1. Introduction

This paper is based on my contribution to the Pama-Nyungan (PN) panel discussion at the Workshop on Reconstruction and Subgrouping in Australian Languages (ICHL 2001). As a member of the panel I was asked to summarise my thoughts on the following question:

Is such an entity (i.e. PN) (or something of similar scope) justifiable according to the methods of traditional historical linguistics, on the basis of current knowledge?

In my original response I argued that PN is not yet an established genetic entity because it has generally been treated as a subgroup of a higher order Australian family, rather than as a family to be determined independently of any conjectured distant genetic connections to other Australian languages. In my view, the validity of Pama-Nyungan as a genetic entity needs to be demonstrated using language family, rather than subgroup, criteria because there is no known larger entity of which these languages could form a branch.

In this paper, I re-present my argument with some background information. The structure is as follows: in section 2 I review traditional methods in historical linguistics; section 3 presents PN and the Australian classification; section 4 is my conclusion.

2. Methods of traditional historical linguistics

The discussion question refers specifically to traditional methods. In presenting my view on the issue of whether PN is an established genetic grouping, I therefore see as my first task describing what these methods are, and what their application involves. The first part is relatively easy. An inspection of historical linguistics textbooks reveals that the only method that everyone accepts as reliable in establishing genetic entities is quite clearly the ‘comparative method’ (CM). Everyone also agrees, I believe, that two types of genetic entity can be determined using this method:

1. one can determine a ‘family’ on the basis of similarities between languages that are interpreted as being 'retentions' from a common ancestor, because they are unlikely to be due to chance, nature or borrowing/diffusion;
2. one can determine a ‘subgroup of a family’ on the basis of similarities that are interpreted as ‘shared innovations’: they are found only in a particular subset of languages of the family, and they are unlikely to have occurred independently in several languages or to be the result of borrowing/diffusion.

1 I would like to thank Alan Dench, Bethwyn Evans, Eva Lindström, Malcolm Ross and the editors for reading this paper. The usual disclaimers apply.
2 By ‘nature’ I mean iconic or universal features of language. For a discussion of natural resemblance see Harrison (2003).
3 The similarities that we interpret as retentions are, of course, not just similarities but strict correspondences that hold between languages.
Subgrouping therefore logically follows the establishment of a family and is determined on the basis of a special type of similarity. Although everyone accepts the CM, there are certain differences in opinion when it comes to its application. One of the most important disagreements, in my view, regards the point at which a family can be considered established, which is dependent on the type of similarities favoured as heuristic evidence. A full review of the literature on this topic is beyond the scope of this paper, but I will briefly summarise three views found in recent publications. In evaluating whether or not PN is an established genetic entity it is important to keep in mind that, although there is a single uncontroversially accepted method, controversy lies in what the method actually entails.

2.1 Paradigmatic, or other individual-identifying evidence, establishes genetic relatedness

Johanna Nichols (1996:41) argues that the classic application of the CM, as presented in Meillet (1965, 1967) for example, has two separate components, the first being the heuristic component and, the second, linguistic comparison, i.e. the work that establishes correspondences and produces reconstructions.

She describes the heuristic component as an assumption of relatedness based on evidence that identifies a unique individual proto-language, that is, features whose “probability of multiple independent occurrence among the world’s languages … [are] so low that for practical purposes … [they] can be regarded as unique and individual” (1996:48). She goes on to demonstrate that the type of evidence that reaches the individual-identifying threshold is usually paradigmatic (i.e. morphosyntactic), and ideally that there should be whole intersecting systems (multidimensional paradigmaticity).

The main aim of the comparative work that follows (the identification of sound correspondences and the subsequent reconstruction of lexical and grammatical elements), she argues, is that of producing the detailed family structure. However, it is quite clear that Nichols expects lexical comparison to provide further evidence of genetic relatedness in the form of cognates displaying regular sound correspondences. In a different publication she writes that observed (paradigmatic) similarities which do not yield systematic lexical confirmation are problematic because the evidence is, in a sense, conflicting. In discussing the proposed grouping of Altaic (Turkic, Mongolian and Tungusic), for which, in her opinion, there appears to be paradigmatic evidence but no lexical cognates, she concludes the following:

