

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

## Coevoluzione di lingue e geni: il caso dell'Indo-Europeo e degli Indo-Europei

**This is a pre print version of the following article:**

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/1524630> since 2015-09-10T13:08:22Z

*Publisher:*

Scienze e lettere

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

VISTO: SI STAMPI

(Firma Autore).....

ALBERTO PIAZZA \*

COEVOLUZIONE DI LINGUE E GENI:  
IL CASO DELL'INDO-EUROPEO E DEGLI INDO-EUROPEI

ABSTRACT. – *TITOLO INGLESE*. In several studies by Cavalli-Sforza and myself, the spread of anatomically modern man was reconstructed on the basis of genetic and linguistic pieces of evidence: the main conclusion was that these two approaches reflect a common underlying history, the history of our past still frozen in the genes of modern populations. I introduced the expression 'genetic history' to point out that if today we find many genes showing the same geographical patterns in terms of their frequencies, this may be due to the common history of our species. Specific cases of structural analogies between linguistic and genetic geographical patterns have been explored that supply further and more updated information. It is important, however, to emphasize at the outset that evidence for coevolution of genes and languages in human populations does not suggest by itself that some genes of our species determine the way we speak; this coevolution may simply be due to a common mode of transmission and mutation of genetic and linguistic units of information and common constraints of demographic factors. While our oral contribution at the Meeting has been devoted to the coevolution of genes and languages at the global level of our species, the following written contribution will deal with the more specific case of Indo-European languages and people. As a matter of fact, the first linguistic family to be recognised was Indo-European. Its recognition, and the fact that the scholars of linguistics in the western world spoke Indo-European tongues, eventually resulted in its being very extensively investigated. The Indo-European family also has the distinction of exemplifying the first model of tree of descent, shortly after Darwin introduced tree models for explaining biological evolution as multiplication with modification. August Schleicher (1863) published the first linguistic tree of the Indo-European family as an example of his *Stammbaumtheorie* (tree theory). A few years later, his student Johannes Schmidt (1872) generated the alternative *Wellentheorie* (wave theory). Tree models emphasise the independent evolution of languages originating from a common source, but spoken in separate areas and not communicating with each other in more than a limited way. Wave models emphasise the exchange of single words and

\* Dipartimento di Scienze Mediche – Università degli Studi di Torino – Via Verdi, 8 – 10124 TORINO. alberto.piazza@unito.it

other features between neighbouring groups, either among the local dialects of a single language or among different languages. Neither model can give a complete representation of the historical and geographic change of languages. In this contribution a tree model analysis of a set of linguistic lexical data will be presented and discussed in the context of already published genetic data. In spite of the fact that genetic data are not able to provide a robust tree of descent for the European populations, we have been able to show on a quantitative basis that European genes still identify a possible candidate area in Ukraine from which Indo-European (and Indo-Europeans) originated and diffused.

RIASSUNTO. – La nostra storia passata è fatta di insediamenti e quando le persone si muovono, non solo muovono i loro geni, ma anche le loro idee. Poiché geni ed idee sono condizionati da fattori spesso identici, è naturale trovare una correlazione tra geni e cultura. Che biologia e linguistica avessero dei legami stretti, era già stato intuito da Darwin che nel quattordicesimo capitolo dell' *Origine della specie* scrisse: «*If we possessed a perfect pedigree of the mankind, a genealogical arrangement of the races of man would afford the best classification of the various languages now spoken throughout the world*». La capacità discriminante dei dati prodotti dalla genetica molecolare moderna dà corpo alla profezia di Darwin e permette di usare l'albero filogenetico della nostra specie per inferire le prime suddivisioni di un ipotetico albero linguistico. Lingue e geni contribuiscono a comprendere la storia dell'uomo che è un processo che non si può riprodurre e ripetere in laboratorio e ad apprezzare la sua diversità: strumenti - le une e gli altri - che danno ad alcune voci silenziose del passato una possibilità di parlare ancora. Per offrire al lettore un contributo scritto più puntuale ci si limiterà ad esporre alcuni risultati di un'analisi inedita su un sottoinsieme del quadro complessivo, quello che riguarda lingue e geni in Europa. L'analisi congiunta dei dati lessicali di 84 lingue europee e dei marcatori genetici saggiati in aree geografiche corrispondenti porta a confermare su base quantitativa l'ipotesi che i geni europei identificano in Ucraina una possibile area geografica da cui hanno avuto origine e si sono diffusi l'Indo-Europeo (e gli Indo-Europei).

## INTRODUZIONE

La nostra storia passata è fatta di insediamenti e quando le persone si muovono, non solo muovono i loro geni, ma anche le loro idee. Poiché geni ed idee sono condizionati da fattori spesso identici, è naturale trovare una correlazione tra geni e cultura. Ma vi sono spiegazioni dell' analogia anche più sottili. Se analizzassimo più nei particolari i meccanismi evolutivi biologici e culturali della nostra specie, scopriremmo molte corrispondenze. Considerando il linguaggio come fenomeno culturale per eccellenza, dei quattro meccanismi evolutivi che controllano il nostro cambiamento genetico: la *migrazione*

influenza geni e linguaggi nello stesso modo inducendo fenomeni di diffusione; la *selezione* agisce in entrambi i campi: la selezione naturale favorendo il tipo biologico più adatto a sopravvivere e la selezione culturale favorendo la mutua intelligibilità lessicale e fonetica del linguaggio; *mutazione* ed innovazione linguistica svolgono la stessa funzione: sono cambiamenti che, una volta verificatisi casualmente in singoli individui, vengono poi adottati da altri individui. Quello che tuttavia rende profondamente diverso il gene dalla cultura è il meccanismo di trasmissione. Mentre in biologia la trasmissione dell'informazione può avvenire solo da genitore a figlio, la trasmissione culturale può avvalersi di altri meccanismi. Cavalli-Sforza e Feldman (1981) ne hanno esemplificati quattro:

1. *Da genitore a figlio (verticale)*. È il meccanismo di diffusione dell'informazione più lento ma è anche quello che conserva la variabilità da individuo ad individuo.
2. *Da un individuo all'altro (orizzontale)*. È un meccanismo analogo a quello del contagio in una epidemia, per cui l'informazione si diffonde rapidamente.
3. *Da uno a più individui (per esempio da un insegnante ai suoi allievi o da un leader sociale ai suoi seguaci)*. È il meccanismo più efficace per diffondere un'innovazione in un gruppo sociale.
4. *Da più individui ad un singolo individuo*. È il meccanismo della pressione sociale: comunemente sfavorisce il diffondersi di una innovazione.

Qual è il meccanismo che ha più peso? Probabilmente dipende dal tipo di cultura e dal momento storico. Se il meccanismo 1 ha avuto nel passato un ruolo preminente e fornisce una spiegazione alle associazioni tra geni e lingue che ancora oggi siamo in grado di ricostruire, è chiaro che oggi l'avvento delle comunicazioni di massa favorisce e favorirà sempre più un meccanismo di tipo 3. Alla famiglia che ha garantito fino ad oggi non solo la trasmissione dei geni ma anche quella della cultura si sostituisce la figura del *leader*, la trasmissione dell'informazione dal singolo a più individui. Ora il meccanismo culturale *da uno a più individui* rappresenta l'analogo culturale più estremo di deriva genetica prima menzionato: è come se un solo individuo trasmettesse i suoi geni a tutta una popolazione, sicché scomparirebbe la variabilità genetica. In altre parole: la deriva culturale potrebbe rendere assai più veloce la diffusione dell'informazione, ma ad un prezzo rischiosissimo, quello che l'informazione diventi una sola.

