Introduction to
the Study of Language
-Ling 1-
Fall 2020

Giuseppina Silvestri

g.silvestri@ucla.edu

Week 10 - December 9, 2020

Announcements

- Analysis Paper 2 is due on Friday December 11 (11:59 PM)

- SONA experiments for experimental credit close tomorrow Thursday December 10

Roadmap for today's class

- 60 minutes
- by the end: notes on Final exam

- last 10/15 minutes: *your* time to ask me questions
    - orally by using the raise-hand button
    - by sending me a private message in the Zoom chat
    - questions can be on any topic of any live lecture

❖ Do genetic diversity, language variation, and geographical distances correlate? How?
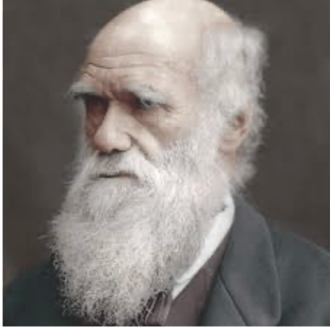
Darwin's dream
(Darwin 1859, 422)
Part 1

If we possessed a **perfect pedigree of mankind**,

a genealogical arrangement of the *races* of man

would afford the **best classification of the various languages**

now spoken throughout the world;

Darwin's dream
(Darwin 1859, 422)
Part 2

And if all extinct languages,

and all **intermediate and slowly changing dialects**

had to be included,

such an arrangement would, I think,

be **the only possible one**.

Darwin's dream

- the evolution and distribution of languages share many fundamental aspects with the evolution and distribution of human populations

- the history and the spread of a language in time and space overlap with the history and the spread of the respective speakers' groups

    - consider languages that disappeared, but populations' genes survived

Is it possible to pursuit Darwin's *dream*?

- In the last few decades, **geneticists** (Cavalli-Sforza, Barbujani) compared the classification of genetic features with larger linguistic families

    - results *not* accepted by linguists

- Linguists never faced such an issue experimentally, until a few years ago

    - **Parametric Comparison Method** (Longobardi and Guardiano 2009)

    - probabilistic and universally valid, applicable to every human language

    - linguistic variation reduced to differences among syntactic features

    - syntactic features assumed to be **discrete** and **universal** "parameters"

Research Project (York University, 2013-2018)

Meeting Darwin's last challenge:
toward a global tree of human languages and genes

https://www.york.ac.uk/language/research/projects/completed/langelin/project/

The Comparative Method:
an old one and a new one

❖ Old

    - regularities in sound changes

    - genealogical classification of languages

❖ New

    - syntactic features/properties as *comparanda*

    - deeper phylogenetic/genealogical classification of languages

Comparing populations and languages:
is this even possible?

- Populations = Genes (DNA)

- Languages = Linguistic characters

    => Syntactic features as characters

Examples from Noun Phrase domain

1. languages with number features vs languages without number features

2. languages with person features vs languages without person features

3. languages with adjectives before N vs languages with adjectives after N

    ETC

Syntactic Characters

TABLE A			Sic	Cal	It	Sal	Sp	Fr	Ptg	Rm	Gri	BoG	Grk	E	D	Da	Ice
1	± gramm. person	FGP	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	± gramm. number +FGP	FGN	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
3	± gramm. gender +FGN	FGG	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
4	± NP over D +FGP	NOD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	± feature spread to N +FGN	FSN	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
6	± numb. on N (BNs) +FSN	FNN	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+
7	± gramm. partial def	DGP	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
8	± gramm. def +DGP	DGR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
9	± strong person +FGP, +DGR, -NOD	NSD	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
10	± free null partitive Q +FNN	DPQ	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-
11	± gramm. dist. art. -FSN or -FNN or +DGR	DDA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	± def-checking N +DGR	DCN	-	-	-	-	-	-	-	+	-	-	-	-	-	+	+
13	± def spread to N +DCN, -NSD	DSN	0	0	0	0	0	0	0	0	0	0	0	0	0	-	+
14	± def on relatives +DGR	DOR	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15	± D-controlled infl. on N +FSN	DIN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	± plural spread from cardinals +FSN	CPS	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
17	± gramm. boundedness	CGB	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	± strong article +DGR, +FNN, -CGB	CGR	+	+	+	+	+	0	+	+	+	+	+	+	+	+	-
19	± bounded.-checking N +CGB	CCN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	+NSD	DNN	-	-	-	-	+	-	+	0	-	-	-	0	0	0	0
21	± structured APs	AST	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
22	± feature spread to struct. APs +FSN, + AST	FFS	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+
23	± feature spread to pred. APs +FGN	FSP	+	+	+	+	+	+	+	+	+	+	+	-	-	+	+
24	± D-controlled infl. on A -NSD, +FFS	ADI	0	0	0	0	0	0	0	0	0	0	0	0	+	+	+
25	± DP over relatives	ADR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
26	± relative extrap. -ADR	AER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

The power of Syntax

(Dialectal) syntax retains
a significant **historical signal**

erc

Long-Range Comparison between Genes and Languages Based on Syntactic Distances

Vincenza Colonna[a]   Alessio Boattini[b]   Cristina Guardiano[c]   Irene Dall'Ara[a]
Davide Pettener[b]   Giuseppe Longobardi[d]   Guido Barbujani[a]

[a]Dipartimento di Biologia ed Evoluzione, Università di Ferrara, Ferrara, [b]Dipartimento di Biologia Evoluzionistica Sperimentale, Università di Bologna, Bologna, [c]Dipartimento di Scienze del Linguaggio e della Cultura, Modena, and [d]Laboratorio di Linguistica e Antropologia Cognitiva, Dipartimento di Storia e Culture, Trieste, Italy

Significant D_{SYN} / D_{GEN}

correlation

r=**0.68**

(Colonna et al 2010, 16 pop.)

