

DON'T MISS

Contadino: scarpe grosse... e modello fino Posted 1 month ago

Search site

[Home](#) > [Divulgazione](#) > [Focus](#) > Le origini del popolo malgascio, alcune certezze e qualche mistero

## Le origini del popolo malgascio, alcune certezze e qualche mistero

On July 18, 2017

Con imbarcazioni simili a quelle della foto qui a lato (a bilanciere dell'etnia Vezo in Madagascar) ma molto più grandi, gli antenati dei malgasci arrivarono dall'Indonesia. Ma chi erano veramente questi antenati? E quando sono arrivati? Maurizio Serva ci racconta come, con l'uso di alcuni algoritmi lessicostatistici, sia possibile fare un po' di luce su questi misteri.



**Preambolo.** Se un giorno vi dovesse capitare di fare un viaggio in Madagascar, vi accorgereste che i malgasci hanno un aspetto che può variare da quello di un asiatico del sud-est a quello di un africano. In effetti il patrimonio genetico della popolazione ha una composizione che è per il 50/70% africana e per il 30/50% indonesiana. A questa eterogeneità genetica corrisponde una sostanziale omogeneità linguistica, infatti tutti i dialetti del malgascio appartengono al gruppo delle lingue indonesiane occidentali della famiglia linguistica austroasiatica. Questa famiglia comprende oltre 1200 lingue tra le quali il tagalog, il malese, le lingue indigene di Taiwan e le lingue polinesiane come il maori, l'hawaiano e la lingua rapanui. In particolare c'è una parentela molto stretta tra il malgascio e la lingua Maanyan del Borneo sud-orientale mentre la componente genetica di origine indonesiana sembra indicare una maggiore affinità con le popolazioni Banjar, anch'esse del Borneo sud-orientale. Anche i tratti più rilevanti della cultura malgascia come i culti tradizionali, le ceremonie funebri, l'architettura delle abitazioni e gli strumenti musicali possono essere fatti risalire ad antenati indonesiani.

C'è quindi un consenso generale sul fatto che il nucleo fondante della popolazione malgascia provenga dal Borneo, mentre la componente africana sembra essersi aggiunta in un secondo tempo. Un grande mistero però sta nel fatto che le popolazioni Maanyan e Banjar che vivono nel Borneo sud-orientale non possiedono le conoscenze e la tecnologia per una lunga navigazione marittima dall'Indonesia al Madagascar e quindi ci si chiede come possano aver intrapreso questo viaggio. Una possibile spiegazione è che gli antenati dei malgasci costituivano la ciurma borneana di pochi esperti navigatori malesi e in effetti nel malgascio resta traccia di vocaboli malesi per i manufatti e le attività relative alla navigazione.

Accanto a questo quadro storico, sul quale c'è una sostanziale unanimità tra gli esperti, ci sono altri aspetti della storia del popolo malgascio e del suo stanziamento in Madagascar che sono controversi. Ad esempio sembra che i marinai indonesiani raggiunsero il Madagascar con una navigazione dall'Indonesia tra i mille e i duemila anni fa, ma oltre alla data anche il percorso e il luogo dello sbarco sono ancora dibattuti. Come anche si discute ancora sulla composizione etnica di questi colonizzatori, dato che nel malgascio ci sono deboli tracce lessicali di lingue parlate in Sulawesi, Java, Sumatra, Malesia e forse anche nelle Filippine.

Con il nostro lavoro [1] JM. Serva, F. Petroni, D. Volchenkov & S. Wichmann: *Malagasy dialects and the peopling of Madagascar*, Journal of the Royal Society Interface 9, 54-67, (2012). abbiamo cercato di risolvere questi problemi insieme ad altri, come ad esempio la classificazione interna delle varianti locali della lingua. Il lavoro si basa su uno studio lessicostatistico dei dialetti malgasci, però, prima di poter discutere i risultati trovati, bisogna fare un passo indietro e parlare del metodo che abbiamo utilizzato.

**Distanze.** Le lingue evolvono nel tempo come organismi: si riproducono (lingue romanze dal latino), mutano (italiano moderno da quello medioevale) e si estinguono (dalmatico nel 1900). È quindi possibile, in linea di principio, assegnare una distanza per ogni coppia di lingue in analogia con la distanza genetica tra due organismi. Supponendo che il vocabolario sia l'analogo del DNA, le distanze possono essere valutate dalle differenze lessicali.

