‘New mathematical methods’ in linguistics constitute the greatest intellectual fraud in the discipline since Chomsky

Nijmegen, 8th May, 2015

MPI

Roger Blench

Chief Research Officer, Kay Williamson Educational Foundation
McDonald Institute for Archaeological Research, Cambridge
Visiting Professor, University of New England, Armidale
Academic Visitor, Rajiv Gandhi University, Arunachal Pradesh

What's the issue?

- Historical linguistics, and increasingly other areas of linguistics, have been convulsed by the application of 'new mathematical methods' based on Bayesian phylogenies
- These papers are usually published in 'hard science' journals rather than linguistics outlets
- And indeed they claim to put historical and phylogenetic linguistics on a scientific basis
- Given that the numbers of papers in this area are increasing and making even broader claims it is worth asking
- What is the value of the results thus obtained?
- How can we test their validity?
- Is it an improvement on previous methods, and if so in what way?

Is this science?

- For a method or disciplinary procedure to be deemed scientific it seems it should meet some minimum criteria, including;
  - a) reproducibility. Other researchers should reach comparable results using analogous datasets and/or methods
  - b) falsifiability. It should be possible to clearly state the conditions under which the conclusions would be falsified or shown to be erroneous
  - c) It should able to explain why alternative hypotheses can be excluded
- It is relatively easy to show that on present showing none of these conditions are, or possibly can be met. If this is so, then the editor of Science has presumably been bamboozled

If it appears in ‘Science’ then apparently …

- …helped crack open two areas—animal cognition and historical linguistics—long regarded by many as black boxes, impenetrable to the scientific method. Because languages change at unpredictable rates, analyzing their relationships was for many linguists more of an art than a science. But by applying evolutionary methods borrowed from genetics, Gray and his colleagues are transforming the discipline, shaping it into a science of prehistory. His group has unraveled the histories of the Austronesian and Proto-Indo-European languages and peoples, and traced their migrations over vast distances.

- Science, 19 September 2014
  Virginia Morell

A bit of background I

- Mathematical methods in language classification go back at least to Dumont d'Urville in the 1830s
- But they really take off with lexicostatistics and later glottochronology with Swadesh
- Their 'era' was the 1950s to the early 1990s, after which the accumulated objections of the scholarly community overtook the idea and it went underground for a while
- Most of us concluded that lexicostatistics was good for a 'first look' at relatedness but was not going to tell us anything complex or unexpected
- Despite being championed by Greenberg in his 1987 book and the 'new Russian' school
- However, by the mid-1990s, with new computing power available, access to more sophisticated statistics, the first mathematical phylogenies began to appear
- In some ways, these were a solution looking a problem. They were not introduced to resolve existing problems in historical linguistics

A bit of background II

- These techniques were initially associated with, above all, the New Zealand school.
- The authors sought to publish in high-prestige science journals and the papers were burdened with extensive mathematical apparatus
- Which apparently convinced the editors to send the papers to statisticians not linguists, for refereeing
- Academic publishing is an extreme Thatcherite free market, and one interpretation is that these papers had found a niche to pass under the radar
- They were never going to get the sort of critique offered by conventional historical comparativists
- High-ranked papers lead to better jobs and more grants, so why not?
- The recommendation that we do more of the same old stuff is hardly exciting to the next generation
- And, let us face it, Indo-European in particular, is locked into tedious circular debates with apparently little progress

A bit of background III

- So these new approaches, with their spectacular graphics, seemed to be something fresh
- And the authors were careful to insist this was ‘not lexicostatistics’ (although it was based on cognate judgments)
- So many of the usual rules about reading the literature and explaining why these new models trumped opposing models were discarded in the excitement
- But the methods and results should be subjected to the same critical gaze we would give to other similar claims
- And this is where the parallel with Chomsky comes in, as indeed, it is impossible to falsify any result from generative analysis

Why classify languages? I

- A rather fundamental question is, why classify languages at all?
- Language classification has a long history, which, not atypically for academic discourse, often obscures its purpose.
- Possibly the earliest clear statement of language relatedness is the 10th century *Risāla* of Yehuda Ibn Quraysh, who compared the phonology and morphology of Hebrew, Aramaic and Arabic.
- From the sixteenth century, when large catalogues of the languages of the world begin, attempts accelerated, using a quasi-genetic framework, although often without explicit justification.
- What, however, was the point of such classifications? Why not just list languages alphabetically, or by region?
- Classification is something that particularly appeals to middle-aged white males, and can be of the same genre as categorising tracks on an iPod or knowing an unsettling amount about train timetables.
- Often, as in the biological sciences, justifications for classification have followed significantly later than the exercises themselves.

