COMPARATIVE AND ‘DIVERSITY LINGUISTICS’

WHERE NEXT?

Paul Heggarty

Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig

Paul.Heggarty@gmail.com — http://eva-mpg.academia.edu/PaulHeggarty
A WIND OF CHANGE?
SIGN OF THE TIMES?

Bernard Comrie, (typological) linguist

Russell Gray, “evolutionist”

... interested in “language universals” ... and “language typology”. Why are language universals and cross-linguistic variation the way they are? Various phenomena are studied across a wide range of languages ... field work is an important tool...

... bring together biologists, linguists and social scientists to apply cutting-edge ... computational advances from the natural sciences while still maintaining the highest standards of scholarship from the humanities ... [to resolve] long-standing questions about human history that were previously deemed difficult, or even completely intractable.
ILL WIND — OR BREATH OF FRESH AIR?

• Linguistics being outdone — even led? — by other disciplines?

• Led ... by non-linguistic tools and models:
  – From biological or mathematical sciences.
  – Uns suited to language?

• Led ... in which direction?

• Change of focus, even of whole objective?
  – Language for language’s sake ...
    → Language for human (pre)history.
BE NOT AFRAID!

LINGUISTS, MEET NUMBERS ...

- Change in methods: numbers. But ...
  - You can’t get good numbers without the qualitative analysis.
  - Typological and universal tendencies also need quantitative answers.
HOW DID IT COME TO THIS?

Make the news.
Get the funding.
Call the shots.

• Old question — hot news.

• Linguistics began with a question posed in … 1786.

• Still no answer!

• Huge new advances … from outside linguistics.
Mysterious Indo-European homeland may have been in the steppes of Ukraine and Russia

By Michael Balter | 13 February 2015 2:15 pm | 77 Comments

What do you call a male sibling? If you speak English, he is your “brother.” Greek? Call him “frater.” Sanskrit, Latin, Old Irish? “Brother,” “frater,” or “brither,” respectively. Ever since the mid-17th century, scholars have noted such similarities among the so-called Indo-European languages, which span the world and number more than 400 if dialects are included. Researchers agree that they can probably all be traced back to one ancestral language, called Proto-Indo-European (PIE). But for nearly 20 years, scholars have debated vehemently when and where PIE arose.

Two long-awaited studies, one described online this week in a preprint and another scheduled...
Massive migration from the steppe was a source for some Indo-European languages in Europe.

Wolfgang Haak, Iosif Lazaridis, Nick Patterson, Nadin Rohland, Swapan Mallick, Bastien Llamas, Guido Brandt, Susanne Nordenfelt, Eadaoin Harney, Kristin Stewardson, Qiaomei Fu, Alissa Mittnik, Eszter Bánffy, Christos Economou, Michael Francken, Susanne Friederich, Rafael Garrido Pena, Fredrik Hallgren, Valery Khartanovich, Aleksandr Khokhlov, Michael Kunst, Pavel Kuznetsov, Harald Meller, Oleg Mochalov, Vayacheslav Moiseyev et al.

Received 29 December 2014 | Accepted 12 February 2015 | Published online 02 March 2015

Haak et al. (2015)

“a steppe origin of at least some of the Indo-European languages of Europe.”
A fictionalized but familiar conversation ...  

"So can you name any real discovery made by generative syntax?"

"How about hierarchical phrase structure, which obeys its own principles but interacts with just about everything else in language?"

"What about case theory and other deep principles that shape the phrases of languages"?

"Or uniform locality conditions that cut across distinct constructions?"

"The discovery in all these areas of a deep unity amidst the diversity of the syntax of the world's languages?"

"Very interesting, but can you name any real discovery made by generative syntax?"

LSA, Boston, January 2013, plenary by David Pesetsky (MIT)

“Что дѣлать? What is to be done?”
Languages differ so fundamentally from one another at every level of description ... that it is very hard to find any single structural property they share. The claims of Universal Grammar ... are either empirically false, unfalsifiable, or misleading ...

