Genetic biasing in language diversity and universals

Dan Dediu
The Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands
Dan.Dediu@mpi.nl

Conference on Phonetic Universals
Max Planck Institute for Evolutionary Anthropology, Leipzig
30 October 2010
Overview

- Genes, language diversity and universals
- Genetic biasing: the idea
- Some computer and experimental models
- An early linguistic proposal
- Tone, *ASPM* and *Microcephalin*
- The stability of *Tone*
- Conclusions
Genes, language diversity and universals

- **Analogies** between genetics (population & evolutionary) and linguistics:
  - Absolute universal ~ fixed allele/gene
  - Variation → “tendencies”
- **Causal** relationship:
  - genes → linguistic distributions
  - understanding variation is essential for understanding universals
The Genetic Biasing of Language

Language is **shaped by** the

- **Production** and **perception** systems
- **Brain/cognition**
- **Social & cultural** factors and processes

→ some have *genetic components*

Adaptation to these through **cultural evolution**

⇒ **genetic biases can affect language**

*But no “gene for Chinese”!*
Computer and experimental models

- Models of **language evolution**: *small biases* have a *complex dynamics when culturally transmitted* (Kirby et al., 2007; Dediu, 2008, 2009)


- **Experimental amplification** of biases through *cultural chains*: compositionality, systematicity, etc. (Smith & Wonnacott, 2010; Kirby et al., 2007, etc.)
Italian and Yoruba vowel systems

- 7-vowel systems (/i e ɛ a ɔ o u/)
- $F_2$ lower in Italian
- Due to differences in mouth opening

“This does not, of course, imply that a Yoruba could not learn perfect Italian. *Any individual speaker could compensate* for the overall, statistical, difference in headshape” (Ladefoged 1984: 86)
Tone, ASPM & Microcephalin

Two Genes Involved in Brain Development

- **ASPM** (1q31), **Microcephalin** (8p23)
- Recessive primary microcephaly
- Probably affects the *number of asymmetric divisions*
- *Accelerated evolution* in the lineage leading to humans
- **Derived haplogroups** (ASPM-D, MCPH-D):
  - *Recent* origin
  - Marked *geographic structure*
  - Recent/ongoing *natural selection* (???)
  - Unknown phenotypic effects
**Tone, ASPM & Microcephalin**

*Database, Methods & Results*

- **49 populations, 981 genetic markers, 26 binary features**
  - $r_{ASPM-Tone} = -0.53$, $r_{MCPH-Tone} = -0.54$, $p < 0.05$, top 1.4%
  - **logistic regression**: $p < 0.05$, Nagelkerke's $R^2 = 0.53$, 73% corr. classif., in top 2.7%.

**Mantel corr:** *geography, genetics, typology & history:*

- controlling geography: $r = 0.291$, $p = 0.003$
- & controlling history: $r = 0.283$, $p = 0.000$
Tone, ASPM & Microcephalin

Summary of Findings

Scatter plot of ASPM x MCPH by Tone

- India?
- Australia?

Dediu, 30 Oct 2010
Tone, ASPM & Microcephalin

Proposed Mechanism

1. Genetics → Ontogeny/environmental modulation → Individual linguistic bias
2. Language transmission → Language change

Population

Legend:
- Population
- Language with two possible alternative states
- Frequency of allele A
- Relative strengths of the biases
Tone, ASPM & Microcephalin

Other Supporting Evidence & Future Work

- Christiansen, Kelsey & Tomblin (pc. 2008): preliminary association between an ASPM SNP and several measures (probably) related to phonology
- Dediu (2010): tone seems to be more stable than expected
- Effects on brain anatomy (Rimol et al., 2010)
- Work in progress:
  - Operationalization of the bias & genetic association studies
  - Historical characteristics of tone
  - More specific effects on brain structure
Stability of Tone

The idea

- **Vocabulary** (cognacy): can be very stable (Pagel et al. 2007, Greenhill et al. 2010)
- **Typology**: can also be very stable (Hunley et al. 2008) but *differs across language families* (Greenhill et al. 2010)

- What about **global tendencies**? → small number of language families
- The genetic biasing hypothesis → *tone should be stable*
Stability of Tone