Why classify languages? II

- Classification of languages feeds into two debates, historical and typological. Historical linguists need to know what is related to what to create a narrative about the past, whether reconstructing individual words, or trying to unravel a pattern of diversification.
- Many historical linguists wish to correlate their findings with other disciplines, such as archaeology and genetics, although this is not essential to the endeavour.
- Typologists, in a sense, want the reverse; to show that languages have generalisable features which cannot be confined within classificatory cages.
- Many of the modernist papers do not really have much to say about the purpose of classification, except where they continue and link it to dates, geography or human genetics.
- But this seems to be a fundamental procedural error; how you go about classifying languages should be intimately linked with the meaningfulness of result you wish to achieve.

Why classify languages? III

- As problems with simple tree models multiplied, historical linguists were increasingly burdened with ever more complex graphic representations of their ideas.
- This inevitably reflected a more nuanced way of doing historical linguistics, which was linked to the intended output, to show language history in a rich way, positing connections between related languages but incorporating histories of borrowing, analogical restructuring and similar features.
- Such methods are by definition unresponsive to simple mathematical or statistical treatment because they invoke cultural specificity.
- Unfortunately, however exciting this type of monograph was to specialists of a phyletic group or an area, it looked rather like a sea of mud to those gazing in from the outside.
- To geneticists or archaeologists, the *Stammbaum* remained attractive, since it was visually similar to their own graphics.
- Numerical calculations were more familiar than lengthy and often inconclusive discussions of cognacy versus borrowing.

Some assumptions

- Linguistic trees produced by Bayesian phylogenetics have some underlying assumptions..
- Cognacy is usually applied to lexical cognacy, the judgment that two lexemes are historically related. In principle, however, it can apply to other areas of linguistics, such as phonology, morphology, or paradigms, such as pronominal sets.
- This in turn makes significant covert assumptions about the nature of the procedure, namely that;
  - a) cognacy judgments are ‘correct’, that consensus can be reached about their accuracy
  - b) that the proposed cognates are genuinely inherited from a proto-language and can thus be reliably distinguished from borrowings
  - c) that language is an autonomous system, resembling more a physical than a cultural system
  - d) and that relationships can be expressed as a series of binary splits
- Moreover, trees as presently published have ‘bare’ nodes, that is, there is no actual evidence to support the node such as isoglosses or phonological shifts.

A crucial wrong assumption

- The most important erroneous assumption is a) the fixity of cognate judgments
- Cognate judgments vary between authors, between language phyla and over time.
- Indeed it could be argued that nuancing cognate judgments is one of the most important tasks of historical linguistics.
- Anyone who has done long-term work on compiling cognates for a phylum knows that you often go back and reverse judgements, and indeed some of your former judgments seem mysterious.
- But more generally cognates are the product of preconceptions.
- Consider the example of Nilo-Saharan. Opinions about the validity of this phylum run all the way from Ehret (1600 starred forms) through Bender (107 proto-forms) to the Glottolog, which does not accept the reality of the phylum (and thus assumes all the claimed proto-forms are chance or lookalikes
- The point is not the judgment to be made about correctness, only that with this level of variability, this is not a transparent, repeatable scientific process

‘I always find more cognates after a good lunch’ – the late Hal Fleming

- This should of course not be true, but it is. When cognacy judgments are disputable, whether you judge two forms to be cognate is in part reflective of your mood, the weather and so on

Anomalies versus norms I

- Languages can be classified according to general principles, here called norms and anomalies.
- Norms are where a language is supposed to have a standard feature which can be compared to the same in another language. Most commonly this consists of a wordlist of ‘basic’ vocabulary items, quite often the 100-word Swadesh list, although many variants have been proposed
- Historically speaking, the use of anomalies or idiosyncrasies much precedes the use of norms. It was early observed that the appearance of irregular paradigms across several languages, such as the comparative series ‘good, better, best’, was unlikely to be due to chance or borrowing, and thus such anomalies were more reliable indicators of the relatedness of languages.
- However, anomalies have two problems. They rely on an in-depth knowledge of the languages under discussion, something which is not available for many languages of the world. Secondly, they do not produce any index or degree of relatedness.

