

UNIVERSITÀ DELLA CALABRIA



UNIVERSITÀ DELLA CALABRIA

Dipartimento di Biologia, Ecologia e Scienze della Terra

Dottorato di Ricerca in Scienze e Ingegneria dell'Ambiente,  
delle Costruzioni e dell'Energia

Ciclo  
XXIX

**Aspetti genetici dell'evoluzione  
linguistica nelle popolazioni umane:  
un'analisi comparativa e  
interdisciplinare**

Settore Scientifico Disciplinare BIO-18

Coordinatore: Chiar.mo Prof. Pietro Pantano

Firma Pietro Pantano

Tutor: Chiar.mo Prof. Giuseppe Passarino

Firma Giuseppe Passarino

Co-Tutor: Prof.ssa Serena Dato

Firma Serena Dato

Dottoranda: Dott.ssa Mara Vigna

Firma Mara Vigna

*“Nada podrá medir el poder que oculta una palabra. Contaremos sus letras, el tamaño que ocupa en un papel, los fonemas que articulamos con cada sílaba, su ritmo, tal vez averigüemos su edad; sin embargo, el espacio verdadero de las palabras, el que contiene su capacidad de seducción, se desarrolla en los lugares más espirituales, etéreos y livianos del ser humano.”*

Álex Grijelmo

La seducción de las palabras

## Abstract

This research is an interdisciplinary work, aimed at exploring the mutual contributions from genetics and linguistics to the innately interdisciplinary topic of human evolution. In brief, as for their genomes, humans are 99% identical to their closest relatives, the non-human primates, but in the remaining 1% lie the crucial differences that led our genus to distinguish itself from other branches in its phylogenetic tree. Among all traits pertaining to *Homo sapiens*, species-specific cognitive abilities clearly represent a step up the ladder of primate's evolution, and among these, language stands out for its exclusively human characteristic: namely, the ability to create and produce a communication system which is recursive and potentially unlimited, starting from a finite set of elements. But, which are the variations inside our DNA that allowed and shaped human evolution? Where are they located and on which mechanisms do they act? To shed light on these questions, we chose three genes (*HARI*, *FOXP2* and *ASPM*) directly associated to cognition, brain development and language acquisition, and analysed them through genomic methodology, such as cross-species comparison of homology and multiple alignment. By comparing the human version of these sequences to their non-human primate and non-primate mammalian homologues, we confirmed their high level of conservation, but also discovered specific and localized changes in their nucleotide sequences within the human genome, probably representing interesting sites to further search for distinctively human evolutionary traits.

On the other side, genetics and linguistics historically collaborated for the complex problem of the phylogenetic classification of people and languages around the world, and in particular, the last part of this thesis is dedicated to the long and fruitful work by Professor Luigi Luca Cavalli Sforza and his team, whose interdisciplinary approach has been the key for this research: by fusing theories, methods and models from disciplines apparently far from each other, we explored a rich and inspiring research topic, and, hopefully, added a further piece to the puzzle of the much wider and essentially interdisciplinary research field investigating the evolution of human being.

# Indice

<b>Introduzione</b> .....	<b>1</b>
<b>Capitolo I</b>	
<b>Le origini del linguaggio. Presupposti biologici per l'evoluzione del linguaggio tra le capacità cognitive umane.</b> .....	<b>3</b>
0. Introduzione .....	4
1. Presupposti biologici per lo sviluppo del linguaggio: quando emerge il linguaggio umano?.....	6
1.1 Le dimensioni del cervello umano.....	6
1.2 La corteccia cerebrale e le “aree del linguaggio” .....	8
2. Discussione e conclusioni.....	10
<b>Capitolo II</b>	
<b>Cosa ci rende umani. Analisi comparativa di geni coinvolti nello sviluppo cognitivo nel genoma dell'uomo e di primati non umani</b> .....	<b>12</b>
0. Introduzione .....	13
1. Comparare linguaggio e genoma: sistemi complessi in evoluzione.....	13
2. Il ruolo della bioinformatica e della genomica: alla ricerca di ciò che ci rende umani.....	16
2.1 La scoperta di HAR1 .....	18
2.2 Esseri parlanti: il ruolo di FOXP2.....	21
2.3 ASPM e lo sviluppo cerebrale.....	25
3. Analisi comparativa dei tre geni in primati e non-primati.....	26
4. Risultati .....	30
4.1 HAR1: Risultati dell'allineamento multiplo della sequenza nucleotidica .....	30
4.2 FOXP2: Risultati dell'allineamento multiplo della sequenza nucleotidica e aminoacidica .....	32
4.3 ASPM: Risultati dell'allineamento multiplo della sequenza nucleotidica e aminoacidica .....	38
5. Conclusioni e sviluppi futuri.....	45

### **Capitolo III**

#### **Evoluzione dei linguaggi, dei geni e delle popolazioni umane:**

<b>uno sguardo interdisciplinare.....</b>	<b>46</b>
0. Introduzione .....	47
1. Cultura e biologia: il ruolo della cultura nell'evoluzione .....	48
2. Linguistica e biologia: il ruolo del linguaggio nell'evoluzione .....	50
3. L'origine del linguaggio: monogenesi o filogenesi? .....	53
4. Il caso delle origini dei Nativi Americani.....	55
5. Conclusioni .....	62
<b>Conclusioni generali .....</b>	<b>63</b>
<b>Bibliografia.....</b>	<b>65</b>
<b>Appendice:</b>	
<b>Le reti del linguaggio nei processi cognitivi emozionali in</b>	
<b>Shakespeare.....</b>	<b>71</b>

# Introduzione

La ricerca presentata in questo studio è il frutto di un lavoro interdisciplinare volto ad esplorare i contributi reciproci della genetica e della linguistica allo studio dell'evoluzione umana, e si focalizza sulle origini del linguaggio inteso in senso stretto, come la capacità esclusivamente umana di produrre e utilizzare un sistema comunicativo ricorsivo e potenzialmente illimitato a partire da un insieme di elementi linguistici limitati.

Circa sei milioni di anni fa visse l'antenato comune tra uomo e scimpanzé, ed è intorno a quella data che è avvenuta la scissione filogenetica che ha portato l'evoluzione degli ominidi fino alla comparsa del genere *Homo*. È all'interno di questo ramo dell'evoluzione che sono ricercate le origini del linguaggio umano, con datazioni che ruotano intorno all'epoca della comparsa dell'*Homo habilis* o dell'*Homo erectus*. L'arco di tempo che va dall'*Homo habilis* ad oggi, in particolare, comprende alcuni degli stadi evolutivi imprescindibili al linguaggio, da quanto risulta dalle evidenze archeologiche e da una serie di caratteristiche biologiche necessarie ad un potenziale sviluppo del linguaggio. Una breve trattazione riguardo alcuni di questi prerequisiti biologici relativi principalmente all'evoluzione del volume cerebrale, del processo di encefalizzazione e della corteccia cerebrale è oggetto del primo capitolo di questa tesi.

Le capacità cognitive dell'uomo rappresentano uno stacco evolutivo piuttosto drastico nella catena dell'evoluzione dei primati, che ha fatto sì che l'uomo si distinguesse per caratteristiche non riscontrabili in nessuno dei suoi antenati, prima fra tutte, il linguaggio. Partendo dalla consapevolezza che il genoma umano è identico al 99% a quello degli altri primati, in questa ricerca ci siamo focalizzati su quelle regioni del genoma che possano contraddistinguere la nostra specie da organismi geneticamente simili, come gli altri primati, e questi rispetto ad altri più lontani lungo la scala evolutiva, dove siano localizzate e su quali meccanismi agiscano. In sintesi, quali sono i geni che ci rendono "umani"?

A tal fine, nel secondo capitolo si descrive lo studio genomico delle tre sequenze geniche, *HARIA*, *FOXP2* e *ASPM*, che sono state prescelte come oggetto di analisi per la loro diretta correlazione con la cognizione umana, l'evoluzione cerebrale e lo sviluppo del linguaggio. Le tre sequenze sono state comparate con sequenze omologhe in altri organismi (primati non umani e mammiferi non primati), con l'obiettivo di rivelare le variazioni significative nel genoma umano, potenzialmente causa del "salto cognitivo" avvenuto nel genere *Homo*.

La chiave di lettura dell'intero lavoro di ricerca che ha portato alla stesura di questa tesi è l'interdisciplinarietà. La fusione di metodi, modelli e approcci disciplinari apparentemente molto lontani tra loro, quali la linguistica e la genetica, incorporando informazioni cruciali anche dalla paleoantropologia, ha aperto un campo di ricerca fertile e ricco di prospettive. Uno dei modelli di ispirazione per questo tipo di "pozione" è stato sicuramente il lavoro decennale del gruppo di ricerca del Professore Luigi Luca Cavalli Sforza, che tuttora continua a sovrapporre genetica delle popolazioni, linguistica storica, antropologia e archeologia in una visione ampia dell'evoluzione umana che abbraccia campi e discipline diverse ma ugualmente necessarie. Il terzo capitolo di questa tesi è dedicato a questo approccio interdisciplinare, con un approfondimento sul lavoro di Cavalli Sforza e del suo team e sulle possibilità di collaborazione tra scienze umanistiche e scienze biologiche nello studio del genere umano.

Gli argomenti di questa tesi sono stati elaborati sotto forma di contributi in volume, per il quindicesimo numero della collana "Filosofia e Scienza" (Aracne ed.), come indicato nel riferimento bibliografico relativo al secondo e al terzo capitolo. Una precedente pubblicazione riguarda invece un filone di ricerca separato (ma altrettanto interdisciplinare), approfondito durante il primo anno di dottorato a partire dagli argomenti della tesi di laurea magistrale, riguardo l'analisi semantica di due tragedie shakespeariane attraverso l'uso della scienza delle reti, e riportato in appendice del presente lavoro.

## **Capitolo I**

**Le origini del linguaggio.  
Presupposti biologici per l'evoluzione del  
linguaggio tra le capacità cognitive umane.**

## 0. Introduzione

L'evoluzione del linguaggio nell'uomo è strettamente correlata all'evoluzione di alcuni elementi biologici essenziali al suo sviluppo: innanzi tutto, i cambiamenti nella dimensione e nella struttura del cervello, così come del tratto respiratorio, del tratto vocale e della laringe. Circa 2 milioni di anni fa, alcune specie umane (*Homo habilis* prima e *Homo erectus* dopo) erano probabilmente già dotate di un encefalo compatibile con lo sviluppo di una qualche forma rudimentale di linguaggio, seppur fortemente limitata dalle costrizioni fisiologiche legate al posizionamento della laringe e al controllo della respirazione; solo intorno a 400.000 anni fa, la struttura del tratto vocale degli ominidi ha probabilmente iniziato ad acquisire una forma tale da permettere, successivamente, lo sviluppo di una forma di linguaggio più complessa, con la comparsa dell'*Homo sapiens* circa 200.000 anni fa.

L'arco di tempo che va dall'*Homo habilis* ad oggi (il ramo più ampio dell'albero in Figura 1) è interessante per studiare l'origine del linguaggio umano, poiché comprende alcuni degli stadi evolutivi imprescindibili allo sviluppo di questa abilità cognitiva complessa e tipicamente umana, tra i quali, appunto, un aumento significativo delle dimensioni dell'encefalo e il riposizionamento del tratto vocale.

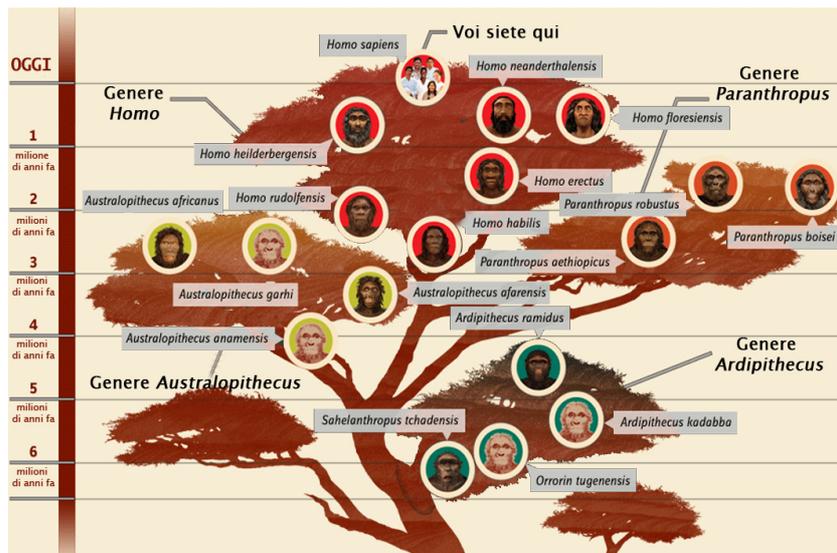


Figura 1: Albero filogenetico umano a partire dal punto di scissione uomo-scimpanzé (circa 6 milioni di anni fa). Immagine riadattata da: Smithsonian's National Museum of Natural History.

Le supposte capacità linguistiche degli ominidi precedenti all'*Homo sapiens* sono state studiate e discusse a fondo in letteratura, alla ricerca delle possibili origini del linguaggio: ad esempio, le abilità linguistiche correlate alla dimensione dell'encefalo e al controllo respiratorio nell'*Homo habilis* e nell'*Homo erectus* sono state studiate da Deacon (1997), Tobias (1998) e Wynn (1998), le proprietà del tratto respiratorio dell'*Homo neanderthalensis* sono state analizzate da Laitman e collaboratori (1996) e Franciscus (1999), e in alcuni studi, come in Lieberman (1992, 1993) e Lieberman & McCarthy (2007), sono state approfondite proprio le severe limitazioni morfologiche e fisiologiche specifiche dei Neanderthal in relazione alla possibilità di produzione linguistica. Anche studi più recenti hanno indagato la possibilità che qualche forma di linguaggio sia nata prima della comparsa dell'*Homo sapiens* intorno all'epoca tra *Homo habilis* e *Homo erectus*, soprattutto per quanto riguarda lo studio del legame che intercorre tra linguaggio verbale e non verbale, attraverso l'analisi dei pattern neuronali coinvolti nella produzione linguistica e degli "speech-related hand gestures" (Whishaw et al. 2010).

Il presente studio si concentrerà nello specifico sul ruolo del cervello e della sua evoluzione nel genere *Homo*, da un punto di vista interdisciplinare. Di seguito, al fine di analizzare alcune delle componenti biologiche ritenute necessarie per lo sviluppo del linguaggio e inquadrarlo all'interno dell'evoluzione degli ominidi, sono sintetizzati alcuni studi importanti relativi, nello specifico, alle dimensioni del cervello, all'evoluzione della corteccia cerebrale e di aree cerebrali più strettamente correlate al linguaggio.

# 1. Presupposti biologici per lo sviluppo del linguaggio: quando emerge il linguaggio umano?

## 1.1 Le dimensioni del cervello umano

Il volume cranico e le dimensioni dell'encefalo sono considerati variabili importanti per l'evoluzione del cervello umano, essendo tra le caratteristiche che contribuiscono a renderlo atto alla cognizione umana. Di conseguenza, sia nel campo della linguistica che in quelli della genetica e dell'antropologia, è stata prodotta molta letteratura riguardo la capacità cranica e le dimensioni cerebrali negli ominidi, allo scopo di definire le limitazioni che la biologia ha imposto sulle capacità cognitive dei primati e le possibili abilità cognitive degli stessi, con risultati spesso discordi.

Ciò che sappiamo con sufficiente certezza, è che il cervello umano è l'organo che ha subito l'evoluzione più rapida della quale siamo a conoscenza oggi, all'incirca triplicando il suo volume nel corso dell'evoluzione umana. La tabella riassuntiva delle dimensioni corporee e cerebrali dei vari membri del genere *Homo* (Tabella 1) mostra come dall'*Homo habilis* all'*Homo sapiens* il volume cerebrale sia aumentato da 552 cm<sup>3</sup> a 1355 cm<sup>3</sup>, dato molto significativo soprattutto se rapportato al peso corporeo, che invece non subisce la stessa intensità di crescita:

SPECIE	EPOCA STIMATA (milioni di anni fa)	DIMENSIONI CORPOREE (kg)	DIMENSIONI CEREBRALI (cm <sup>3</sup> )
<i>Homo sapiens</i>	0-0,2	53	1355
<i>H. neanderthalensis</i>	0,03-0,25	76	1512
<i>H. heidelbergensis</i>	0,3-1	62	1198
<i>H. erectus</i>	0,2-1,9	57	1016
<i>H. ergaster</i>	1,5-1,9	58	854
<i>H. rudolfensis</i>	1,8-2,4	-	752
<i>H. habilis</i>	1,6-2,3	34	552

Tabella 1: Dimensioni cerebrali e corporee nell'evoluzione del genere *Homo*, dall'*Homo habilis* all'*Homo sapiens*. Immagine rielaborata da: Enciclopedia della scienza e della tecnica Treccani.

La correlazione con le dimensioni corporee è necessaria per sottolineare che i nostri antenati (e noi stessi) non erano caratterizzati “soltanto” da cervelli più

grandi, ma anche da un alto Quoziente di Encefalizzazione (QE), una misurazione generalmente collegata all'intelligenza e ottenuta dal rapporto tra la massa cerebrale effettiva e quella stimata in proporzione alla massa corporea (Klein 2016). Tale aumento delle dimensioni cerebrali è significativo rispetto al peso corporeo anche nel confronto con gli altri mammiferi, come indicato nel grafico in Figura 2.

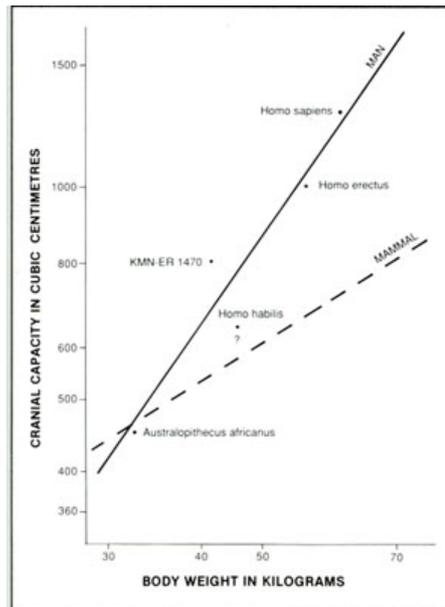


Figura 2: Rapporto tra dimensioni del cervello e dimensioni corporee di fossili del genere *Homo* rispetto allo stesso rapporto nei mammiferi: si nota l'aumento notevole della capacità cranica della nostra specie a parità di peso corporeo, rispetto agli altri mammiferi. Immagine modificata da: <https://answersingenesis.org/human-evolution/cranial-capacity-and-endocranial-casts/>.

La dimensione del cervello degli ominidi è andata incontro ad un aumento significativo rispetto alle dimensioni del corpo durante il Pleistocene, tra 2 milioni e 300,000 anni fa (Boyd & Silk 2000; Deacon 1997), in corrispondenza con la comparsa dell'*Homo habilis*. Di conseguenza, secondo molti studiosi, tra i quali il già citato Deacon (1997), le prime abilità di linguaggio umano, sarebbero da imputare proprio all'*Homo habilis*, il quale, sulla base dei ritrovamenti fossili, sarebbe già stato dotato di un cervello potenzialmente in grado di produrre una forma rudimentale di linguaggio. Più nello specifico, il cervello dell'*Homo habilis* presentava già alcune caratteristiche "umane"; oltre all'ingrandimento stesso, anche il tipico pattern dei solchi cerebrali, le caratteristiche fessure che delimitano le circonvoluzioni cerebrali, l'asimmetria

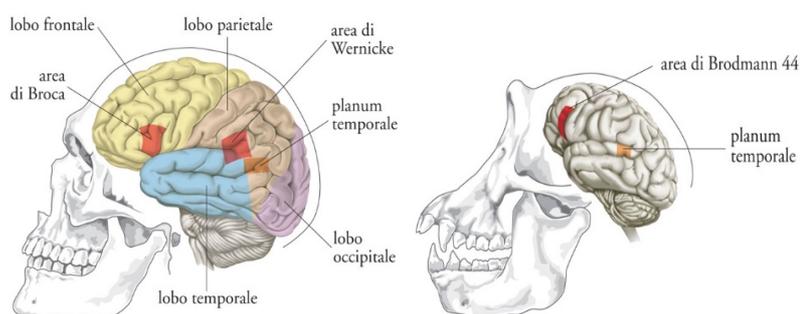
tra i due emisferi e la presenza delle aree cerebrali fisicamente corrispondenti a quelle di Wernike e di Broca, notoriamente correlate alla produzione e comprensione linguistica (Tobias 1998), che saranno approfondite nel paragrafo seguente.

### *1.2 La corteccia cerebrale e le “aree del linguaggio”*

Sebbene la dimensione del cervello umano sia un elemento interessante per uno studio dello sviluppo cognitivo, questo aumento volumetrico da solo non dà sufficiente contezza delle capacità cognitive umane, le cui origini sono ben lontane dall'essere completamente spiegate e descritte. Un elemento che più nel dettaglio fornisce informazioni sui meccanismi cerebrali alla base del linguaggio è la corteccia cerebrale, perché ospita due specifiche aree che sono state direttamente collegate alla processazione linguistica: l'area di Broca, sul lobo frontale, e l'area di Wernicke, sul lobo temporale, associate rispettivamente alla produzione e alla comprensione linguistica.

Gli studi di neuroimaging hanno dimostrato che le due aree, più che dividersi nettamente i compiti, collaborano attivamente alla processazione del linguaggio, coordinando il controllo motorio necessario alla vocalizzazione (Rosenfield 2013), e che partecipano all'abilità linguistica insieme a diverse altre aree cerebrali egualmente necessarie (Binder 2015). Nonostante ciò, mantengono una posizione privilegiata per lo studio dell'evoluzione del linguaggio, per via delle impronte che i cappucci di queste due aree corticali lasciano all'interno della superficie cranica, rilevabili poi dagli studi paleoantropologici tramite i calchi endocranici (detti “endocast”). Per la sua conformazione, è l'area di Broca la più studiata dai paleoantropologi, perché l'impronta ricostruita tramite gli endocast è più affidabile di quella del cappuccio di Wernicke. Entrambe le aree sono oggetto di studio della neurogenetica e neurolinguistica, e con l'aiuto dei ritrovamenti fossili analizzati dalla paleoantropologia, si è in grado oggi di affermare che queste aree erano presenti già a partire dall'*Homo habilis* (Tobias 1987) e addirittura precedentemente nella linea evolutiva degli ominidi (Spoceter et al. 2010),

chiarendo dunque che non esiste una relazione univoca tra la presenza di queste due aree cerebrali e lo sviluppo del linguaggio umano, ma che queste aree svolgono funzioni più ampiamente correlate alla comunicazione non verbale. In Figura 3, alcune aree cerebrali di interesse per lo studio delle funzioni cognitive superiori umane sono comparate con aree corrispondenti nel cervello dello scimpanzé. In particolare, l'area di Broca (e l'area di Brodmann 44 che ne fa parte), l'area di Wernicke e il planum temporale.



**Figura 3: Comparazione di alcune aree cerebrali di interesse nel cervello dell'uomo e dello scimpanzé. Immagine rielaborata da: Enciclopedia della scienza e della tecnica Treccani.**

In particolare per l'area di Broca, è evidente un ruolo nel controllo motorio della mano e della bocca e nella comprensione dei comportamenti sociali, grazie alla funzione empatica svolta dai neuroni specchio, la cui presenza è cospicua in quest'area (Rizzolatti & Sinigaglia 2006), sottolineando come queste aree si siano probabilmente molto sviluppate perché il loro ruolo nel miglioramento della comunicazione apportava un chiaro vantaggio evolutivo per lo scambio di informazioni vitali all'interno di un quadro di comunicazione sociale. Le evidenze raccolte e analizzate finora, dunque, non sono sufficienti ad affermare con certezza che le origini del linguaggio umano siano databili a cavallo tra l'*Homo habilis* e l'*Homo erectus*, che rappresentano probabilmente, a livello linguistico e cognitivo in generale, una sorta di fase intermedia tra lo stadio dei primati non umani e quello più vicino, dell'*Homo neanderthalensis*. I soli dati fossili, pur fornendo numerosi indizi sulle dimensioni cerebrali e sulla struttura del cervello negli ominidi antecedenti all'*Homo sapiens*, non forniscono risposte definitive sulle funzioni svolte neppure da quelle aree che

possiamo definire “moderne” e “umane” nella loro forma. Utilizzando le parole di Wynn (1998) riguardo le due famose aree di Wernike e di Broca nell’*Homo erectus*: “we can safely conclude that those general areas of the *Homo erectus* brain are modern in shape, but must be more cautious about identifying the presence of speech production and comprehension areas”.

## **2. Discussione e conclusioni**

Non esiste al momento una risposta netta sull’emergere di tale capacità cognitiva, un tempo esatto prima o dopo del quale la si possa localizzare in maniera netta, così come non esiste un’area o un gene del linguaggio. Il cervello è infatti un complesso sistema dinamico e interconnesso, nel quale l’interazione è la chiave, al di là del volume fisico delle parti che lo costituiscono. Le evidenze sull’attività cerebrale dimostrano con sempre maggiore chiarezza e precisione che le funzioni svolte dal cervello sono il frutto della sua dinamicità e quindi dell’attivazione simultanea di processi complessi, non limitati esclusivamente ad una specifica area deputata ad una specifica funzione.

Il focus della ricerca neuroscientifica delle ultime decadi è stato appunto la propensione interna del sistema cervello alla complessità, spostando l’attenzione dallo studio specializzato di singole aree (con le loro specifiche funzioni associate) ad un’analisi su ampia scala dei meccanismi mente-cervello (Bassett & Gazzaniga 2011). In questa luce, lo studio del cervello come rete interconnessa, ha prodotto risultati significativi per quanto concerne le sue connessioni strutturali, mappandole nel cosiddetto Connettoma (Sporns 2011), e correlandole poi alle specifiche funzioni svolte. Così, l’anatomia dei circuiti cerebrali in cervelli sani e danneggiati ha fornito risposte importanti sul generale funzionamento di questo organo e sulla sua interrelazione con la cognizione e il comportamento umano (Honeya, Thiviergeb & Sporns, 2010).

Una trattazione estesa di questi argomenti non è necessaria in questo studio, ma questo breve appunto ha lo scopo di sottolineare che i prerequisiti biologici

sopracitati rappresentano solo alcuni degli aspetti dell'evoluzione umana, e che questi hanno progressivamente generato le abilità cognitive di cui l'uomo è capace, in concomitanza con altri fattori altrettanto importanti.