Structural differences should instead be accepted for what they are, and integrated into a new approach to language and cognition that places diversity at centre stage ... Chomsky’s notion of Universal Grammar (UG) has been mistaken ... for a set of substantial research findings about what all languages have in common.
Generative linguists following Chomsky have claimed that linguistic diversity must be constrained by innate parameters that are set as a child learns a language. . . . [Our] findings . . . that — at least with respect to word order — cultural evolution is the primary factor that determines linguistic structure.
Mapping the Origins and Expansion of the Indo-European Language Family

Remco Bouckaert, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard, Quentin D. Atkinson

Bouckaert et al. (2012)
Gray et al. (2009)

Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement

R. D. Gray,1 A. J. Drummond,2 S. J. Greenhill1

Science
Currie et al. (2010)

Linguistics as a reference framework for human cultural (pre)history?
Language transmission. Glossogenetic studies relate Y-chromosome DNA types with language in the indicated regions. Such correlation is not observed for mtDNA, which is inherited from the mother. “Melanesian” designates non-Malayo-Polynesian languages in New Guinea, where Malayo-Polynesian is spoken in coastal pockets. The hatched region shows Bantu as a branch within the Niger-Congo language family.

Male vs. female lines match differently with language lineages.
IF YOU CAN’T BEAT ’EM, JOIN ’EM?

“gene-language congruence ... by formal syntax ... brought to bear on historical issues.”

Centre for Linguistic History and Diversity

- Workshop on Language Variation and Change and Cultural Evolution
- The New Historical Linguistics and the World of Annotated Corpora

Current partners beyond York are:
- University of Campinas (Brazil)
- University of Pennsylvania (USA)
FROM LANGUAGES TO HISTORY ... BY NUMBERS?

1. Encode
   turn language data into numbers

2. Number Crunch
   statistical & phylogenetic analysis & visualisation

3. Interpret
   what the results mean for (pre)history

- At each stage:
  - Concerns, problems, dangers, false analogies.
  - Opportunities, scope for huge advances.
Stage 1:

Think of a number, any number

Quantity vs. Quality?
PUTTING MEANINGFUL NUMBERS ON LANGUAGE?


Maddieson (2013: WALS 2a): Vowel Quality Inventories
Quantification, Rule 1: Do Not ‘Bin’ Continuous Data

"the WALS data are binned into ranges ..."

(Atkinson 2011: SI 2)


Values

<table>
<thead>
<tr>
<th>Size</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small (2-4)</td>
<td>93</td>
</tr>
<tr>
<td>Average (5-6)</td>
<td>287</td>
</tr>
<tr>
<td>Large (7-14)</td>
<td>184</td>
</tr>
</tbody>
</table>

When vowel qualities are counted in this way in the sample of languages surveyed for this chapter, the average number of vowels in a language is just fractionally below 6. The smallest vowel quality inventory recorded is 2 and the largest 14.

Maddieson (2013: WALS 2a)

  mod. large [26-33] large [34+]  

- Tone: no tone simple tone complex tone ...
When 7 = 13 ... But Not 5

- English: 13 = ●
- German: 14 = ●
- Spanish: 5 = ○
- Latin: 5 = ○
  (5 long + 5 short)
- Italian: 7 = ●
  = 5 basic, + /ɛ/ /ɔ/ if stressed

● = ● i.e. 7 = 13 = 14
● ≠ ○ i.e. 7 ≠ 5

→ In vowel quality inventory, Italian is represented as ... 
  - Identical to English, German = most extreme languages in sample.
  - Completely different to Spanish, Latin = just on other side of mean (6).
Comrie (2013: WALS 98a): Alignment of Case Marking of Full Noun Phrases
But the main recurrent difficulty is that in many languages, different kinds of full noun phrases partake of different case marking patterns. For instance, in Spanish the accusative marker, the preposition *a*, is found (roughly) only with specific, animate noun phrases, so that strictly speaking a noun phrase like the male proper name *Juan* has a nominative–accusative case marking system, while the inanimate noun phrase *el libro* ‘the book’ has a neutral case marking system, as illustrated partially in (7).