Anomalies versus norms II

- It is therefore unsurprising that mathematical methods have focused on norms, as they produce comparable results across different language families and the type of visuals attractive to journals.
- The link with chronology is also persuasive; glottochronology was once regarded as an exploded hypothesis but has made its return in a variety of new publications.
- The catch is that comparison of norms appears to be the subject of a large number of objections, many of which were raised when lexicostatistics was first propounded. These need not be rehearsed in detail, but can be summarised as follows;

Anomalies versus norms III

a) The assumption that all languages change at a regular rate is unproven at best, and there is a significant body of evidence to suggest it is false

b) The assumption that cognacy based on inheritance can be reliably distinguished from borrowing among closely related languages

c) The assumption that phonaesthetic processes do not act to make concepts such as body parts phonologically and structurally similar in ways that bypass inherited patterns

d) Regionalisms. In some regions of the world, such as Australia and the Amazon, there are lexical items found in unrelated language families which retain a common phonological shape. We have no idea why this occurs and the items themselves are different, but they must be excluded when comparing languages

Reproducibility and methodological opacity

- In usual science results are tested by the ability of others to reproduce the results under identical conditions
- This isn't always the case, cosmologists clearly have to deal with one-off events
- Obviously if we use exactly the same dataset and algorithms, then reproducibility is trivial
- But if each experiment uses a variant on the algorithm, then the problem is methodological opacity, namely an inability to explain why one result is 'better' than another
- Linguistic data should be fairly stable, though we can experiment with different subsets, such as pronouns or grammatical morphemes
- Usually these produce 'different' results; obviously this is the case with conventional historical linguistics
- But in that case we need to explain and possibly discount certain types of data

Reproducibility and methodological opacity

- For example, we can easily show that certain words in Arawakan are amazingly stable across what is otherwise a highly diverse family
- ‘water’, ‘tongue’, first and second person pronouns. These are effectively useless for subgrouping purposes
- But the analyses keep producing different results, different trees, predictions about homelands
- Because authors use slightly different datasets, pick up cognacy judgments from prior researchers or use their own, consider grammatical data or only use lexical data, use variations on the methods of calculation etc.
- The study of Tupian by Chousou-Polydouri (2014) indeed reproduces a series of trees by previous authors which are distinct from her results