Che biologia e linguistica avessero dei legami stretti, era già stato intuito da Darwin, che nel quattordicesimo capitolo dell' *Origine della specie* scrisse:

*If we possessed a perfect pedigree of the mankind, a genealogical arrangement of the races of man would afford the best classification of the various languages now spoken throughout the world; and if all extinct languages, and all intermediate and slowly changing dialects, were to be included, such an arrangement would be the only possible one. Yet it might be that some ancient language had altered very little and had given rise to few new languages, whilst others had altered much owing to the spreading, isolation, and state of civilization of the several co-descended races, and had thus given rise to many new dialects and languages. The various degrees of difference between the languages of the same stock, would have to be expressed by groups subordinate to groups; but the proper or even the only possible arrangement would still be genealogical; and this would be strictly natural, as it would connect together all languages, extinct and recent, by the closest affinities, and would give the filiation and origin of each tongue.*

La capacità discriminante dei dati prodotti dalla genetica molecolare moderna dà corpo alla profezia di Darwin e permette di usare l'albero filogenetico della nostra specie per inferire le prime suddivisioni di un ipotetico albero linguistico. La presentazione orale che ho inteso offrire nel Seminario ne rappresenta l'elaborazione più aggiornata. Le famiglie linguistiche di più antica data devono essere africane: la Khoisan è probabilmente la più antica, e la Afro-Asiatica la più recente. La ricostruzione che oggi si può inferire offre maggiori dettagli, anche se il terreno interdisciplinare in cui avviene il confronto tra geni e lingue suscita dubbi e critiche talvolta incandescenti soprattutto tra i cultori delle discipline linguistiche. In ogni caso utili: il nostro genoma ricorda la sua evoluzione passata con una meravigliosa ricchezza di dettagli che è riflessa anche nelle lingue che parliamo. Lingue e geni contribuiscono a comprendere la storia dell'uomo che è un processo che non si può riprodurre e ripetere in laboratorio e ad apprezzare la sua diversità: strumenti - le une e gli altri - che danno ad alcune voci silenziose del passato una possibilità di parlare ancora.

Per offrire al lettore un contributo scritto più puntuale ma altrettanto ambizioso mi limiterò ad esporre alcuni risultati di un'analisi inedita su un sottoinsieme del quadro complessivo, quello che riguarda lingue e geni in Europa

Può sembrare singolare che uno studioso che non si occupa professionalmente di lingua e di linguistica tratti un argomento di linguistica così controverso quale questo ma ho avuto la fortuna, in un ambiente accademico dalle ramificazioni complesse, comprensibilmente diffidente e geloso del suo contesto disciplinare, poco aperto ad avventure glottocronologiche quale mi pareva essere quello della linguistica italiana, di incontrare - complice l'ombra lunga, ricca di fermenti di quel lungimirante ed innovativo maestro torinese di linguistica storica che fu Benvenuto Terracini - una docente di linguistica interessata, attenta, vivace, generosa di suggerimenti: Bice Mortara Garavelli. La sua disponibilità non solo mi fu preziosa per conoscere le posizioni di stampo prevalentemente strutturalista della linguistica in Italia (e non solo), ma fu anche il segno di una virtù rara in noi ricercatori e docenti: la capacità di cogliere il nuovo in qualunque disciplina si trovi, pronti a mettere in discussione le proprie conoscenze pur di arricchirle entro un orizzonte più vasto. Mi riferisco particolarmente al caso della classificazione e dell'origine delle lingue indo-europee nonché dei popoli che le parlano.

L'Indo-europeo, com'è noto, fu la prima famiglia linguistica ad essere riconosciuta e ricostruita. Il fatto che gli studiosi di linguistica nel mondo occidentale parlassero lingue indo-europee, fece sì che questa famiglia venisse studiata con maggiore impegno.

La famiglia linguistica indo-europea si distingue anche perché ha costituito il primo esempio di modello di *albero filogenetico*, poco dopo che Darwin ebbe introdotto tale modello per spiegare l'evoluzione biologica come un processo di fissioni successive di specie, causate da modificazioni biologiche e ambientali: August Schleicher (1863) pubblicò il primo albero linguistico della famiglia indo-europea come esempio della sua *Stammbaumtheorie*, una teoria evolutiva secondo cui da una madre lingua originaria si sarebbero differenziate, per divisioni successive, tutte le lingue indo-europee. Alcuni anni dopo il suo discepolo Johannes Schmidt (1872) propose una teoria evolutiva alternativa che chiamò *Wellentheorie*. I modelli ad albero sottolineano l'evoluzione indipendente delle lingue da un'origine unica: una volta separate, sono parlate in aree geografiche distinte senza comunicazione tra loro se non in modo limitato.

I modelli di tipo ondulatorio pongono in risalto il fenomeno di scambio di singole parole e altri caratteri linguistici tra gruppi vicini, sia tra dialetti locali all'interno di un'unica lingua sia tra lingue diverse. Nessuno dei due modelli

può fornire un quadro completo dell'evoluzione storica e geografica delle lingue.

#### I DATI LINGUISTICI

Dyen, Kruskal e Black (1992) hanno studiato le affinità lessicali dei 200 significati selezionati da Swadesh in base alla loro "universalità" e invarianza temporale (lista di Swadesh) in 84 lingue indo-europee (escludendo quelle estinte). I risultati della loro analisi identificano chiaramente sei raggruppamenti:

- *Indiano* (Kashmiro, Marathi, Gujarati, Hindi, Panjabi, Lahnda, Bengalese, Nepalese, Khaskura, Singalese)
- *Iranico* (Osseto, Afgano, Waziri, Persiano, Tadziko, Belucistano, Wakhi)
- *Celtico* (Irlandese A, Irlandese B, Gallese N, Gallese C, Bretone L, Bretone SE, Bretone ST)
- *Balto-Slavo* (Lituano O, Lituano ST, Lettone, Sloveno, Lusaziano L, Lusaziano U, Ceco, Slovacco, Ceco E, Ucraino, Bielorusso, Polacco, Russo, Macedone, Bulgaro, Serbocroato)
- *Romanzo* (Rumeno, Valacco, Italiano, Ladino, Provenzale, Francese, Vallone, Francese, Creolo C, Francese, Creolo D, Sardo N, Sardo L, Sardo C, Spagnolo, Portoghese, Brasiliano, Catalano)
- *Germanico* (Tedesco, Olandese della Pennsylvania, Olandese, Afrikaan, Fiammingo, Frisone, Svedese Up, Svedese VL, Svedese L, Danese, *Riksmal*, Islandese, Faroese, Inglese)

Dopo aver eliminato alcune lingue geograficamente vicine (Albanese C e K, Armeno L, Bretone ST, Fiammingo, Greco D, K, e MD, Irlandese B, Lituano O, Lusaziano L, Svedese Up e VL), nonché alcuni dialetti importati (Afrikano, Brasiliano, Zingaro, Olandese della Pennsylvania, Valacco) e le lingue creole (Creolo Francese C e D, Takitaki) la nostra analisi si è limitata a 63 lingue.

## DISTANZE LESSICALI, TEMPI DI SEPARAZIONE E ALBERI LINGUISTICI

Il modello glottocronologico di Swadesh (1955) prevede che i tempi di separazione di due lingue si possano stimare dalla percentuale di significati con la stessa radice lessicale (in inglese *cognate*), mediante una semplice funzione esponenziale negativa tra tale percentuale e tempo. Questa funzione, del tutto simile a quella che si usa per descrivere il decadimento radioattivo, assume che la probabilità di sostituzione di una parola con un'altra con lo stesso significato è costante nel tempo, in lingue diverse e per parole diverse. Kruskal *et al.* (1973) hanno dimostrato che i tassi di cambiamento variano invece notevolmente da significato a significato; la funzione empirica che lega la percentuale di significati con la stessa radice e il tempo di separazione non è esponenziale negativa, ma ha una concavità verso l'alto, per cui è molto facile che le stime dei tempi remoti siano erronee per difetto (in pratica, troppo brevi). Egli calcolò anche una stima empirica del tasso di cambiamento nel tempo dei singoli significati.

Dopo aver trasformato le percentuali di significati con la stessa radice per ogni coppia di lingue prese in esame in distanze lessicali, abbiamo ricostruito con tre metodi statistici diversi [*average linkage* (o UPGMA: Sokal e Michener 1958) che presuppone che il lessico delle varie lingue evolva con una velocità costante; *unione tra vicini* (o NJ: Saitou e Nei, 1987) che non fa riferimento a questa ipotesi; e un nuovo metodo che permette la valutazione statistica di percorsi evolutivi diversi] un albero evolutivo che, partendo da un'unica lingua madre (la *radice* dell'albero), conduce attraverso una serie di separazioni successive (*filogenesi*) alle lingue che si parlano oggi (*foglie* dell'albero). In Figura 1 è illustrato un albero filogenetico delle lingue indo-europee la cui scala temporale, rappresentata nell'asse orizzontale, tiene conto della variabilità dei tassi evolutivi dei singoli significati.