Il metodo utilizzato dalla lessicostatistica, sviluppato da Morris Swadesh negli anni '50, misura le distanze usando la percentuale di parole imparate nelle due lingue [2 ]M. Swadesh:

ISCRIVITI A ALLA MADD-LETTER E RICEVI TUTTI GLI AGGIORNAMENTI

Inserisci la tua Email:

Iscriviti

Visita questo gruppo google

Partecipa al nostro gruppo Facebook

 Guarda l'ultima Madd-Letter [del 26 luglio 2017](#)

Guarda tutte le Madd-Letters

È USCITO IL NUMERO 2/2017 DI ARCHIMEDE

Archimede



COMICS&amp;SCIENCE - THE BABBAGE ISSUE



ARTICOLI RECENTI SU MADDMATHS!

 Archimedia 2/2017: Un kebab con Khayyām  
 25/07/2017

Doppio oro e diciottesimo posto per l'Italia alle IMO 2017 24/07/2017

Risposte ai problemi di Archimede EUREKA del numero 1/2017 24/07/2017

Sui boomerang, triangoli e la curvatura dell'universo 23/07/2017

Il bizzarro intreccio di topologia e fisica dei materiali 21/07/2017

3D lo spazio: il teatro geometrico 21/07/2017

Le origini del popolo malgascio, alcune certezze e qualche mistero 18/07/2017

Le due teste del tiranno, un libro di Marco Malvaldi (recensione) 16/07/2017

Morta Maryam Mirzakhani, prima donna a vincere una Medaglia Fields 15/07/2017

Data Scientist...uno, nessuno, centomila: breve tutorial linguaggio R 13/07/2017

NEWS



Il bizzarro intreccio di topologia e fisica dei materiali

*Lexicostatistic dating of prehistoric ethnic contacts*, Proceedings of the American Philosophical Society 96, 452-463, (1952). Le parole imparentate sono quelle con stessa origine storica e uguale significato (ad esempio *homme* e *uomo*). Il punto debole di questo metodo è quello del giudizio soggettivo. Ad esempio, la parola spagnola *leche* e la parola greca *gala* sono imparentate (dal latino *lac/lactis* e dal greco antico *gala/galactos*), ma bisogna conoscerne la storia altrimenti la parentela che viene stabilita da alcuni esperti non è riconosciuta da altri. Questa discordanza si verifica facilmente per le lingue la cui storia è poco nota come quelle australiane o Khoi-San.

In verità l'idea di misurare la distanza tra lingue sembra essere molto più antica (ma dimenticata). Infatti l'esploratore francese Dumont D'Urville, durante il suo viaggio a bordo dell'Astrolabe dal 1826 al 1829, raccolse liste di parole di varie lingue e nel suo lavoro sulla divisione geografica del Pacifico [3] D. D'Urville: Sur les îles du Grand Ocean, Bulletin de la Société de Géographie 17, 1-21, (1832). propose un metodo abbastanza simile a quello della lessicostatistica moderna.

La nostra strategia automatizzata [4] M. Serva, F. Petroni: *Indo-European languages tree by Levenshtein distance*, EuroPhysics Letters 81, 68005 (2008). ha il vantaggio di evitare la soggettività. Anche noi consideriamo coppie di parole con lo stesso significato nelle due lingue, ma invece di stabilire se sono imparentate, ne misuriamo la distanza di Levenshtein e la normalizziamo dividendo per il numero di caratteri della più lunga delle due parole.

La distanza di Levenshtein tra due parole è semplicemente il numero minimo di inserzioni, cancellazioni o sostituzioni di un singolo carattere che sono necessarie per trasformare una parola nell'altra. Ad esempio, se consideriamo le due parole *homme* e *uomo* si vede facilmente che la loro distanza di Levenshtein è 3, infatti:

*homme* → *uomme* → *uome* → *uomo*

mentre la normalizzazione si ottiene dividendo per 5, che è il numero di caratteri della più lunga delle due parole. Quindi la distanza di Levenshtein normalizzata è  $3/5 = 0,6$ .

Vediamo la cosa in simboli: sia  $\alpha_i$  la parola corrispondente al significato  $i$  nella lingua  $\alpha$  e  $\beta_i$  la parola corrispondente allo stesso significato  $i$  nella lingua  $\beta$ ; la distanza di Levenshtein normalizzata tra le due parole è

$$D(\alpha_i, \beta_i) = \frac{D_L(\alpha_i, \beta_i)}{L(\alpha_i, \beta_i)}$$

dove  $D_L(\alpha_i, \beta_i)$  è la distanza di Levenshtein e  $L(\alpha_i, \beta_i)$  è il numero di caratteri della più lunga delle due parole. Il risultato è un numero compreso tra 0 (parole identiche) e 1 (massima differenza possibile).