## Capitolo II

### **Cosa ci rende umani. Analisi comparativa di geni coinvolti nello sviluppo cognitivo nel genoma dell'uomo e di primati non umani**

Il contenuto del seguente capitolo è stato pubblicato come contributo in volume per la Collana “Filosofia e Scienza”, Aracne editrice, come indicato:

VIGNA M., DATO S., PASSARINO G. (2017), *Cosa ci rende umani. Analisi comparativa di geni coinvolti nello sviluppo cognitivo nel genoma dell'uomo e di primati non umani*, in BRUNI A.C., COLONNELLO P., DATO S., PASSARINO G. (a c. di) (2017), *Interdisciplinarietà, complessità e questioni epistemiche*, “Filosofia e Scienza”, 15, Aracne editrice, pp. 21 - 40.

## **0. Introduzione**

Tra tutte le caratteristiche umane, le capacità cognitive rappresentano evidentemente un grande salto in avanti nella catena dell'evoluzione, una sorta di "rivoluzione cognitiva" che ha fatto sì che, in un lasso di tempo relativamente breve, se considerato nell'ottica dell'evoluzione delle specie, l'uomo si distinguesse in maniera così drastica dai suoi più prossimi antenati. Il linguaggio umano è la manifestazione più evidente della peculiarità delle facoltà cognitive dell'uomo e si distacca fortemente da tutti i sistemi comunicativi sviluppati dalle altre specie. Ma quali sono i suoi correlati biologici? In quest'ottica, il genoma può essere considerato la chiave d'interpretazione delle piccole e grandi "novità" evolucionistiche che ci rendono umani e tra le quali spicca, appunto, il linguaggio.

L'incontro tra discipline apparentemente distanti come la genetica e la linguistica, genera interessanti spunti di ricerca per entrambi i campi. La letteratura riporta come attraverso la comparazione di specifici segmenti del genoma umano con quello delle scimmie antropomorfe e di altri mammiferi non-scimmie sia possibile ritrovare indizi sul genoma di eventi che hanno permesso di distinguere l'uomo dai suoi antenati non umani. Il punto di partenza di questi studi è il gap sorprendentemente piccolo che differenzia il genoma umano da quello delle scimmie, un gap dalle dimensioni non proporzionali alle effettive differenze psicofisiche che ci rendono umani.

## **1. Comparare linguaggio e genoma: sistemi complessi in evoluzione**

Genoma e linguaggio sono entrambi caratterizzati da interazioni dinamiche e multi-scala tra i loro elementi costitutivi e, anche, da processi evolutivi intrinseci. Entrambi sono costituiti da unità minime che interagiscono in strutture gerarchiche di complessità crescente: nella struttura del DNA come in quella del linguaggio, si rileva la proprietà combinatoria nella formazione di stringhe di informazione a partire da singole unità o elementi (proprietà chiave,

ad esempio, nelle teorie dei modelli connessionisti di percezione e produzione del linguaggio, come quello descritto da McClelland & Elman 1986).

Oltre ai meccanismi combinatori, i due sistemi condividono ad esempio la caratteristica crescita della complessità, fenomeno intrinseco dei sistemi evolutivi, biologici e non. L'evoluzione nei sistemi complessi è caratterizzata solitamente da un aumento della complessità del sistema stesso, sia esso biologico o non biologico, che può consistere sia in un aumento degli elementi costitutivi del sistema che in un aumento dei meccanismi di interazione tra essi. Questo meccanismo di crescita può essere riassunto in tre fasi principali:

*i.* Durante un primo step nell'evoluzione sia dei genomi che delle lingue, la complessità crescente è conseguita attraverso un aumento della quantità di elementi costitutivi dei sistemi, ovvero quelli che hanno la funzione fondamentale di veicolare informazione. Per la biologia, si tratta di geni, mentre per la linguistica, di sostanti e verbi (le cosiddette "parole piene"). La lingua cinese si presta bene come esempio per via della sua lunga storia evolutiva di più di 6000 anni e per il grande numero di caratteri che la compongono. Nel 2004, Zhang ha descritto come da un primo stadio della lingua scritta (le testimonianze incise sugli ossi oracolari), composto da circa 5000 caratteri, il cinese antico si sia arricchito fino a raggiungere circa 47000 caratteri, al fine di aumentare le proprie capacità descrittive (ad esempio, raggiungendo fino a 70 caratteri per distinguere i vari tipi di cavallo). Statisticamente, è stato fissato a 22000 il numero minimo di caratteri da conoscere per raggiungere una copertura del 99,99% del cinese antico (Zhang 2004). All'interno dell'evoluzione dei genomi si osserva la stessa tendenza: gli organismi più antichi e semplici, come i procarioti, compiono le funzioni necessarie alla sopravvivenza e alla proliferazione con un corredo genetico di circa qualche migliaia di geni, mentre organismi più complessi arrivano a presentarne fino ad un ordine di grandezza in più, circa 50000 nel caso del mais (Messing et al. 2004).

*ii.* In una seconda fase evolutiva, in entrambi i sistemi si è registrato un riarrangiamento interno degli elementi costitutivi, volto comunque ad una crescente complessità globale. In questa fase, il cambiamento non implica un

aumento quantitativo ma una risistemazione di elementi già esistenti, e può addirittura caratterizzarsi per una diminuzione del numero di blocchi costitutivi del sistema: il cinese moderno ha così snellito il suo corredo lessicale, necessitando di circa 5000 caratteri per coprire efficacemente il 99,99% del suo patrimonio semantico (contro i 22000 del cinese antico). Rispetto alla lingua antica, i suoi elementi più usati sono stati combinati insieme per formare concetti nuovi in maniera più economica. L'evoluzione dei genomi presenta un simile processo di sintesi funzionale: i mammiferi, come il topo e l'*Homo sapiens*, hanno all'incirca 25000 geni, molti di meno rispetto a organismi meno complessi, come ad esempio il riso, che ha 43000 geni o il mais, 59000, dimostrando dunque che la quantità degli elementi veicolanti informazioni non è indice di una maggiore o minore complessità dell'organismo stesso.

*iii.* Il terzo step evolutivo vede come protagonisti non gli elementi veicolo di informazione, cioè i geni o le “parole piene”, ma gli elementi regolatori e modificatori, elementi cosiddetti “virtuali”. Nei linguaggi, gli avverbi, le preposizioni o le congiunzioni sono considerate “parole vuote”, perché non veicolanti un significato di per sé, ma comunque necessarie alla costruzione del significato. Nei genomi, un paragone può essere fatto con il ruolo svolto dall'RNA non codificante. Tornando all'esempio descritto finora, mentre il cinese antico presenta pochissimi elementi regolatori, questi sono invece molto più utilizzati nel cinese moderno, che presenta meno caratteri, ma organizzati in maniera più economica. Allo stesso modo, l'espressione dei geni nei mammiferi è regolata dall'attività di RNA, talora molto piccoli, con funzione di regolazione genica.

In conclusione, in entrambi i casi si nota la tendenza al miglioramento delle performance dei sistemi economizzando gli elementi che li compongono e ottimizzandone le funzioni. Tale prospettiva interdisciplinare è utile a dimostrare quanto la contaminazione tra scienze possa arricchire la ricerca con punti di vista alternativi a quelli classici. Il codice genetico si presenta come un linguaggio biologico che codifica la vita nelle sue varie forme e le differenzia attraverso le caratteristiche fenotipiche. Il linguaggio umano, insieme alle altre

facoltà cognitive superiori, è una di quelle che in maniera più evidente distingue l'uomo dagli altri organismi nella scala evolutiva.

## **2. Il ruolo della bioinformatica e della genomica: alla ricerca di ciò che ci rende umani**

L'evoluzione è un cambiamento genetico che ha luogo nel tempo. Perché si verifichi, prima di tutto è necessario che abbiano luogo delle variazioni genetiche e in seguito che le forze evolutive modifichino nel corso del tempo la proporzione delle varianti genetiche: la variazione genetica è perciò la base di tutti i cambiamenti evolutivi. Le sequenze di nucleotidi che compongono il DNA di tutti gli esseri viventi contengono la chiave di lettura dell'evoluzione dei vari organismi, delle loro peculiarità e dei loro comportamenti: una sorta di libro della vita, contenente le istruzioni necessarie allo sviluppo di quel sistema complesso che è un organismo vivente. È in quelle sequenze di nucleotidi che risiedono le differenze intra e inter-specie, definendo e guidando i processi evolutivi che hanno dato un'impronta allo sviluppo della vita stessa.

Il paragone tra i vari genomi d'interesse è reso possibile dagli sforzi della comunità scientifica mondiale, che nel corso degli anni si è focalizzata sul sequenziamento del genoma di molteplici organismi, fino ad arrivare, tramite le ricerche di progetti internazionali come lo Human Genome Project<sup>1</sup> (HGP), a mappare l'intero genoma umano. Da qui, il lavoro di comparazione tra genomi è inevitabilmente diventato un campo di ricerca ambito per tutti coloro che sono interessati allo studio delle origini dell'uomo, rappresentando, inoltre, una fonte preziosa di informazioni per ricercatori provenienti da discipline e backgrounds diversi tra loro. La genetica, e nello specifico la genomica, ha sia offerto risposte e conferme che aperto nuove linee di ricerca a campi anche molto distanti dalle scienze biologiche, come l'antropologia, la linguistica, la psicologia cognitiva o le neuroscienze, fra gli altri.

---

<sup>1</sup> Tutte le informazioni riguardo HGP sono disponibili sul sito del National Human Genome Research Institute: <https://www.genome.gov/10001772/all-about-the-human-genome-project-hgp/>

Il sequenziamento dei vari genomi ha permesso sia l'allineamento puntuale di sequenze provenienti da specifici geni che dell'intero genoma di organismi differenti, definendo la percentuale di variabilità che differenzia ogni essere vivente. Ciò che sorprende è che, quantitativamente, questa percentuale di variabilità non è enorme. Per fare un esempio, uno degli organismi più studiati nel paragone con l'uomo è il *Pan troglodytes*, uno scimpanzé, ovvero il più vicino tra gli antenati dell'uomo tuttora viventi, molto più simili all'uomo rispetto ad altri mammiferi ma comunque sufficientemente distanti dal punto di vista cognitivo. Il genoma del *Pan troglodytes* e dell'*Homo sapiens* corrispondono per il 99%, ciò significa che il cuore delle differenze tra noi ed uno scimpanzé risiede all'interno di una piccolissima parte di DNA: circa 15 milioni su 3 miliardi di nucleotidi che compongono il nostro DNA contengono, in teoria, la chiave dell'evoluzione dell'uomo nei 6 milioni di anni che segnano il suo distacco dalla linea evolutiva del *Pan troglodytes*. Inoltre, di questi 15 milioni di varianti, solo un'ulteriore piccolissima percentuale di modificazioni ha un effetto dal punto di vista funzionale, rappresentando dunque una mutazione in grado di determinazioni significative a livello dell'organismo.

Tra le varie sfide che ha comportato la disponibilità di enormi quantità di dati di sequenza, una delle più interessanti ha riguardato il miglioramento delle capacità di allineamento dei genomi, arrivando ad ottenere alti livelli di precisione nell'individuazione di quella piccolissima percentuale di cambiamenti tra organismi diversi. Tramite l'uso di mezzi informatici, è inoltre possibile distinguere tra i cambiamenti genetici che non hanno effetti significativi nello sviluppo di un organismo (nel senso che non ne migliorano né peggiorano il funzionamento o l'adattabilità), e quelli che invece influenzano direttamente le funzioni geniche. Generalmente, la velocità alla quale avvengono le mutazioni nei genomi è un dato indicativo della loro stessa significatività: le mutazioni casuali tendono a manifestarsi ad una velocità standard, che viene quindi spesso usata dai software come segno d'identificazione del primo gruppo di mutazioni, definite "neutrali". I punti del genoma in cui le variazioni mostrano un tasso di mutazione fuori dalla media rappresentano invece loci di alto interesse, poiché denotano generalmente la

non casualità delle mutazioni stesse, e dunque un possibile effetto di selezione positiva. L'evoluzione, com'è noto, tende a favorire le mutazioni genetiche che sostengono la vita, lo sviluppo, le capacità di riproduzione e sopravvivenza di un organismo, che in sintesi ne migliorano la *fitness*.

Un aspetto qualitativo di notevole importanza che è emerso dal sequenziamento del genoma di molti microrganismi e animali, consiste nel fatto che la localizzazione delle sostituzioni delle basi di DNA è molto importante e può avere un rilievo maggiore del numero delle sostituzioni stesse. In particolare, è emerso come i grandi cambiamenti fenotipici fra noi e le scimmie non si sono determinati mediante un'accelerazione complessiva dell'orologio molecolare, ma a causa di rapidi cambiamenti in siti di maggiore impatto sul funzionamento dell'organismo. Secondo gli studi più recenti, proprio all'interno delle regioni del genoma che hanno subito una rapida evoluzione, potrebbe risiedere quella scintilla di umanità che ha plasmato l'uomo, allontanandolo drasticamente dai suoi più vicini antenati (Pollard et al. 2006).

### *2.1 La scoperta di HAR1*

Sulla base di questa linea di pensiero, nel 2004 un team di ricercatori della University of California, guidato da Katherine Pollard, ha individuato una stringa all'interno del codice genetico umano che presentava un tasso di mutazioni insolitamente alto rispetto al *Pan troglodytes*, definendo la regione d'interesse HAR1 (Highly Accelerated Region o anche Human Accelerated Region 1). La scoperta di HAR1 ha aperto il campo alla ricerca di queste sequenze fortemente mutate nel genoma umano ma allo stesso tempo, altamente conservate negli stadi evolutivi precedenti alla comparsa dell'uomo: il gruppo di Pollard ha identificato fino a 202 regioni del genoma umano che hanno subito modificazioni così significative da rappresentare una chiave di lettura importante per l'evoluzione biologica tipicamente umana (Pollard et al. 2006). La particolarità di queste regioni è che sono per la maggior parte non

codificanti, sottolineando ancora una volta l'importanza dello studio del cosiddetto "junk DNA" (DNA non codificante) nell'ambito dell'identificazione di elementi regolatori del codice genetico.

Il gene *HARI* non codifica per una proteina, bensì per un RNA con funzioni regolatorie, e come tale è il primo caso documentato di RNA sottoposto a selezione positiva. Il fatto che sia stato praticamente congelato per centinaia di milioni di anni indica che questa sequenza ha una funzione importante, che poi ha subito un'improvvisa revisione all'interno del genoma umano. In particolare, una specifica sezione di 118 basi è diventata oggetto di paragone con altre specie, come il topo, mostrando così la sua particolare storia evolutiva: quest'area, infatti, ha subito pochissime modifiche fino all'arrivo dell'uomo. Dalla gallina fino allo scimpanzé, la sequenza è rimasta pressoché identica per circa 300 milioni di anni, differenziandosi di sole 2 basi. Al contrario, come si può notare dalle immagini riassuntive in Figura 4 e 5, nel ben più breve lasso di tempo che separa lo scimpanzé dall'uomo, le differenze sono arrivate a 18 basi: la straordinaria accelerazione appare evidente.

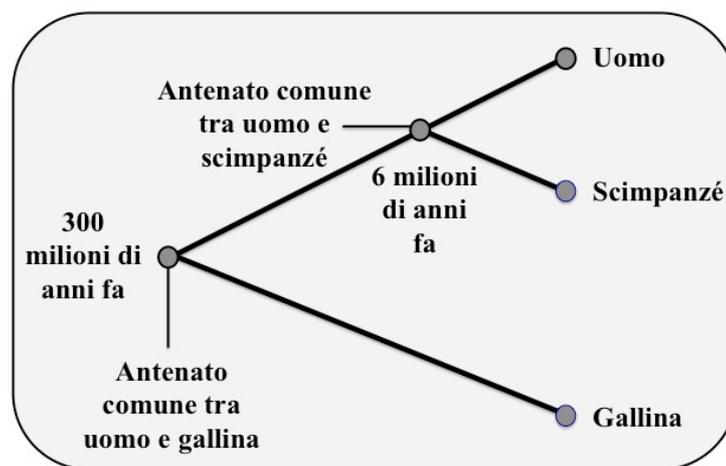


Figura 4: Albero genealogico che indica il punto di divergenza uomo-gallina e uomo-scimpanzé. Immagine rielaborata da Pollard (2009).

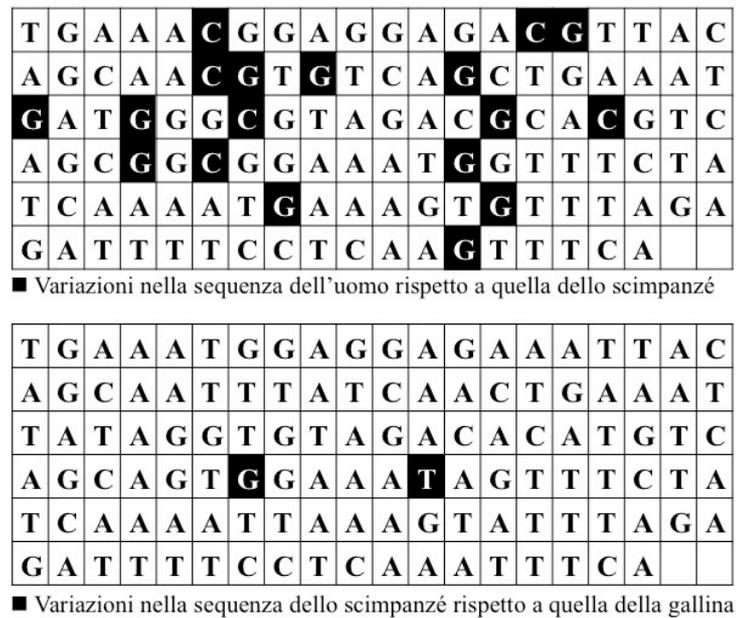


Figura 5: Dettaglio delle mutazioni nucleotidiche, riscontrate paragonando una sequenza di 118 basi, della regione HAR1, tra i genomi degli organismi indicati. Dal basso: scimpanzé vs. gallina (2 mutazioni), uomo vs. scimpanzé (18 mutazioni). Immagine rielaborata da Pollard (2009).

Ciò che si può dedurre già da queste informazioni è che HAR1 è una regione altamente conservata, perché presente in tutti gli organismi analizzati lungo la scala evolutiva, e allo stesso tempo che la sua rapida evoluzione nel genoma umano è un indicatore di una qualche specifica funzione che è stata improvvisamente modificata nel lignaggio umano.

E dunque, qual è questa funzione? Studi successivi dimostrarono che *HAR1* è attivo come elemento regolatorio in uno specifico tipo di neurone necessario ad un corretto sviluppo della corteccia cerebrale, che ne definisce pattern e circonvoluzioni, contribuendo allo sviluppo di una corteccia sana. Un malfunzionamento di questi neuroni è riscontrato nello sviluppo della lissencefalia, una malformazione molto grave, spesso mortale, della corteccia cerebrale, che si presenta eccessivamente liscia e priva delle sue tipiche circonvoluzioni corticali (Pollard et al. 2006b). Da un punto di vista patologico, il malfunzionamento del metabolismo di *HAR1* potrebbe essere correlato alla schizofrenia (Tolosa et al. 2008), come sostiene anche recentemente Xu et al. (2015): alcune componenti dell'architettura genetica della schizofrenia sembrano essere specie-specifiche, ovvero, in questo caso,

pertinenti solo alla specie umana, e quindi riconducibili all'evoluzione del lignaggio umano e alla specifica evoluzione di elementi genici tipicamente umani, come *HARI*.

Da un punto di vista più strettamente morfologico, Beniaminov e collaboratori (2008) hanno riportato differenze significative tra la tipica forma a quadrifoglio assunta da *HARI* nell'uomo rispetto a quella a forcina nelle scimmie antropomorfe: questa differenza morfologica tra le strutture secondarie del gene è probabilmente indice di differenze funzionali della molecola tra i due organismi.

Sebbene non codifichi per una proteina, *HARI* svolge quindi un ruolo importante nella formazione di un cervello sano. Mentre un tempo l'attenzione della genetica era focalizzata principalmente su quella piccola frazione di DNA codificante (circa il 2%), tralasciando il resto, la ricerca ha ormai chiaramente dimostrato che la parte di DNA non codificante è un'immensa e ricca area di studio, sebbene le sue funzioni non siano direttamente "visibili" in prodotti come le proteine, ma giocando un ruolo chiave di regolazione nei processi genetici in generale.

Ciò che il caso di *HARI* ci insegna, inoltre, è che all'interno dei genomi, spesso non è la quantità delle mutazioni avvenute che definiscono l'entità del cambiamento evolutivo, ma il posizionamento di queste mutazioni in zone specifiche, punti chiave del genoma in oggetto, proprio come è accaduto per *HARI*.

## *2.2 Esseri parlanti: il ruolo di FOXP2*

Decadi di ricerche hanno permesso di identificare le basi anatomiche e fisiologiche di quelle differenze a livello cerebrale che potrebbero essere in parte responsabili delle caratteristiche uniche del cervello umano. Diversi attributi dell'*Homo sapiens* derivano presumibilmente da un insieme di funzionalità neurali assenti o fortemente ridotte in altre specie: la nostra specie si sforza incessantemente di inventare dispositivi sempre più utili, migliorare i

sistemi sociali, produrre nuove opere d'arte. La capacità creativa deriva da una flessibilità cognitiva che ci permette, tra l'altro, di formare un numero potenzialmente illimitato di nuove parole e frasi, nonché strumenti, arte, forme di danza e musica.

Se si parla di quelle caratteristiche che più marcatamente distinguono l'uomo dai suoi più recenti antenati, e dagli altri animali in generale, non si può non pensare al complesso sistema di comunicazione sviluppato dagli esseri umani. Il linguaggio inteso come la capacità di generare un numero infinito di espressioni a partire da un set finito di elementi linguistici è una capacità esclusivamente umana: molti animali, non solo mammiferi, hanno senza dubbio un loro sistema comunicativo, efficace generalmente a svolgere o coadiuvare funzioni di base relative al compimento di azioni necessarie alla sopravvivenza. Esempi classici sono la danza delle api, che varia a seconda del messaggio da veicolare, generalmente relativo alla scoperta e allo sfruttamento di una fonte di cibo o alla ripartizione dei lavori nell'alveare (Beekman et al. 2008), o le vocalizzazioni di alcuni primati, oggetto di studio interdisciplinare perché fonte di informazioni preziose sullo sviluppo del linguaggio nell'uomo (Ghazanfar & Miller 2006).

Nessun sistema comunicativo animale, finora, si è tuttavia dimostrato tanto complesso quanto quello umano, che ha caratteristiche uniche, come la capacità di esprimere in potenza qualunque tipo di pensiero. Le origini di questa particolare facoltà cognitiva sono ancora molto dibattute, passando per ipotesi contrastanti, da chi intende il linguaggio come una naturale evoluzione delle abilità cognitive dei primati, positivamente selezionata perché utile alla sopravvivenza della specie, (Cheney & Seyfarth 2005; Zuberbühler 2006; Chater & Christiansen 2010) all'ipotesi innatista che sostiene l'esclusività umana di questa capacità, attuabile solo grazie all'esistenza di un dispositivo specifico che rende atti allo sviluppo del linguaggio (Chomsky 1965; Hauser et al. 2002). Al di là delle varie ipotesi e teorie, la genetica in questo senso rappresenta ancora una volta una fonte preziosa di informazioni: gli studi sui geni coinvolti nel linguaggio aiutano a fare chiarezza su quanto ci sia di

biologico e genetico, seppur con influenze ambientali, nell'abilità di comprendere e produrre il linguaggio.

La prima correlazione tra genetica e sviluppo del linguaggio è rappresentata dalla scoperta del diretto coinvolgimento di un gene, *FOXP2*, nella capacità di produrre correttamente il linguaggio (Lai et al. 2001). La storia di *FOXP2* inizia con la scoperta di una mutazione in comune tra gli individui di un grande pedigree di tre generazioni (anonimamente definito Famiglia KE) colpiti da un grave disturbo del linguaggio che comporta sia problemi nell'articolazione dei suoni che nell'assimilazione di regole morfologiche, oltre che un QI sotto la media. I ricercatori identificarono che si trattava di una transizione G-A nell'esone 14 del gene *FOXP2*, che determinava il cambiamento aminoacidico da arginina a istidina nel codone 553 della proteina (R553H). Studi di neuroimaging hanno rivelato nei portatori della mutazione anomalie nella morfologia e nell'attività dei gangli basali; studi embriologici hanno dimostrato che sia la versione umana di questo gene che quella murina modulano lo sviluppo dei gangli basali e di altre strutture sub-corticali. Dalla scoperta di Lai e colleghi, nel 2001, *FOXP2* è stato considerato per molto tempo il “gene del linguaggio”, definizione in realtà approssimativa e sensazionalistica: per la genetica è chiaro che non esiste un gene del linguaggio, deputato allo sviluppo o al sottosviluppo di questa abilità, ma che, come per qualunque altra caratteristica o funzione, i geni e i processi coinvolti sono molteplici e intricati (Bishop 2002).

*FOXP2* ha sicuramente rappresentato un importante indizio per gli studi sul linguaggio, riaccendendo il dibattito tra linguisti e aggiungendo prospettive nuove alla linguistica stessa, ma come dimostrato successivamente, è un gene altamente conservato, presente in moltissime specie che non hanno sviluppato alcun tipo di linguaggio e coinvolto in molteplici funzioni al di là di quelle linguistiche. In ogni caso, sebbene sia riconosciuto che il linguaggio sia un'abilità cognitiva molto complessa, *FOXP2* rimane un punto di partenza importante per un'analisi complessiva e interdisciplinare delle varie sfaccettature del fenomeno “linguaggio”, in quanto primo caso di un gene direttamente collegato con lo sviluppo di questa attività. In aggiunta a questo,

le alterazioni della facoltà del linguaggio sono considerate generalmente sintomo di alcune malattie mentali, e per questo motivo *FOXP2* è stato spesso studiato come un possibile locus di suscettibilità per vari tipi di patologie. Una ricerca condotta su tre gruppi di pazienti cinesi affetti da schizofrenia/ depressione/ bipolarismo ha evidenziato il legame con queste patologie (Li et al. 2013). Anche per l'autismo è stata ricercata una possibile correlazione con mutazioni di *FOXP2* e i pareri della comunità scientifica sono tuttora contrastanti: la maggior parte degli studi non ha rilevato una correlazione significativa (Gauthier et al. 2003), sebbene Gong e collaboratori (2004) abbiano riscontrato associazione con l'autismo per alcuni SNPs (Single Nucleotide Polymorphisms) di *FOXP2*.