(7) **Spanish**

a. *María vio a Juan.*

Mary see.AOR.3SG ACC John  
‘Mary saw John.’

b. *María vio el libro.*

Mary see.AOR.3SG the book  
‘Mary saw the book.’

↑ ‘Qualitative’ justification (‘personal a’)...
... but ‘anti-quantitative’:

A. The policy that has been followed in assigning such languages to types has been \textbf{to maximize the occurrence of overt case marking}. Thus, if a language has an optional accusative case marker, or one that occurs only under certain specified circumstances, then this has been given priority and taken as critical. This policy decision needs to be taken into account consistently in interpreting the maps \ldots Thus, Spanish and Burmese come out as accusative, Araona and Gooniyandi as ergative, and Hindi as tripartite.

\begin{itemize}
  \item "Maximise ... priority ... critical" $\rightarrow$
  \begin{itemize}
    \item All = $\bullet$. Any = $\bullet$. Any = all.
    \item 0.01 is closer to 1 than to 0. 0.01 is 1. 1\% = 100%.
  \end{itemize}
\end{itemize}
QUALITATIVE OR QUANTITATIVE?

• An atlas for display purposes (APiCS too) ...

• ... but being used as a database for quantitative purposes.

• Other issues:
  – (Mis)used for inferences about genealogy ...
    ... but WALS ‘families’ very controversial:
      e.g. *Khoisan, *Altaic, *Australian, *Nilo-Saharan, etc...

  – Coverage of languages sparse (avg. under 3%) and inconsistent.

• We need new databases dedicated for quantitative uses.

• Qualitative or quantitative? “It doesn't have to be this way ...”
NEW DATABASES: QUALITATIVE AND QUANTITATIVE

GLOTTOBANK: world-scale databases, specifically for quantitative applications...

- GramBank
  - Harald Hammarström, Hedvig Skirgård
- LexiBank
  - Simon Greenhill
- PhonoBank
  - Mattis List
- IELex and URAlex
  - Michael Dunn
- Syncretism in paradigms
  - Nick Evans
STAGE 2: CRUNCHING THE NUMBERS

TOOLS & MODELS

STATISTICS & PHYLOGENETICS
CLIMBING DOWN FROM THE TREES?

• Being led by the tools and models?

• Tree idealisation: a concern with new phylogenetic models?

• Far more of a problem for \textit{traditional} historical linguistics…
  
  \textit{e.g.} Best-researched ‘LOL’ families in world: agreed trees?

• Wild goose chase: \textit{no} tree reflects historical realities of speaker populations.
THE TREE MODEL VS. REAL POPULATION HISTORY...

• Human societies do not live (‘evolve’) only in binary branching relationships.

• So nor do their languages.   (Cause-and-effect relationship.)

- Exploding a myth:  Comparative Method ≠ Trees!
- It’s precisely the comparative method that confirms data not tree-like!

- Alex François   (Société Linguistique de Paris, 17th January 2015).
HISTORIES NOT TREE-LIKE: A NEGLIGIBLE FRUSTRATION?

Indic, Arabic, ‘Chinese’, Bantu, Mayan, Quechua, Algonquian, Italy, Scandinavia, Switzerland — formerly much more of Europe ...
BAYESIANISM: HANDLING AND MEASURING UNCERTAINTY

Bouckaert (2015, last Friday)

Ringe et al. (2002)

‘Distribution’ of Indo-European phylogenies

single ‘perfect phylogeny’ (no Germanic!)

Which is more realistic?

Language-tree divergence times support the Anatolian theory of Indo-European origin.