A comparison of Tupian trees

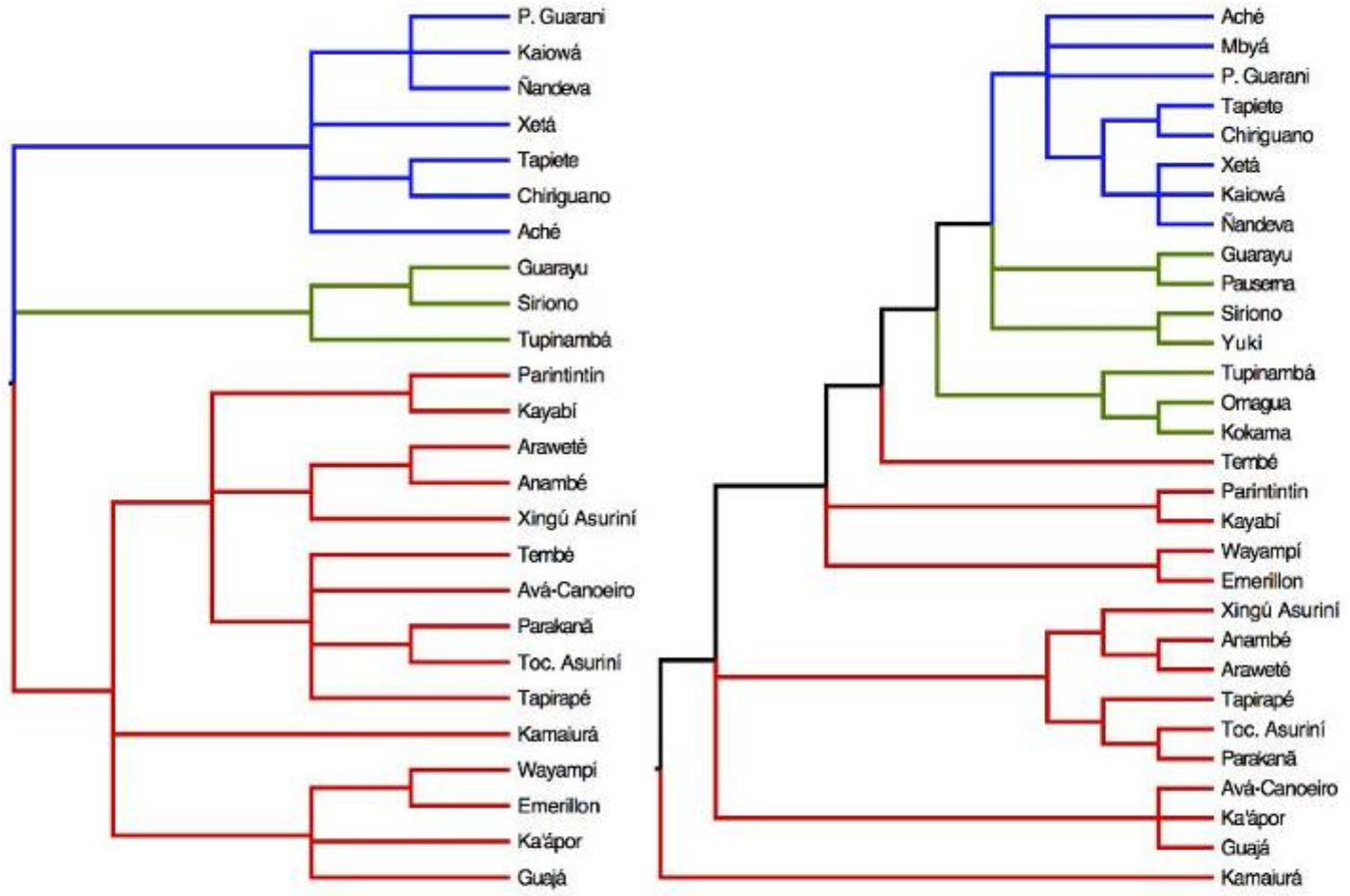


Figure 3.10: Comparison of higher structure between our classification (on the right) and Rodrigues and Cabral (2002) (on the left). Corresponding areas of the trees are colored the same.

Methodological opacity

- And it is completely opaque why one should be preferred to another
- Because the results are not tested against results achieved by other methods, and because there is no decision-making procedure to decide in the case of a mismatch
- The same is even more true of Bantu, which has been the subject of lexicostatistical and then Bayesian analysis since Henrici (1973) and perhaps before
- Other versions include Piron (1997), Bastin, Coupez & Mann (1999), Holden & Gray (2006), Grollemund (2012 and fc.)
- Most recently a strikingly amateur attempt by Currie et al. (2013) which betrays a remarkable lack of familiarity with the basic literature
- Bantu in particular has been subject to extensive internal borrowing, and language levelling episodes

Falsifiability

- Reproducibility and falsifiability in this case are two sides of the same coin. If successive papers do not reproduce the results and this is acceptable, then by definition they cannot be falsified
- Is this just a case that science moves on? As our methods improve of course we get different results
- Well, no, because how can we tell that one result is better than another? In this case there is no heuristic, no instrumentalist method of judging
- Except that whether it matches the results from other methods?
- In which case there was no point in doing it in the first place

How have these developments impacted conventional historical linguistics?

- Have specialists in key phyla, Indo-European, Austronesian and Bantu been galvanised by these new insights?
- Well, no. Generally speaking the ‘findings’ have been completely ignored. Ironically, Bob Blust, who supplied the data and cognacy judgments for the early trees of Austronesian has consistently subsequently ignored the results in favour of conventional historical comparative methods.
- (See his Samalic and ‘Macro-Philippines’ hypotheses)
- Bantuists have similarly blanked out on the various (and of course contradictory trees) presented by a range of authors
- Indo-European specialists have begun to object..

Developing expertise

- The cognacy judgments on which the calculations depend are made by scholars with a wide variation in expertise
- This is important, as there is presumably a general relationship between the expertise of a scholar in a particular phylum and their cognacy judgments
- Experts are more likely to detect incorporated morphology, loanwords between already related languages, and issues with misleading transcription.
- Amateur compilations don’t avoid these and indeed sometimes incorporate error-checking mechanisms to get around the problem of undetected borrowing. [These may in future be useful to police in discerning undetected murders]
- This seems strangely unscientific, since other types of science do not usually incorporate the assumption that experiments have been conducted by amateurs

The problem of coherent and fragmented phyla

- Among the world's language phyla, there is a spectrum between those which are disputed and those which are generally accepted
- At one end is Austronesian or Mayan, whose cognate sets can be easily established from Taiwan to Aoterea, as it were, and Nilo-Saharan or Altaic, where the sceptics think all the proposed cognate sets are lookalikes as these are not phyla
- Typically, the advocates of the new mathematical methods have worked with the more transparent phyla, where scholars largely agree. Clearly where there are major disagreements between professional linguists the data must be hard to process and most importantly, you must follow the view of an individual
- If you have to cherry-pick your phylum, this does not seem to be a very scientific procedure