…”either the ‘Altaic’ languages are genetically related (in which case they must be quite closely related, for the pronoun similarities are comparable to those that obtain within a single branch of Indo-European such as Germanic or Balto-Slavic; yet the lexical stock proves to offer very few potential cognates, much unlike the situation in Germanic or Balto-Slavic) or they are not (in which case the pronominal similarities are due to chance, which seems highly implausible). (Nichols 1992:4)

At the very low level, genetic relatedness may be ‘self-evident’. That is, there may still be a degree of mutual intelligibility between language varieties, and there may also be oral or written traditions that confirm this. In certain cases, a subgroup of a family may therefore be known as a genetic entity before it is established as a subgroup of that family (see e.g. Nichols 1996 on Slavic).

There are also disagreements about which types of similarities can be used for subgrouping arguments, but I will not go into this here (see e.g. Harrison 2003, Hetzron 1976, Nichols 1996, Fox 1995, Watkins 1966).
2.2 Regular sound correspondences in lexical comparison establish genetic relatedness

Harrison (2003) argues that, since to establish genetic relatedness it must be demonstrated that similarities are unlikely to be due to nature, the comparativist must concentrate on the most symbolic of linguistic signs: individual lexical items. The CM is therefore described by him as a tool designed to operate on lexical data alone, the ‘domain restriction’ being necessary for the exclusion of natural resemblances that are of no use for genetic arguments. The assumption that sound change is regular is then required by the CM to eliminate the possibility that the lexical resemblances (similar forms with similar meanings/interpretations) are due to chance and, to a lesser extent, borrowing/diffusion. That is, the regularity assumption allows the comparativist to establish potential cognates as real cognates: chance resemblances will not produce a regular pattern of sound correspondence and, unless borrowing has occurred on a massive scale, patterns of correspondence will also help to identify loans.

Harrison (2003:229) briefly discusses grammatical oddities, or the type of paradigmatic evidence that linguists have sometimes put forward as evidence of genetic relatedness, and concludes that, although he understands the motivation of the linguists involved, “[t]he principal virtue of the comparative method is just that its logic doesn’t demand that we seek out oddities, but regularities”.

2.3 Genetic relatedness is established when there is more than one type of evidence

Thomason and Kaufman (1988) argue that genetic relatedness cannot be established on one type of evidence alone, whether lexical or paradigmatic, and Campbell (1997, 2003b) also suggests that it is preferable to have different types of evidence in support of a proposal of genetic relatedness.

In his conclusions regarding which methodological principles and procedures are valid to determine family relationship, Campbell (1997:259) writes: “[p]rincipal among these are reliance on regular sound correspondences in basic vocabulary and patterned grammatical evidence involving submerged features or shared aberrancy,…”. That he believes relatedness is best confirmed using a combination of these different types of evidence becomes clear from his careful illustration of the various ways in which even the most reliable features can sometimes have explanations other than genetic relatedness. Most interesting is, perhaps, his discussion of apparent shared aberrancies, or idiosyncratic grammatical correspondences of the type that appear to be individual-identifying, but which, in fact, have non-genetic explanations (Campbell 1997:219-221).

---

6 His lexical evidence includes grammatical morphemes and morphologically complex words.

7 Under Johanna Nichols’ interpretation of the CM, because genetic relatedness is established before the comparative work begins, the similarity condition (i.e. the requirement that potential cognates must be similar in meaning as well as in form) is not part of the operational procedure. Nichols writes, for example, “once relatedness is assumed, then the labour-intensive process of working out the correspondences and cognate sets begins. Since relatedness is assumed, this lexical work makes the further assumption that any vocabulary set displaying the regular sound correspondences is in fact cognate, however far-fetched the semantic correspondences” (1996:41).