Anche *FOXP2* è un gene molto antico e presenta un iter evolutivo simile alle altre sequenze presentate. Il dibattito su una possibile data di apparizione della versione “umana” di *FOXP2*, ovvero quella che contribuì a rendere possibile la facoltà di linguaggio umana, è ancora acceso. L'analisi della sua storia molecolare dimostra che, nei circa 130 milioni di anni che separano l'antenato comune dell'uomo e della scimmia dal topo, è avvenuta un'unica variazione di aminoacido nella proteina FOXP2, mentre nei 6 milioni di anni circa che separano uomo e scimmia, questa proteina nell'uomo ha modificato due dei suoi aminoacidi, rimasti invece invariati nello scimpanzé e negli altri primati (Enard et al. 2002). La proteina umana FOXP2 differisce dalla sequenza del gorilla e dello scimpanzé in appunto due residui, conosciuti come T303N e N325S, e dalle sequenze dell'orangutan e del topo in tre e quattro residui rispettivamente. Studi più recenti basati sul sequenziamento di *FOXP2* ritrovato in fossili di Neanderthal (Krause et al. 2007), sostengono che questi cambiamenti siano condivisi con i nostri parenti estinti più vicini, i Neanderthals appunto, suggerendo che dunque, sottoposti a selezione positiva, essi risalgano a prima del comune antenato dell'uomo e del Neanderthal (esistito circa 300.000-400.000 anni fa).

In conclusione, non esisterebbe un “gene umano” del linguaggio, poiché il gene si trova in altre specie: tuttavia, c'è la possibilità che le due sostituzioni

nella proteina FOXP2 caratterizzanti la discendenza umana siano di importanza funzionale per l'origine del linguaggio.

### 2.3 *ASPM e lo sviluppo cerebrale*

La visione che il cervello umano non rappresenti un break qualitativo rispetto a tutte le altre specie, ma piuttosto il risultato di un trend che hanno affrontato anche altri primati trova riscontro in molti studi. Cambiamenti progressivi si sono realizzati per un periodo prolungato di 60-70 milioni di anni nella linea evolutiva che porta dai primati ancestrali all'uomo moderno, sebbene il tasso di cambiamento più importante si sia registrato solo negli ultimi. La grandezza della scatola cranica e del cervello dell'*Homo sapiens*, triplicata nell'arco di 2 milioni di anni, è un tratto distintivo rispetto agli altri primati ed è considerato uno dei cambiamenti cruciali che ha contribuito allo sviluppo del linguaggio e di altre abilità cognitive superiori umane. Sebbene questo salto evolutivo sia riconosciuto come un fenomeno fondamentale, le sue basi genetiche sono ancora poco conosciute e un'analisi genomica può aiutare a capire a quali modificazioni tipiche del genoma possa essere attribuita la peculiare crescita dell'encefalo umano rispetto a quello degli altri primati.

*ASPM* (Abnormal SPindle-like Microcephaly associated o Abnormal spindle microtubule associated), collocato sul cromosoma 1, è uno degli esempi più studiati di un gene correlato alle dimensioni del cervello. È in generale riconosciuto come un elemento genico che incide positivamente sulla fitness dell'organismo uomo (Zhang 2003), essendo stato conservato per milioni di anni ed è considerato una delle “novità” genetiche che ci rendono umani. Perché? Studi genetici hanno dimostrato il ruolo di questo gene, insieme ad altri, nello sviluppo di dimensioni del cervello nella norma (Bond et al. 2002), attraverso esperimenti condotti su casi di microcefalia, dove la dimensione del cervello può ridursi fino al 70%. Uno studio del 2007 (Dediu & Ladd 2007) ha inoltre analizzato la possibile correlazione esistente tra *ASPM* e lo sviluppo delle lingue tonali, tramite uno studio di popolazioni che mostra

come le aree linguistiche tonali correlino con un'alta frequenza di una variante di *ASPM* nella popolazione.

Come *HARI*, anche *ASPM* ha una storia che dimostra il processo di selezione positiva a cui è stato sottoposto. La sua apparizione risalirebbe a circa 5 milioni 800 mila anni fa, dunque in prossimità temporale con il punto di divergenza uomo-scimmie antropomorfe (Mekel-Bobrov et al. 2005). Nello studio di Evans e collaboratori (2004), la storia di *ASPM* è stata analizzata nel dettaglio allo scopo di dimostrare che la sua evoluzione è significativamente accelerata nella discendenza che dalle scimmie antropomorfe porta all'uomo. I test statistici hanno rilevato che *ASPM* tende a fissare una sostituzione aminoacidica vantaggiosa ogni 300.000 – 400.000 anni dal punto di divergenza uomo-scimpanzé, rafforzando l'ipotesi del suo forte ruolo nell'evoluzione dell'*Homo sapiens*, in particolare dell'ingrandimento dell'encefalo che lo caratterizza.

### **3. Analisi comparativa dei tre geni in primati e non-primati**

Partendo dai presupposti delineati finora, abbiamo concentrato la nostra attenzione di linguisti e genetisti sull'analisi dei tre geni descritti, *FOXP2*, *ASPM* e *HARI*, direttamente collegati alla cognizione, allo sviluppo cerebrale e del linguaggio umano. Allo scopo di identificare le regioni geniche maggiormente mutate lungo la scala evolutiva, abbiamo sottoposto le sequenze nucleotidiche di tutti e tre i geni e aminoacidiche degli ultimi due, in quanto codificanti proteine, ad un processo di allineamento multiplo, considerando 5 organismi oltre all'uomo: quattro primati (*Pan troglodytes*, *Gorilla gorilla*, *Pongo abelii*, *Macaca mulatta*), e un altro mammifero vertebrato, il *Mus musculus* (*Rattus norvegicus*, in alcuni casi). In seguito all'allineamento, le sequenze sono state utilizzate per la costruzione di alberi filogenetici, che ci consentono di confrontare il tasso di mutazione delle sequenze sia all'interno dei primati, che dei mammiferi come *Rattus norvegicus* e *Mus musculus* con

l'uomo, ed infine con un organismo evolutivamente più distante, come il *Gallus gallus*.

Nei prossimi paragrafi sono presentati i metodi utilizzati per l'allineamento delle sequenze e i risultati ottenuti per ognuna delle sequenze analizzate.

### *3.1 Metodologie usate per l'analisi bioinformatica delle sequenze genomiche*

Il primo passo per lo studio dei geni prescelti è rappresentato dal retrieval, ovvero il recupero delle sequenze dai database:

*i. Download della sequenza genomica* – Il punto di partenza è Ensembl, il database costruito all'interno dell'Ensembl Project nella versione della sua ultima release, Ensembl 2016<sup>2</sup>. Ensembl è un sistema per l'annotazione, l'analisi e lo storage di genomi, continuamente ampliato e aggiornato, corredato di strumenti bioinformatici per l'allineamento e l'analisi genomica in generale (Yates et al., 2015). Nella sequenza di riferimento umana (ultima release: GRCh38)<sup>3</sup> sono state ricercate e selezionate le sequenze d'interesse.

*ii. Analisi di omologia in BLAST* – In questa fase, la sequenza genomica umana, scaricata da Ensembl in formato FASTA, viene sottoposta ad analisi di omologia in BLAST (Basic Local Alignment Search Tool), direttamente dal sito dell'NCBI<sup>4</sup>, "The National Center for Biotechnology Information" (NCBI Resource Coordinators, 2016). BLAST è un programma che ricerca regioni di similarità tra sequenze genomiche (Johnson et al. 2008); basato sull'algoritmo sviluppato da Altschul e collaboratori (Altschul et al. 1990), BLAST si basa sull'assunto che in sequenze omologhe - anche notevolmente divergenti - possono essere individuate regioni che mostrano una similarità statisticamente significativa. Nel confronto tra sequenze nucleotidiche è attribuito punteggio positivo alle identità e negativo alle non identità. Per l'analisi delle sequenze nucleotidiche si può utilizzare l'algoritmo più comune, blastn, oppure gli algoritmi megablast e discontinuous megablast, ottimizzati per l'allineamento

---

<sup>2</sup> Sito ufficiale dell'Ensembl Project: [www.ensembl.org](http://www.ensembl.org)

<sup>3</sup> Tutte le informazioni tecniche sullo Human assembly sono consultabili online su: [http://www.ensembl.org/Homo\\_sapiens/Info/Annotation](http://www.ensembl.org/Homo_sapiens/Info/Annotation)

<sup>4</sup> Sito ufficiale: <http://www.ncbi.nlm.nih.gov/blast>

di lunghe sequenze non particolarmente divergenti. Nel nostro caso, sono state scelte le sequenze genomiche di 4 organismi modello: *Pan troglodytes* (scimpanzé), *Gorilla gorilla* (gorilla), *Macaca mulatta* (macaco) e *Mus musculus* (topo) o *Rattus norvegicus* (ratto), a seconda della disponibilità all'interno del database, e sono state paragonate tramite allineamento globale mediante discontinuous megablast con la sequenza di *Homo sapiens*. Discontinuous megablast è più sensibile di blastn, e utilizzato particolarmente per confronti cosiddetti cross-species. Per ciascun organismo analizzato, fin dallo scaricamento in formato FASTA delle sequenze dal sito di Ensembl si è dovuto scegliere di analizzare tutto il gene o un blocco di esso, scegliendo in quest'ultimo caso il frammento genico più lungo.

*iii. Allineamento multiplo delle sequenze aminoacidiche* – Quest'analisi è stata condotta per le sequenze geniche codificanti proteine, quali *FOXP2* e *ASPM*. Lo stesso sito NCBI contiene al link di HomoloGene<sup>5</sup> sia allineamenti pairwise che multipli di sequenze aminoacidiche. Sebbene molto utile, il sito tuttavia analizza solo una variante proteica di tutte quelle determinate da un singolo gene; per questo, al fine di analizzare l'allineamento e successivamente costruire gli alberi filogenetici, le sequenze aminoacidiche corrispondenti ad una specifica variante, la più lunga di norma, sono state scaricate in formato FASTA dal sito di NCBI, e sottoposte ad allineamento mediante ClustalW2. Questo algoritmo, sviluppato tra gli altri sul modello dell'EBI<sup>6</sup>, è una risorsa che consente il Multiple Sequence Alignment (MSA) di tre o più sequenze di lunghezza simile. Dai risultati dell'allineamento si deducono le regioni contenenti i cambiamenti aminoacidici rispetto alla sequenza di riferimento, nel nostro caso la sequenza umana.

*iv. Alberi filogenetici* – Al fine di rappresentare mediante un albero di tipo filogenetico la distanza tra i genomi dei diversi organismi analizzati, le sequenze aminoacidiche scelte e allineate sono state quindi caricate e analizzate in Geneious 10<sup>7</sup> (Kearse et al. 2012), una suite bioinformatica che

---

<sup>5</sup> HomoloGene database su NCBI: <https://www.ncbi.nlm.nih.gov/homologene>

<sup>6</sup> Sito ufficiale: <http://www.ebi.ac.uk/Tools/msa/>

<sup>7</sup> Sito ufficiale: <http://www.geneious.com/>

raccoglie numerosi strumenti analitici per la biologia molecolare. Nel nostro caso, l'analisi filogenetica è usata come supporto analitico per una comprensione più rigorosa dell'evoluzione dei geni/proteine presi in esame, coinvolti nello sviluppo cognitivo. L'albero è un grafo bidimensionale che mostra le relazioni evolutive esistenti tra diverse entità (geni, sequenze, organismi) e rappresenta uno strumento utile a mostrare la distanza tra organismi rispetto alla sequenza genetica analizzata, quantificando e riportando la divergenza tra due nodi sulla lunghezza degli archi (Branch Length), che normalmente rispecchia la lontananza tra organismi in termini di evoluzione. Un albero si compone di nodi terminali, che rappresentano oggetti esistenti; nodi interni o punti di divergenza o biforcazioni, che rappresentano ipotetici antecessori; un nodo iniziale o radice che rappresenta l'antecessore di tutti i bracci e linee che congiungono i vari nodi. Per la costruzione degli alberi filogenetici delle proteine *FOXP2* e *ASPM*, il programma costruisce una matrice di distanza tra sequenze, che misura il numero di sostituzioni per sito: in questo modo le distanze nell'albero possono essere interpretate come il numero medio di caratteristiche che differiscono su tutti i siti analizzati in una regione allineata. Il modello di distanza (o modello di evoluzione molecolare) sul quale si basa la costruzione degli alberi per le sequenze aminoacidiche è il Jukes Cantor (Jukes & Cantor 1969), che assume che tutti gli aminoacidi della sequenza abbiano la stessa frequenza (0,05) e che per ogni elemento il tasso di sostituzioni sia eguale. Il metodo di clustering scelto per la ricostruzione degli alberi è UPGMA, un sistema di clustering basato su "Unweighted Pair Group Method using arithmetic Average" (Michener & Sokal 1957). L'algoritmo raggruppa le sequenze a partire dalle più simili ed aggiungendo successivamente un nodo all'albero, producendo alla fine alberi con radice ed ultrametrici.

## 4. Risultati

I risultati e le statistiche indicati nei paragrafi successivi sono estrapolati dai valori degli allineamenti ottenuti con i metodi appena descritti. Per ogni gene, è presentato, per maggiore chiarezza, un piccolo identikit contenente le informazioni essenziali sulla sequenza usata, la lunghezza e la posizione cromosomica.

### 4.1 *HARI*: Risultati dell'allineamento multiplo della sequenza nucleotidica

Come abbiamo recedentemente descritto, il lavoro di questa tesi parte dalle osservazioni di Pollard e collaboratori (2006) di regioni genomiche che hanno subito una sorta di “accelerazione” nel loro tasso mutazionale nell’uomo rispetto a organismi più in basso lungo la scala evolutiva. Riprendendo il loro lavoro, e ricercando nei databases specializzati la sequenza del gene *HARI* umano, si trovano due varianti, A e B; solo la prima ha implicazioni con lo sviluppo cognitivo, per cui il tratto preso in esame per l’analisi genomica è *HAR1A*, che come si può evincere dalla tabella di sintesi (Tabella 2) e dall’immagine sulla posizione cromosomica (Figura 6), è un long non-coding RNA posizionato sul cromosoma 20.

<i>HAR1A</i>	
<b>DEFINIZIONE</b>	<b>Highly Accelerated Region 1A</b>
<b>Organismo</b>	<i>Homo sapiens</i>
<b>Codice identificativo</b>	NR_003244
<b>Tipo</b>	long non-coding RNA
<b>Paia di basi</b>	2794
<b>Cromosoma</b>	20q13.33

Tabella 2: Identikit della sezione di sequenza *HAR1A*.

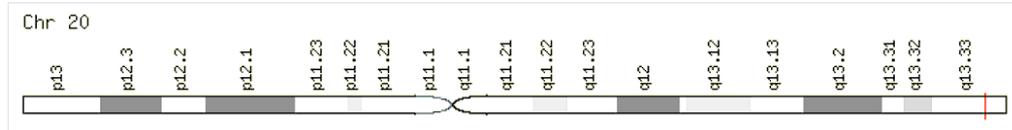


Figura 6: Posizione genomica di *HARIA*. Fonte: Ensembl.

La Figura 7 riporta i risultati dell'allineamento di *HARIA* tra gli organismi scelti. È innanzitutto importante sottolineare come la sequenza umana trovi omologia con frammenti di sequenze degli altri organismi, spesso con cloni di sequenze genomiche piccoli e senza funzione ancora nota, a dimostrazione del fatto che si tratti di un campo in esplorazione. Dall'analisi di un frammento comune tra uomo e organismi scelti pari a 105 basi si possono notare, evidenziate nei riquadri in rosso, le singole sostituzioni interspecie, e mentre si evince chiaramente la similarità che caratterizza le sequenze dei cinque organismi, nel paragone la sequenza umana, si notano modifiche sostanziali nella sequenza nucleotidica.



Figura 7: Risultato dell'allineamento multiplo della sequenza nucleotidica di *HARIA*. Nei riquadri in rosso sono evidenziate le sostituzioni nucleotidiche inter-specie per questa sezione di sequenza.

Come sintetizzato in Tabella 3, la percentuale di variabilità è simile tra i primati, sebbene piuttosto alta rispetto all'atteso dell'1-2% che caratterizza il genoma umano rispetto a quello degli altri primati, con i quali condivide appunto il restante 98% di elementi genomici. Al contrario, la differenza tra topo e primati non umani è quasi nulla, pari a sole 2 mutazioni, evidenziando il congelamento della sequenza riportato dalla letteratura.

<i>HARIA</i>		
<b>ORGANISMO</b>	<b>N° Variazioni</b>	<b>Frequenza relativa</b>
<i>Pan troglodytes</i>	15/105	0,143
<i>Gorilla gorilla</i>	15/105	0,143
<i>Pongo abelii</i>	15/105	0,143
<i>Macaca mulatta</i>	14/105	0,133
<i>Mus musculus</i> vs <i>Macaca mulatta</i>	2/105	0,019

Tabella 3: Descrizione della livello di variabilità tra differenti genomi, risultante dall'allineamento nucleotidico di *HARIA* (sequenza di 105 pb) in *Homo sapiens* vs *Pan troglodytes*, *Gorilla gorilla*, *Pongo abelii*, *Macaca mulatta* e *Mus musculus*.vs *Macaca Mulatta*

#### 4.2 *FOXP2*: Risultati dell'allineamento multiplo della sequenza nucleotidica e aminoacidica

L'identikit della sequenza del gene *FOXP2* con le sue informazioni essenziali (Tabella 4) e una grafica della posizione del gene in Figura 8 mostrano come questo sia un gene molto lungo localizzato sul cromosoma 7:

<i>FOXP2</i>	
<b>DEFINIZIONE</b>	<b>Forkhead box P2</b>
<b>Organismo</b>	<i>Homo sapiens</i>
<b>Codice identificativo</b>	ENSG00000128573
<b>Tipo</b>	Protein coding
<b>Paia di basi</b>	607.445
<b>Cromosoma</b>	7q31.1
<b>Treascritti</b>	26

Tabella 4: Identikit di *FOXP2* da Ensembl.

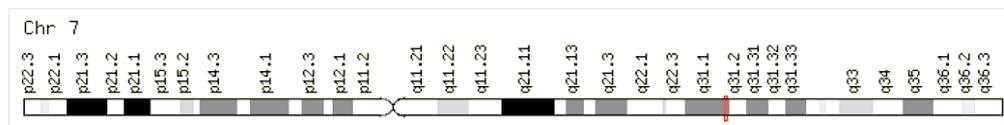


Figura 8: Posizione genomica di *FOXP2*. Fonte: Ensembl.

Per *FOXP2*, sono di seguito presentati i risultati ottenuti attraverso l'allineamento multiplo sia della sequenza nucleotidica del gene che della sua versione tradotta in proteina, dunque la sequenza aminoacidica.

La Figura 9 riporta i risultati dell'allineamento di *FOXP2* tra gli organismi scelti: come prima, nei riquadri in rosso, sono evidenziate le singole sostituzioni interspecie. Confrontando tra di loro le sequenze alle posizioni indicate, notiamo come questa volta la differenza maggiore è tra primati e non primati, mentre la sequenza umana assomiglia moltissimo a quella degli altri primati.

#### FOXP2

human	tgcag-tgcTgtgtAgtggccatgatgactccccGgtgatcacccctcagcaAatgc
Pan Tr.	tgcagGtgctgtgtcagtggccatgatgactccccaggtgatcacccctcagcaaatgc
Macaca M.	tgcagGtgctgtgtcagtggccatgatgactccccaggtgatcacccctcagcaaatgc
Gorilla G.	tgcagGtgctgtgtcGgtggccatgatgactccccaggtgatcacccctcagcaaatgc
Rattus N.	tgcagGtgctGgtgtcGgtggccatgatgactccccAgtgatcacccctcagcaGatgc
human	agcagatccttcagcaAcaAgtcctgtctcctcagcagctAcaAgccctTctccaAcaAc
Pan Tr.	agcagatccttcagcaacaagtccctgtctcctcagcagctCcaagcccttctccaacaad
Macaca M.	agcagatccttcagcaacaagtccctgtctcctcagcagctCcaagcccttctccaacaGc
Gorilla G.	agcagatccttcagcaacaagtccctgtctcctcagcagctCcaagcccttctccaacaad
Rattus N.	agcagatccttcagcaGcaGgtcctgtctcctcagcagctCcaGgccctCctccaGcaGc

Figura 9: Sezione dell'allineamento della sequenza nucleotidica di *FOXP2*, nella quale si evidenziano le variazioni tra genomi.

<http://www.altronauti.com/>

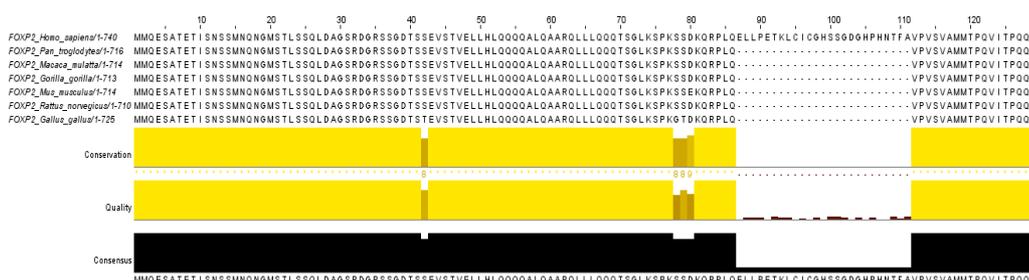
Andando a valutare quantitativamente, nel confronto delle intere sequenze nucleotidiche sottoposte ad allineamento, emerge l'alta percentuale di conservazione, con una percentuale di variabilità tra uomo e altri primati molto bassa, che si aggira intorno all'1%, scendendo addirittura sotto questa soglia nel confronto tra *Homo sapiens* e *Pan troglodytes*. In Tabella 5, il dettaglio della variabilità:

<b>FOXP2 – Allineamento sequenza nucleotidica</b>				
<b>ORGANISMO</b>	<b>Identità <i>H. sapiens</i>/altro organismo</b>	<b>N° Variazioni</b>	<b>Frequenza relativa*</b>	<b>N° Gaps**</b>
<i>Pan troglodytes</i>	210268/213395 (0,99)	3127	0,014	954 (0%)
<i>Gorilla gorilla</i> *	42677/43566 (0,98)	889	0,020	270 (0%)
<i>Macaca mulatta</i> *	46674/49723 (0,94)	3049	0,061	896 (1%)
<i>Mus musculus</i> *	13108/16417 (0,80)	3309	0,201	904 (5%)

**Tabella 5: Valori di variabilità ottenuti dall'allineamento nucleotidico di *FOXP2* umano (Block 1)\* contro *Pan troglodytes*, *Macaca mulatta*, *Gorilla gorilla* e *Mus musculus*. \*\*numero di gaps rispetto alla sequenza umana.**

Se questo risultato rispecchia la conservazione di sequenza di un gene importante tra vari organismi lungo la scala evolutiva, dall'analisi dei risultati ottenuti con l'allineamento della sequenza aminoacidica della proteina *FOXP2*, e' possibile trarre alcune osservazioni, che riteniamo significative:

A) come mostrato in Figura 10, una intera regione o stretch di 25 aminoacidi, corrispondenti dunque a 75 nucleotidi, è presente solo nella sequenza umana di *FOXP2* e dunque assente anche nel parente più prossimo *Pan troglodytes*.



**Figura 10: Risultato dell'allineamento multipla della sequenza aminoacidica di *FOXP2*.**

B) per *FOXP2* si può osservare chiaramente una maggiore conservazione all'interno dei primati e una più marcata divergenza con il *Rattus norvegicus*, in linea con l'ipotesi che il gene esistesse in una versione moderna già molto



In totale, 34 variazioni aminoacidiche sono invece riscontrabili nel confronto tra la sequenza umana contro quella del *Pan troglodytes*, a rappresentazione dei primati, e con la sequenza del *Rattus norvegicus*, di cui 33 positive, ovvero che cambiano anche la classe dell'aminoacido (Tabella 6).

<b>FOXP2 – Allineamento sequenza aa</b>		
<b>ORGANISMO</b>	<b>N° tot. variazioni</b>	<b>N° variazioni positive</b>
<i>Pan troglodytes</i>	3	2
<i>Rattus norvegicus</i>	34	33

Tabella 6: Valori di variabilità ottenuti dall'allineamento della sequenza aminoacidica di *FOXP2* contro *Pan troglodytes* e *Rattus norvegicus*.

