Introduction

§1. Three important issues facing Australian historical linguistics:

1. exceptionality
2. areality (and language contact, and their effects on reconstruction)
3. reconstruction methods

Today: State of the art and future directions in Australian prehistory; a look at the issues and some promising methods.

§2. Outline:

• a brief overview of Australian languages
• the basis of claims of exceptionality in Australia
• the contribution of areality and language contact and their effects on reconstruction
• a case study of computational work in conjunction with the comparative method.

§3. Language Background: ‘Australian’ as a family

• 28 Phyllic families
• c. 250 languages
• greatest diversity in North
• small and large families

Pama-Nyungan

• 90% of Australia’s land mass
• c. 150 languages
• > 20 subgroups
• many singleton (unclassified) languages

§4. Basis for classification

• Curr 1886: common lexical items
• Schmidt (1919): common lexical items, pronouns: ‘southern’ and ‘northern’ languages (but it’s unclear whether he means genetic relationships)
• O’Grady et al. 1966a,b: Lexicostatistics + informal consideration of morphology (e.g. Yawuru and Karajarri): note that all subsequent classifications are based on these works.
• Dixon 2002: Criteria unknown
• Bowern and Koch 2004: Comparative method for assorted subgroups of Pama-Nyungan and Northern families.

Therefore: most of the (sub)groups haven’t been explicitly justified with anything more than a basic vocabulary list. This is work for the future.
Exceptionality

§5. The origins of ‘exceptional’ claims about Australian languages

- nonconfigurational (e.g. Warlpiri; Hale 1983)
- syntactically ergative (e.g. Dyirbal; Dixon 1972)
- VC syllable structure (e.g. Arrernte; Breen and Pensalfini 1999)
- unreconstructible or otherwise historically difficult (Pama-Nyungan; Dixon 2001, 2002)

That is, they exhibit structures which are different from languages elsewhere in the world (no surprise – that’s true for all languages!)

§6. Specific historical claims:

- Australian languages have a dynamic which predisposes them to cyclic shift and extensive parallel development. (This is unverifiable and unlikely; requires unscientific assumptions.)
- the families are too old (> 40,000 years of settlement)
- there’s been too much contact
- we can’t get the right data

Data: exceptional?

§7. Data sources

- Linguists (most > 1955)
- Missionaries (much pre-1970)
- Station owners (and other parties without training in linguistics) (19th Century; eastern Australia)

Note: very few native speaker linguists (exceptions include Laura Dixon, Ephraim Bani and Leonora Adidi). It is a continent documented by outsiders.

This effects the amount of data we have for the languages, and it multiplies uncertainties. These all feed back into difficulties in reconstruction (e.g. irregularity in sound change might be data error).

§8. Is the data state exceptional?

- Probably not especially so.
- But it does affect reconstruction.

Age of the family: Pama-Nyungan and Proto-Australian

There are claims (e.g. by Dixon) that Pama-Nyungan is an archaic family whose speakers spread out shortly after the initial colonisation of Australia, and who have been in situ just about ever since. There are variations on this.

§9. Current distribution of languages vs colonisation

- Humans have been in Australia for at least 40,000 years (Mulvaney and Kamminga 1999).
- It is unclear whether there has been more than one migration.
- **If the current distribution is related, Australia is indeed exceptional:**
  - extreme linguistic conservatism. We would need to explain why these languages show the same sorts of similarities that we find in language families an order of magnitude younger. One way to explain this is by a combination of cultural conservatism and extreme language contact.
  - static populations…
  - little affected by climate change
- This is directly contradicted by the archaeological record.
- Sutton, Evans, McConvell, Jones, etc: Pama-Nyungan spread linked to ‘intensification’ (Evans and Jones 1997, McConvell and Evans 1997, Sutton 1990)

Therefore whatever we say about initial colonisation in Australian languages, we don’t need to take it into account when we are thinking about lower level reconstruction. After all, the presence of Neanderthals in Europe is irrelevant to Indo-European and Finno-Ugric.

§10. Is the time depth exceptional?

- Probably not. We don’t know.
- If a version of the McConvell/Evans/Jones/Sutton scenario is true, there’s stuff that needs explaining, but Pama-Nyungan itself isn’t exceptionally old.
Areality and Contact

Claims of areality

$11$. Australia as a linguistic area A ‘hotbed of multilingualism’:

- Heath 1978: linguistic diffusion in Arnhem Land
- Dench 2001: it’s impossible to tell whether a given change is due to diffusion, parallel innovation, or shared innovation.