8 Campbell prefers to approach questions of genetic relatedness in terms of “strength of hypothesis” and “level of confidence warranted in making judgements” rather than in terms of a strict dichotomy of established/unestablished. He explains: “A proposal may present evidence that is sufficient to attain a certain level of plausibility but not sufficient to eliminate all doubt” (Campbell 1997:261); therefore, “it is more accurate to view unconfirmed proposals of family relationship as falling somewhere along a continuum ranging from the highly probable to the very unlikely” (Campbell 1997:260-261).
Thomason and Kaufman’s argument for why evidence from the various linguistic levels is needed stems from a different definition of genetic relatedness. They argue that genetic relatedness should be defined socially, rather than linguistically, as what results from ‘normal linguistic transmission’. They describe this as the process by which a language is:

“…passed on from parent generation to child generation and/or via peer group from immediately older to immediately younger, with relatively small degrees of change over the short run…”

(emphasis theirs, Thomason and Kaufman 1988:9-10)

Languages arising by means other than normal linguistic transmission, they claim, are not genetically related to any antecedent system. They argue that an interruption in normal linguistic transmission should give rise to “…a lack of correspondence among the various subsystems of the language, most probably between the lexicon as a whole and the grammar as a whole” (emphasis theirs) (Thomason and Kaufman 1988:11). In their view, the requirement of evidence from different subsystems is therefore necessary for the identification of languages that lie outside the dimension of genetic relatedness, and to distinguish them from those that have undergone contact-induced change over the course of several generations.

3. Pama-Nyungan and the Australian classification

In 1966, O’Grady, Voegelin and Voegelin (OVV) published a lexicostatistical classification of Australian languages (see Koch this volume, Chapter 2 for details) and it was here that the genetic unit of PN was first proposed, as one of the 29 phylic families within the Australian macro-phylum. Lexicostatistics has been used as a starting point in the classification of other language families, such as Austronesian. However, since the method has been rejected by the vast majority of linguists because they consider it to be unreliable, any groupings identified this way are to be taken as hypothetical groupings only, to be confirmed by the CM.

The internal classification of Austronesian, for example, has changed dramatically from Dyen’s (1965) lexicostatistical schema. On the contrary, although the OVV classification has been modified, the Australian lexicostatistical groupings have generally not been ‘reestablished’ using the CM; this was of course the aim of the workshop on which this volume is based. Rather, some of OVV’s groupings have been ‘defended’ through the identification of phonological and morphosyntactic features, whose distribution more or less coincides with the lexicostatistical groupings, and certain languages have been included or excluded from original groupings, on the basis of such features.

9 For a rather different opinion see Croft (2000:chapter 8).

10 Much of the data collection and results were actually the work of Ken Hale, who is credited with the discovery of Pama-Nyungan and who coined the name for this genetic entity using the words for “person” in the languages found at the two extremes of the family’s territory: pama in the north-east and nyungar in the south west. Stephen Wurm also collaborated on the project and provided much of the data for Eastern languages (see O’Grady and Hale, this volume).

11 For a discussion of the problematic nature of all the assumptions at the basis of lexicostatistics, see Campbell (1998:177-186).
3.1 Defending Pama-Nyungan

O’Grady, Voegelin and Voegelin write that the low cognate density present at their macro-phylum level is suggestive of only a very remote relationship and that it makes the reconstruction of all, or even most, of the phonemic contrasts of the parental phylum rather difficult. They seem to be suggesting that it may not be possible to establish the Australian macro-phylum through application of traditional methods and, therefore, that their family level groupings are the highest level for which the CM may bring confirmation (OVV 1966:15-16). Following this, it might be expected that PN, one of their family level groupings, would have been treated as a family, and not a subgroup, in comparative work that followed; that its validity would be confirmed on the basis of ‘correspondences’, not ‘innovations’.

Dixon (1970b, 1980), who understandably rejected PN as an established genetic grouping because it was the product of lexicostatistics, nevertheless based his critiques on the idea of ‘PN as a branch of a higher-order Australian family’. He pointed out that, although there were many similarities, there appeared to be no innovations unique to this group of languages that established them as a subgroup (see also Dixon 1990). Perhaps in response to this criticism, researchers who believed PN to be justifiable have since dedicated themselves mainly to the task of demonstrating the existence of PN innovations. Two 1988 publications present the features that are still regarded as crucial evidence for PN as a subgroup. Blake (1988) provides a survey and reconstruction of pronoun paradigms. He reconstructs two sets of pronouns, one for the northern languages (often referred to as non-PN), the other for PN languages. Although he does not explicitly state this, he implies that, since there are obvious similarities between the two, they must stem from a common source and that the differences between the two reconstructed sets constitute some of the innovations that justify each branch. For PN, the most important of these features was the reconstructed first person dual form *ngali*.