Quando le sequenze aminoacidiche vengono usate per ricostruire la distanza tra gli organismi, mediante un albero con radice, si ottiene il seguente cladogramma:

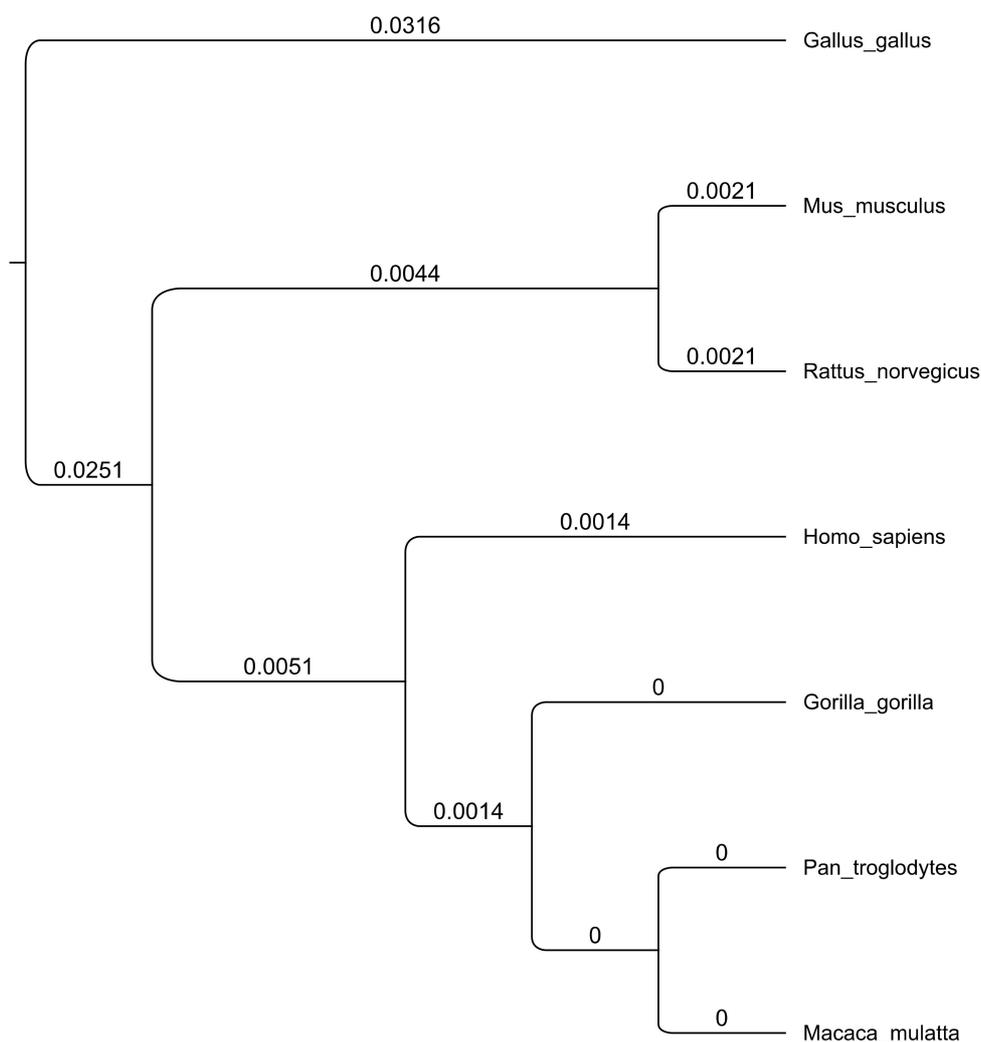


Figura 13: *FOXP2*. Albero filogenetico con radice, costruito con metodo UPGMA.

I branches dell'albero indicano le percentuali di sostituzioni per sito, riportate anche in forma tabellare in basso (Tabella 7); come è possibile notare anche dalle percentuali di identità tra sequenze riportate nella matrice in Tabella 8, maggiore omologia è riscontrabile nel confronto tra primati e tra questi e il genoma del *Mus musculus*, mentre gli organismi più lontani sembrerebbero il *Rattus norvegicus* e più ancora il *Gallus gallus*. Si nota altresì la clusterizzazione dei primati, con pochissime varianti all'interno dei primati

non umani, indice probabilmente di una versione del gene/proteina con caratteristiche molto simili prima dello sviluppo di cognitive più evolute.

FOXP2	<i>H. sapiens</i>	<i>P. troglodytes</i>	<i>M. mulatta</i>	<i>G. gorilla</i>	<i>M. musculus</i>	<i>R. norvegicus</i>	<i>G. gallus</i>
<i>H. sapiens</i>		0.003	0.003	0.003	0.004	0.017	0.103
<i>P. troglodytes</i>	0.003		0.000	0.000	0.013	0.013	0.058
<i>M. mulatta</i>	0.003	0.000		0.000	0.014	0.013	0.055
<i>G. gorilla</i>	0.003	0.000	0.000		0.015	0.016	0.057
<i>M. musculus</i>	0.004	0.013	0.014	0.015		0.004	0.056
<i>R. norvegicus</i>	0.017	0.013	0.013	0.016	0.004		0.051
<i>G. gallus</i>	0.103	0.058	0.055	0.057	0.056	0.051	

Tabella 7: Matrice delle distanze relativa all'albero filogenetico di FOXP2 in figura 11, che riporta la percentuale di sostituzioni per sito.

FOXP2	<i>H. sapiens</i>	<i>P. troglodytes</i>	<i>M. mulatta</i>	<i>G. gorilla</i>	<i>M. musculus</i>	<i>R. norvegicus</i>	<i>G. gallus</i>
<i>H. sapiens</i>		95%	95%	86%	96%	93%	90%
<i>P. troglodytes</i>	95%		98%	90%	98%	97%	94%
<i>M. mulatta</i>	95%	98%		90%	99%	98%	93%
<i>G. gorilla</i>	86%	90%	91%		90%	89%	85%
<i>M. musculus</i>	96%	98%	99%	90%		97%	93%
<i>R. norvegicus</i>	94%	97%	98%	89%	97%		92%
<i>G. gallus</i>	90%	94%	93%	85%	93%	92%	

Tabella 8: Matrice delle distanze relativa all'albero filogenetico di FOXP2 in figura 11, che riporta la percentuale di identità tra una specie e l'altra.

#### 4.3 ASPM: Risultati dell'allineamento multiplo della sequenza nucleotidica e aminoacidica

Come per le due sequenze precedenti, riportiamo le caratteristiche essenziali di ASPM nella Tabella 9, seguite dalla posizione sul cromosoma 1 in Figura 14:

<i>ASPM</i>	
<b>DEFINIZIONE</b>	<b>Abnormal Spindle Microtubule Assembly</b>
<b>Organismo</b>	<i>Homo sapiens</i>
<b>Codice identificativo</b>	ENSG00000066279
<b>Tipo</b>	Protein coding
<b>Paia di basi</b>	62.566
<b>Cromosoma</b>	1q31.3
<b>Treascritti</b>	4

Tabella 9: Identikit di *ASPM* da Ensembl

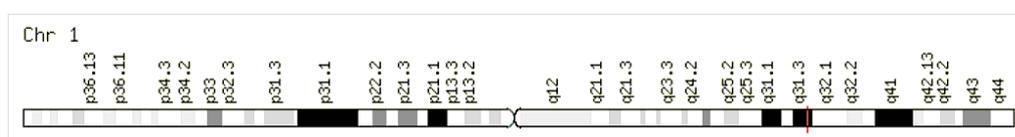


Figura 14: Posizione genomica di *ASPM*. Rielaborata da: Ensembl.

I risultati ottenuti attraverso l'allineamento multiplo della sequenza nucleotidica sono riassunti in Figura 15, che mostra una porzione delle intere sequenze nucleotidiche prese in analisi.

#### ASPM

human	ttcaggc <b>GC</b> Att <b>Cag</b> ggg <b>TAT</b> gaaa <b>Ac</b> Tagaa <b>GA</b> catt <b>Gaa</b> G <b>GT</b> atgcat <b>TCCT</b> :Tg
Pan Tr.	ttcaggc <b>AT</b> Ttt <b>Tag</b> ggg <b>Aa</b> A <b>gaaa</b> G <b>ct</b> agaaga <b>catt</b> A <b>aa</b> A <b>TC</b> atgcat <b>ATAG</b> :Cg
Macaca F.	ttcaggc <b>AT</b> Ttt <b>Tag</b> ggg <b>Aa</b> A <b>gaaa</b> G <b>ct</b> agaaga <b>catt</b> A <b>aa</b> A <b>TC</b> atgcat <b>GTAG</b> :Cg
Gorilla G.	ttcaggc <b>AT</b> Ttt <b>Tag</b> ggg <b>Aa</b> A <b>gaaa</b> G <b>ct</b> agaaga <b>catt</b> A <b>aa</b> A <b>TC</b> atgcat <b>ATAG</b> :Cg
Rattus N.	ttcaggc <b>gc</b> Ctt <b>Tag</b> ggg <b>CG</b> t <b>gaaa</b> G <b>c</b> Agaa <b>AC</b> catt <b>gaa</b> g <b>C</b> tatgcat <b>Gc</b> Tt <b>ctg</b>
human	caac <b>Cct</b> Tattcagag <b>Tag</b> Gtttaga <b>Tc</b> AT <b>ta</b> Ct <b>GG</b> GagGaga <b>AG</b> att <b>cAt</b> TT <b>Cct</b> Ca
Pan Tr.	caac <b>Tct</b> Cattcagag <b>Gag</b> Atttaga <b>Ac</b> TC <b>ta</b> Atg <b>Atg</b> ag <b>Aga</b> agatt <b>cCt</b> C <b>tc</b> T <b>ct</b> ca
Macaca F.	caac <b>cc</b> Cattcagag <b>Gag</b> Atttaga <b>Ac</b> TC <b>ta</b> Atg <b>Atg</b> ag <b>Aga</b> agatt <b>cCt</b> G <b>tc</b> T <b>ct</b> ca
Gorilla G.	caac <b>Tct</b> Cattcagag <b>Gag</b> Atttaga <b>Ac</b> TC <b>ta</b> Atg <b>Atg</b> ag <b>Aga</b> agatt <b>cCt</b> G <b>tc</b> T <b>ct</b> ca
Rattus N.	caac <b>cc</b> ttattcagag <b>ta</b> AA <b>tt</b> taga <b>Gc</b> T <b>AG</b> a <b>At</b> Ag <b>ta</b> ag <b>aga</b> CA <b>att</b> c <b>C</b> tt <b>Gc</b> T <b>ct</b> ca
human	AA <b>aaa</b> G <b>c</b> T <b>act</b> A <b>tt</b> TT <b>Gt</b> Tcagag <b>Gaa</b> ata <b>TC</b> g <b>Agc</b> -CAC <b>c</b> Att <b>TGT</b> G <b>c</b> CA <b>AA</b> CATA <b>aa</b> A
Pan Tr.	a <b>Gaa</b> ag <b>c</b> CG <b>ct</b> a <b>tt</b> tt <b>GAtt</b> cagag <b>Aaa</b> ata <b>tc</b> g <b>Gc</b> d <b>Aca</b> T <b>c</b> -ttt <b>gt</b> A <b>c</b> A <b>aa</b> G <b>cat</b> Ca <b>C</b>
Macaca F.	a <b>Gaa</b> Act <b>Gct</b> G <b>tt</b> GG <b>Att</b> cagag <b>Aaa</b> ata <b>tc</b> g <b>Gc</b> d <b>Aca</b> T <b>c</b> -ttt <b>gt</b> A <b>c</b> A <b>aa</b> G <b>cat</b> Ca <b>C</b>
Gorilla G.	a <b>Gaa</b> Act <b>Gct</b> a <b>tt</b> tt <b>GAtt</b> cagag <b>Aaa</b> ata <b>tc</b> g <b>Gc</b> d <b>Aca</b> T <b>c</b> -ttt <b>gt</b> A <b>c</b> A <b>aa</b> G <b>cat</b> Ca <b>C</b>
Rattus N.	G <b>aaa</b> ag <b>ct</b> G <b>ct</b> a <b>tt</b> C <b>ct</b> G <b>c</b> agag <b>gaa</b> ata <b>C</b> Ag <b>ad</b> CAC <b>cc</b> T <b>tt</b> AT <b>g</b> C <b>aa</b> GCATA <b>aa</b> T

Figura 15: Allineamento nucleotidico di *ASPM*. Nei riquadri in rosso sono evidenziate le sostituzioni nucleotidiche inter-specie.

Lo studio della conservazione di una porzione di sequenza codificante evidenzia una percentuale di variazioni uomo-primati maggiore rispetto alla

media, mentre si osserva maggiore conservazione all'interno del gruppo delle scimmie antropomorfe a livello nucleotidico. I dettagli sulle percentuali di variabilità sono riassunti in Tabella 10.

<i>ASPM</i> – Allineamento sequenza nucleotidica				
<b>ORGANISMO</b>	<b>Identità <i>H. sapiens</i>/altro organismo</b>	<b>N° Variazioni</b>	<b>Frequenza relativa*</b>	<b>N Gaps**</b>
<i>Pan troglodytes</i>	43583/44072 (0,99)	489	0,011	170 (0%)
<i>Macaca mulatta</i>	44678/47944 (0,93)	3176	0,066	907 (1%)
<i>Gorilla gorilla</i> *	43685/44626 (0,98)	941	0,021	398 (0%)
<i>Mus musculus</i> *	2839/3636 (0,78)	797	0,219	57 (1%)

Tabella 10: Valori di variabilità ottenuti dall'allineamento nucleotidico di *ASPM* umano (Block 1)\* contro *Pan troglodytes*, *Macaca mulatta*, *Gorilla gorilla* e *Mus musculus*. . \*\*numero di gaps rispetto alla sequenza umana.

A livello di proteina, la prima osservazione da fare riguarda la differenza tra primati e non primati a livello di sequenza aminoacidica, come in Figura 16:

<b>ASPM</b>			
<u>Homo sapiens</u>	1	-MANRRVGRGCWEVSPTEERRPPA <b>GLRGPAA</b> EEEE-ASSPPVLSLSHFCSRSP	48
<u>Pan Troglodytes</u>	1	-MANRRVGRGCWEVSPTEERRPPA <b>GLRGPAA</b> EEEE-ASSPPVLSLSHFCSRSP	48
<u>Macaca mulatta</u>	1	-MANRRVGRGCWEVSPTEERRPPA <b>ALRGPAT</b> EEEE-ASSPPVLFSLHFCSRSP	48
<u>Rattus norvegicus</u>	1	-MAT <b>VQAAYSSEESGRRARRDP</b> ----- <b>AAE-DAS</b> PPVLLLSHFCAVP	40
<u>Homo sapiens</u>	49	FLCFGDVLL <b>GA</b> SRTLSLALDNPNEEVAEVKISHFPAADLGFVSQRFCVL	98
<u>Pan Troglodytes</u>	49	FLCFGDVLL <b>GA</b> SRTLSLALDNPNEEVAEVKISHFPAADLGFVSQRFCVL	98
<u>Macaca mulatta</u>	49	FLCFGDVLL <b>GD</b> SRTLSLALDNPNEEVAEVKISHFPAADLGFVSQRFCVL	98
<u>Rattus norvegicus</u>	41	FLCFGDV <b>RVGTSR</b> TRSLV <b>LHNPHAEPLHVELSL</b> LRAAG <b>QGGFASPNRCEL</b>	90
<u>Homo sapiens</u>	2353	RYQALKQASVVIQQQY <b>Q</b> ANRA <b>AKL</b> QRQHLYLR <b>QRH</b> SAVILQAAFRGMKTRR	2402
<u>Pan Troglodytes</u>	2353	RYQALKQASVVIQQQY <b>Q</b> ANRA <b>ATL</b> QRQHLYLR <b>QRH</b> SAVILQAAFRGMKTRR	2402
<u>Macaca mulatta</u>	2353	RYQALKQASVVIQQQY <b>R</b> ANRA <b>AKL</b> QRQHLYLR <b>QR</b> SAVILQAAFRGVKTRR	2402
<u>Rattus norvegicus</u>	2321	RYQH <b>LKR</b> ASVVI <b>QKQYRA</b> HRA <b>AE</b> LQR <b>QLFVR</b> QR <b>QAAL</b> TLQAAFRGV <b>KARN</b>	2361
<u>Homo sapiens</u>	2453	QFL <b>HLR</b> KAAITIQSSY---RRLMVKKLQEMQRAAVLIQATFRM <b>YRTYIT</b>	2499
<u>Pan Troglodytes</u>	2453	QFL <b>HLR</b> KAAITIQSSY---RRLMVKKLQEMQRAAVLIQATFRM <b>HRTYIT</b>	2499
<u>Macaca mulatta</u>	2453	QFL <b>QLR</b> KAAITIQSSY <b>CTI</b> RRLMVKKLQEMQRAAVLIQATFRM <b>HRTCVT</b>	2502
<u>Rattus norvegicus</u>	2421	QFL <b>QLR</b> KAAITIQSS <b>F</b> ---RRLMVKK <b>QLQE</b> RHRA <b>ALI</b> QATFRM <b>HRTYVR</b>	2467
<u>Homo sapiens</u>	2550	LREK-HK <b>ASIV</b> IQSTYRMYRQYCFY <b>Q</b> KLQWATKIIQEKYRANKKK <b>QKVFQ</b>	2598
<u>Pan Troglodytes</u>	2550	LREK-HK <b>ASII</b> IQSTYRMYRQYCFY <b>R</b> KLQWATKIIQEKYRANKKK <b>QKAFQ</b>	2598
<u>Macaca mulatta</u>	2553	LREK-HK <b>AAII</b> IQSTYRMYRQYCFY <b>Q</b> KLQWATKIIQEKYRANKKK <b>QKALQ</b>	2601
<u>Rattus norvegicus</u>	2496	<b>SRE</b> E-HR <b>AAVR</b> IQSTYRMHRQRCVY <b>Q</b> LRW <b>AKVI</b> QKKYRANK <b>RR-QDLP</b>	2543
<u>Homo sapiens</u>	2599	HNELKK--ETCVQAGFQDMNIKKQIQE-QHQAAIIQKHCKAFKIRKHLYL	2645
<u>Pan Troglodytes</u>	2599	HNELKK--ETCVQAGFQDMNIKKQIQE-QHQAAIIQKHCKAFKIRKHLYL	2645
<u>Macaca mulatta</u>	2602	HNELKK--ETCVQASFDQDMNIQKQIQE-QHQAAIIQKHCKAFKIRKHLYL	2648
<u>Rattus norvegicus</u>	2544	MAC <b>KEE</b> -- <b>TPLTQ</b> ML <b>FQGS</b> NMA <b>KVQL</b> -QHDA <b>AVI</b> IQKH <b>FR</b> AF <b>KA</b> -----	2585
<u>Homo sapiens</u>	2646	HLRATVVS <b>I</b> QRRYRKLTA <b>VRTQ</b> -----AVICIQSYR <b>GF</b> KVR <b>K</b> DIQ	2686
<u>Pan Troglodytes</u>	2646	HLRATVVS <b>I</b> QRRYRKLTA <b>VRTH</b> -----AVICIQSYR <b>GF</b> KVR <b>K</b> DIQ	2686
<u>Macaca mulatta</u>	2649	HLRATVVS <b>I</b> QRRYRKLTA <b>VRTQ</b> -----AVICIQSYR <b>GF</b> KVR <b>R</b> DIQ	2689
<u>Rattus norvegicus</u>		-----	

Figura 16: Allineamento multiplo di ASPM (aa). Sono mostrate alcune sezioni significative.

Questa differenza è dovuta sia a sostituzioni aminoacidiche (come ad esempio si nota nella parte della sequenza da 1 a 40 oppure nella riga tra 2496 e 2543 in Figura 16, cosiccome alla mancanza di intere stretches di aminoacidi, per cui si produce una proteina più corta nel *Rattus norvegicus*, quando avviene ad una estremità della sequenza, oppure una proteina totalmente diversa, quando avviene al suo interno. Ovviamente, ciò si traduce in una maggiore distanza tra la sequenza umana e quella del *Rattus norvegicus*, appunto, con un numero maggiore di sostituzioni anche positive (vedi Tabella 11).

Tra *Homo sapiens* e *Pan troglodytes* esistono tuttavia numerosi cambiamenti in aminoacidi, di cui fino a 26 positivi, per i quali è ragionevole ipotizzare un ruolo importante nella funzione della proteina e che possono rappresentare i punti di “salto” evolutivo in termini cognitivi che stiamo cercando.

<i>ASPM</i> – allineamento sequenza aa		
<b>ORGANISMO</b>	<b>N° Tot. variazioni</b>	<b>N° Variazioni positive</b>
<i>Pan troglodytes</i>	41	26
<i>Rattus norvegicus</i>	1087	687

Tabella 11: Variabilità ottenuta dall'allineamento di *ASPM* umano (aa) contro *Pan troglodytes* e *Rattus norvegicus*.

Se un numero maggiore di sostituzioni non sinonime rispetto a quelle sinonime è indice di una selezione darwiniana positiva, è possibile affermare che dallo scimpanzé all'uomo è avvenuto un cambiamento sostanziale della sequenza, atto a migliorare la funzione della proteina nella nostra specie.

Quando le sequenze aminoacidiche vengono usate per ricostruire la distanza tra gli organismi, mediante un albero con radice, si ottiene il seguente cladogramma:

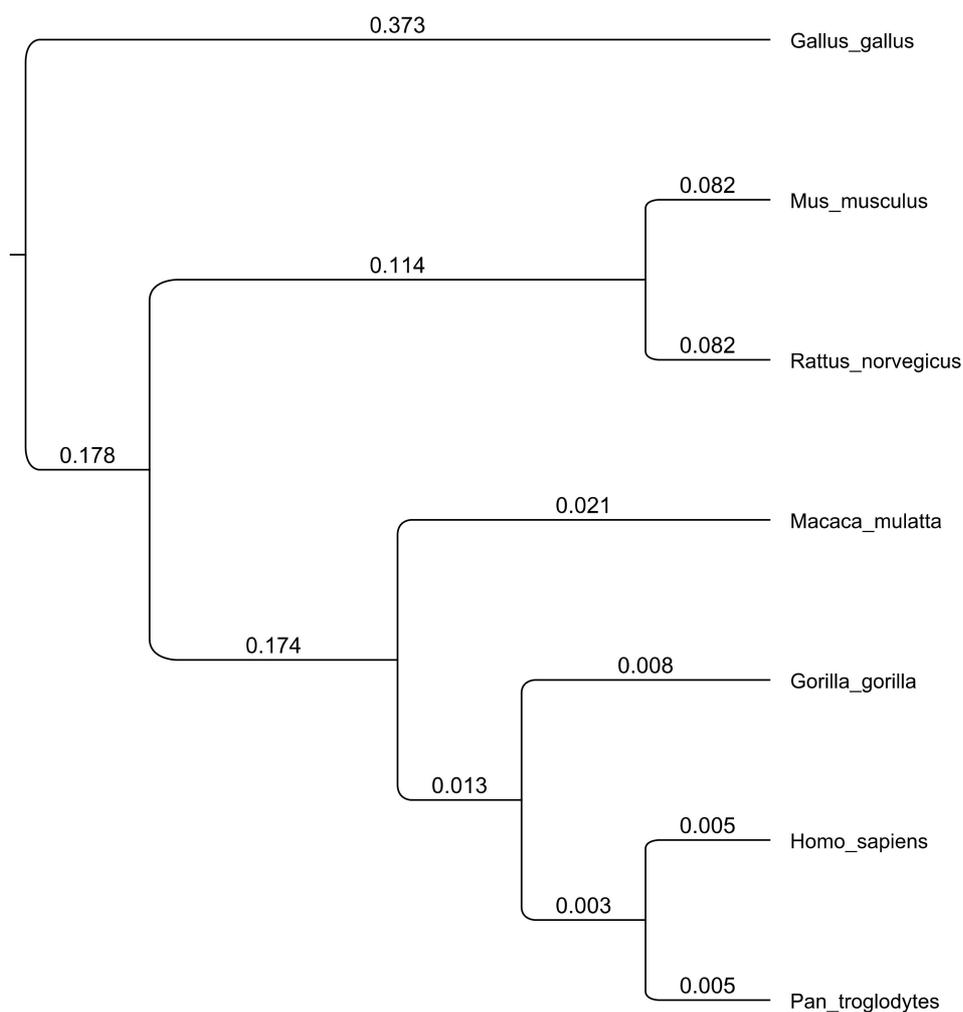


Figura 17: ASPM. Albero filogenetico con radice, costruito con metodo UPGMA.

Come è possibile notare anche dalle matrici delle distanze in Tabella 12 (percentuale di sostituzioni per sito) e in Tabella 13 (percentuale di identità tra le sequenze allineate), maggiore omologia è riscontrabile nel confronto tra primati e tra questi e il genoma del *Mus musculus*, mentre gli organismi più lontani sembrerebbero il *Rattus norvegicus* e più ancora il *Gallus gallus*.

ASPM	<i>H. sapiens</i>	<i>P. troglodytes</i>	<i>M. mulatta</i>	<i>G. gorilla</i>	<i>M. musculus</i>	<i>R. norvegicus</i>	<i>G. gallus</i>
<i>H. sapiens</i>		0.010	0.042	0.016	0.383	0.398	0.712
<i>P. troglodytes</i>	0.010		0.042	0.017	0.386	0.400	0.714
<i>M. mulatta</i>	0.042	0.042		0.044	0.385	0.396	0.711
<i>G. gorilla</i>	0.016	0.017	0.044		0.383	0.400	0.716
<i>M. musculus</i>	0.383	0.386	0.385	0.383		0.163	0.813
<i>R. norvegicus</i>	0.398	0.400	0.396	0.400	0.163		0.815
<i>G. gallus</i>	0.712	0.714	0.711	0.716	0.813	0.815	

Tabella 12: Matrice delle distanze dall'albero filogenetico di ASPM, che riporta la percentuale sostituzioni per sito.

ASPM	<i>H. sapiens</i>	<i>P. troglodytes</i>	<i>M. mulatta</i>	<i>G. gorilla</i>	<i>M. musculus</i>	<i>R. norvegicus</i>	<i>G. gallus</i>
<i>H. sapiens</i>		99%	96%	98%	61%	61%	48%
<i>P. troglodytes</i>	99%		96%	98%	61%	60%	48%
<i>M. mulatta</i>	96%	96%		95%	61%	60%	48%
<i>G. gorilla</i>	98%	98%	95%		60%	60%	47%
<i>M. musculus</i>	61%	61%	61%	60%		83%	42%
<i>R. norvegicus</i>	60%	60%	60%	60%	83%		42%
<i>G. gallus</i>	48%	48%	48%	47%	42%	42%	

Tabella 13: Matrice relativa all'albero filogenetico di ASPM, che riporta la percentuale di identità della sequenza tra una specie e l'altra.

## 5. Conclusioni e sviluppi futuri

L'uso della genomica comparativa per l'analisi dell'evoluzione cognitiva rappresenta un notevole vantaggio rispetto agli approcci più tradizionali, focalizzati sulla comparazione dell'anatomia cerebrale, evidenziando importanti cambiamenti evolutivi difficili da identificare a livello di fenotipo, a causa della infinita complessità dei meccanismi coinvolti nello sviluppo cognitivo. In attesa di saperne di più sul ruolo che sequenze non codificanti come quelle a rapida evoluzione come *HARI* abbiano nello sviluppo cognitivo, possiamo già dire che la comparazione qualitativa del DNA di *Pan troglodytes* ed *Homo sapiens* ha evidenziato grandi differenze in geni estremamente conservati lungo l'evoluzione, come *FOXP2* e *ASPM* e che, probabilmente, il prosieguo degli studi continuerà ad approfondire la base genetica di ciò che ci ha reso umani. L'analisi preliminare di *HARI*, come esempio di gene non codificante proteine, e di *ASPM* e *FOXP2*, come esempi di geni codificanti proteine, ha infatti evidenziato la localizzazione in specifiche regioni di sequenza, tra *Homo sapiens*, *Pan troglodytes* e mammiferi più distanti, non primati, di hotspot di mutazione. Tutte queste regioni sono accomunate dal fatto di essere regioni sottoposte a selezione positiva, per cui si può ipotizzare un loro coinvolgimento in funzioni tipicamente "umane", come quelle caratterizzanti il salto cognitivo tra uomo e suoi parenti più prossimi.