Ad ogni separazione sono indicati i valori statistici di *bootstrap*, percentuali che indicano la frazione degli alberi trovati con quella separazione, quando i dati sono ricampionati a caso: tanto più sono vicini al 100%, quanto più quella separazione è riproducibile, cioè non un semplice prodotto del caso. La filogenesi delle divisioni in grandi famiglie linguistiche evidenziata dalle 63 lingue analizzate nella fig. 1 è ricostruita nella fig. 2. I numeri associati alle foglie indicano quante lingue sono incluse in ogni raggruppamento, mentre la loro posizione indica il tempo in cui il raggruppamento o *cluster* incomincia a ramificarsi nelle singole lingue.



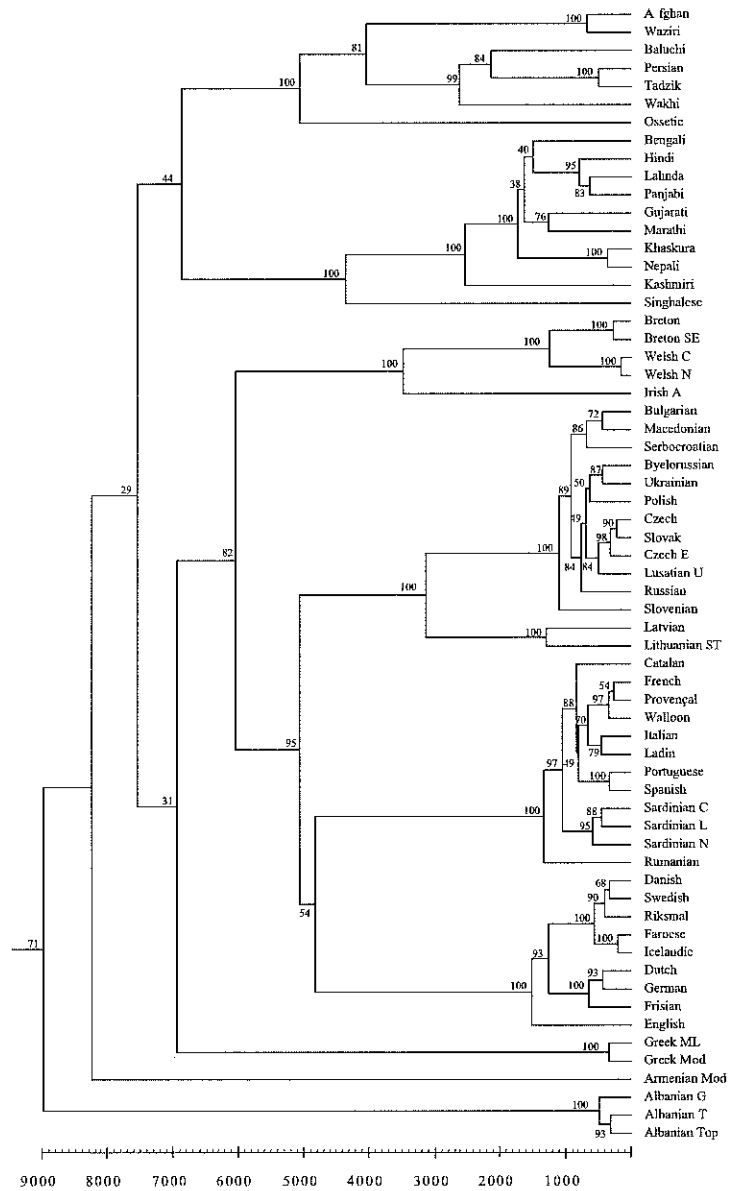


Fig. 1. - Albero filogenetico basato sulle distanze lessicali di 63 lingue Indo-Europee. I numeri sui rami dell'albero indicano in percentuale il grado di riproducibilità del raggruppamento a destra del ramo secondo il test statistico del *bootstrap*. La scala temporale sotto l'albero è calibrata in modo tale da tenere conto della velocità di sostituzione di ogni lessema (v. testo).

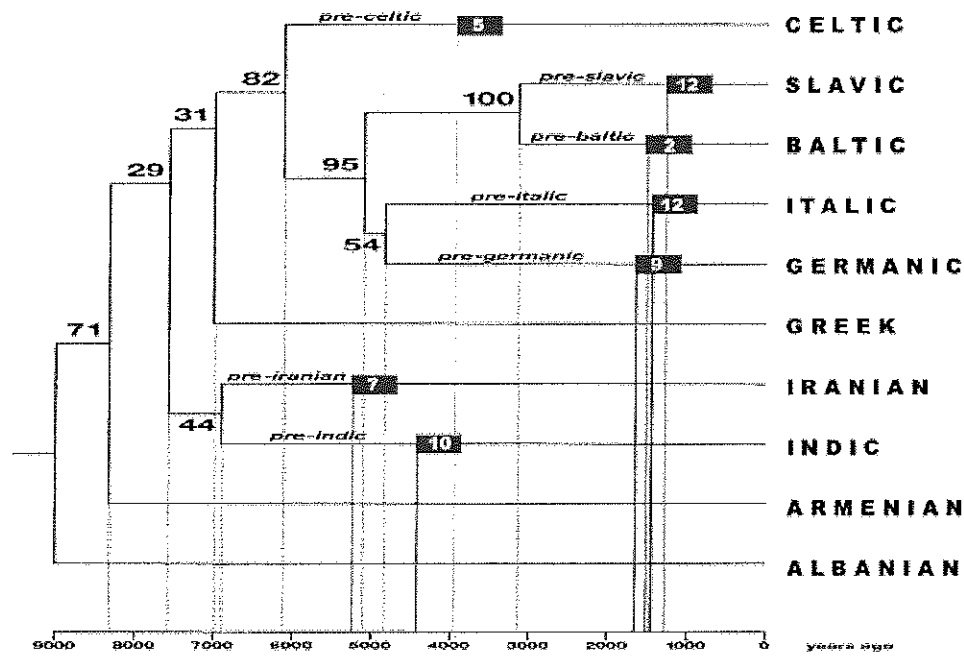


Fig. 2. - Albero filogenetico che indica la separazione in famiglie linguistiche maggiori suggerita dall'albero in fig. 1. I numeri entro i quadrati indicano il numero di lingue che si separano e la posizione del quadrato il tempo in cui avviene la separazione.

### PRESTITI LINGUISTICI

All'interno dei *cluster* in cui si raggruppano le varie lingue, alcune assumono una posizione isolata, come l'Inglese all'interno del cluster germanico, quasi certamente a causa di numerose parole «imprestate» dalle lingue romanze; il Rumeno all'interno delle romanze, probabilmente a causa di prestiti da parte delle lingue slave; e forse lo Sloveno per le parole imprestate dall'Italiano. Le posizioni separate dell'Irlandese, Cingalese, e Osseto potrebbero avere la stessa origine. La matrice delle distanze lessicali tra le 63 lingue permette di fare un'analisi di queste ipotesi con i semplici calcoli riportati in tab. 1. Si noti che le distanze lessicali che abbiamo calcolato non debbono essere considerate un campione casuale valido per un vocabolario completo, poiché i 200 significati scelti da Swadesh sono stati scelti proprio perché conservati nel tempo più di qualsiasi altro significato scelto a caso.

La tabella 1 conferma i prestiti linguistici già indicati aggiungendone altri possibili, come ad esempio l'Inglese dal Balto-Slavo; l'Irlandese dal Germanico e il Rumeno dal Germanico. Nel caso dell'Osseto, una o tutte le famiglie linguistiche Germanico, Indiano, Romanzo e Slavo sono candidate a prestiti, ma potrebbero aver contribuito anche altre lingue non indo-europee. Nel caso dello Sloveno, sono candidati a prestiti sia il Germanico sia il Romanzo.