Infine la distanza tra le due lingue  $\alpha$  e  $\beta$  si ottiene semplicemente mediando tra tutte le coppie  $\alpha_i, \beta_i$  con uguale significato  $i$  in una lista con  $M$  significati ( $i = 1, 2, \dots, M$ )

$$D(\alpha, \beta) = \frac{1}{M} \sum_{i=1}^M D(\alpha_i, \beta_i)$$

dove  $M$  tipicamente è 100 o 200. Anche la distanza tra due lingue  $D(\alpha, \beta)$  è un numero compreso tra 0 e 1. Se il numero di lingue è  $N$ , allora l'output è una matrice simmetrica  $N \times N$  con diagonale nulla, i cui  $\frac{N(N-1)}{2}$  elementi sopra (o sotto) la diagonale sono le distanze tra tutte le coppie di lingue.

**Sbarco.** Il dataset da noi raccolto all'inizio del 2010, costituito da un vocabolario di  $M = 200$  parole per  $N = 23$  dialetti, copre tutte le aree dell'isola ed è probabilmente la più grande collezione di liste comparative per i dialetti malgasci.

A partire da questi dati abbiamo calcolato la distanza tra tutte le  $\frac{N(N-1)}{2} = 253$  diverse coppie di dialetti malgasci, dalle quali ci aspettavamo di ottenere risultati relativi alle relazioni interne che a loro volta possono dare informazioni su data e modalità dell'insediamento. Abbiamo anche calcolato le distanze tra i dialetti malgasci e alcune lingue austronesiane (tra queste il Maanyan) che eventualmente possono fornire informazioni sulla composizione etnica dei fondatori indonesiani.

La matrice delle distanze  $D(\alpha, \beta)$  contiene tutte le informazioni ma queste non sono manifeste ed è quindi necessario estrarne la parte rilevante per renderle leggibili. L'approccio standard a questo problema è quello di trasformare la matrice in un albero filogenetico.



Marta Maryam Mirzakhani, prima donna a vincere una Medaglia Fields



Ecco come velocizzare le stampe 3d sfruttando la matematica



Contadino: scarpe grosse... e modello fino



I numeri sono innati o frutto della cultura?

## FAKE NEWS



La prova matematica dell'esistenza di Dio?



Ecco il "birra tour" perfetto tra i pub inglesi



Coda al supermercato? Ecco come scegliere la più veloce



Ricercatore italiano scambiato per terrorista a causa di un'equazione



McDonald e le combinazioni «infinte» degli ingredienti nei panini

## EVENTI



Matematica in Salone

"Scrivo in modo che si possa sempre vedere l'interno..."

Posted 2 months ago 0

## ULTIMI COMMENTI

Luca on Odifreddi risponde (con replica dai firmatari della lettera precedente)

Roberto Natalini on È uscito il numero 2/2017 di Archimede

enrico mijotto on È uscito il numero 2/2017 di Archimede

Privacy & Cookies Policy

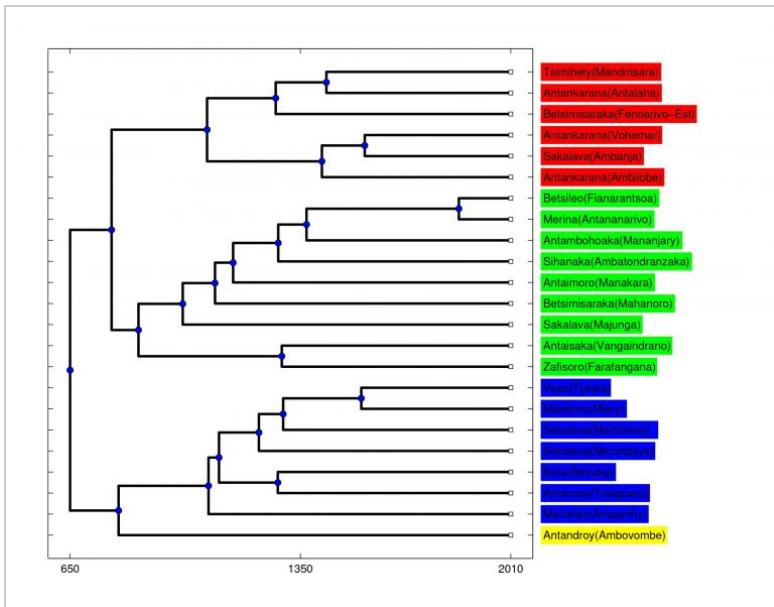


Figura 1. Albero filogenetico dei dialetti del Madagascar. Il tempo caratteristico è stato scelto in modo che la radice sia fissata nell'anno 650 dell'era volgare. La ripartizione filogenetica dei dialetti ha una corrispondenza geografica come può essere osservato comparando con la figura 2 (sinistra).