Claims: people borrow extensively from their neighbours, and they borrow anything.

It’s true that there are many areal phenomena in Australia. In Eastern Arnhem Land, exogamy has led to widespread multilingualism, to the extent that it’s not unusual for older people to be fluent in five or more languages and for linguistically very diverse conversations with extensive codeswitching to take place. It’s also true that there’s areal patterning at the local level for lots of features (Miller 1972 on Western Desert, my work in Arnhem Land, see also Schebeck 2002).

However, we need a much better idea of how geography, area and contact work in Australia before we can say for certain. It’s not the case that ‘anything goes’. If we assume that, we’re back to our circular exceptional assumptions. There’s no particular reason to assume that shared irregularities (e.g. suppletive morphology) are easily borrowed in Australia.

Two things stand out for further consideration: ideas of space, and patterns of multilingualism,

Desiderata for areal studies

$12$. 1. A better model of geography: who was in contact?
   2. A better idea of continent-wide multilingualism: who spoke what language(s)?

$13$. Example 1: not all geography is created equal

- desert sand dunes funnel traffic in particular directions (cf Arabana-Wangkangurru; Hercus 1994)
- river systems are important travel conduits in semi-arid areas
- borrowing is sensitive to factors other than geography (geography is only one factor in variation)

$14$. Example 2: Australia is multilinguistically diverse

- Arnhem Land: community-wide multilingualism
- Lake Eyre Basin: key multilingual ‘translators’; widespread bilingualism
- Western Desert: clan multilectalism but monolingualism
- Dampier Peninsula: asymmetric bi/multilingualism

Summary

$15$. Interim Summary

- There’s interesting stuff going on
- It’s complex
- And it won’t be solved by single methods

$16$. How to sort this all out?

- careful comparison using comparative method
- areal analysis
- quantification of results
- engagement with the meta-arguments (e.g. plausibility)

Now, this is all too complex for a single talk.

Karnic case study

In the time remaining I want to briefly present some results of recent research applying methods from computational biology to linguistic data. There are many issues that arise when formalising a heuristic (I along with many other historical linguists regard the comparative method as a heuristic (or set of problem-solving techniques) and not an algorithm). Work on this is at an early stage and we don’t know all the special issues which arise when using methods developed for different data.

$17$.
§18. Some problems with Karnic data

1. Conflicting subgrouping (6 different proposals over the last 80 years; cf. Breen (2007));
2. Conflicting correspondence sets;
3. Hard to find clear splits between groups (few bunching isoglosses).
4. Hard to define what’s archaic, because of lack of knowledge (at this stage) of higher order groups.

Question: How can we sort out the relative contribution of areal and genetic features?

Data and Method

§19. Karnic Data

- 615 correspondence sets (∝ 'character sets') from 19 Karnic languages/varieties
- lexical items, nouns, verbs, in a variety of semantic fields.
- nominal morphology, verb morphology
- pronouns

§20. Method: NNets

- Coding cognate forms
- Code the correspondence sets for part of speech, semantic field
- Compile networks for subsets of the data
- Not coded for archaisms vs innovation here
- Multistate form-meaning correspondences

§21. Expectations

<table>
<thead>
<tr>
<th>Tree-like evolution</th>
<th>Contact/diffusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>part of speech has no effect on tree</td>
<td>part of speech subsets produce different trees</td>
</tr>
<tr>
<td>semantic field has minimal effect independent of geography</td>
<td>semantic field affects topology skewing correlates with geography</td>
</tr>
</tbody>
</table>

§22. Nominals only
§23. Verbs only

§24. ‘Stable’ vocabulary (body parts, adjectives, verbs)

§25. Good loan candidates (flora, fauna, artifacts)

Conclusions

§26. Conclusions for Karnic

- illustration of representations: neat way of showing network-like behaviour
- What have we been able to infer for Karnic?
  - some areal contribution (cf Reuther data; flora/fauna)
  - all the methods find the lowest level groups
- Evidence for a ‘Northern Karnic’ node is ambiguous
- Most of the data sets don’t find Northern Karnic (it’s there in pronominal data and verbal data)
- Composition of Central Karnic is stable

§27. Further directions

- this was a preliminary illustration
- there are data coding issues (may issues arise when things are formalised)
- starting from a premise of exceptionality compounds problems, and clouds the similarities; starting with assumptions of uniformity doesn’t have to result in a ‘one size fits all’ language model
- We can model different aspects of language structure to assess the contribution of various descent paths (and networks allow that representation); this feeds into mapping of social networks (but it’s not the same as mapping social networks).
References