Studi funzionali sulle regioni identificate, assieme al supporto bioinformatico di programmi di predizione di struttura delle proteine, per valutare la conservazione strutturale di domini specifici, potranno chiarire in futuro quali effetti può determinare ogni modificazione, a livello della singola proteina.

Uno spunto di ricerca di natura diversa ma altrettanto interessante e interdisciplinare è rappresentato dalla possibilità di correlare l'analisi genetica delle popolazioni umane alla diffusione geografica delle lingue. Questa seconda possibilità è stata considerata all'interno della tesi a livello preliminare e il frutto di questo studio è riportato nel capitolo successivo.

## Capitolo III

# **Evoluzione dei linguaggi, dei geni e delle popolazioni umane: uno sguardo interdisciplinare**

Il contenuto del seguente capitolo è stato pubblicato come contributo in volume per la Collana “Filosofia e Scienza”, Aracne editrice, come indicato:

VIGNA M., PASSARINO G., DATO S. (2017), *Evoluzione dei linguaggi, dei geni e delle popolazioni umane: uno sguardo interdisciplinare*, in BRUNI A.C., COLONNELLO P., DATO S., PASSARINO G. (a c. di) (2017), *Interdisciplinarità, complessità e questioni epistemiche*, “Filosofia e Scienza”, 15, Aracne editrice, pp. 41-57.

## 0. Introduzione

The nineteenth century witnessed a fundamental transformation of man's perception of itself, and his place in nature, that was perhaps as significant as the Copernican Revolution some 300 years earlier. In a number of disparate fields, including linguistics, geology, paleontology, and biology, a static, theologically inspired view of nature was replaced by a dynamic scientific approach. The most celebrated breakthrough was of course Charles Darwin's discovery that species may evolve gradually into other species through the operation of forces such as natural selection. Furthermore, Darwin, and linguists of the day like August Schleicher, did not fail to note the similarities between their respective disciplines.

Merritt Ruhlen (1991)

Lo studio dell'evoluzione umana è un tema di grande interesse per varie scienze e passa attraverso il contributo prezioso di discipline diverse, dimostrando come l'interdisciplinarietà possa davvero rappresentare la via per una maggiore comprensione delle dinamiche evolutive dell'uomo. Il lavoro decennale del team del genetista Luigi Luca Cavalli Sforza è un esempio storico dei risultati che si possono ottenere sovrapponendo metodi e saperi di natura diversa nella ricostruzione delle origini della specie umana. Biologia, linguistica, statistica, antropologia, demografia: campi di ricerca diversi che grazie agli allievi di Cavalli Sforza, continuano a collaborare e forniscono dati da armonizzare per ricomporre il quadro dell'evoluzione umana. In questo capitolo sono esplorati in particolare i contributi reciproci della genetica e della linguistica allo studio delle migrazioni della specie umana, tracciando un'unione tra fattori biologici e culturali nella distribuzione delle popolazioni nei diversi continenti.

## **1. Cultura e biologia: il ruolo della cultura nell'evoluzione**

Il concetto di cultura è molto dibattuto e soprattutto complesso da racchiudere in una definizione, perché coinvolge molteplici aspetti e definisce in maniera marcata l'evoluzione di una specie. Cultura, infatti, non è prerogativa dell'uomo: se intesa come l'insieme di conoscenze, competenze, modelli di comportamento e scoperte tecnologiche trasmesse da una generazione all'altra o da un individuo all'altro, la cultura è osservabile anche negli animali. Insieme con comportamenti innati e caratteristiche biologiche, la cultura rappresenta quella parte di conoscenza che non si esaurisce nell'arco di una vita, ma si somma di generazione in generazione, trasferendo e accumulando informazione. È ciò che ci permette di accumulare il sapere e trarre profitto dalle esperienze passate senza doverle azzerare e ripetere da capo.

La cultura non è un elemento di interesse soltanto per l'antropologia o la sociologia, ma accomuna campi apparentemente più distanti, come la biologia. Ad esempio, è possibile tracciare un'analogia tra cultura e genoma, come sistemi capaci di conservare le informazioni utili e trasmetterle di generazione in generazione, con lo scopo di migliorare la fitness e l'adattamento dell'individuo o della specie all'ambiente circostante. Le due forze in gioco nel quadro di questa analogia sono quindi la selezione naturale e la selezione "culturale", che opera secondo leggi meno definite, basandosi spesso sulle scelte dei singoli individui ma sottostando comunque all'esame finale della selezione naturale, che favorirà o sfavorirà le scelte culturali adoperate da un individuo o da un gruppo sulla base dei suoi criteri, ossia del vantaggio o svantaggio che esse apportano alla specie

Dall'epidemiologia, invece, la cultura prende in prestito il concetto di trasmissione, in particolare di trasmissione verticale e orizzontale, per descrivere i meccanismi di diffusione della stessa. La cultura, come le malattie, può trasmettersi verticalmente da genitore a figlio, come accade con la trasmissione di caratteri genetici: in questo caso, l'informazione di carattere culturale è trasferita in maniera lenta, poiché l'unità temporale di riferimento

corrisponde ad una generazione. Il legame biologico tra l'individuo di una generazione che "passa" il dato culturale all'individuo della generazione successiva non è strettamente necessario: ad esempio, un figlio adottivo può assimilare il dato con la stessa ricettività di un figlio biologico. In una visione più ampia del legame genitoriale, la trasmissione culturale verticale procede attraverso epoche e popolazioni antiche, passando le informazioni, gli usi e i costumi nei secoli, con l'aiuto della tradizione orale e, in maniera ancora più efficace e duratura nel tempo, della tradizione scritta. La trasmissione orizzontale, invece, è più simile ai processi di contagio tra individui non imparentati tra loro: una sorta di epidemia culturale, con tempi evidentemente più brevi rispetto ai processi di trasmissione verticale, anche se non fulminei come quelli tipici di un contagio diretto.

In generale, le caratteristiche culturali sono più flessibili rispetto a quelle geneticamente determinate: basti pensare alla religione o alla politica come forze determinanti l'identità di un individuo, che possono subire variazioni anche piuttosto rapide, come nel caso di una conversione o di un cambio di ideologia o affiliazione politica. Nonostante ciò, esistono elementi culturali più resistenti alle modifiche esterne, specie se instaurati in periodi critici della formazione di un individuo, come ad esempio i primi anni di vita. Il linguaggio è uno di questi: le lingue apprese dopo la pubertà, per quanto possano raggiungere livelli ottimali di apprendimento, difficilmente possono arrivare ad essere equiparabili alla propria lingua madre, che è invece appresa nei primissimi anni di vita, nei quali si gettano, tra l'altro, le basi della personalità. In sintesi, le caratteristiche culturali apprese nel periodo critico dell'imprinting sono generalmente conservate più a lungo e di conseguenza, più difficili da eliminare o modificare successivamente.

È evidente come la cultura nel suo senso più ampio abbia un enorme impatto potenziale per l'adattamento biologico, che coinvolge l'abilità di apprendimento ovvero la capacità di imparare. Conservare e trasmettere la conoscenza è fondamentale per l'evoluzione della vita a tutti i livelli, perché facilita l'adattamento all'ambiente circostante, aggiungendosi al bagaglio di conoscenze biologiche innate in "dotazione" per ogni organismo. Cavalli

Sforza illumina il concetto di cultura come mezzo di adattamento tramite dei semplici esempi che ne rivelano il potenziale: il senso dell'olfatto e del gusto, per quanto abilmente programmati nell'evoluzione per tenerci lontani dall'ingerire sostanze dannose, non sono sempre sufficienti da soli. In alcuni casi, è necessario apprendere, oltre che sentire attraverso i sensi: la trasmissione di conoscenze botaniche, per esempio, può rivelarsi utile alla sopravvivenza, insegnandoci quali piante siano tossiche o commestibili, laddove i sensi non possono dare ulteriori informazioni. Caratteristiche biologiche e caratteristiche culturali si delineano dunque come elementi egualmente essenziali nella formazione dell'identità di un singolo individuo, di un popolo o di una specie.

Nel caso dell'uomo, è sicuramente il linguaggio a rappresentare l'elemento chiave nel trasferimento della cultura a livello inter- e intra-generazionale, aumentando l'efficienza comunicativa e permettendo così alla nostra specie di adattarsi all'ambiente circostante e velocizzandone l'evoluzione, fino a fornire all'uomo una sorta di "marcia in più" rispetto alle altre specie (Cavalli-Sforza 2001).

## **2. Linguistica e biologia: il ruolo del linguaggio nell'evoluzione**

Il linguaggio rappresenta allo stesso tempo sia un mezzo di diffusione della cultura che un suo prodotto, sebbene non esclusivo: l'abilità linguistica è frutto di processi culturali ma anche biologici, essendo il risultato di scelte operate dalla selezione naturale su caratteristiche innate dell'organismo, di tipo anatomico e neurologico, in primis, come trattato nel primo capitolo di questa tesi.

Oltre che per le caratteristiche morfologiche necessarie allo sviluppo di questa complessa abilità cognitiva, il linguaggio e la biologia hanno importanti elementi di paragone che ne caratterizzano l'evoluzione. Come sottolineato nel paragrafo precedente, l'evoluzione linguistica rappresenta una sottocategoria

della più ampia evoluzione culturale e nonostante le evidenti differenze con il genoma, è possibile parlare di “coevoluzione” dei due sistemi. Per coevoluzione intendiamo che questi fenomeni si sono sviluppati e diramati seguendo traiettorie sufficientemente simili e spesso combacianti, tali da far supporre una reciproca influenza l’uno sull’altro. A tal proposito, Cavalli-Sforza sottolinea come per influenza dei geni sul linguaggio umano non si intenda un’influenza della genetica sulla facilità di apprendimento di una specifica lingua rispetto ad un’altra, ma ci si riferisca invece alla facoltà di apprendere qualunque lingua, in assenza di deficit biologici o ambientali (come ad esempio, rispettivamente, disturbi a livello cognitivo o la mancata esposizione ad una sufficiente quantità di input linguistico).

Un primo esempio di analogia tra i due sistemi si può trovare nel fenomeno della mutazione. Nel genoma, così come nel linguaggio, il meccanismo delle mutazioni è molto rilevante: in entrambi i sistemi, un cambiamento all’interno di un individuo può diffondersi man mano alla popolazione. Le mutazioni genetiche vengono trasmesse di generazione in generazione e possono addirittura stabilizzarsi nel genoma tanto da sostituire la versione originale. Così nel linguaggio, a volte basta l’aggiunta di una singola consonante o vocale per cambiare completamente il significato di una parola (in inglese ad es. wood, legno cambia in wool, lana). Le differenze con i sistemi linguistici sono comunque molto evidenti: mentre il genoma è ben protetto dai cambiamenti (le mutazioni sono rare, spesso sfavorite dalla selezione naturale o protette da meccanismi di riparazione, e si trasmettono solo tramite un legame parentale diretto), le lingue sono molto più flessibili, perché molto più esposte ai cambiamenti apportati dai singoli individui. Inoltre, i cambiamenti o le “mutazioni” linguistiche sono di gran lunga più comuni di quelli genetici e possono essere trasmessi senza necessità di un legame biologico tra individui.

Un'altra forza che influisce in maniera simile sull’evoluzione linguistica e su quella genetica è l’isolamento, che può consistere in una condizione geografica o sociale. Due popolazioni isolate e distinte tendono a differenziarsi sia geneticamente che linguisticamente: il fattore dell’isolamento, che sia dovuto a barriere ecologiche o culturali, riduce le probabilità di contatto tra popolazioni,

ad esempio attraverso il matrimonio, facendo sì che le popolazioni evolvano gradualmente in direzioni diverse. La differenziazione genetica, così come quella linguistica, avviene ad un ritmo graduale e costante e, almeno in linea di principio, le due traiettorie coincidono in molti punti. Ma ancora una volta, date le caratteristiche della variabile “cultura”, meno predicibili e misurabili di quelle della variabile “genoma”, l’evoluzione delle lingue è soggetta a fattori più flessibili di quelli biologici e può quindi subire modifiche più rapide, che deviano la traiettoria comune. Ad esempio, sotto la spinta di forze politiche, la lingua di una popolazione può essere rimpiazzata con un’altra in maniera forzata e anche relativamente veloce, come spesso accade con altri fattori culturali, come la religione di un popolo. Un potere dall’alto, una dittatura o una monarchia, ad esempio, può imporre alla maggioranza una nuova lingua, influenzando così fortemente l’evoluzione linguistica di una popolazione, seppur non influenzando in maniera massiva l’evoluzione genetica della stessa.

Un classico esempio di questo “slittamento” di traiettoria evolutiva è rappresentato dalla lingua ungherese: appartenente al ramo ugrofinnico della famiglia delle lingue uraliche, questa lingua è parlata in territori geograficamente collegati alla famiglia delle lingue indo-europee, come le lingue slave, germaniche e romanze. Quando i Magiari conquistarono l’Ungheria imposero la loro lingua alla popolazione, che parlava la lingua romanza del luogo. A questo forzato e massiccio cambiamento linguistico non corrispose un altrettanto massiccio cambiamento genetico della popolazione, dal momento che i conquistatori rappresentavano solo il 30% della popolazione. L’intrusione genetica non fu dunque delle stesse proporzioni di quella linguistica: infatti, mentre la lingua romanza precedente alla conquista magiara è stata sostituita completamente dalla lingua ungherese, solo il 10% dei geni ungheresi possono essere ricondotti al genoma magiario.

Un esempio del fenomeno opposto, interessante inoltre per gli effetti genetici evidenti dell’isolamento, è invece il complesso quadro linguistico e genetico sardo. Al contrario di ciò che è avvenuto con il magiario, infatti, il sardo è una lingua neolatina, al pari dell’italiano, seppure con le sue evidenti particolarità, dovute alla maggiore conservazione linguistica che ha interessato

l'isola proprio grazie alla sua "insularità", e che l'ha resa uno scrigno linguistico prezioso che conserva tracce antiche dei sostrati che ne compongono la lingua. È grazie a questa caratteristica che il sardo odierno mostra tuttora i segni del sardo antico (o paloesardo) parlato prima della romanizzazione: una lingua non indoeuropea, imparentata con il paleobasco, con il quale ancora condividerebbe elementi fonetici e lessicali. È noto a tal proposito il lavoro di comparazione e ricostruzione del corpus toponomastico sardo a cura del filologo e linguista Eduardo Blasco Ferrer (Ferrer 2011). Il processo di romanizzazione ha causato la scomparsa del paleosardo a favore della lingua romanza che oggi conosciamo: ma a questa forte trasformazione linguistica non corrisponde una altrettanto forte trasformazione nel patrimonio genetico, che si caratterizza per un estremo livello di clusterizzazione dovuto a fenomeni di deriva, come dimostrato, tra gli altri, da un importante studio di genetica di popolazione condotto su 1204 soggetti sardi attraverso la genotipizzazione del cromosoma Y (Francalacci et al. 2013).

La sostituzione totale di una lingua avviene quindi molto più rapidamente e radicalmente sotto l'influsso di un forte potere politico, come d'altronde è avvenuto anche nelle Americhe (caso che analizzeremo meglio più avanti), mentre l'isolamento e una forte identità sociale e culturale rappresentano fattori di conservazione linguistica e genetica.

### **3. L'origine del linguaggio: monogenesi o filogenesi?**

Facendo un passo indietro rispetto all'evoluzione del linguaggio e delle sue molteplici manifestazioni (le lingue), una questione molto dibattuta è stata per gli studiosi l'origine stessa di tutte le lingue. Se è possibile immaginare uno scenario in cui un primo esemplare di uomo, con il suo patrimonio genetico, ha iniziato a migrare e popolare la terra, dando origine a popolazioni diverse con patrimoni genetici sempre più diversi man mano che ci si allontana dal punto d'origine, è altrettanto possibile immaginare un simile scenario riguardo le lingue di queste popolazioni? La questione dell'origine delle lingue è ancora controversa e dibattuta nella comunità linguistica. È stato subito evidente come

le somiglianze tra lingue anche molto lontane siano statisticamente così rilevanti da non poter essere considerate casuali e da richiedere dunque una spiegazione diversa.

Negli anni post-strutturalismo linguistico, l'attenzione dei linguisti si è spostata dallo studio di quegli elementi che differenziano e caratterizzano le lingue alla ricerca di quegli elementi di base che invece ne accomunano il funzionamento. Questi elementi, che ritroviamo in ogni area del linguaggio, dalla fonologia alla semantica, sono definiti "universali linguistici" e coinvolgono sia elementi del linguaggio (in tutte le lingue esiste un elemento per indicare l'emittente del messaggio) che alcune sue proprietà (ad esempio, la necessità di attuare una modifica di una frase dichiarativa per costruirne una interrogativa). Da cosa deriva questa universalità del linguaggio? La tesi più accreditata fonde le intuizioni di ipotesi diverse: una delle più conosciute è sicuramente quella del LAD (Language Acquisition Device), strutturata all'interno della teoria innatista di Noam Chomsky, che definisce l'acquisizione del linguaggio come una capacità innata dell'essere umano (e quindi, potenzialmente, di qualunque lingua), che spiegherebbe a livello biologico l'esistenza degli universali linguistici: l'acquisizione del linguaggio sarebbe connaturata nella struttura e funzionalità del cervello umano (Chomsky 1965).

Un'altra ipotesi, che fusa con la teoria chomskyana rappresenta un solido terreno di ricerca, è l'ipotesi monogenetica. Contrariamente alle originarie direttive della Société de linguistique de Paris, che nel suo statuto costitutivo<sup>8</sup> scoraggiava apertamente i linguisti dal ricercare una lingua universale, antenata comune di tutte le lingue, il fascino di questa ipotesi ha attratto molti ricercatori che nei decenni hanno continuato a lavorare in questa direzione. Sebbene non sia possibile ricostruire questo ipotetico antenato linguistico (come è stato fatto ad esempio con altre proto-lingue, come l'Indoeuropeo), il linguista e antropologo dell'università di Stanford Joseph H. Greenberg propose ed utilizzò il metodo della "mass comparison" per risalire alle

---

<sup>8</sup> Fonte: <http://www.slp-paris.com/spip.php?article5>

parentele tra lingue ad un livello più profondo di quelle stabilite dalle filogenie già conosciute. Il metodo prevedeva, appunto, una comparazione di massa di un grande numero di elementi lessicali appartenenti ad un grande numero di lingue contemporaneamente, sacrificando una visione molto dettagliata in cambio di una prospettiva più ampia sulle connessioni familiari che collegano più lingue in esame. Questo metodo, testato su una serie di termini che per Greenberg costituiscono una sorta di vocabolario essenziale di tutte le lingue, gli ha permesso ad esempio di operare una classificazione generale delle lingue africane, fortemente avversata nel momento in cui fu presentata alla comunità linguistica, eppure oggi largamente accettata ed utilizzata (Greenberg 1948).

I pareri discordanti della comunità linguistica sono giustificati se si tiene conto che il dato linguistico si differenzia da altri tipi di dati scientifici ed è più difficilmente misurabile secondo categorie nette e precise. Inoltre, soprattutto in campo fonologico e semantico (e più lentamente, sintattico), i cambiamenti di una lingua sono frequenti e rapidi, tanto da renderla irriconoscibile nell'arco di poche migliaia di anni, come accade, ad esempio, con le lingue romanze parlate attualmente rispetto al loro comune avo, il latino. Fortunatamente, però, alcune parole o parti del discorso, risultano meno intaccate dal passare del tempo e dal normale ritmo evolutivo di un linguaggio, e tra queste vi sono appunto i termini relativi a quelle aree semantiche considerate più "resistenti" al cambiamento selezionate nello studio comparativo alla base della classificazione di Greenberg.

#### **4. Il caso delle origini dei Nativi Americani**

L'integrazione tra linguistica e biologia ed in particolar modo della genetica è stata fondamentale nella verifica della classificazione di lingue native. Molto interessante a tal proposito è lo studio delle popolazioni dei Nativi Americani. Con questo termine si indicano le popolazioni esistenti sul suolo degli Stati Uniti in epoca pre-colombiana: si tratta di un insieme di gruppi etnici, i cui progenitori sopravvivono tuttora, spesso suddivisi in tribù caratterizzate da una

estrema diversità linguistica e culturale. Storicamente, si stima che il popolamento iniziale del Nord America a partire dall'Asia nord-orientale cominciò approssimativamente 15-18 mila anni fa. Secondo la teoria dello stretto di Bering, i primi uomini arrivarono in Nord America dall'Asia attraverso un corridoio di terra chiamato Beringia, che connetteva la Siberia orientale e l'Alaska e univa i due continenti. Questi uomini si sarebbero poi spostati più a sud fino ad abitare tutto il continente, diversificandosi in migliaia di etnie e tribù differenti. La maggior parte dei nativi presenta caratteristiche somatiche affini alle popolazioni asiatiche: occhi allungati, zigomi sporgenti, quasi assenza di barba e capelli per lo più scuri e lisci. Queste caratteristiche portarono gli antropologi a ipotizzare la loro origine da antichi asiatici che attraversarono lo stretto di Bering nella preistoria.

Questa ipotesi fu poi confermata da successivi studi genetici, nei quali i ricercatori esaminarono campioni del DNA dei Nativi Americani alla ricerca di marker genetici capaci di definire i cosiddetti aplogruppi, ovvero un insieme di varianti genetiche o polimorfismi che li rendono specifici e differenti tra di loro, relativi sia alla linea ereditaria materna (DNA mitocondriale o mtDNA) che a quella paterna (cromosoma Y). Questo tipo di studi si basa sul fatto che a livello genetico, popolazioni che presentano specifici aplogruppi possano essere correlati a variazioni "regionali" e quindi permettano di risalire a popolazioni insediate in specifiche regioni geografiche. A questo scopo, i genetisti hanno analizzato il DNA delle popolazioni del vecchio mondo alla ricerca di un aplogruppo che potesse rappresentare il ramo originario dal quale, attraverso fenomeni miscelanei come l'isolamento, il mescolamento o le mutazioni casuali, le tre macro-famiglie genetiche dei popoli americani potessero essersi diramate e diffuse. Il profilo genetico che più probabilmente può aver originato quelli dei Nativi Americani è risultato essere quello delle popolazioni della Siberia meridionale, probabilmente stanziate ad est del fiume Enisej, confermando così l'ipotesi del popolamento delle Americhe per migrazioni dall'Asia attraverso lo stretto di Bering.

Nel panorama della ricerca italiana, decisivi in tal senso furono gli studi basati sull'analisi della variabilità del genoma mitocondriale da parte dei

ricercatori Torroni e Achilli, allievi del già citato Prof. Luigi Luca Cavalli-Sforza. I ricercatori hanno gettato nuova luce sull'origine degli Indiani d'America, confermando appunto che i loro antenati arrivarono in Nord America in una serie di eventi migratori, scaglionati nel tempo (Achilli et al. 2013). Riassumendo, alla luce delle scoperte antropologiche e genetiche, la prima di queste migrazioni dall'Asia fu di gran lunga la più rilevante e seguendo la costa del Pacifico, dalla Beringia in poche migliaia di anni raggiunse la parte meridionale del Sud America. Le altre non si spinsero così a Sud, ma ebbero un ruolo rilevante nella formazione del patrimonio genetico di una serie di popolazioni della parte settentrionale del Nord America, in particolare degli Algonchini, dei Na-Dene e degli Eschimesi. Grazie alle caratteristiche di eredità uniche del genoma mitocondriale, trasmesso per sola via materna, i ricercatori hanno compreso che oltre alla ormai ben nota espansione dalla Beringia lungo la costa del Pacifico, un secondo gruppo sarebbe penetrato sempre dalla Beringia nel Nord America, attraverso il corridoio di terra ad est delle Montagne Rocciose tra i ghiacciai canadesi. Espansioni successive, flussi migratori e mescolamento tra le popolazioni modificarono ulteriormente il patrimonio genetico delle popolazioni native, portando al popolamento delle Americhe e alla diversità linguistica e culturale ancora oggi riscontrabile nei Nativi Americani.