TABELLA 1. – *Medie e deviazioni standard delle proporzioni di radici lessicali comuni tra popolazioni (colonne) che nell'albero filogenetico di fig. 1 si collocano in una posizione isolata rispetto alla loro famiglia linguistica e le popolazioni sia della loro stessa famiglia linguistica (in neretto) sia delle altre famiglie linguistiche (in corsivo quelle maggiori). La prima colonna rappresenta la somiglianza media di tutte le lingue entro le varie famiglie linguistiche: è sempre maggiore delle somiglianze descritte nella stessa riga.*

	Media	Inglese	Irlandese	Rumeno	Osseto	Sloveno	Cingalese
Albanese	0.840	0.120±0.008	0.099±0.005	0.112±0.003	0.090±0.003	0.148±0.014	0.095±0.003
Celtico	0.563	0.170±0.009	<b>0.326±0.019</b>	0.191±0.026	0.106±0.005	0.175±0.009	0.096±0.007
Germanico	0.683	<b>0.570±0.026</b>	<i>0.180±0.010</i>	<i>0.232±0.014</i>	<i>0.141±0.008</i>	<i>0.245±0.016</i>	0.128±0.006
Greco	0.869	0.146±0.016	0.132±0.011	0.147±0.011	0.100±0.002	0.163±0.020	0.107±0.001
Indiano	0.490	0.146±0.010	0.117±0.007	0.160±0.013	<i>0.147±0.017</i>	0.183±0.015	<b>0.260±0.021</b>
Iranico	0.355	0.124±0.021	0.100±0.013	0.126±0.021	<b>0.227±0.043</b>	0.134±0.017	0.124±0.024
Romanzo	0.704	<i>0.235±0.007</i>	0.163±0.008	<b>0.606±0.029</b>	<i>0.141±0.006</i>	0.222±0.010	<i>0.143±0.005</i>
Balto-Slavo	0.621	<i>0.232±0.015</i>	<i>0.178±0.010</i>	<i>0.213±0.013</i>	0.133±0.004	<b>0.606±0.126</b>	<i>0.145±0.010</i>

#### VALIDITÀ DELLE STIME DEI TEMPI DI SEPARAZIONE

Le lunghezze dei rami possono essere stime attendibili dei tempi di separazione? Questa domanda ne implica un'altra: entro quali limiti sono valide le assunzioni sulle quali si basano le stime?

Abbiamo stimato le distanze lessicali tra lingue assumendo che seguano una distribuzione statistica particolare: la distribuzione gamma, che tiene conto della variazione dei tassi di cambiamento delle singole parole. La speranza principale di saggiare l'efficacia del modello paragonando le stime dei tempi con tempi di separazione noti, rimane tale perché questi ultimi sono in gran parte sconosciuti. In ogni caso si debbono tenere presenti alcune schematizzazioni.

Tra queste ramificazioni, poche corrispondono a tempi di separazione puntuali. Raramente l'origine di una nuova lingua è improvvisa; normalmente è un processo continuo che attraversa fasi di cambiamento più o meno veloci. L'introduzione di una lingua «straniera» da parte dei conquistatori di un Paese può essere improvvisa, ma la transizione verso una nuova lingua sarà certamente lenta e potrebbero passare secoli prima che la vecchia lingua scompaia del tutto.

Abbiamo accennato al fatto che una lingua con molti prestiti da lingue di diverse famiglie linguistiche può separarsi nell'albero filogenetico prima di tutti gli altri membri della propria famiglia linguistica (Cavalli Sforza, Menozzi e Piazza, 1994). Perciò non è certo che l'origine delle lingue germaniche debba essere posta veramente sul ramo da cui deriva l'Inglese (a causa dei prestiti linguistici numerosi potrebbe distinguersi in un tempo apparentemente antecedente) o sul ramo più tardo che separa Olandese, Tedesco, e Frisone dalle lingue scandinave. Lo stesso problema si pone per l'origine delle lingue romanze; si deve porre all'inizio del ramo che conduce al Rumeno o al ramo più tardo che porta al Sardo, o ancora dopo?

L'origine sia temporale sia spaziale delle lingue indo-europee è stata fonte di discussioni appassionate e di una multiforme varietà di ipotesi, soprattutto per quanto riguarda l'area geografica di origine (si veda *i.e.* Mallory, 1989).

Comunemente l'origine nel tempo viene fatta risalire a circa 6500 - 6000 anni fa, su basi non molto rigorose. È pur vero che la percentuale di significati con radici comuni corrispondente a questo tempo di separazione nella curva glottocronologica classica di Swadesh (conservazione dell'86% di parole ogni 1000 anni) dà una conservazione del 16% di significati con radici comuni tra due lingue dopo 6000 anni, che non è molto lontano dal valore osservato circa il 10%: l'Albanese, che si separa per primo negli alberi delle FIGURE 1 e 2, mette in evidenza percentuali di significati con radici comuni con tutte le altre lingue che variano tra il 7,1% e il 18%. Tuttavia 6000 anni potrebbero essere una stima accettabile solo se il tasso di cambiamento per ogni parola fosse

costante. La stima di 9000 anni proposta nella fig. 1 è quella ottenuta adottando tassi di cambiamento variabile da parola a parola: la distribuzione statistica gamma delle distanze lessicali è stata adattata ai tassi di cambiamento dei 200 significati della lista di Swadesh che per le lingue indo-europee sono stati calcolati in Kruskal *et al.* (1973).

#### TEMPO E LUOGO DI ORIGINE DELLE LINGUE INDO-EUROPEE

Nel 1987 Colin Renfrew avanzò l'ipotesi che gli agricoltori che incominciavano ad espandersi dall'Anatolia circa 9000 anni fa parlassero lingue indo-europee. La sua ipotesi si basava su quella proposta originariamente da Ammermann e Cavalli-Sforza (1973, 1984) che l'espansione dell'agricoltura nel Neolitico dalla Mezzaluna Fertile fosse dovuta all'espandersi degli agricoltori stessi e non soltanto alla tecnologia agricola, e alla considerazione che i popoli che migrano conservano la propria lingua, se possibile. Questa affermazione è senza dubbio corretta quando chi migra si stabilisce in una zona disabitata o poco abitata. Per quanto riguarda gli agricoltori del Neolitico che si spostavano dalla Mezzaluna Fertile verso Ovest, Nord e Est, è molto probabile che essi si muovessero in grandi gruppi, poiché l'agricoltura (lo stimolo per i loro spostamenti) permetteva una rapida espansione ed il raggiungimento di densità di popolazioni fino ad allora impensabili.

L'idea di un movimento vero e proprio degli agricoltori e non di una semplice trasmissione culturale delle tecniche agricole fu provata da più analisi della struttura genetica dell'Europa (Menozzi *et al.*, 1978; Sokal e Menozzi, 1982; Sokal *et al.*, 1991). L'ipotesi di Renfrew venne invece criticata dalla maggior parte degli studiosi di lingue indoeuropee (vedi: Mallory 1989; Lehmann 1993: 283-288) e non ebbe una accoglienza favorevole rispetto alle ipotesi precedenti, ora generalmente associate al nome di un'altra archeologa, Marjia Gimbutas (1970; 1985): gli Indo-Europei emigrarono verso l'Europa dalla steppa del Ponto, nella Russia meridionale, dal Dniepr al Volga (che lei chiamò «Kurgan», dal nome russo dei tumuli che ricoprivano le tombe e che sono molto comuni in questa zona) fin dall'età del Bronzo, cioè circa 5500 anni fa. Tuttavia non bisogna dimenticare che Dolgopolsky (1988), Gamkrelidze e Ivanov (1991) avevano collocato l'origine delle lingue indo-europee in Anatolia per ragioni del tutto diverse.

La struttura genetica dell'Europa moderna sembrerebbe confermare che molto probabilmente un'espansione importante e più di una migrazione ebbero origine da quella regione geografica (Cavalli-Sforza *et al.*, 1993, 1994). I dati genetici non possono fornire datazioni precise associabili a movimenti migratori, a maggiore ragione nella zona dei Kurgan, uno dei complessi preistorici più grandi d'Europa, da cui molto probabilmente ebbero origine ulteriori espansioni di popolazioni molto dopo l'età del Bronzo. In quell'area troviamo, circa 6000 anni fa, la cultura di Sredni-Stog; più tardi (5500-4500 anni fa) le culture Yamnaya precedentemente chiamate culture *pit-grave* (*tomba a fossa*) che si estendevano dalla parte meridionale del fiume Bug fino al fiume Ural. A partire da circa 5000 anni fa incominciamo a trovare prove della presenza in queste culture di carri a due e a quattro ruote (Anthony, 1995: 561 dà  $4370 \pm 120$  anni fa e  $4440 \pm 40$  anni fa per due carri rinvenuti in tombe Yamna rispettivamente nelle regioni del basso Dniepr e del fiume Kuban). I morti venivano sepolti in una fossa (in Russo *yamna*, da cui Yamnaya) sdraiati sulla schiena e con le ginocchia ripiegate. Questo tipo di sepoltura molto caratteristica rimase in uso lungamente, anche quando gli Indo-Europei si erano ormai dispersi lontano da questa zona.