The problem of rooted trees I

- There is a major question of why we should want to classify languages at all but presumably one reason is to know whether a particular language is part of a phylum or a subgroup
- Unfortunately, this exactly the one thing mathematical methods cannot achieve. Trees must be ‘rooted’, i.e. they start from the assumption that the languages in the dataset are related
- Usually an outgroup language or languages is included, and its remoteness on the tree is an indicator of its non-relatedness
- But it is easy to see where this can lead to a wrong result. Cham languages in Nigeria have borrowed heavily fundamental lexicon from neighbouring unrelated Chadic languages [Tangale] so much so that lexical counts make them closer to Chadic than Adamawa.
- Usual historical linguistics, by looking at morphology, can easily decode this, but ‘blind’ judgment will simply be mistaken

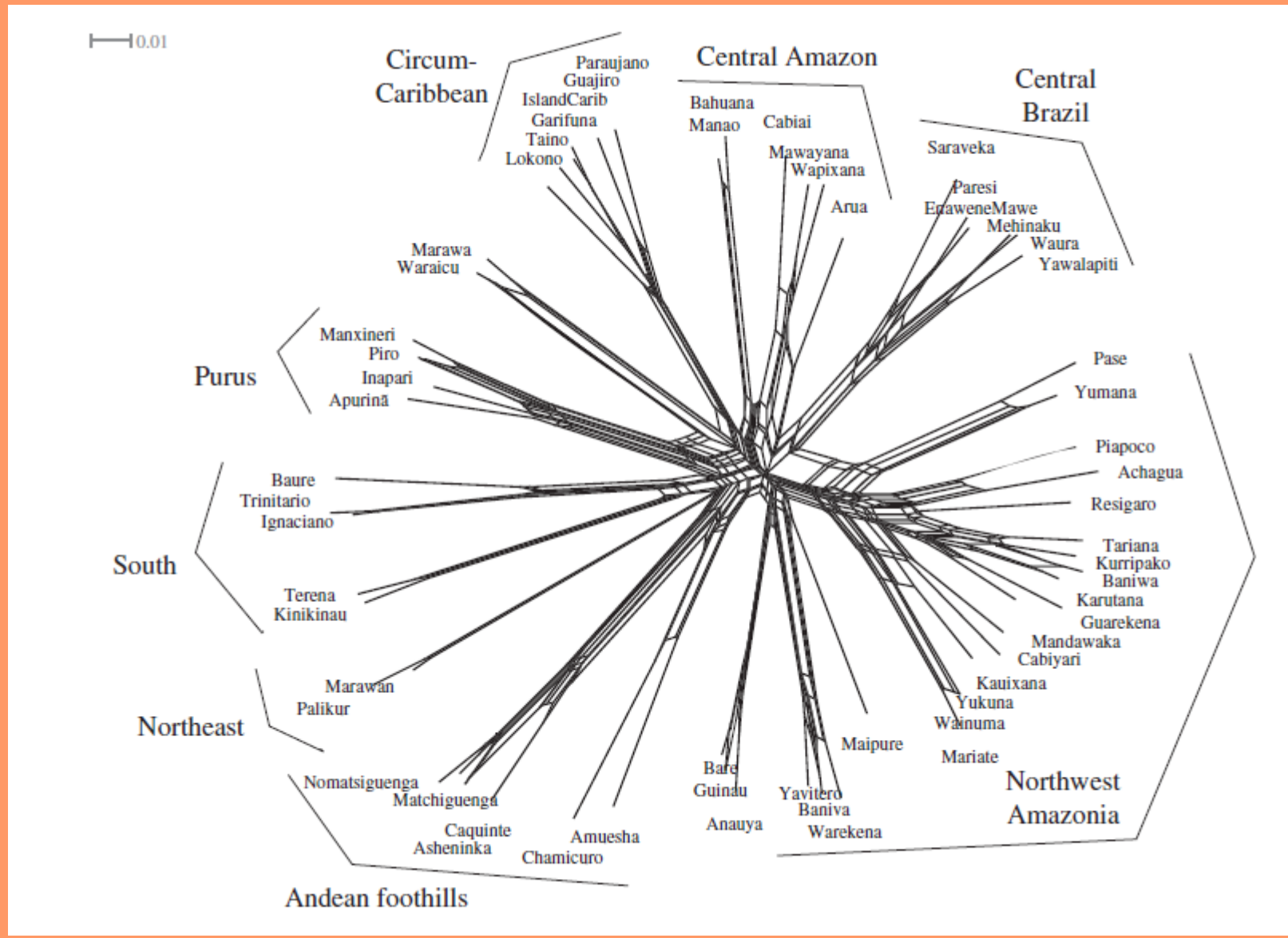
The problem of rooted trees II

- Rooted trees cannot easily test ‘unrooted’ hypotheses.
- For example, Austronesian outliers [Kokota, Utupuan etc.] in the Solomon Islands show extremely low lexical cognacy with Oceanic, barely above chance
- Grammatically, they resemble Oceanic, despite a lack of cognate segmental morphemes
- Despite this, most specialists, present author excepted, consider they ‘must’ be Oceanic, for partly non-linguistic reasons
- Bayesian phylogenies have no method to resolve this issue, because they cannot consider the possibility that these languages are unrelated but have fundamental lexicon borrowings
- So, when there is something we really would like to know about relatedness, the methods prove useless