La genetica, dunque, ha rappresentato una risorsa preziosa per confermare ipotesi e teorie di natura storica ed antropologica. Grazie a questi studi, negli anni furono identificati cinque aplogruppi originari per linea materna (A, B, C, D e X, in Figura 18) e due per linea paterna (C e Q, in Figura 19). Conducendo successive analisi sulla variabilità del DNA dei nativi americani, la genetica ha fornito all'antropologia informazioni preziose anche per una possibile datazione delle tre ondate migratorie dall'Asia, dimostrando come il DNA dei gruppi che popolarono il Nord America (come i Na-Dene) si differenzi da quello delle popolazioni che penetrarono nel Sud e Centro America, perché risalente a migrazioni successive rispetto alle prime due grandi ondate.<sup>9</sup>

---

<sup>9</sup> Fonte: <https://ows.edb.utexas.edu/site/hight-kreitman/land-bridge-theory>

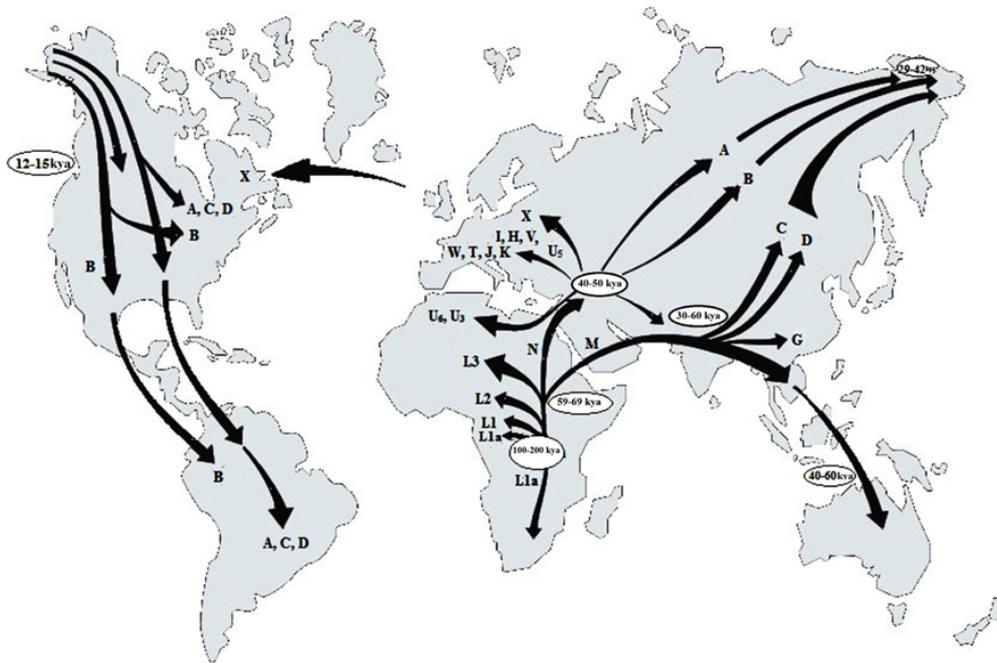
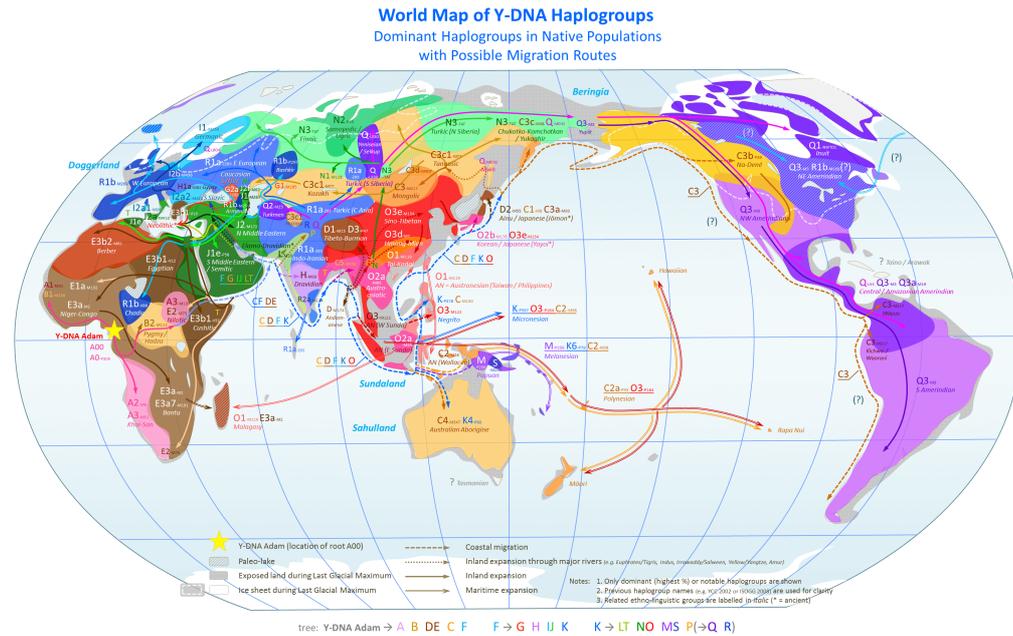


Figura 18: Distribuzione geografica degli aplotipi umani di DNA mitocondriale (indicati con le lettere sulla mappa) e possibili rotte migratorie (indicate con le frecce) a partire dal continente africano. Immagine rielaborata da Witas e Zawicki (2004).

In particolare la presenza del sub-aplogruppo X2a in Nord America ha confermato la migrazione delle popolazioni dall'Asia orientale, considerando che solo i moderni abitanti della Siberia meridionale, dai monti Altai in Occidente fino al fiume Amur in Oriente, presentano lo stesso profilo genetico.

D'altro canto, la distribuzione di mutazioni specifiche del cromosoma Y indica che alcune popolazioni amerindie sono state isolate dopo la colonizzazione iniziale della regione. Na-Dené, Inuit e le popolazioni indigene dell'Alaska, seppur accomunate dalla presenza dell'aplogruppo Q-M242, si distinguono da altri indigeni amerindi per diverse varianti autosomiche e del mtDNA, suggerendo che i popoli che per primi si insediarono all'estremità settentrionale del Nord America e della Groenlandia derivavano da eventi migratori più tardivi rispetto a quelli che penetravano più a sud del continente americano, raggiungendo così una conclusione simile a quella dei linguisti.



**Figura 19: Possibili rotte di migrazione verso le Americhe, degli aplogruppi prevalenti del cromosoma Y nelle popolazioni native d'America: sono indicati i percorsi via mare, quelli attraverso la fascia costiera del Pacifico e la possibile via trans-Atlantica. Immagine da: [https://en.wikipedia.org/wiki/Genetic\\_history\\_of\\_indigenous\\_peoples\\_of\\_the\\_Americas](https://en.wikipedia.org/wiki/Genetic_history_of_indigenous_peoples_of_the_Americas).**

In effetti, in parallelo alla diffusione dei geni, linguisti e antropologi spiegano la presenza di molte lingue diverse come dovuta a migrazioni multiple, lungo la costa del Pacifico e provenienti dal Nord-Est dell'Asia, di popolazioni che parlavano già lingue diverse tra loro. Le lingue indigene d'America presentano una grande variabilità, e il numero di unità filogenetiche (o famiglie linguistiche) consolidate è assai elevato: circa 80 famiglie, più un numero considerevole di lingue isolate. Già il lavoro di Cavalli-Sforza aveva suggerito come l'inclusione delle informazioni linguistiche, assieme a quelle genetiche, era in grado di migliorare notevolmente l'appartenenza di un individuo a un cluster genetico. Come precedentemente discusso, la classificazione delle lingue è stato argomento molto dibattuto; in particolare, la classificazione delle lingue delle popolazioni native americane è stato uno dei casi più contesi per i linguisti, tanto da portare a diverse classificazioni gerarchiche delle lingue Amerindie, tra cui la più famosa, la già citata classificazione di Greenberg (Greenberg 1987). Secondo Greenberg, in accordo con vari studi di natura archeologica sulla ricostruzione delle

migrazioni sul suolo americano, le lingue parlate dalle popolazioni americane pre-colombiane possono essere raggruppate in tre famiglie: Eschimese-auletina, Na-Dene e Amerindia. Il gruppo di Greenberg arrivò a questa classificazione comparando centinaia di lingue tra loro ed esaminando una lista di parole correlate a concetti universali e basilari, come i pronomi personali, le parti del corpo o la natura, aree semantiche che risentono meno dei normali processi di contaminazione linguistica. Questa teoria, come accennato sopra, suscitò non poche polemiche, perché considerata troppo semplicistica rispetto all'alto tasso di diversificazione delle numerose lingue in questione, ma nonostante ciò, risulta tuttora una traccia valida per un'iniziale classificazione delle lingue native, sotto molti punti di vista e soprattutto per uno studio comparato con la genetica. La corrispondenza tra le grandi famiglie linguistiche in cui si suddividono le popolazioni umane e la loro evoluzione genetica è illustrata in Figura 20.

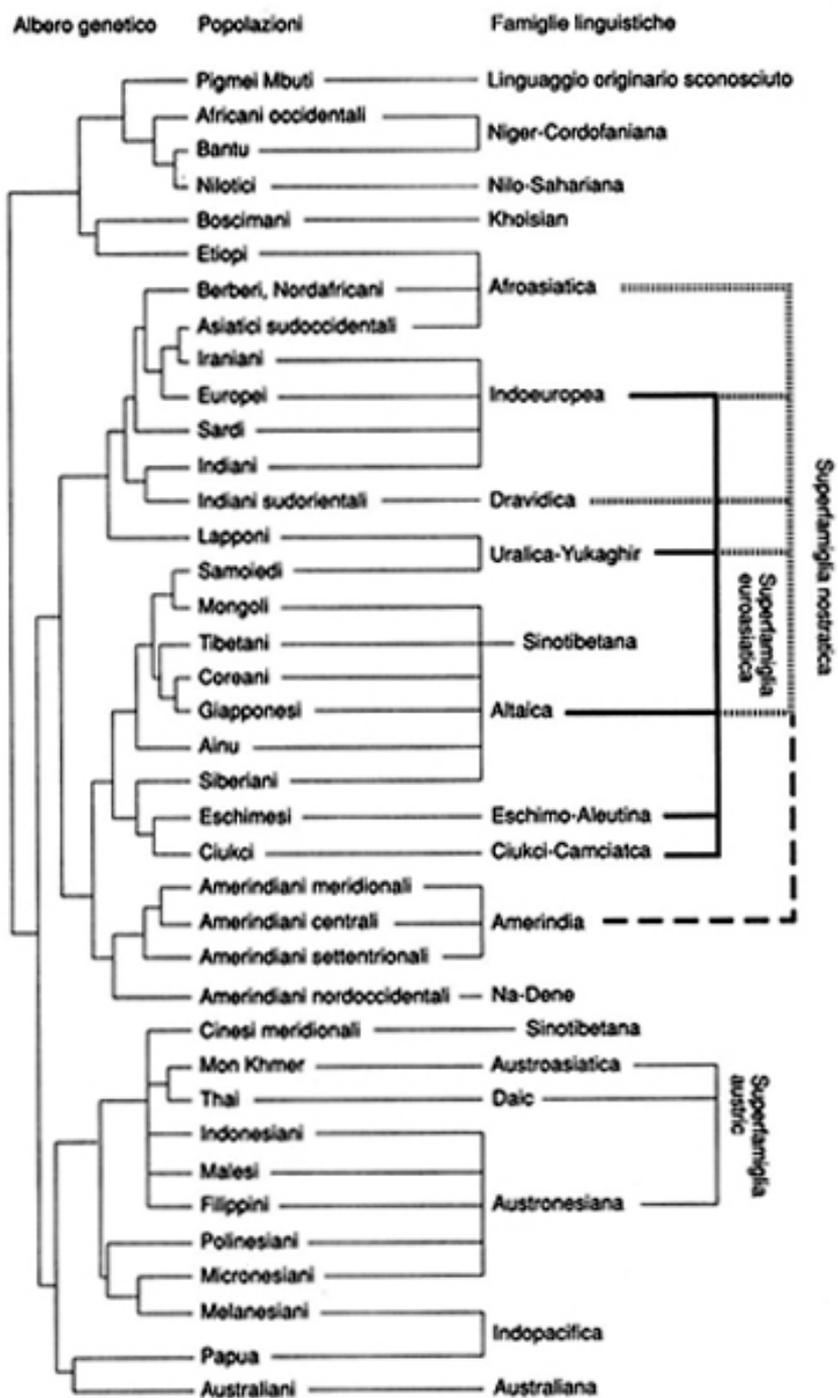


Figura 20: Albero filogenetico che riporta, per ogni popolazione geneticamente distinta, la corrispondente famiglia linguistica. Si noti come all'interno delle popolazioni dei Nativi Americani, sia la genetica che la linguistica differenzino almeno tre filoni presenti, ovvero Amerindi, Eskimo e Na-Dene. Immagine rielaborata da Cavalli-Sforza et al. (1988).

## 5. Conclusioni

Il metodo della comparazione di massa (chiamato oggi anche comparazione multilaterale) ha rappresentato una grande svolta per la linguistica classica. Normalmente, le filogenie venivano indagate attraverso una comparazione a tappeto di un gran numero di elementi di due lingue per volta, in modo da poter escludere con certezza le somiglianze casuali tra due lingue, e definirne così, un passo alla volta, il grado di parentela. Sebbene grazie a questo metodo dettagliato i linguisti abbiano ottenuto grandi risultati (l'intera classificazione attuale dell'Indoeuropeo è avvenuta tramite questo approccio), una pecca è rappresentata dal fatto che questo metodo non permetta di risalire più indietro di questo livello di parentela. Il metodo proposto e attuato da Greenberg, al contrario, pur essendo meno dettagliato e meno inattaccabile da un punto di vista scientifico, ha permesso all'analisi filogenetica di ampliare lo sguardo sulla ricostruzione degli alberi genealogici delle lingue del mondo, e di costruire così un ponte con la ricostruzione degli alberi genealogici umani attraverso gli studi sul genoma e sulle migrazioni umane a partire dal continente africano.

In quest'ottica, genetica e linguistica (ancor più se unitamente a conoscenze di natura geografica e antropologica) diventano alleati per lo studio delle origini dell'uomo, fornendosi informazioni e conferme a vicenda, allo scopo di rafforzare sempre di più la conoscenza attuale della natura umana.

## Conclusioni generali

L'evoluzione cognitiva dell'uomo si presenta come un campo di ricerca affascinante, sul quale possono incrociarsi con effetti più che positivi gli studi provenienti da discipline di natura diversa. La combinazione di saperi, metodi e modelli appartenenti alla linguistica e alla genetica è stato il fil rouge che ha collegato le diverse questioni affrontate in questo lavoro di ricerca. Nell'ambito delle facoltà cognitive superiori, il linguaggio è sicuramente una delle più complesse: se inteso non in senso generico come sistema di comunicazione, ma in senso stretto come sistema combinatorio ricorsivo e potenzialmente illimitato, la capacità di produrre e comprendere il linguaggio è esclusiva del genere umano. Di conseguenza, questa particolare abilità cognitiva si presta bene come oggetto di analisi per la ricerca di quegli elementi del genoma umano che sono mutati nel corso dell'evoluzione dei primati e hanno fatto sì che l'uomo sviluppasse il linguaggio, che così nettamente lo distingue dalle altre specie.

La recente scoperta di aree del genoma fortemente conservate all'interno dell'albero filogenetico umano, anche nel paragone con organismi più lontani, come il topo, che hanno subito una forte accelerazione nel genoma umano rispetto al tasso mutazionale standard, ha aperto prospettive particolarmente interessanti per uno studio in ottica evolutiva e cognitiva. Queste evidenti accelerazioni nell'orologio molecolare rappresentano loci di forte interesse, perché potrebbero indicare punti del genoma nei quali sono avvenute quelle modificazioni essenziali che hanno plasmato il genoma dell'uomo: in breve, ciò che ci rende umani. Tale rapido accumulo di variazioni sembra essere avvenuto sia in regioni non codificanti, come *HARI*, che codificanti proteine, come *FOXP2* e *ASPM*.

*HARI* è l'esempio di sequenza genomica tramandata praticamente identica per circa 300 milioni di anni, per poi subire un'accelerazione mutazionale significativa nei soli 6 milioni di anni che dallo scimpanzé portano alla comparsa dell'uomo. Correlata ad una formazione corretta delle

circonvoluzioni cerebrali, *HARI* ha un ruolo chiave nello sviluppo di un cervello sano e quindi di capacità cognitive nella norma.

D'altro canto, *FOXP2*, scelto per la sua storica (e controversa) correlazione diretta con il linguaggio e con un elevato livello di conservazione tra organismi, mostra interessanti fenomeni di inserzione/delezione nella sequenza proteica, caratterizzanti la discendenza umana e molto probabilmente funzionali per il linguaggio. *ASPM*, correlato alle dimensioni dell'encefalo, mostra invece maggiore tasso di variabilità nucleotidica già all'interno dei primati, candidandola tra le regioni genomiche più interessanti per lo studio del salto cognitivo compiuto dall'uomo.

L'approccio interdisciplinare che ha caratterizzato questo studio si è rivelato efficace per esplorare un campo di ricerca che abbraccia discipline anche molto distanti fra loro, e che generalmente tendono a non collaborare in maniera effettiva, come la genetica e la linguistica. Se la genetica fa luce sulla componente biologica del linguaggio, sulla sua natura di fenomeno evolutivo, allo stesso modo la linguistica fornisce una cornice nella quale collocare le scoperte biologiche che riguardano le capacità linguistiche, chiarendo funzioni e caratteristiche di questa proprietà così "geneticamente" umana. La combinazione di queste due discipline è stata la chiave di volta per affrontare la complessa questione della classificazione delle lingue e delle popolazioni umane. All'esplorazione di questo tema è dedicato il terzo capitolo di questa tesi.

Questo lavoro, con la sua amalgama di modelli e approcci interdisciplinari, vuole aggiungere un pezzo al più ampio puzzle della ricerca di quei fattori che caratterizzano l'evoluzione umana.

## Bibliografia

ACHILLI A., PEREGO U.A., LANCIONI H., OLIVIERI A., GANDINI F., KASHANI B.H., BATTAGLIA V., GRUGNI V., ANGERHOFER N., ROGERS M.P., HERRERA R.J., WOODWARD S.R., LABUDA D., SMITH D.G., CYBULSKI J.S., SEMINO O., MALHI R.S., TORRONI A. (2013), *Reconciling migration models to the Americas with the variation of North American native mitogenomes*, in “Proceedings of the National Academy of Sciences”, Vol. 110, n. 35, pp. 14308-14313.

ALTSCHUL S.F., GISH W., MILLER W., MYERS E.W., LIPMAN D. J. (1990), *Basic local alignment search tool*, in “Journal of molecular biology”, 215, 3, 403-410.

BASSETT D.L., GAZZANIGA M.S. (2011), *Understanding complexity in the human brain*, in “Elsevier, Trends in Cognitive Sciences”, 15, 5.

BEEKMAN M., GLOAG R.S., EVEN N., WATTANACHAIYINGCHAREON W., OLDROYD B.P. (2008), *Dance precision of *Apis florea*—clues to the evolution of the honeybee dance language?*, in “Behavioral Ecology and Sociobiology”, 62, 8, 1259-1265.

BENIAMINOV A., WESTHOF E., KROL A. (2008), *Distinctive structures between chimpanzee and human in a brain noncoding RNA*, in “RNA”, 14, 7, 1270-1275.

BINDER, J.R. (2015), *The Wernicke area. Modern evidence and a reinterpretation*, in “Neurology”, 85, 24, 2170-2175.

BISHOP D.V. (2002), *Putting language genes in perspective*, “TRENDS in Genetics”, 18, 2, 57-59.

BOYD R., SILK J.B. (2000), *How Humans Evolved*, W.W. Norton & Company, Inc., New York.

BOND J., ROBERTS E., MOCHIDA G.H., HAMPSHIRE D.J., SCOTT S., ASKHAM J.M., SPRINGELL K., MAHADEVAN M., CROW Y. J., MARKHAM A.F., WALSH C.A., WOODS C.G. (2002), *ASPM is a major determinant of cerebral cortical size*, in “Nature genetics”, 32, 2, 316-320.

CAVALLI-SFORZA, L. L., PIAZZA, A., MENOZZI, P., & MOUNTAIN, J. (1988), *Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data*, Proceedings of the National Academy of Sciences, 85(16), 6002-6006.

CAVALLI-SFORZA L.L. (2001), *Genes, peoples, and languages*, University of California Press, Oakland, CA.

CHATER N., CHRISTIANSEN M.H. (2010), *Language evolution as cultural evolution:*

*how language is shaped by the brain* in “Wiley Interdisciplinary Reviews: Cognitive Science”, 1, 5, 623-628.

CHENEY D.L., SEYFARTH R.M. (2005), *Constraints and preadaptations in the earliest stages of language evolution*, in “The Linguistic Review”, 22, 2-4, 135-159.

CHOMSKY N. (1965), *Aspects of the Theory of Syntax*, MIT Press, Cambridge MA.

DEACON T.W. (1997), *The Symbolic Species*, W.W. Norton & Company, Inc., New York.

DEDIU D., LADD D.R. (2007), *Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin*, in “Proceedings of the National Academy of Sciences”, 104, 26, 10944-10949.

ENARD W., PRZEWORSKI M., FISHER S.E., LAI C.S., WIEBE V., KITANO T., MONACO A.P., PÄÄBO S. (2002), *Molecular evolution of FOXP2, a gene involved in speech and language*, in “Nature”, 418, 6900, 869-872.

EVANS P.D., ANDERSON J. R., VALLENDER E. J., GILBERT S. L., MALCOM C. M., DORUS S., LAHN B. T. (2004), *Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans*, in “Human Molecular Genetics”, 13, 5, 489-494.

FERRER, E.B. (2011), *Il sostrato paleosardo: fine d'un rebus*, in “Zeitschrift für romanische Philologie” (ZrP), 127, 1, 63-110.

FISHER S.E. (2007), *Molecular windows into speech and language disorders*, in “Folia Phoniatica et Logopaedica”, 59, 3, 130-140.

FRANCALACCI P., MORELLI L., ANGIUS A., BERUTTI R., REINIER F., ATZENI R., PILU R., BUSONER F., MASCHIO A., ZARA I., SANNA D., USELI A., URRU M.F., MARCELLI M., CUSANO R., OPPO M., ZOLEDZIEWSKA M., PITZALIS M., DEIDDA F., PORCU E., PODDIE F., KANG H.M., LYONS R., TARRIER B., GRESHMAN J.B. LI B., TOFANELLI S., ALONSO S., DEI M., LAI S., MULAS S., WHALEN M.B., UZZAU S., JONES C., SCHLESSINGER D., ABECASIS G.R., SANNA S., SIDORE C., CUCCA F., (2013), *Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny*, in “Science”, 341, 6145, 565-569.

FRANCISCUS R.G. (1999), *Neanderthal nasal structures and upper respiratory tract ‘specialization.’*, in “Proceedings of the National Academy of Sciences”, 96, 1805-1809.

GAUTHIER J., JOOBER R., MOTTRON L., LAURENT S., FUCHS M., DE KIMPE V., ROULEAU G.A. (2003), *Mutation screening of FOXP2 in individuals diagnosed with autistic disorder*, in “American Journal of Medical Genetics Part A”, 118, 2, 172-175.

GHAZANFAR A.A., MILLER C. T. (2006), *Language evolution: loquacious monkey brains?*, in “Current Biology”, 16, 20, R879-R881.

GONG X., JIA M., RUAN Y., SHUANG M., LIU J., WU S., GUO Y., YANG J., LING Y., YANG X., ZHANG D. (2004), *Association between the FOXP2 gene and autistic disorder in Chinese population*, in “American Journal of Medical Genetics Part B: Neuropsychiatric Genetics”, 127, 1, 113-116.

GREENBERG J. (1948), *The Classification of African Languages*, in “American Anthropologist”, 50, 1, 24-30.

GREENBERG J. H. (1987), *Language in the Americas*, Stanford University Press, Stanford, CA.

HAUSER M.D., CHOMSKY N., FITCH W.T. (2002), *The faculty of language: what is it, who has it, and how did it evolve?*, in “Science”, 298, 5598, 1569-1579.

HONEYA C.J., THIVIERGEB J.P., SPORNS O. (2010), *Can structure predict function in the human brain?*, in “NeuroImage”, 52, 3, 766–776.

JAKOBSON R. (1973), *Essais de linguistique générale. 2. Rapports Internes et Externes du Language*, Les Éditions de Minuit, Paris.

JAY F., FRANÇOIS O., BLUM M.G. (2011), *Predictions of native American population structure using linguistic covariates in a hidden regression framework*, in “PLoS One”, 6, 1, e16227.

JOHNSON M., ZARETSKAYA I., RAYTSELIS Y., MEREZHUK Y., MCGINNIS S., MADDEN T.L. (2008), *NCBI BLAST: a better web interface*, in “Nucleic acids research”, 36 (suppl\_2), W5-W9.

JUKES T.H., CANTOR C.R. (1969), *Evolution of protein molecules*, in “Mammalian protein metabolism”, 3, 21, 132.

KEARSE M., MOIR R., WILSON A., STONES-HAVAS S., CHEUNG M., STURROCK S., BUXTON S., COOPER A., MARKOWITZ S., DURAN C., THIERER T., ASHTON B., MENTJIES P., DRUMMOND A. (2012), *Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data*, in “Bioinformatics”, 28, 12, 1647-1649.

KLEIN R.G. (2016), *Language and human evolution*, in “Journal of Neurolinguistics”, [http:// dx.doi.org/10.1016/j.jneuroling.2016.11.004](http://dx.doi.org/10.1016/j.jneuroling.2016.11.004)

KRAUSE J., LALUEZA-FOX C., ORLANDO L., ENARD W., GREEN R.E., BURBANO H.A., HUBLIN J.J., HÄNNI C., FORTEA J., DE LA RASILLA M., BERTRANPETIT J. (2007), *The derived FOXP2 variant of modern humans was shared with Neandertals*, in “Current biology”, 17, 21, 1908-1912.

LAI C.S.L., FISHER S.E., HURST J.A., VARGHA-KHADEM F., MONACO A.P. (2001), *A forkhead-domain gene is mutated in a severe speech and language disorder*, in “Nature”, 413, 519–523.

LAITMAN J.T., REIDENBERG J.S., MARQUEZ S., GANNON P.J. (1996), *What the nose knows: New understandings of Neanderthal upper respiratory tract specializations*, in “Proceedings of the National Academy of Sciences”, 93, 10543- 10545.

LI T., ZENG Z., ZHAO Q., WANG T., HUANG K., LI J., LI Y., LIU J., WEI Z., WANG Y., FENG G., HE L., SHI Y. (2013), *FoxP2 is significantly associated with schizophrenia and major depression in the Chinese Han Population*, in “The World Journal of Biological Psychiatry”, 14, 2, 146-150.

LIEBERMAN P. (1992), *On Neanderthal Speech and Neanderthal Extinction*, in “Current Anthropology”, 33, 4, 409-410.

LIEBERMAN P. (1993), *On the Kebra KMH 2 Hyoid and Neanderthal Speech*, in “Current Anthropology”, 34, 2, 172-175.

LIEBERMAN P., MCCARTHY R. (2007), *Tracking the evolution of language and speech: Comparing Vocal Tracts to Identify Speech Capabilities*, in “Expedition: The magazine of the University of Pennsylvania”, 49, 2, 15-20.

MANTEGNA R.N., BULDYREV S.V., GOLDBERGER A.L., HAVLIN S., PENG C.K., SIMONS M., STANLEY H.E. (1994), *Linguistic features of noncoding DNA sequences*, in “Physical review letters”, 73, 23, 3169.

MANTEGNA R.N., BULDYREV S.V., GOLDBERGER A.L., HAVLIN S., PENG C.K., SIMONS M., STANLEY H.E. (1995), *Systematic analysis of coding and noncoding DNA sequences using methods of statistical linguistics*, in “Physical Review E”, 52, 3, 2939.

MCCLELLAND J.L., ELMAN J.L. (1986), *The TRACE model of speech perception*, in “Cognitive psychology”, 18, 1, 1-86.

MEKEL-BOBROV N., GILBERT S.L., EVANS P.D., VALLENDER E.J., ANDERSON J. R., HUDSON R.R., TISHKOFF S.A., LAHN B.T. (2005), *Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens*, in “Science”, 309, 5741, 1720-1722.

MESSING J., BHARTI A.K., KARLOWSKI W.M., GUNDLACH H., KIM H.R., YU Y., WEI F., FUKS G., SODERLUND K.A., MAYER K.F.X., WING R.A. (2004), *Sequence composition and genome organization of maize*, in “Proceedings of the National Academy of Sciences of the United States of America”, 101, 40, 14349-14354.