Alcune scoperte archeologiche a Dereivka, un sito a sud di Kiev sul Dniepr (Anthony e Brown, 1991) hanno mostrato che le date più antiche sono quelle proposte originariamente dalla Gimbutas, e che l'addomesticamento precoce del cavallo nella steppa e la costruzione di carri da guerra potrebbe aver fornito al popolo Yamnaya una notevole capacità di muoversi e di fare la guerra. Ad una cultura della steppa simile a questa (probabilmente la cultura di Andronov o una precedente, vicino al Mare di Aral), si attribuiscono le migrazioni che, iniziate forse prima di 4000 anni fa, portarono all'occupazione dell'Iran e dell'India da parte di popoli che parlavano una lingua indo-europea.

Abbiamo suggerito varie volte e da tempo (Cavalli-Sforza *et al.*, 1993; Cavalli-Sforza *et al.*, 1994; Piazza *et al.*, 1995) che le ipotesi di Renfrew e Gimbutas non dovrebbero considerarsi alternative: potrebbero essere compatibili, come è stato anticipato da uno studioso di storia antica vissuto in un'epoca pre-radiocarbonio, Schrader, fin dal 1890: «gli Indo-Europei praticavano già l'agricoltura in un luogo tra il Dniepr e il Danubio dove si sviluppò il linguaggio agricolo del ramo Europeo» (citato da Lehmann, 1993: p. 279).

L'insediamento nella steppa da parte degli agricoltori del Neolitico dev'essere avvenuto dopo la loro emigrazione dall'Anatolia, e se le espansioni ebbero inizio 9500 anni fa dall'Anatolia e 6000 anni fa dalla regione della cul-

tura Yamnaya, la migrazione verso la regione del Volga-Don dall'Anatolia, probabilmente attraverso i Balcani, durò circa 3500 anni. Nel Volga-Don si sviluppò una cultura completamente nuova, soprattutto pastorale, stimolata dall'ambiente poco favorevole all'agricoltura tradizionale, che offriva però nuove, attraenti possibilità.

Perciò la nostra ipotesi è che le lingue indo-europee siano derivate da una espansione secondaria dalla regione della cultura Yamnaya dopo che gli agricoltori del Neolitico, forse provenienti dall'Anatolia e stabilitisi lì, ebbero sviluppato un regime di pastorizia nomade. Questa ipotesi può anche giustificare l'idea di William Sturtevant, secondo la quale l'Indo-Ittita ha due rami: l'Anatolico e l'Indo-Europeo (Sturtevant 1962), ma non offre una soluzione completa al problema perché non vi sono tracce della lingua che questi coloni possono aver parlato.

Come si comporta il nostro albero lessicale rispetto a questa ipotesi? La radice dell'albero in fig. 1 separa gli Albanesi dagli altri, con una riproducibilità del 71% calcolata col metodo statistico del *bootstrapping*. Il ramo più antico dopo questo è l'Armeno. L'interpretazione più semplice è che la lingua dei primi agricoltori che emigrarono dall'Anatolia sopravviva oggi in due lingue, l'Albanese e l'Armeno, che si allontanarono dalle lingue indo-europee più antiche con cambiamenti linguistici divergenti, ma le cui popolazioni corrispondenti sono rimaste relativamente vicine al punto di origine comune. Questa ricostruzione non è in contrasto con la topologia data in Warnow (1997), dove l'Albanese può separarsi dall'albero in un intervallo di tempo incerto che va da una separazione che precede tutte le altre ad un raggruppamento con l'Armeno e il Greco. Tutti i primi gruppi di individui di origine anatolica che parlavano una lingua indo-europea vennero sommersi da arrivi posteriori di gruppi provenienti da regioni di cultura Yamnaya dotati di una forte organizzazione sociale e militare, che li mettevano in grado di sottomettere le comunità agricole locali e di imporre la propria lingua. Una eccezione potrebbero essere i Greci che sembrerebbero legati più alle popolazioni europee che alle popolazioni di cultura Yamnaya. È anche possibile, come risulta nell'albero di Warnow (1997) dove il Greco si unisce all'Armeno, che la lingua antenata al Greco fosse quella di un terzo gruppo di agricoltori di origine anatolica.

Se associamo alla prima separazione dell'albero l'inizio dell'espansione degli agricoltori anatolici pre-indo-europei, circa 9000 anni fa (v. anche fig. 1), possiamo calcolare che l'origine del ramo europeo risalga a circa 6000 anni fa. I tre o quattro rami maggiori (pre-Celtico, pre-Baltico-Slavo, pre-Ita-

lo-Germanico; o pre-Celtico, pre-Balto-Slavo, pre-Romanzo, pre-Germanico, se si ritiene che il sottogruppo di breve durata Pre-Romanzo e Pre-Germanico sia un'incongruità dovuta a prestiti non individuati, come si è accennato) potrebbero corrispondere entro certi limiti alle ondate migratorie ipotizzate da Gimbutas (1991: c.10), ma le datazioni archeologiche sono troppo scarse per fornire associazioni certe. È ragionevole pensare che una prima migrazione corrisponda al primo ramo, i pre-Celti (6000 anni fa secondo l'albero), che si insediarono per primi e proseguirono verso Ovest. Infatti ciò che resta della loro lingua vive ancora oggi nei lembi estremi dell'area geografica originariamente occupata. Essendo i primi, ebbero tempo di stabilirsi nell'Europa centrale e procedere verso Ovest. In origine potrebbero anche essersi stabiliti a Est e aver abbandonato la parte orientale del continente sotto la pressione di arrivi successivi. Più tardi ebbero il vantaggio di essere tra i primi a sviluppare la cultura dell'Età del Ferro, e poterono sviluppare una grande comunità anche linguistica. Prima del dominio romano si estesero in mezza Europa, dalla Spagna alla Francia, alla maggior parte delle Isole Britanniche, l'Italia del Nord e l'Europa centrale.

Una seconda espansione (ca. 5000 anni fa) potrebbe essersi divisa in due ondate principali, una che si estendeva verso l'Europa meridionale dando luogo alle lingue Italiche, una delle quali (il Latino) successivamente venne diffusa dall'Impero romano nei territori che dominava; l'altra che giunse nell'Europa del Nord, dando vita alle lingue germaniche, che comprendono quelle scandinave. I gruppi italici pre-romanzi e germanici corrispondono probabilmente a due migrazioni indipendenti dalla stessa regione; la prima ebbe inizio meno di 5000 anni fa, mentre si può presumere che la seconda - più probabilmente l'espansione Germanica - sia successiva, poiché ha popolato un territorio più orientale. Il complesso Remedello-Rinaldone in Italia (5200-4500 anni fa) reca tutti i segni di un'invasione indo-europea: un nuovo stile di ceramica, un nuovo rito di sepoltura, l'introduzione di un'aristocrazia guerriera, della metallurgia, del cavallo e del carro (Cardarelli, 1993: pp. 386-387). L'arrivo in Inghilterra rappresenta l'ultimo lembo dell'ondata germanica (con mercenari anglo-sassoni che si stabilirono in Inghilterra nel V-VI secolo d.C. e soppressero insieme alla locale lingua romanza la maggior parte di ciò che rimaneva del Celtico).

L'ultima ondata migratoria preistorica verso ovest è la balto-slava. La separazione tra le lingue baltica e slava risale a circa 3000 anni fa. Questa datazione conferma la ipotesi fondata su dati archeologici che l'unità balto-slava



si deve far risalire a circa 4000-3000 anni fa, ed è compatibile con l'ipotesi che il popolo balto-slavo sia associato alla cultura Komarov, che sappiamo essere stata presente a quell'epoca nella zona intorno al tratto meridionale del fiume Bug. Il fatto che allora i popoli di lingua balto-slava occupassero (e occupino tuttora) l'estremità orientale dell'Europa, è un'ulteriore conferma del fatto che erano gli immigrati più recenti.

I gruppi Iraniano e Indiano hanno un breve tratto iniziale comune, e si possono considerare un unico cluster che si divise molto presto (7000 anni fa). I loro antenati potrebbero avere occupato oasi diverse nell'Asia centrale prima di muoversi verso Sud. La ramificazione singalese può sembrare precoce a causa della mescolanza successiva con popoli di lingue dravidiche, ma la data di origine (un poco più antica di 4000 anni fa) suggerita dall'albero, coinciderebbe con la prima penetrazione del subcontinente Indiano da parte di gente legata alla cultura di Andronov. L'occupazione delle varie regioni indiane non è compatibile con quanto suggerisce l'albero, da 2500 a 500 anni fa.

La differenziazione successiva in lingue locali suggerita dall'albero non è del tutto soddisfacente: le prime lingue romanze a separarsi, circa 1200 anni fa, sembrano essere le tre varianti sarde. In realtà la Sardegna fu occupata per la prima volta dai Romani circa 2000 anni fa, ma la differenziazione del Sardo da altre lingue Romanze non può essere incominciata tanto presto. La conquista romana della Sardegna significò soltanto l'inizio dell'uso del Latino.