Per raggiungere questo obiettivo è utile (ma non obbligatorio) trasformare ogni distanza in un tempo di separazione (tempo dall'antenato comune). La regola fondamentale della glottocronologia assume che le differenze si accumulino con un tasso costante nel tempo  $\frac{1}{\tau}$ . Da questa sola ipotesi si ricava

$$T(\alpha, \beta) = -\tau \ln[1 - D(\alpha, \beta)]. \quad (1)$$

Come si vede, ad una distanza nulla corrisponde un tempo di separazione nullo mentre alla distanza massima ( $D(\alpha, \beta) = 1$ ) corrisponde un tempo di separazione infinito. Il valore del tempo caratteristico  $\tau$  può essere fissato se si conosce il tempo di separazione relativo anche a una sola tra tutte le coppie di lingue. Nel caso dei dialetti del malgascio questo non è possibile per cui, come vedremo, abbiamo dovuto usare un nuovo criterio per determinare  $\tau$ .

Un albero filogenetico può essere facilmente costruito utilizzando diversi algoritmi. Nella figura 1 mostriamo il risultato ottenuto utilizzando l'algoritmo UPGMA (Unweighted Pair Group Method with Arithmetic Mean).

Gli alberi sono onnipresenti nelle rappresentazioni delle tassonomie delle lingue, tuttavia non rivelano tutte le informazioni contenute nella matrice delle distanze. La ragione è che un albero non può rappresentare tutte le complesse interazioni tra lingue (prestiti lessicali, migrazioni interne ...). Abbiamo quindi introdotto un altro approccio di tipo geometrico (Analisi delle Componenti Strutturali). Questo approccio codifica le informazioni della matrice in uno spazio astratto  $n$ -dimensionale dove sono rappresentate le posizioni delle lingue.

Sarebbe veramente troppo lungo descrivere questa rappresentazione, ma la cosa interessante è che la varianza della componente radiale è linearmente associata al tempo che separa le lingue di una famiglia dalla protolingua. Abbiamo verificato che il tasso di crescita lineare della varianza è verosimilmente lo stesso per tutti i gruppi linguistici, per cui abbiamo potuto concludere che la protolingua malgascia ha iniziato a differenziarsi 1360 anni fa (questo ci permette anche di fissare il  $\tau$  nella formula (1)). I nostri risultati suggeriscono quindi che lo sbarco degli antenati della popolazione attuale sia avvenuto circa nell'anno 650 dell'era volgare. Questo risultato è stato successivamente confermato da altri autori con ricerche di tipo genetico.

Per determinare il centro della dispersione dei dialetti, e quindi il luogo dello sbarco abbiamo preso in prestito un'idea molto usata in biologia: il luogo di origine di una specie è quello con maggiore diversità biologica. Non descriviamo in dettaglio questo metodo, ma il risultato è che la lingua malgascia ha iniziato a differenziarsi nel sud-est dell'isola che è quindi anche il luogo di approdo dei fondatori indonesiani.

Una conferma indipendente di questo viene da un fatto ben noto. C'è infatti una corrente nell'Oceano Indiano che collega Sumatra con la costa sud-est del Madagascar. Quando il monte Krakatoa esplose nel 1883, la pomice fu trasportata sulla costa sud-orientale vicino a Mananjary. La stessa area vide l'arrivo di alcuni rottami delle navi che navigavano tra Java e Sumatra e che furono bombardate dalle forze aeree giapponesi nella seconda guerra mondiale.

Infine, per convincere il lettore della sensatezza dei nostri risultati abbiamo aggiunto la figura 2,

Giuseppe Ornaghi on #2 Una versione elementare della Congettura di Riemann

Sandra Lucente on Possibile che funzioni continue siano discontinue? E viceversa?

FABRY2 on C'è veramente un nuovo approccio alla Congettura di Riemann?

Giovanni Passalacqua on Possibile che funzioni continue siano discontinue? E viceversa?

Sandra Lucente on Possibile che funzioni continue siano discontinue? E viceversa?

Sandra Lucente on Possibile che funzioni continue siano discontinue? E viceversa?

G on Possibile che funzioni continue siano discontinue? E viceversa?

dove accanto alla geografia del Madagascar (sinistra) ne abbiamo ricostruita una fittizia (destra), basandoci solo sulla matrice delle distanze lessicali [5] M. Serva, D. Vergni, D. Volchenkov, A. Vulpiani: *Recovering geography from a matrix of genetic distances*, EuroPhysics Letters, 118, 48003 (2017).. Un po' come se un viaggiatore cercasse di risalire alla geografia del Madagascar sedendosi ad un tavolo per parlare di calcio o di rugby con persone provenienti da varie parti dell'isola.