Evans (1988) presents further evidence of innovations in the identification of a shared sound change, the laminalisation of word-initial apical consonants. Overall, he lists the following set of innovations in support of PN as a subgroup:

- (a) ergative *-ngku/-lu* and locative *-ngka/-la*, with their distinct allomorphy,
- (b) the nominalizer in *-NHDHA*,
- (c) a new pronoun system,
- (d) certain verb inflections, and
- (e) with Gunwingguan, whatever system of verbal inflections gave rise to the conjugation markers. (Evans 1988:94)

Alpher (1990) gives further verbal evidence for PN in the form of reconstructed paradigms.

---

12 The reconstructions in Dixon (1980), attributed to a proto-Australian, have been pointed out as being mostly based on data from Pama-Nyungan languages, and therefore not really ‘proto-Australian’ (see e.g. Blake 1988:8).

13 The set of pronouns reconstructed for Pama-Nyungan languages is, for the most part, in agreement with the Proto-Australian reconstructions in Dixon (1980).

14 Heath (1997) explicitly refers to a proto Pama-Nyungan and a proto non-Pama-Nyungan (Northern).

15 On the basis of the reconstructed pronoun paradigms, certain changes were made to the composition of the Pama-Nyungan ‘subgroup’. The Tangic languages, that on lexical counts had been included in Pama-Nyungan, are now excluded due to the absence of ngali and the existence of distinctively Northern forms in their pronoun systems. Yanyuwa, on the contrary, originally left out of Pama-Nyungan, has been included on the basis of the pronominal evidence and identified as being most closely related to the Warluwarric languages. Karwa and Waanyi, very closely related languages, not included in Pama-Nyungan by O’Grady, Voegelin and Voegelin, appear to have a ‘mixed’ pronoun system with some PN and some Northern forms. Evans clearly excludes them from Pama-Nyungan but sees them as the most closely related languages to the subgroup and, therefore, as sharing some of the innovations associated with Pama-Nyungan languages (see e.g. Evans 1995, Evans and Jones 1997). Dixon (2001) argues that ngali is better understood as a form that has diffused over a large contiguous area. This hypothesis is not discussed further here, this paper being largely based on the original panel presentation. Dixon (2001) appeared in print only after the ICHL conference.

16 See Alpher (this volume) for a critique of the laminalisation innovation.
3.2 The Australian macro-phylum or Australian family

The problem with the interpretation of PN as a subgroup, as already suggested, is that the Australian family, of which it is supposedly a branch, is not established under any of the interpretations of the CM discussed above.

As mentioned in section 3.1, O’Grady, Voegelin and Voegelin themselves pointed out that there are not enough lexical similarities to arrive at regular sound correspondences and to reconstruct a proto-Australian phonological system. The Australian family is therefore not established (or, in my view, possible to establish) under the interpretation of the CM that requires lexical comparison and regular sound correspondences.

Furthermore, although Blake (1988) discusses the similarities between the two pronoun systems reconstructed for Australian languages, he does not reconstruct a proto-Australian pronominal system. He also points out that the other idiosyncratic evidence given in support of PN, the case marking allomorphy, causes problems for comparison with non-PN languages because many northern languages lack these cases, being largely head-marking. There doesn’t appear to be any paradigmatic evidence, or morphosyntactic evidence, that could establish an Australian family under Johanna Nichols’ interpretation of the CM either, let alone the interpretation that requires evidence from different linguistic subsystems.

4. Conclusion

As stated at the beginning of this paper, I maintain that, at present, PN cannot be considered an established genetic entity because the evidence so far proposed in its favour relies too heavily on the assumption that all Australian languages are related.

There is sufficient evidence to suggest that we are highly likely to be dealing with a group of genetically related languages, for example, the recurring set of pronouns and the shared case allomorphy briefly discussed above: further investigation of these features may ultimately demonstrate that these reach Johanna Nichols’ individual-identifying threshold. However, all of the approaches described in section 2 consider regular sound correspondences in the lexicon crucial in establishing/confirming genetic relatedness. Therefore, more detailed comparison and reconstruction, of both phonology and morphology, must be undertaken before PN can be considered to have been established in accordance with the methods of traditional linguistics.

17 Alpher (this volume) is the first presentation of sound correspondences in PN.
Miceli — References