La lingua italiana incominciò ad assumere la sua forma attuale circa 900-1000 anni fa: perciò l'ipotesi che le lingue romanze si siano separate 800 anni fa non è illogica, come non lo è quella del Sardo, ma la separazione di 500 anni fa del Provenzale e del Francese appare troppo recente. Probabilmente se si tenesse conto in modo esplicito degli scambi lessicali tra popolazioni geograficamente vicine (per es. usando il metodo di Embleton (1986)) le date di separazione recenti potrebbero essere meglio stimate.

Una considerazione interessante a sostegno del nostro albero è che l'ordine delle separazioni dovrebbe essere congruente con la geografia delle popolazioni coinvolte. Infatti non è irragionevole pensare che le separazioni più recenti collochino le popolazioni separate ad una distanza geografica minore dal luogo di origine della loro separazione (che per il raggruppamento principale del nostro albero è probabilmente la steppa del Ponto): per cui le separazioni più recenti dovrebbero aver luogo in un'area geografica più vicina ad essa. L'ordine geografico da Ovest a Est della steppa è chiaramente Celtico,

Germanico a Nord e Romano a Sud, poi Balto-Slavo, che corrisponde all'ordine dei rami della fig. 2.

#### DATI LINGUISTICI E DATI GENETICI

Possiamo paragonare i dati lessicali fin qui analizzati con i dati genetici? Il paragone è tutt'altro che banale per una serie di ragioni.

Un problema preliminare è quello di trovare i dati genetici per ognuna delle 63 popolazioni che parlano le 63 lingue che formano il database lessicale prima esaminato. Si tratta di una ricerca quasi senza speranza con i dati attuali, ma l'abbiamo aggirata usando le mappe geografiche delle frequenze dei 95 geni analizzati in Europa pubblicate in Cavalli-Sforza *et al.* (1994): abbiamo interpolato la superficie di ogni frequenza genica nel baricentro delle aree geografiche che delimitano l'uso di ognuna delle 63 lingue in esame. Abbiamo quindi ricostruito una matrice di dati completa, per cui ognuna delle 63 popolazioni ha assegnati 95 valori approssimati di frequenze geniche.

È stato dimostrato ormai da tempo (Cavalli-Sforza *et al.*, 1993, 1994; Piazza, 1993) che la rappresentazione delle relazioni genetiche tra le popolazioni mediante un albero evolutivo non è una buona rappresentazione se applicata all'Europa. Ciò è dovuto principalmente al modello evolutivo mediante il quale si ricostruisce l'albero che non permette che i suoi rami, una volta separati in popolazioni differenti, possano poi di nuovo convergere in una sola popolazione, cioè si possa verificare una «mescolanza» genetica tra le popolazioni corrispondenti ai rami dopo la loro separazione. La storia delle popolazioni europee fa ritenere che l'Europa non soddisfi questo presupposto fondamentale del modello ad albero. In realtà, è proprio a causa di questo problema che abbiamo usato il sistema di elaborare *mappe a curve di livello* delle frequenze geniche su una superficie geografica (Menozzi *et al.*, 1978) per descrivere la struttura genetica dell'Europa.

In ogni caso, l'albero genetico derivato dai dati delle frequenze geniche interpolate per le 63 popolazioni che parlano Indo-Europeo non fornisce più informazioni di quello pubblicato in Cavalli-Sforza *et al.* (1994: p. 268; i dati originali erano esattamente gli stessi) e indica le popolazioni che sono più isolate dalle altre: la Sardegna si separa prima ed è una delle poche separazioni statisticamente «robuste» (cioè affidabili) secondo l'analisi del *bootstrap*.

Il nostro interesse principale è quello di saggiare l'esistenza o meno di una possibile congruenza tra i dati linguistici e quelli genetici. La struttura ad albero si è dimostrata non valida per una corretta rappresentazione genetica delle popolazioni europee, per cui abbiamo lavorato direttamente sulle matrici delle distanze lessicali e genetiche da cui derivano gli alberi, facendo due analisi. In primo luogo abbiamo calcolato un coefficiente di correlazione tra tutte le distanze linguistiche e genetiche, e lo abbiamo analizzato col test di Mantel (1967) per valutare la significatività statistica. Abbiamo trovato una correlazione del 17% perciò non alta (il massimo è il 100%) ma significativamente diversa da una correlazione nulla. Si noti che il valore del coefficiente di correlazione stima la frazione lineare di una correlazione tra due serie di distanze, ma non esistono ragioni a priori per cui le distanze linguistiche (lessicali) e genetiche debbano essere associate in modo lineare. Per controllare questo punto abbiamo applicato alle due serie di distanze un'analisi statisticamente più efficace e più sofisticata, secondo il metodo proposto da Schoenemann e Carrol (1970).

Abbiamo ottenuto in questo modo una correlazione del 69%, molto più elevata di quella ottenuta precedentemente, che indica una correlazione non lineare non trascurabile tra i dati genetici e linguistici (lessicali) in Europa.

## CONCLUSIONI

L'evoluzione delle lingue può essere rappresentata sotto forma di albero provvisto di radice (la lingua madre da cui tutte le altre sono originate) i cui nodi interni denotano lingue ancestrali, in un processo del tutto simile a quello che dà luogo all'evoluzione genetica delle popolazioni umane. È noto che il tasso di cambiamento dei lessemi può essere diverso da lessema a lessema, e tale variabilità modifica la funzione esponenziale negativa tradizionalmente adottata per porre in relazione le percentuali di significati con la stessa radice lessicale presenti in una coppia di lingue con il loro tempo di separazione, e perciò inficia ogni misura di datazione: la novità della nostra analisi è quella di aver usato le informazioni empiriche sulla variazione di tali tassi di cambiamento a seconda dei significati per correggere questa relazione. L'analisi così riformulata dovrebbe fornire datazioni più attendibili, e in effetti sembrerebbe associare nuove datazioni a eventi linguistici di importanza non secondaria nell'evoluzione delle lingue indo-europee. L'albero ottenuto dai dati di Dyen

*et al.*, indica poche lingue di origine molto antica, forse associate all'espansione degli agricoltori dall'Anatolia, ed un *cluster* che raggruppa molte lingue probabilmente originatosi nelle steppe del Ponto nella Russia meridionale, in accordo con le ipotesi di Gimbutas e Anthony. I sottogruppi di questo gruppo più esteso si sono separati nello stesso ordine in cui sono geograficamente collocati in Europa oggi.

Anche se siamo ben consapevoli di quanto diversi siano i meccanismi di trasmissione delle informazioni linguistiche e biologiche, siamo stati stimolati da analisi precedenti (Cavalli-Sforza *et al.*, 1988, 1992; Piazza *et al.*, 1995) a paragonare la struttura informativa fornita da un albero filogenetico generato da dati lessicali con quella fornita usando dati genetici, e, elemento più interessante, a cogliere le opzioni di entrambe le ricostruzioni per meglio inferire quando e dove abbiano avuto origine i popoli che parlano lingue indo-europee. Riassumiamo risultati e proposte:

1. Lavori precedenti in campo genetico (Cavalli-Sforza *et al.*, 1994; Piazza *et al.*, 1995) hanno dimostrato che una componente importante della struttura genetica dell'Europa potrebbe essere associata alla possibile espansione iniziale di alcune popolazioni che parlavano lingue indo-europee da una regione geografica corrispondente a quella in cui si insediò la prima cultura Yamnaya (5500-5300 anni fa) in Ucraina, tra il medio o basso Volga e il fiume Dniepr. Si tratta della prima cultura della steppa potenzialmente in grado di sviluppare un'efficace economia pastorale della steppa *a*) addomesticando animali da pascolo, *b*) andando a cavallo, *c*) usando veicoli a ruote: per tutte queste ragioni di potenziale superiorità culturale, quell'area e quella cultura sono state considerate candidate ideali per annettervi l'insediamento di un popolo che parlasse una lingua proto-indo-europea (Anthony, 1995). È possibile che nella vasta area di cultura Yamnaya si parlassero più lingue della famiglia che includeva il proto-Indo-Europeo, e che il proto-Indo-Europeo abbia sostituito gradualmente tali lingue sorelle attraverso un processo di predominio da parte di élites socio-culturali che si imposero con la forza. L'anatolico potrebbe essere derivato da una delle lingue sorelle non ancora completamente assorbita o sostituita durante questo processo (Anthony, comunicazione personale), e questa lingua potrebbe essere stata parlata da alcuni gruppi di agricoltori neolitici la cui diffusione antecedente dal Vicino Oriente è anche documentata da dati genetici (Menozzi *et al.*, 1978; Sokal *et al.*, 1991).