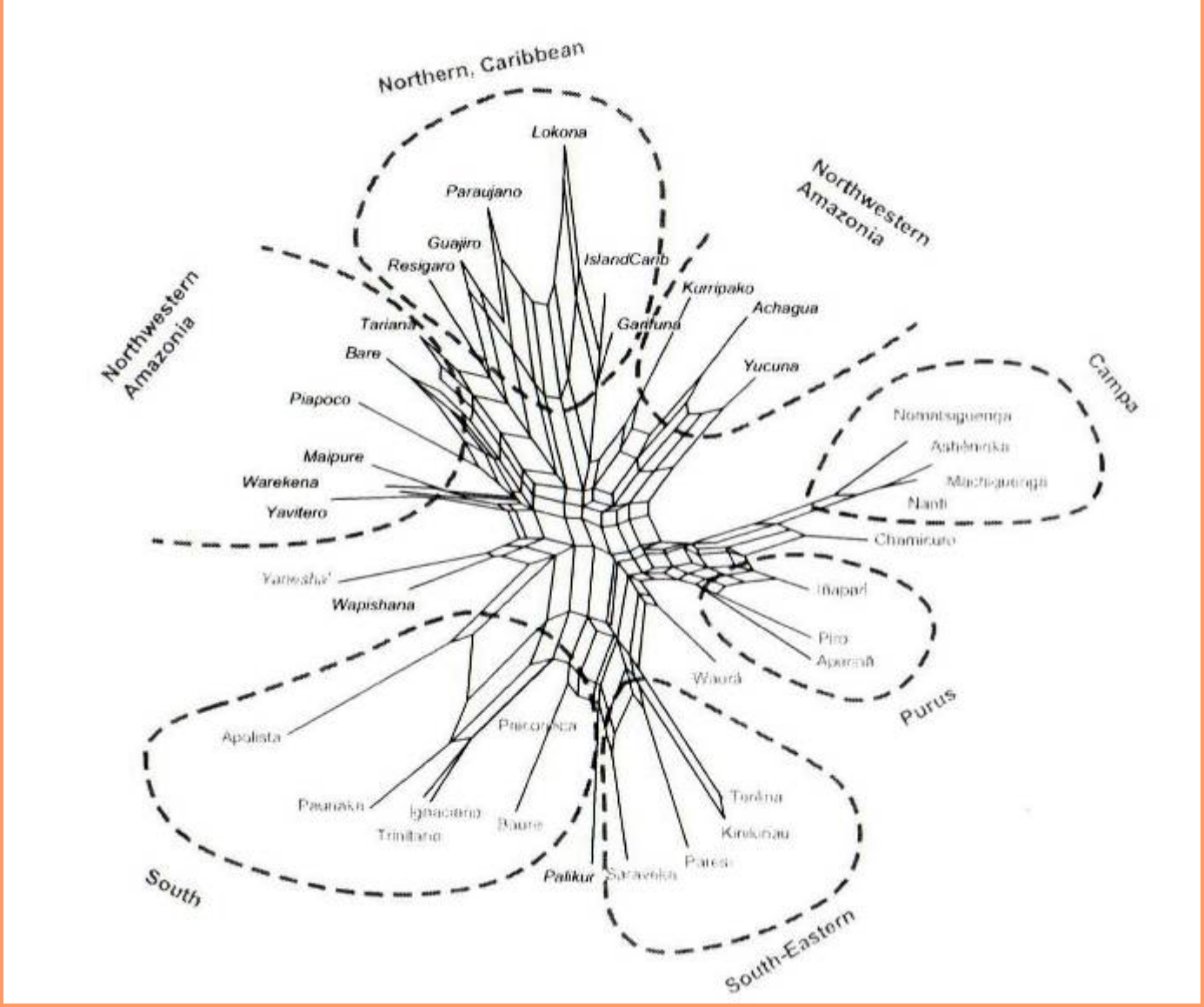
Actual problems: the Arawakan database

- If you are working on trying to understand a large and complex phylum and trying to create a large-scale database, then numerous problems arise that can only be resolved through critical scholarship
- The example here is Arawakan languages, a phylum of some 76 languages (50+ living, another 25 dead)
- A phylogeny of Arawakan has been presented in Walker and Ribeiro (2011) represented as a ‘neighbour net’ starburst
- They say ‘Estimates of the Arawak homeland exclude Northwest Amazonia and are bi-modal, with one potential homeland on the Atlantic seaboard and another more likely origin in Western Amazonia’
- No consideration of the type of integration with archaeological and other evidence in Eriksen and Danielsen (2014), which also introduces *ad hoc* explanations from conventional historical linguistics to explain exceptionalist results

Arawakan neighbour net (2011)



Arawakan personal pronoun starburst (2014)



Actual problems: the Arawakan database

- A nexus of actual problems that have been encountered

- Completely different lexemes cited by different authors, even in basic vocabulary. Baure is an egregious example where only 80% of lexemes are common.
- Many languages have doublets even for basic vocabulary; Do we cite the one which is cognate with Arawakan (?tempting) Do we use the source we have reason to believe is more reliable?