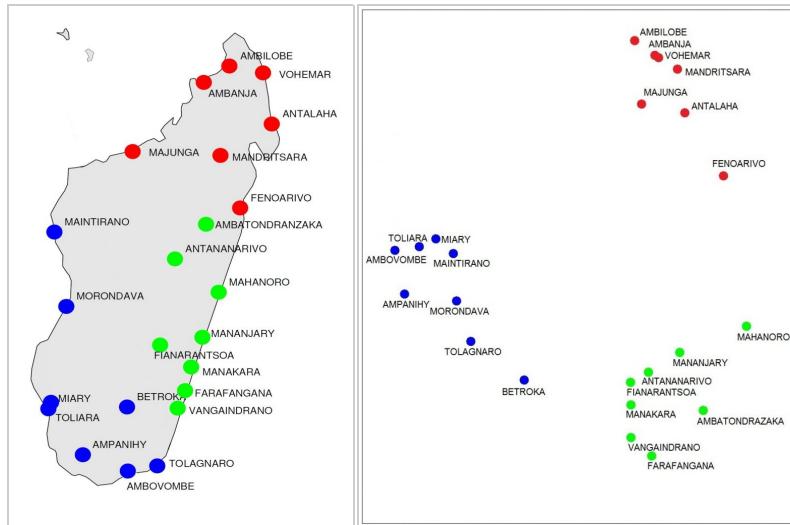


Figura 2. A sinistra la geografia del Madagascar, a destra una geografia ricostruita a partire da dati lessicali.

**Misteri.** Abbiamo visto come molte informazioni possono essere estratte da semplici dati lessicali, inclusi data e luogo di sbarco dei fondatori, tuttavia molti misteri ancora permangono.

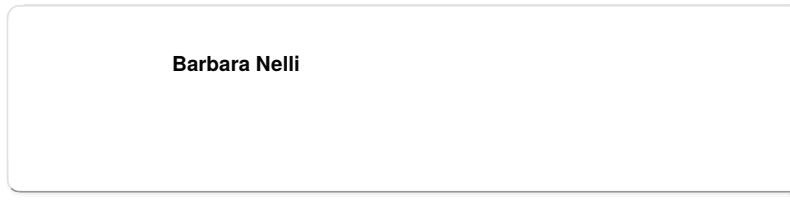
Ad esempio, usando le distanze dialetti malgasci/lingue austronesiane, siamo stati in grado di confermare che il malgascio ha affinità con altre lingue indonesiane e filippine, ma non siamo stati in grado di risalire alla composizione etnica dei fondatori. Ma il mistero più grande è quello relativo agli abitanti che hanno preceduto i fondatori. Il Madagascar era abitato prima dell'arrivo degli antenati degli odierni malgasci. Ma da chi? da indonesiani di una precedente migrazione? da Bantu? da popoli affini ai Khoi-San? o addirittura dai discendenti di una precoce emigrazione di Sapiens dall'Africa?

I malgasci si riferiscono a questo mitico popolo come i Vazimba. Abbiamo cercato di dare una risposta raccogliendo i dati linguistici relativi al popolo Mikea (l'unico popolo pre-agriculturale del Madagascar). La gente di questa etnia è da molti sospettata di essere in qualche modo imparentata con i Vazimba, ma dalla nostra analisi non è emerso niente di significativo. Non abbiamo però rinunciato, e questa storia è ancora aperta.

*Maurizio Serva*

Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica, Università dell'Aquila

**Barbara Nelli**



## Bibliografia

1. ↑ M. Serva, F. Petroni, D. Volchenkov & S. Wichmann: *Malagasy dialects and the peopling of Madagascar*, Journal of the Royal Society Interface 9, 54-67, (2012).
2. ↑ M. Swadesh: *Lexicostatistic dating of prehistoric ethnic contacts*, Proceedings of the American Philosophical Society 96, 452-463, (1952)
3. ↑ D. D'Urville: Sur les îles du Grand Ocean, Bulletin de la Société de Géographie 17, 1-21, (1832).
4. ↑ M. Serva, F. Petroni: *Indo-European languages tree by Levenshtein distance*, EuroPhysics Letters 81, 68005 (2008).
5. ↑ M. Serva, D. Vergni, D. Volchenkov, A. Vulpiani: *Recovering geography from a matrix of genetic distances*, EuroPhysics Letters, 118, 48003 (2017).