Across Language Families: Genome Diversity Mirrors Linguistic Variation Within Europe

Giuseppe Longobardi,[1,2] Silvia Ghirotto,[3] Cristina Guardiano,[4] Francesca Tassi,[3] Andrea Benazzo,[3] Andrea Ceolin,[1] and Guido Barbujani[3]*

[1]Department of Language and Linguistic Science, University of York, York, UK
[2]Department of Humanities, University of Trieste, Trieste, Italy
[3]Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy
[4]Department of Communication and Economics, University of Modena-Reggio Emilia, Modena, Italy

KEY WORDS parametric comparison method; genome-wide diversity; single-nucleotide polymorphisms; human evolutionary history

ABSTRACT   Objectives: The notion that patterns of linguistic and biological variation may cast light on each other and on population histories dates back to Darwin's times; yet, turning this intuition into a proper research program has met with serious methodological difficulties, especially affecting language comparisons. This article takes advantage of two new tools of comparative linguistics: a refined list of Indo-European cognate words, and a novel method of language comparison estimating linguistic diversity from a universal inventory of grammatical polymorphisms, and hence enabling comparison even across different families. We corroborated the method and used it to compare patterns of linguistic and genomic variation in Europe. Materials and Methods: Two sets of linguistic distances, lexical and syntactic, were inferred from these data and compared with measures of geographic and genomic distance through a series of matrix correlation tests. Lin-guistic and genomic trees were also estimated and compared. A method (Treemix) was used to infer migration episodes after the main population splits. Results: We observed significant correlations between genomic and lin-guistic diversity, the latter inferred from data on both Indo-European and non-Indo-European languages. Con-trary to previous observations, on the European scale, lan-guage proved a better predictor of genomic differences than geography. Inferred episodes of genetic admixture fol-lowing the main population splits found convincing corre-lates also in the linguistic realm. Discussion: These results pave the ground for previously unfeasible cross-disciplinary analyses at the worldwide scale, encompassing populations of distant language families. Am J Phys Anthropol 157:630–640, 2015. © 2015 The Authors American Journal of Physical Anthropology Published by Wiley Periodicals, Inc.

Significant D_{SYN} / D_{GEN}

correlation **in Europe**

r=**0.60**

(Longobardi et al 2015, 15 pop.)

X

13

Genes and languages in Southern Italy

- Within Italo-Romance dialects, D_{GEN} becomes virtually **uninformative**

- demographic structure of Southern Italy shows a rather **uniform genetic landscape** (Sarno et al 2016)

- there is just one significant non-Romance exception

❖ **Why**?

Genes, syntax, and geography

- Syntactic and genomic diversities have **a different distribution through space**:

-
  - ➢ **genes** correlate with **geography** much more **strongly** than syntax

    - D_{GEN}/D_{GEO} (r=**0.63** in Europe, r=**0.88** in Eurasia)
    - D_{SYN}/D_{GEO} (r=**0.42** in Europe, r=**0.47** in Eurasia)

  - ❖ Genetic diversity depends more heavily on geographical distances

Genes and languages in Southern Italy

❖Why do we see a rather uniform genetic landscape?

➢In contexts of intense population admixture **genetic uniformation** happens

➢ In such contexts **genetic uniformation** occurs **more than linguistic uniformation**

❖Why is D_{GEN} virtually uninformative in southern Italy?

➢if **D_{GEN}** correlates strongly with **D_{GEO}**, when the latter becomes very short, D_{GEN} will tend to 0

Syntax and Geography in Southern Italo-Romance

- **D_{SYN} / D_{GEO}** r =**0.75**

❖ in Southern Italo-Romance the distribution of D_{SYN} **connected** to that of D_{GEO}

At the level of micro-comparison,

geographical factors are **reflected in syntactic distances**.

Is Darwin's prediction confirmed?

- Darwin's prediction: large-scale **genetic diversity-language variation congruence**

- *At that scale* Darwin's prediction is **confirmed**

- Darwin's prediction about gene-language is **not confirmed** at *the microvariation level*

About the Final (1)

- **OPENS** on **Tuesday 12/15 at 6 PM** (PST)
- **CLOSES** on **Thursday 12/17 at 6PM** (PST)

     => 48-hour window

- timed => 3 hours
- there will be **no make-up** final for any circumstances.
- anyone who misses the final exam will be given an **Incomplete**.

- **cumulative** (week 1- week 10)
- about the same number of questions as the Midterm

About the Final (2)

- Prepare about Week 9 and Week 10 contents by:

    - doing the readings from the textbook

        Week 9: pp. 326-361

        Week 10: pp. 365-373 and 16-21

    - attending the Sections

About the final (3)

- Contents of Week 9 and 10 to focus on:

      - Comparative Method (Neogrammarians)

      - Family Tree model

      - sound change correspondences

      - Great Vowel Shift (English)

      - Indo-European Family

      - evolutionary value of language

      - evolution of larynx and other organs (concept of Exaptation)

      - Phylogeny (and Ontogeny)

EVALS still open

close on Saturday, December 12 (8AM)

- Exercise **your right** of expressing your opinion

- **Help** your instructors

STAY SAFE !