- Language change over time makes the language look more Arawakan; e.g. Resigaro. This is probably to be explained through knowledge of the context of elicitation
- Incorporated morphology in citations forms seems to have confused cognacy judgments

Actual problems: the Arawakan database

- Variable rates of borrowing detection; cf. comments in Eriksen & Danielsen (2014)
- If we had perfect information about frequency of utterance, context of elicitation and so on, no doubt we could resolve these
- But in the real world, typical sources give us no means to resolve these and we need to make hasty decisions for coding

Borrowing versus levelling

- One of the practical objections raised to these types of models is the difficulties of accounting for borrowing/copying and language levelling
- Even in a well-known phylum such as Austronesian, scholars such as Bob Blust and Laurie Reid have disagreed about the amount of levelling in the Philippines
- You can model a ‘realistic’ amount of borrowing (30%) and assume there is no borrowing within closely related branches (?realistic)

PROCEEDINGS OF THE ROYAL SOCIETY B

Proc. R. Soc. B (2009) **276**, 2299–2306
doi:10.1098/rspb.2008.1944
Published online 18 March 2009

Does horizontal transmission invalidate cultural phylogenies?

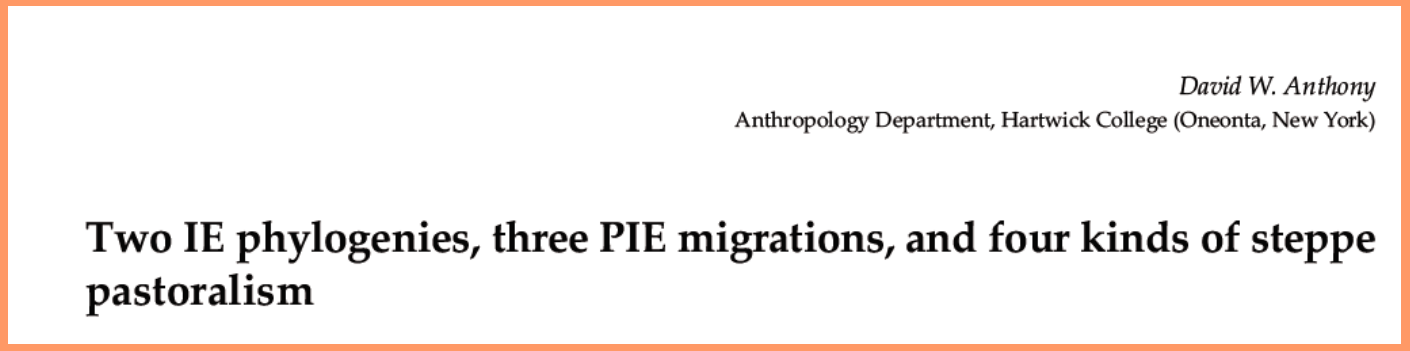
Simon J. Greenhill[1,*], Thomas E. Currie[2] and Russell D. Gray[1]

