente di inferire sui meccanismi e sulle modalità di evoluzione di una popolazione e di osservare in modo concreto le trasformazioni da essa subite nel tempo.

Sulla base della probabilità che individui presi a caso tra più popolazioni portino lo stesso cognome è possibile valutare il grado di diversificazione o di similarità tra esse esistente. Possono così essere tratte inferenze sulle relazioni che tra esse intercorrono o che si sono verificate nel passato: se le popolazioni derivano da una comune matrice culturale, come si sono differenziate nel tempo, se tra esse vi è stato, e in quale misura, interscambio di individui.

Summary

Biodemographic research using surname data

The origin of surnames, and their different typologies, depends on the context and cultural features of the population in which they formed. Since their fixation and the establishment of specific norms regulating their transmission, they are assimilable to allelic forms of the same gene and analyzable accordingly. Surnames can therefore be useful in defining the "identity" of a given population, which is cultural and biological at the same time. Their typology, number, frequency distribution can be consequently used to describe and study human populations.

The knowledge of the surname structure allow the researcher to carry out analyses on the internal

structure of the population, its variability, isolation degree, marriage pattern, estimate of migrations, etc. The possibility to study also historical populations, starting from the period of fixation of surnames, allow to make hypotheses on the forms and mechanisms of evolution and to describe really the structural changes of populations over time.

The probability to find two people in different populations carrying the same surname makes possible to evaluate the distance between two (or more) populations. Hypotheses can then be drawn on their interrelationships in the past, how, for example, they differentiated each other over time, and what amount of migratory flows occurred among them.