Principles & Methods

- **Bayesian phylogenetics** (2 software packages: MrBayes 3 & BayesLang) → rate estimation
- As many **lg.fams** as possible (41 in total) from 2 classifications (Ethnologue & WALS)
- As many **typological features** as possible (68 polymorphic & 86 binary; WALS)
- As many **outgroups** as possible (23 language isolates)

- Estimate rates → rank features → compare ranks across software packages, data codings, outgroups,...
Stability of *Tone*

**Results**

- **Similar results across:**
  - Lg. classifications: $0.96 \leq r \leq 0.99$, $p<10^{-10}$
  - Outgroups: $0.49 \leq r \leq 0.92$, $p<10^{-6}$, mean $r=0.78$, PC$_1$ 79%

- **Binary coding:**
  - Strong agreement: $0.59 \leq r \leq 0.98$, $p<10^{-8}$, mean $r=0.78$, PC$_1$ 81.4%
  - *Tone2* (8 of 86), *Tone1* (23 of 86)

- **Polymorphic coding:**
  - Strong agreement: $0.51 \leq r \leq 0.99$, $p<10^{-5}$, mean $r=0.71$, PC$_1$ 76.1%
  - *Tone* (8 of 68)
Stability of Tone

Binary vs polymorphic codings

mean $r=0.61$, median $p=6.5\cdot10^{-9}$, PC$_1$ (agreement) 67.4%, PC$_2$ (bin vs poly) 16.1%

▲ = Word Order, ○ = Phonology, ∆ = Nominal Syntax, ● = Morphology, □ = Verbal Categories, ■ = Nominal Categories, + = Simple Clauses, and All = all types of features combined.
Stability of *Tone*

**Results & conclusions**

- *Tone* is very stable as **poly feat** \( t_{56} = 9.7, p = 1.35 \cdot 10^{-13} \)
- *Tone2* is very stable as **bin feat** \( t_{70} = 12.04, p < 2.2 \cdot 10^{-16} \) and **overall** \( t_{70} = 12.27, p < 2.2 \cdot 10^{-16} \)
- *Tone1* relatively stable as **bin feat** \( t_{70} = 4.35, p = 4.5 \cdot 10^{-5} \) and **overall** \( t_{70} = 4.35, p = 6.7 \cdot 10^{-9} \)

- Many **caveats** (data quality and quantity, appropriateness of strict phylogenetic methods to typology, language isolates as outgroups, etc.) → **suggestive at best!!!**
Stability of Tone

Comparing with other methods (Work in progress)

- Other methods for typological stability:
  - M. Parkvall (2008) Which parts of language are the most stable? *STUF* 61:34–250
    - stable if showing strong vertical signal and weak propensity to borrowing
    - related vs unrelated languages → stable features shared more between related languages
Stability of Tone

Comparing with other methods (Work in progress)

- After **removing outliers**
- Pretty strong and significant **correlations** between feature stabilities ($0.39 \leq r \leq 0.63$, $p < 0.002$)
- **PCA**: $PC_1$ (the commonality) explains 63.5% of variance
- **Tone** stability:

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Rank (out of 61)</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>7</td>
<td>11.47%</td>
</tr>
<tr>
<td>P1</td>
<td>42</td>
<td>68.85%</td>
</tr>
<tr>
<td>P2</td>
<td>32</td>
<td>52.45%</td>
</tr>
<tr>
<td>WH</td>
<td>18</td>
<td>29.50%</td>
</tr>
<tr>
<td>Agreement ($PC_1$)</td>
<td>22</td>
<td>36.06%</td>
</tr>
</tbody>
</table>
Conclusions

- Probably parallels between population/evolutionary genetics and linguistic typology are profitable
- *Universals* → statistical tendencies due to many factors:
  - **constraints** on cognition, perception, production → possible genetic biases, which are a type of:
    - **cultural evolutionary processes** such as random drift, “evolutionary inertia” and selection
- Exploration of the linguistic “design space” →
  - time for cultural evolution
  - wrong metaphor? Maybe the design space is in fact dynamic?
- What are we generalizing to?
End

... for More Info:


-----------


... Many Thanks to:


The University of Edinburgh, UK; ESRC UK; The Max Planck Institute for Psycholinguistics, Nijmegen

And to the organizers of the Conference on Phonetic Universals