Borrowing versus levelling II

- and ‘prove’ that it does not invalidate your ideas, but this is to miss the point in a radical way
- The point being that even dense scholarship cannot always unpick borrowing uncontroversially
- How much less amateur cognacy judgments
- The general pattern in recent times has been to acknowledge that borrowing is far more common than has formerly been acknowledged
- Linguists who have tried to use these procedures on actual language families (for example, Atlantic) have found the trees to be far from robust
- It seems that a few changes in assumptions about borrowing change the tree markedly
- Robustness can only be shown on real language families, and incorporate real debates about cognacy and borrowing

Fightback

- Most historical linguists have frankly ignored these productions, probably hoping they'll go away
- And anyway, no high-profile journal is going to publish a paper saying we should simply do more of what we usually do
- However, Indo-European specialists have been more annoyed than most and there have been a series of papers by well-known scholars objecting to the results,
- Jim Mallory, David Anthony and others have pointed to simple errors of scholarship, as well as a failure to take into account alternative narratives



David W. Anthony
Anthropology Department, Hartwick College (Oneonta, New York)

Two IE phylogenies, three PIE migrations, and four kinds of steppe pastoralism

Fightback

J. P. Mallory
Queen's University (Belfast)

Twenty-first century clouds over Indo-European homelands

- As Mallory says 'If there are any lessons to be learned, it is that every model of Indo-European origins can be found to reveal serious deficiencies as we increase our scrutiny. One is reminded of Daniel Kahneman's observation:
- "It is the consistency of the information that matters for a good story, not its completeness. Indeed, you will often find that knowing little makes it easier to fit everything you know into a coherent pattern" (Kahneman 2011, 87).'

OK, we can fix these

- The phylogenist will presumably argue that these can all be fixed
- And indeed they can, with enough careful historical scholarship, awareness of the conventions of the sources and the likely reliability of individual scholars
- The ‘new Russian’ school of lexicostatistics is attempting to reinstate the method through precisely this type of fine-tooth comb method
- But it cannot be fixed by tinkering with the mathematics, because this isn’t the problem in the first place
- But then the problem becomes; what have we gained? If everything has to be checked through the usual lens of the fine grained comparative method, all we have is fancy graphics, not new understanding

So..

- It seems that ‘new mathematical methods’ don’t pass any usual tests for science
- Methodological opacity means that it is almost impossible to judge between results, and thus hard to know how to falsify them
- To this extent they strongly resemble the ‘modelling’ papers that make claims about stuff that happened in the past, for which no amount of empirical data (usually dismissed as ‘anecdotal’) is deemed as counter-evidence
- And they do not adequately (or at all) explain contrary results arrived at by the comparative method and ‘*Worte und Sachen*’
- If the new mathematical methods don’t pass any test for science is this a problem?
- The historical-comparative method definitely has non-science aspects but is still able to advance knowledge

So..

- Building on past findings in a pyramidal fashion
- So perhaps the outputs are still useful in assisting our thinking. But the evidence is weak that historical linguists make any use of the findings
- Some linguists do think it is interesting, and has helped them think about the classification of the languages they study
- But this reduces the value to the enthusiasms of individuals
- And linguistics has a great history of rapidly adopted and discarded approaches
- A science-like approach shouldn’t depend on the aesthetic preferences of individuals, it should be aiming for universal relevance

Dealing with some objections

- Some of the objections raised against the arguments in this talk were;
- It is negative about the use of mathematics. Not so. Mathematics is a tool like any other, but it is poorly adapted to the sorts of phenomena found in language diversification because of the importance of anomalous structures. It is possible to envisage mathematical treatments that would analyse these.
- The graphic outputs help you visualise the data better. This is really misleading, as it assumes that languages diversify according to arborescent or starburst-like structures, and that they do not change patterns during the course of diversification. Both of these we know to be false
- Therefore, the visualisations, far from being helpful, are actually misleading.

On the positive front

- If we are to develop tools to describe patterns of language diversification, for successful modelling a number of outputs are required. These include;

  a) A method of deciding whether languages are related, which can identify isolates or distinct phyla. An aspect of this is the potential to conclude that relatedness is undecidable.

  b) A method for ascribing patterns of language diversification to basic modes, with their subtypes, and including the possibility that the pattern flips partway through the diversification process

  c) A method for identifying iconic regional lexicon or morphosyntax and excluding it

  d) A method for identifying, tabulating and weighting anomalies, both in the process of language classification and establishing inter-branch borrowing

  e) A method for weighting different hypotheses about levels of borrowing and thus their impact on language diversification models

  f) A model which can incorporate diversification modes ‘flipping’ with the process of language expansion

So can anything be salvaged from this?

er, No

and Chomsky?

- No offence to Chomsky, I’m a huge admirer of his politics

THANKS

- To the many people who have unwittingly contributed to this talk