MICHENER C.D., SOKAL R.R. (1957), *A quantitative approach to a problem in classification*, in "Evolution", 11, 2, 130-162.

NCBI RESOURCE COORDINATORS (2016), *Database resources of the National Center for Biotechnology Information*, in "Nucleic Acids Research", 44, D7–D19.

POLLACK R. (1994), *Signs of life. The language and meanings of DNA*, Penguin Books, London.

POLLARD K. (2009), *What makes us human?*, in "Scientific American", 300, 5, 44-49.

POLLARD K.S., SALAMA S.R., KING B., KERN A.D., DRESZER T., KATZMAN S., SIEPEL A., PEDERSEN J.S., BEJERANO G., BAERTSCH R., ROSENBLOOM K.R., KENT J., HAUSSLER D. (2006), *Forces shaping the fastest evolving regions in the human genome*, in "PLoS Genet", 2, 10, e168.

POLLARD K.S., SALAMA S.R., LAMBERT N., LAMBOT M.A., COPPENS S., PEDERSEN J.S., KATZMAN S., KING B., ONODERA K., SIEPEL A., KERN A.D., DEHAY C., IGEL H., ARES M.JR., VANDERHAEGHEN P., HAUSSLER D. (2006b), *An RNA gene expressed during cortical development evolved rapidly in humans*, in "Nature", 443, 7108, 167-172.

RIZZOLATTI, G., & SINIGAGLIA, C. (2006), *So quel che fai: il cervello che agisce e i neuroni specchio*, R. Cortina ed..

ROSENFELD D.B. (2013), *Neurology of speech and language*, in "Animal Models of Speech and Language Disorders", Springer New York, 3-12.

RUHLEN M. (1991), *A guide to the world's languages: Classification* (Vol. 1), Stanford University Press, Stanford, CA.

SAITOU N., NEI M. (1987), *The neighbor-joining method: a new method for reconstructing phylogenetic trees*, in "Molecular biology and evolution", 4, 4, 406-425.

SPOCTER ET AL. (2010), *Wernicke's area homologue in chimpanzees and its relation to the appearance of modern human language* in "Proceedings of the Royal Society of London" B, 277, 1691, 2165-2174.

SPORNS O. (2011), *The human connectome: a complex network*, in "Annals of the New York Academy of Sciences", 1224, 1, 109-125.

TOBIAS P.V. (1987), *The brain of Homo habilis: A new level of organization in cerebral evolution*, in "Journal of Human Evolution", 16, 741-761.

TOBIAS P.V. (1998), *Evidence for the Early Beginnings of Spoken Language*, in "Cambridge Archaeological Journal", 8, 72-78.

TOLOSA A., SANJUÁN J., LEAL C., COSTAS J., MOLTÓ M. D., DE FRUTOS R. (2008), *Rapid evolving RNA gene HAR1A and schizophrenia*, in “Schizophrenia research”, 99, 1, 370-372.

YATES A., AKANNI W., AMODE M.R., BARRELL D., BILLIS K., CARVALHO-SILVA D., CUMMINS C., CLAPHAMP., FITZGERALD S., GIL L., GARCÍA GIRÓN C., GORDON L., HOURLIER T., HUNT S.E., JANACEK S.H., JOHNSON N., JUETTEMANN T., KEENAN S., LAVIDAS I., MARTIN F.J., MAUREL T., MCLAREN W., MURPHY D.N., NAG R., NUHN M., PARKER A., PATRICIO M., PIGNATELLI M., RAHTZ M., SINGH RIAT H., SHEPPARD D., TAYLOR K., THORMANN A., VULLO A., WILDER S.P., ZADISSA A., BIRNEY E., HARROW J., MUFFATO M., PERRY E., RUFFIER M., SPUDICH G., TREVANION S.J., CUNNINGHAM F., AKEN B.L., ZERBINO D.R., FLICEK P. (2015). *Ensembl 2016*, in “Nucleic acids research”, 44(D1), D710-D716.

WHISHAW I.Q., SACREY L.A.R., TRAVIS S.G., GHOLAMREZAEI G., KARL J.M. (2010), *The functional origins of speech-related hand gestures*, in “Behavioural brain research”, 214, 2, 206-215.

WITAS H. W., ZAWICKI P. (2004), *Mitochondrial DNA and human evolution: a review*, in “Anthropological Review”, 67, 97–110.

WYNN T. (1998), *Did Homo erectus Speak?*, in “Cambridge Archaeological Journal”, 8, 78-81.

XU K., SCHADT E.E. POLLARD K.S., ROUSSOS P., DUDLEY J.T. (2015), *Genomic and network patterns of schizophrenia genetic variation in human evolutionary accelerated regions*, in “Molecular biology and evolution”, 32, 5, 1148-1160.

ZHANG Z.C. (2004), *Large-scale character-usage statistical analysis on Chinese ancient characters*, in “Proceedings of Third Symposium on Databases of Chinese Writing and History”, National Committee for Programming Publishing House and National League of Ancient Book Publishing House, Beijing, China, 107–119.

ZHANG J. (2003), *Evolution of the human ASPM gene, a major determinant of brain size*, in “Genetics”, 165, 4, 2063-2070.

ZUBERBÜHLER K. (2006), *Language evolution: The origin of meaning in primates*, in “Current biology”, 16, 4, R123-R125.

## **Appendice: Le reti del linguaggio nei processi cognitivi emozionali in Shakespeare**

Alcuni degli argomenti trattati nella tesi di laurea magistrale sono stati approfonditi nel primo anno di dottorato e ampliati in collaborazione con il gruppo ESG (Evolutionary Systems Group) dell'Università della Calabria, e riassunti in un articolo dal titolo "*Hamlet and Othello wandering in the web: inferences from network science on cognition*", presentato a COGNITIVE 2016, "The Eighth International Conference on Advanced Cognitive Technologies and Applications". L'articolo è allegato di seguito e disponibile su [www.thinkmind.org](http://www.thinkmind.org)<sup>10</sup>.

In questo articolo, la scienza delle reti è stata utilizzata come principio esplicativo e strumento metodologico per ottenere una comprensione più profonda della cognizione emotiva, all'interno delle tragedie shakespeariane di Otello e Amleto. E' stato costruito un lessico emozionale dei termini ritrovati e analizzati nei due testi letterari, a partire dal quale sono state create reti semantiche esemplificative. Ogni nodo della rete rappresenta un termine emozionale, mentre la presenza di un link tra due nodi indica una relazione sinonimica tra gli elementi lessicali. Ulteriori termini del lessico emozionale sono stati analizzati allargando la ricerca alla repository di WordNet, allo scopo di ricostruire la complessa struttura di questa sezione di lessico e i suoi molteplici collegamenti con i sensi, il linguaggio del corpo, e la cognizione.

I risultati delle varie simulazioni, confrontati con i dati estratti manualmente, hanno confermato l'ipotesi che la struttura del comportamento emotivo dipende da un'interazione tra dinamiche cognitive su vari livelli, dimostrando inoltre la validità dell'utilizzo della scienza delle reti, come mezzo per confrontarsi con dati reali attraverso le simulazioni al computer, con l'obiettivo di capire le regole generali che operano sui processi cognitivi.

---

<sup>10</sup> Scaricabile al link seguente:

[https://www.thinkmind.org/index.php?view=article&articleid=cognitive\\_2016\\_6\\_30\\_40045](https://www.thinkmind.org/index.php?view=article&articleid=cognitive_2016_6_30_40045)

# Hamlet and Othello Wandering in the Web

## Inferences from Network Science on Cognition

Francesca Bertacchini

Department of Mechanical Engineering, Energy and  
Management  
University of Calabria  
e-mail: fbertacchini@unical.it

Patrizia Notaro, Mara Vigna, Antonio Procopio,  
Pietro Pantano, Eleonora Bilotta

Department of Physics  
University of Calabria  
e-mails: patrizia.notaro@unical.it, mara.vigna@unical.it,  
antonio.procopio@unical.it, pietro.pantano@unical.it,  
eleonora.bilotta@unical.it

**Abstract**—Network theory was used to gain a deeper understanding of emotional cognition, pretending that Shakespearian tragedies of Othello and Hamlet were life-like linguistic contexts. A manual segmentation of both plays was carried out for defining a lexicon of emotional words represented by networks. Using ad hoc developed computational methods, further instances of the emotional lexicon networks were compared with WordNet's manually extracted data. The results revealed organizations of emotional terms' neighborhoods, evidencing the emergence of emotional patterns, with symmetries and breaches of symmetries, thus giving a strong support to comprehend the dynamics operating on speech production cognitive processes.

**Keywords**—language; Shakespeare; network; semantics.

### I. INTRODUCTION

Network science considers the organizational principles of complex structures found in different fields of research, from physics to biology and social sciences. Its starting point is the graph, a mathematical structure made of nodes connected by links. Network science has recently contributed to the study of language as a real-life problem, by analyzing the statistical properties of words associations [1], thesauri [2], syntactic dependencies networks [3], modeling the properties of these structures [1] and proposing abstract models for increasing the general knowledge of language functions. Though networks were already associated to cognition using different methods [4][5], network science greatly expanded cognitive science domains, improving knowledge about brain connectivity [6], representing semantics in the human brain [7] and studying semantic organization of memory [1][8][9]. Despite the huge number of studies aiming at analyzing language networks, cognitive processes underlying these networks have only been studied at the phonological level [10], for spoken word recognition [11], or failed lexical retrieval [12], thus leaving away the study of lexical semantic organization. In this paper, network theory was used as an explanatory principle and a methodological tool to gain a deeper understanding of emotional cognition into the Shakespearian tragedies of Othello and Hamlet as the best exemplar of narrative text, pretending they were real life situations. We considered

mental lexicon as the parameter space of cognitive processes related to speech production. We considered the mental lexicon as a small-world network, and lexical retrieval as a search through that network, similar to the PageRank algorithm [13] searches through the World-Wide Web. Our aim was to represent the organization of words in memory in order to develop a more comprehensive view of emotional cognition dynamics and to see how the emotional speech could be understood and represented as dynamical system by network theory. Data obtained through the manual method was subsequently compared and analyzed by computational systems searching through WordNet, to examine the growth of semantic networks with networks obtained by manually segmenting both plays. Networks organized in structures that span from simple to complex mathematical configurations were detected. They represented patterns of semantically similar neighborhoods of emotional words, showing symmetries and motifs [14]. The organization of these semantic networks has been considered as dynamical attractors that work on the connections of semantic similarity. If two nodes share many common properties then it will be very likely that these two nodes share several links. Their concepts are definitely connected. The attractors also showed different dynamical patterns, used as a representation of emotional dynamics in the plays' characters and of productive processes in the verbal production.

The paper is organized as follows. After a subsection on the theoretical approaches to cognitive processes involved in emotional speech and behavior, and basic notions about network science, the methods used to analyze text as real life situations are described in Section II, specifying the two different types of analysis: the manual and the computational one. Results on the main theme of this paper are then deeply discussed in Section III. Conclusions and further developments are presented in Section IV, closing the work.

#### A. Basic assumptions on emotions and network science

Emotions as psychological occurrences have exclusive traits, they are embodied, displayed into stereotyped behavioral patterns of facial expressions and conduct, stimuli driven and related to all aspects of cognition. Besides, emotions are significantly connected to the social-cultural situations in which they occur [15][16]. According to many researches [17][18][19][20], a stimulus in the environment

triggers a chain reaction in humans. According to Frijda [21], emotions trigger cognitive structures, which are characteristic of a given emotional experience, by a process known as cognitive tuning. Human language represents the means to reach a deep knowledge about emotions and body language allowing furthermore to understand the other participants cognitive mental processing of ideas and concepts [22]. Linguistic properties have often been analyzed through network science, interpreting language as a graph. A classic definition of network is a set of nodes connected by edges [23][24][25]. Considering  $G$  as an ordered set  $G = \{N, E\}$ , where  $E = \{e_{ij} = (x_i, x_j) \mid x_i, x_j \in N\}$  is a set of paired elements of  $N$ , among which a relationship is established. The components belonging to  $N$  are called nodes, or vertices, while the components belonging to  $E$  are called edges. At the words level, the single term is considered a node, while the connections among words is the synonymic relation. Crucial features of semantic organization can be detected and explained with networks models. To analyze the size of a semantic network, the total number of nodes ( $n$ ) and edges ( $m$ ) is calculated. Size in terms of nodes may also be critical for the structure of word relationships. Another significant feature of semantic networks is the degree of a word ( $k$ ), which identifies its number of links and highlights the weight that a single term has as a source of ties (possible connections and influences of a word with other words in the whole structure). For very large networks, the degree distribution deviates significantly from the Poisson distribution, as highlighted for the World Wide Web [24], for the Internet [26], and for metabolic networks [27]. The work of Barabási et al. [23] defines these networks as scale free. This structure includes few words which are highly interconnected nodes (hubs) participating to a large number of interactions and other terms connecting to the network by following a preferential attachment to these hubs. The mathematical approach based on network science's principles used in this experiment represented a useful tool to model and analyze cognition.

## II. METHODS

This Section describes the two different methods used.

### A. How to

In order to obtain a cognitive organization of emotional terms, these steps were followed: starting from a text a list of emotional terms was manually extracted; from this list several networks were built as described in the next subsection; finally a comparison between these networks with the network extracted from WordNet [28][29][44] was made to depict and analyze how the emotions are configured in the text.

### B. Manual research method

Two different methods were used for analyzing emotional cognition. The first one employed a lexicon of about 2000 emotional words, drawn directly and manually from the tragedies of Hamlet and Othello, referring to the five basic emotions of 'anger', 'love', 'sadness',

'joy'/'enjoyment' and 'fear'. A first linguistic modeling was applied to this data, organizing them in tables divided by emotion and five grammatical categories: verbs, names, adjectives, adverbs and sentences (i.e., metaphors or idiomatic expressions). The annotation system that was used is showed in Figure 1.

From this first step, a second kind of modeling was built: the linguistic data was transformed in numerical data, where every element was associated to its synonyms, in order to create five separated undirected graphs, containing all the words selected for each emotion, with links to every synonym in the text. The next step was to complete the nodes in the networks, using different kinds of codes:

- a color code to indicate grammatical categories;
- a numerical code for the same purpose (i.e., Names 1, Adjectives 2): it allowed the extraction of quantitative information about the distribution of the nodes in the text;
- an alphabetical code of small letters to indicate acts and scenes (i.e., a=sc.1);
- an alphabetical code of capital letters indicating the character speaking.

### C. Computational Research Methods

In order to define a computational model inspired to biological cognition, the psychological resource WordNet was downloaded from Princeton University's website [44]. Developed by George Miller and collaborators [28][29], the resource is an attempt to capture psycholinguistic theory within a linguistic system, for defining and modeling the meaning of words and associations between meanings. The system was then transformed into a network, presenting 139,999 nodes (at the time the system has been downloaded for the present study). The following computational models, completely adherent with the WordNet system, were developed:

- R1\_Antonyms collected words which stand in gradual opposition with a chosen term, have complementary meanings, or are the opposite, as different and symmetrical terms;
- R2\_Hypernyms created the hierarchical structure of the words, taxonomically classifying them into types and subtypes;
- the function R3\_Hyponyms collected terms correlated by the 'IT IS A' relationship;
- R4\_Entailment identified all the terms that were in some way required, from a conceptual point of view, by the starting term;
- R5\_Similar collected semantic resemblance between terms;
- R6\_Synonyms searches for words with share partial sharing of semantic content, denotative and connotative meaning, are perfectly interchangeable;
- R7\_Related identified the terms correlated to the one chosen for the study;
- R8\_Overview identified a mix of all the above semantic dynamics for the term under study, for all syntactic categories.

The above detailed formal models were implemented in the software (SWAP), which performs search in the WordNet database, and then calculates a network in which nodes are connected by edges according to a chosen R.

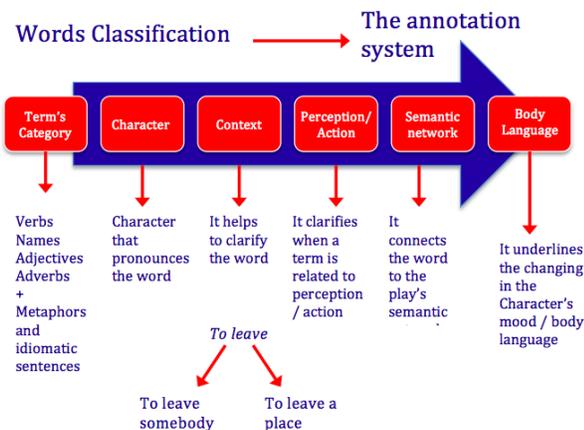


Figure 1. The annotation system used for creating the dictionary and the emotional words' classification system.

We thought of these computational models as 'emotional-terms-hunters bots', or search engines (software robots) of emotional terms, virtually able to 'hunt' all the semantic connections related to emotional lexicon. Investigating the system with different computational models allowed the detection of distinct characteristics due to the particular network topology: choosing to move only within paths of the network that converged to few nodes or a single node, the number of nodes visited decreases, as in the case of the computational model R3\_Hyponyms. But, if the paths diversified or specialized (as in the case of computational models such as R2\_Hypernyms and R6\_Synonyms), the number of visited nodes increased. This could suggest that the neighbors of synonyms and hypernyms are more densely populated while those of hyponyms are sparser. The total network held steady, while the visited part changed. But, if the visited nodes were considered as possible states of the system, with the computational models seeking synonyms and hypernyms, then the number of states eligible by the system was wider than in the case of the system using the computational model to search for hyponyms.

### III. RESULTS AND DISCUSSION

Section III highlights and discusses major results.

#### A. Manual Research Network Results

The manual annotation method of emotional terms disclosed several features of emotional cognition for every character (see an example in Table I).

Considering parameters  $\lambda$  and  $\gamma$ , where  $\lambda$  is the ratio between the average path length (L) and the path length of the equivalent random network and  $\gamma$  is the ratio between the network clustering coefficient and the random one clusterization, the small-world index  $\sigma$  is then calculated as the ratio between the clustering coefficient and the path length ratios. If  $\gamma \gg 1$ ,  $\lambda \approx 1$ ,  $\sigma > 1$  [30], the network

has a small-world structure. Statistical data reported in Table II shows such parameters for emotional terms in both tragedies: the emotional language in Hamlet follows a small-world structure and the same is for 'fear' and 'joy' in Othello. What emerged was that verbs related to the main emotions, for example 'to fear' in both tragedies, were generally pronounced by the main characters in every act. Positive emotion words were rare and they generally appeared only in negative contexts. The verb 'to love' and the substantives 'passion' and 'happiness' in the tragedy of Othello, were always related to the idea of 'hate', and to the character of Iago. The inner emerging dynamics are invisible to a classical approach, the association of the character of Iago to the idea of 'hate' and 'anger' could be rather predictable, but the semantic slip of the idea of 'hate' hiding in term such as 'passion' and 'love' are not completely detectable without considering different level of complexity and different computational models, such as the one proposed to the present study. In Hamlet, the main character, according to his dramatic role in the play, had never pronounced the most meaningful substantives of the network of the term 'enjoyment'. Later, the computational approach allowed to calculate and to visualize the networks dynamics in Othello and Hamlet, detecting the main organization involved in this process, compared to the growth model and the communities of the WordNet networks.

#### B. Computer Simulation Results: Network Statistics for Emotional Terms

The statistical values for the analyzed sample of networks, for the 5 emotional terms, till 5 levels, together with the communities of each network, were calculated. As for Othello and Hamlet, such values revealed a small-world structure and a broad scale behavior (see as example Tables III and IV). It was observed that the relationship between nodes varies from one term to the next, with a number of similarities among terms. It should be noted that for certain terms the overall number of nodes and edges is very high, and the number of nodes and number of edges are correlated. It also stands out that the term 'fear' is the one with the absolute highest number of nodes and edges, at all the levels that were taken into consideration (Table V and Figures 2, 3 and 4).

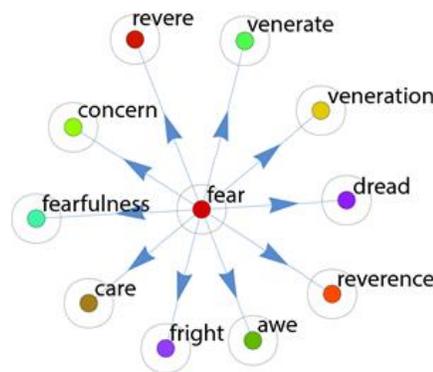


Figure 2. First level of the experimental run with the computational system R8\_Overview.



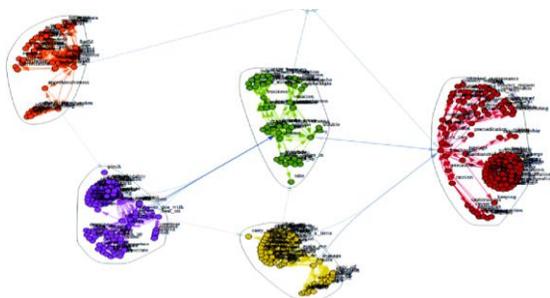


Figure 4. Third level using the function R8\_Overview, for the term ‘fear’, with a greater growth rate of both the network and the communities.

1) *The Growth of the Network Structure Through the Function R8\_Overview*

Networks analysis was carried out using the software Mathematica. The program to achieve communities of networks with this software is given in the Appendix C of supplementary materials [43]. This network is composed by a mix of names and verbs, related to the term ‘love’ in all its varieties of manifestation and behavior.

The first levels of the system and the communities of the network (Figure 5) showed strongly connected terms, sharing mutual meaning relationships: links are bidirectional. In fact, there is a clustering of terms based on the communities to which they belong. The verbs of the first experimental run for the word ‘love’ have reciprocal links, which create an emerging structure with terms very close from the semantic point of view. The emerging structure is a complete graph (see following paragraphs for details). The second community presented different relationships, less dense and more distant, and moreover, with a different spatial organization. From a cognitive perspective, these differences in topology may imply a greater clustering of terms (denser neighborhoods [10]), resulting in a higher availability of the terms of one community compared to another.

In the second level of the network, the number of communities arose from 2 to 4 and, as before, the dynamical reorganization of meanings was complex. The denser community didn’t change throughout the growth processes, while the other community acquired new terms. These additions let the three new communities emerge, i.e., three new network topologies. A growth that was significantly different from Barabási and Albert’s preferential attachment [23] was observed in this case, given that each node had the possibility to develop in multiple dimensions of the semantic space.

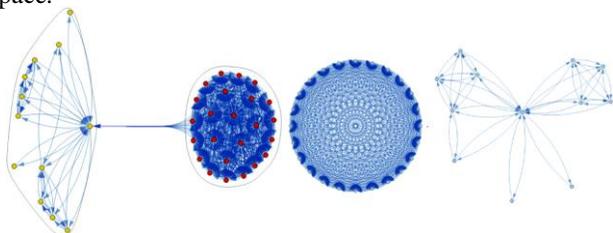


Figure 5. Communities of the first growth level of the term ‘love’. The whole network is showed on the left, and the communities on the right.

At level 4, this phenomenon kept getting more and more complex: here, the computational system collected a network composed by 290 nodes and 2546 edges, structured over the 13 communities.

Many of the communities among the 13 formed in this level, though still belonging to the term ‘love’, created in turn second level structures, which reproduced the structures observed at the first depth level of search algorithm. Topologically, similar structures emerged in this network, with the same spatial organization, but with different terms.

These structures started to grow in neighborhoods characterized by highly connected terms, or in neighborhoods, which were globally poorly dense, but locally dense. From this analysis, generative growth models were detected: they worked combining their own growth mechanism to other existing ones, thus realizing creative cognitive processes of terms association.

2) *The Associative Dynamics Obtained through the Functions R2\_Hypernyms, R3\_Hyponyms and R6\_Synonyms*

The simulations with the functions R2\_Hypernyms, R3\_Hyponyms and R6\_Synonyms revealed interesting behaviors of terms organizations. Again, using ‘love’ as an example, simulations were run for the three functions at level 4, in order to understand the general dynamics of aggregation and the related network topologies that sustained such functions.

For the function R2\_Hypernyms, the computational simulations gave the network (Figure 6) showing behaviors of terms aggregation, which were similar to the synonyms. As for the function R3\_Hyponyms, instead, the topological organizations of the emerging communities were different. The relationship between nodes and edges had the same nature for synonyms and hypernyms, changing for the hyponyms network.

As regards the function R2\_Hypernyms, it showed behaviors of terms aggregation that are similar to the synonyms, as can be seen in Figure 7.

As for the function R3\_Hyponyms, instead, the topological organizations of the emerging communities are different, as showed in Figure 8. In this last network, communities were not linked: hypernyms network showed a greater amount of communities as for the other two functions, but these communities were composed by terms settled in non-densely populated neighborhoods, which somehow linked themselves to the key terms of every detected sub- community.

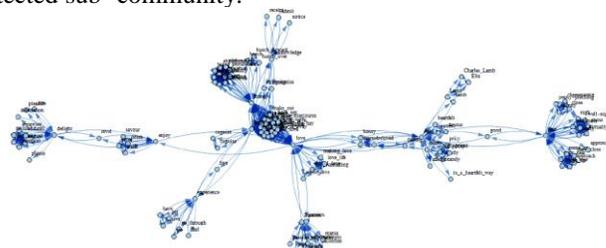


Figure 6. Network obtained by R2\_Hypernyms starting from the term ‘love’, and realizing a structure composed by 125 nodes and 1142 edges.

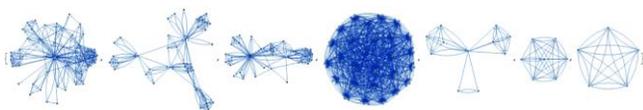


Figure 7. Communities of the hypernyms network of 'love'. The structure had 168 nodes and 1480 edges.

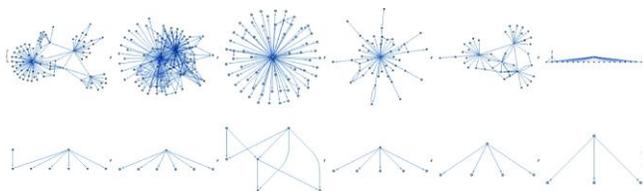


Figure 8. Hyponyms network, presenting 370 nodes and 758 edges. The aggregational structures are limited to few terms, often not interconnected.