2. Se questo modello (compatibile con i dati archeologici e genetici) sia stato in grado di generare un processo plausibile di cambiamento della lin-

gua, i dati lessicali di Dyen *et al.* (1992) potrebbero testimoniare in modo coerente. Quando le distanze lessicali sono state corrette tenendo conto per ciascuno dei 200 significati della lista di Swadesh usato da Dyen *et al.* (1992) della loro velocità di cambiamento, abbiamo ottenuto un albero con una radice datata a circa 9000 anni fa e successive separazioni nel seguente ordine: *a)* Albanese, *b)* Armeno, *c)* Indo-Iraniano, *d)* Greco ed *e)* il sottogruppo Celtico, Italo-Germanico, Balto-Slavo. L'Albanese è una lingua ancora poco studiata ma si conosce che ha subito forti influssi dal Latino, dal Turco, dal Greco e dallo Slavo. È molto probabile che sia cambiato a tal punto dal Proto-Indo-Europeo o forse dal Pre-Indo-Europeo da poter rappresentare una possibile radice dell'albero. Anche l'Armeno ed il Greco hanno una lunga storia: le loro date di origine molto antiche (tra 8000 e 6000 anni fa) possono essere dovute ad una origine pre-anatolica o a prestiti da lingue vicine o a entrambe le eventualità. La data di 6000 anni fa che segna l'origine di tutte le altre lingue corrisponde alla cultura di Sredny Stog (Ucraina, 6500-5500 anni fa) in cui il cavallo venne addomesticato per la prima volta, e alla cultura Yamnaya successiva (5600-4200 anni fa) quando comparvero i primi carri da trasporto. Anche l'origine del Pre-Indo-Europeo sembra molto antica, forse corrispondente alla cultura dei *Kurgan* (a sud del Mar Caspio, 5000-4250 anni fa) e alla cultura Andronov successiva (4000-3000 anni fa). L'unità balto-slava risale a più di 3000 anni fa, il che potrebbe corrispondere alla cultura Komarov nell'area meridionale del fiume Bug (ca. 3500 anni fa).

3. Per ragioni spiegate altrove, i dati genetici non sono in grado di rappresentare le popolazioni europee in una struttura a forma di albero. Si è calcolata una correlazione tra distanze lessicali e genetiche, e si è elaborato un test di congruità tra i due gruppi di dati. Abbiamo trovato una correlazione bassa ma statisticamente significativa tra le due distanze, molto probabilmente perché i dati lessicali e genetici non sono correlati in modo lineare. Infatti quando i due gruppi di dati vengono trasformati in modo non lineare, senza variare le relazioni tra le distanze, abbiamo ottenuto una correlazione molto più elevata, circa il 70%. Quando si è sottratto l'effetto delle distanze geografiche tra popolazioni/lingue (popolazioni e lingue a distanze più vicine tendono ad essere più simili), abbiamo ottenuto un coefficiente di correlazione più basso, circa il 27%, ma ancora statisticamente significativo. Questi due risultati costituiscono una conferma quantitativa rigorosa delle nostre considerazioni precedenti: i geni europei identificano in Ucraina una possibile area geografica da cui hanno avuto origine e si sono diffusi l'Indo-Europeo (e gli Indo-Europei).

## BIBLIOGRAFIA

- AMMERMAN A.J., CAVALLI-SFORZA L.L., 1973. *A population model for the diffusion of early farming in Europe*. In: C. RENFREW (ed.), *The explanation of Culture Change*. Duckworth, London: 343-357.
- AMMERMAN A.J., CAVALLI-SFORZA L.L., 1984. *Neolithic Transition and the Genetic of Populations in Europe*. Princeton University Press, Princeton.
- ANTHONY D.W., 1995. *Horse, wagon & chariot: Indo-European languages and archaeology*. *Antiquity*, 69: 554-565.
- ANTHONY D.W., BROWN D., 1991. *The origins of horseback riding*. *Antiquity*, 65: 22-36.
- BONET M., PHILLIS C., WARNOW T., YOOSEPH S., 1996. *Constructing evolutionary trees in the presence of polymorphic characters*. In: *Proceedings of the Twenty-eight annual ACM Symposium on the Theory of Computing*. Philadelphia: 220-229.
- BERGSLAND K., VOGT H., 1962. *On the validity of glottochronology*. *Current Anthropology*, 3: 115-153.
- ASTOLFI P., PIAZZA A., KIDD K.K., 1978. *Testing of evolutionary independence in simulated phylogenetic trees*. *Sist. Zool.*, 27: 391-400.
- CARDARELLI A., 1993. *Le età dei metalli nell' Italia settentrionale*. In: A. GUIDI, M. PIPERNO (eds.) *Italia preistorica*. Laterza, Bari: 366-419.
- CAVALLI-SFORZA L.L., FELDMANN M.W., 1981. *Cultural transmission and evolution: A quantitative approach*. Princeton University Press, Princeton.
- CAVALLI-SFORZA L.L., PIAZZA A., 1975. *Analysis of evolution: evolutionary rates, independence and treeness*. *Theoretical Population Biology*, 37: 3-25.
- CAVALLI-SFORZA L.L., WANG W.S.-Y., 1986. *Spatial distance and lexical replacement*. *Language*, 62: 38-55.
- CAVALLI-SFORZA L.L., PIAZZA A., MENOZZI P., MOUNTAIN J., 1988. *Reconstruction of human evolution: bringing together genetic, archaeological and linguistic data*. *Proc. Natl. Sci. USA*, 85: 6002-6006.
- CAVALLI-SFORZA L.L., MINCH E., MOUNTAIN J., 1992. *Coevolution of genes and languages revisited*. *Proc. Natl. Sci. USA*, 89: 5620-5624.
- CAVALLI-SFORZA L.L., MENOZZI P., PIAZZA A., 1993. *Demic Expansions and Human Evolution*. *Science*, 259: 639-646.
- CAVALLI-SFORZA L.L., MENOZZI P., PIAZZA A., 1994. *The History and Geography of Human Genes*. Princeton University Press, Princeton.
- CAVALLI-SFORZA L.L., MINCH E., 1997. *Paleolithic and Neolithic Lineages in the European Mitochondrial Gene Pool*. *Am. J. Hum. Genet.*, 61: 247-251.
- DARWIN C.R. 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. 1st ed. John Murray, London: 371.
- DOLGOPOLSKY A.B. 1988. *The Indo-European homeland and lexical contacts of Proto-Indo-Europeans with other languages*. *Mediterr. Lang. Rev.*, 3: 7-31.

- DYEN I., KRUSKAL J.B., BLACK P., 1992. *An Indoeuropean classification: a lexicostatistical experiment*. Transactions of the American Philosophical Society, 82: Part 5. American Philosophical Society, Philadelphia.
- EMBLETON S.M., 1986. *Statistics in historical linguistics*. Quantitative Linguistics, 30: 1-194.
- EFRON B., TIBSHIRANI R.J., 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- FELSENSTEIN J., 1985. *Confidence limits on phylogenies: an approach using the bootstrap*. Evolution, 39: 783-791.
- FORSTER P., TOTH A., BANDELT H.-J., 1998. *Evolutionary Network Analysis of Word Lists: Visualising the Relationships between Alpine Romance Languages*. Journal of Quantitative Linguistics, 5: 174-187.
- FORT J., MENDEZ V., 1999. *Time-delayed Theory of the Neolithic Transition in Europe*. Physical Review Letters, 82(4): 867-870.
- GAMKRELIDZE T.V., IVANOV V.V., 1991. *The early history of Indo-European languages*. Sci.Am., 262(3): 110-116.
- GIMBUTAS M., 1970. *Proto-Indo-European culture: the Kurgan culture during the fifth, fourth, and third millennium B.C.* In: G. Cardona, H.M. Hoenigswald, A. Senn (eds.), *Indo-European and Indo-Europeans*. University of Pennsylvania Press, Philadelphia: 155-195.
- GIMBUTAS M., 1985. *Primary and secondary homeland of the Indo-Europeans*. Journal of Indo-European Studies, 13: 185-202.
- GIMBUTAS M., 1991. *The Civilizations of the Goodness: the World of Old Europe*. Harper, San Francisco.
- GREENBERG J.H., 1987. *The Languages of Americas*. Stanford University Press, Stanford (CA).
- GUGLIELMINO-MATESSI C.R., VIGANOTTI C., CAVALLI-SFORZA L.L., 1979. *Climate and the evolution of skull metrics in man*. Am. J. Phys. Anthropol., 50: 549-564.
- HIERNAUX J., 1968a. *Bantu expansion: the evidence from physical anthropology confronted with linguistic and archaeological evidence*. J. Afr. Hist., IX(4): 505-515.
- HIERNAUX J., 1968b. *La Diversité Humaine en Afrique Subsaharienne: Recherches Biologiques*. Institut de Sociologie, Université Libre, Bruxelles.
- JOHNSON N.L., KOTZ S., 1970. *Continuous univariate distributions - 1*. Wiley, New York.
- KIMURA M., WEISS G.H., 1964. *The stepping-stone model of population structure and the decrease of genetic correlation with distance*. Genetics, 49: 561-576.
- KRUSKAL J.B., DYEN I., BLACK P., 1973. *Some results from the vocabulary method of reconstructing language trees*. In: I. DYEN (ed.), *Lexicostatistics in genetic linguistics*. Mouton, The Hague: 30-55.
- LALOUEL J.M., 1973. *Topology of population structure*. In: N.E. MORTON (ed.), *Genetic Structure of Populations*. University of Hawaii Press, Honolulu: 139-149.