- Time range of farming not Steppe hypothesis.
STAGE 3: INTERPRETATION

FROM DATA ANALYSES TO HUMAN (PRE)HISTORY

- Languages with no data not shown.
- Mercator = area distorting projection, which ...
... “SHOULD NEVER BE USED FOR DENSITY VISUALISATION PURPOSES”


• Eckert IV equal area projection, buffered Thiessen Tessellation.

• Languages with no data all included and shown as such.
INTERPRETING STATISTICS:  
ANYONE FOR FISHING?

Everett (2013):  
*Evidence for direct geographic influences on linguistic sounds: the case of ejectives*

“62% of languages with ejectives are located in high elevation ‘zones’, which are defined here as major regions greater than 1500 m in altitude, plus land within 200 km of such a region”

Chen (2013):  
*The effect of language on economic behavior: evidence from savings rates, health behaviors, and retirement assets.*

“Lies, damned lies, and statistics.”
PATTERNS IN LANGUAGE DIVERSITY: NOT JUST FAMILIES …

Traditional ‘family preference’, especially for work on prehistory, but …

… linguistics has far more to say on human origins and interactions.

Patterns on all other ‘diversity dimensions’ of linguistic panorama.
**DADDY, WHERE DO LINGUISTIC AREAS COME FROM?**

General principle: linguistic effects ← real-world causes.

**LANGUAGE FAMILIES**  ← expansive, **divergent processes**

**LINGUISTIC AREAS**  ← **convergent processes**.

Clear-cut: Member of family, yes or no?

Diffuse: core vs. peripheral members.
PATTERNS AND CAUSATION: THE CASE OF “ALTAIC”

- **Core vs. Periphery**
  Altai vs. Uralic, Korean, Japanese
  = Pattern typical of convergence areas.

- ‘Mobility’, nomadism, very low density ...
  → Family ‘spread zone’ (→ divergence)?
  Or → Intense long-range contact → convergence (Steppe ‘confederations’).

☒ A diverging ‘Altaic’ family. ☑ A North Eurasian convergence area.
PATTERNS ON DIFFERENT DIMENSIONS: OVERLAPS & CONTRASTS

DIVERGENT LANGUAGE FAMILIES — LINGUISTIC CONVERGENCE AREAS

↑ Güldemann (2010):
“Sprachraum” and geography: linguistic macro-areas in Africa
LANGUAGE STRUCTURES AND THE HOLY GRAIL

• ‘Ultra-stable’ structures / parameters → reveal deepest families, prehistory?

↓ Phylogeny of Austronesian

Gray et al. (2009) ↓

↑ Structural isoglosses within Austronesian

Donohue & Denham (2010)
WHEN STRUCTURES ARE MORE STABLE THAN FAMILIES.

MASS LANGUAGE SHIFT

The same deep structural features:

• Resistant to internal change:
  = Genealogically most stable.
  — So long as transmission is normal ...

• Resistant even through language shift:
  → Carried over into new language:
  = Genealogically least stable.
  = ‘Stable’ in speaker population, even when they switch genealogy.

• Features so stable structurally that they are .... unstable ‘genealogically’...
  → Less diagnostic of deep genealogy than ‘Austronesian’ lexis!
WELCOME TO THE ‘NEW LINGUISTICS’

1. New databases (‘GlottoBank’):
   – World-scale, fullest coverage.
   – Specifically for quantitative uses.

2. New ‘number-crunching’ models and analyses:
   – Constantly refined to get closer to modelling how languages behave.

3. New cross-disciplinary scope and co-operation:
   – Ancient DNA, archaeological science…
Some papers on some of these themes:

http://eva-mpg.academia.edu/PaulHeggarty

Paul.Heggarty@gmail.com
REFERENCES


www.cambridge.org/ec/academic/subjects/archaeology/prehistory/cambridge-world-prehistory


http://dx.doi.org/10.2307/415281