What emerged were structured dynamics hiding in both computational and cognitive modeling, similar to a mental lexicon organization, shaped by the processes of clusterization according to semantic fields and grammatical categories or according to a set of stimulus-reaction families, which activated different cognitive spreading activation dynamics of meaning.

The clusterization of emotional terms at the first level appeared as a structured set of three main collections linked to the central term love. The emerging idea is that terms clustered according to grammatical categories, where verbs topologically and quantitatively represented the main nucleus and two outlying clusters composed by adjectives and substantives fork, changing the main geometry. At the second level of growth, the communities arose with the grammatical clusterization of substantives and related hypernyms, hyponyms and synonyms. The third level totally broke the initial topology of the network, creating new ramifications, which semantically blended. The term love reached its whole geometry covering with brand-new meanings. The emerging grammatical phenomenon is the absence of relevant new verbs, in fact, newly appeared categories were substantives, adjectives and adverbs. The existing set of adjectives near the meaning of 'dear' increased its cluster, strengthened its connotation and mixed grammatical categories. At this level, emerging phenomena focused on specific growing principle and structured dynamics hiding in computational modeling and similar to a mental lexicon organization.

*E. The Resulting Dynamics for Othello and Hamlet*

The emergence of communities was analyzed in the five networks related to the five basic emotions in Hamlet and Othello as well (Figure 9).

*1) 'anger'*

Starting with Othello (Figures 10 and 11), the first level of connection of the central term was related to the concept of 'evil'. The growing process of the synonym network proceeded in a shading semantic path, sliding from the notion of 'fury' and 'rage', to the conceptualization of 'bad' and 'evil'. First high degree nodes arose as central architecture of meaning where the emotional lexicon was shaped. At an advanced growth level 'foul' created a new geometry, connecting the meaning of 'evil' and 'bad' to the

concept of 'madness' with the term 'foul'. The creative process seemed to model the shade of meaning according to the cognitive dynamics involved in the play. There were different levels of synonymy, which modeled the topology of the network.

The second major clustering of the network reorganized the semantic space, moving from the concept of 'madness' to the cluster with the highest degree node 'caitiff'. The new shade of meaning shuttled to another dimension and affected the initial denotation of 'anger', according to the well-known plot of Othello, where the idea of 'evil' and 'anger' was strictly related to 'dreadful traitors'. The term 'anger' was analyzed in detail in Hamlet too, through the visualization of three time steps of its network evolution, from a single node to a small synonymic network.

At time step 1 (see Figure 12), 'anger' linked itself to other key terms such as 'rage', 'choler' or 'wrath', reproducing results that were similar to WordNet networks, where they were considered pure synonyms of 'anger'. Due to diachronic mutations of semantic areas of the terms, in this step other words were included as synonyms, such as 'gall', 'distemper' or 'venom'. These elements gave remarkable hints about the author perception of the emotion 'anger', always focused on poisonous and contagious elements. Unsurprisingly, at the second level of the network (Figure 13) terms as 'poison' and 'contagion' appeared, progressively expanding the semantic area covered by the initial node 'anger'. It is interesting to note the immediate presence of the term 'madness' in the network, already at level 1 of its evolution and growing at level 2, linking to other existent nodes such as 'choler'. Furthermore, 'lunacy' was added as node to the network configuration. 'madness' and 'poison' (which are the two leitmotifs in Hamlet's plot) were well-represented concepts in Shakespearian language production networks.

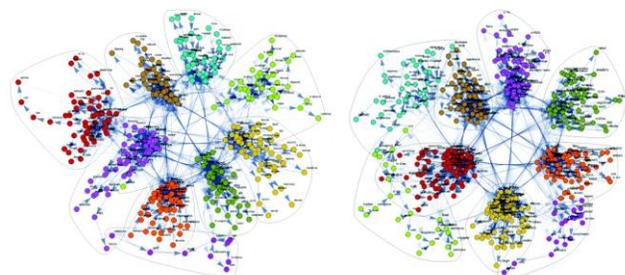


Figure 9. Communities of emotional words in Othello (left) and Hamlet (right).

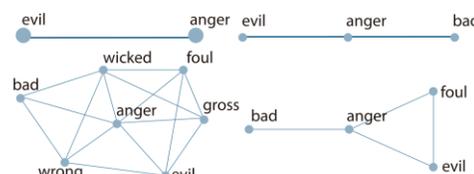


Figure 10. The network of emotion 'anger' in Othello at levels 1, 2, 3 and 5, clockwise.

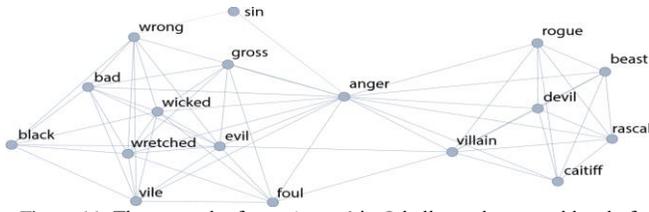


Figure 11. The network of term ‘anger’ in Othello, at the second level of clustering on the left.

1) ‘fear’ and ‘joy’

‘fear’ and ‘joy’ were analyzed as well. Two levels of growth were found by using both methods: the computational and the manual one showed few similar results. The term ‘joy’ reflected the same dynamics of growth. In the computational model, the starting point was a first setting of direct synonyms, which clearly belonged to the semantic field of ‘joyfulness’. As seen in Figure 14, the manual model, at the first level, reproduced a structure of cognitive organization similar to WordNet communities, connecting the word ‘joy’ with a specific set of word (‘delight’ and ‘pleasure’).

The second level detached from the probability rules of the computational model, clustering the terms according to different cognitive processes, used in the creation of the tragedy’s plot. Then, the term ‘joy’ acquired a different shade, growing in a non-linear sense, towards the semantic field of ‘satisfaction’ and ‘bliss’ with a hidden idea of ‘revenge’. This outcome was due to the differences in the structural properties of the two networks: the computational one grew by searching into a bigger database (WordNet) and looking for various kind of connections, while the manual one focused on the search of strong (mainly synonymic) links inside the pool of words used by Shakespeare in Hamlet. It is noticeable that while the first network allowed a general exploration of the semantic space of the ideal term ‘fear’, the second network was strictly indicative of Shakespearean lexical choices. The relationship between the same term in the two networks reflected the relationship of a ‘type’ with its ‘token’, where a ‘type’ represents a prototypical concept, and the token represents one of its possible real occurrences (in this case, the one chosen by the author of the tragedy).

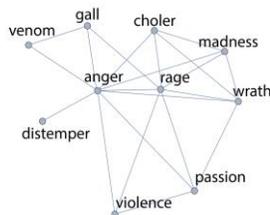


Figure 12. The network of the term ‘anger’ in Hamlet, at level 1.

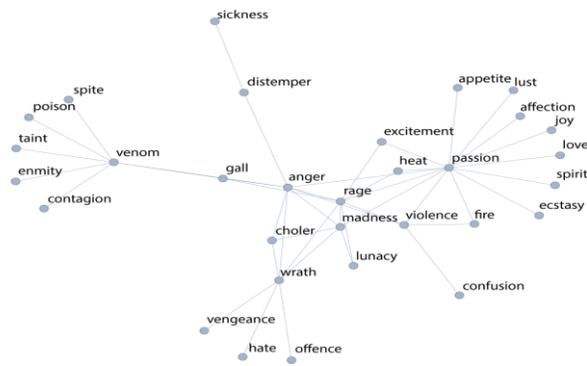


Figure 13. The node ‘anger’ in Hamlet, at level 2.

2) Body and Emotions Dynamics in Othello and Hamlet

Physiological changes due to specific emotions represent reparatory reactions to specific events [31]. Most of these modifications are highly visible, such as increase/decrease in heart rate, changes in skin complexion, etc. A deep analysis of the metaphors and images consciously or unconsciously associated to emotional states helped clarifying the way emotions are perceived. In the networks, we found a good example for this phenomenon, useful to discover how the characters of the plays sensed emotions. In Figure 15, the cluster in the network revealed how ‘mourning’ and ‘sorrow’ were expressed through a group of interconnected verbs. Sorrow expression was linked to hearing and voice. Thanks to clusters visualization, it was easy to study verbs such as: ‘to mourn’, ‘to weep’, ‘to grieve’, and ‘to lament’. They shared different auditory variations of the same sense, through which ‘sorrow’ became concrete and audible. Tables opened the possibility to enter in the character’s inner sensing and understand how an emotion shaped its body and mind.

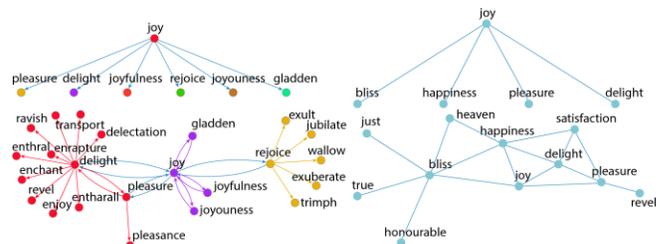


Figure 14. Growth levels Comparative analysis (first top and second level bottom) of the term ‘joy’: manual (right) and computational (left) methods.

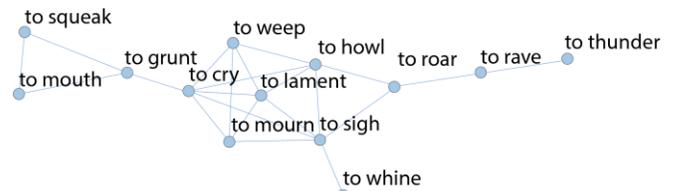


Figure 15. ‘sadness’ (Hamlet). Focus on the cluster of verbs representing pain and its expressions.

3) The Emotional Continuum in Cognitive Phenomena: Dynamical Transitions

Generally, emotions are described as a discrete phenomenon [32][33]. In real situations, subjects experiment

a continuous passage from one emotion to another, giving the change of the variables (internal and external factors) that influence cognitive behavior [34].

We use networks and communities of emotional words as indicators of these complex changes in Shakespearean characters. In fact, networks visualizations made visible the intersection points between different emotional words. It allowed us to identify those terms recurring in every (or almost every) emotional phenomenon, as can be seen in Table VI for Hamlet and Othello.

TABLE VI. CROSS-EMOTIONAL TERMS IN HAMLET AND OTHELLO

Hamlet	'fear'	'anger'	'sadness'	'love'	'joy'
'blood'	+	+	+	+	-
'soul'	-	+	+	+	-
'heart'	-	+	+	+	+
'passion'	-	+	+	+	-
'madness'	+	+	+	+	-
'fire'	+	+	+	+	-
'bosom'	-	+	+	+	-
'heat'	-	+	+	+	-
Othello					
'blood'	-	+	+	-	-
'soul'	-	+	+	+	+
'heart'	+	+	+	+	+
'love'	-	+	+	+	-
'happiness'	-	+	+	+	+

F. Emerging Motifs as Cognitive Attractors

After this initial work, we found that cognitive networks, not only created structured connections between terms, but such terms, far from being arranged randomly, had an organization. We assisted to the emergence of self-organizing structures [35][36]. In discrete dynamic systems, such as Cellular Automata, said motifs are known as gliders [37], and are responsible for the stowing and transportation of information [38][39]. Similarly, the growth of linguistic networks produced increasingly complex structures, showing that there existed many modalities of network growth: at the global level, within the social/cultural expression of emotions, and at the community of words level, within the large domain of the characters dialogue. Well organized patterns were observed, recalling platonic solids and other multidimensional geometric figures. Each motif was preceded by the starting emotional noun and followed by a number. The latter defined the order of appearance of the motif within the growth of the emotional linguistic network. An iconographic apparatus was associated to each motif: two images that present the constitutive elements of the motif itself in two dimensions, one with linguistic tags and one without, and one three-dimensional image. Some explanatory examples are provided below.

G. Complete Graphs, Symmetries and Breaches of Symmetries

Complete graphs were composed of n nodes, each of which was connected to all of the other ones. These motifs possessed both reflective and translational symmetries. Since the detailed study of said symmetries went beyond

this article’s objectives, for further details please see [40]. For explanatory purposes, it must be pointed out that polygons have two dimensions, prisms are three-dimensional objects (whose volume is composed of a finite number of planar polygons), and polytopes have various dimensions (4, 5 . . . n . . .), known as polychorons. The concept of polytopes extends that of prism to various dimensions. Some of the emerging motifs that were found belonged to the prism category and others to the polytopes category.

A second category of motifs was similar to the previous one, although some of the edges were removed: this resulted in the loss of some symmetries, also known as breach of symmetries. The latter had a very complex organizational nature; for example, the breach of two symmetries, if appropriately situated, could restore symmetry [41]. This category of motifs highlighted the presence of non-conventional symmetries of great interest to the previously presented cases. Many of these motifs presented a pair of elements joined by a double bond that pointed to other nodes situated in a plane orthogonal to the pair’s axis. For example, AngerEP\_C1 presented a pentagonal structure orthogonal to the pair’s axis. The most outstanding symmetry in this structure was the 72° rotation, as in the traditional pentagon. Similarly, AngerEP\_C3 presented an octagon on the plane orthogonal to the pair’s axis, with a rotational symmetry of 42°. Examples of said motifs are presented in Figure 16 with the labels on the left, the link directions in the middle, and 3D structure on the right. AngerEP\_B1 (containing the term ira, synonym of anger and degenerated to other meanings) and DisgustEP\_B1 underwent a break of symmetry.

H. Motif Composition

The emergence and breach of symmetries provided information regarding the extraordinary compositional abilities of the organizational and semantic structure of language. The sequence of motifs and patterns highlighted the typical nature of a generative grammar, significantly recalling the language of chemistry and the formation of biological macromolecules. Figure 17 shows some examples of the motifs AngerEP\_D1 and AngerEP\_D2, which identify attractors of emotional terms.

A variant of this structure was found in DisgustEP\_D3, where a 3-simplex and a triangle related to the pair’s axis. This case also presented a double breach in symmetry that reproduced another symmetry. Motif DisgustEP\_D4 was more complex and difficult to visualize (and represent). Parallel to a plane containing a triangle, a square developed on a plane to one side, and a triangle to the other side, resulting in a breach of symmetry. The motif DisgustEP\_D5 highlighted the emergence of a complex structure, starting from a complex 3-motif graph, from which several bonds with other motifs unraveled, situating themselves in a variety of manners across space. The motif ExpectationEP\_D1 was particularly interesting: two

complete graphs composed of a 9-simplex and a 3-node motif, joined around the axis of a node pair. The two nodes connected, one opposite the other, to the central pair and to one of the motifs, respectively.

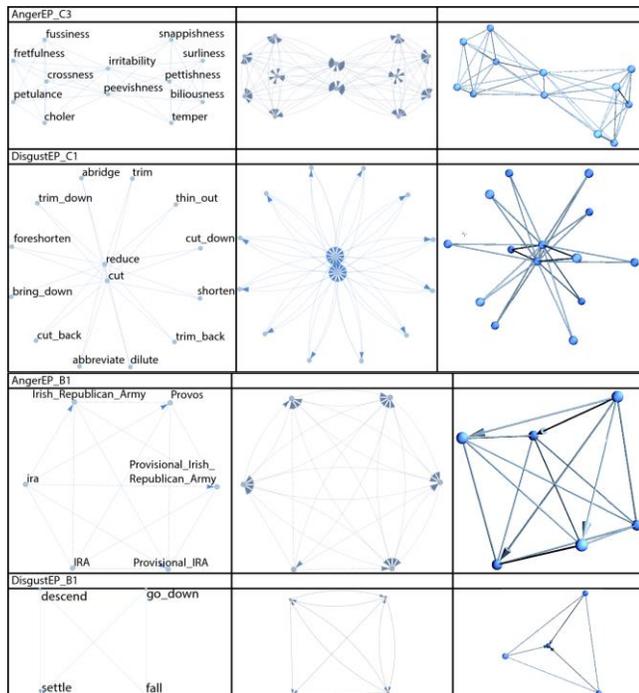


Figure 16. The figure shows examples of symmetries that intercept the semantic organization of two emotional terms.

This kind of visualization highlighted a typical feature of emotions: they tend to blend and mix in a sort of continuum [42], with no clear borders.

#### IV. CONCLUSIONS

This analysis was a non-conventional study of texts full of emotional cognition: Hamlet and Othello, the two well-known Shakespeare’s tragedies. The aim was to convert these masterpieces in a multi- dimensional field of exploration in order to understand how the creative processes in the Shakespeare mind modeled an emotion, how the emotional words semantic network was involved in the process, and how a character, like any other person, felt and expressed an emotion inside communicative contexts.

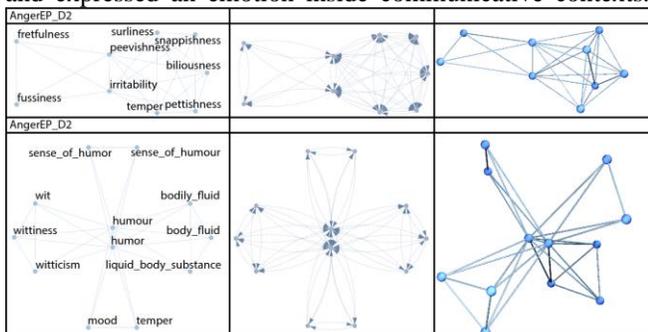


Figure 17. This collection of images represents two different semantic organizations for the emotional term ‘anger’.

The first step was the creation of a list of names, verbs, adjectives, adverbs and sentences for five specific emotions, manually segmented from both plays. Terms referred to ‘anger’, ‘love’, ‘sadness’, ‘joy’ and ‘fear’ and their related semantic fields were selected. The list included specific characteristics for each term, such as who pronounced it, when it appeared in the plays, its context and the perceptual modality by which it was expressed. Then, complex systems theory was applied to the analysis of emotional cognition and these words were transformed into networks. The investigation was further explored by interconnecting the first skeleton of words with WordNet, thus allowing the flourishing of emotional words organization patterns, emerged from the interaction of many basic components at several levels of complexity.

The investigation highlighted the following issues:

- emotional states represented dynamical conditions explained through linguistic and behavioral complex configurations;
- transitions existed at different emotional states;
- linguistic and behavioral attractors were detected as stable configurations or related words patterns, which demonstrated complex organization, with symmetries and breaches;
- such attractors were identified by a set of variables describing environmental, cognitive and physical features of the main characters of both plays;
- patterns variations demonstrated the same features of chaotic systems: small changes in variables’ values involved abrupt qualitative behavioral changes (both in corporeal and cognitive states changes). The transition from one attractor to another could represent phase transitions from an emotional state to another;
- complex organization motifs can be found in both the plays and in the WordNet search networks, that not only highlighted the complexity of the neighborhoods of each emotional term, but that such neighborhoods came with special organized structures, with different geometrical features, which arose during linguistic production.

The emotional terms networks’ growth showed a different level of complexity for the various terms, which far from being strictly ordered, created many cognitive organization patterns. Results showed the evolution of the story plot connected to the growth of the emotional cognition networks, at the single character, and at the global level, particularly represented by transitions among different kinds of emotions, that demonstrated the complexity of the emotional dynamics in the characters of both plays.

The innovative contribution of this paper was to highlight how these elements are brought together in an extraordinary variety of patterns and dynamics, which according to us can represent an expression of cognitive organization of emotion.

#### REFERENCES

- [1] M. Steyvers and J. B. Tenenbaum, “The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth”, *Cognitive Science*, vol. 29, no. 1, 2005, pp. 41-78.
- [2] M. Sigman and G. A. Cecchi, “Global organization of the wordnet lexicon”, *Proceedings of the National Academy of Sciences*, vol. 99, no. 3, 2002, pp. 1742-1747.

- [3] R. F. i Cancho, "The structure of syntactic dependency networks: insights from recent advances in network theory", *Problems of Quantitative Linguistics*, 2005, pp. 60-75.
- [4] M. R. Quillian, "Word concepts: a theory and simulation of some basic semantic capabilities", *Behavioral Science*, vol. 12, no. 5, 1967, pp. 410-430.
- [5] A. M. Collins and E. F. Loftus, "A spreading-activation theory of semantic processing", *Psychological review*, vol. 82, no.6, 1975, p. 407.
- [6] O. Sporns, "Networks of the Brain", MIT press, 2011.
- [7] A. G. Huth, S. Nishimoto, A. T. Vu, and J. L. Gallant, "A continuous semantic space describes the representation of thousands of object and action categories across the human brain," *Neuron*, vol. 76, no. 6, 2012, pp. 1210-1224.
- [8] T. L. Griffiths, M. Steyvers, A. and Firl, "Google and the mind predicting fluency with pagerank," *Psychological Science*, vol. 18, no. 12, 2007, pp. 1069-1076.
- [9] T. T. Hills, M. Maouene, J. Maouene, A. Sheya, and L. Smith, "Categorical structure among shared features in networks of early-learned nouns," *Cognition*, vol. 112, no. 3, 2009, pp. 381-396.
- [10] M. S. Vitevitch, "What can graph theory tell us about word learning and lexical retrieval?," *Journal of Speech, Language, and Hearing Research*, vol. 51, no. 2, 2008, pp. 408-422.
- [11] M. S. Vitevitch, G. Ercal, and B. Adagarla, "Simulating retrieval from a highly clustered network: Implications for spoken word recognition," *Frontiers in psychology*, vol. 2, 2011, p. 369.
- [12] M. S. Vitevitch, K.Y. Chan, and R. Goldstein, "Insights into failed lexical retrieval from network science," *Cognitive Psychology*, vol. 68, no. 1, 2014, pp. 1-32.
- [13] L. Page, S. Brin, R. Motwani, T. and Winograd, "The pagerank citation ranking: bringing order to the web." *Stanford Digital Libraries Working Paper*, 1998.
- [14] R. Milo et al. "Network motifs: Simple building blocks of complex networks," *Science*, vol. 298, no. 5594, 2002, pp. 824-827.
- [15] N. H. Frijda, "The Emotions," Cambridge University Press, 1986.
- [16] A. H. Fischer, "Emotion Scripts: A Study of the Social and Cognitive Facets of Emotions," DSWO Press, 1991.
- [17] A. Ortony, "The Cognitive Structure of Emotions," Cambridge University Press, 1990.
- [18] R. S. Lazarus, "Thoughts on the relations between emotion and cognition," *American Psychologist*, vol. 37, no. 9, 1982, p. 1019.
- [19] K. R. Scherer, "Emotion as a process: function, origin and regulation," *Social Science Information/sur les Sciences Sociales*, vol. 21, no. 4-5, 1982, pp. 555-570.
- [20] R. Plutchik, "The nature of emotions. human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *American Scientist*, vol. 89, no. 4, 2001, pp. 344-350.
- [21] N. H. Frijda, "The different roles of cognitive variables in emotion," *Cognition in Individual and Social Contexts*, 1989, pp. 325-337, Amsterdam: Elsevier Science.
- [22] F. Sharifian, R. Dirven, N. Yu., and S. Niemeier, "Introduction: the Mind Inside the Body," *Culture, body, and Language: Conceptualizations of Internal Body Organs across Cultures and Languages*, vol. 3, no. 24, 2008, Berlin: Mouton de Gruyter.
- [23] A.L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, 1999, pp. 509-512.
- [24] R. Albert, H. Jeong, and A.L. Barabási, "Internet: diameter of the world-wide web," *Nature*, vol. 401, no. 6749, 1999, pp. 130-131.
- [25] A.L. Barabási, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, no. 7039, 2005, pp. 207-211.
- [26] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," In *ACM SIGCOMM Computer Communication Review*, vol. 29, no. 4, 1999, pp. 251-262.
- [27] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási, "The large-scale organization of metabolic networks," *Nature*, vol. 407, no. 6804, 2000, pp. 651-654.
- [28] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K.J. Miller, "Introduction to wordnet: An on-line lexical database\*," *International Journal of Lexicography*, vol. 3, no. 4, 1990, pp. 235-244.
- [29] G. A. Miller, "Wordnet: A lexical database for English," *Communications of the ACM*, vol. 38, no. 11, 1995, pp. 39-41.
- [30] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, 1998, pp. 440-442.
- [31] R. W. Levenson, "Autonomic specificity and emotion," *Handbook of Affective Sciences*, vol. 2, 2003, pp. 212-224.
- [32] P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, 6, no. 3-4, 1992, pp. 169-200.
- [33] L. F. Barrett, "Discrete emotions or dimensions? The role of valence focus and arousal focus," *Cognition and Emotion*, vol. 12, no. 4, 1998, pp. 579-599.
- [34] T. A. Walden and M. C. Smith, "Emotion regulation," *Motivation and Emotion*, vol. 21, no. 1, 1997, pp. 7-25
- [35] C. G. Langton, "Artificial Life: an Overview," MIT Press, 1995.
- [36] E. Bilotta and P. Pantano, "Cellular Automata and Complex Systems: Methods for Modeling Biological Phenomena," IGI Global, 2010.
- [37] S. Wolfram, "Universality and complexity in cellular automata," *Physica D: Nonlinear Phenomena*, vol. 10, no. 1, 1984, pp. 1-35.
- [38] C. G. Langton, "Computation at the edge of chaos: phase transitions and emergent computation," *Physica D: Nonlinear Phenomena*, vol. 42, no. 1, 1990, pp. 12-37.
- [39] C. G. Langton, "Studying artificial life with cellular automata," *Physica D: Nonlinear Phenomena*, vol. 22, no. 1, 1986, pp. 120-149.
- [40] H. S. M. Coxeter, "Regular Complex Polytopes Volume 1," Cambridge University Press, 1974.
- [41] I. Stewart and M. Golubitsky, "Fearful Symmetry: Is God a Geometer?," Courier Corporation, 2010.
- [42] R. Plutchik, "The Psychology and Biology of Emotion," Harper Collins College Publishers, 1994.
- [43] <https://goo.gl/NkNS13>, Access date: November 2015.
- [44] <http://wordnet.princeton.edu/>, Access date: January 2015.

Questo piccolo frammento di ricerca esiste grazie al lavoro di molti altri pezzetti di puzzle, oltre a me. Casa mia e tutti gli altri luoghi in cui mi sono sentita a casa, gli amici e i colleghi vecchi e nuovi, di cubi e paesi diversi, i professori che non hanno perso l'amore per la "loro" scienza e ti si siedono ancora accanto per spiegartelo. E anche se non l'avrei mai detto, gli antagonisti che, nel loro piccolo, mi hanno mostrato chi non voglio diventare. Li ringrazio tutti di cuore senza nominarli, sanno bene di essere parte del puzzle.