- LEHMANN W.P., 1993. *Theoretical Bases of Indo-European Linguistics*. Rutledge, London.
- MALLORY J.P., 1989. *In Search of the Indo-Europeans: Language, Archaeology and Myth*. Thames & Hudson, London.
- LEVINE M., 1990. *Dereivka and the problem of horse domestication*. *Antiquity*, 64: 627-640.
- LEVINE M., 1999. *The origin of horse husbandry on the Eurasian steppe*. In: M. Levine, Y. Rassamakin, A. Kislenko, N. Tatarintseva (eds.) *Late prehistoric exploitation of the Eurasian stepp*. McDonald Institute for Archaeological Research, Cambridge: 5-58.
- MALÉCOT G., 1950. *Quelques schémas de probabilités sur la variabilité des population naturelles*. *Ann. Univ. Lyon Sci.*, A13: 37-60.
- MANTEL N., 1967. *The detection of disease clustering and a generalised regression model*. *Cancer Research*, 27: 209-220.
- MENOZZI P., PIAZZA A., CAVALLI-SFORZA L.L., 1978. *Synthetic maps of human gene frequencies in Europe*. *Science*, 201: 786-792.
- MOSCATI S., 1991. *I Celti*. Bompiani, Milano.
- PIAZZA A., 1993. *Who are the Europeans?* *Science*, 260: 1757-1759
- PIAZZA A., RENDINE S., MINCH E., MENOZZI P., MOUNTAIN J., CAVALLI-SFORZA L.L., 1995. *Genetics and the origin of the European languages*. *Proc. Natl. Acad. Sci. USA*, 92: 5836-5840.
- RENFREW C., 1987. *Archaeology and language: the puzzle of Indo-European origins*. Jonathan Cape, London.
- RENFREW C., 1998. *All the king's horses: assessing cognitive maps in later European prehistory*. In: S. MITHEN (ed.), *Creativity in Human Evolution and Prehistory*. Routledge, London: 260-284.
- RENFREW C., MCMAHON A., TRASK L. (eds.), 2000. *Time Depth in Historical Linguistics*. McDonald Institute for Archaeological Research, Cambridge, UK.
- RICHARDS M., CÔRTE-REAL H., FORSTER P., MACAULAY V., WILKINSON-HERBOTS H., DE-MAINE A., PAPIHA S., HEDGES R., BANDELT H.J., SYKES B., 1996. *Paleolithic and neolithic lineages in the European mitochondrial gene pool*. *Am. J. Hum. Genet.*, 59: 185-203.
- RICHARDS M., MACAULAY V., SIKES, B., PETTITT, P. HEDGES R., FORSTER P., BANDELT H.-J. 1997. *Reply to Cavalli-Sforza and Minch*. *Am. J. Hum. Genet.*, 61: 1251-254.
- RICHARDS M., OPPENHEIMER S., SYKES B., 1998. *mtDNA suggests Polynesian origins in Eastern Indonesia*. *Am. J. Hum. Genet.*, 63: 1234-1236.
- RICHARDS M., MACAULAY V., HICKEY E., VEGA E., SYKES B., GUIDA V., RENGO C. *et al.*, 2000. *Tracing European Founder Lineages in the Near Eastern mtDNA Pool*. *Am. J. Hum. Genet.*, 67: 1251-1276.
- RUHLEN M., 1987. *A Guide to the World's Languages*. Stanford University Press, Stanford (CA).



- SAITOU, N., NEI M., 1987. *The neighbor-joining method: a new method for reconstructing phylogenetic trees*. *Mol. Biol. Evol.*, 4(4): 406-425.
- SANKOFF D., 1973. *Mathematical developments in lexicostatistical theory*. In: T. SEBEEK (ed.), *Current Trends in Linguistics*. Mouton, The Hague, 11: 93-113.
- SCHLEICHER A., 1863. *Die Darwinische Theorie und die Sprachwissenschaft*. H. Böhlau, Weimar.
- SCHMIDT J., 1872. *Die Verwandtschaftsverhältnisse der Indogermanischen Sprachen*. H. Böhlau, Weimar.
- SCHÖNEMANN P.H., CARROL R.M., 1970. *Fitting one matrix to another under choice of a central dilatation and a rigid motion*. *Psychometrika*, 35: 245-255.
- SEIELSTAD M.T., MINCH E., CAVALLI-SFORZA L.L., 1998. *Genetic evidence for a higher female migration rate in humans*. *Nat. Genet.*, 20: 278-280.
- SEMINO O., PASSARINO G., BREGA A., FELLOUS M., SANTACHIARA-BENERECETTI A.S., 1996. *A view of the Neolithic demic diffusion in Europe through two Y-chromosome-specific markers*. *Am. J. Amer. Genet.*, 59:964-968.
- SEMINO O., PASSARINO G., OEFNER P.J., LIN A.A., ARBUZOVA S., BECKMAN L.E., DE BENEDICTIS G., FRANCALACCI P., KOUVATSI A., LIMBORSKA S., MARCIKIAE M., MIKA A., MIKA B., PRIMORAC D., SANTACHIARA-BANERECETTI A.S., CAVALLI-SFORZA L.L., UNDERHILL P.A., 2000. *The Genetic legacy of paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective*. *Science*, 290: 1155-1159.
- SOKAL R.R., MICHENER C.D., 1958. *A statistical method for evaluating systematic relationship*. *Univ. Kansas Sci. Bull.*, 38: 1409-1438.
- SOKAL R.R., MENOZZI P., 1982. *Spatial autocorrelations of HLA frequencies in Europe support demic diffusion of early farmers*. *Am. Nat.*, 119: 1-17.
- SOKAL R.R., ODEN N.L., WILSON C., 1991. *Genetic evidence for the spread of agriculture in Europe by demic diffusion*. *Nature*, 351: 143-145.
- STAROSTIN S., 2000. *Comparative-historical linguistics and lexicostatistics*. In: C. RENFREW, A. MCMAHON, L. TRASK (eds.). *Time Depth in Historical Linguistics*. McDonald Institute for Archaeological Research, Cambridge (UK), 1: 223-259.
- STURTEVANT E.H., 1962. *The Indo-Hittite Hypothesis*. *Language*, 38: 105-110.
- SYKES B., 1999 *The molecular genetics of European ancestry*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 354: 131-138.
- SYKES B., LEIBOFF A., LOW-BEER J., TETZNER S., RICHARDS M., 1995. *The origins of the Polynesians: an interpretation from mitochondrial lineage analysis*. *Am. J. Hum. Genet.*, 57: 1463-1475.
- SWADESH M., 1952. *Lexico-statistic dating of prehistoric ethnic contacts*. *Proceedings of the American Philosophical Society*, 96: 452-463.
- SWADESH M., 1955. *Towards greater accuracy in lexicostatistical dating*. *International Journal of American Linguistics*, 21: 121-137.
- WARNOW T., 1997. *Mathematical approaches to comparative linguistics*. *Proc. Natl. Acad. Sci. USA*, 94: 6585-6590.